

Tendencias y características de los viajeros que visitan la ciudad de Pereira por medio de técnicas de minería de datos.

Diego Fernando Salcedo Toro



Universidad Autónoma De Manizales

Maestría en Gestión y Desarrollo de Proyectos de Software

Manizales, Caldas – Colombia

Noviembre 2015

Tendencias y características de los viajeros que visitan la ciudad de Pereira por medio de técnicas de minería de datos.

Diego Fernando Salcedo Toro

Trabajo de Grado para optar al título de Magister en Gestión y Desarrollo de Proyectos de Software

Asesor Temático y Metodológico Msc. Jairo Iván Vélez Bedoya.

Universidad Autónoma De Manizales

Maestría en Gestión y Desarrollo de Proyectos de Software

Manizales 2015

Tabla de contenido

Introducción	5
Referente Contextual	7
Descripción del Área Problemática	7
Antecedentes	11
Formulación del problema	22
Objetivos	23
Objetivo General.....	23
Objetivos Específicos	23
Resultados esperados	24
Estrategia metodológica	25
Metodología	25
Pruebas	28
Presupuesto	29
Desarrollo	30
Referente Teórico	30
Desarrollo del Proceso de Minería de Datos con el modelo CRISP-DM	72
Comprensión del Dominio	72
Comprensión de los datos.....	75
Preparación de los datos	81
Modelado	101
Evaluación.....	191

Despliegue.....	194
Conclusiones	198
Recomendaciones	200
Bibliografía	203
Anexos	220

Lista de figuras

Figura 1. The knowledge destination framework	14
Figura 2. The knowledge destination framework architecture	15
Figura 3. The knowledge destination framework	17
Figura 4. The knowledge destination framework architecture	18
Figura 5. Tareas genéricas y salidas del modelo CRISP-DM.....	28
Figura 6. Arquitectura de Big Data.....	43
Figura 7. Elementos de un BI	44
Figura 8. Ruta de un BI.....	46
Figura 9. Arquitectura básica de un sistema ETL.....	47
Figura 10. Técnicas de Clustering	57
Figura 11. Algoritmo de particionado de K-Means	61
Figura 12. Clustering de un grupo de objetos utilizando K-Means; para (b) unos centros actualizados y una reasignación de objetos (La media de cada clúster está marcada por a+)	61
Figura 13. Proceso Normalizar datos con pesos numéricos.	82
Figura 14. Carga datos a RapidMiner	100
Figura 15. Importación de datos a RapidMiner	100
Figura 16. Proceso RapidMiner para el Clustering.....	103
Figura 17. Filtrado y eliminación de Outliers.....	104
Figura 18. Validar el número óptimo de clústers para los datos proceso 1.	105
Figura 19. Ciclo para validar los clústers.....	105

Figura 20. Validación con medidas internas parte 1.....	106
Figura 21. Validar el número óptimo de clústers proceso 2.	106
Figura 22. Proceso para generar reglas de asociación.	108
Figura 23. Proceso de analisis Descriptivo.....	109
Figura 24 . Datos para analisis descriptivo.	110
Figura 25 . Metadata 5123 Viajeros.....	112
Figura 26. Metadata 2500 Viajeros.....	113
Figura 27 . Metadata 100 Viajeros.....	114
Figura 28. Metadata 3116 Viajeros.....	114
Figura 29. Totales estadisticos de los Viajeros.	115
Figura 30. Histograma Genero-Edad.	119
Figura 31. Histograma Genero-Estado Civil.	120
Figura 32. Histograma Genero-Nivel de Estudios.....	120
Figura 33. Histograma Genero-Ocupación.....	121
Figura 34. Histograma Genero-Fuente de Ingresos.	121
Figura 35.Histograma Genero-Número de visitas a la ciudad.....	122
Figura 36. Histograma Edad-Ocupación.....	123
Figura 37. Histograma Edad-Nivel de Estudios.	123
Figura 38. Histograma Edad-Estado Civil.....	124
Figura 39. Histograma Edad-Fuente de Ingresos.....	124
Figura 40. Clustering validación 1.....	127
Figura 41. Filtrado y eliminación de Outliers proceso 2 modificado.	130
Figura 42. Validación con medidas internas parte 1.....	133

Figura 43. Ciclo para validar los clústers.....	135
Figura 44. Indices de validación de clusters.	136
Figura 45. Distribucion de centroides con k=4 Proceso 3.	137
Figura 46. Validar el número óptimo de clústers para los datos proceso 1.	138
Figura 47. Validar el número óptimo de clústers proceso 2.	138
Figura 48. Plotter del índice Silhouette.....	141
Figura 49. Tabla del indice Silhouette y el número de clusters.	142
Figura 50. Proceso para generar reglas de asociación.	143
Figura 51. Quinto Proceso Clustering.....	144
Figura 52. Gráfico de Desviación de todos los clusters.....	147
Figura 53. Estadística del cluster cero.	148
Figura 54. Reglas de inducción del Cluster cero.	149
Figura 55. Datos de validación de las reglas del clúster cero.	150
Figura 56. Estadística del cluster uno.	155
Figura 57. Reglas de inducción del Cluster uno.	155
Figura 58. Datos de validación de las reglas del clúster uno.	156
Figura 59. Estadística del cluster dos.....	161
Figura 60. Reglas de inducción del Cluster dos.....	162
Figura 61. Datos de validación de las reglas del clúster dos.	163
Figura 62. Estadística del cluster tres.	168
Figura 63. Reglas de inducción del Cluster tres.	168
Figura 64. Datos de validación de las reglas del clúster tres.	169
Figura 65. Distribución de Outliers	174

Figura 66. Proceso para Análisis de la Percepción.....	176
Figura 67. Reglas del Clúster 0 sobre percepción.....	178
Figura 68. Reglas del Clúster 1 sobre percepción.....	178
Figura 69. Reglas del Clúster 2 sobre percepción.....	179
Figura 70. Reglas del Clúster 3 sobre percepción.....	179

Lista de tablas

Tabla 1. Estadísticas de turismo en el triángulo del café a Mayo de 2014	8
Tabla 2. Lista de categorías de productos característicos del turismo y de industrias turísticas ..	31
Tabla 3. Casos de partida para el análisis de clustering.....	53
Tabla 4. Artículos comprados por los clientes.....	66
Tabla 5. Criterios de Evaluación de Algoritmos.....	71
Tabla 6 . Campos de datos personales de los encuestados	76
Tabla 7. Campos de motivos de viaje y clasificación.....	77
Tabla 8. Campos discretizados de la edad	83
Tabla 9. Campos discretizados del número de viajes a la ciudad.....	84
Tabla 10. Campos discretizados de la fecha de viaje.....	84
Tabla 11. Campos discretizados del número de días de permanencia.....	85
Tabla 12. Campos discretizados del número de compañeros de viaje.....	85
Tabla 13. Campos discretizados del grado de satisfacción [Ocio]	86
Tabla 14. Campos discretizados del grado de satisfacción [Oferta cultural].....	86
Tabla 15. Campos discretizados del grado de satisfacción [Conservación Monumentos]	87
Tabla 16. Campos discretizados del grado de satisfacción [Señalización turística].....	87
Tabla 17. Campos discretizados del grado de satisfacción [Cultura de la ciudadanía]	88
Tabla 18. Campos discretizados del grado de satisfacción [Oferta de alojamiento]	88
Tabla 19. Campos discretizados del grado de satisfacción [Calidad en la atención recibida].....	89
Tabla 20. Campos discretizados del grado de satisfacción [Disponibilidad alojamiento]	89

Tabla 21. Campos discretizados del grado de satisfacción [Calidad del alojamiento]	90
Tabla 22. Campos discretizados del grado de satisfacción [Precio del alojamiento]	90
Tabla 23. Campos discretizados del grado de satisfacción [Información oficina turismo]	91
Tabla 24. Campos discretizados del grado de satisfacción [Información en el hotel]	91
Tabla 25. Campos discretizados del grado de satisfacción [Información dada por transeúntes] .	92
Tabla 26. Campos discretizados del grado de satisfacción [Información dada por la policía].....	92
Tabla 27. Campos discretizados del grado de satisfacción [Información dada por la policía de tránsito]	93
Tabla 28. Campos discretizados del grado de satisfacción [Información dada por la policía de carreteras].....	93
Tabla 29. Campos discretizados del grado de satisfacción [Información turística en vallas, carteles, volantes, folletos, guías, etc.].....	94
Tabla 30. Campos discretizados del país de origen por organización territorial	95
Tabla 31. Campos discretizados del gasto estimado.....	96
Tabla 32. Sociodemográfico de Género.....	116
Tabla 33. Sociodemográfico de Edades.....	116
Tabla 34. Sociodemográfico de Estado Civil.	116
Tabla 35. Sociodemográfico de Nivel de Estudios.	117
Tabla 36. Sociodemográfico de Ocupación.....	117
Tabla 37. Sociodemográfico de Fuente de Ingresos.	118
Tabla 38. Sociodemográfico de Ubicación Territorial.	118
Tabla 39. Tabla uno de la percepción del gasto de los viajeros.....	180
Tabla 40. Tabla dos de la percepción del gasto de los viajeros	181

Tabla 41. Tabla tres de la percepción del gasto de los viajeros	182
---	-----

Lista de anexos

Anexo 1. Presupuesto Global Y Discriminado.....	220
Anexo 2. Resultados, productos esperados y potenciales beneficiarios	223
Anexo 3. Cuadro de impactos esperados.....	226
Anexo 4. Modelo inicial de la encuesta	230
Anexo 5. Modelo encuesta aplicada	234
Anexo 6. Tablas de pruebas	246

Resumen

El sector del turismo es uno de los más grandes a nivel mundial, generando 212 millones de empleos contribuyendo al PIB de muchos países.

Según las estadísticas del ministerio de comercio el total de llegadas de turistas internacionales a la ciudad de Pereira es mayor que el de Armenia y Manizales, esto sin tener en cuenta los turistas que ingresaron por una frontera, datos que se han mantenido desde el año 2011 hasta la actualidad, año 2014.

En el Quindío se cuenta con un observatorio de turismo el cual cada vez ha ido incorporando nuevos datos para sus estadísticas, observándose en el año 2014 que se tienen en cuenta a los actores de turismo, como lo son los prestadores de servicios hoteleros, restaurantes, artesanías, servicios de guías.

La ciudad de Pereira no cuenta con un observatorio de turismo ni estudios sobre sus viajeros, las entidades encargadas del Turismo de Pereira no tienen en cuenta dicha variable ya que en sus sitios web y en su entidad no hay información recopilada sobre estos.

Por esto hizo un análisis de la ciudad, para poder determinar datos estadísticos como: Procedencia de los visitantes, propósito de la visita, servicios utilizados, tipos de comidas preferidas, tipo de transporte utilizado, medio por el cual se enteraron de la ciudad, nivel de satisfacción en la ciudad, costos de la ciudad.

Con los datos recopilados se realizó una segmentación y descripción de hábitos de los viajeros realizando análisis de clúster, posteriormente se clasificaron de acuerdo a tendencias y características aplicando reglas de asociación y finalmente se agruparon por el nivel de gastos,

percepción y satisfacción luego de visitar la ciudad por medio de un análisis de clúster.

Este proceso de extracción de conocimiento a partir de los datos KDD, utilizado actualmente por el sector empresarial, educativo y científico, se desarrolló utilizando la metodología CRISP-DM y fue implementado sobre la aplicación de uso libre RapidMiner y comprende desde la recolección de la información en línea y la extracción de información de la base de datos diseñada para tal fin, su transformación, validación, el cálculo de los clústers y las reglas de asociación.

Abstract

The tourism sector is one of the largest worldwide, generating 212 million jobs contributing to the GDP of many countries.

According to statistics from the ministry of commerce, total international tourist arrivals to the city of Pereira is greater than that of Armenia and Manizales , this regardless of tourists who entered by a border , data that has been maintained since 2011 to the present, 2014.

Quindío has a tourism observatory which increasingly has incorporated new data for its statistics observed in 2014 that takes into account the actors of tourism, such as providers of hotel services, restaurants, crafts, guide services.

The city of Pereira does not have a tourism observatory and studies on its rating, entities of Tourism in Pereira do not consider this variable as on their websites and in their entity, there is no information collected about them.

Therefore, an analysis of the city, to determine statistics such as: Origin of visitors, purpose of visit, services used, types of favorite foods, type of transport used, satisfaction level in the city, customer perception about costs in the city.

With the collected data segmentation and description of travel behavior performing cluster analysis was performed subsequently classified according to trends and characteristics applying association rules and finally grouped by the level of expenditure, perception and satisfaction after visiting the city through cluster analysis.

This process of extracting knowledge from the KDD data currently used by the business, educational and scientific sector, was developed using the CRISP -DM methodology and was implemented on the application of free use RapidMiner, process uses collection of information

online and extracting information from the database designed for this purpose, transformation, validation, calculation of clusters and association rules.

Introducción

En la actualidad los datos digitales cada vez ocupan más espacio. Los requerimientos básicos de instalación de un programas necesita de un espacio de almacenamiento elevado y por otro lado su contraparte (El Hardware) ha bajado los costos de las maquinas, discos duros, memorias USB y demás medios de almacenamiento masivo, todo esto también gracias a los servicios en la nube. (IDC, 2014).

En esta época todas las organizaciones manejan gran cantidad de datos, algunos de manera organizada, otros utilizan diferentes sistemas y almacenan datos redundantes, desorganizados y con errores. Las nuevas organizaciones almacenan pocos datos en comparación a las que tienen una mayor trayectoria, las que llevan más tiempo por lo general tienen una gran cantidad de datos históricos. Podemos decir que toda empresa, institución o negocio requiere del almacenamiento de datos para poder realizar informes y tener más claridad de su información y hacer balances de sus gestiones (ComputerWorld, 2014).

Cuando los reportes se quedan en la parte transaccional, se hace evidente un método diferente, unos algoritmos más complejos, una metodología que convierta dichos datos en información más útil (IDC, 2012).

Dichas metodologías presentan el conocimiento implícito y previamente desconocido en conocimiento explícito, conocimiento del cual se pueda derivar un análisis para proyecciones, encontrar patrones en los datos, realizar asociaciones con los datos, agrupaciones de clúster, encontrar tendencias, entre otras operaciones. Estas metodologías se encuentran bajo el nombre de minería de datos, el cual tiene bases muy fuertes de la inteligencia artificial y el análisis

estadístico. Con esas bases se crea una ciencia la cual recoge, modifica y limpia los datos para posteriormente extraer conocimiento(Microsoft, 2014).

En el triángulo del café, (Quindio O. T., 2014) luego de realizar un estudio del sector del turismo se pudo observar que las entidades públicas y privadas de la región tienen dichos datos, algunos completos, algunos incompletos, algunos por sector hotelero, otros por sector de salidas a través del aeropuerto. En este sector se puede hacer un estudio de minería de datos el cual permita encontrar patrones, grupos y realizar predicciones muy útiles para la región, para fomentar y mejorar el turismo.

El estudio parte de la información recopilada en internet mediante encuesta en línea, el total de registros recopilados es de 60 mil ya que en la ciudad de Pereira no hay información estadística sobre los aspectos requeridos.

El desarrollo del trabajo siguió la metodología CRISP-DM, un estándar internacional en proyectos de minería de datos, presentando para cada fase las parametrizaciones y resultados obtenidos (Chapman, 2007).

Se utilizó únicamente software de uso libre, entre ellos MySQL como motor de base de datos para almacenar la Información de los turistas y el software de análisis RapidMiner para implementar el proceso de ETL y el flujo del proceso de minería de datos.

Finalmente se presentan el análisis de resultados y las conclusiones.

Referente Contextual

Descripción del Área Problemática

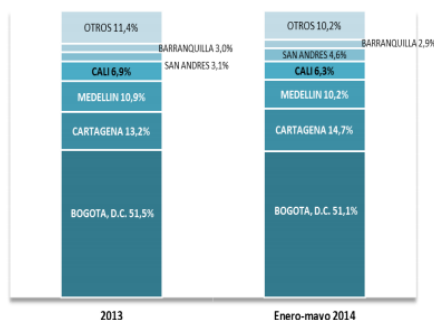
La industria del turismo es una de las más grandes a nivel mundial generando cerca de 212 millones de empleos y representando una importante contribución al producto interno bruto de muchos países año 2011 (Comercio, 2007) y año 2013 (Council, 2012). En una industria de estas dimensiones, la tecnología de información “TI” ha jugado un papel fundamental como medio para mejorar la eficiencia de las organizaciones y la entrega del servicio.

Según las estadísticas del ministerio de comercio del año 2011. (Turismo M. d., 2010) podemos observar el total de llegadas de turistas internacionales a la ciudad de Pereira es mayor que el de Armenia y Manizales, esto sin tener en cuenta los turistas que ingresaron por una frontera.

En el año 2014 (Turismo M. y., 2014) se mantiene dicha tendencia, por lo cual es clara la necesidad de contar con datos estadísticos de Pereira y sus visitantes.

Tabla 1. Estadísticas de turismo en el triángulo del café a Mayo de 2014

Destino reportado	Principal ciudad de destino				
	Numero de llegadas y variación anual			Participación porcentual	
	2013	Enero-mayo 2014	Mayo 2014	2013	Enero-mayo 2014
BOGOTA, D.C.	888.852	380.587	75.639	7,5%	10,4%
CARTAGENA	227.034	109.677	19.454	11,8%	14,0%
MEDELLIN	187.552	75.757	15.096	17,7%	16,0%
CALI	119.327	47.186	9.412	11,0%	14,2%
SAN ANDRES	53.835	34.404	6.438	5,3%	32,1%
BARRANQUILLA	52.387	21.234	3.676	4,9%	8,4%
CUCUTA	30.592	13.107	2.252	15,9%	23,3%
SANTA MARTA	21.688	8.773	1.594	-1,2%	-4,3%
BUCARAMANGA	19.183	6.871	1.103	-7,7%	-0,5%
PEREIRA	17.974	5.226	1.092	2,9%	-21,0%
OTROS	107.876	42.040	7.798	0,9%	8,3%
TOTAL	1.726.300	744.862	143.554	8,5%	12,0%



Fuente: MinTic. (2014). Principal ciudad de destino. Extranjeros no residentes, Mayo 2014.

En el Quindío se cuenta con un observatorio de turismo en cual se analiza constantemente el flujo de turistas, uso de medios de transporte, y entre los años 2013-2014 han agregado la satisfacción de los visitantes, adicionalmente ya tienen en cuenta a los otros actores del turismo: restaurantes, hoteles, artesanos (Quindio C. d., 2014).

En Caldas en el presente Plan de Desarrollo Manizales 2012-2015 “Gobierno en la Calle” 2012 (Manizales, 2012) se contempla la promoción del turismo de montaña, el observatorio de aves, ecoturismo. Se plantea el emprender estrategias de promoción de la ciudad, estrategias de desarrollo empresarial y estrategias de productos turísticos, en el momento no hay un observatorio de turismo o de turistas.

En Pereira la información existente sobre los turistas desde el año 2011 es recopilada mediante los PIT (Puestos de Información Turística), los cuales se encuentran ubicados en 3 puntos estratégicos de la ciudad, en el Aeropuerto Matecaña, en El centro comercial Alcides

Arévalo y en el teatro Lucy tejada, pero dicha información no es publica, ni se conoce, ni se publica, ya que el propósito de los PIT es brindar información mas no recopilarla (Pereira F. a., 2014).

El instituto de cultura y fomento al turismo se encarga de la recopilación de la información de los PIT, además de la información que se recopila del terminal de transportes y de los hoteles (Pereira F. a., 2014).

En el plan de desarrollo Municipal de la ciudad de Pereira del año 2008-2011 denominado “Pereira Región de Oportunidades” (Esap, 2008), en la Pagina 50 de dicho plan, se habla de un proyecto a 4 años sobre un estudio estratégico prospectivo el cual tiene unos puntos similares al plan realizado y ejecutado en el Quindío (Plan de Desarrollo Turístico Quindío 2020), pero dicho plan nunca se vio reflejado en ningún sitio.

En el nuevo plan desarrollo 2012-2015 denominado “Por una Pereira mejor” se tiene lo siguiente (Catolica, 2014):

De acuerdo con la medición de la cuenta satélite de turismo realizada por la Cámara de Comercio de Pereira en convenio con la Alcaldía, se encontró que para el año 2010, el turismo generó una producción en términos monetarios de \$72.3 mil millones, esto implica que el Turismo de la ciudad aporta al Producto Interno Bruto el 1.1%¹ del total de la economía, resultado que se encuentra dentro de los parámetros de un turismo aún por desarrollar, toda vez que el resultado Nacional es de 1.89% como proporción del PIB.

Esto indicaría una oportunidad de desarrollo para el turismo y la economía, en el

corto, mediano y largo plazo con una oferta central que se consolidaría con el “Parque Temático de Flora y Fauna” acompañado de otros lugares de tipo regional en el marco del paisaje Cultural Cafetero, se trata entonces de ofertas complementarias en un corredor territorial que sin duda será la estrategia correcta, pues es válido que el turismo aun esta por desarrollar en la ciudad y el Departamento, en las actividades directas y relacionadas como el transporte, Alimentación, alojamiento y agencias de viajes principalmente como lo indica la estructura de producción sectorial.(p,394)

Desde la perspectiva de las cuentas satélites para estudios estadísticos, se analizan variables como: turismo, entorno habitual, visitantes, turismo interno, turismo receptor, turismo emisor, turismo interior, turismo nacional, consumo turístico, productos característicos, productos conexos al turismo, productos no específicos. De acuerdo con las cuentas satélites, se hacen estudios estadísticos, pero no se analiza a profundidad los grupos de personas, las reglas de asociación de acuerdo con las características, no se determina el nivel de satisfacción ni se puede caracterizar de acuerdo a varios aspectos del viajero (Estadística, 2007).

Recordemos que la minería de datos se aleja de la estadística al no necesitar crear un modelo o hipótesis para luego estimarlo, tampoco se requiere tomar muestras ya que la minería de datos permite el uso de todos y cada uno de los registros(LUCENA, 2005).

Por lo tanto se hace necesario para todas las empresas de la ciudad de Pereira, para la alcaldía, para la cámara de comercio y demás entidades el poder caracterizar los visitantes y de esta manera mejorar aspectos desconocidos para todos o conocidos para unos pocos.

En la ciudad de Armenia se podría utilizar la información histórica y actual para hacer un estudio de minería de datos el cual nos brindara nuevos horizontes, esto también es posible en Pereira y Manizales, luego de recopilar información importante teniendo en cuenta variables de los diferentes sectores económicos y sociales.

Antecedentes

Enhancing Tourism Intermediaries with the Data Mining Process . En Taiwán se realizó un estudio para la industria del turismo, dicho proyecto permitía administrar y mejorar las relaciones con los clientes del turismo a largo plazo, directamente los beneficiarios eran los relacionados con el turismo: restaurantes, artesanos, hoteles, hostales, vendedores, comercio, proveedores de servicios turísticos internos, guías turísticos, sitios recreativos, museos, bares, discotecas, etc. (Intermediaries, 2012).

Los taiwaneses se enfocaron en promocionar Taiwán como destino turístico, esto implicó una estrategia para las agencias de viajes para que jugaran un rol importante en la operación del turismo. Ellos actuaron como intermediarios entre proveedores de hoteles, transporte, viajes y servicios para consumidores. De esta manera las agencias de viajes son reconocidas por el gobierno como parte fundamental de la cadena de distribución en el incremento de ingreso y salida de turistas (Intermediaries, 2012).

De esta manera las agencias se interesaron por los viajeros y mejoraron sus plataformas teniendo como base un CRM. Con el estudio realizado, todos los administradores estuvieron de acuerdo que el mayor beneficio de la minería de datos era la de retener los clientes en cada una de las agencias. Luego del estudio realizado, el autor del estudio se dio cuenta que las agencias de viajes no aplicaron adecuadamente la minería de datos, de acuerdo con esto añadieron un

cuarto proceso valioso para las agencias de turismo intermediarias (Intermediaries, 2012) de la siguiente manera:

Paso 1. Obtener información de los clientes. En este paso, es posible para las agencias de viajes obtener información valiosa de los tiquetes, reservaciones y procesos de registro. Esta información incluye solicitudes especiales de las acomodaciones y la duración de la estadía, esto ayuda a los agentes de viaje a modificar sus campañas de marketing e incrementar las reservas.

Paso 2 Clasificar los datos recolectados de los clientes. Utilizando la minería de datos para compilar datos recolectados, se clasifica a los clientes en segmentos predefinidos que permiten un tamaño y estructura adecuados para ser monitoreados. Al final, las agencias de viajes pueden analizar los datos demográficos para sectorizar los clientes los cuales relacionan las compras y demandas de los clientes.

Paso 3 Identificando segmentos objetivos. La minería de datos es utilizada para identificar grupos de clientes con comportamiento similar y medir las reacciones de los clientes a las diferentes ofertas. La estrategia del CRM puede visualizarse con más claridad luego de este paso y es utilizado para establecer relaciones de larga duración con los clientes.

Paso 4 Mejorando los servicios. Los clientes de alta rentabilidad se pueden mantener por el uso de la minería de datos desde que dichas técnicas ayuden a mejorar el servicio de acuerdo con las preferencias de cada cliente. Por ejemplo, una agencia de viajes crea una campaña directa al email y se lo envía únicamente a los clientes objetivos.

De acuerdo con el proceso realizado de forma empírica por la entidad, (Lin, 2012) las agencias de viaje mejoran sus estrategias de marketing y maximizan sus ganancias.

A knowledge destination framework for tourism sustainability: A business intelligence application from Sweden. (Fuchs, Abadzhiev, & Svensson, 2013) Este proyecto fue realizado a partir de las diferentes fuentes de información de datos de viajes de clientes que están dispersos en los destinos turísticos (Ejemplo: servidor web de los turistas, comportamiento de navegación web, base de datos de transacciones, datos de feedback, etc.), estos datos valiosos normalmente permanecen sin uso. Así, la creación de conocimiento y la adquisición de procesos en los destinos turísticos pueden ser significativamente extendidos aplicando métodos de minería de datos (Liu, 2008), (Pitman, 2010).

De esta manera se considera una nueva alternativa para descubrir las preferencias de los consumidores de productos turísticos, particularmente en hoteles y restaurantes, esto utilizando las opiniones de los sitios Web como TripAdvisor.

framework (Fuchs, Abadzhiev, & Svensson, 2013) se basa en varias capas:

Capa de creación del conocimiento. Aplica métodos de extracción de información, convierte los orígenes de datos a una forma accesible para los subproveedores. Por ejemplo, en el lado del cliente/turista, el conocimiento se genera por mecanismos de feedback como encuestas y evaluaciones en línea. Después el conocimiento implícito puede ser explícito visualizando información de seguimiento de turistas. (Ejemplo: comportamiento de las búsquedas en la web) a través de minería. En el lado del proveedor, el conocimiento sobre productos, perfiles, información disponible, información sobre recursos y calidad de los recursos, así como calidad de vida de los residentes y satisfacción en el trabajo.

La capa de aplicación del conocimiento. Ofrece servicios a usuarios finales que inteligentemente informan acerca de los recursos de los destinos, elementos de los proveedores y actividades de los clientes. Así las aplicaciones para clientes / turistas particularmente comprenderá los servicios de localización que apoyan la construcción de comunidad y el consumo inteligente (por ejemplo, a través de servicios de recomendación por ser sensible al contexto, así como de adaptación para el usuario).

En cambio, los servicios inteligentes para los proveedores de destino interesados, se centran en aplicaciones de inteligencia de negocios relacionados con el turismo (Cho, Knowledge discovery techniques in database marketing for the tourism industry, 2002) ; (Olmeda, 2002).

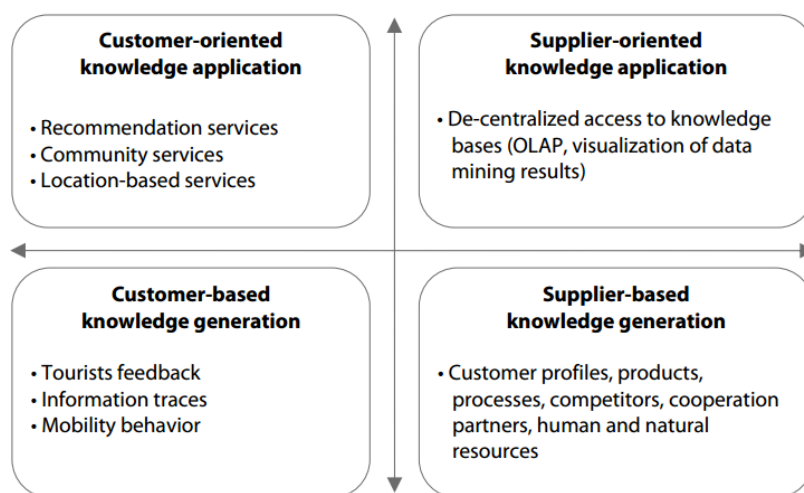


Figura 1. The knowledge destination framework

Fuente: (Fuchs M. &, 2009) (2009). Knowledge discovery techniques in database marketing for the tourism industry

La arquitectura del framework del conocimiento:

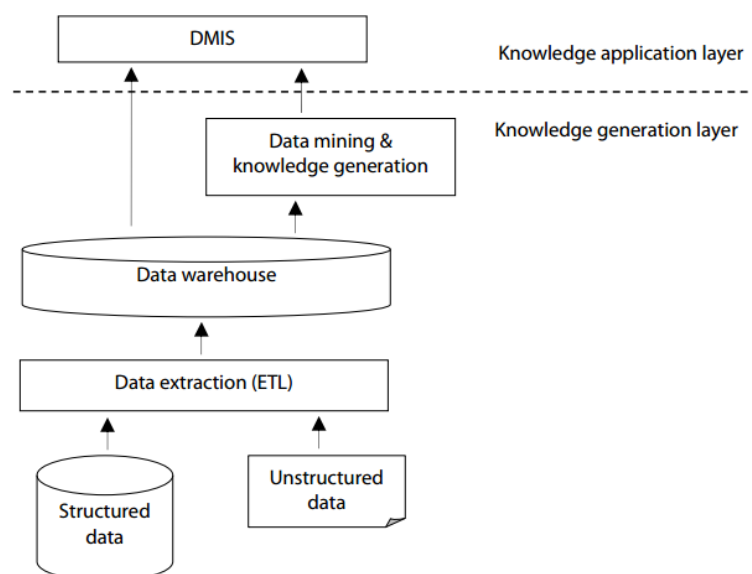


Figura 2. The knowledge destination framework

Fuente: (Fuchs M. &, 2009) (2009). Knowledge discovery techniques in database marketing for the tourism industry. [Figura]. Recuperado de Knowledge discovery techniques in database marketing for the tourism industry

A supplier-oriented knowledge application - The destination management information system (DMIS). En el diseño de un DMIS, el conocimiento basado en el cliente se crea a través de técnicas de segmentación de clientes y servicio de evaluación del desempeño (Ritchie, 2002); (Cho, Knowledge discovery techniques in database marketing for the tourism industry, 2002); (Pyo, 2005) .Los datos recogidos, almacenados, analizados y visualizados en los DMIS incluyen datos demográficos de los turistas características de la compra, motivos y las percepciones de la marca, así como el uso de la información de los clientes y los patrones de consumo del producto, respectivamente (Fuchs M. &, 2009).

Dado que el uso eficaz de una DMIS requiere no sólo las aplicaciones de tecnología sofisticada, pero particularmente demanda establecer un aprendizaje organizacional, es crucial para integrar las partes interesadas públicas y privadas en para definir los requisitos de conocimientos.

En el estudio de establecer el " Destino de Turismo Neutro " (Thierstein, 2000); (McDonough, 2002); (Gössling, 2009), los estudiosos del turismo obtuvieron suficiente evidencia empírica necesaria para caracterizar el turismo sostenible con el medio ambiente (Dolnicar S. &., Selective marketing for environmentally sustainable tourism, 2008), (Dolnicar S. &., Selective marketing for environmentally sustainable tourism, 2008); (Reinsberg, 2010). En consecuencia, los turistas con la " huella ecológica más pequeña " relativamente se caracterizan por los siguientes atributos:

1) Socio- demográficos (Dolnicar S. &., Selective marketing for environmentally sustainable tourism, 2008) (Dolnicar S. C., 2008). :

- Mediana edad hasta tercera edad
- Educación superior

2) El comportamiento en el viaje y estilos de las vacaciones

- Sitios de campamento
- Apartamentos privados

3) Las actitudes positivas hacia actividades basadas en la naturaleza

- Apreciación de la naturaleza y el disfrute de la belleza natural y el paisaje
 - Interacción con la naturaleza y la preferencia por actividades basadas en la

naturaleza (por ejemplo, senderismo, ciclismo, pesca, escalada, observación de la naturaleza, etc.).(p. 677)

En el documento y en la investigación realizada hay fundamentos teóricos y empíricos para enfocar la encuesta hacia los puntos más relevantes para tener en cuenta por el sector público y privado en cuanto a las características de los turistas. Los clientes dejan evidencias de sus actividades en la web como, búsquedas, planeación de viajes, reservaciones, consumos, feedback en sus sitios webs o blogs (Höpken W. F., 2014).

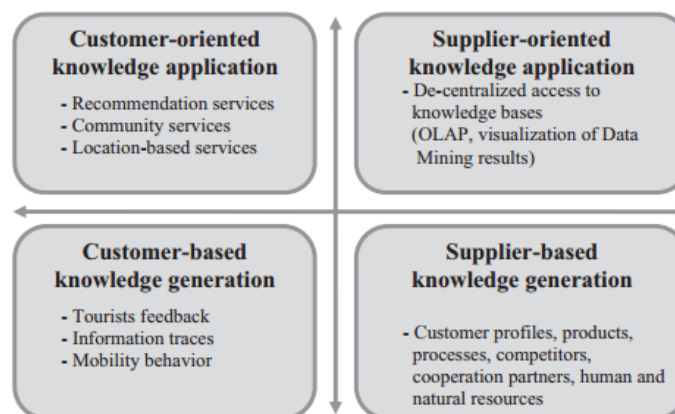


Figura 3. The knowledge destination framework

Fuente: (Höpken W. F., 2014) (2014). The Knowledge destination- Applying methods of business intelligence to tourism.

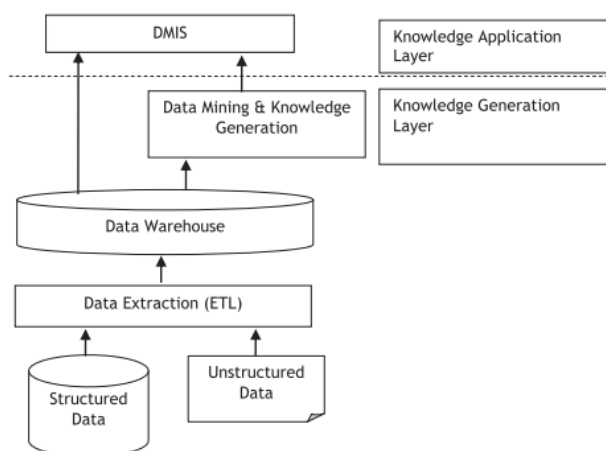


Figura 4. The knowledge destination framework architecture

Fuente: (Höpken W. F., 2014). The Knowledge destination- Applying methods of business intelligence to tourism.

Atributos Relevantes en el Perfil del Turista del Interior en la Ciudad de Mar del Plata. En un artículo publicado en la red de revistas científicas de Red de Revistas Científicas de América Latina, el Caribe, España y Portugal (Redalyc, 2011) habla sobre un estudio realizado sobre “Atributos Relevantes en el Perfil del Turista del Interior en la Ciudad de Mar del Plata” en su resumen (Mallo P. E., y otros) dice:

El propósito del trabajo consistía en presentar los grupos de turistas del interior con características distintivas para su identificación durante la temporada de verano 2004-2005 en la ciudad de Mar del Plata a través de la utilización de técnicas de aprendizaje no supervisado. (P.2)

Las variables que se tuvieron en cuenta señalan los aspectos más relevantes en el trazado

de los rasgos distintivos de aquellas personas que eligen dicha ciudad como destino de sus vacaciones, utilizando para este análisis a los turistas que residen en las provincias del interior de Argentina, dicho trabajo es empírico, seleccionando a priori las personas por su lugar de procedencia. En su metodología de desarrollo se puede observar que la información necesaria para el proceso de minería de datos era inexistente, por lo tanto utilizaron varias técnicas de recopilación de información (Mallo P. E., y otros) para su posterior procesamiento:

- Encuesta Coincidental:

Lugar de toma de datos: Puntos fijos en estaciones de servicio, Estación de ómnibus, Estación de Trenes y Aeropuerto, se encuesta a los turistas en el momento de la partida de la ciudad.

Tamaño de la muestra: 300 casos semanales

Muestreo: probabilístico según ajuste al peso relativo de los medios de transporte utilizados por los turistas para ingresos a la ciudad.

- Encuesta de Ocupación Hotelera

Muestra: 117 hoteles

Estratificación: por categorías y por zonas

Medio de Recolección: relevamiento telefónico

- Encuesta de Perfil y Comportamiento

Lugar de toma de datos: balnearios de la ciudad

Muestreo: probabilístico por conglomerados.

Tamaño de la muestra: 1.100 casos proyectados, 1.072 casos realizados. (p.7)

Justificación

La ciudad de Pereira cuenta con la entidad pública “Instituto de Cultura y al turismo” para realizar obras de teatro, generar, fortalecer y ejecutar proyectos de cultura y turismo, orientados al reconocimiento de nuestra cultura y la de otros grupos sociales y étnicos, permitiendo el acceso y disfrute de las diversas expresiones culturales y atractivos turísticos (PEREIRA, 2014).

La información existente sobre los turistas de la ciudad de Pereira, desde el año 2011 es recopilada mediante los PIT (Puestos de Información Turística), los cuales se encuentran ubicados en 3 puntos estratégicos de la ciudad, en el aeropuerto Matecaña, en El centro comercial Alcides Arévalo y en el teatro Lucy tejada. El instituto de cultura y fomento al turismo se encarga de la recopilación de la información de los PIT, además de la información que se recopila del terminal de transportes y de los hoteles (PEREIRA, 2014).

Dicha información en realidad no está soportada en ningún Sistema de información, las encuestas son pequeñas muestras, no hay un observatorio de turismo, la entidad no encuentra interés en realizar dichos estudios, esto desde el año 2011 en el cual se hicieron los primeros acercamientos en los cuales se comprometen a gestionar con las personas encargadas del turismo, las diferentes encuestas y recopilación de datos sobre los viajeros de la Ciudad.

Hasta el momento la información real de los turistas no se encuentra recopilada, solo se obtiene información de las cuentas satélites de turismo (Quindio O. T., 2014). Dichas cuentas cada vez abarcan más indicadores de turismo para determinar satisfacción (Estadística, 2007). Estos datos cuentan con estadísticas y gráficos para análisis, pero no hay un estudio de minería

de datos para determinar una caracterización, clustering y reglas de asociación de acuerdo con determinadas características de los viajeros.

En el Quindío se cuenta con un observatorio de turismo en cual se analiza constantemente el flujo de turistas, uso de medios de transporte, y entre los años 2013-2014 han agregado la satisfacción de los visitantes, adicionalmente ya tienen en cuenta a los otros actores del turismo: restaurantes, hoteles, artesanos (Quindio C. d., 2014).

La ciudad de Armenia y el Quindío en general obtendrían bastantes beneficios de un proyecto de minería de datos, ya que ellos cuentan con bastante información. Claro que se puede observar que los datos son muestras estadísticas, pero no hay una minería de datos estricta como tal. Ya que en el triángulo del café no hay un estudio de minería de datos para avanzar en el análisis de datos, ya que la estadística se queda en el muestreo e interpretación a pequeña escala, se hace necesario un estudio de minería de datos el cual utilice toda la información y nos brinde nuevas relaciones entre los datos (Lang., 2010).

Pero no solamente los 3 departamentos podrían aprovechar el potencial de un análisis de minería de datos, todos los sectores relacionados con la ciudad se verían beneficiados, como podría ser el sector hotelero, restaurantes, centros comerciales, universidades, tiendas de barrio, etc. Todos están relacionados directa o indirectamente con los locales y visitantes. (Comercio, 2007)

De acuerdo con lo anterior es importante determinar el por qué y él para que de los visitantes de la ciudad de Pereira, aportando información valiosa la cual servirá para hacer un análisis y brindar oportunidades de negocios a otros sectores que tal vez no se han tenido en cuenta.

Formulación del problema

¿Cómo podemos agrupar y definir reglas de asociación de los viajeros de la ciudad de Pereira para poderlos caracterizar?

La información existente sobre los turistas de la ciudad de Pereira, desde el año 2011 es recopilada mediante los PIT (Puestos de Información Turística), los cuales se encuentran ubicados en 3 puntos estratégicos de la ciudad, en el aeropuerto Matecaña, en El centro comercial Alcides Arévalo y en el teatro Lucy tejada (PEREIRA, 2014).

Hasta el momento la información real de los turistas no se encuentra recopilada, solo se obtiene información de las cuentas satélites de turismo (Quindio C. d., 2014). Dichas cuentas cada vez abarcan más indicadores de turismo para determinar satisfacción (Estadística, 2007).

Ya que la estadística se limita al muestreo e interpretación a pequeña escala, se hace necesario un estudio de minería de datos el cual utilice toda la información, agrupe, defina nuevas relaciones entre los datos, defina reglas, asocie datos de acuerdo con determinadas características (Lang., 2010).

Objetivos

Objetivo General

Establecer las principales tendencias y características de los viajeros que visitan la ciudad de PEREIRA por medio de técnicas en minería de datos.

Objetivos Específicos

- Segmentar y describir hábitos de viajeros que visitan la ciudad de Pereira por medio de un Análisis de Clúster para la industria Turística de Pereira.
- Establecer relaciones y tendencias existentes entre las características de los viajeros y los aspectos que rodean su visita a la ciudad de Pereira por medio de reglas de asociación.
- Determinar en los viajeros que visitan la Ciudad de Pereira el nivel de gastos, percepción y satisfacción luego de visitar la Ciudad por medio de un análisis de clúster.

Resultados esperados

- Vista minable de datos a partir de la base de datos de los viajeros de la ciudad de Pereira.
- Flujo de trabajo con el proceso de ETL automatizado en una herramienta de uso libre que genere la vista minable.
- Documento con la segmentación y descripción de hábitos de los viajeros que visitan la ciudad de Pereira por medio de análisis de clúster.
- Documento con análisis de tendencias y relaciones entre los viajeros y los aspectos que rodean su visita por medio de reglas de asociación.
- Documento con informe de percepción, satisfacción y nivel de gastos de los viajeros de la ciudad de Pereira por medio de clúster. (Ver ¡Error! No se encuentra el origen de la referencia.)

En la metodología CRISP-DM se tiene en cuenta la validación y verificación de los datos (Chapman, 2007). Para el caso de este proyecto, la comparación con otros resultados de otros estudios no es posible ya que en la ciudad no un estudio similar, los datos más similares son los del observatorio de Turismo del Quindío (Quindio O. T., 2014). Ya que esta región tiene una capital con unos aspectos sociales similares pero con diferentes orígenes. En el desarrollo del proyecto se realiza un contraste de los resultados con la realidad: noticias, estadísticas, percepción. Como ejemplo, si en los informes oficiales se habla sobre reducción del desempleo pero el resultado de la minería es de pobreza, se realizara un contraste como analistas de datos de

ambos informes.

Estrategia metodológica

Metodología

Para esta investigación se requiere de una metodología independiente de las herramientas de minería de datos la cual contenga los procesos de revisión del negocio, permitiendo organizar, estructurar y definir los parámetros iniciales del negocio (Santos, 2005).

La metodología requerida para este tipo de investigación basa en la metodología CRISP-DM por su uso con herramientas libre y herramientas de pago (Chapman, 2007), se requieren las siguientes tareas acordes con el alcance del proyecto.

Comprensión del dominio. Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto y exigencias desde una perspectiva institucional, luego convirtiendo este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos:

- Establecimiento de los objetivos del proyecto.
- Evaluación de la situación.
- Establecimiento de los objetivos de la minería de datos.

Comprensión de los datos (Chapman, 2007). La fase de entendimiento de datos comienza con la colección de datos inicial y continúa con las actividades que le permiten familiarizarse primero con los datos, identificar los problemas de calidad de datos, descubrir los primeros conocimientos en los datos, y descubrir subconjuntos interesantes para

formar hipótesis en cuanto a la información oculta (Chapman, 2007).

- Recopilación inicial de datos.
- Descripción de los datos.
- Exploración de los datos.
- Verificación de calidad de datos.

Preparación de los datos. La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto de datos final (los datos que serán provistos en las herramientas de modelado) de los datos en brutos iniciales. Las tareas de preparación de datos probablemente van a ser realizadas muchas veces y no en cualquier orden prescripto. Las tareas incluyen la selección de tablas, registros, y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan (Chapman, 2007).

- Selección de los datos.
- Limpieza de datos.
- Construcción de datos.
- Integración de datos.

Modelado. En esta fase, varias técnicas de modelado son seleccionadas y aplicadas, y sus parámetros son calibrados a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de datos. Por lo tanto, volver a la fase de preparación de datos es a menudo necesario (Chapman, 2007).

- Selección de la técnica de modelado.

- Diseño de la evaluación.
- Construcción del modelo.
- Evaluación del modelo.

Evaluación. En esta etapa del proyecto, se ha construido un modelo (o modelos) que parecen tener la alta calidad desde la perspectiva de análisis de datos.

Antes del proceder al despliegue final del modelo, es importante evaluarlo a fondo y la revisión de los pasos ejecutados para crearlo, para comparar el modelo correctamente obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no ha sido suficientemente considerada. En el final de esta fase, una decisión en el uso de los resultados de minería de datos debería ser obtenida (Chapman, 2007).

- Evaluación de resultados
- Revisar el proceso
- Validación de resultados obtenidos

Despliegue. La creación del modelo no es generalmente el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento ganado tendrá que ser organizado y presentado en el modo en el que el cliente pueda usarlo. Ello a menudo implica la aplicación de modelos "vivos" dentro de un proceso de toma de decisiones de una organización, por ejemplo, en tiempo real la personalización de página Web o la repetida obtención de bases de datos de mercadeo. Dependiendo de los requerimientos, la fase de desarrollo puede ser tan simple como la generación de un informe o tan compleja como la realización repetida de un proceso cruzado de minería de datos a

través de la empresa. En muchos casos, es el cliente, no el analista de datos, quien lleva el paso de desarrollo. Sin embargo, incluso si el analista realizara el esfuerzo de despliegue, esto es importante para el cliente para entender de frente que acciones necesita para ser ejecutadas en orden para hacer uso de los modelos creados actualmente (Chapman, 2007).

- Generación de informe final
- Documento con experiencias.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background Business Objectives Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i>	Select Data <i>Rationale for Inclusion/ Exclusion</i>	Select Modeling Techniques <i>Modeling Technique Modeling Assumptions</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</i>	Plan Deployment <i>Deployment Plan</i>
Assess Situation <i>Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</i>	Describe Data <i>Data Description Report</i>	Clean Data <i>Data Cleaning Report</i>	Generate Test Design <i>Test Design</i>	Review Process <i>Review of Process</i>	Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Determine Data Mining Goals <i>Data Mining Goals Data Mining Success Criteria</i>	Explore Data <i>Data Exploration Report</i>	Construct Data <i>Derived Attributes Generated Records</i>	Build Model <i>Parameter Settings Models Model Descriptions</i>	Determine Next Steps <i>List of Possible Actions Decision</i>	Produce Final Report <i>Final Report Final Presentation</i>
Produce Project Plan <i>Project Plan Initial Assessment of Tools and Techniques</i>	Verify Data Quality <i>Data Quality Report</i>	Integrate Data <i>Merged Data</i>	Assess Model <i>Model Assessment Revised Parameter Settings</i>		Review Project <i>Experience Documentation</i>
		Format Data <i>Reformatted Data Dataset Dataset Description</i>			

Figura 5. Tareas genéricas y salidas del modelo CRISP-DM

Fuente: Chapman. (2007). Ilustración del modelo Crisp-Dm.

Pruebas

Las pruebas se definen en el apartado “Generar el diseño de la prueba”, en el capítulo Desarrollo siguiendo la metodología CRISP-DM.

Presupuesto

Para el desarrollo del presente proyecto no se contempla un presupuesto muy alto las herramientas requeridas son de uso gratuito y las herramientas de apoyo que son de serán aportadas por el autor. (Ver

Bibliografía

A. K. Jain, M. M. (1999). Data clustering: a review. En M. M. A. K. Jain, *ACM Computing Surveys* (págs. 264-323).

A., M. E. (1 de 12 de 2000). *Revista de Investigación en Psicología*. Obtenido de <http://sisbib.unmsm.edu.pe>:
http://sisbib.unmsm.edu.pe/bvrevistas/investigacion_psicologia/v03_n2/pdf/a08v3n2.pdf

ABS. (2002). Measuring a Knowledge Based Economy and Society. *Australian Bureau of Statistics*.

Acharya., S. M. (2003). *Data Mining: Multimedia, Soft Computing and Bioinformatics*. Wiley Inter-Science.

Agrawal, R. e. (1993). *Mining Association Rules between Sets of Items in Large Database*. Recuperado el Octubre 15 de 2014, de <http://rakesh.agrawal-family.com>:
<http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>.

Aguilar, L. J. (2013). Big Data. En L. J. Aguilar, *Análisis de grandes volúmenes de datos en organizaciones* (págs. 153-210). Mexico: Alfaomega.

Ahmed, P. K. (2002). *Learning Through Knowledge Management*. Oxford: Butterworth.

Airbnb. (s.f.). *Airbnb*. Obtenido de <https://www.airbnb.es/>: <https://www.airbnb.es/>

Ayala, H. (2000). Surprising Partners: Hotel Firms and Scientists Working Together to. *Cornell Hotel and Restaurant Administration Quarterly*, 42-57.

- Bahra, N. (2001). *Competitive Knowledge Management*. Basingstoke: Palgrave.
- Bernal, F. (2012). *Departamento de Ciencias de la Computacion e I.A.* Obtenido de Departamento de Ciencias de la Computacion e I.A.: <http://elvex.ugr.es/decsai/intelligent/slides/dm/D2%20Association.pdf>
- Bornhorst, T. R. (2010). Determinants for DMO & destination success: an empirical examination. *Tourism Management*, 572-589.
- Bouncken, R. (2002). Knowledge Management for Quality Improvement in Hotels. *The Haworth Hospitality Press*, págs. 25-29.
- C, P., & Santín, D. (2007). *Mineria de Datos. Técnicas y Herramientas*. Madrid: Paraninfo.
- CAMPBELL, M. (2009). *base IV Guía de Autoenseñanza*. España. : McGraw Hill – Interamericana.
- Carson, D. a. (2004). Strategic Knowledge Management in Tourism. *paper presented to the CAUTHE conference*.
- Catolica, U. (1 de MAyo de 2014). www.ucp.edu.co, 1. Recuperado el 10 de Septiembre de 2014, de www.ucp.edu.co: http://www.ucp.edu.co/ucp_docs/ppr/documentos-generales/PLAN_DESARROLLO-pereira.pdf
- Chapman, P. e. (2007). *CRISP-DM 1.0 - Step-by-step data mining guide*. Recuperado el 15 de Octubre de 2014, de <http://www.crisp-dm.org/CRISPWP-0800.pdf>: <http://www.crisp-dm.org/CRISPWP-0800.pdf>
- Cho, V. &. (2002). Knowledge discovery techniques in database marketing for the tourism industry. *Journal* , 109-131.
- Cho, V. &. (2002). Knowledge discovery techniques in database marketing for the tourism

- industry. *Journal of Quality Assurance in Hospitality & Tourism*, 109-131.
- Choi, T. a. (2000). Towards a Knowledge Discovery Framework for Yield Management in the. *Hospitality Management*, 17-31.
- Chu, F. L. (2004). Forecasting tourism demand: a cubic polynomial approach. *Tourism Management*, 209-218.
- Comercio, M. d. (1 de 1 de 2007). *Ministerio de comercio*, 2008. Recuperado el 1 de Septiembre de 2014, de Ministerio de comercio:
<http://www.mincomercio.gov.co/econtent/documentos/turismo/2008/PlanSectoria2007-2010->
- ComputerWorld. (10 de 04 de 2014). *ComputerWorld*. Obtenido de ComputerWorld:
<http://www.computerworld.es/tendencias/el-universo-digital-se-expande-acelerado-por-el-crecimiento-de-los-datos>
- Cooper, & Ruhanen. (2003). Developing a Knowledge Management Approach to Tourism Research.
- Cooper, C. (5 de Abril de 2005). *www.elsevier.com/locate/atoures*. Recuperado el 24 de Septiembre de 2014, de *www.elsevier.com/locate/atoures*:
<http://www.sciencedirect.com/science/article/pii/S0160738305000915>
- Cooper, C. (4 de Enero de 2006). *Knowledge management and tourism*. Recuperado el 25 de Septiembre de 2014, de <http://www.sciencedirect.com>:
<http://www.sciencedirect.com/science/article/pii/S0160738305000915>
- Council, W. T. (1 de Noviembre de 2012). *wttc.org*, 1. Recuperado el 1 de Septiembre de 2014, de *wttc.org*: <http://www.wttc.org/~media/files/reports/benchmark>

- Crompton, J. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, 408-424.
- Crompton, J. y. (1997). Motives attending festival events. *Annals of Tourism Research*, 425-439.
- DANE. (1 de 08 de 2015). <http://www.dane.gov.co>. Obtenido de <http://www.dane.gov.co>:
http://www.dane.gov.co/files/investigaciones/boletines/ipc/cp_ipc_feb15.pdf
- Daniel, C. K. (4 Mar 2009). *Sistemas de información para la toma de decisiones*. México. :
McGraw-Hill.
- Davidson, C. a. (2002). *Knowledge Management*. Auckland: Tandem.
- Dean, J. (2014). *Big data, Data mining and Machine Learning - Value Creation for Business Leaders ans Practitioners*. Wiley & SAS Business Series.
- Devesa Fernández María, L. G. (1997). Motivación, satisfacción y lealtad en el turismo: el caso de un destino de interior. *Revista Electronica de Motivación y Emoción*.
- Devesa, M. y. (2006). Determinantes de la satisfacción percibida en el turismo rural. En M. y. Devesa, *Determinantes de la satisfacción percibida en el turismo rural*. Valencia: Tirant Lo Blanch.
- Dolnicar, S. &. (2008). Selective marketing for environmentally sustainable tourism. *Tourism Management*, 672-680.
- Dolnicar, S. &. (2008). Selective marketing for environmentally sustainable tourism. *Tourism Management*, 672-680.
- Dolnicar, S. C. (2008). Environment-friendly tourists: what do we really know about them? *Journal of Sustainable Tourism*, 197-210.
- Duque Méndez, N. D. (2005). *Universidad Nacional de Colombia*. Obtenido de Universidad Nacional de Colombia:

<http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060029/lecciones/cap8-1.html>

E, B. G. (2003). *Lic. en sistemas de información. Almacenes de datos y Minería de Datos.*

Obtenido de Lic. en sistemas de información. Almacenes de datos y Minería de Datos:

<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm>

Esap. (21 de Mayo de 2008). <http://cdim.esap.edu.co>. Recuperado el 1 de Septiembre de 2014, de <http://cdim.esap.edu.co>:

<http://cdim.esap.edu.co/BancoMedios/Documentos%20PDF/pd%20-%20plan%20de%20desarrollo%20-%20pereira%20-%20risaralda%20-%202008%20-%202011.pdf>

Estadística, D. A. (27 de Diciembre de 2007). <https://www.dane.gov.co>, 1. Obtenido de <https://www.dane.gov.co>:

https://www.dane.gov.co/files/investigaciones/fichas/pib/ficha_sateli_turismo.pdf

Feng, Y. L.-x.-y. (2010). Design of the Low-cost Business Intelligence System Based on Multi-agent. . *International Conference of Information Science and Management Engineering.*

Franks, B. (2012). Taming the big data tidal wave. En *Finding opportunities in huge data streams with advanced analytics*. New Jersey: Wiley.

Fu, R. X. (2008). Research on the Organizational Decision Support System for Small & Medium-Size Enterprise Based on Agent. *International Symposiums on Information Processing*. Obtenido de Research on the Organizational Decision Support System for Small & Medium-Size Enterprise Based on Agent.

Fuchs, M. &. (2004). Destination benchmarking - an indicator-system's potential for exploring

- guest satisfaction. *Journal of Travel Research*, 212-225.
- Fuchs, M. &. (2009). Data mining in tourism (In German: „Data Mining im Tourismus"). *Praxis der Wirtschaftsinformatik*, 73-81.
- Fuchs, M. C. (2011). Destination Brand Equity Modelling and Measurement. *A case from Sweden* (págs. 95-116). Emerald.: Goldsmith.
- Fuchs, M., Abadzhiev, A., & Svensson, B. (2013). A knowledge destination framework. *Scientific Journals of Croatia*.
- Gamble, P. M. (2000). Customer Knowledge Management in the Travel Industry. *Journal of Vacation Marketing*, 83-91.
- Google. (24 de Julio de 2013). <https://code.google.com/p/google-refine/>. Recuperado el 11 de Octubre de 2014, de google-refine/: <https://code.google.com/p/google-refine/>
- Goossens, C. (2000). Tourism information and pleasure motivation. *Annals of Tourism Research.*, 301-321.
- Gössling, S. (2009). Carbon neutral destinations: a conceptual analysis. *Journal of Sustainable Tourism*, 17-37.
- Grabova, Darmont, Chauchat, & Zolotaryova. (Junio de 2010). *ACM Digital Library*.
Recuperado el 27 de Octubre de 2014, de ACM Digital Library:
<http://dl.acm.org/citation.cfm?id=1893180>
- Gretzel, U. &. (2004). Implementing a knowledge-based tourism marketing information system. *Journal of Information Technology & Tourism*, 245-255.
- Guptara, P. (2000). Why Knowledge Management Fails. *Knowledge Management Review*, 26-30.

- H Witten, I., Eibe, F., & A. Hall, M. (2011). *Data Mining: Practical Machine Learning Tool and Techniques*. Morgan Kaufmann.
- Hallin, C. A. (2008). Knowledge management in the hospitality industry: a review of empirical research. *Tourism Management*, 366-381.
- Hastie, T. T. (2009). *The Elements of Statistical Learning*. New York: Springer.
- Hawkins, D. (2000). Transferring Tourism Knowledge. *Madrid: WTO Tourcom Conference*.
- Hayes-Roth, F. a. (1994). The State of Knowledge-Based Systems. *Communications of the ACM*, 27-39.
- Höpken, W. F. (2011). The knowledge destination – a customer information-based . *Information and Communication Technologies in Tourism 2011*, 417-429.
- Höpken, W. F. (2014). The Knowledge destination- Applying methods of business intelligence to tourism. *Encyclopedia of business analytics and optimization*, 307-321.
- <http://www.mividafreelance.com>. (s.f.). <http://www.mividafreelance.com>. Obtenido de <http://www.mividafreelance.com>: <http://www.mividafreelance.com/intercambiar-tu-casa-con-viajeros-tendencia-vacacional-auge/>
- Hunday, D. (2011). *Performance comparison of Apriori and FP-Growth algorithms in generating association rules*. Paris: Department of Computer Science.
- IBM. (18 de 06 de 2012). *Ibm Developers*. Recuperado el 2014, de Ibm Developers: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- IDC. (1 de Diciembre de 2012). <http://www.emc.com>. Recuperado el 1 de Octubre de 2014, de IDC EMC: <http://www.emc.com/about/news/press/2012/20121211-01.htm>
- IDC. (2014). <http://www.emc.com>. Recuperado el Octubre de 2014, de IDC EMC:

<http://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf>

Intermediaries, E. T. (23 de Febrero de 2012). *Enhancing Tourism Intermediaries*. Recuperado el

15 de Septiembre de 2014, de Enhancing Tourism Intermediaries:

www.ipcsit.com/vol45/025-ICIKM2012-M0061.pdf

Jared, D. (2014). *Big Data, Data Mining and Machine Learning, value creation for business leaders and practitioners*. Wiley.

Jiawei han, M. K. (2012). 3rd edition data mining concepts and techniques. En M. K. Jiawei han, *3rd edition data mining concepts and techniques*. Morgan Kaufmann.

Jiawei, H., Micheline, K., & Jian, P. (2012). *Data Mining : Concepts and Techniques*. Morgan Kaufmann.

Jones, A. (2001). Knowledge Capitalism—Business. *Oxford University Press*.

Kahle, E. (s.f.). Implications of “New Economy” Traits for the Tourism Industry. *Knowledge Management in Hospitality and Tourism* (págs. 5-23). The Haworth Hospitality Press.

Kimball, R., Caserta, J., & 2004. (s.f.). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. En R. & Kimball. Indianapolis: Wiley Publishing, Inc.

Kindsein. (1 de 8 de 2015). <http://www.kindsein.com/es/29/1/666/>. Obtenido de

<http://www.kindsein.com/es/29/1/666/>: <http://www.kindsein.com/es/29/1/666/>

Kuo, N.-W. (15 de Enero de 2011). *A Holistic Customer Experien*, 1. Recuperado el 15 de Septiembre de 2014, de A Holistic Customer Experien:

www.aicit.org/AISS/ppl/AISS103PPL.pdf

Lang., K. N. (2 de Febrero de 2010). *The increasing focus on managing relationships and*

customer retention, 1. (Technol, Editor, Entrep, Productor, & Afr) Recuperado el 15 de Septiembre de 2014, de The increasing focus on managing relationships and customer retention: www.ajol.info/index.php/jolte/article/download/51992/40627

Larose, D. T. (2005). *Discovering Knowledge in Data*. John Wiley & Sons.

Larose, D. T. (2005). *Discovering Knowledge in Data an Introduction to Data Mining*. . New Jersey: Jhon Wiley & Sons, Inc Publication.

Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics*. Canada: Wiley.

Law, R. (1998). Room occupancy rate forecasting – A neural network approach. *International Journal of Contemporary Hospitality Management*, 234-239.

Li, X. Z. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses. *International Conference on Computational Science*.

Lin, T.-C. (25 de Enero de 2012). *International Conference on Information and Knowledge Management*. Recuperado el 2 de Septiembre de 2014, de International Conference on Information and Knowledge Management: <http://www.ipcsit.com/vol45/025-ICIKM2012-M0061.pdf>

Liu, B. (2008). *Web Data Mining* (2. New York: Springer.

LUCENA, M. A. (01 de 01 de 2005). Análisis comparativo entre métodos estadísticos y de minería de datos. Madrid. Obtenido de <http://www.iit.upcomillas.es/pfc/resumenes/42b98a68079d0.pdf>

Magnini, V. H. (2003). Data mining for hotel firms: use and limitations. *Cornell Hotel and Restaurant Administration Quarterly*, 94-105.

Mallo, P. E., Artola, M. A., Zanfrillo, A. I., Morettini, M., & Galante, M. (2007). *Sistema de*

Información Científica. Recuperado el 1 de Octubre de 2014, de www.redalyc.org:
www.redalyc.org/pdf/276/27611106.pdf

Manizales, A. d. (15 de Enero de 2012). *www.manizales.gov.co*, 1. Recuperado el 1 de
 Septiembre de 2014, de www.manizales.gov.co:
<http://www.manizales.gov.co/dmd/pd/doc1.pdf>

Mario, P., Adoración, M., & Esperanza, M. (2000). *Diseño de bases de datos relacionales*.
 Alfaomega.

McDonough, W. &. (2002). *Cradle to Cradle. Remaking the Way We Make Things*. Toronto:
 Douglas & McIntyre.

Merv Adrian. (2011). <http://www.teradatamagazine.com>. Recuperado el 23 de Octubre de 2014,
 de <http://www.teradatamagazine.com>:
<http://www.teradatamagazine.com/v11n01/Features/Big-Data/>

Microsoft. (1 de Febrero de 2014). <http://msdn.microsoft.com>. Recuperado el 1 de Octubre de
 2014, de Microsoft: <http://msdn.microsoft.com/es-es/library/ms174949.aspx>

Min, H. M. (2002). A data mining approach to develop the profile of hotel customers.
International Journal of Contemporary Hospitality Management, 274-285.

Min, W. (22 de Marzo de 2010). *www.aicit.org*, 5. Recuperado el 15 de Septiembre de 2014, de
www.aicit.org: www.aicit.org/JCIT/paper_detail.html?q=269

Mining, J. H. (2000). *Concepts and Techniques*. Morgan Kaufmann Publishers Inc.

Moore. (1 de 11 de 2014). *Wikipedia*. Obtenido de Wikipedia:
http://es.wikipedia.org/wiki/Ley_de_Moore

Morales, D. R. (2008). Passenger name record data mining based cancellation forecasting for

revenue management. *Innovative Applications of O.R.*, 554-562.

Olmeda, I. &. (2002). Data mining techniques and applications for tourism internet marketing.

Journal of Travel & Tourism Marketing, 1-20.

Oracle. (15 de Enero de 2014). *Oracle NoSQL Database*. Recuperado el 23 de Octubre de 2014,

de Oracle NoSQL Database: <http://www.oracle.com/technetwork/database/database-technologies/nosqldb/overview/index.html>

otun, E. d. (28 de 7 de 2015). <http://www.eldiario.com.co>. Obtenido de

<http://www.eldiario.com.co>: <http://www.eldiario.com.co/seccion/LOCAL/c-mo-suena-pereira-de-noche-1507.html>

P, J. R. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster

analysis. *Journal of Computational and Applied Mathematics*, 20:53-65.

Pereira, A. d. (15 de Enero de 2014). www.pereira.gov.co, 1. Recuperado el 1 de Septiembre de

2014, de www.pereira.gov.co: www.pereira.gov.co/es/inicio.html

Pereira, F. a. (11 de 03 de 2014). <http://pereiraculturayturismo.gov.co>. Obtenido de

<http://pereiraculturayturismo.gov.co>:

http://pereiraculturayturismo.gov.co/index.php?option=com_content&view=article&id=350:puntos-de-informacion-turistica-reforzaran-la-atencion-de-propios-y-visitantes-durante-la-semana-santa&catid=89:noticiasinicio

PEREIRA, I. D. (15 de Octubre de 2014). *INSTITUTO DE CULTURA DE PEREIRA*.

Recuperado el 15 de Octubre de 2014, de <http://www.pereiraculturayturismo.gov.co>:

http://www.pereiraculturayturismo.gov.co/index.php?option=com_content&view=article&id=91&Itemid=631

Pitman, A. Z. (2010). Web usage mining in tourism – a query term analysis and. *Information and Communication Technologies in* , 393-403.

Press, E. (1 de 8 de 2015). <http://www.europapress.es>. Obtenido de <http://www.europapress.es>:
<http://www.europapress.es/sociedad/noticia-60-hijos-viven-padres-estan-parados-son-hombres-20150602142815.html>

Prisa., E. P. (24 de 9 de 2013). *elpais.com*. Obtenido de elpais.com:
http://elviajero.elpais.com/elviajero/2013/09/19/actualidad/1379610088_258672.html

Pujari, A. K. (2001). *Data Mining Techniques*. Universities Press.

Pyo. (2002). Knowledge Discovery Database for Tourist Destinations. *Journal of*, (págs. 396–403).

Pyo, S. (2005). Knowledge map for tourist destinations. *Tourism Management*, 583-594.

Quindio, C. d. (1 de Mayo de 2014). www.eam.edu.co, 28. Recuperado el 1 de Septiembre de 2014, de www.eam.edu.co:
<http://www.eam.edu.co/centrodeinvestigaciones/politicaspUBLICAS/Observatorio%20Tur%20C3%ADstico>

Quindio, O. T. (1 de Febrero de 2014). <http://www.eam.edu.co>. Recuperado el 15 de Octubre de 2014, de <http://www.eam.edu.co>:
<http://www.eam.edu.co/centrodeinvestigaciones/politicaspUBLICAS/Observatorio%20Tur%20C3%ADstico%20del%20Quind%20C3%ADo%202014.pdf>

Quindio., T. (1 de 5 de 2015). <http://www.turismoquindio.com>. Obtenido de <http://www.turismoquindio.com>:
http://www.turismoquindio.com/fincaspanaca_herreria14.php

- Quintero K. (2015). Ideafreelance1: Alquilar tu casa o habitación a viajeros. Obtenido de:
<http://www.mividafreelance.com/intercambiar-tu-casa-con-viajeros-tendencia-vacacional-auge/>
- Rainardi, V. (2007). Building a data warehouse with examples in SQL Server. En V. Rainardi.
Nueva York: Apress.
- Redalyc. (05 de 10 de 2011). *Red de Revistas Científicas de América Latina y el Caribe, España y Portugal*. Obtenido de Universidad Autónoma de México:
<http://www.redalyc.org/home.oa>
- Reinsberg, C. &-T. (2010). Characterizing environment friendly tourists. *Master Thesis*.
University of Stavanger. The Norwegian school of Hotel Management.
- República, B. d. (1 de 9 de 2014). <http://www.banrep.gov.co>. Obtenido de
<http://www.banrep.gov.co>:
http://www.banrep.gov.co/docum/Lectura_finanzas/pdf/dtser_208.pdf
- Risaralda., A. d. (1 de 06 de 2015). *Invest In Pereira*. Obtenido de Invest In Pereira:
<http://www.investinpereira.org/es/ipaginas/ver/98/ubicacion-geoestrategica-privilegiada/>
- Ritchie, R. J. (2002). A framework for an industry supported destination marketing information system. *Tourism Management*, 439-454.
- Roberto, N., & Marina, D. L. (abril de 2009). <http://www.scielo.org.co/>. Recuperado el 27 de Septiembre de 2014, de <http://www.scielo.org.co/>:
<http://www.scielo.org.co/pdf/iei/v29n1/v29n1a08.pdf>
- Sambamurthy, V. &. (2005). Information technologies and knowledge management.

Management , 1-7.

Santos, M. &. (2005). Data Mining – Descubierta de Conhecimento em Bases de Dados. *FCA Publisher*.

Schianetz, K. &. (2007). The learning tourism destination: the potential of a learning organization approach for improving the sustainability of tourism destinations. *Tourism Management*, 1485-1496.

Shawn, G., & Williams, A. (2009). Knowledge transfer and management in tourism organisations. *Tourism Management*, 325-335.

Sinnexus. (1999). <http://www.sinnexus.com>. Obtenido de <http://www.sinnexus.com>: http://www.sinnexus.com/business_intelligence/

Soares, S. (2012). *Big data governance. An Emerging Imperative*. MC Press Online.

Sveiby, E. (15 de Marzo de 2001). *What is Knowledge Management?* Recuperado el 15 de Septiembre de 2014, de *What is Knowledge Management?*: <http://www.sveiby.com/articles/>

Thierstein, A. &. (2000). The Sustainable Region: A Model for Action. *Series of works published by the Institute for Public Services and Tourism Contributions to the Regional Economy*.

Tiempo, E. (27 de 09 de 2015). *El Tiempo Noticias*. Obtenido de El Tiempo Noticias: <http://www.eltiempo.com/politica/justicia/trafico-de-personas-en-colombia/16388090>

TripAdvisor. (1 de 9 de 2015). <https://www.tripadvisor.co>. Obtenido de <https://www.tripadvisor.co>: <https://www.tripadvisor.co/HotelsList-Pereira-Hoteles-Baratos-zfp8016993.html>

Turismo, M. d. (1 de Septiembre de 2010). */www.mincomercio.gov.co*, 1. Recuperado el 1 de Septiembre de 2014, de *www.mincomercio.gov.co*:

<http://www.mincomercio.gov.co/eContent/documentos/turismo/2010/2010-TurismoNoviembre.pdf>

Turismo, M. y. (8 de Abril de 2014). *www.mincit.gov.co*, 1. Recuperado el 2014, de *www.mincit.gov.co*: <http://www.mincit.gov.co/descargar.php?id=71370>

Turismo, O. M. (22 de Enero de 2007). *www.unwto.org*, 1. Recuperado el 8 de Septiembre de 2014, de *www.unwto.org*: <http://media.unwto.org/es/content/entender-el-turismo-glosario-basico>

Turismo., M. I. (1 de 10 de 2015). *http://colombia.co*. Obtenido de <http://colombia.co>: <http://www.colombia.co/turismo/turismo-por-regiones/bogota-se-recorre-en-bicicleta.html>

Uber. (1 de 5 de 2015). <https://www.uber.com/es/cities/bogota>. Obtenido de <https://www.uber.com/es/cities/bogota>: <http://www.sociosbogota.com/vehiculos-aceptados.html>

unstats. (2 de Mayo de 1998). <http://unstats.un.org>, 1. Recuperado el 2 de Septiembre de 2014, de <http://unstats.un.org>: http://unstats.un.org/unsd/publication/SeriesM/SeriesM_67rev1S.pdf

Valhondo, D. (2003). *Gestión del conocimiento: Del mito a la realidad*. Madrid: Ediciones Díaz de Santos.

Vanguardia, P. (27 de 9 de 2015). <http://www.vanguardia.com>. Obtenido de <http://www.vanguardia.com>: <http://www.vanguardia.com/colombia/329543-eje->

cafetero-un-destino-de-clase-mundial

Velazco, R. H. (2007). *Almacenes de datos (Datawarehouse)*. Obtenido de Almacenes de datos (Datawarehouse).: <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>

Vlahogianni, E. I. (2010). Advanced computational approaches for predicting tourist arrivals. In T. *Advanced computational approaches for predicting tourist arrivals*. (págs. 309-324). Vienna: InTech.

Wang, Y. &. (2007). Conceptualizing and evaluating the functions of destination marketing systems. . *Journal of Vacation Marketing*, 187-203.

Watson, H. J., & Wixom, B. H. (Septiembre de 2007). *Ieee Computer Society*, 4. Recuperado el 23 de Octubre de 2014, de The community for Technology.

wikipedia.org. (11 de Mayo de 2014). *MongoDB*. Recuperado el 23 de Octubre de 2014, de MongoDB: <http://en.wikipedia.org/wiki/MongoDB>

Wöber, K. W. (1998). Global statistical sources- TourMIS. *Pacific Tourism Review*, 273-286.

Xie, H. (2009). The Research on Knowledge Management of Small and Medium-Sized IT Enterprises Based on Data Mining. *International Conference on Information Management, Innovation Management and Industrial Engineering*.

Yu, L. y. (2006). A comparative analysis of international tourists' satisfaction in Mongolia. *Tourism Management*, 1331-1342.

Yüksel, A. y. (2002). Measurement of tourist satisfaction with restaurant services A segment based approach. . *Journal of Vacation Marketing*., 52-68. .

Zaki, M., & Meira, Jr, W. (2014). *Data Mining and Analysis*. Cambridge.

Anexos

).

Desarrollo

Referente Teórico

Entender el turismo: conceptos básicos. La comunidad internacional se reunió en torno al turismo, definiendo un marco conceptual y análisis de la economía del turismo; un esfuerzo que duró casi tres años (2005-2007). El consenso internacional que siguió, en forma de las Recomendaciones internacionales aprobadas por Naciones Unidas, establece los conceptos, las definiciones, clasificaciones y el conjunto básico de datos e indicadores que deben formar parte de cualquier Sistema nacional de Estadísticas de Turismo (Turismo O. M., 2007).

Cuenta satélite de turismo (CST). La Cuenta Satélite de Turismo (que se describe en Cuenta satélite de turismo: Recomendaciones sobre el marco conceptual, 2008) es, aparte de las Recomendaciones internacionales para las estadísticas de turismo 2008, la segunda recomendación internacional sobre estadísticas de turismo que se ha desarrollada en un marco de coherencia con el Sistema de Cuentas Nacionales. Ambas recomendaciones son coherentes entre sí y constituyen el marco conceptual para medir y analizar el turismo como una actividad económica (Turismo O. M., 2007).

Terminología

Turista (o visitante que pernocta): Un visitante (interno, receptor o emisor) se clasifica como turista (o visitante que pernocta), si su viaje incluye una pernoctación (Turismo O. M., 2007).

Excursionista (o visitante del día): Un visitante (interno, receptor o emisor) se clasifica como turista (o visitante que pernocta), si su viaje incluye una pernoctación, o como visitante del día (o

excursionista) en caso contrario (Turismo O. M., 2007).

Industrias turísticas. Las industrias turísticas (también conocidas como actividades turísticas) son aquellas que generan principalmente productos característicos del turismo. Los productos característicos del turismo son aquellos que cumplen uno o ambos de los siguientes criterios:

- El gasto turístico en el producto debería representar una parte importante del gasto total turístico (condición de la proporción que corresponde al gasto/demanda).
- El gasto turístico en el producto deberían representar una parte importante de la oferta del producto en la economía (condición de la proporción que corresponde a la oferta). Este criterio supone que la oferta de un producto característico del turismo se reduciría considerablemente si no hubiera visitantes. (Turismo O. M., 2007)

Tabla 2. Lista de categorías de productos característicos del turismo y de industrias turísticas

Productos	Industrias
1. Servicios de alojamiento para visitantes	1. Alojamiento para visitantes
2. Servicios de provisión de alimentos y bebidas	2. Actividades de provisión de alimentos y bebidas
3. Servicios de transporte de pasajeros por ferrocarril	3. Transporte de pasajeros por ferrocarril
4. Servicios de transporte de pasajeros	4. Transporte de pasajeros por carretera

Productos	Industrias
por carretera	
5. Servicios de transporte de pasajeros	5. Transporte de pasajeros por agua
por agua	
6. Servicios de transporte aéreo de pasajeros	6. Transporte aéreo de pasajeros
7. Servicios de alquiler de equipos de transporte	7. Alquiler de equipos de transporte
8. Agencias de viajes y otros servicios de reservas	8. Actividades de agencias de viajes y de otros servicios de reservas
9. Servicios culturales	9. Actividades culturales
10. Servicios deportivos y recreativos	10. Actividades deportivas y recreativas
11. Bienes característicos del turismo, específicos de cada país	11. Comercio al por menor de bienes característicos del turismo, específicos de cada país
12. Servicios característicos del turismo, específicos de cada país	12. Otras actividades características del turismo, específicas de cada país

Fuente: Organización Mundial del Turismo OMT. (2008). Lista de categorías de productos característicos del turismo y de industrias turísticas, Enero 2008

La ciudad de Pereira: generalidades

Localización. El Municipio de Pereira está localizado a 4 grados 49 minutos de latitud norte, 75 grados 42 minutos de longitud y 1.411 metros sobre el nivel del mar; en el centro de la región occidental del territorio colombiano, en un pequeño valle formado por la terminación de un contra fuerte que se desprende de la cordillera central. Su estratégica localización central dentro de la región cafetera, lo ubica en el panorama económico nacional e internacional, estando unido vialmente con los tres centros urbanos más importantes del territorio (Pereira A. d., 2014).

Población. Consta de 488.839 personas de las cuales 410.535 se encuentran en el área urbana localizadas en 19 comunas y 78.304 en el área rural en 12 corregimientos (Pereira A. d., 2014).

Geografía El Municipio de Pereira, cuenta con pisos térmicos que van desde las nieves perpetuas (Nevado de Santa Isabel a 5.200 msnm) en límites con el Departamento del Tolima, hasta pisos cálidos a 900 msnm y a orillas del río Cauca. Por lo tanto, presenta distintas alternativas de uso agrícola (Pereira A. d., 2014).

Existen bosques para protección de cuencas, zonas de diversificación y medias conocidas como la zona cafetera y zonas cálidas con actividad ganadera y agrícola (piña, caña de azúcar, caña panelera y pasto) (Pereira A. d., 2014).

La ciudad de Pereira se encuentra a una altura promedio de 1.411msnm y cuenta con una temperatura promedio de 21°C. (Pereira A. d., 2014)

Clima . El suelo de Pereira se distribuye según sus climas así:

Clima cálido el 9.9 %, clima medio el 60.7 %, clima frío el 11.5%, páramo 17.7%, su precipitación media anual es de 2.750 mm (Pereira A. d., 2014).

Esta característica climática y la conformación de los suelos, brinda también una variedad en la cobertura vegetal y paisajística, potencializando el municipio de Pereira con una de las biodiversidades más ricas de la nación. No obstante, la ciudad se presenta como zona alta vulnerabilidad sísmica por el tipo de suelos que la conforman y por las fallas geológicas que la atraviesan (Pereira A. d., 2014).

Técnicas de almacenamiento de datos. La mayoría de las decisiones que se toman en una empresa u organización, se hace con base en la información obtenida de los datos que se tienen almacenados sobre la actividad de dicha organización (Daniel, 4 Mar 2009); generalmente éste almacenamiento está consolidado como una base de datos en primera instancia, la cual puede ser transformada en una bodega de datos para facilitar la toma de decisiones, pero inclusive y mejor aún se puede hacer una búsqueda inteligente de patrones o tendencias con las herramientas de Minería de Datos (C & Santín, 2007).

A continuación se hará una descripción de los tipos de almacenamientos de datos en la actualidad, los procesos de descubrimiento del conocimiento y las técnicas de minería de datos a utilizar en este proyecto.

Bases de datos. Una base de datos, es el almacenamiento organizado de datos que tienen una dependencia y que han sido recolectados y explotados por una organización o empresa en particular, dicha explotación se hace con programas creados para la manipulación de la misma (Daniel, 4 Mar 2009).

El término base de datos fue escuchado por primera vez en 1.963 en un simposio realizado en Estados Unidos California, donde se definió como “un conjunto de información relacionada que se encuentra agrupada o estructurada” (CAMPBELL, 2009); aunque realmente no se puede afirmar que sea un conjunto de información ya que en realidad son datos, los cuales como tal no son información a menos que se les intervenga para que lo sean, pero por ser la primera vez que se utilizó, se puede decir que es aceptable tal definición.

Para dar una información más detallada sobre las bases de datos es bueno hacer énfasis en cómo surgieron y para qué han sido utilizadas, se comenzará por hacer saber que éstas no han sido como hoy se presentan ya procesadas en computador, sino que esta información se tenía almacenada en papel como en archivos o bibliotecas por ejemplo; en las cuales se encuentra información y está organizada de acuerdo a unas métricas utilizadas por los archivistas/bibliotecarios; en las empresas, los datos que almacenaban eran demasiados y a la hora de necesitar hacer uso de estos podría resultar engorroso dependiendo del orden con el que los tuviesen guardados; quizás es a raíz de este problema en cuanto a tiempo, eficiencia y espacio que surgen las bases de datos en formato digital (Mario, Adoración, & Esperanza, 2000).

Otra de las definiciones para bases de datos la da el autor (Daniel, 4 Mar 2009) quien dice que: “Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.”. (p.60)

Desde el punto de vista informático, la base de datos es un sistema formado por un conjunto de datos almacenados en discos que permiten el acceso directo a ellos y un conjunto de programas que manipulen ese conjunto de datos.

Cada base de datos se compone de una o más tablas que guarda un conjunto de datos. Cada tabla tiene una o más columnas y filas. Las columnas guardan una parte de la información sobre cada elemento que se requiera guardar en la tabla, cada fila de la tabla (CAMPBELL, 2009).

Bodegas de datos, También conocidas como Almacenes de Datos, es un concepto relativamente nuevo, orientado al manejo de grandes volúmenes de datos, provenientes de diversas fuentes, de muy diversos tipos. Estos datos cubren largos períodos de tiempo, lo que trae consigo que se tengan diferentes esquemas de las bases de datos fuentes. La concentración de esta información está orientada a su análisis para apoyar la toma de decisiones oportunas y fundamentadas (Duque Méndez, 2005).

Una Bodega de Datos, es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales sirven de soporte en el proceso de toma de decisiones de la administración (Duque Méndez, 2005).

Su nombre, Bodega de Datos de ahora en adelante BGDs., se asocia con una colección de datos de gran volumen, provenientes de sistemas en operación y otras fuentes, después de aplicarles procesos de análisis, selección y transferencia de datos seleccionados. Su misión consiste en, a partir de estos datos y apoyado en herramientas sofisticadas de análisis, obtener información útil para el soporte a la toma de decisiones (Velazco, 2007).

En síntesis una BGDs. es una gran colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, y cuya actividad se centra en la *Toma de Decisiones* -es decir, en el análisis de la información- en vez de en su captura. Una vez reunidos los datos de los sistemas fuentes se guardan durante mucho tiempo, lo que permite el acceso a

datos históricos; así los almacenes de datos proporcionan al usuario una interfaz consolidada única para los datos, lo que hace más fácil escribir las consultas para la toma de decisiones (Velazco, 2007).

Según lo anterior se puede observar como las bodegas de datos, están más orientadas para ayudar al usuario en la toma de decisiones, diferente a las bases de datos las cuales tienen como principal objetivo la captura y almacenamiento de datos, siendo esto lo necesario para el desarrollo de éste proyecto.

Objetivos de las bodegas de datos. A continuación se muestran los objetivos que las BGDs. tienen para la utilización de datos de una empresa:

- Proveer una visión única de los clientes en toda la empresa.
- Poner tanta información comercial como sea posible en manos de tantos usuarios diferentes como sea posible.
- Mejorar el tiempo de espera que insumen los informes habituales.
- Monitorear el comportamiento de los clientes.
- Predecir compras de productos.
- Mejorar la capacidad de respuesta a problemas comerciales.
- Incrementar la precisión de las decisiones.
- Aumentar la productividad.
- Incrementar y distribuir las responsabilidades. (E, 2003)

Base de datos NOSQL. Base de datos no relacional, distribuida, de alto rendimiento y altamente escalable. Estas bases de datos son una categoría de sistemas de bases de datos que no

utilizan SQL como lenguaje de consulta principal. Existen cuatro grandes categorías: Almacenes clave-valor, almacenes de documentos, almacenes de grafos y familias de bases de datos columnares (racle, 2014).

Tipos de bases de datos NOSQL. Ya existe una gran cantidad de bases de datos NOSQL, a veces difíciles de clasificar debido a las diferencias entre soluciones. Algunos productos pueden tomar características de varias fuentes, como es el caso de la base de datos Cassandra, una de las más utilizadas y que tiene propiedades de dos grupos. Pero en general se suelen agrupar en cuatro grandes categorías.

- Orientadas a clave-valor
- Orientada a documentos
- Orientada a grafos
- Orientada a columnas y a BigTable. (Cassandra, 2015)

Aunque los precursores del movimiento NoSql son google BigTable y Amazon Dynamo, ambas son de código cerrado, pero en la actualidad existe una alternativa de código abierto denominada MongoDB la cual a la fecha es el sistema de NOSQL más popular (wikipedia.org, 2014).

Bases de datos clave-valor. El modelo clave-valor procede del modo de acceso a memoria en la programación en ensamblador. La dirección de la posición de memoria actúa como el valor que almacena en esa dirección de memoria. Otro ejemplo típico del modelo valor-clave es el concepto de tabla *hash* que tiene una función que transforma la clave de un índice utilizado para encontrar su valor asociado (Aguilar, 2013).

Los almacenamientos key-value asocian una clave única (key) al valor que se quiere guardar (value). Varias implementaciones de los almacenamientos key-value tienen funcionalidad adicional, pero a un nivel básico, el key-value solo requiere una clave y un valor. Este tipo de bases de datos suelen ser extremadamente rápidas y óptimas para una gran cantidad de accesos. Su intención es la de simplemente guardar una gran serie de “claves” con su valor asociado, lo cual da una potente flexibilidad ante datos no estructurados. La razón de utilizar este patrón (modelo) es la velocidad (Aguilar, 2013).

Bases de datos orientadas a grafos. Estas bases de datos organizan la información en grafos dirigidos. Son óptimas para hacer operaciones de consulta sobre las relaciones entre miembros y son extremadamente rápidas. La información se almacena dividiéndola en trozos más básicos, nodos (chunks) y estableciendo relaciones. Este tipo de bases de datos es muy eficiente para el caso de múltiples contactos (Facebook o LinkedIn) o los mensajes propios de Twitter (Aguilar, 2013).

En esta arquitectura los objetos se conocen como nodos y aristas (edges) aunque sirven los roles de entidad-relación de la arquitectura estándar SQL. Los nodos contienen propiedades que describen el dato real contenido en cada objeto. Un diagrama de una base de datos orientada a grafos es muy similar a los diagramas de objetos que utilizamos en la programación orientada a objetos.

La ventaja más grande es su velocidad en cierto tipo de transacciones, aquellas que implican relaciones, dado que no requiere el típico procesamiento intensivo de Join. (Aguilar, 2013)

Bases de datos orientadas a documentos. Estos abarcan una gran colección de formatos y codificación binaria. Los formatos de marcación estándar tales como XML y JSON se combinan con formatos propietarios como PDF de Adobe o Word de Microsoft. El lenguaje de consultas UnQL se desarrolló principalmente para consulta de documentos y objetos de datos marcados en JSON (Aguilar, 2013).

Los almacenes de documentos guardan la información como un listado de documentos desestructurados. Al acceder a un documento, se pueden ingresar en un número no especificado de campos con sus respectivos valores. Son muy rápidos para recuperar toda la información asociada al documento y tienen un esquema de datos muy flexible. Sin embargo, suelen ser lentos para hacer consultas donde se buscan todos los documentos con un determinado campo, ya que estos no suelen tener índices (Aguilar, 2013).

Guardan documentos heterogéneos que pueden ser:

- Textuales XML, formatos JSON, BSON, YAML, con la finalidad de almacenar información con estructura laxa y cambiante.
- Formatos binarios como PDF o Word. Ideal para servicios que requieren almacenar transacciones o datos enviados desde dispositivos móviles que usen JSON para intercambiar información. La información se almacena teniendo en cuenta dicha información más una “meta información” que la encapsula y clasifica por categorías (Aguilar, 2013).

Big Data. Big data o grandes volúmenes de datos o macro datos es un término que se escucha bastante la actualidad: en las comunidades científicas, en la radio, en la televisión, en conferencias sobre IOT – Internet of Things “Internet de las Cosas” (Cisco, 2015).

Los datos han ido creciendo de modo exponencial los últimos años, durante el 2011 se

crearon 1,8 zettabytes de datos (1 billón de gigabytes) según la consultora IDC y esta cifra se dobla cada dos años. Como dato significativo, Wal-Mart posee bases de datos con una capacidad de 2,5 peta bytes y procesa más de un millón de transacciones por hora (Aguilar, 2013).

Los Big Data están por todos lados y el no tenerlos en cuenta dará una desventaja significativa frente a otros que si lo tengan en cuenta, como señala (Franks, 2012): “Es imperativo que las organizaciones persigan agresivamente la captura y análisis de estas nuevas fuentes de datos para alcanzar los conocimientos y oportunidades que ellos ofrecen”.

Definición de big data. No hay unanimidad en la definición del concepto pero si un acuerdo entre todos los actores en denominarla: grandes volúmenes de datos, diferentes fuentes de información y análisis de dichos volúmenes de datos.

Adrian Merv, analista de la consultora Gartner da una definición en la revista *Teradata Magazine* de mayo de 2011 (Merv Adrian, 2011): “Big data excede el alcance de los entornos de hardware de uso común y herramientas de software para capturar, gestionar y procesar los datos dentro de un tiempo transcurrido tolerable para su población de usuarios”.

Otra definición bien de McKinsey Global Institute (Merv Adrian, 2011): “Big data se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar”.

Características del Big Data. Big data al igual que la nube (Cloud) abarca diversas tecnologías. Los datos de entrada de un Big Data pueden provenir de redes sociales, Logs, registros de servidores Web, sensores de flujos de tráfico, imágenes digitales, imágenes satelitales, flujos de audio y de radio, transacciones bancarias, MP3 de música, contenido de páginas web, escaneado de documentos de administración, caminos o rutas de GPS, telemetría de

automóviles, datos de mercado financiero (Aguilar, 2013).

IBM plantea que el Big Data abarca tres grandes dimensiones conocidas como el modelo “Modelo de las tres V” (3 v o v³): volumen, velocidad, variedad. Todos los autores coinciden en las 3 v, sin embargo IBM considera una cuarta V, denominada como Veracidad y otros autores añaden una quinta V denominada como Valor (IBM, 2012).

La Arquitectura del BIG DATA. La arquitectura de referencia se compone de dos grandes categorías: arquitectura de Big Data y gobierno de Big Data que debe integrarse con las infraestructuras existentes y coexistir con las acciones del gobierno de los datos tradicionales. La arquitectura se apoya en varios componentes que se organizan en torno a capas de la arquitectura (Aguilar, 2013).



Figura 6. *Arquitectura de Big Data*

Fuente:

(Aguilar,

2013). Ilustración de la Arquitectura del Big Data.

Inteligencia de negocios . La Inteligencia de Negocios es una solución tecnológica cuyo objetivo primordial es el apoyar la toma de decisiones, esta solución tecnológica está conformada por metodologías y herramientas que permiten cumplir dicho objetivo. Sin embargo, la Inteligencia de negocios puede adaptarse a diversos alcances dependiendo de las necesidades del negocio. La inteligencia de Negocios busca transformar los datos en información para finalmente transformar la información en conocimientos (Sinnexus, 1999).

La Inteligencia de Negocios en términos generales debiera contener los elementos que se muestran en la imagen.



Figura 7. *Elementos de un BI*

Fuente: (Sinnexus, 1999). Elementos de un BI.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc. (Sinnexus, 1999) .

Los principales productos de Business Intelligence que existen hoy en día son (Sinnexus, 1999):

- Cuadros de Mando Integrales (CMI)
- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)

Por otro lado, los principales componentes de orígenes de datos en el Business Intelligence que existen en la actualidad son:

- Datamart
- DataWarehouse

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un DataWarehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos (Sinnexus, 1999).

En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados (Kimball, Caserta, & 2004).

En definitiva, una solución BI completa permite:

- Observar ¿qué está ocurriendo?
- Comprender ¿por qué ocurre?
- Predecir ¿qué ocurriría?
- Colaborar ¿qué debería hacer el equipo?
- Decidir ¿qué camino se debe seguir?



Figura 8. *Ruta de un BI*

(Sinnexus, 1999). Elementos de un BI.

Extracción, transformación y carga ETL. ETL es un conjunto de procesos que recuperan datos desde un conjunto de sistemas fuente, los transforman y los cargan en un sistema de destino. La transformación puede ser usada para cambiar los datos de acuerdo al formato y criterio del sistema destino, generando nuevos valores en el sistema destino. La mayoría de sistemas ETL tienen mecanismos para limpieza de los datos antes de agregarlos, basándose en reglas de calidad de los datos (Rainardi, 2007).

Los sistemas ETL añaden valor a los datos de las siguientes formas:

- Remueven errores y corrigen datos vacíos.
- Proveen medidas de la confiabilidad de los datos.
- Almacenan el flujo de datos transaccionales de manera segura.
- Ajustan datos de múltiples fuentes para ser usados en conjunto.
- Estructuran datos para ser utilizados por herramientas de usuario final (Kimball, Caserta, & 2004).

La arquitectura básica de un sistema de ETL se muestra en la Figura:

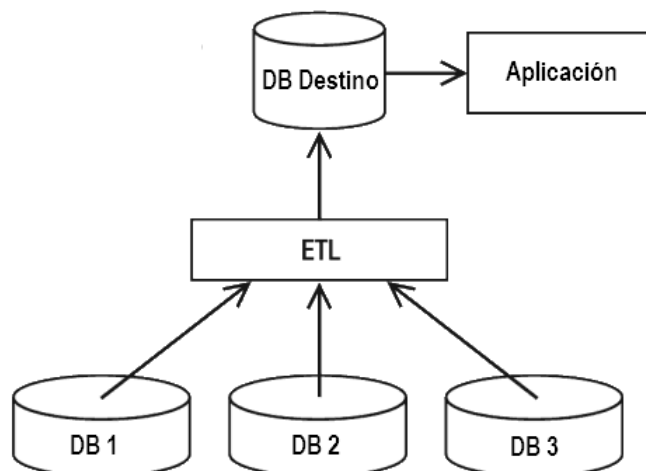


Figura 9. Arquitectura básica de un sistema ETL

Fuente: (C & Santín, 2007). Figura de la arquitectura de un sistema ETL. Técnicas y Herramientas.

El proceso de ETL, aunque no es estrictamente necesario para la minería de datos, suele ser necesario cuando se tratan grandes volúmenes de datos, cuando provienen de fuentes heterogéneas, cuando son cambiantes en el tiempo o cuando los algoritmos exigen ciertos formatos en la información de entrada (C & Santín, 2007).

Minería de Datos y Extracción de conocimiento a partir de datos .Debido al crecimiento en los volúmenes de información que experimentan las empresas y gracias a la disminución del costo en los equipos de cómputo y los sistemas de almacenamiento, muchas organizaciones cuentan hoy en día con unos vastos conjuntos de datos con información estratégica oculta, y que sin embargo, no puede ser descubierta utilizando los métodos de consulta o estadísticos tradicionales (Moore, 2014).

La minería de datos utiliza nuevas técnicas para buscar la información, muchas basadas en procesos estadísticos o de inteligencia artificial, que permite identificar patrones y relaciones en la información para crear modelos que permitan caracterizar y hasta predecir el comportamiento de esta; la extracción de conocimiento (KDD, por sus siglas en inglés) abarca la minería de datos, pero además incluye la preparación de la información y la interpretación de los resultados con el fin de dar significado al modelo planteado (Pujari, 2001).

El conocimiento no se obtiene simplemente por tener un conjunto de datos, estos solamente son la materia prima; que al dársele un significado dentro de un contexto pasa a ser información, y que una vez un experto elabora o identifica un modelo y permite interpretar esa información que aporte un nuevo valor, este pasa a ser conocimiento (Valhondo, 2003).

Etapas de KDD. Las etapas de KDD inician con un flujo de datos y finalizan con un conocimiento extraído a partir de esos datos.

- Selección: seleccionar o segmentar los datos de acuerdo a su relevancia con base en los objetivos planteados para la extracción de conocimiento.
- Preprocesado: se refiere a la limpieza de la información, eliminando información innecesaria, consolidando información y unificando formatos que provenga de múltiples fuentes.
- Transformación: consiste en transformar la información recolectada con el objeto de que pueda ser utilizable y navegable (consultar la información desde diferentes puntos de vista) preparándola para la etapa de minería de datos.
- Minería de datos: extracción de patrones en los datos.
- Interpretación y evaluación: convertir los patrones encontrados en conocimiento, que sirvan como fundamento para la toma de decisiones.
- Visualización de datos: busca ofrecer profundidad en el análisis y un entendimiento más intuitivo de los datos y sus patrones definidos en la minería de datos (Pujari, 2001).

La minería de datos es la principal etapa de KDD ya que se enfoca en la búsqueda de las relaciones y patrones ocultos, por lo que se profundiza en el presente documento.

Alcance de la minería de datos

- Predicción automatizada de tendencias y comportamientos.
- Descubrimiento automatizado de modelos previamente desconocidos.

La minería de datos produce cinco tipos de información:

- Asociaciones.
- Secuencias.
- Clasificaciones.
- Agrupamientos.
- Pronósticos (Pujari, 2001).

Áreas de investigación en la minería de datos

- Estadísticas: uno de los fundamentos de la minería de datos es la estadística, su amplia difusión y fundamentación teórica la valida, sin embargo, sus resultados deben ser interpretados por expertos debido a su complejidad en la interpretación. Se suelen usar modelos estadísticos como los lineales para llevar a cabo la minería de datos.
- Máquinas de aprendizaje: es un proceso automatizado de aprendizaje, tomando el aprendizaje como el equivalente a reglas basadas en observaciones, la generalización de comportamientos basada en ejemplos.

- Aprendizaje supervisado: es el aprendizaje basado en ejemplos, el cual produce una función que establece la correspondencia entre la entrada y la salida deseada (ejemplo).
- Aprendizaje sin supervisión: es el aprendizaje basado en la observación y el descubrimiento, por lo que no existe un conocimiento a priori, no existe una categorización de la información.
- Programación matemática (Jiawei, Micheline, & Jian, 2012).

Técnicas de minería de datos. Las técnicas más representativas de la minería de datos son:

- Descubrimiento de reglas de asociación: utilizado para descubrir hechos que ocurren en común dentro de un contexto, del tipo, cuando un cliente compra un producto X, suele comprar el producto Y.
- Agrupamiento o Clustering: trata de agrupar datos en diferentes grupos, de tal forma que los datos en cada grupo tengan tendencias o patrones comunes.
- Descubrimiento de reglas de clasificación o Árboles de decisión: sirven para representar un conjunto de reglas que determinan como se clasifican los elementos, donde cada nueva rama contiene una condición para el conjunto de datos que agrupa.
- Redes neuronales: forma de aprendizaje basada en la forma en que funciona el sistema nervioso de los animales, basado en la interconexión de neuronas con estímulos ponderados.
- Regresión lineal: rápida en la búsqueda de relaciones de datos pero limitado su uso a espacios bidimensionales. (Jiawei, Micheline, & Jian, 2012)

Clustering. El Clustering consiste en agrupar una colección dada de patrones no etiquetados en un conjunto de grupos. En este sentido, las etiquetas están asociadas con los grupos, pero las categorías se obtienen únicamente de las propiedades de los datos (H Witten, Eibe, & A. Hall, 2011).

- Definición. “El clustering representa la división de datos en grupos de objetos similares llamados clúster” (Acharya., 2003). Clustering es también conocido como clasificación no supervisada, en donde no se tienen asignación de grupos a clases ya predefinidas, sino que los grupos se van creando de acuerdo a las características de los datos.

Los grupos o clústers, son un conjunto de objetos que comparten características similares y juegan un papel muy importante en la manera en como la gente analiza y describe el mundo que los rodea. De forma natural, el humano se encarga de dividir objetos en grupos (clustering) y asignar objetos particulares a dichos grupos (clasificación) (H Witten, Eibe, & A. Hall, 2011).

Clustering es una de las técnicas más útiles para descubrir conocimiento oculto en un conjunto de datos. En la actualidad el análisis de clustering en minería de datos ha jugado un rol muy importante en una amplia variedad de áreas tales como: reconocimiento de patrones, análisis de datos espaciales, procesamiento de imágenes, cómputo y multimedia, análisis médico, economía, bioinformática y biometría principalmente. Esto ha hecho posible que el análisis de clustering se considere como una de las mejores técnicas para obtener conocimiento y realizar exploraciones en los datos (Mining, 2000).

- Análisis de clustering. Un problema de análisis de clustering, parte de un conjunto de casos u objetos cada uno de los cuales está caracterizado por varias variables (Ver Tabla 8.

Campos discretizados de la edad).

A partir de dicha información se trata de obtener grupos de objetos, de tal manera que los objetos que pertenecen a un grupo sean muy homogéneos entre sí y, por otra parte, la heterogeneidad entre los distintos grupos sea muy elevada. Expresado en términos de variabilidad hablaríamos de minimizar la variabilidad dentro de los grupos para al mismo tiempo maximizar la variabilidad entre los distintos grupos (H Witten, Eibe, & A. Hall, 2011).

Tabla 3. Casos de partida para el análisis de clustering

$C_1 \dots C_i \dots C_k$
$O_1 \ x_1^1 \ \dots x_i^1 \ \dots x_k^1$
.....
$O_2 \ x_2^j \ \dots x_i^j \ \dots x_k^j$
.....
$O_n \ x_1^N \ \dots x_i^N \ \dots x_k^N$

Denotando por $O = O_1, \dots, O_N$ al conjunto de N objetos, se trata de dividir O en k grupos o clústers, C_1, \dots, C_k .

A partir del planteamiento de un problema de clustering, las actividades del análisis de clustering típicamente involucran los siguientes pasos:

- Representación de patrones. Se refiere al establecimiento del número de clases, número de patrones, y el número, tipo y tamaño de las características disponibles para el algoritmo de clustering.
- Definición de proximidad. La proximidad de los patrones es usualmente medida por una función distancia entre un par de datos.

- Clustering. La etapa de agrupamiento puede desarrollarse en un gran número de formas. Se pueden utilizar agrupamientos de clústers jerárquicos, particionales y otros más abarcan métodos probabilísticos o de teoría de grafos.
- Abstracción de datos. Es el proceso de extraer una representación simple y compacta del conjunto de datos.
- Verificación de resultados. Consiste en validar el análisis de clustering realizado evaluando los resultados obtenidos. (A. K. Jain, 1999) (p.3)

Características de los algoritmos de clustering Las características deseables de la mayoría de los algoritmos de clustering son las siguientes:

- Escalabilidad. La mayoría de los algoritmos de clustering trabajan de manera apropiada con un número pequeño de observaciones (hasta 200 aproximadamente), mientras que se necesita una gran escalabilidad para realizar agrupamiento de datos en bases con millones de observaciones.
- Habilidad para trabajar con distintos tipos de atributos. Muchos algoritmos se han diseñado para trabajar sólo con datos numéricos, mientras que en una gran cantidad de ocasiones, es necesario trabajar con atributos asociados a tipos numéricos, binarios, discretos y alfanuméricos.
- Descubrimiento de clústers con formas arbitrarias. La mayoría de los algoritmos de clustering se basan en la distancia euclidiana, lo que tiende a encontrar clústers todos con forma (circular) y densidad similares. Es importante diseñar algoritmos que puedan establecer clústers de formas arbitrarias.

- Requerimientos mínimos en el conocimiento del dominio para determinar los parámetros de entrada. La herramienta no debería solicitarle al usuario que introduzca la cantidad de clases que quiere considerar, ya que dichos parámetros en muchas ocasiones no son fáciles de determinar, y esto haría que sea difícil controlar la calidad del algoritmo.
- Habilidad para tratar con datos ruidosos. La mayoría de las BD contienen datos con comportamiento extraño, datos faltantes, desconocidos o erróneos. Algunos algoritmos de clustering son sensibles a tales datos y pueden derivarlos a clústers de baja calidad.
- Insensibilidad al orden de las observaciones de entrada. Algunos algoritmos son sensibles al orden en que se consideran las observaciones. Por ejemplo, para un mismo conjunto de datos, dependiendo del orden en que se ingresen los valores se dará el resultado.
- se analicen, los clústers devueltos pueden ser diferentes. Es importante entonces que el algoritmo sea insensible al orden de los datos, y que el conjunto de clústers devuelto sea siempre el mismo.
- Alta dimensionalidad. Una BD o DW (DataWarehouse) puede contener varias dimensiones o atributos, por lo que es bueno que un algoritmo de clustering pueda trabajar de manera eficiente y correcta no sólo en repositorios con pocos atributos, sino también en repositorios con un alto espacio dimensional, o gran cantidad de atributos.
- Clustering basado en restricciones. Es un gran desafío el agrupar los datos teniendo en cuenta no sólo el comportamiento, sino también que satisfagan ciertas restricciones.
- Interpretación y uso. Los usuarios esperan que los resultados del clustering sean comprensibles, fáciles de interpretar y de utilizar. (H Witten, Eibe, & A. Hall, 2011) (p.446)

Con estas características, se busca diseñar algoritmos más flexibles que sean capaces de manipular una gran variedad de requerimientos de acuerdo a las necesidades de los usuarios

Técnicas de clustering. Los algoritmos de agrupación de clustering varían entre sí por las reglas heurísticas que utilizan y el tipo de aplicación para el cual fueron diseñados. La mayoría de ellos se basa en el empleo sistemático de distancias entre vectores (objetos a agrupar) así como entre clústers o grupos que se van formando a lo largo del proceso de clustering. Las características básicas por las que los algoritmos de clustering pueden ser clasificados son en función de:

- El tipo de dato que manejan (numérico, categórico y/o mixto).
- El criterio utilizado para medir la similitud entre los puntos.
- Los conceptos y técnicas de clustering empleadas (ej. lógica difusa, estadísticas) (H Witten, Eibe, & A. Hall, 2011).

En la literatura existen una gran cantidad de técnicas de clustering que varían de acuerdo a la arquitectura que utilizan. Una clasificación general divide los algoritmos en: clustering particional, clustering jerárquico, clustering basado en densidad y clustering basado en Grid (A. K. Jain, 1999).

Para cada una de las categorías presentadas en la Figura 10. *Técnicas de Clustering*, existen una variedad de sub-clasificaciones que presentan algoritmos con diferentes técnicas para encontrar clústers en los datos. Algunas de ellas son:

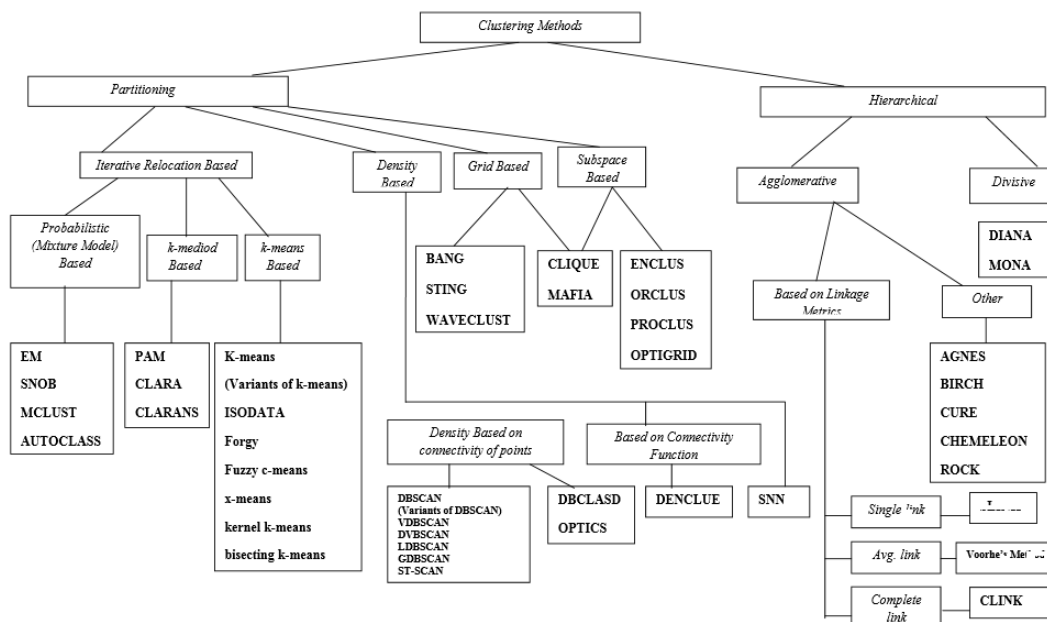


Figura 10. Técnicas de Clustering

Fuente: (A. K. Jain, 1999). Figura de las Técnicas de Clustering

- Técnicas estadísticas, basadas en la utilización de medidas de similitud y análisis estadístico para agrupar los datos.
- Técnicas conceptuales, basadas en la clasificación de características cualitativas de los datos.
- Técnicas excluyentes, basadas en el agrupamiento de datos sin traslape, es decir un dato única y exclusivamente puede pertenecer a una sola clase. La mayoría de los algoritmos de clustering se basan en esta técnica.
- Técnicas con traslapes, basadas en técnicas de lógica difusa que consideran grados de pertenencia en los datos. Los objetos pueden pertenecer a más de una clase (H Witten, Eibe, &

A. Hall, 2011).

Muchas técnicas de clustering están disponibles a un alto nivel, sin embargo, las técnicas de clustering se pueden dividir en enfoques jerárquicos y enfoques particionales. Los enfoques particionales se acercan intentando dividir o particional los datos en regiones o clústers donde los puntos de una región o de un grupo son más similares a los puntos de la misma región o clúster, mientras que es diferente de los puntos en otros (Jiawei han, 2012).

Un enfoque de partición divide el espacio en regiones y calcula la distancia de los puntos al centro de la región. Moviendo los centros utilizando técnicas para minimizar las distancias de centro a los puntos se puede encontrar una partición óptima. Ejemplos de este tipo de algoritmo son medios K-Means y K-Medoids. Estos enfoques son más fuertes cuando los puntos están dispuestos en clúster de forma esférica de cada uno de densidad similar (Mining, 2000).

Los enfoques jerárquicos son más adecuados para algunos problemas, incluyendo situaciones en las que los puntos de datos no están dispuestas en clústers esféricos. Estos también pueden dar muy buenos resultados, pero puede ser más costoso computacionalmente y sufrir de la desventaja de que una vez que una operación de división de combinación se ha realizado, no se puede revertir y esto puede causar problemas en los que los datos cuando estos tienen ruido (Mining, 2000).

K-Means Método basado en centroides. Supongamos que un conjunto de datos (data set), D , contiene n objetos en el espacio euclidiano. Métodos de particionamiento distribuyen los objetos en D en k clústers, C_1, \dots, C_k , de esta manera C_i pertenece a D y $C_i \cap C_j = \emptyset$ para $(1 \leq i, j \leq k)$. Una función objetivo se utiliza para evaluar la calidad de partición de modo

que los objetos dentro de un clúster son similares entre sí, pero diferentes a los objetos en otro clúster. Esto es, la función objetivo apunta a una alta similitud entre clústeres y baja similitud intercluster (Jiawei han, 2012).

Una técnica de partición basada en centroide utiliza el centroide de un clúster, C_i , para representar ese clúster. Conceptualmente, el centro de un clúster es su punto central. El centroide puede ser definido de varias maneras, tales como por la media o medoide de los objetos (o puntos) asignado al clúster (H Witten, Eibe, & A. Hall, 2011).

La diferencia entre un objeto p pertenece a C_i y C_i , el representante del clúster, se mide por $\text{dist}(p, c_i)$, donde $\text{dist}(x, y)$ es la distancia euclidiana entre dos puntos x e y . La calidad del grupo C_i puede ser medida por la variación intracluster, que es la suma de error al cuadrado entre todos los objetos en C_i y el centroide c_i , definida como:

$$E = \sum_{i=1}^k \sum_{p \in C_i} \text{dist}(p, c_i)^2,$$

Donde E es la suma del cuadrado del error de todos los objetos en el data set; p es el punto en el espacio representando el objeto dado y c_i es el centroide del clúster C_i (ambos p y c_i son multidimensionales). En otras palabras, para cada objeto en el clúster, la distancia del objeto al centro del clúster es elevado al cuadrado y las distancias sumadas. Esta función trata de hacer que el clúster k resulte lo más compacto y separado posible (H Witten, Eibe, & A. Hall, 2011).

Optimizando la variación dentro del clúster es computacionalmente desafiante. En el peor caso, se tendría que enumerar una serie de posibles particiones que son exponenciales al número

de clústers y comprobar los valores de variación intraclúster. Se ha mostrado que el problema es NP-hard en el espacio euclidiano inclusive para dos clústers (ej. $K=2$). Adicional, el problema es NP-hard para un numero general de clústers k incluso en espacio euclidiano 2D.

Si el número de clúster K y la dimensionalidad del espacio d son corregidos, el problema puede ser resuelto en un tiempo $O(n^{dk+1} \log n)$, en donde n es el número de objetos. Para sobrepasar la prohibición del costo computacional para la solución exacta, se realiza un acercamiento codicioso. Un primer ejemplo es el algoritmo K-Means, el cual es simple y muy comúnmente utilizado (H Witten, Eibe, & A. Hall, 2011).

El algoritmo *K-Means* mejora iterativamente la variación dentro del clúster. Para cada clúster, se calcula la nueva media usando los objetos asignados al clúster en la iteración anterior. Todos los objetos son luego reasignados utilizando la media actualizada como los nuevos centros de los clústers. Las iteraciones continúan hasta que la asignación es estable, es decir, los clústers formados en la ronda actual son los mismos que los que se forman en la anterior ronda. El procedimiento de *K-Means* se resume en la

Figura 7

Figura 11. Algoritmo de particionado de K-Means

Algorithm: k -means. The k -means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input:

- k : the number of clusters,
- D : a data set containing n objects.

Output: A set of k clusters.

Method:

- (1) arbitrarily choose k objects from D as the initial cluster centers;
- (2) **repeat**
- (3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
- (4) update the cluster means, that is, calculate the mean value of the objects for each cluster;
- (5) **until** no change;

Figura 11. Algoritmo de particionado de K-Means

Fuente: (Jiawei han, 2012). Figura del Algoritmo K-Means.

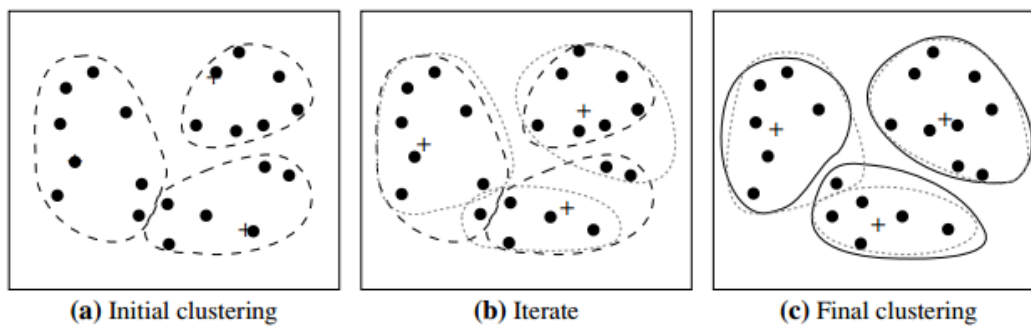


Figura 12. Clustering de un grupo de objetos utilizando K-Means; para (b) unos centros actualizados y una reasignación de objetos (La media de cada clúster está marcada por a+)

Fuente: (Jiawei han, 2012). Figura del Algoritmo K-Means.

K-Medoids técnica basada en representación de objetos. El algoritmo de k-medias es sensible a los valores atípicos debido a que tales objetos están lejos de la mayoría de los datos y por lo tanto, cuando se asigna a un clúster este puede distorsionar dramáticamente el valor medio del clúster. Esto afecta de forma inadvertida la asignación de otros objetos al clúster.

Este efecto es particularmente exacerbado debido a la utilización de la función del Error cuadrado (squared-error) y como se observa en el siguiente Ejemplo: (Jiawei han, 2012)(p.454)

- Un inconveniente de k-medias. Considere la posibilidad de seis puntos en el espacio 1-D que tiene los valores 1, 2, 3, 8, 9, 10 y 25 respectivamente. Intuitivamente, mediante inspección visual, podemos imaginar los puntos divididos en los clústers {1, 2, 3} y {8, 9, 10}, donde se excluye el punto 25 ya que parece ser un caso atípico. ¿Cómo K-Means particiona los valores? Si aplicamos K-Means utilizando $k = 2$ y la ecuación de la figura (9), el particionado {{1, 2, 3}, {8, 9, 10, 25}} tiene la variación dentro del clúster. (Jiawei han, 2012)(p.455)

$$(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2 + (8 - 13)^2 + (9 - 13)^2 + (10 - 13)^2 + (25 - 13)^2 = 196,$$

Dado que la media del Clúster {1, 2, 3} es 2 y la media de {8, 9, 10, 25} es 13. Compare esto con el particionado {{1, 2, 3, 8}, {9, 10, 25}} para que k-medias compute la variación del clúster interno como:

$$(1 - 3.5)^2 + (2 - 3.5)^2 + (3 - 3.5)^2 + (8 - 3.5)^2 + (9 - 14.67)^2 + (10 - 14.67)^2 + (25 - 14.67)^2 = 189.67,$$

Dado que 3,5 es la media del clúster {1, 2, 3, 8} y 14,67 es la media del clúster {9, 10, 25}.

La última partición tiene la variación más baja dentro del clúster; por lo tanto, el método K-Means asigna el valor 8 a un clúster distinto del que contenga 9 y 10 debido al punto externo 25. Por otra parte, el centro del segundo clúster (14.67), está sustancialmente lejos de todos los miembros del clúster. (Jiawei han, 2012)(p.456)

"¿Cómo podemos modificar el algoritmo de k-medias para disminuir dicha sensibilidad a los valores atípicos?"

En lugar de tomar el valor medio de los objetos en un clúster como punto de referencia, podemos recoger objetos reales para representar a los clústers, utilizando un objeto representante de cada clúster.

Cada objeto restante se asigna al clúster del que el objeto más representativo

Es el más similar. El método de partición se realiza entonces basa en el principio de minimizando la suma de las diferencias entre cada objeto \mathbf{p} y su correspondiente objeto representativo. Es decir, se utiliza un criterio de error absoluto, definido como,

$$E = \sum_{i=1}^k \sum_{\mathbf{p} \in C_i} \text{dist}(\mathbf{p}, \mathbf{o}_i),$$

Donde E es la suma del error absoluto para todos los objetos \mathbf{p} en el conjunto de datos, y \mathbf{O}_i es el objeto representante de C_i . Esta es la base para el método K-Means, en los cuales de objetos n en los clústers k minimizan el error absoluto.

Cuando $K=1$, podemos encontrar la mediana exacta en el tiempo $O(n^2)$.

Sin embargo, cuando k es un número positivo en general, el problema del método k -media es NP-hard. (Jiawei han, 2012)(p.457)

“¿Qué método es o más robusto: K -Means o K -Medoids?” El método K - Medoids es más robusto que k -medias en la presencia de ruido y los valores atípicos porque un medoide es menos influenciada por los valores atípicos u otros valores extremos que una media. Sin embargo, la complejidad de cada iteración en el algoritmo K -Medoids es $O(k(n-k)^2)$ (Jiawei han, 2012).

Para valores grandes de n y k , tal consumo computacional se vuelve muy costoso y mucho más costosos que el método K -Means. Ambos métodos requieren que el usuario especifique, el número de clústers k (Jiawei han, 2012).

Reglas de asociación. Las reglas de asociación juegan un papel importante en muchas aplicaciones de minería de datos, tratando de encontrar interesantes patrones en bases de datos. A fin de obtener estas asociaciones, los grupos de datos (data set) deben ser generados con anterioridad. Los algoritmos más comunes que se utilizan para este tipo de acciones son el A-priori (que generan los grupos de datos frecuentes y las reglas de asociación) y crea reglas de asociación FP-Growth (FP-Growth genera frecuentes juegos de artículos, que luego son utilizados para crear reglas de asociación y generar reglas de asociación) (Jared, 2014).

Aunque el algoritmo A-priori procesa los datos de una forma diferente a partir de los algoritmos FP-Growth y crea reglas de asociación eliminando los conjuntos de artículos (data set) que no son frecuentes (con un mínimo soporte, más pequeño que el mínimo apoyo especificado), existe una correlación significativa ($p < 0,05$) entre los resultados del proceso

generado a través del respectivo algoritmo, haciendo evidente a través de la línea de regresión , en el caso de soporte independiente, respectivamente a través de las líneas de regresión , en el caso de las tres variantes del valor *min_support* (Hunday, 2011).

Estrategias en las reglas de asociación. Las estrategias según varios estudios y los autores, es el uso de los dos algoritmos, el A priori y el FP-Growth. Una de las estrategias utilizadas es la de reducción del número de candidatos, el número de transacciones y el número de comparaciones, de esta manera obtener un buen conjunto de reglas de asociación (Bernal, 2012).

- Reducir en número de candidatos (M)
Uso de técnicas de poda
Ejemplo: Algoritmos A priori y DHP (Direct Hash)
- Reducir el número de transacciones (N)
Reducir N conforme aumenta el tamaño del itemset
Ejemplo: Algoritmo AprioriTID
- Reducir el número de comparaciones (NM)
Uso de estructuras de datos eficientes para almacenar los candidatos o las transacciones, de forma que no haya que comparar cada candidato con todas las transacciones (Bernal, 2012).

Adicionalmente se cuenta con varios criterios a tener en cuenta, como: accesos a la base de datos, costo computacional, tiempo de ejecución y rendimiento, los cuales se analizan para cada algoritmo y posteriormente seleccionar el más ajustado a los requerimientos del presente proyecto (Roberto & Marina, 2009).

Reglas de asociación en la transacción de negocios. Actualmente, con la masiva cantidad de datos que las organizaciones recolectan en sus procesos de negocio el descubrimiento de asociaciones interesantes en los registros de transacciones puede ayudar para la toma de decisiones en los procesos de marketing (Jared, 2014).

En el ejemplo típico para reglas de asociación, “el análisis de canasta de mercado”, supóngase que un granjero local ha puesto un stand de verduras y está ofreciendo los siguientes artículos: {espárragos, frijoles, brócoli, maíz, pimientos verdes, calabazas, tomates}, a este conjunto de artículos lo denotaremos I, y en la Tabla 4. Artículos comprados por los clientes, se mostrarán los artículos comprados (Larose D. T., 2005).

Tabla 4. Artículos comprados por los clientes

No	Contenido Canasta
1	Brócoli, pimienta, maíz
2	Espárragos, calabaza, maíz
3	Maíz, tomates, frijoles,
4	Pimienta, maíz, tomates, frijoles
5	Frijoles, espárragos, frijoles,
6	Calabaza, espárragos, frijoles,
7	Tomates, maíz,
8	Brócoli, tomates, pimienta
9	Calabaza, espárragos, frijoles
10	Frijoles, maíz
11	Pimienta, brócoli, frijoles,
12	Espárragos, frijoles, calabaza
13	Calabaza, maíz, espárragos,
14	Maíz, pimienta, tomates,

Fuente: Artículos comprados por clientes 1 Abril 2009

En el conjunto D de transacciones representadas en la Tabla 4. Artículos comprados por

los clientes⁴, cada transacción (T) en D representa un conjunto de artículos contenidos en I. Suponga que se tiene un conjunto particular de artículos A (Ej., frijoles y calabazas), y otro conjunto de artículos B (Ej., espárragos). Luego una regla de asociación toma la forma de (A =>B), donde el antecedente A y el consecuente B son subconjuntos propios de I, y A y B son mutuamente excluyentes.

Existen dos medidas asociadas a una regla de asociación: soporte y confianza, que le dan validez a la misma. El soporte para una regla de asociación particular A=>B es la proporción de transacciones en D que contienen A y B (Larose D. T., 2005).

$$\text{Soporte} = P(A \cap B) = \frac{\text{No. Transacciones Con A y B}}{\text{No. Total Transacciones}}$$

La confianza C de la regla de asociación A=>B es una medida de exactitud de la regla, determinada por el porcentaje de transacciones en D que contienen A y B (Larose D. T., 2005).

$$\text{Confianza} = P(B|A) = \frac{P(A \cap B)}{P(A)}$$

El analista puede preferir reglas que tengan alto soporte o alta confianza, o usualmente ambas. Las reglas fuertes son las que reúnen o superan ciertos soportes mínimos y criterios de confianza. Por ejemplo, un analista interesado en encontrar qué artículos del supermercado se compran juntos, puede establecer un nivel de soporte mínimo de 20% y un nivel de confianza mínimo del 70%. (Jared, 2014)

Por otro lado, en detección de fraude o de terrorismo, se necesitaría reducir el nivel de soporte mínimo a 1% o menos, ya que comparativamente pocas transacciones son fraudulentas o relacionadas con terrorismo.

Un itemset es un conjunto de artículos contenidos en I, y un k-itemset es un itemset que contiene k artículos; por ejemplo, { frijoles, calabazas } es un 2-itemset, y { brócoli, pimienta verde, maíz } es un 3-itemset, cada uno de los estantes de vegetales puestos en I. La frecuencia Φ del conjunto de artículos (itemset) es simplemente el número de transacciones que contienen el conjunto de artículos particular. Un conjunto de artículos frecuente es aquel que ocurre al menos un cierto mínimo número de veces, teniendo una frecuencia de conjunto de artículos, por ejemplo: suponiendo que $\Phi = 4$, los conjuntos de artículos que ocurren más de cuatro veces se dice que son frecuentes; denotamos el conjunto de K-ItemSets como F_k (Jared, 2014).

Las reglas de asociación para minería de grandes bases de datos son procesos de dos pasos:

- Encontrar todos los conjuntos de artículos frecuentes, es decir, aquellos con frecuencia $\geq \Phi$.
- Del conjunto de artículos frecuentes, generar reglas de asociación que satisfagan las condiciones mínimas de soporte y confianza (Larose D. T., 2005).

FP-Growth (frequent pattern growth: crecimiento de patrones frecuentes). Este algoritmo está basado en una representación de árbol de prefijos de una base de datos de transacciones llamada Frequent Pattern Tree (Jiawei han, 2012).

La idea básica del algoritmo FP-Growth puede ser descrita como un esquema de

eliminación recursiva: en un primer paso de preprocesamiento se borran todos los ítems de las transacciones que no son frecuentes individualmente o no aparecen en el mínimo soporte de transacciones, luego se seleccionan todas las transacciones que contienen al menos un ítem frecuente, se realiza esto de manera recursiva hasta obtener una base de datos reducida (Zaki & Meira, Jr, 2014).

Al retorno, se remueven los ítems procesados de la base datos de transacciones en la memoria y se empieza otra vez, y así con el siguiente ítem frecuente. Los ítems en cada transacción son almacenados y luego se ordena descendentemente su frecuencia en la base de datos. Después de que se han borrado todos los ítems infrecuentes de la base de datos de transacciones, se pasa al árbol FP. Un árbol FP es básicamente de prefijos para las transacciones, esto es: cada camino representa el grupo de transacciones que comparten el mismo prefijo, cada nodo corresponde a un ítem (Zaki & Meira, Jr, 2014).

Todos los nodos que referencian al mismo ítem son referenciados juntos en una lista, de modo que todas las transacciones que contienen un ítem específico pueden encontrarse fácilmente y contarse al atravesar la lista. Esta lista puede ser accesada a través de la cabeza, lo cual también expone el número total de ocurrencias del ítem en la base de datos (Jiawei han, 2012).

La evaluación de éste algoritmo es la siguiente:

- **Accesos en la base de datos:** este algoritmo no requiere de la generación de candidatos, por lo tanto, precisa de pocos accesos a la base de datos (Performance Evaluation of Apriori and FP-Growth Algorithms).

- Costo computacional: el algoritmo está basado en una representación de árbol de prefijos de una base de datos de transacciones, por lo tanto necesita de la creación de un árbol de prefijos; sin embargo, la creación de dicho árbol no requiere de un costo computacional elevado (Performance Evaluation of Apriori and FP-Growth Algorithms).
 - Tiempo de ejecución: este algoritmo busca patrones frecuentes con una corta búsqueda recursiva de prefijos, lo que en tiempo de ejecución es muy superior al del A priori, ya que no requiere de constantes accesos a la base de datos (Performance Evaluation of Apriori and FP-Growth Algorithms).
 - Rendimiento: puede generar un árbol FP-Tree de una base de datos proyectada si el árbol inicial no se puede alojar completamente en la memoria principal, lo que le permite adecuarse a los recursos disponibles (Performance Evaluation of Apriori and FP-Growth Algorithms).
- De acuerdo a lo anterior, se decide aplicar el algoritmo FP-Growth ya que tiene ventajas operacionales sobre los otros al no necesitar de la generación de ítems candidatos y ser computacionalmente más rápido (Jiawei han, 2012).

Entre las razones por la que se seleccionó este algoritmo es debido a que requiere de pocos accesos a la base o bodega de datos. Este algoritmo está basado en una representación de árbol de prefijos de una base de datos de transacciones; sin embargo, la creación de dicho árbol no requiere de un costo computacional elevado (Zaki & Meira, Jr, 2014), (Performance Evaluation of Apriori and FP-Growth Algorithms), (A comparative Analysis of Association Rules Mining Algorithms).

El algoritmo busca patrones frecuentes con una corta búsqueda recursiva de prefijos, lo

que en tiempo de ejecución es muy superior al Apriori, ya que no requiere constantes accesos a la base o bodega de datos. (Zaki & Meira, Jr, 2014) (Performance Evaluation of Apriori and FP-Growth Algorithms) (A comparative Analysis of Association Rules Mining Algorithms)

Tabla 5. Criterios de Evaluación de Algoritmos

	Accesos en la			Rendimiento
ALGORITMO	Base de Datos	Costo	Tiempo de Ejecución	
Apriori		Es costoso debido a		
	Requiere de varios accesos a	que el número de Subconjuntos	Consume mucho	Tiene algunas mejoras en el rendimiento. La memoria se puede
	Requiere de varios accesos a	La memoria puede	Los accesos a la BD pueden llegar a	volver insuficiente, afectando el
			La reducción de las operaciones	
ECLAT	Requiere de dos	Tiene un elevado	de E/S disminuye el	Tiene algunas mejoras en el
	Requiere de un	Tiene un elevado	agrupamiento que	Computacionalmente es
	acceso en la	costo en memoria. No requiere de	infiere en el En tiempo de	más eficiente que el A
	Requiere de	un costo	ejecución es muy	Se adecua fácilmente a los

Fuente: (Performance Evaluation of Apriori and FP-Growth Algorithms) .(2013). Tabla de criterios de evaluación de los Algoritmos de Reglas de Asociación 5-Mayo 2013.

Se observa que la técnica de reglas de asociación es la que más se adecúa para el

desarrollo del presente proyecto, en el caso de nuestra investigación se quiere descubrir relaciones entre los viajeros de la ciudad de Pereira.

Además esta técnica sugiere una búsqueda por toda la base de datos, realizando una clasificación en cada barrido, por lo tanto no hay límite establecido para la cantidad de datos que puede manejar, busca las características presentes en los viajes realizados, las cuales pueden tener atributos de diferentes tipos, por lo tanto no es necesario hacer una conversión a un tipo de datos específico.

La capacidad predictiva de la técnica depende de las medidas establecidas de confianza y soporte, ya que esta técnica se basa en el conteo de ocurrencias posibles entre las combinaciones de ítems en la tabla de transacciones, y posee gran escalabilidad ya que realiza un barrido por la base de datos, por lo que puede operar sin mayores problemas con un número grande de datos.

Desarrollo del Proceso de Minería de Datos con el modelo CRISP-DM

Comprensión del Dominio

Objetivos institucionales de los encargados de turismo. Los objetivos de las personas encargadas de turismo son los mismos que se plantean en la investigación, es decir, objetivos de establecimiento de tendencias y características de los viajeros a la ciudad de Pereira.

Especificando algunos, tenemos:

- Recolectar los datos de los viajeros.
- Clasificar los datos recolectados de los viajeros.
- Identificar los segmentos objetivos.
- Segmentar y describir hábitos de viajeros que visitan la ciudad de Pereira por medio de

un Análisis de Clúster para la industria Turística de Pereira.

- Establecer relaciones y tendencias existentes entre las características de los viajeros y los aspectos que rodean su visita a la ciudad de Pereira por medio de reglas de asociación.
- Determinar en los viajeros que visitan la Ciudad de Pereira el nivel de gastos, percepción y satisfacción luego de visitar la Ciudad por medio de un análisis de clúster.

Para el turismo es evidente el cambio del entorno y los requerimientos por parte de los proveedores y los consumidores. Los destinos cambian constantemente debido a situaciones, por ejemplo, después de los ataques del 11 de Septiembre (Korstanje, 2009).

Evaluación de la Situación

- Recursos disponibles: El único recurso disponible es el autor del proyecto, el cual diseñó las encuestas teniendo en cuenta modelos de encuestas, facilidad de uso y herramientas software para automatizar encuestas.
- Datos: Los datos con los que se cuenta son encuestas realizadas Online mediante su publicación en un sitio Web y redes sociales.

La información con la que se cuenta en dicha base de datos de manera es: (Ver

).

- Datos personales del viajero. (edad, sexo, procedencia, ocupación, estado civil, fuentes de ingreso).
- Datos sobre el viaje. (No. de veces que ha viajado, No. días permanencia, No. de compañeros de viaje, tipo de alojamiento, tipo de transporte, motivo viaje, actividades realizadas).
- Datos de satisfacción. (medios de transporte, sitios de alimentación, gasto estimado, calificación del gasto, aspectos de impacto al llegar, motivos para regresar).

Recursos Computacionales

Hardware: Los recursos computacionales para el procesamiento y análisis de la información, son proporcionados por el autor del proyecto:

- Computador portátil Acer 4750 Core i5 2.33 - 2.88 GHz con 8 GB de RAM.

Software: Se utiliza software de uso gratuito, como bases de datos MySQL, Herramientas de Office Online como Google Docs, Google refine para limpieza de datos, sistemas de minería de datos como RapidMiner, lenguajes y entornos de programación de aplicaciones estadísticas y minería de datos como Lenguaje R y R-Studio, entre otras que se requieren en alguna etapa del proceso CRISP-DM.

La publicación de los resultados no contiene información sensible de los viajeros.

Determinación de los objetivos de la minería de datos

- Segmentar y describir hábitos de viajeros que visitan la ciudad de Pereira por medio de un Análisis de Clúster para la industria Turística de Pereira.

- Establecer relaciones y tendencias existentes entre las características de los viajeros y los aspectos que rodean su visita a la ciudad de Pereira por medio de reglas de asociación.
- Determinar en los viajeros que visitan la Ciudad de Pereira el nivel de gastos, percepción y satisfacción luego de visitar la Ciudad por medio de un análisis de clúster.

Comprensión de los datos

Recolección de datos iniciales. El conjunto de datos se recolectó completamente Online mediante encuesta, dicha información se alojó posteriormente en un servidor local para hacer el análisis de los datos. La información se encuentra almacenada en una sola tabla con todos sus campos, ya que de esta manera se logró control sobre los datos ingresados, evitando que los encuestados dejaran datos incompletos o digitaran datos erróneos, esto no evita los datos poco comunes, como por ejemplo, un viajero con menos de 12 años y un ingreso ALTO, dichos datos como parte de este proyecto se considera “Outliers” o datos que alteran la muestra (Larose & Larose, 2015), pero forman parte de un grupo a analizar, otro caso son los campos en los cuales se puede ingresar un dato escrito (campo Otros).

Este tipo de datos se manejan en la etapa de preparación de los datos.

Cada usuario o viajero encuestado tiene un solo registro en la base de datos.

A continuación se describe el proceso de recolección de datos:

- Se diseñó la encuesta restringiendo los campos para que el encuestado solo pueda seleccionarlos dando clic con el mouse, no se pueden digitar.

La encuesta se diseñó en idioma español e inglés.

- El acceso a los datos se realizó mediante conexión al sitio web en donde se encontraba alojada la encuesta (Google Forms para diseñar la encuesta y mensajes privados, públicos y links en redes sociales como Facebook y tripadvisor.com).
- A partir de Google Forms se exportaron los datos directamente a Excel.
- Se utilizó MySQL de forma local para crear un repositorio de datos y poderlos manipular, modificar, realizar consultas, exportar a otros formatos y también

posteriormente conectarnos a ella mediante RapidMiner.

- Se exportaron las tablas que contienen la información personal, datos sobre el viaje y datos de satisfacción a un archivo Excel.
- Se validó la información corroborando que no se presentaran errores al exportar.

Descripción de los datos. En la siguiente tabla se muestran los campos de la encuesta:

Tabla 6 . Campos de datos personales de los encuestados

CAMPO	TIPO	TAMAÑO	DESCRIPCIÓN
Email	Texto	50	Email
Edad	Int	10	Edad
Sexo	Texto	10	Sexo
Ciudad	Texto	50	Ciudad de Origen
País	Texto	50	País de Origen
Ocupación	Texto	50	Independiente, empleado, pensionado, etc.
Estado civil	Texto	20	Soltero, casado, viudo, divorciado, unión libre.
Fuentes_de_ingreso	Texto	20	Empleado, independiente, pensionado.
Num_días_permanencia	Int	10	Numero
Num_companeros_viaje	Int	10	Numero

Fuente: Elaboración propia

Tabla 7. Campos de motivos de viaje y clasificación

CAMPO	TIPO	TAMAÑO	DESCRIPCIÓN
Núm_veces_viajado_a_Pereira	Int	10	Numero
Motivo_viaje	Texto	50	Turismo, comercio, aventura, educación, docencia, cultural, otras.
Tipo_sitio_alojamiento	Texto	50	Hotel, familia, amigos, Motel, Otros.
Actividad_desarrollada	Texto	50	Turismo, comercio, aventura, educación, docencia, cultural, otras.
Actividad_que_planeo_y_ejecuto	Texto	50	Turismo, comercio, aventura, educación, docencia, cultural, otras.
Medio_de_transporte_utilizado_viaje	Texto	50	Transporte utilizado para llegar: avión, transporte privado, publico, otro.
Medio_de_transporte_utilizado_interno	Texto	50	Transporte utilizado de uso interno: avión, transporte privado, publico, otro.

Lugares_de_alimentación	Texto	50	Restaurante, comida rápida, calle, familia, amigos, otros.
Gasto_estimado	Long	10	Valor en pesos o dólares.
Calificación_del_gasto	Texto	50	Muy bajo, bajo, medio, medio alto, alto, muy alto.
Aspecto_que_lo_impacto_de_la_ciudad	Texto	50	Contaminación, amabilidad, seguridad, inseguridad, ruido, tranquilidad, indigencia, educación, poca educación, racismo, violencia, mucho desarrollo infraestructura, poco desarrollo infraestructura, clima agradable, café, otros.
Aspecto para regresar a la ciudad	Texto	50	

Fuente: Elaboración propia

Ficha Técnica: Viajeros a la ciudad de Pereira antes de validar

- Número total de registros: 50.000.
- Total de atributos: 42
- Países: *de todos los continentes.*

- Fecha de visita: *antes del 2000 y después del 2000.*
- Origen de viajero: Nacional e internacional
- Tipo de viajero: Turista o visitante.
- Edades: Desde los 12 años de edad y menos hasta más los 95 años y mucho más.
- Nivel socioeconómico: Todos se tienen en cuenta.
- Nivel de estudios: Todos se tienen en cuenta.
- Religión: No se tiene en cuenta.

Exploración y Validación de los datos. Para el proceso de limpieza inicial se utilizó la herramienta Google Refine, que permite explorar de manera ágil los datos y realizar ajustes con múltiples opciones disponibles (Google, 2013).

La encuesta se modificó (Ver

) con respecto a la primera versión (Ver

) ya que en un principio al recolectar los datos se encontraron muchas ambigüedades, texto incompleto lo cual dificultaba el determinar los países, aspectos de la ciudad y aspectos para regresar a la ciudad.

La encuesta se validó con anterioridad para evitar redundancia de datos, datos incompletos o error al ingresar la información, las medidas que se tomaron fueron:

- Los datos se validaron para evitar información incompleta, es decir, se puede seleccionar de listas y radio buttons.
- En la encuesta todos los campos eran obligatorios.
- El campo edad va desde 13 años a 90 años, una instancia para los menores de 13 años y una instancia para los mayores de 90 años.
- El campo país se encuentra sin ciudad, de esta manera poder clasificar más fácilmente a los viajeros.
- En el campo: número de veces que ha viajado a la ciudad se colocó una instancia (0) cero para determinar que es el primer viaje realizado a la ciudad, esto se aclaró en el formato.
- Se modificaron los nombres de las columnas de datos de las preguntas 25 a la 27(Nivel de satisfacción) y cada sub pregunta de la 25 a la 27(Escala de satisfacción 1-5) dando un total de 18 columnas, esto, para evitar errores al exportar los datos a MySQL.

Se listan las circunstancias encontradas al examinar y validar los datos:

- En algunas preguntas de la encuesta se agregó la opción de ingresar los datos manualmente, es decir, una opción denominada “Otros”.
- Dicha opción “otros” se filtró y modifíco con el software Google refine, ya que se

observaron en algunas muestras datos no relacionados con la encuesta o el tipo de pregunta.

Luego de realizar las anteriores validaciones quedo un total de 5123 registros para continuar con el proceso CRISP-DM

Viajeros a la ciudad de Pereira luego de validar

- Número total de registros: 5.123.
- Total de atributos: 42
- Países: de todos los continentes.
- Fecha de visita: antes del 2000 y después del 2000.
- Tipo de viajero: Nacional e internacional
- Tipo de viajero: Turista o visitante.
- Edades: Desde los 12 años de edad y menos hasta más los 95 años y mucho más.
- Nivel socioeconómico: Todos se tienen en cuenta.
- Nivel de estudios: Todos se tienen en cuenta.
- Religión: No se tiene en cuenta.

Preparación de los datos

Selección de los datos y Limpieza de datos. Para la selección y limpieza de la información se continuó el uso de Google Refine y creo una base de datos en MySQL en la cual aplicar ajustes a los datos mediante la creación de varias tablas para cada objetivo del proyecto.

- La primera tabla contiene todo los datos limpios de las encuestas. (Datos contenidos en las 27 preguntas).

- La segunda tabla contiene todos los datos personales y datos del viaje de cada uno de los viajeros. (Preguntas 1-8 de datos personales y de la 9-21 sobre datos del viaje).
- La tercera tabla contiene campos: datos principales, datos de nivel de gastos y datos de percepción sobre los aspectos de la ciudad (Preguntas 1-3, 7-27)

(Ver

) encuesta utilizada.

Las Tablas 5 a la 28 muestran como fueron discretizados los valores de los diferentes atributos usados, todos los datos de la encuesta se discretizaron con su respectiva columna de “Valor” y “numero”, ya que algunos algoritmos no permiten valores nominales, algunas tablas de datos no se encuentran en el documento pero si se encuentran en el archivo Excel de viajeros, el proceso utilizado en RapidMiner se describe a

continuación:

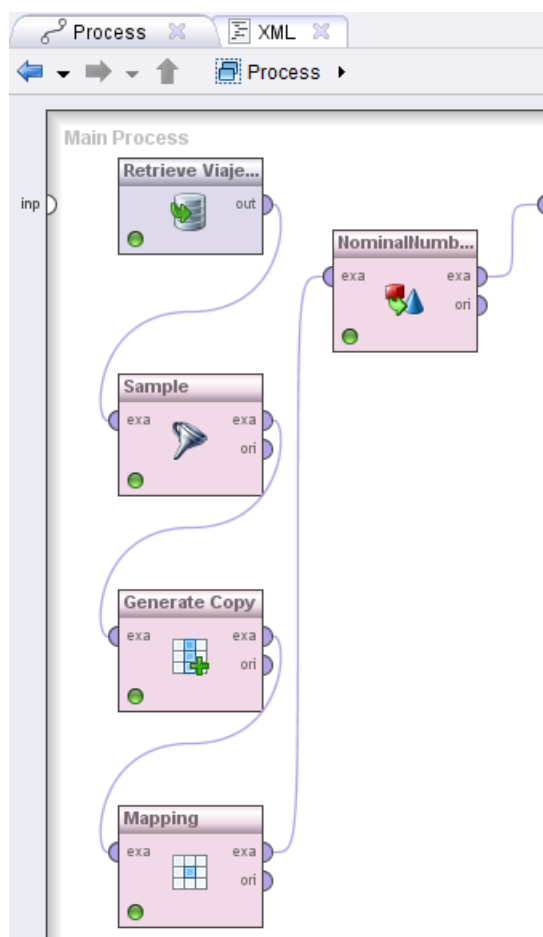


Figura 13. Proceso Normalizar datos con pesos numéricos.

Fuente: Resultado obtenido desde el software RapidMiner.

Descripción del proceso:

- Retrieve Viajeros: (ReadExcel) Realiza la lectura del archivo CSV ya procesado.
 - Sample: En este proceso podemos seleccionar una muestra de todo el conjunto de datos del DataSet.
 - Generate Copy: Realiza una copia de todo el DataSet, de esta manera tener el dato original y el Nuevo dato numérico normalizado.
 - Mapping: Definimos todos los parámetros para realizar la normalización, en este proceso los parámetros son los datos Nominales con su valor numérico, para tener más claridad observar una de las tablas normalizadas, Ej. Tabla 8. Campos discretizados de la edad
- NominalNumbersToNumerical: Convierte los datos numéricos con formato texto en un formato numérico para poder ser utilizados por los algoritmos de minería de datos.

Los datos resultantes se pueden observar en la descripción de las siguientes tablas:

Tabla 8. Campos discretizados de la edad

2.- Edad	Valor	Peso.
Menos de 12 y hasta 17 años.	MENOR_18	1
Entre 18 y 21	ENTRE_18_21	2
Entre 22 y 30	ENTRE_22_30	3
Entre 31 y 40	ENTRE_30_40	4
Entre 41 y 50	ENTRE_40_50	5
Entre 51 y 60	ENTRE_50_60	6

Entre 61 y 70	ENTRE_60_70	7
Entre 71 y 80	ENTRE_70_80	8
Entre 81 y 90	ENTRE_80_90	9
Mayor de 91 incluyendo la edad de 91.	MAYOR_90	10

Fuente: Elaboración propia.

Tabla 9. Campos discretizados del número de viajes a la ciudad

9.- Número de veces que ha viajado a la ciudad de Pereira	Valor	Peso.
Entre 0 y 1	MUY_BAJO	1
Entre 2 y 10	BAJO	2
Entre 11 y 20	MEDIO	3
Entre 21 y 30	ALTO	4
Más de 30	MUY_ALTO	5

Fuente: Elaboración propia.

Tabla 10. Campos discretizados de la fecha de viaje

10.- Fecha de Viaje a la Ciudad	Valor	Peso.
Fecha menor al año 2000	ANTES_2000	1
Fecha entre 2000 y 2003	ENTRE_2000_2003	2
Fecha entre 2004 y 2007	ENTRE_2004_2007	3

Fecha entre 2008 y 2011	ENTRE_2008_2011	4
Fecha entre 2012 y 2013	ENTRE_2012_2013	5
Actual	ANO_2014	6

Fuente: Elaboración propia.

Tabla 11. Campos discretizados del número de días de permanencia

12.- Número de días de permanencia	Valor	Peso.
Entre 0 y 1	MUY_BAJO	1
Entre 2 y 10	BAJO	2
Entre 11 y 20	MEDIO	3
Entre 21 y 30	ALTO	4
Más de 30	MUY_ALTO	5

Fuente: Elaboración propia.

Tabla 12. Campos discretizados del número de compañeros de viaje

13.- Numero de compañeros de viaje	Valor	Peso.
Entre 0 y 1	MUY_BAJO	1
Entre 1 Y 2	BAJO	2
Entre 3 y 10	MEDIO	3
Entre 11 y 20	ALTO	4

21 en adelante	MUY_ALTO	5
----------------	----------	---

Fuente: Elaboración propia.

Tabla 13. Campos discretizados del grado de satisfacción [Ocio]

25.- Grado de Satisfacción- Aspectos	Valor	Peso.
Generales sobre la ciudad. [Oferta de Ocio y Entretenimiento]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 14. Campos discretizados del grado de satisfacción [Oferta cultural]

25.- Grado de Satisfacción- Aspectos Generales sobre la ciudad.	Valor	Peso.
[Oferta Cultural]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 15. Campos discretizados del grado de satisfacción [Conservación Monumentos]

25.- Grado de Satisfacción- Aspectos	Valor	Peso.
Generales sobre la ciudad. [Conservación de los Monumentos]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 16. Campos discretizados del grado de satisfacción [Señalización turística]

25.- Grado de Satisfacción- Aspectos	Valor	Peso.
Generales sobre la ciudad. [Señalización Turística]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 17. Campos discretizados del grado de satisfacción [Cultura de la ciudadanía]

25.- Grado de Satisfacción- Aspectos	Valor	Peso.
Generales sobre la ciudad. [Cultura de la Ciudadanía]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 18. Campos discretizados del grado de satisfacción [Oferta de alojamiento]

26.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Alojamiento [Oferta de Alojamiento]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 19. Campos discretizados del grado de satisfacción [Calidad en la atención recibida]

26.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Alojamiento [Calidad en la atención recibida]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 20. Campos discretizados del grado de satisfacción [Disponibilidad alojamiento]

26.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Alojamiento [Disponibilidad de sitios de alojamiento]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 21. Campos discretizados del grado de satisfacción [Calidad del alojamiento]

26.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Alojamiento [Calidad del alojamiento]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 22. Campos discretizados del grado de satisfacción [Precio del alojamiento]

26.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Alojamiento [Precio del alojamiento]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 23. Campos discretizados del grado de satisfacción [Información oficina turismo]

27.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Información turística brindada por diferentes sectores. [Información en la oficina de turismo]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 24. Campos discretizados del grado de satisfacción [Información en el hotel]

27.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Información turística brindada por diferentes sectores. [Información en el hotel]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 25. Campos discretizados del grado de satisfacción [Información dada por transeúntes]

27.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Información turística brindada por diferentes sectores. [Información por personas en la calle]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 26. Campos discretizados del grado de satisfacción [Información dada por la policía]

27.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Información turística brindada por diferentes sectores. [Información por parte de la policía.]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 27. Campos discretizados del grado de satisfacción [Información dada por la policía de tránsito]

27.- Grado de Satisfacción sobre la ciudad. Información turística brindada por diferentes sectores. [Información por parte de la policía de tránsito]	Valor	Peso.
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 28. Campos discretizados del grado de satisfacción [Información dada por la policía de carreteras]

27.- Grado de Satisfacción sobre la ciudad. Información turística brindada por diferentes sectores. [Información por parte de la policía de carreteras.]	Valor	Peso.
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 29. Campos discretizados del grado de satisfacción [Información turística en vallas, carteles, volantes, folletos, guías, etc.]

27.- Grado de Satisfacción sobre la ciudad.	Valor	Peso.
Información turística brindada por diferentes sectores. [Información turística en vallas, carteles, volantes, folletos, guías, etc.]		
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Fuente: Elaboración propia.

Tabla 30. Campos discretizados del país de origen por organización territorial

4.- País de origen	Valor
Norteamérica	NORTEAMERICA
América Central	CENTROAMERICA
América del Sur	SURAMERICA
Islas del caribe	ANTILLAS
África del Norte	AFRICA_DEL_NORTE
África Occidental	AFRICA_OCCIDENTAL
África Central	AFRICA_CENTRAL
África Oriental	AFRICA_ORIENTAL
África Meridional	AFRICA_MERIDIONAL
Asia del Norte	ASIA_DEL_NORTE
Asia Central	ASIA_CENTRAL
Asia Oriental	ASIA_ORIENTAL
Asia Suroriental	ASIA_SURORIENTAL
Asia Meridional	ASIA_MERIDIONAL
Oceanía	OCEANIA
Europa	EUROPA

Fuente: Elaboración propia.

Tabla 31. Campos discretizados del gasto estimado

21.- Gasto estimado	Valor	Peso.
\$1 - \$100 Dollars	MUY_BAJO	1
\$100 - \$300 Dollars		
\$300 - \$500 Dollars	BAJO	2
\$500 - \$700 Dollars	MEDIO	3
\$700 - \$900 Dollars		
\$900 - \$1.100 Dollars		
\$1.100 - \$1.500 Dollars	MEDIO_ALTO	4
\$1.500 - \$2.000 Dollars		
\$2.000 - \$3.000 Dollars	ALTO	5
\$5.000 or more dollars	MUY_ALTO	6

Fuente: Elaboración propia.

La Tabla 31. Campos discretizados del gasto estimado,Tabla 31. Campos discretizados del gasto estimado se discretiza teniendo en cuenta el salario mínimo en el País y en el resto de países de Suramérica, con este salario de referencia se determina qué valor es alto, que valor es medio y que valor es bajo.

La tabla se discretiza para análisis posteriores pero se aclara el nivel de gasto y su clasificación en la pregunta 22 del

, ya que es difícil determinar que es alto o bajo si no se pregunta directamente al viajero.

Viajeros a la ciudad de Pereira

- Número total de registros: 5.123.
- Total de atributos: 42
- Países: de todos los continentes.
- Fecha de visita: antes del 2000 y después del 2000.
- Tipo de viajero: Nacional e internacional
- Tipo de viajero: Turista o visitante.
- Edades: Desde los 12 años de edad y menos hasta más los 95 años y mucho más.
- Nivel socioeconómico: Todos se tienen en cuenta.
- Nivel de estudios: Todos se tienen en cuenta.
- Religión: No se tiene en cuenta.

Otra tarea muy importante para la limpieza de datos es la detección de Outliers, dichos Outliers son valores poco frecuentes o valores que se encuentran por fuera de los cuartiles, dichos valores se encuentran en los extremos, para este proceso de minería de datos, la detección de Outliers se realiza en la fase de

Fuente: Proceso Obtenido desde el software RapidMiner.

Modelado, ya que por las características de la investigación se requiere analizar los datos por partes.

Construcción de datos e Integración de datos. Esta etapa tuvo como objetivo construir un conjunto de datos o vista minable, los datos de cada viajero se estructuraron en un solo registro, dichos datos se normalizaron para evitar redundancias y poder realizar un análisis más efectivo, a continuación se aclara la estructura de los datos minables y el porqué de algunas normalizaciones.

Nuestra tabla consolidada contiene todos los datos de cada uno de los viajeros en un solo registro, estos fueron normalizados en la etapa anterior “ *Selección de los datos y Limpieza de datos* “. Los datos agrupados se relacionan a continuación:

- Información de identificación del viajero: nombre, edad, sexo, país de origen, estado civil, nivel de estudios.
- Información financiera del viajero: Ocupación, fuentes de ingreso.
- Información sobre el viaje: número de veces que ha viajado a la ciudad, fecha del viaje, motivo del viaje, número de días de permanencia, número de compañeros de viaje, como

se enteró de la ciudad, lugar de alojamiento durante el viaje, actividad planeada desarrollada a cabalidad.

- Información sobre los medios de transporte utilizados en el viaje: medio de transporte utilizado para llegar a la ciudad, medio de transporte utilizado para transportarse en la ciudad.
- Información sobre la alimentación en la ciudad: Tipo de sitio de alimentación, su comida favorita.
- Información sobre el gasto: gasto estimado antes de la visita, calificación del gasto de la visita.
- Información sobre el impacto al llegar a la ciudad: aspecto que más lo impactó cuando llegó a la ciudad, aspecto que lo motivarían a regresar.
- Información sobre el grado de satisfacción: Sobre los aspectos de la ciudad, sobre el alojamiento, sobre información turística brindada a los viajeros.

La construcción de los datos agrupados, consolidados y normalizados se describen en los siguientes 4 pasos:

1. En la base de datos MySQL: el procesamiento inicial se realizó en la base de datos MySQL, debido al alto volumen de registros de viajeros (más de 50.000 registros).
2. En el software google refine se realizó la limpieza de los datos como se definió en la etapa anterior.
3. En una hoja de Excel en formato CSV se guardaron los 5.123 registros de la limpieza inicial, se guardó en dicho formato ya que la versión Community de RapidMiner no permite la

conexión directa a la base de datos MySQL. Dichos registros se exportaron ejecutando una consulta JOIN de todas las tablas en la base de datos y así unir todos los registros en una sola tabla.

4. En el software RapidMiner se diseña un proceso en el cual se importan los datos del archivo Excel con formato CSV. Ver Figura 14. Carga datos a RapidMiner

Descripción del proceso:

- Read Excel: Lee e importa los datos a RapidMiner desde una Fuente con formato CSV.

Ver Figura 15. *Importación de datos a RapidMiner*

Figura 14. Carga datos a RapidMiner



Fuente: Proceso Obtenido desde el software RapidMiner.

Figura 15. Importación de datos a RapidMiner

Data import wizard - Step 4 of 4

This wizard guides you to import your data.
Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.

Preview uses only first 100 rows.
 Date format:

	N	O	P	Q	R	S	T	U	V	W	X	Y
Attribute	polyno...	text	text	text	text	text	text	polyno...	text	text	text	text
Role	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute

Attribute	Role	Value
mpañero:	Como se en	Lugar alojan
DIO	Agencia de v	Amigos / Fri
JO	Empresa do	Familia / Far
DIO	Internet / Inte	Hotel / Hotel
JO	Agencia de v	Hostal / Hos
JO	Empresa do	Hostal / Hos
Y_BAJO	Television /	Hostal / Hos
DIO	Empresa do	Hostal / Hos
JO	Empresa do	Amigos / Fri
JO	Agencia de v	Hotel / Hotel
DIO	Agencia de v	Hostal / Hos
JO	Agencia de v	Amigos / Fri
Y_BAJO	Empresa do	Familia / Far
Y_BAJO	Internet / Inte	Hotel / Hotel
DIO	Agencia de v	Amigos / Fri

Ignore errors
 Show only errors

Row	Column	Error	Original value	Message
-----	--------	-------	----------------	---------

Fuente: Proceso Obtenido desde el software RapidMiner.

Modelado

Seleccionar técnicas de modelado. La primera técnica seleccionada para dar respuesta a nuestro primer objetivo y tercer objetivo, el cual es el de segmentar y describir los hábitos de los visitantes y viajeros a la ciudad de Pereira y segmentar de acuerdo con el nivel de satisfacción, debemos tener en cuenta las características de los datos, por tal motivo se realiza un análisis de Clustering.

La técnica de Clustering K-Means se seleccionó sobre otras ya que es un método no supervisado lo cual indica que no se requiere de un DataSet de entrenamiento sino que el algoritmo identifica grupos de acuerdo con sus características y los agrupa.

Con el proceso y modelo en RapidMiner de Clustering se agregaron otras tareas antes y después de ejecutar el Clustering:

- Análisis EDA, ver “Proceso inicial de análisis exploratorio de datos (*EDA*):”, es decir, un análisis exploratorio de los datos, se realiza con modelos matemáticos y se analiza de forma gráfica, con este proceso determinamos: distribución de variables y de datos, datos estadísticos, detección de Outliers. Dentro de nuestro análisis EDA realizamos uno que denominamos EDA_Sociodemografico, en el cual describimos la distribución de viajeros por el sexo, la edad, nivel de estudios y la situación laboral.
- Ejecutamos el modelo de Clustering para obtener grupos de datos o viajeros.
- Ejecutamos el modelo de Clustering para obtener grupos de datos o viajeros, dicho modelo se aplica con variables tipo Rol de Clúster y sin variables tipo Rol de Clúster.
- Posteriormente aplicamos reglas de asociación para cada Clúster con el fin de encontrar

y describir hechos que ocurren en común dentro del conjunto de datos, también diversas relaciones entre las variables.

- Adicional a las reglas de asociación, visualizamos los resultados mediante los gráficos de desviación y graficando los centroides de cada clúster, en algunos casos se visualizan los diagramas de árbol o arboles de decisión.

Una variante del uso de Clustering que analizamos es el planteado en el punto 3 del párrafo anterior, este hace referencia a definir una variable con un Rol Clúster con lo cual se determinan las relaciones entre las variables aplicadas a dicho rol, es decir, si tenemos un grupo de datos de estudiantes con sus calificaciones y un campo que indica si aprobó o no el semestre y requerimos agrupar de acuerdo a sus características de las notas, la variable con el rol Clúster para agrupar debería ser la variable que indica si aprobaron o no el semestre.

La segunda técnica seleccionada para determinar nuestro segundo objetivo el cual es describir las relaciones y tendencias entre las características de los viajeros y los aspectos de su visita, es el algoritmo FP-Growth por medio de reglas de asociación.

Las reglas de asociación no son tan diferentes a las reglas de clasificación excepto por que pueden predecir cualquier atributo, no solo la clase, dando la libertad de predecir la combinación de atributos.

Las reglas de asociación se aplican a un DataSet el cual permite determinar reglas y manejar datos de validación como la cobertura la cual es el número de instancias por las cuales se predice correctamente la regla y la confianza es el número de instancias que se predicen correctamente la regla.

Se aplican las técnicas de Clustering y reglas de asociación a los grupos de datos que se fueron filtrando poco a poco, de esta manera se buscaron las combinaciones que nos proporcionaron información valiosa de los viajeros.

Generar el diseño de prueba. Para validar la calidad del modelo y de los datos analizados, la calidad de los clústers generados, la medida adecuada de clústers de acuerdo con su distribución, la calidad de los datos normalizados, la calidad de las reglas de asociación y su nivel de confianza utilizamos varias medidas.

El diseño inicial (

) toma los datos del archivo CSV resultante del proceso de limpieza de la etapa de “Preparación de los datos” de modelo CRISP-DM, el algoritmo K-Means de Clustering se configura tomando un valor de $K=5$, $\text{max_runs} = 10$, tipo de medida= BregmanDivergences y divergencia= SquaredEuclidianDistance, siendo el valor de $k=5$ aleatorio para aplicarle a los 5.123 registros de los viajeros.

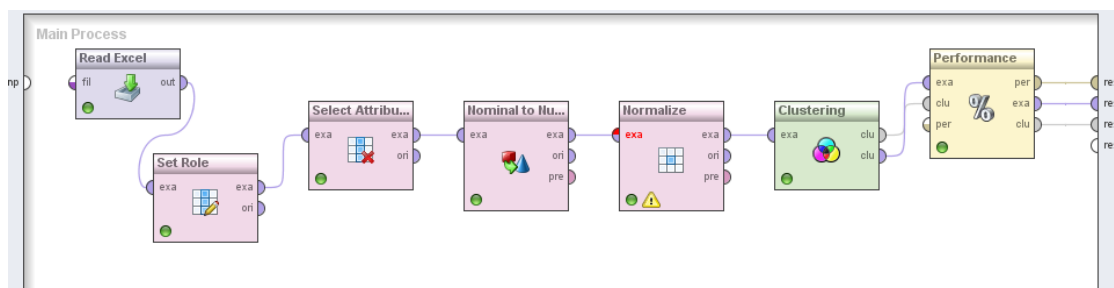


Figura 16. Proceso RapidMiner para el Clustering

Fuente: Proceso Obtenido desde el software RapidMiner.

El segundo modelo se encuentra en la

Figura 17, se toma el archivo completo, se analiza de forma gráfica y matemáticamente para posteriormente eliminar los Outliers, generándose un nuevo archivo de viajeros para analizar.

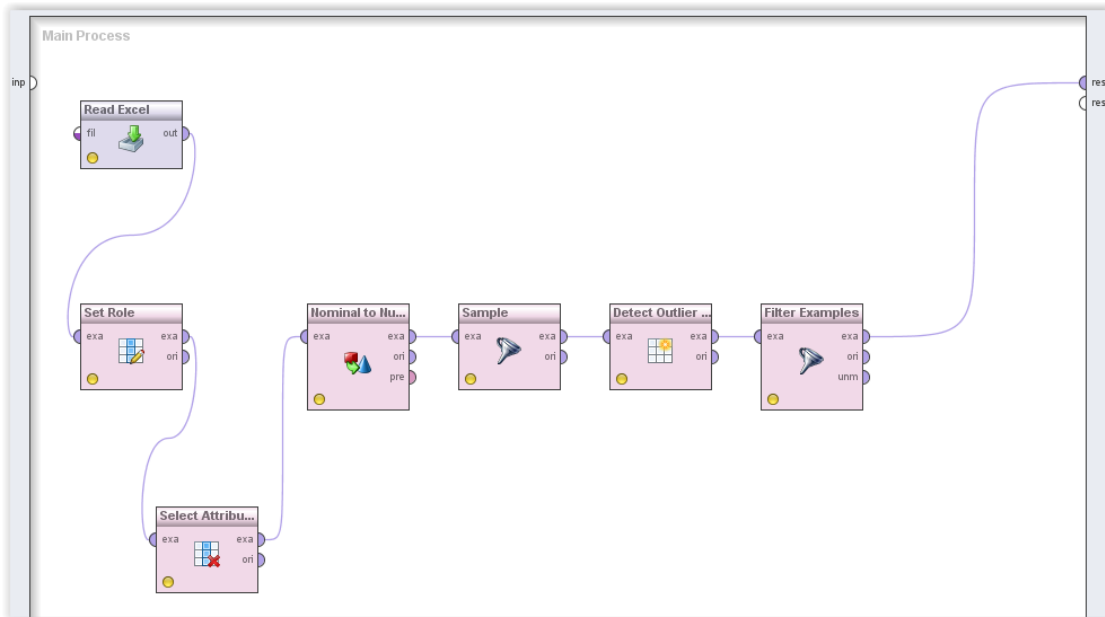


Figura 17. Filtrado y eliminación de Outliers.

Fuente: Proceso Obtenido desde el software RapidMiner.

En el tercer modelo Figura 42 se evalúan los clústers K seleccionados al azar, posteriormente con el modelo de la Figura 43 se valida la cantidad ideal de clústers para posteriormente volver a validar el modelo de la Figura 46, Figura 47. Validar el número óptimo de clústers proceso 2.y Figura 21. Validar el número óptimo de clústers proceso 2.

Los aspectos a evaluar se realizan con las medidas de validación internas especificadas para los clústers, teniendo en cuenta que un buen clúster es el cual todos los puntos están cercanos los unos a los otros y cada clúster debe estar separado del otro.

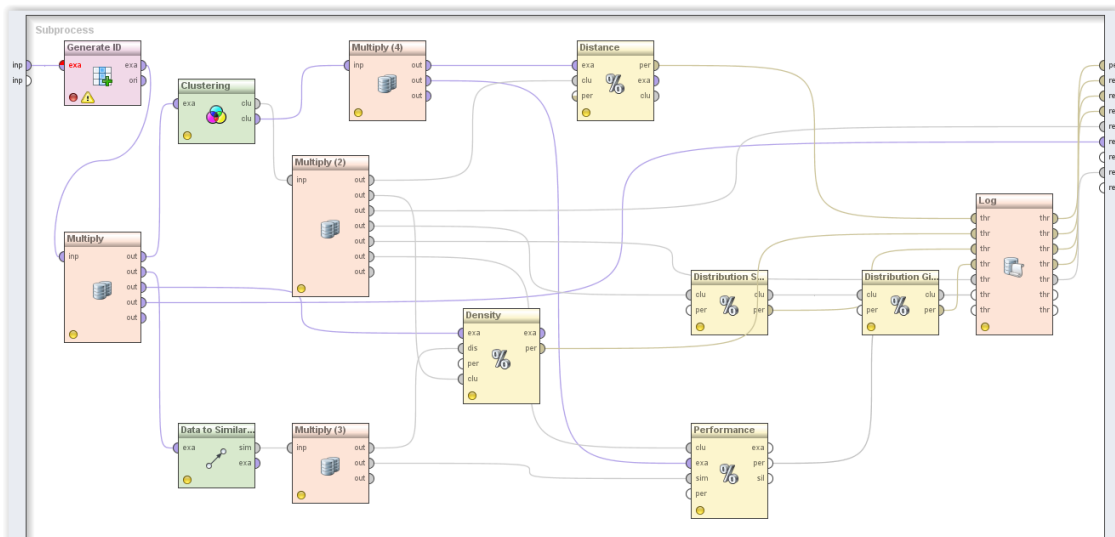


Figura 18. Validar el número óptimo de clústers para los datos proceso 1.

Fuente: Proceso Obtenido desde el software RapidMiner.

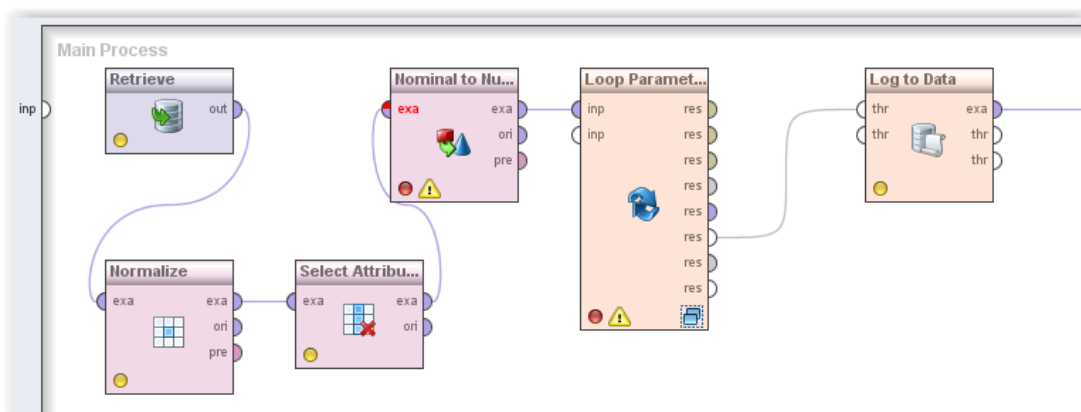


Figura 19. Ciclo para validar los clústers.

Fuente: Proceso Obtenido desde el software RapidMiner.

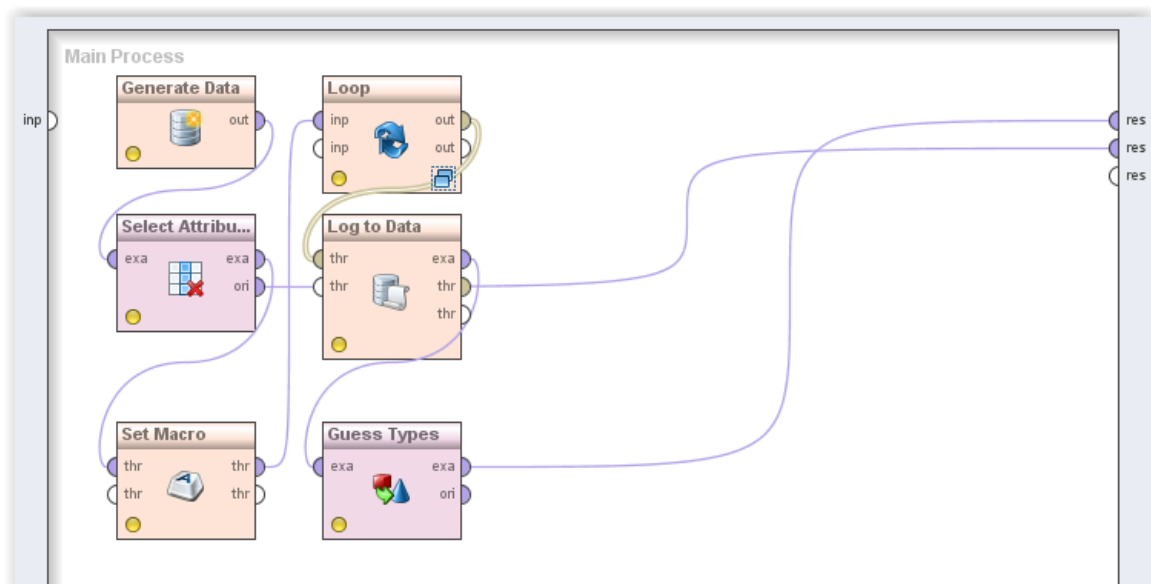


Figura 20. Validación con medidas internas parte 1.

Fuente: Proceso Obtenido desde el software RapidMiner.

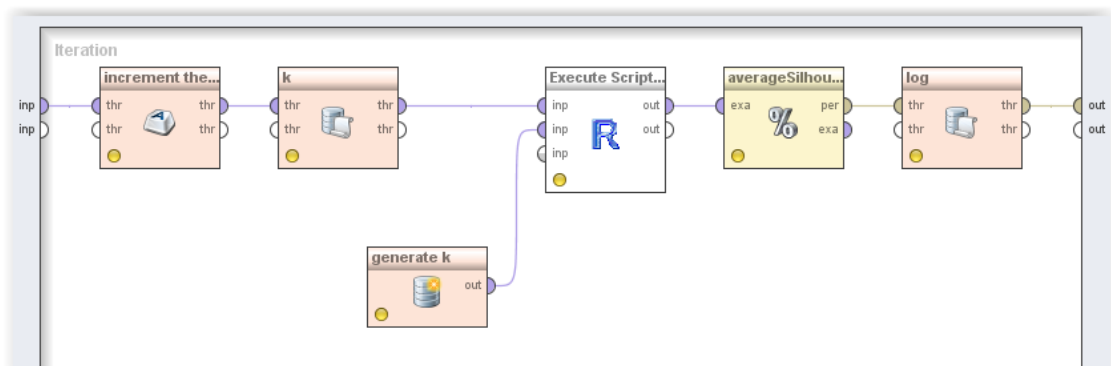


Figura 21. Validar el número óptimo de clústers proceso 2.

Fuente: Proceso Obtenido desde el software RapidMiner.

Las medidas internas de validación de clustering que se utilizaron: AverageWithin centroid

distance: Considera cada clúster y calcula la distancia promedio de cada punto en el clúster al centroide.

Clúster density: Considera cada clúster y encuentra la distancia promedio entre todos los pares de puntos en el clúster y los multiplica por el número de puntos en el clúster, el resultado es el equivalente a la distancia por punto entre el clúster, similar a la densidad.

Sum of squares ítem distribution: Considera cada clúster y divide el número de puntos en el clúster por el número de punto de datos en todos los clústers.

Distribution Gini (ítem distribution performance): Mide la diferencia entre valores de una frecuencia de distribución, ejemplo, niveles de salarios de personas.

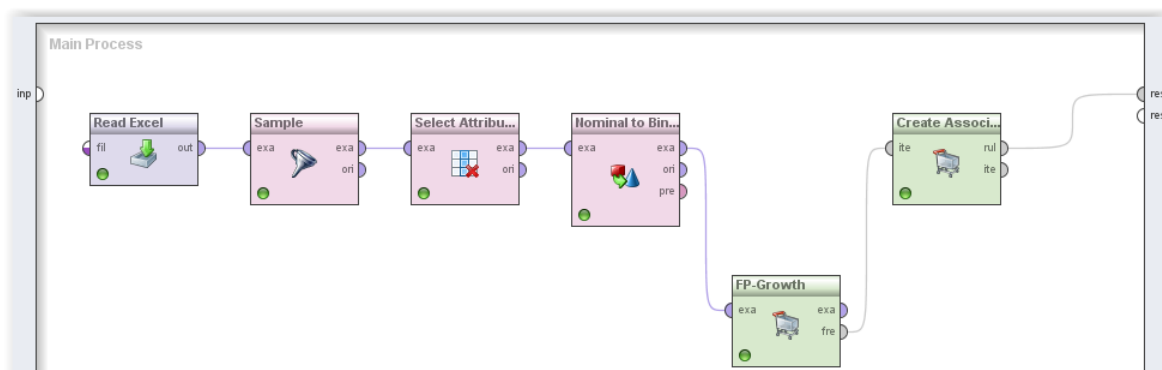
Davies Bouldin: Para cada clúster determina cual otro clúster tiene el máximo ratio de promedio intracluster para puntos en dos clústers a la distancia entre clústers. Obteniendo esta máxima, los promedia todos.

Performance Silhouette: Siendo el índice Silhouette el que nos da el valor de K con el mejor rendimiento para nuestros clústers, ya que es un método de interpretación validación de datos en los clústers (P, 1987).

Para la validación de las reglas de asociación tenemos en cuenta los siguientes aspectos:

- Las reglas de asociación descubren y describen hechos de los datos, como las relaciones entre un conjunto de datos y de esta manera encontrar las reglas que predicen la ocurrencia de un ítem.

- Las medidas reales para determinar una buena regla de asociación se dan con las medidas de Soporte y confianza, dichas medidas son el resultado de aplicar el algoritmo FP-Growth.
- Aplicamos FP-Growth a nuestro DataSet para obtener las relaciones y validar el soporte y la confianza, el diagrama se encuentra en la Figura 22. Proceso para generar reglas de



asociación.

Fuente: Proceso Obtenido desde el software RapidMiner.

Construcción del modelo

Desarrollo del Objetivo 1 : Segmentar y describir hábitos de viajeros que visitan la ciudad de Pereira por medio de un Análisis de Clúster para la industria Turística de Pereira

Para la aplicación del modelo se utilizó:

- Archivo CSV con el DataSet de los viajeros.

Figura 22. Proceso para generar reglas de asociación.

- R
api

dMiner Studio 6.2 Community Edition y RapidMiner 5.3.015 AGPL V3 de 64bits.

- Plugin: R Scripting.
- Plugin: R Extension.

- Plugin: Weka Extension.
- Plugin: Anomaly Detection.
- Plugin: Clustering.
- Plugin: WhiBo.
- R Language for Statistical Computing V3.2.2 64 bits.

Proceso inicial de análisis exploratorio de datos (EDA): El primer proceso que realizamos fue para determinar gráficamente como están distribuidos los datos, cual es la media, datos totales por categoría, variables representativas y adicionalmente determinamos Outliers, posteriormente determinamos como se agruparon los datos, es decir para realizar el primer análisis de Clustering. Ver Figura 23. Proceso de analisis Descriptivo.

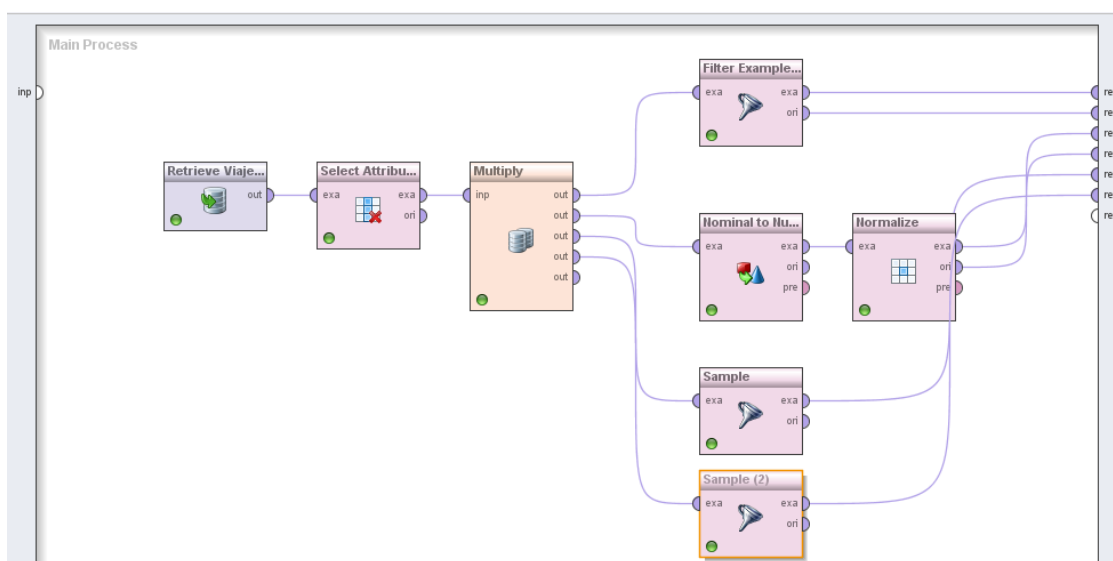


Figura 23. Proceso de analisis Descriptivo.

Fuente: Proceso Obtenido desde el software RapidMiner.

En el proceso anterior tenemos los operadores:

Retrieve ViajerosV3: Obtiene los datos de los viajeros almacenados con anterioridad en el repositorio local del computador.

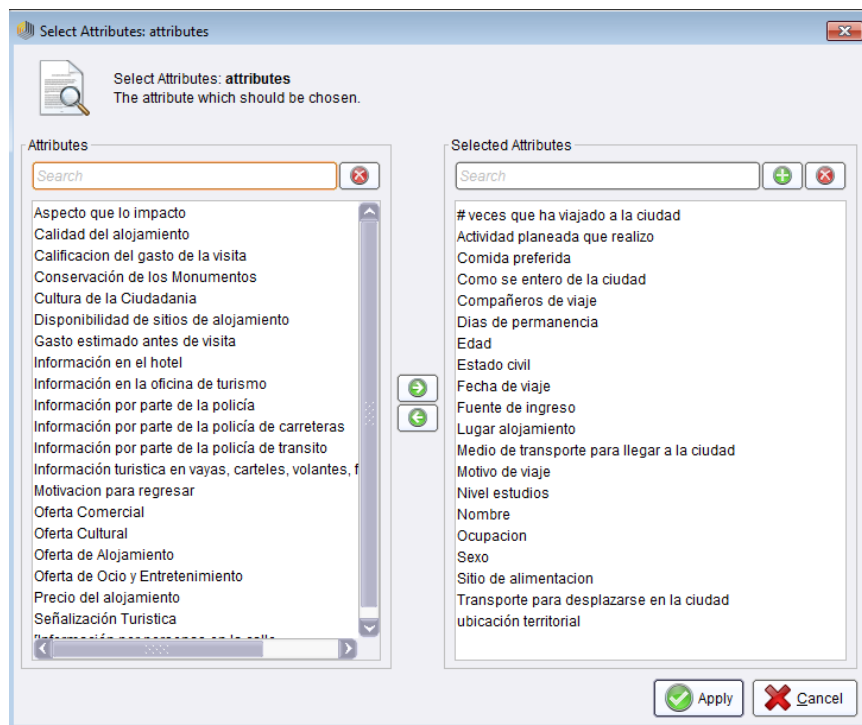


Figura 24 . Datos para analisis descriptivo.

Fuente: Proceso Obtenido desde el software RapidMiner.

Select Attributes: Seleccionamos únicamente los atributos que hacen referencia a los datos personales y datos que rodean la visita a la ciudad. Ver Figura 24 . Datos para analisis descriptivo.

Multiply: Divide el DataSet con los atributos seleccionados con anterioridad hacia 4 operadores: Filter Example Range, Nominal to Numerical, Sample y Sample 2.

Los 4 operadores determinan diferentes salidas del Dataset:

Filter Example Range: Permite seleccionar una muestra de los datos, en este caso seleccionamos únicamente los primeros 100 registros para su análisis.

Nominal to Numerical: Convierte todos los valores categóricos polinomiales seleccionados en la Figura 24 . Datos para analisis descriptivo. A datos numéricos para su posterior análisis. Dichos datos se normalizan mediante el proceso Normalize.

Sample: Se configura como “Absolute”, con lo cual se seleccionan los primeros 2500 registros del DataSet, no se seleccionaron de forma aleatoria.

Sample(2): Se seleccionó un sub DataSet del 0.6, es decir, del 60% de los datos del DataSet original de forma aleatoria.

Analizando los resultados del proceso tenemos la relación de los MetaDatos de los 4 resultados, en la Figura 25 . Metadata 5123 Viajeros. Observamos que la normalización realizada con anterioridad no permite que los datos se alteren y que esté muy bien definida la escala y la cantidad.

Data View Meta Data View Plot View Advanced Charts Annotations

ExampleSet (5123 examples, 1 special attribute, 19 regular attributes)

Name	Statistics
Nombre	avg = 2562 +/- 1479.027
ubicación territorial	mode = NORTEAMERICA (346), least = ANTILLAS (295)
Edad	mode = ENTRE_50_60 (731), least = ENTRE_80_90 (147)
Sexo	mode = M / Male (2571), least = F / Female (2552)
Estado civil	mode = Soltero / Single (1353), least = Divorciado / Divorced (1183)
Nivel estudios	mode = Técnico / Technician (644), least = Maestria / Masters (446)
Ocupacion	mode = Pensionado-Jubilado / Pensioner-Retired (1266), least = Empleado / Employee (921)
Fuente de ingreso	mode = Pensionado-Jubilado / Pensioner-Retired (1279), least = Empleado / Employment (571)
# veces que ha viajado a la ciudad	mode = MEDIO (1646), least = MUY_ALTO (164)
Fecha de viaje	mode = ENTRE_2012_2013 (880), least = ANO_2014 (813)
Motivo de viaje	mode = Deportes / Sports (878), least = Docencia / Teaching (52)
Dias de permanencia	mode = MEDIO (1971), least = MUY_ALTO (82)
Compañeros de viaje	mode = ALTO (2287), least = MEDIO (742)
Como se entero de la ciudad	mode = Empresa donde labora / Your employer. (1040), least = Television / TV (1009)
Lugar alojamiento	mode = Amigos / Friends (1336), least = Hotel / Hotel (1251)
Actividad planeada que realizo	mode = Aventura / Adventure tourism (866), least = Docencia / Teaching (826)
Medio de transporte para llegar a la ciudad	mode = Avión / Traveling by plane (1719), least = Transporte publico / Public transport (1696)
Transporte para desplazarse en la ciudad	mode = Bicicleta / Bicycle (1311), least = Caminar / Walk (1252)
Sitio de alimentacion	mode = Restaurantes / Restaurants. (1047), least = Comida rápida en la calle. / Fast food in the street. (997)
Comida preferida	mode = Carnes Rojas / Meat (1041), least = Pollo / Chicken (1008)

Figura 25 . Metadata 5123 Viajeros.

Fuente: Proceso Obtenido desde el software RapidMiner.

Al realizar el mismo ejercicio con el sub Dataset de 100 viajeros, el de 2500 y el de 3116 se puede observar la misma tendencia.

La presente investigación agrupa datos de acuerdo al Método Clustering y no determinamos anticipadamente valores para margen de error en estadística, además de que no conocemos la muestra total del país, ya que hay viajeros que ingresan por aire, otros por tierra, no sabemos si todos los que diligenciaron la encuesta son realmente extranjeros, no conocemos a ciencia cierta si todos respondieron adecuadamente, no sabemos cuántos de los viajeros son inmigrantes ilegales, ya que en las noticias podemos observar muchos casos en la región (Tiempo, 2015).

Por lo tanto tomamos los datos completos, en el Clustering se toman diferentes variables y se hacen experimentos, de esta manera decidir si analizar nuevamente todos los datos o tomar

una muestra, esto depende del análisis resultante y de las relaciones visibles en los histogramas.

Data View
 Meta Data View
 Plot View
 Advanced Charts
 Annotations

ExampleSet (2500 examples, 1 special attribute, 19 regular attributes)

Name	Statistics
Nombre	avg = 2565.947 +/- 1471.064
ubicación territorial	mode = AFRICA_ORIENTAL (182), least = ASIA_CENTRAL (128)
Edad	mode = ENTRE_30_40 (359), least = ENTRE_80_90 (72)
Sexo	mode = M / Male (1263), least = F / Female (1237)
Estado civil	mode = Casado / Married (650), least = Divorciado / Divorced (600)
Nivel estudios	mode = Ninguno / None (319), least = Doctorado / Doctorate (212)
Ocupacion	mode = Pensionado-Jubilado / Pensioner-Retired (630), least = Estudiante / Student (447)
Fuente de ingreso	mode = Pensionado-Jubilado / Pensioner-Retired (625), least = Empleado / Employment (278)
# veces que ha viajado a la ciudad	mode = MEDIO (794), least = MUY_ALTO (89)
Fecha de viaje	mode = ENTRE_2008_2011 (433), least = ANO_2014 (398)
Motivo de viaje	mode = Deportes / Sports (435), least = Docencia / Teaching (23)
Dias de permanencia	mode = MEDIO (976), least = MUY_ALTO (42)
Compañeros de viaje	mode = ALTO (1140), least = MUY_BAJO (343)
Como se entero de la ciudad	mode = Internet / Internet (519), least = Recomendado / Friend (483)
Lugar alojamiento	mode = Amigos / Friends (654), least = Hostal / Hostal (603)
Actividad planeada que realizo	mode = Comercio / Buy in stores (421), least = Docencia / Teaching (408)
Medio de transporte para llegar a la ciudad	mode = Transporte Privado / private transport (840), least = Transporte publico / Public transport (823)
Transporte para desplazarse en la ciudad	mode = Transporte publico / Public transport (653), least = Caminar / Walk (599)
Sitio de alimentacion	mode = Restaurantes / Restaurants (527), least = Comida rápida en restaurantes / Fast food in restaurants (476)
Comida preferida	mode = Mariscos / Seafood (519), least = Pollo / Chicken (465)

Figura 26. Metadata 2500 Viajeros

Fuente: Proceso Obtenido desde el software RapidMiner.

Data View Meta Data View Plot View Advanced Charts Annotations

ExampleSet (100 examples, 1 special attribute, 19 regular attributes)

Name	Statistics
Nombre	avg = 50.500 +/- 29.011
ubicación territorial	mode = SURAMERICA (11), least = ASIA_CENTRAL (2)
Edad	mode = ENTRE_60_70 (16), least = ENTRE_18_21 (4)
Sexo	mode = M / Male (52), least = F / Female (48)
Estado civil	mode = Viudo / Widowed (39), least = Casado / Married (14)
Nivel estudios	mode = Especialización / Specialization (27), least = Tecnología / Technologist (3)
Ocupacion	mode = Pensionado-Jubilado / Pensioner-Retired (33), least = Estudiante / Student (12)
Fuente de ingreso	mode = Pensionado-Jubilado / Pensioner-Retired (26), least = Empleado / Employment (10)
# veces que ha viajado a la ciudad	mode = ALTO (33), least = MUY_BAJO (5)
Fecha de viaje	mode = ENTRE_2004_2007 (20), least = ANTES_2000 (11)
Motivo de viaje	mode = Deportes / Sports (21), least = Docencia / Teaching (0)
Días de permanencia	mode = MEDIO (45), least = ALTO (0)
Compañeros de viaje	mode = MEDIO (47), least = ALTO (0)
Como se entero de la ciudad	mode = Recomendado / Friend. (27), least = Television / TV (13)
Lugar alojamiento	mode = Familia / Family (32), least = Hotel / Hotel (20)
Actividad planeada que realizo	mode = Comercio / Buy in stores (20), least = Cultural / Culture (12)
Medio de transporte para llegar a la ciudad	mode = Transporte publico / Public transport (38), least = Avión / Traveling by plane (26)
Transporte para desplazarse en la ciudad	mode = Bicicleta / Bicycle (37), least = Caminar / Walk (17)
Sitio de alimentacion	mode = Familia / Family (24), least = Amigos / Friends (15)
Comida preferida	mode = Pastas o pizzas / Italian food. (22), least = Mariscos / Seafood (18)

Figura 27 . Metadata 100 Viajeros.

Fuente: Resultado Obtenido desde el software RapidMiner.

Data View Meta Data View Plot View Advanced Charts Annotations

ExampleSet (3116 examples, 1 special attribute, 19 regular attributes)

Name	Statistics
Nombre	avg = 2556.425 +/- 1470.797
ubicación territorial	mode = SURAMERICA (219), least = ANTILLAS (179)
Edad	mode = ENTRE_30_40 (459), least = ENTRE_80_90 (87)
Sexo	mode = M / Male (1558), least = M / Male (1558)
Estado civil	mode = Soltero / Single (810), least = Divorciado / Divorced (719)
Nivel estudios	mode = Técnico / Technician (400), least = Doctorado / Doctorate (269)
Ocupacion	mode = Pensionado-Jubilado / Pensioner-Retired (769), least = Empleado / Employee (558)
Fuente de ingreso	mode = Pensionado-Jubilado / Pensioner-Retired (786), least = Empleado / Employment (338)
# veces que ha viajado a la ciudad	mode = MEDIO (1010), least = MUY_ALTO (102)
Fecha de viaje	mode = ENTRE_2012_2013 (548), least = ANO_2014 (501)
Motivo de viaje	mode = Deportes / Sports (542), least = Docencia / Teaching (27)
Días de permanencia	mode = MEDIO (1198), least = MUY_ALTO (55)
Compañeros de viaje	mode = ALTO (1398), least = MUY_BAJO (461)
Como se entero de la ciudad	mode = Internet / Internet (651), least = Agencia de viajes / Travel agency (597)
Lugar alojamiento	mode = Hostal / Hostal (798), least = Hotel / Hotel (751)
Actividad planeada que realizo	mode = Cultural / Culture (552), least = Docencia / Teaching (488)
Medio de transporte para llegar a la ciudad	mode = Avión / Traveling by plane (1063), least = Transporte publico / Public transport (1016)
Transporte para desplazarse en la ciudad	mode = Transporte publico / Public transport (805), least = Transporte privado / Private transport (756)
Sitio de alimentacion	mode = Familia / Family (642), least = Comida rápida en la calle. / Fast food in the street. (608)
Comida preferida	mode = Carnes Rojas / Meat (644), least = Pastas o pizzas / Italian food. (602)

Figura 28. Metadata 3116 Viajeros.

Fuente: Resultado Proceso Obtenido desde el software RapidMiner.

Ahora describimos las variables de datos personales, datos del viaje a la ciudad y datos sociodemográficos del DataSet original de 5123 viajeros. Ver Figura 29. Totales estadísticos de los

Name	Range
Nombre	[1.000 ; 5123.000]
ubicación territorial	ASIA_DEL_NORTE (324), ASIA_MERIDIONAL (320), SURAMERICA (335), ASIA_ORIENTAL (322), AFRICA_ORIENTAL (337), EUROPA (309), NORTEAMERICA (346), OCEANIA (334), CENTROAMERICA (330), ASIA_SURORIENTAL (304), AFRICA_CENTRAL
Edad	ENTRE_60_70 (724), ENTRE_70_80 (725), ENTRE_30_40 (724), ENTRE_50_60 (731), MENOR_18 (402), ENTRE_18_21 (317), ENTRE_40_50 (690), ENTRE_22_30 (663), ENTRE_80_90 (147)
Sexo	M / Male (2571), F / Female (2552)
Estado civil	Divorciado / Divorced (1183), Viudo / Widowed (1258), Soltero / Single (1353), Casado / Married (1329)
Nivel estudios	Especialización / Specialization (599), Secundaria / High school (577), Doctorado / Doctorate (450), Ninguno / None (640), Primaria / Elementary School (589), Universitario / College degree (582), Maestría / Masters (446), Técnico / Technician (644), Tec
Ocupación	Pensionado-Jubilado / Pensioner-Retired (1266), Desempleado / Unemployed (1058), Empleado / Employee (921), Estudiante / Student (941), Independiente / Independent (937)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (1279), Familia / Family-Parents (782), Estudiante / Student (802), Independiente / Independent (783), Desempleado / Unemployed (906), Empleado / Employment (571)
# veces que ha viajado a la ciudad	ALTO (1583), MEDIO (1646), BAJO (1415), MUY_ALTO (164), MUY_BAJO (315)
Fecha de viaje	ANO_2014 (813), ENTRE_2012_2013 (880), ANTES_2000 (853), ENTRE_2008_2011 (858), ENTRE_2004_2007 (842), ENTRE_2000_2003 (877)
Motivo de viaje	Deportes / Sports (878), Turismo / Tourism (852), Educación / Education (832), Aventura / Adventure tourism (861), Comercio / Buy in stores (855), Cultural / Culture (793), Docencia / Teaching (52)
Días de permanencia	MEDIO (1971), MUY_BAJO (334), BAJO (1894), ALTO (842), MUY_ALTO (82)
Compañeros de viaje	MEDIO (742), BAJO (1339), MUY_BAJO (756), ALTO (2287)
Como se entero de la ciudad	Agencia de viajes / Travel agency (1017), Empresa donde trabaja / Your employer (1040), Internet / Internet (1026), Television / TV (1009), Recomendado / Friend (1031)
Lugar alojamiento	Amigos / Friends (1336), Familia / Family (1269), Hotel / Hotel (1251), Hostal / Hostel (1267)
Actividad planeada que realizo	Cultural / Culture (863), Comercio / Buy in stores (844), Aventura / Adventure tourism (866), Docencia / Teaching (826), Educación / Education (860), Turismo / Tourism (864)
Medio de transporte para llegar a la ciudad	Transporte publico / Public transport (1696), Avión / Traveling by plane (1719), Transporte Privado / private transport (1708)
Transporte para desplazarse en la ciudad	Transporte publico / Public transport (1295), Transporte privado / Private transport (1265), Bicicleta / Bicycle (1311), Caminar / Walk (1252)
Sitio de alimentación	Familia / Family (1026), Amigos / Friends (1034), Comida rápida en la calle / Fast food in the street (997), Comida rápida en restaurantes / Fast food in restaurants (1019), Restaurantes / Restaurants (1047)
Comida preferida	Comida picante o condimentada / Spicy food (1018), Mariscos / Seafood (1035), Pollo / Chicken (1008), Carnes Rojas / Meat (1041), Pastas o pizzas / Italian food (1021)



Viajeros.

Fuente: Resultado Obtenido desde el software RapidMiner.

Ya que los datos que analizamos hacen referencia a personas con género, edad, nivel de estudios, ingresos, situación laboral, el análisis lo denominaremos EDS_sociodemografico, dicho estudio contiene varias tablas en las cuales podremos observar que contienen datos de frecuencias y una tabla de frecuencias relativas, a continuación:










Figura 29. Totales estadísticos de los Viajeros.

Tabla 32. Sociodemográfico de Género.

<i>Sexo</i>	<i>Total de Viajeros</i>	<i>%</i>
<i>M / Male</i>	 2571	50,2%
<i>F / Female</i>	 2552	49,8%
<i>Totales:</i>	5123	100,0%





Fuente: Elaboración propia.

Tabla 33. Sociodemográfico de Edades.

<i>Rango de Edades</i>	<i>Total de Viajeros</i>	<i>%</i>
<i>MENOR_18</i>	 402	7,8%
<i>ENTRE_18_21</i>	 317	6,2%
<i>ENTRE_22_30</i>	 663	12,9%
<i>ENTRE_30_40</i>	 724	14,1%
<i>ENTRE_40_50</i>	 690	13,5%
<i>ENTRE_50_60</i>	 731	14,3%
<i>ENTRE_60_70</i>	 724	14,1%
<i>ENTRE_70_80</i>	 725	14,2%
<i>ENTRE_80_90</i>	 147	2,9%
<i>Totales:</i>	5123	100,0%










Fuente: Elaboración propia.

Tabla 34. Sociodemográfico de Estado Civil.

<i>Estado Civil</i>	<i>Total de Viajeros</i>	<i>%</i>
<i>Soltero</i>	 1353	26,4%
<i>Casado</i>	 1329	25,9%
<i>Viudo</i>	 1258	24,6%
<i>Divorciado</i>	 1183	23,1%
<i>Totales:</i>	5123	100,0%






Fuente: Elaboración propia.

Tabla 35. Sociodemográfico de Nivel de Estudios.

<i>Nivel de Estudios</i>	<i>Total de Viajeros</i>	<i>%</i>
<i>Ninguna</i>	 640	12,5%
<i>Secundaria</i>	 577	11,3%
<i>Primaria</i>	 589	11,5%
<i>Técnico</i>	 644	12,6%
<i>Tecnólogo</i>	 596	11,6%
<i>Pregrado</i>	 582	11,4%
<i>Especialización</i>	 599	11,7%
<i>Maestría</i>	 446	8,7%
<i>Doctorado</i>	 450	8,8%
<i>Totales:</i>	5123	100,0%







Fuente: Elaboración propia.

Tabla 36. Sociodemográfico de Ocupación.

<i>Ocupación</i>	<i>Total de Viajeros</i>	<i>%</i>
<i>Estudiante</i>	 941	18,4%
<i>Empleado</i>	 921	18,0%
<i>Independiente</i>	 937	18,3%
<i>Pensionado-Jubilado</i>	 1266	24,7%
<i>Desempleado</i>	 1058	20,7%
<i>Totales:</i>	5123	100,0%

















Fuente: Elaboración propia.

Tabla 37. Sociodemográfico de Fuente de Ingresos.

Fuente de Ingresos	Total de Viajeros	%
Estudiante	 802	15,7%
Familia	 782	15,3%
Empleado	 571	11,1%
Independiente	 783	15,3%
Pensionado-Jubilado	 1279	25,0%
Desempleado	 906	17,7%
Totales:	5123	100,0%

Fuente: Elaboración propia.

Tabla 38. Sociodemográfico de Ubicación Territorial.

Ubicación Territorial	Total de Viajeros	%
ASIA_DEL_NORTE	 324	6,3%
ASIA_MERIDIONAL	 320	6,2%
SURAMERICA	 335	6,5%
ASIA_ORIENTAL	 322	6,3%
AFRICA_ORIENTAL	 337	6,6%
EUROPA	 309	6,0%
NORTEAMERICA	 346	6,8%
OCEANIA	 334	6,5%
CENTROAMERICA	 330	6,4%
ASIA_SURORIENTAL	 304	5,9%
AFRICA_CENTRAL	 321	6,3%
AFRICA_DEL_NORTE	 311	6,1%
AFRICA_OCCIDENTAL	 312	6,1%
ANTILLAS	 295	5,8%
ASIA_CENTRAL	 314	6,1%
AFRICA_MERIDIONAL	 309	6,0%
Totales:	5123	100,0%

Fuente: Elaboración propia.

A continuación se describen varios Histogramas con 2 variables relacionadas, la variable género, la variable edad y las demás variables de datos personales consideradas anteriormente, con dichos diagramas tendremos más claridad sobre el perfilamiento y características de los viajeros.

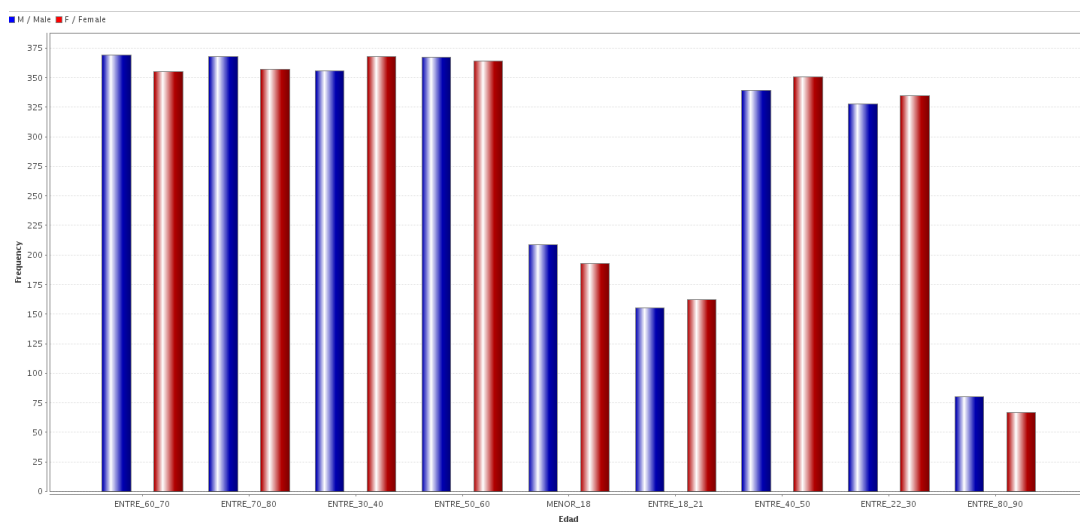


Figura 30. Histograma Genero-Edad.

Fuente: Resultado obtenido desde el software RapidMiner.

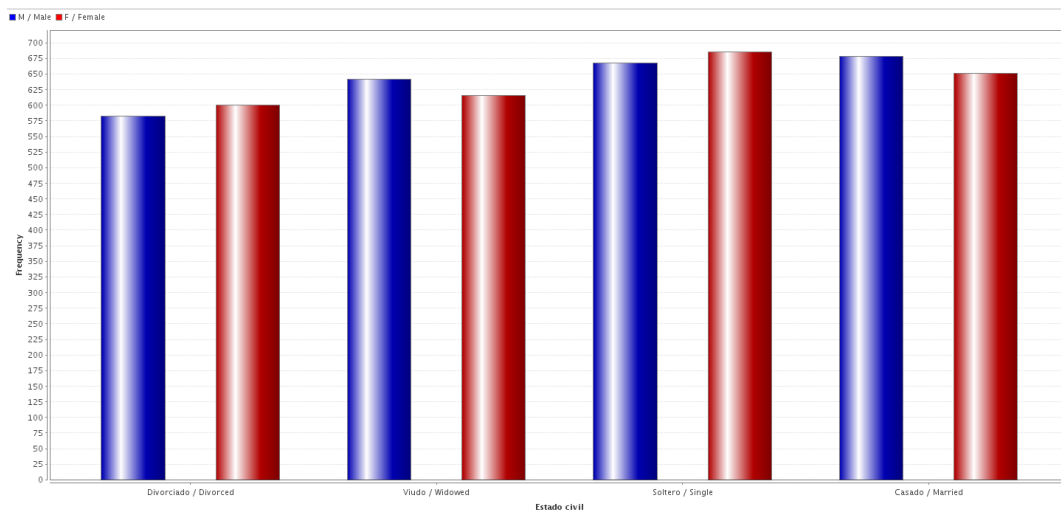


Figura 31. Histograma Genero-Estado Civil.

Fuente: Resultado obtenido desde el software RapidMiner.

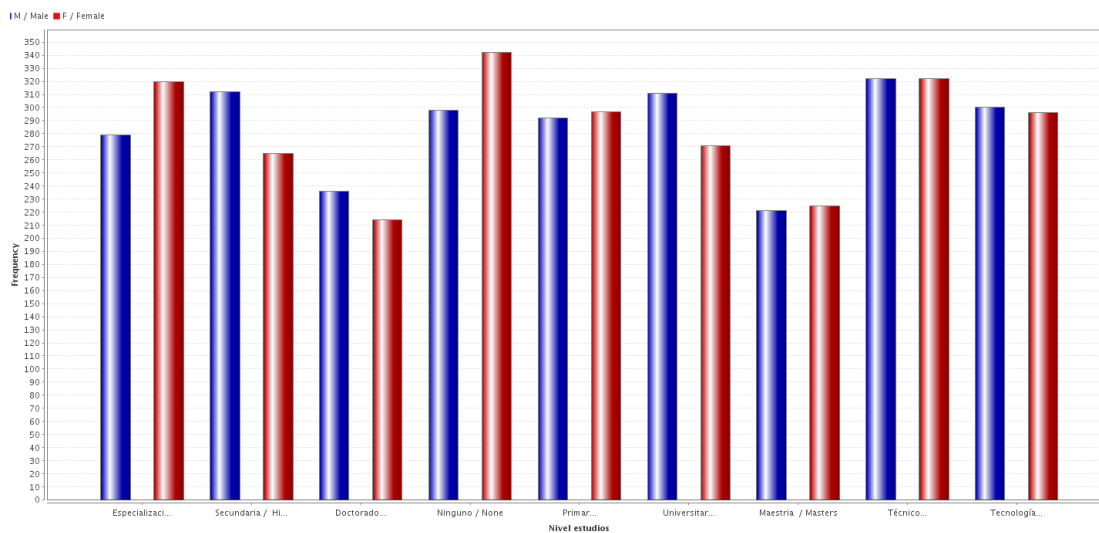


Figura 32. Histograma Genero-Nivel de Estudios

Fuente: Resultado obtenido desde el software RapidMiner.

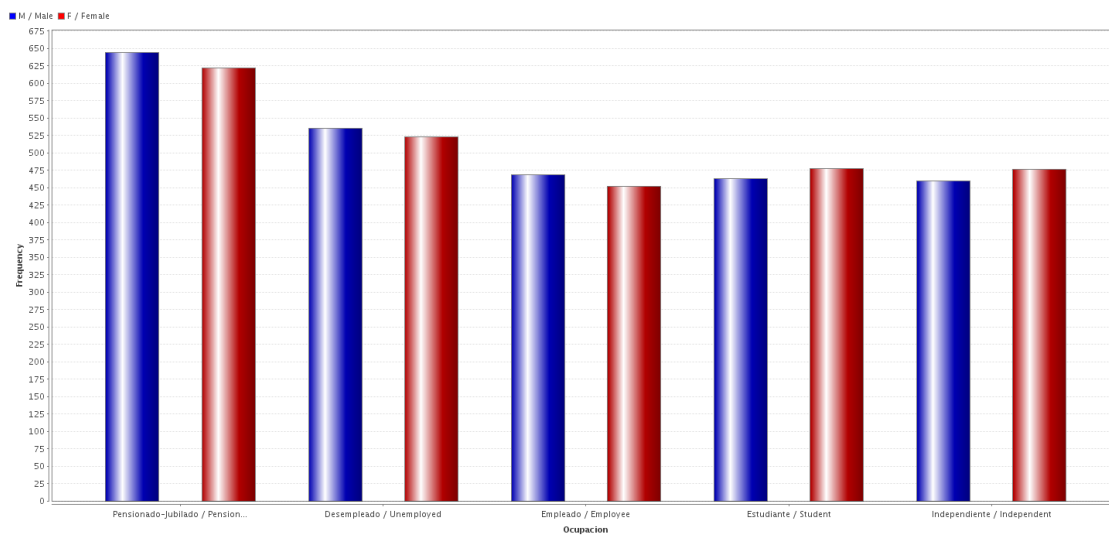


Figura 33. Histograma Genero-Ocupación.

Fuente: Resultado obtenido desde el software RapidMiner.

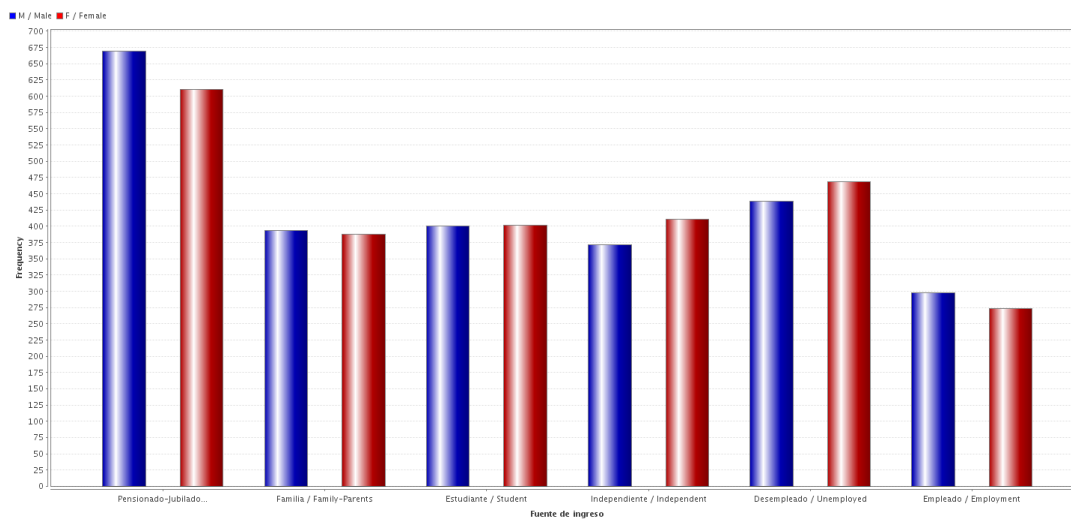


Figura 34. Histograma Genero-Fuente de Ingresos.

Fuente: Resultado obtenido desde el software RapidMiner.

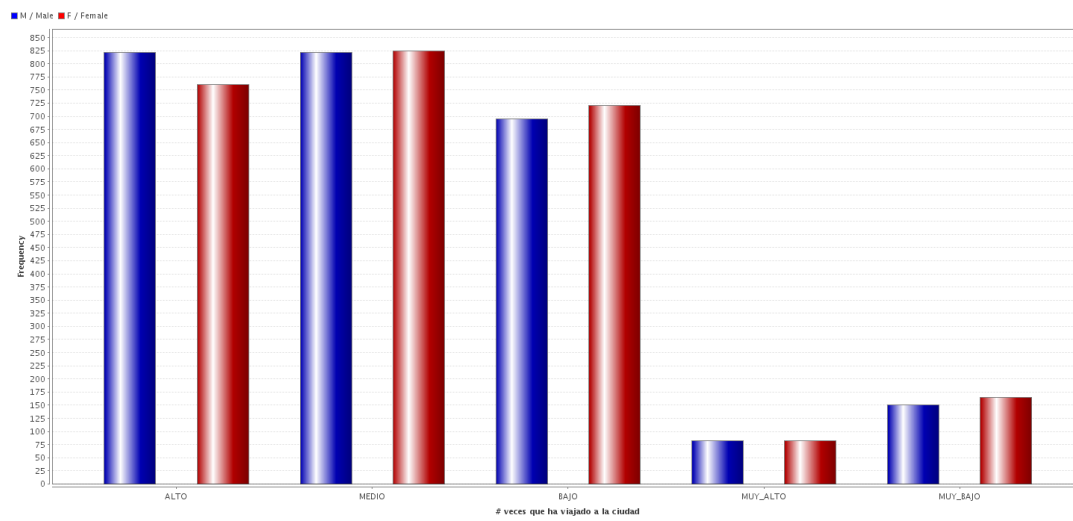


Figura 35. Histograma Genero-Número de visitas a la ciudad.

Fuente: Resultado obtenido desde el software RapidMiner.

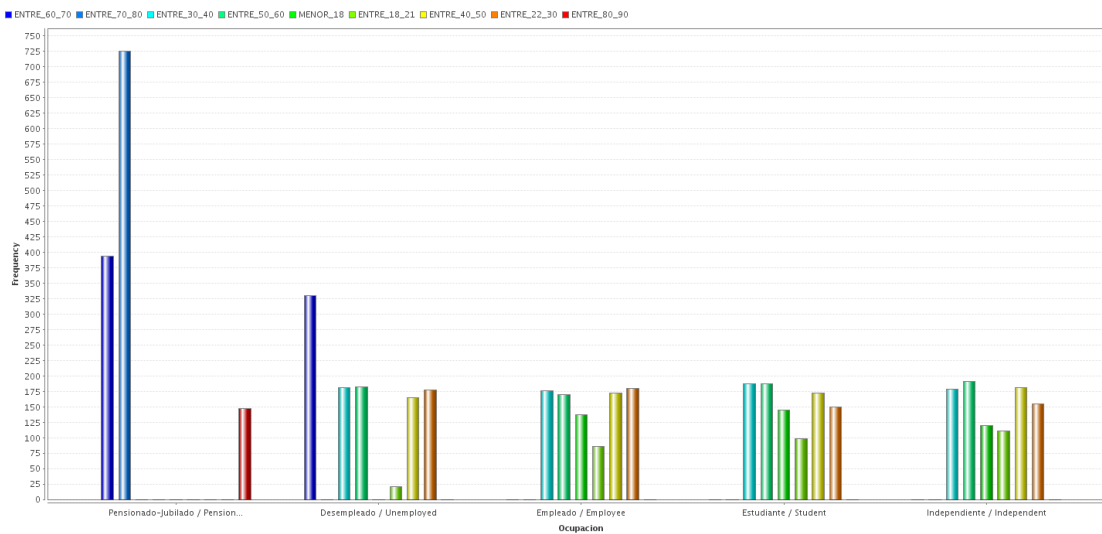


Figura 36. Histograma Edad-Ocupación.

Fuente: Resultado obtenido desde el software RapidMiner.

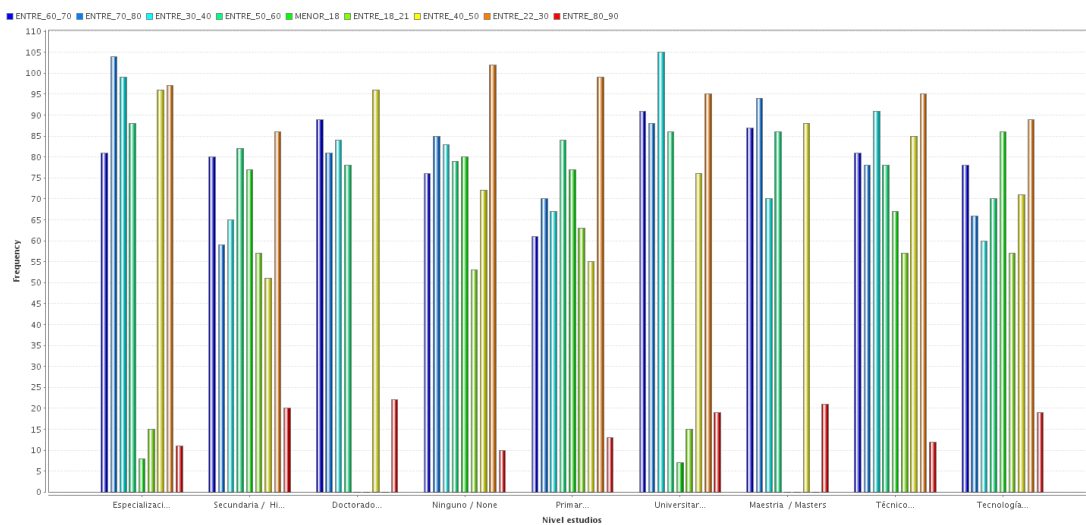


Figura 37. Histograma Edad-Nivel de Estudios.

Fuente: Resultado obtenido desde el software RapidMiner.

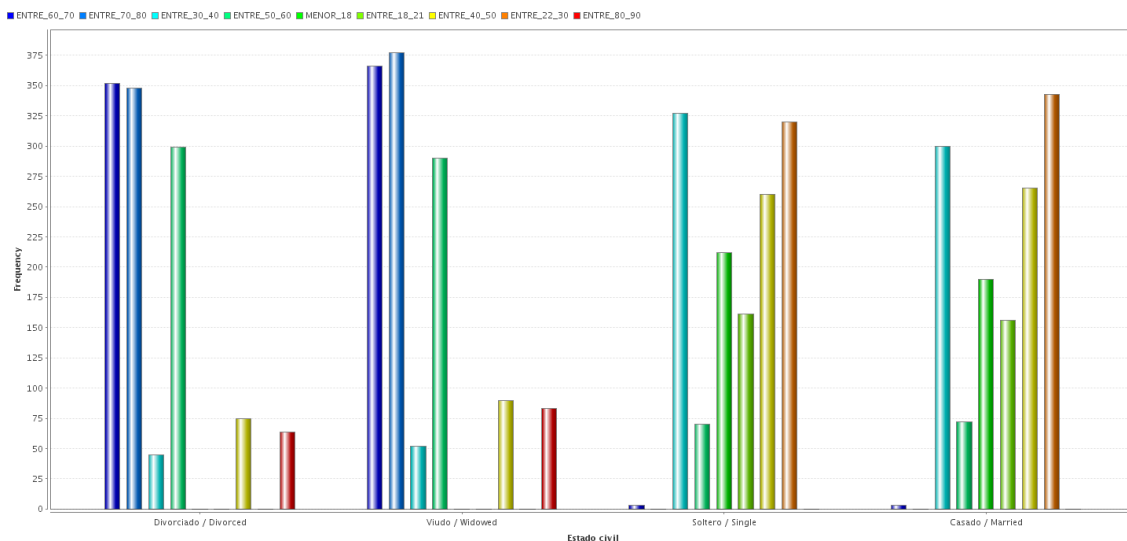


Figura 38. Histograma Edad-Estado Civil.
Fuente: Resultado obtenido desde el software RapidMiner.

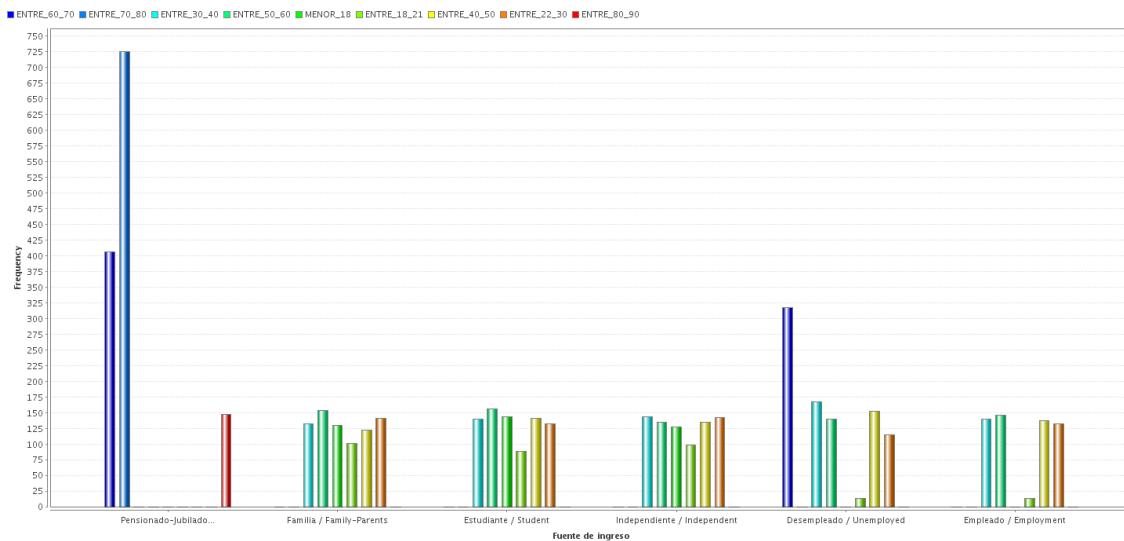


Figura 39. Histograma Edad-Fuente de Ingresos.
Fuente: Resultado obtenido desde el software RapidMiner.

Recordemos que el análisis de minería de datos requiere de conocimiento del sector en donde se realiza el estudio, a continuación se describen algunas observaciones con unas posibles

causas.

La variable número de visitas es importante (Ver Figura 35.Histograma Genero-Número de visitas a la ciudad.), esta tiene una variación marcada entre los viajeros, es interesante analizar las características de las personas que regresan nuevamente a la ciudad.

Se puede observar que hay una marcada tendencia en el número de visitas, los datos más extremos o con pocos datos son:

- Alto número de viaje
- Bajo número de viaje.

Argumentando sobre la ciudad, esta tendencia nos puede indicar que el grupo de poco número de viajes no le interesa regresar o que viajan por necesidad a la ciudad, puede ser que en algún momento llegaron a la ciudad de paso y por algún motivo se quedaron durante varios días (Risaralda., 2015).

El otro grupo de viajeros muy alto, es interesante analizarlo con Clustering ya que nos marcaría la tendencia de ese grupo de personas, inicialmente podríamos argumentar que es posible que visiten a algún amigo, familiar, estén realizando negocios en la ciudad, sean propietarios de una casa, finca, local comercial o apartamento en la ciudad, puede ser que les genere algún tipo de interés conocer a fondo la cultura de la ciudad, estén realizando algún proyecto social, etc.

Otras variables a analizar son la fuente de ingresos y la ocupación (Figura 34. Histograma Genero-Fuente de Ingresos. y Figura 36. Histograma Edad-Ocupación., la ciudad se beneficia con la visita y con el gasto que realizan los viajeros, la descripción de las personas que visitan la ciudad de acuerdo con sus ingresos nos ayudan a inferir aspectos que

permiten desarrollar actividades para que dichas personas regresen o que informe a personas conocidas con lo cual se incrementarían las visitas, aspectos a tener en cuenta:

- El mayor número de visitantes está en el grupo pensionados-jubilados.
- El menor grupo de visitantes esta en de empleados.
- El tercer grupo que más visita es el de desempleados, pero con una diferencia corta entre dicho grupo y el de estudiantes.

La edad, el estado civil y el nivel de estudios brindan datos importantes sobre el nivel socioeconómico, nivel de preparación académica y datos familiares de los viajeros, en la Figura 37. Histograma Edad-Nivel de Estudios.Figura 38. Histograma Edad-Estado Civil.Figura 39. Histograma Edad-Fuente de Ingresos. observamos tendencias diferenciadoras entre las edades de los viajeros, algunos análisis que se pueden inferir:

- Tenemos un grupo numeroso de viudos y divorciados en edades superiores a los 60 años.
- En el nivel de estudios observamos que no hay menores de edad con Doctorado, tampoco con especialización, según el promedio general de estudios de varios países parece correcto, también nos dice que ningún niño genio ha visitado la ciudad.
- En los grupos de desempleados, estudios de primaria, secundaria y pregrado hay diferentes rangos de edades.
- Las personas que más visitan la ciudad son los casados y solteros en rangos de edad inferiores a los 60 años.
- Una característica del grupo de casados es que tenemos menores de edad con ese estado civil, en muchos países es legal con consentimiento de los adultos, en muchos otros no es

permitido, por lo tanto parece normal dicho grupo.

- La fuente de ingresos por edad es variable, el menor grupo es el de los empleados (Figura 39. Histograma Edad-Fuente de Ingresos.), pocos menores de edad con ingresos de empleado, esto contrasta y no corresponde con el histograma de ocupación (Figura 36. Histograma Edad-Ocupación.) en la cual se observa un alto número de empleados menores de edad. Dicha variable puede causar ruido para el análisis, se filtrará la ocupación en el proceso de Clustering a continuación.

Proceso inicial de Clustering. El proceso inicial de Clustering (Figura 40. Clustering validación 1) tiene el propósito de crear los primeros clústers para ser analizados y ajustar el modelo y/o los datos.

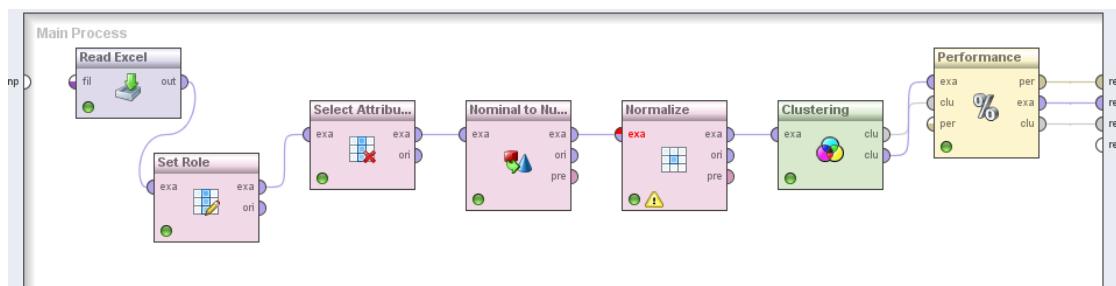


Figura 40. Clustering validación 1

Fuente: Proceso Obtenido desde el software RapidMiner.

Este proceso determino los primeros clústers sin supervisión, posteriormente se eliminan los Outliers, se realizar análisis con algunas variables, se eliminan inconsistencias entre los datos y se observan las relaciones entre los datos de cada Clúster con diagrama de Árbol y reglas de asociación.

La explicación del modelo:

ReadExcel: Realiza la lectura del archivo CSV ya procesado.

Set Rol: Se selecciona el rol definido como CLUSTER al campo que agrupara los datos, en este caso se utilizó el Rol ID para el campo Nombre, de esta manera tener el Código o identificador de cada registro del DataSet.

SelectAttributes: Se seleccionan los atributos necesarios para generar el clúster, en la primera etapa se definen los campos con los datos de los viajeros como: edad, estado civil, nivel de estudios, ocupación, fuente de ingresos.

Nominal_to_numerical: Convertir los atributos de formato nominal a número ya que los algoritmos de Clustering no permiten valores nominales, polinomiales y texto, solo numéricos. En este proceso se ejecutó 2 veces utilizando los parámetros de codificación de los valores nominales en “dummy coding” y “unique integers”.

Normalize: La normalización modifica los valores numéricos convertidos en el proceso anterior (Nominal to Numerical) y los convierte a una misma escala para ser procesados, en este caso seleccionamos la opción: “z-transformation” y “range transformation”, siendo el rango mínimo 0 y el rango máximo 1.

Clustering: Ejecuta el algoritmo K-Means, se seleccionó $k=4$ y $\text{max runs}=10$ lo cual equivale a inicializar 10 veces de forma aleatoria el método K-Means. El tipo de medida utilizada para generar los clústers fue BregmanDivergences con una divergencia de distancia euclídea, siendo esta la recomendada para Clustering (Dean, 2014).

Performance: Evalúa el modelo del clúster basado en la distribución de los ejemplos, es decir en las distancias entre ellos.

El resultado del proceso genero los datos de rendimiento (Performance): “Dummy coding y range transformation”

PerformanceVector:

Avg. within centroid distance: -6.396

Avg. within centroid distance_cluster_0: -6.902

Avg. within centroid distance_cluster_1: -6.511

Avg. within centroid distance_cluster_2: -6.265

Avg. within centroid distance_cluster_3: -5.820

Davies Bouldin: -3.347

Performance: “unique integers y range transformation”

PerformanceVector:

Avg. within centroid distance: -0.751

Avg. within centroid distance_cluster_0: -0.781

Avg. within centroid distance_cluster_1: -0.720

Avg. within centroid distance_cluster_2: -0.707

Avg. within centroid distance_cluster_3: -0.771

Davies Bouldin: -1.708

La interpretación de los resultados de acuerdo con los autores indica que mientras más bajo sea el índice Bouldin es mucho mejor la agrupación, ya que la técnica de Clustering por K-mean es una técnica no supervisada, esta se maneja a consideración del analista de datos. Como se puede ver, los mejores índices están utilizando la conversión nominal to numerical con enteros y la normalización con una transformación de rango entre 0-1 (Dean, 2014).

Segundo Proceso de Clustering: Este proceso tiene como propósito realizar la verificación y eliminación de los Outliers.

Se verificaron los datos normalizados mediante el peso asignado, dichos pesos fueron seleccionados cuidadosamente, por ejemplo: para el número de viajes de la Tabla 9. Campos discretizados del número de viajes a la ciudad, el mayor puntaje lo obtienen los viajes con un número mayor a 30, en otro ejemplo, al analizar la Tabla 8. Campos discretizados de la edad de las edades de los viajeros, tenemos un peso en un rango de [1-10] teniendo en cuenta que los menores de edad tienen un puntaje de 1 al depender económicamente de los familiares o de sus padres, ese es el común de las poblaciones, en caso de que algún menor de edad tenga sus propios ingresos, dicho dato es tratado como un Outlier ya que altera la media de la población, si hay un grupo numeroso de menores con sus propios ingresos, el agrupamiento por Clustering nos ayuda a separar dichos datos para su análisis.

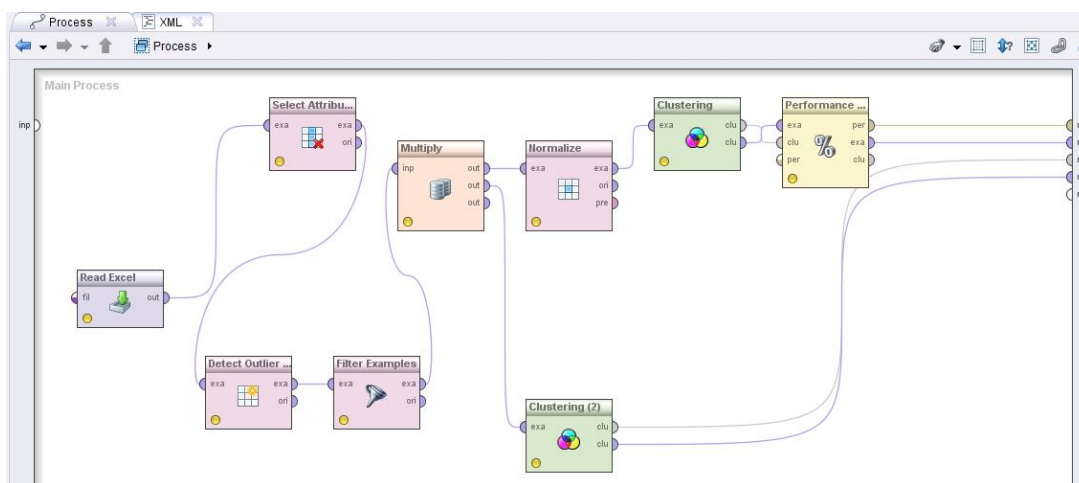


Figura 41. Filtrado y eliminación de Outliers proceso 2 modificado.

Fuente: Proceso Obtenido desde el software RapidMiner.

La explicación del modelo:

ReadExcel: Realiza la lectura del archivo CSV ya procesado.

SelectAttributes: Se seleccionan los atributos necesarios para generar el clúster, en la primera etapa se definen los campos con datos de hábitos, en la siguiente etapa de Clustering los de percepción y satisfacción. En este proceso se seleccionan los datos de hábitos (número de veces que ha viajado, motivo viaje, días de permanencia, compañeros de viaje, como se enteró de la ciudad, lugar de alojamiento, actividad que realizo, medio de transporte para llegar a la ciudad, medio de transporte utilizado en la ciudad, sitio de alimentación, comida preferida). En el proceso se seleccionan las variables con su respectivo peso numérico.

Detect Outliers: Mediante el algoritmo de distancias euclideas, se determina eliminar un máximo de 10 Outliers teniendo en cuenta un análisis de 10 vecinos.

Filter Examples: Se crea la condición: outlier=false para que solamente obtengamos en la salida del proceso los datos más compactos y eliminemos los más lejanos al clúster.

Multiply: Separa el resultado de la eliminación de los Outliers para validar de dos formas el Cluster, la primera utilizando el Clustering con el algoritmo K-Means, el segundo utilizando la normalización de los datos, es decir, los datos numéricos se normalizan utilizando el método (range transformation) determinando que todos los valores quedaran en escala de números entre [0-1]

Normalize: Normalizamos todos los datos numéricos seleccionados en el proceso “Select attributes”, Cuando seleccionados esta opción, aparecen otros dos parámetros (min, max). La transformación por rango normaliza en el rango [min,max], nosotros seleccionamos un rango numérico entre [0-1].

Performance: Este operador se utiliza para evaluar el rendimiento de la evaluación de los centroides en el método Clustering. Los criterios seleccionados con el proceso son:

Avg within centroid distance y Davies Bouldin.

El resultado del proceso sin normalización:

PerformanceVector:

Avg. within centroid distance: -11.881

Avg. within centroid distance_cluster_0: -11.773

Avg. within centroid distance_cluster_1: -12.029

Avg. within centroid distance_cluster_2: -11.960

Avg. within centroid distance_cluster_3: -11.699

Avg. within centroid distance_cluster_4: -12.278

Avg. within centroid distance_cluster_5: -11.938

Avg. within centroid distance_cluster_6: -11.733

Avg. within centroid distance_cluster_7: -11.656

Davies Bouldin: -2.418

El resultado del proceso con normalización:

PerformanceVector:

Avg. within centroid distance: -0.894

Avg. within centroid distance_cluster_0: -0.934

Avg. within centroid distance_cluster_1: -0.879

Avg. within centroid distance_cluster_2: -0.857

Avg. within centroid distance_cluster_3: -0.925

Avg. within centroid distance_cluster_4: -0.920

Avg. within centroid distance_cluster_5: -0.846

Avg. within centroid distance_cluster_6: -0.918

Avg. within centroid distance_cluster_7: -0.869

Davies Bouldin: -2.293

El resultado del índice Davies Bouldin indica el promedio de similaridad entre cada clúster, cada clúster debe ser compacto y separado de los demás, el índice mientras más pequeño indica que es mucho mejor, por lo tanto el resultado normalizado tiene mejor puntaje.

Comparando el primer proceso Clustering(Proceso *inicial de Clustering*) en el cual utilizamos los datos discretizados nominales con el (Segundo Proceso de Clustering:) en el cual utilizamos los datos con pesos, el Índice Bouldin indica una mejor distribución en el primer caso.

Tercer Proceso de Clustering: Este proceso tiene como propósito validar el rendimiento de los clústers mediante las medidas internas:

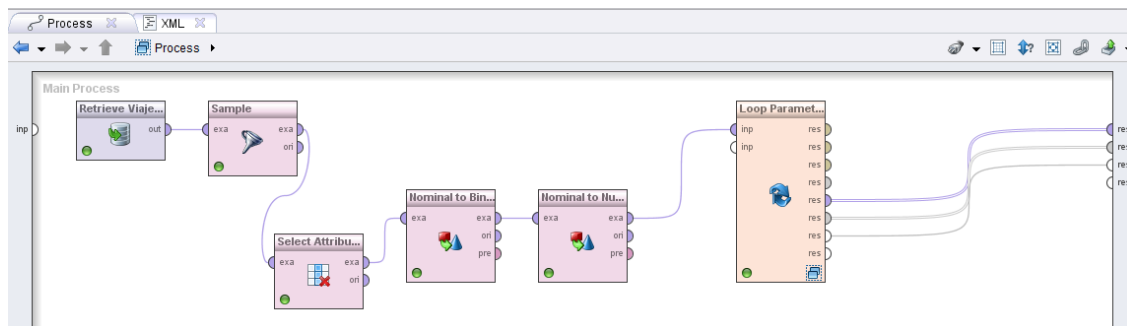


Figura 42. Validación con medidas internas parte 1.

Fuente: Proceso Obtenido desde el software RapidMiner.

La explicación del modelo:

Retrieve: Realiza la lectura del archivo CSV ya procesado resultante del modelo anterior

Figura 17 en cuál es el DataSet con los valores nominales, no el DataSet con los pesos numéricos.

Sample: Permite seleccionar un subDataSet para realizar experimentos con parte de la información.

SelectAttributes: Se seleccionan los atributos necesarios para generar el clúster, en la primera etapa se definen los campos con datos de hábitos, en la siguiente etapa de Clustering los de percepción y satisfacción.(Fuente de ingreso, nivel de estudios, Número de veces que ha viajado, motivo viaje, días de permanencia, compañeros de viaje, como se enteró de la ciudad, lugar de alojamiento, actividad que realizo, medio de transporte para llegar a la ciudad, medio de transporte utilizado en la ciudad, sitio de alimentación, comida preferida).

Nominal_to_binomial: Convertir los atributos de formato nominal a valor true/false, cada uno de los atributos se convierte a dicho valor, generándose gran cantidad de columnas, ya que los algoritmos de clustering no permiten valores nominales, polinomiales y texto, solo numéricos.

Nominal_to_numerical: Convertir los atributos de formato nominal anteriormente convertidos (true/false) a número ya que los algoritmos de Clustering no permiten valores nominales, polinomiales y texto, solo numéricos.

Loop Parameter: Ejecuta todo el ciclo de Clustering mediante un ciclo el cual inicia en 1 y

termina en 20, esto quiere decir que ejecuta 20 veces el ciclo y cada vez cambia el número de k en el clúster k_means.

Log to Data: Transforma los datos generados por el ciclo anterior “Loop Parameter” en un ExampleSet el cual pueda ser utilizado por otros operadores del proceso.

GenerateID: Genera un identificador para los elementos del clúster.

Multiply: Divide la entrada para múltiples salidas.

Clustering: Genera los clústers mediante el algoritmo K-Means, tomando como parámetro k=20, max runs= 20, utilizando local random seed = 2, measure types= BregmanDivergences, divergence = SquaredEuclideanDistance.

Distance: Evalúa el rendimiento del centroide basado en métodos de Clustering, indica la distancia media de cada elemento del clúster hacia su centroide.

Density: Entrega una lista de criterios de rendimiento basado en la densidad de los clústers. Esto indica el número de puntos en un radio específico.

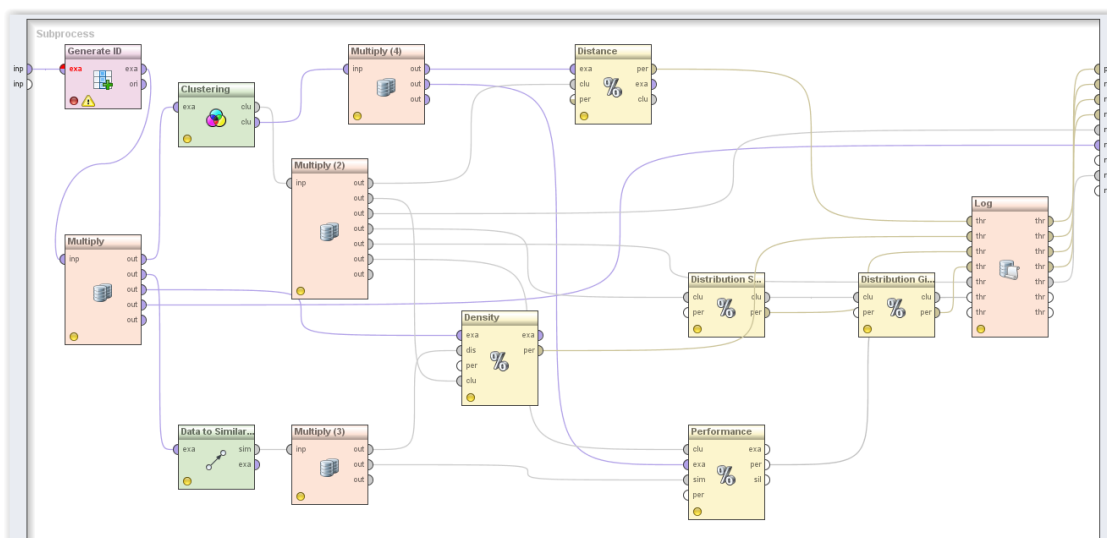


Figura 43. Ciclo para validar los clústers.

Fuente: Proceso Obtenido desde el software RapidMiner.

Distribution SoS: (Ítem distribution performance) Evalúa el modelo del clúster basado en la distribución de sus ejemplos, es decir, que tan bien están distribuidos los datos sobre los clústers. El coeficiente utilizado es la suma de sus cuadrados.

Distribution Gini: Evalúa el modelo del clúster basado en la distribución de sus ejemplos, es decir, que tan bien están distribuidos los datos sobre los clústers. Realiza la misma validación que el proceso anterior, la diferencia es que en este caso utiliza el coeficiente Gini, siendo esta una medida de dispersión la cual mide las inequidades entre los valores de una distribución de frecuencia, siendo 0 el valor con el cual se identifica, totalmente iguales y 1 totalmente no iguales.

Performance Silhouette: Computa con el algoritmo Silhouette el ancho de un clúster y el promedio con todo el DataSet. Esto quiere decir que el índice Silhouette permite determinar que tan bien se encuentran asignados los clústers par aun punto particular.

Silhouette maneja cohesión (que tan compactos se encuentran los clústers) y separación (separación entre clústers), debido a esto, dicho índice es muy importante para tener en cuenta.

Table View
 Plot View
 Advanced Charts

Valores (19 rows, 7 columns)

DaviesBouldin	avgWithinDistance	k	clusterDensity	itemDistrib...	Gini	silhouetteln...
3.972	6.010	2	-9849.199	0.554	1.000	0.053
3.368	5.678	3	-5728.166	0.333	1.000	0.078
3.339	5.592	4	-4264.689	0.250	1.000	0.078
3.307	5.508	5	-3531.651	0.208	1.000	0.062
3.406	5.422	6	-3017.749	0.179	1.000	0.054
3.398	5.342	7	-2829.722	0.169	1.000	0.051
3.826	5.273	8	-2595.630	0.155	1.000	0.053
3.691	5.172	9	-2066.326	0.125	1.000	0.064
3.465	5.085	10	-1804.612	0.110	1.000	0.072
3.499	5.048	11	-1613.387	0.099	1.000	0.069
3.257	4.972	12	-1442.445	0.089	1.000	0.073
3.135	4.902	13	-1250.399	0.078	1.000	0.080
3.159	4.875	14	-1189.236	0.075	1.000	0.073
3.441	4.868	15	-1187.869	0.075	0.999	0.066
3.213	4.827	16	-1103.893	0.070	0.999	0.059
3.239	4.809	17	-1033.625	0.065	0.999	0.059
3.081	4.786	18	-968.353	0.061	0.999	0.058
3.452	4.767	19	-972.217	0.061	0.999	0.054
3.161	4.749	20	-891.611	0.057	0.999	0.057

Figura 44. Índices de validación de clusters.

Fuente: Resultado obtenido desde el software RapidMiner.

De acuerdo con el resultado del proceso tres de Clustering y la imagen de la Figura 44. Índices de validación de clusters. tenemos 3 valores de K ideales para dividir el DataSet, ya que el índice Silhouette se acerca a 1, están más cercanos que los demás resultados, el índice de Gini es 1 lo cual indica que la tendencia de agrupación de cada clúster es 100% diferente entre ellos.

La densidad de los clústers indica que el valor de K=4 tiene más distribuidos los datos en grupos, distribución de los ítems de 0.251.

En la siguiente figura observamos la distribución de los centroides en el caso de K=4.

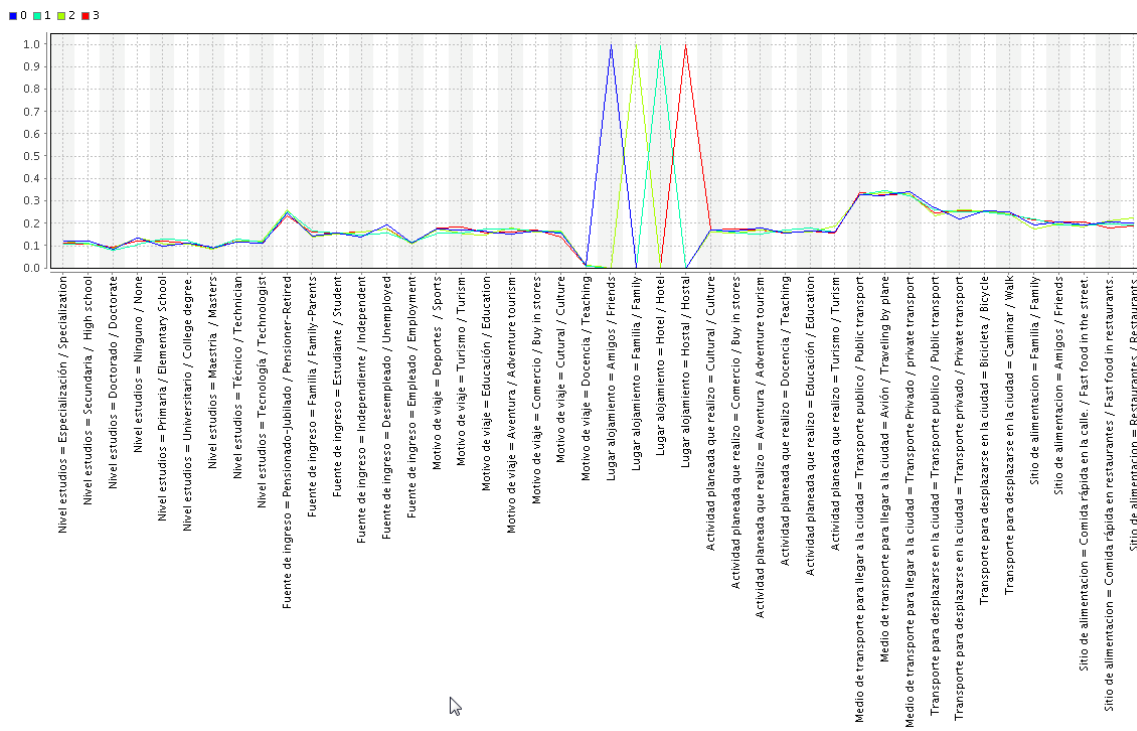


Figura 45. Distribución de centroides con k=4 Proceso 3.

Fuente: Resultado obtenido desde el software RapidMiner.

En el siguiente proceso se valida nuevamente el índice Silhouette con RapidMiner y el lenguaje R.

Cuarto Proceso de Clustering: En el proceso anterior se observó que el índice Silhouette cuando $k=4$ tiende hacia 1, con este nuevo proceso se valida si en realidad esas medidas fueron confiables y verídicas, parte del proceso ejecuta un algoritmo en el lenguaje R para validar Silhouette, de esta manera se realiza el cálculo con 2 herramientas. Figura 46. Validar el número óptimo de clústers para los datos proceso 1. y Figura 47. Validar el número óptimo de clústers proceso 2.

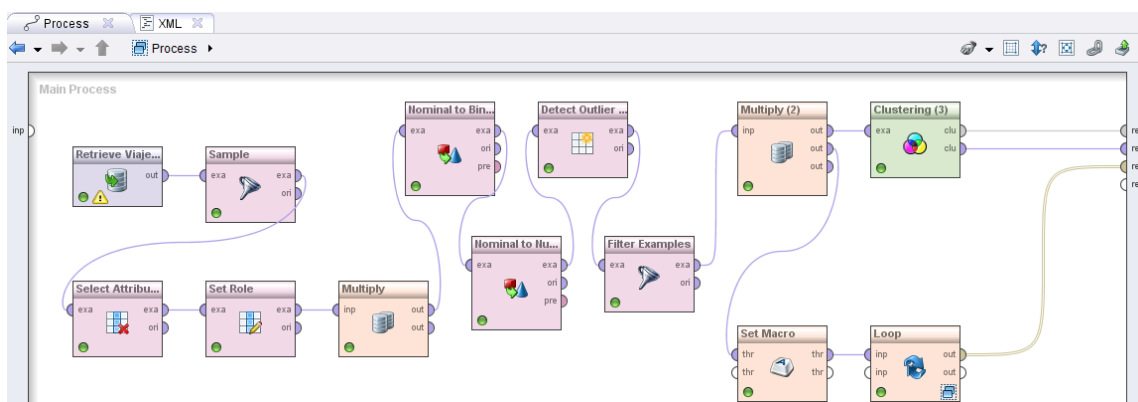


Figura 46. Validar el número óptimo de clústers para los datos proceso 1.

Fuente: Proceso Obtenido desde el software RapidMiner.

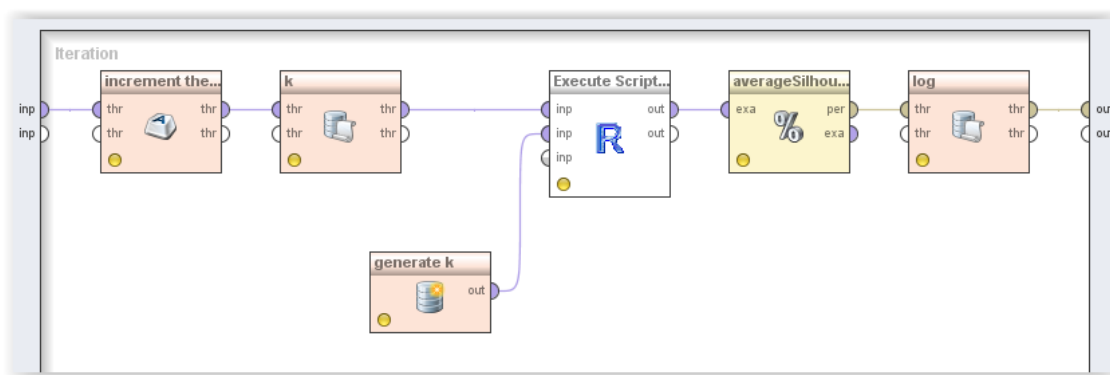


Figura 47. Validar el número óptimo de clústers proceso 2.

Fuente: Proceso Obtenido desde el software RapidMiner.

La explicación del modelo:

Retrieve: Realiza la lectura del archivo CSV con los datos nominales.

SelectAttributes: Se seleccionan los atributos necesarios para generar el clúster, en este caso todos los relacionados con los datos generales de los viajeros y los datos de hábitos. (Edad,

estado civil, sexo, fuente de ingreso, nivel de estudios, número de veces que ha viajado, motivo viaje, días de permanencia, compañeros de viaje, como se enteró de la ciudad, lugar de alojamiento, actividad que realizo, medio de transporte para llegar a la ciudad, medio de transporte utilizado en la ciudad, sitio de alimentación, comida preferida) La ocupación la omitimos debido a las inconsistencias encontradas en el Proceso inicial de análisis exploratorio de datos (EDA):.

Set Rol: Se selecciona el rol definido como ID al campo que agrupara los datos, en este caso se utilizó el Rol ID para el campo Nombre, de esta manera tener el Código o identificador de cada registro del DataSet.

Nominal_to_binomial: Convertir los atributos de formato nominal a valor true/false, cada uno de los atributos se convierte a dicho valor, generándose gran cantidad de columnas, ya que los algoritmos de clustering no permiten valores nominales, polinomiales y texto, solo numéricos.

Nominal_to_numerical: Convertir los atributos de formato nominal anteriormente convertidos (true/false) a número ya que los algoritmos de Clustering no permiten valores nominales, polinomiales y texto, solo numéricos.

Detect Outliers: Mediante el algoritmo de distancias euclídeas, se determina eliminar un máximo de 10 Outliers teniendo en cuenta un análisis de 10 vecinos.

Filter Examples: Se crea la condición: outlier=false para que solamente obtengamos en la salida del proceso los datos más compactos y eliminemos los más lejanos al clúster.

Set Macro: Crea una variable kValue y la inicializa en 1.

Loop: Ejecuta un ciclo para validar el número de clústers, se inicia con un total de 20 iteraciones.

Dentro del ciclo (Figura 47):

Increment the counter: La macro “Variable” kValue se incrementa en 1.

K: Almacena el valor de kValue para poderlo procesar y exportar a un archivo.

Execute Script: Ejecuta el proceso de calcular el clúster y validarlo con el algoritmo Silhouette, tiene como entrada la macro Data y la macro k. Tiene como salida la tabla de datos resultante para ser procesada por el siguiente operador.

```
## Carga la libreria.
```

```
library(cluster)
```

```
## contiene los datos de entrada: data frame data y k
```

```
dataNames = names(data)
```

```
kClusters = k$k
```

```
clusteredData = pam(data[dataNames[1:3]],kClusters)
```

```
averageSilhouette = clusteredData$silinfo$avg.width
```

```
x = as.data.frame(averageSilhouette)
```

AverageSilhouette: Genera un vector de rendimiento dadas las propiedades del exampleSet.

Log: Almacena los datos para ser visualizados, las variables a mostrar son averageSilhouette y la macro k con el número de clústers recomendado para dicha medida de Silhouette.

Cabe recordar que los factores de validación, además de un buen clúster con separación y cohesión (Larose & Larose, 2015, pág. 852) Un buen clúster incorpora ambos criterios, el método de validación de clúster Silhouette nos brinda un índice para determinar dichos

parámetros.

El resultado de la ejecución del proceso da como resultado un resultado $k=4$ con el Silhouette = 1.

La distribución de los cuatro clústers:

Cluster Model

Cluster 0: 945 items

Cluster 1: 1277 items

Cluster 2: 1396 items

Cluster 3: 1495 items

Total number of items: 5113

El modelo se encuentra claramente en un gráfico de todos los índices Silhouette, ver Figura 48. Plotter del índice Silhouette. y Figura 49. Tabla del índice Silhouette y el número de clusters.

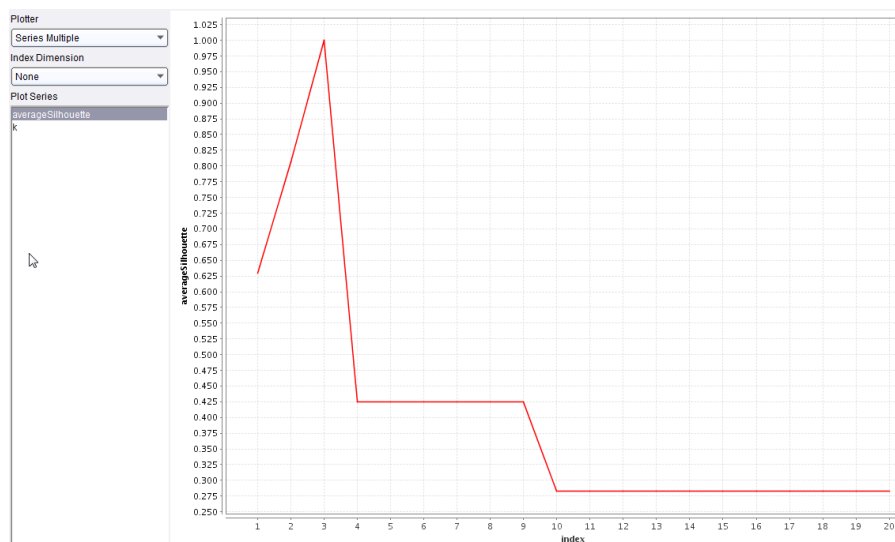


Figura 48. Plotter del índice Silhouette.

Fuente: Resultado obtenido desde el software RapidMiner.

averageSilhouette	k
0.630	2
0.807	3
1	4
0.424	5
0.424	6
0.424	7
0.424	8
0.424	9
0.424	10
0.282	11
0.282	12
0.282	13
0.282	14
0.282	15
0.282	16
0.282	17
0.282	18
0.282	19
0.282	20
0.282	21

Figura 49. Tabla del índice Silhouette y el número de clusters.

Fuente: Resultado obtenido desde el software RapidMiner.

A continuación se determina la configuración de las reglas de asociación y posteriormente el proceso que agrupa todos los requerimientos del objetivo uno de Clustering y el segundo objetivo con las reglas de asociación de la presente investigación.

Proceso para las reglas de asociación: Este proceso tiene como propósito generar reglas de asociación con los clústers generados.

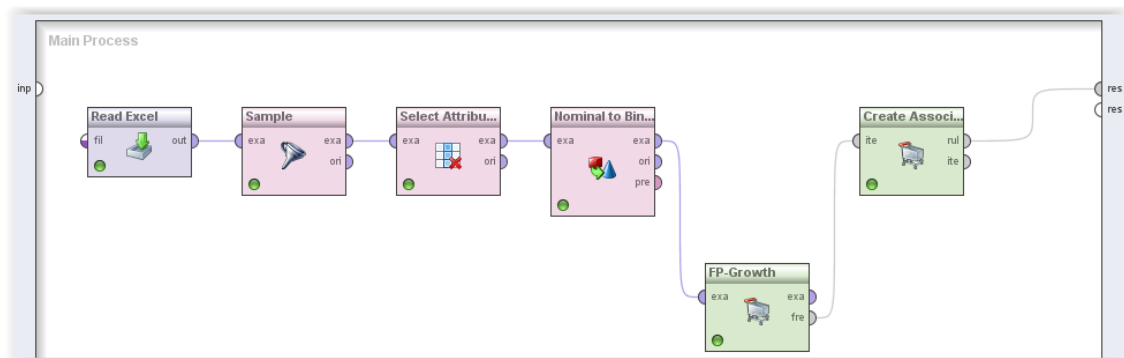


Figura 50. Proceso para generar reglas de asociación.

Fuente: Proceso Obtenido desde el software RapidMiner.

La explicación del modelo:

Read: Realiza la lectura del archivo CSV ya procesado resultante del modelo anterior.

Sample: Toma una muestra de los datos para poder generar más fácilmente las reglas.

SelectAttributes: Se seleccionan los atributos necesarios para calcular la frecuencia de los ítems con el algoritmo FP-Growth.

FP-Growth: Calcula los ItemSets frecuentes de un conjunto de datos ExampleSet, es decir, agrupa los ítems frecuentes.

Nominal to Binomial: Realiza la conversión de los atributos seleccionados ya que el algoritmo solo procesa tipos de datos binomiales.

Create Association Rules: Genera unas reglas de asociación dados los ItemSets generados por el proceso anterior.

Quinto Proceso de Clustering: Este proceso integra todos los anteriores, teniendo en cuenta el resultado validado con RapidMiner y lenguaje R en el cual el número de clústers adecuado para el DataSet es de $K = 4$. En este proceso se validan los Outliers, se realiza la conversión de nominal a binomial y posteriormente a número.

Es de aclarar que todos los modelos desde el primero hasta el quinto se validaron con diferentes valores, combinaciones de atributos, tipos de conversiones, tipos de índices. Los diferentes parámetros de todos los experimentos se encuentran en el (Ver **¡Error! No se encuentra el origen de la referencia.**).

El valore con los mejores índices se encuentra en la última iteración, dicho resultado dio como índice Silhouette = 1.

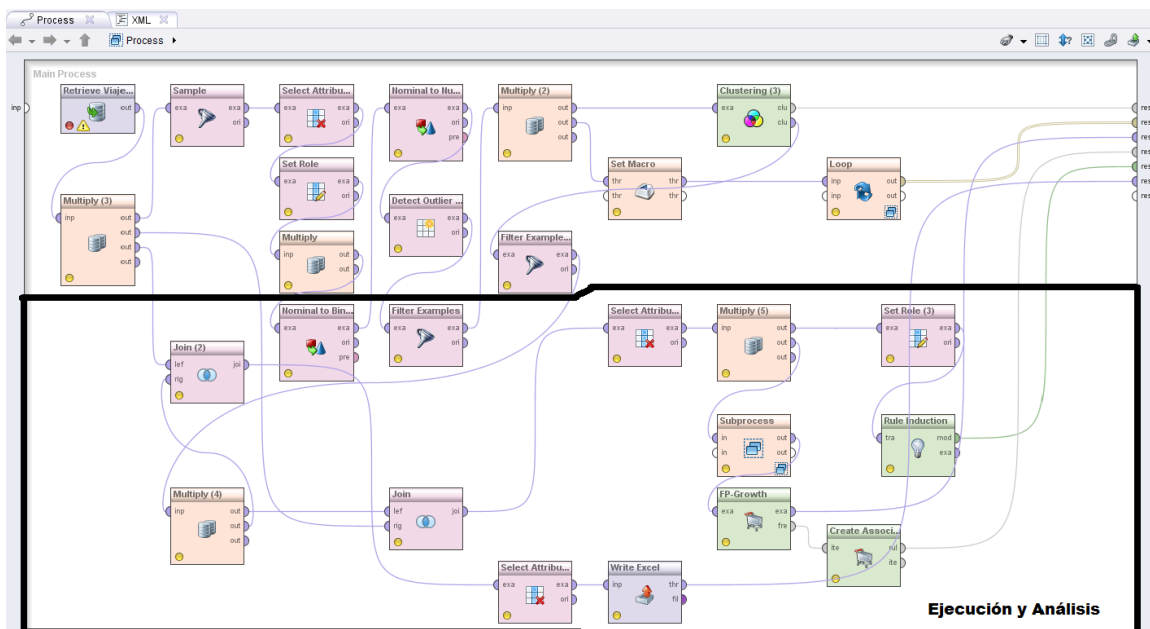


Figura 51. Quinto Proceso Clustering.

Fuente: Proceso Obtenido desde el software RapidMiner.

Explicación del proceso:

Ya que este proceso une los cuatro anteriores se procede a describir únicamente la parte marcada como ejecución y análisis en la Figura 51. Quinto Proceso Clustering.

Join(1) y Join(2): Unen el DataSet original mediante el índice el cual es el nombre, para el dataSet original el nombre es un autonumerico que indica la posición del viajero, el proceso join es igual a un Join en SQL, la otra entrada del proceso es la salida del proceso “Clustering” pero este esta filtrado, es decir, se extrae con el parámetro cluster=1, cluster=2 etc, de esta manera se separan los datos de cada clúster para su ingreso en las reglas de asociación y reglas de inducción.

Este proceso se realiza ya que el DataSet original esta con su formato original “Nominal” y los datos filtrados en formato “Numero”.

Select Attributes: Seleccionamos únicamente los atributos que hacen referencia a los datos personales y datos que rodean la visita a la ciudad. Dichos datos son los originales Nominales.

Ambos Select Attributes realizan la misma actividad.

Los atributos seleccionados son: (Edad, estado civil, sexo, fuente de ingreso, nivel de estudios, número de veces que ha viajado, motivo viaje, días de permanencia, compañeros de viaje, como se enteró de la ciudad, lugar de alojamiento, actividad que realizo, medio de transporte para llegar a la ciudad, medio de transporte utilizado en la ciudad, sitio de alimentación, comida preferida)

Select Attributes y Write Excel: Seleccionan los atributos filtrados y se exportan a un archive excel con lo cual obtenemos cada cluster de forma separada.

Select Attributes, Multiply, set Role, Rule Induction: Primero se seleccionan los atributos filtrados, luego se dividen dichos atributos para ser pasados al proceso Set Role el cual definimos como “Lugar de alojamiento” con atributo rol “Label” de esta manera la regla de inducción nos permite obtener resultados enfocados hacia dicha etiqueta, es decir, todas las reglas de inducción mostraran reglas “IF” que se dan con las demás características de los viajeros que utilizaron los diferentes sitios de alojamiento.

Select Attributes, Multiply, Subprocess(Nominal to Binomial), FP-Growth, Create Association Rule: Primero se seleccionan los atributos filtrados, luego se dividen dichos atributos para ser pasados al proceso mutiply, este filtra directamente y pasa los parámetros al

Subprocess, dicho subprocess contiene “Nominal to Binomial” convirtiendo los datos seleccionados en “Select Attributes” a datos True/False los cuales son los exigidos por el algoritmo FP-Growth, luego de ser analizados por FP-Growth se pasan a la creación de las reglas de asociación con la cual obtenemos las medidas de “support” y ”confidence” de esta manera se analiza las reglas, su calidad y su probabilidad de que ocurran.

El quinto modelo se aplicó a los 3 objetivos ya que integra las validaciones internas de los clústers, filtrado de datos, valores de confianza y confidencia de las reglas de asociación.

En la siguiente etapa de CRISP-DM “Evaluación” se ejecutara la evaluación del modelo y el análisis de los resultados.

Evaluación del modelo

A partir de este punto se desarrolla el Objetivo 2: **Establecer relaciones y tendencias existentes entre las características de los viajeros y los aspectos que rodean su visita a la ciudad de Pereira por medio de reglas de asociación.**

La evaluación del modelo se realiza teniendo en cuenta todos los experimentos en el **¡Error! No se encuentra el origen de la referencia.**, observando los índices de validación más adecuados de acuerdo a la teoría de validación interna de clústers “Las medidas internas de validación de clustering que se *utilizaron:*” y el modelo de la Figura 51. Quinto Proceso Clustering.

Los resultados obtenidos fueron:

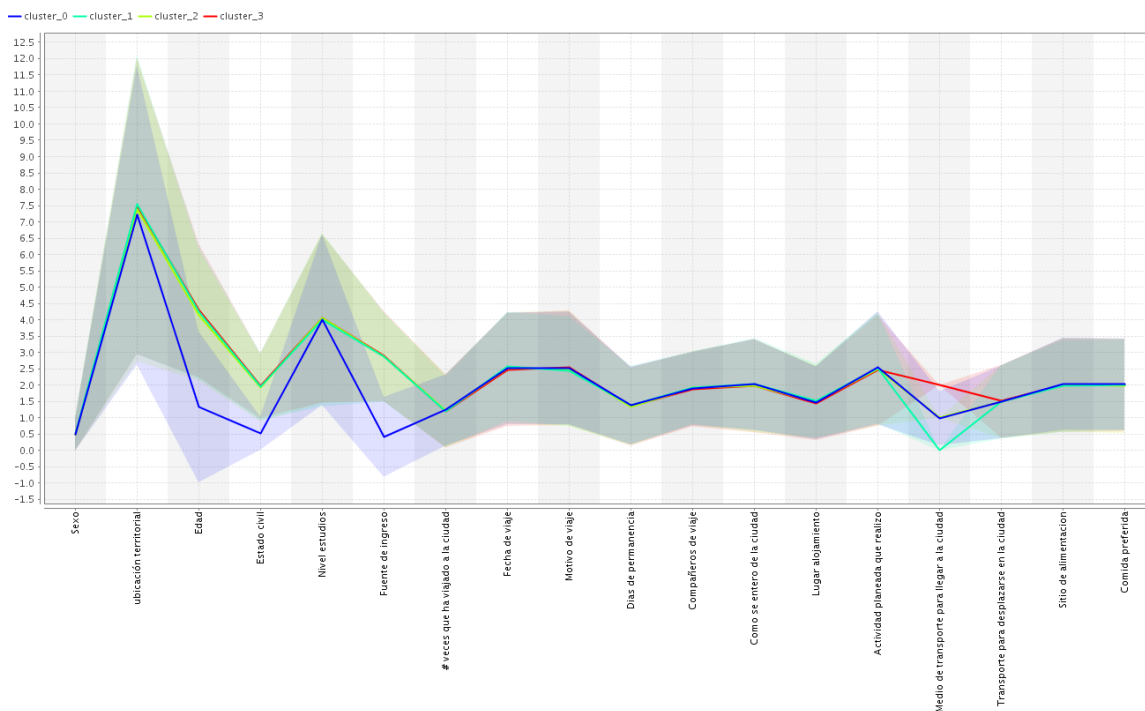


Figura 52. Gráfico de Desviación de todos los clusters.

Fuente: Resultado obtenido desde el software RapidMiner.

En el grafico se observa que la distribución de los cuatro clústers se da con diferencias en casi todas las variables, en las que se puede apreciar que está más marcada la diferencia: Edad, estado civil, nivel de estudios, nivel de ingresos, actividades planeadas realizadas y los medios de transporte utilizados.

Análisis del Clúster 0: El Clúster cero tiene un total de 1426 viajeros.

Name	Range
Nombre	[1.000 ; 5123.000]
outlier	false (1426), true (0)
cluster	cluster_0 (1426), cluster_1 (0), cluster_2 (0), cluster_3 (0)
ubicación territorial	ASIA_DEL_NORTE (95), ASIA_MERIDIONAL (96), SURAMERICA (96), ASIA_ORIENTAL (94), AFRICA_ORIENTAL (102), EUROPA (83), NORTEAMERICA (88), OCEANIA (99), CENTROAMERICA (105), ASIA_S
Edad	ENTRE_60_70 (554), ENTRE_70_80 (725), ENTRE_30_40 (0), ENTRE_50_60 (0), MENOR_18 (0), ENTRE_18_21 (0), ENTRE_40_50 (0), ENTRE_22_30 (0), ENTRE_80_90 (147)
Sexo	M / Male (731), F / Female (695)
Estado civil	Divorciado / Divorced (682), Viudo / Widowed (742), Soltero / Single (1), Casado / Married (1)
Nivel estudios	Especialización / Specialization (181), Secundaria / High school (137), Doctorado / Doctorate (168), Ninguno / None (150), Primaria / Elementary School (123), Universitario / College degree (173), Maestría
Ocupacion	Pensionado-Jubilado / Pensioner-Retired (1265), Desempleado / Unemployed (161), Empleado / Employee (0), Estudiante / Student (0), Independiente / Independent (0)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (1278), Familia / Family-Parents (0), Estudiante / Student (0), Independiente / Independent (0), Desempleado / Unemployed (148), Empleado / Employment (0)
# veces que ha viajado a la ciudad	ALTO (422), MEDIO (461), BAJO (411), MUY_ALTO (50), MUY_BAJO (82)
Fecha de viaje	ANO_2014 (233), ENTRE_2012_2013 (240), ANTES_2000 (241), ENTRE_2008_2011 (228), ENTRE_2004_2007 (241), ENTRE_2000_2003 (243)
Motivo de viaje	Deportes / Sports (237), Turismo / Tourism (238), Educación / Education (217), Aventura / Adventure tourism (255), Comercio / Buy in stores (244), Cultural / Culture (221), Docencia / Teaching (14)
Dias de permanencia	MEDIO (537), MUY_BAJO (92), BAJO (553), ALTO (218), MUY_ALTO (26)
Compañeros de viaje	MEDIO (206), BAJO (368), MUY_BAJO (212), ALTO (640)
Como se entero de la ciudad	Agencia de viajes / Travel agency (265), Empresa donde labora / Your employer (292), Internet / Internet (292), Television / TV (306), Recomendado / Friend (271)
Lugar alojamiento	Amigos / Friends (369), Familia / Family (369), Hotel / Hotel (352), Hostal / Hostal (336)
Actividad planeada que realizo	Cultural / Culture (234), Comercio / Buy in stores (240), Aventura / Adventure tourism (241), Docencia / Teaching (219), Educación / Education (230), Turismo / Tourism (262)
Medio de transporte para llegar a la ciudad	Transporte publico / Public transport (497), Avión / Travelling by plane (452), Transporte Privado / private transport (477)
Transporte para desplazarse en la ciudad	Transporte publico / Public transport (367), Transporte privado / Private transport (334), Bicicleta / Bicycle (392), Caminar / Walk (333)
Sitio de alimentacion	Familia / Family (270), Amigos / Friends (285), Comida rápida en la calle / Fast food in the street (285), Comida rápida en restaurantes / Fast food in restaurants (286), Restaurantes / Restaurants (300)
Comida preferida	Comida picante o condimentada / Spicy food (267), Mariscos / Seafood (292), Pollo / Chicken (288), Carnes Rojas / Meat (291), Pastas o pizzas / Italian food (288)

Figura 53. Estadística del cluster cero.

Fuente: Resultado obtenido desde el software RapidMiner.

RuleModel

```

if Transporte para desplazarse en la ciudad = Transporte privado / Private transport and Motivo de viaje = Deportes / Sports then Familia / Family (7 / 26 / 12 / 12)
if Edad = ENTRE_60_70 and Dias de permanencia = BAJO then Amigos / Friends (71 / 49 / 39 / 52)
if Dias de permanencia = BAJO and Medio de transporte para llegar a la ciudad = Avión / Traveling by plane then Hotel / Hotel (33 / 31 / 41 / 18)
if Actividad planeada que realizo = Turismo / Tourism and Motivo de viaje = Aventura / Adventure tourism then Familia / Family (8 / 18 / 7 / 7)
if Medio de transporte para llegar a la ciudad = Transporte publico / Public transport and Nivel estudios = Primaria / Elementary School then Hotel / Hotel (5 / 8 / 15 / 7)
if Transporte para desplazarse en la ciudad = Transporte privado / Private transport and Edad = ENTRE_80_90 then Familia / Family (2 / 11 / 2 / 5)
if Transporte para desplazarse en la ciudad = Transporte publico / Public transport and Sitio de alimentacion = Amigos / Friends then Familia / Family (9 / 18 / 9 / 7)
if Actividad planeada que realizo = Turismo / Tourism and Edad = ENTRE_80_90 then Hostal / Hostal (0 / 4 / 1 / 9)
if Actividad planeada que realizo = Docencia / Teaching and Compañeros de viaje = BAJO then Hotel / Hotel (5 / 7 / 19 / 5)
if Edad = ENTRE_70_80 and Dias de permanencia = MUY_BAJO then Hostal / Hostal (5 / 4 / 7 / 16)
if Como se entero de la ciudad = Internet / Internet and Compañeros de viaje = BAJO then Amigos / Friends (14 / 7 / 4 / 5)
if Edad = ENTRE_60_70 and Compañeros de viaje = ALTO then Familia / Family (21 / 36 / 21 / 15)
if Motivo de viaje = Aventura / Adventure tourism and # veces que ha viajado a la ciudad = BAJO then Hostal / Hostal (3 / 5 / 5 / 15)
if Estado civil = Viudo / Widowed and # veces que ha viajado a la ciudad = BAJO then Hotel / Hotel (15 / 12 / 29 / 15)
if Sitio de alimentacion = Amigos / Friends and Transporte para desplazarse en la ciudad = Caminar / Walk then Amigos / Friends (12 / 4 / 2 / 3)
if Dias de permanencia = MEDIO and Sitio de alimentacion = Restaurantes / Restaurants. then Hostal / Hostal (10 / 8 / 10 / 23)
if Nivel estudios = Maestria / Masters and Motivo de viaje = Cultural / Culture then Familia / Family (0 / 5 / 1 / 0)
if Medio de transporte para llegar a la ciudad = Avión / Traveling by plane and Actividad planeada que realizo = Aventura / Adventure tourism then Amigos / Friends (7 / 0 / 2 / 1)
if Como se entero de la ciudad = Empresa donde labora / Your employer. and Medio de transporte para llegar a la ciudad = Transporte Privado / private transport then Familia / Family
if Transporte para desplazarse en la ciudad = Caminar / Walk and Como se entero de la ciudad = Recomendado / Friend. then Hotel / Hotel (2 / 5 / 8 / 1)
if Medio de transporte para llegar a la ciudad = Transporte Privado / private transport and Transporte para desplazarse en la ciudad = Bicicleta / Bicycle then Hostal / Hostal (8 / 6)
if # veces que ha viajado a la ciudad = ALTO and Transporte para desplazarse en la ciudad = Bicicleta / Bicycle then Amigos / Friends (13 / 2 / 5 / 4)
if Como se entero de la ciudad = Agencia de viajes / Travel agency and Medio de transporte para llegar a la ciudad = Transporte Privado / private transport then Hotel / Hotel (4 / 2)
if Compañeros de viaje = BAJO and Estado civil = Viudo / Widowed then Familia / Family (5 / 11 / 3 / 4)
if Compañeros de viaje = MUY_BAJO and Nivel estudios = Universitario / College degree. then Hotel / Hotel (1 / 0 / 5 / 1)
if Nivel estudios = Primaria / Elementary School and Estado civil = Divorciado / Divorced then Hostal / Hostal (0 / 0 / 2 / 6)
if Motivo de viaje = Cultural / Culture and Edad = ENTRE_70_80 then Familia / Family (3 / 7 / 2 / 1)
if Medio de transporte para llegar a la ciudad = Avión / Traveling by plane and Estado civil = Divorciado / Divorced then Amigos / Friends (13 / 4 / 6 / 5)
if Nivel estudios = Secundaria / High school and Estado civil = Divorciado / Divorced then Hostal / Hostal (0 / 1 / 0 / 4)
if Motivo de viaje = Comercio / Buy in stores and Transporte para desplazarse en la ciudad = Transporte publico / Public transport then Familia / Family (0 / 4 / 2 / 0)
if Nivel estudios = Ninguno / None and Edad = ENTRE_60_70 then Hostal / Hostal (1 / 0 / 0 / 3)
if Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants. and Como se entero de la ciudad = Internet / Internet then Hotel / Hotel (0 / 0 / 5 / 3)
if Dias de permanencia = MUY_BAJO and Compañeros de viaje = MEDIO then Familia / Family (0 / 4 / 0 / 0)
if Compañeros de viaje = MEDIO and Estado civil = Divorciado / Divorced then Amigos / Friends (6 / 1 / 1 / 2)

```

Figura 54. Reglas de inducción del Cluster cero.

Fuente: Resultado obtenido desde el software RapidMiner.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convict.
120	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.783	0.874	0.940	-1.009	-0.012	0.985	0.896
181	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.783	0.883	0.945	-0.991	-0.012	0.985	0.887
413	Edad = ENTRE_70_80	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.508	1	1	-0.508	0.053	1.116	∞
416	Edad = ENTRE_70_80	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.508	1	1	-0.508	0.057	1.127	∞
421	Edad = ENTRE_70_80	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.508	1	1	-0.508	0.110	1.277	∞
422	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Edad = ENTRE_70_80	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.508	1	1	-0.508	0.057	1.127	∞
423	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Edad = ENTRE_70_80	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.508	1	1	-0.508	0.053	1.116	∞
369	Estado civil = Viudo / Widowed	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.473	0.908	0.969	-0.568	0.006	1.014	1.132
276	Estado civil = Viudo / Widowed	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.465	0.894	0.964	-0.576	0.003	1.007	1.060
174	Estado civil = Divorciado / Divorced	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.422	0.883	0.962	-0.534	-0.006	0.985	0.885
153	Estado civil = Divorciado / Divorced	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.421	0.880	0.961	-0.536	-0.004	0.992	0.939
5	Estado civil = Viudo / Widowed	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.417	0.802	0.932	-0.623	0.010	1.024	1.094
177	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Estado civil = Viudo / Widowed	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.417	0.883	0.962	-0.528	-0.002	0.995	0.963
306	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Estado civil = Viudo / Widowed	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.417	0.897	0.967	-0.513	0.001	1.001	1.012
254	Compañeros de viaje = ALTO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.400	0.891	0.966	-0.498	-0.003	0.994	0.949
194	Compañeros de viaje = ALTO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.397	0.884	0.964	-0.501	-0.001	0.997	0.976
61	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Estado civil = Viudo / Widowed	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.365	0.864	0.960	-0.480	-0.010	0.974	0.829
74	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Estado civil = Divorciado / Divorced	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.365	0.867	0.961	-0.477	-0.012	0.967	0.778
103	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Compañeros de viaje = ALTO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.348	0.870	0.963	-0.452	-0.007	0.981	0.870
133	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Compañeros de viaje = ALTO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.348	0.876	0.965	-0.446	-0.008	0.978	0.839
300	Dias de permanencia = BAJO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.348	0.897	0.971	-0.428	0.000	1.001	1.007
138	Dias de permanencia = BAJO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.340	0.877	0.966	-0.435	-0.004	0.989	0.918
307	Dias de permanencia = MEDIO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.338	0.898	0.972	-0.415	0.004	1.012	1.102
198	Dias de permanencia = MEDIO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.333	0.885	0.968	-0.420	-0.004	0.987	0.899
205	Medio de transporte para llegar a la ciudad = Transporte publico / Public	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.309	0.885	0.970	-0.388	-0.004	0.988	0.905
107	Medio de transporte para llegar a la ciudad = Transporte publico / Public	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.304	0.871	0.967	-0.393	-0.006	0.982	0.877
333	Medio de transporte para llegar a la ciudad = Transporte Privado / private	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.302	0.901	0.975	-0.367	0.002	1.006	1.053
59	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Dias de permanencia = BAJO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.300	0.863	0.965	-0.396	-0.008	0.973	0.824
172	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Dias de permanencia = BAJO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.300	0.882	0.970	-0.380	-0.005	0.985	0.883
235	Medio de transporte para llegar a la ciudad = Transporte Privado / private	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.297	0.889	0.972	-0.372	0.001	1.002	1.016
110	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Dias de permanencia = BAJO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.295	0.871	0.968	-0.381	-0.008	0.972	0.807
192	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Dias de permanencia = BAJO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.295	0.884	0.971	-0.372	-0.001	0.997	0.975
311	# veces que ha viajado a la ciudad = MEDIO	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.290	0.898	0.975	-0.356	0.004	1.012	1.107
242	# veces que ha viajado a la ciudad = MEDIO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.288	0.889	0.973	-0.359	-0.002	0.992	0.938
335	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.286	0.903	0.977	-0.348	0.002	1.007	1.066
336	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.286	0.903	0.977	-0.348	0.005	1.018	1.160
349	# veces que ha viajado a la ciudad = ALTO	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.268	0.905	0.978	-0.324	0.003	1.010	1.095
135	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	Ocupacion = Pensionado-Jubilado / Pensioner-Retired	0.264	0.877	0.971	-0.339	-0.003	0.988	0.916
241	Ocupacion = Pensionado-Jubilado / Pensioner-Retired, Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired	0.264	0.889	0.975	-0.330	-0.002	0.992	0.936

Figura 55. Datos de validación de las reglas del clúster cero.

Fuente: Resultado obtenido desde el software RapidMiner.

Reglas de Asociación: Se describen las primeras cinco reglas de asociación.

[Sitio de alimentación = Amigos / Friends] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.800)

[Actividad planeada que realizo = Comercio / Buy in stores] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.800)

[Estado civil = Viudo / Widowed, Dias de permanencia = BAJO] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.801)

[Nivel estudios = Especialización / Specialization] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.801)

[Estado civil = Viudo / Widowed] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupacion = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.802)

Análisis sobre los resultados del clúster cero:

- Este grupo está conformado por tres grupos de personas entre 60-70, 70-80 y 80-90 años.
- El clúster está conformado por viudos y divorciados.
- Cuando tienen bajo nivel de permanencia coincide que tienen bajo nivel de acompañantes y bajo nivel de viajes anteriores a la ciudad.
- El número de veces que han viajado a la ciudad es bajo, medio y alto.

- Hospedaje en hotel, hostel y familiares.
- Cuando se quedan pocos días es debido a que llegan en avión, en la ciudad para transportarse utilizan transporte privado, seguramente de sus amigos o alquiler, entre las actividades que realizan cuando viajan en su transporte privado son: deportes y aventura.
- Utilizan la bicicleta y caminan en la ciudad, es posible que utilicen las ciclo vías los fines de semana o que su visita sea los fines de semana, también que en las noches utilicen la bicicleta con los grupos de ciclistas que salen diariamente por la ciudad o que permanezcan donde sus amigos en un sector o barrio en el cual puedan caminar o utilizar la bicicleta para ir a la tienda o sectores seguros.

Análisis sobre las reglas de asociación del clúster cero:

De acuerdo con la tabla de la Figura 55. Datos de validación de las reglas del clúster cero., podemos observar que tenemos reglas con un soporte elevado, del 0.417 al 0.783 que indica entre un 41% y un 70% de las transacciones o viajeros que cumplen dicha regla en el clúster cero y un nivel de confianza entre 0.88 a 1 lo que indica que tantas veces sucede dicha regla, en este caso las reglas suceden entre el 80% y 100% de las veces siendo un número alto de probabilidad de que realmente ocurra.

Las siguientes reglas tienen un nivel de confianza entre el 98% y el 100%:

En el 40% de los viajes del clúster cero los viajeros son viudos o divorciados y esa es su fuente de ingresos.

El 70% de los viajeros es pensionado retirado entre 60 y 90 años.

El 34% de los pensionados viajaron un número alto de compañeros de viaje.

El 34% de los viajeros, pensionados tuvo una permanencia media y baja en la ciudad.

La descripción de las actividades planeadas tiene un 98% de confianza:

El 20% de los viajeros realizo turismo de aventura.

El 20% realizo actividades de comercio.

El 15% de los viajeros realizo actividades de docencia.

El 15% de los viajeros realizo actividades de educación.

El 15% de los viajeros realizo actividades turismo.

El 15% de los viajeros realizo actividades de cultura.

Como se enteraron de la ciudad con un nivel de confianza del 88% al 91%:

Agencia de viajes: 20%

Empresa donde labora: 20%

Internet: 20%

Recomendado por un amigo: 20%

Televisión: 20%

Entre los pensionados y divorciados utilizaron medios de transporte para llegar a la ciudad con un nivel de confianza entre el 85% y el 89%:

Avión: 30%

Transporte privado: 40%

Transporte público: 30%

Entre los pensionados y divorciados que utilizaron medios de transporte para desplazarse dentro de la ciudad con un nivel de confianza entre el 82% y el 93%:

Bicicleta: 25%

Caminar:	32%
Transporte privado:	40%
Transporte público:	33%

Entre los pensionados y divorciados que utilizaron diferentes sitios de alimentación con un nivel de confianza entre el 87% y el 89%:

Amigos:	20%
Comida rápida en la calle:	20%
Comida rápida en restaurantes:	20%
Familia:	20%
Restaurantes:	20%

Entre los pensionados viudos y divorciados entre 60 a 90 años se observa unos gustos muy definidos, posiblemente las personas que utilizan los medios de transporte como la bicicleta o que caminan son los que se encuentran entre 60 y 70 años.

La visita a la ciudad se hace en grupos de personas y el tiempo de estadía es corta, utilizan familiares o amigos para la estadía y fue recomendado por amigos y por búsquedas por internet.

Este grupo podría ser mayor en el caso de incentivar más sitios turísticos como museos, actividades culturales para personas de este rango de edades, analizar por parte de los agentes de turismo que es lo más interesante para las personas de dicha edad y de esta manera ofrecer planes más acordes con sus gustos.

Análisis del Clúster 1: El Clúster uno tiene un total de 1195 viajeros.

Name	Range
Nombre	[3.000 : 5118.000]
outlier	falsa (1195), true (0)
cluster	cluster_0 (0), cluster_1 (1195), cluster_2 (0), cluster_3 (0)
ubicación territorial	ASIA_DEL_NORTE (75), ASIA_MERIDIONAL (72), SURAMERICA (70), ASIA_ORIENTAL (73), AFRICA_ORIENTAL (79), EUROPA (67), NORTEAMERICA (74), OCEANIA (85), CENTROAMERICA (80), ASIA_SUROCCIDENTAL (71), AFRICA_CENTRAL (79), AFRICA_DEL_NORTE (75), AFRICA_DEL_SUR (76)
Edad	ENTRE_50_70 (52), ENTRE_70_80 (0), ENTRE_30_40 (238), ENTRE_50_50 (241), MENOR_18 (120), ENTRE_18_21 (109), ENTRE_40_50 (212), ENTRE_22_30 (224), ENTRE_80_90 (0)
Sexo	M/Male (570), F/Female (628)
Estado civil	Divorciado / Divorced (166), Viudo / Widowed (160), Soltero / Single (431), Casado / Married (439)
Nivel estudios	Especialización / Specialization (145), Secundaria / High school (157), Doctorado / Doctorate (81), Ninguno / None (146), Primaria / Elementary School (150), Universitario / College degree (124), Maestría / Masters (89), Técnico / Technician (166), Tecnología / Technologist (13)
Ocupacion	Pensionado-Jubilado / Pensioner-Retired (0), Desempleado / Unemployed (309), Empleado / Employee (294), Estudiante / Student (303), Independiente / Independent (290)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (0), Familia / Family-Parents (255), Estudiante / Student (260), Independiente / Independent (264), Desempleado / Unemployed (228), Empleado / Employment (189)
# veces que ha viajado a la ciudad	ALTO (375), MEDIO (377), BAJO (347), MUY_ALTO (33), MUY_BAJO (64)
Fecha de viaje	ANO_2014 (169), ENTRE_2012_2013 (204), ANTES_2000 (199), ENTRE_2008_2011 (224), ENTRE_2004_2007 (202), ENTRE_2009_2003 (196)
Motivo de viaje	Deportes / Sports (198), Turismo / Tourism (203), Educación / Education (227), Aventura / Adventure tourism (201), Comercio / Buy in stores (192), Cultural / Culture (168), Docencia / Teaching (7)
Días de permanencia	MEDIO (457), MUY_BAJO (76), BAJO (425), ALTO (220), MUY_ALTO (18)
Compañeros de viaje	MEDIO (160), BAJO (310), MUY_BAJO (200), ALTO (526)
Como se entero de la ciudad	Agencia de viajes / Travel agency (240), Empresa donde labora / Your employer (230), Internet / Internet (236), Television / TV (241), Recomendado / Friend (249)
Lugar alojamiento	Amigos / Friends (305), Familia / Family (291), Hotel / Hotel (281), Hostal / Hostel (319)
Actividad planeada que realizo	Cultural / Culture (198), Comercio / Buy in stores (185), Aventura / Adventure tourism (204), Docencia / Teaching (209), Educación / Education (201), Turismo / Tourism (199)
Medio de transporte para llegar a la ciudad	Transporte publico / Public transport (1195), Auto / Travelling by plane (0), Transporte Privado / private transport (0)
Transporte para desplazarse en la ciudad	Transporte publico / Public transport (307), Transporte privado / Private transport (298), Bicicleta / Bicycle (289), Caminar / Walk (302)
Sitio de alimentacion	Familia / Family (245), Amigos / Friends (241), Comida rápida en la calle / Fast food in the street (242), Comida rápida en restaurantes / Fast food in restaurants (239), Restaurantes / Restaurants (229)
Comida preferida	Comida picante o condimentada / Spicy food (215), Mariscos / Seafood (260), Pollo / Chicken (250), Carnes Rojas / Meat (237), Pastas o pizzas / Italian food (234)

Figura 56. Estadística del cluster uno.

Fuente: Resultado obtenido desde el software RapidMiner.

```

RuleModel

if Ocupacion = Desempleado / Unemployed and Nivel estudios = Ninguno / None then Hostal / Hostel ( 9 / 6 / 4 / 19)
if Nivel estudios = Tecnología / Technologist and Sitio de alimentacion = Amigos / Friends then Hostal / Hostel ( 2 / 5 / 6 / 13)
if Edad = ENTRE_50_60 and # veces que ha viajado a la ciudad = ALTO then Amigos / Friends (28 / 13 / 16 / 20)
if Motivo de viaje = Comercio / Buy in stores and Como se entero de la ciudad = Internet / Internet then Hostal / Hostel ( 4 / 2 / 6 / 15)
if Motivo de viaje = Aventura / Adventure tourism and Compañeros de viaje = BAJO then Familia / Family ( 8 / 22 / 9 / 8)
if Como se entero de la ciudad = Empresa donde labora / Your employer. and Dias de permanencia = ALTO then Amigos / Friends ( 20 / 7 / 10 / 7)
if Actividad planeada que realizo = Docencia / Teaching and Dias de permanencia = ALTO then Hostal / Hostel ( 3 / 4 / 6 / 14)
if # veces que ha viajado a la ciudad = ALTO and Edad = ENTRE_30_40 then Familia / Family (13 / 26 / 12 / 13)
if Compañeros de viaje = MUY_BAJO and Fuente de ingreso = Familia / Family-Parents then Amigos / Friends (13 / 5 / 4 / 3)
if Transporte para desplazarse en la ciudad = Transporte publico / Public transport and Edad = ENTRE_40_50 then Hotel / Hotel (10 / 9 / 19 / 11)
if Ocupacion = Empleado / Employee and Nivel estudios = Maestría / Masters then Hostal / Hostel ( 2 / 4 / 2 / 9)
if Como se entero de la ciudad = Agencia de viajes / Travel agency and Motivo de viaje = Comercio / Buy in stores then Amigos / Friends (15 / 4 / 6 / 10)
if Compañeros de viaje = ALTO and Fuente de ingreso = Empleado / Employment then Familia / Family ( 8 / 22 / 11 / 12)
if Transporte para desplazarse en la ciudad = Caminar / Walk and Fuente de ingreso = Independiente / Independent then Amigos / Friends (15 / 5 / 5 / 5)
if Sitio de alimentacion = Familia / Family and Como se entero de la ciudad = Television / TV then Hotel / Hotel ( 7 / 1 / 15 / 6)
if Estado civil = Casado / Married and Compañeros de viaje = BAJO then Familia / Family (14 / 26 / 7 / 16)
if Dias de permanencia = BAJO and Compañeros de viaje = BAJO then Hotel / Hotel ( 5 / 7 / 21 / 10)
if Edad = ENTRE_30_40 and Nivel estudios = Primaria / Elementary School then Amigos / Friends ( 8 / 1 / 2 / 1)
if Dias de permanencia = MEDIO and Sitio de alimentacion = Familia / Family then Hostal / Hostel ( 4 / 9 / 9 / 17)
if Dias de permanencia = MEDIO and Como se entero de la ciudad = Empresa donde labora / Your employer. then Amigos / Friends (14 / 10 / 4 / 6)
if Motivo de viaje = Aventura / Adventure tourism and Estado civil = Divorciado / Divorced then Hotel / Hotel ( 0 / 0 / 5 / 1)
if Actividad planeada que realizo = Docencia / Teaching and Edad = ENTRE_22_30 then Hostal / Hostel ( 2 / 1 / 4 / 7)
if Fuente de ingreso = Independiente / Independent and # veces que ha viajado a la ciudad = ALTO then Familia / Family ( 2 / 10 / 5 / 1)
if Motivo de viaje = Aventura / Adventure tourism and Dias de permanencia = MEDIO then Amigos / Friends ( 9 / 1 / 2 / 2)
if Edad = ENTRE_30_40 and Sitio de alimentacion = Comida rápida en la calle. / Fast food in the street. then Hostal / Hostel ( 4 / 1 / 3 / 9)
if # veces que ha viajado a la ciudad = MEDIO and Compañeros de viaje = ALTO then Familia / Family (12 / 26 / 10 / 7)
if Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants. and Edad = MENOR_18 then Hotel / Hotel ( 1 / 0 / 6 / 1)
if Dias de permanencia = MEDIO and Ocupacion = Desempleado / Unemployed then Hostal / Hostel ( 5 / 5 / 3 / 12)
if Nivel estudios = Técnico / Technician and Estado civil = Soltero / Single then Hotel / Hotel ( 2 / 1 / 9 / 2)
if Edad = ENTRE 40 50 and # veces que ha viajado a la ciudad = ALTO then Amigos / Friends ( 9 / 2 / 4 / 0)
    
```

Figura 57. Reglas de inducción del Cluster uno.

Fuente: Resultado obtenido desde el software RapidMiner.

No.	Premises	Conclusion	Support	Confidence
1	Compañeros de viaje = ALTO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.440	1
2	Días de permanencia = MEDIO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.382	1
3	Estado civil = Casado / Married	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.367	1
4	Estado civil = Soltero / Single	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.360	1
5	Días de permanencia = BAJO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.355	1
6	# veces que ha viajado a la ciudad = MEDIO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.315	1
7	# veces que ha viajado a la ciudad = ALTO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.314	1
8	# veces que ha viajado a la ciudad = BAJO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.290	1
9	Lugar alojamiento = Hostal / Hostal	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.267	1
10	Compañeros de viaje = BAJO	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.259	1
11	Ocupacion = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.258	1
12	Transporte para desplazarse en la ciudad = Transporte publico / Public transport	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.257	1
13	Lugar alojamiento = Amigos / Friends	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.255	1
14	Ocupacion = Estudiante / Student	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.253	1
15	Transporte para desplazarse en la ciudad = Caminar / Walk	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.253	1
16	Transporte para desplazarse en la ciudad = Transporte privado / Private transport	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.249	1
17	Ocupacion = Empleado / Employee	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.246	1
18	Lugar alojamiento = Familia / Family	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.243	1
19	Ocupacion = Independiente / Independent	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.242	1
20	Transporte para desplazarse en la ciudad = Bicicleta / Bicycle	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.242	1
21	Lugar alojamiento = Hotel / Hotel	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.235	1
22	Fuente de ingreso = Independiente / Independent	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.221	1
23	Fuente de ingreso = Estudiante / Student	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.217	1
24	Fuente de ingreso = Familia / Family-Parents	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.213	1
25	Como se entero de la ciudad = Recomendado / Friend.	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.208	1
26	Sitio de alimentacion = Familia / Family	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.205	1
27	Sitio de alimentacion = Comida rápida en la calle. / Fast food in the street.	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.202	1
28	Sitio de alimentacion = Amigos / Friends	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.202	1
29	Edad = ENTRE_50_60	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.202	1
30	Como se entero de la ciudad = Television / TV	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.202	1
31	Como se entero de la ciudad = Agencia de viajes / Travel agency	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.201	1
32	Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants.	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.200	1
33	Edad = ENTRE_30_40	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.199	1
34	Como se entero de la ciudad = Internet / Internet	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.197	1
35	Como se entero de la ciudad = Empresa donde labora / Your employer.	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.192	1
36	Sitio de alimentacion = Restaurantes / Restaurants.	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.191	1
37	Fuente de ingreso = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.191	1
38	Motivo de viaje = Educación / Education	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.190	1
39	Edad = ENTRE_22_30	Medio de transporte para llegar a la ciudad = Transporte publico / Public transport	0.187	1

Figura 58. Datos de validación de las reglas del clúster uno.

Fuente: Resultado obtenido desde el software RapidMiner.

Reglas de Asociación: Se describen las 5 primeras reglas de asociación generadas.

[Compañeros de viaje = ALTO] --> [Medio de transporte para llegar a la ciudad = Transporte público / Public transport] (confidence: 1.000).

[Días de permanencia = MEDIO] --> [Medio de transporte para llegar a la ciudad = Transporte público / Public transport] (confidence: 1.000).

[Estado civil = Casado / Married] --> [Medio de transporte para llegar a la ciudad = Transporte publico / Public transport] (confidence: 1.000).

[Estado civil = Soltero / Single] --> [Medio de transporte para llegar a la ciudad = Transporte publico / Public transport] (confidence: 1.000).

[Días de permanencia = BAJO] --> [Medio de transporte para llegar a la ciudad = Transporte publico / Public transport] (confidence: 1.000)

Análisis sobre los resultados del clúster uno:

- Este grupo está conformado por personas entre menores de 18, entre 18-21, 22-30, 30-40, 40-50, 50-60 y 60-70.
- El clúster está conformado por su mayoría mujeres.
- El estado civil que tiene más viajeros es casado y soltero, lo que le siguen son viudo y divorciado.
- Tenemos viajeros con todos los niveles de estudio y todas las ocupaciones.
- Las principales fuentes de ingresos en orden de cantidad son: Independiente, familia, estudiante, desempleado, empleado.
- El nivel de viajes es alto, medio y bajo.
- Cuando tienen bajo nivel de permanencia coincide que tienen bajo nivel de acompañantes y bajo nivel de viajes anteriores a la ciudad.
- El número de veces que han viajado a la ciudad es bajo, medio y alto.
- Hospedaje principalmente en: hostel y amigos.
- Principalmente cuando el motivo de viaje es aventura se hospedan con familiares y

amigos, cuando es para comercio en un hostel.

- Cuando son técnicos y tecnólogos se alojan en hostales.
- Cuando tienen un nivel de estudios elemental como primaria, se hospedan en hoteles.
- Cuando no tienen estudios se alojan donde los amigos.

Análisis sobre las reglas de asociación del clúster uno: De acuerdo con la tabla de la Figura 55. Datos de validación de las reglas del clúster cero., podemos observar que tenemos reglas con un soporte intermedio, del 0.200 al 0.440 que indica entre un 20% y un 44% de las transacciones o viajeros que cumplen dicha regla en el clúster uno y un nivel de confianza de 1 lo que indica que la credibilidad de que ocurra es del 100%.

La descripción de las actividades planeadas tiene un 100% de confianza:

El 17% de los viajeros realizo turismo de aventura.

El 16% realizo actividades de comercio.

El 18% de los viajeros realizo actividades de docencia.

El 17% de los viajeros realizo actividades de educación.

El 16% de los viajeros realizo actividades turismo.

El 17% de los viajeros realizo actividades de cultura.

Como se enteraron de la ciudad con un nivel de confianza del 100%:

Agencia de viajes: 20%

Empresa donde labora: 20%

Internet: 20%

Recomendado por un amigo: 20%

Televisión: 20%

Entre los viajeros que utilizaron medios de transporte para llegar a la ciudad con un nivel de confianza del 100%:

Avión:	0%
Transporte privado:	0%
Transporte público:	100%

Medios de transporte para desplazarse dentro de la ciudad con un nivel de confianza del 100%:

Bicicleta:	30%
Caminar:	30%
Transporte privado:	10%
Transporte público:	30%

Los diferentes sitios de alimentación con un nivel de confianza del 100%:

Amigos:	20%
Comida rápida en la calle:	20%
Comida rápida en restaurantes:	20%
Familia:	20%
Restaurantes:	20%

En este clúster todos utilizaron el transporte público para llegar a la ciudad.

Entre los estados civiles más representativos tenemos dos: solteros y casados.

Los solteros tienen un tiempo bajo de permanencia y un nivel bajo de compañeros de viaje, mientras que los casados tienen un nivel alto de permanencia y un nivel alto de compañeros de viaje.

Podemos argumentar que los casados viajan con su familia, con sus hijos, primos de sus hijos, con los abuelos de los hijos, de esta manera se sustentaría el uso de transporte público el cual es muy económico a comparación de los tiquetes de avión.

Cuando los viajeros tienen una permanencia alta, el uso de transporte para desplazarse en la ciudad está muy equilibrado, utilizan bicicleta, caminan, transporte público y transporte privado.

En el caso anterior podría ser determinante si hacen salidas de la ciudad o si las visitas son en la zona urbana, el caminar y la bicicleta se podría determinar como principal medio de transporte en barrios o fincas cerca de la ciudad, también alguna vereda o condominios, esto debido a que la ciudad no cuenta con vías para transitar con bicicleta por la ciudad como se realiza en la ciudad de Bogotá o Medellín (Turismo., 2015).

Las entidades de turismo tienen la posibilidad de ofrecer paquetes turísticos familiares de aventura para todas las edades, de esta manera fomentar el uso de hostales, hoteles y ecohoteles de la región, así como lo realiza el Quindío con sus parques y ecohoteles en los cuales pueden comprar acciones (Quindio., 2015).

Otra recomendación para los transportadores privados es la de ofrecer servicio familiar, normalmente el transporte público por medio de taxis solo ofrecen puestos para 4 personas y poca carga, en la región no hay gran cantidad de vehículos de más capacidad como en otras ciudades como Bogotá o Medellín dedicados, como se puede ver en la ciudad de Bogotá en la cual las camionetas y vehículos de gran tamaño aprovechan las plataformas tecnológicas como Uber para localizar clientes (Uber, 2015).

Análisis del Clúster 2: El Clúster dos tiene un total de 1265 viajeros.

Name	Range
Nombre	[4.000 ; 5121.000]
outlier	false (1265), true (0)
cluster	cluster_0 (0), cluster_1 (0), cluster_2 (1265), cluster_3 (0)
ubicación territorial	ASIA_DEL_NORTE (80), ASIA_MERIDIONAL (80), SURAMERICA (80), ASIA_ORIENTAL (83), AFRICA_ORIENTAL (76), EUROPA (68), NORTEAMERICA (104), OCEANIA (66), CENTROAMERICA (73), ASIA_SURORIENTAL (83), AFRICA_CENTRAL (76), AFRICA_DEL_NORTE (82), AF
Edad	ENTRE_60_70 (60), ENTRE_70_80 (0), ENTRE_30_40 (251), ENTRE_50_60 (261), MENOR_18 (142), ENTRE_18_21 (119), ENTRE_40_50 (222), ENTRE_22_30 (210), ENTRE_80_90 (0)
Sexo	M/Male (638), F/Female (627)
Estado civil	Divorciado / Divorced (186), Viudo / Widowed (180), Soltero / Single (456), Casado / Married (443)
Nivel estudios	Especialización / Specialization (136), Secundaria / High school (144), Doctorado / Doctorate (196), Ninguno / None (175), Primaria / Elementary School (158), Universitario / College degree (132), Maestría / Masters (78), Técnico / Technician (197), Tecnología / Technologist (147)
Ocupación	Pensionado-Jubilado / Pensioner-Retired (0), Desempleado / Unemployed (301), Empleado / Employee (307), Estudiante / Student (219), Independiente / Independent (338)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (0), Familia / Family-Parents (274), Estudiante / Student (267), Independiente / Independent (265), Desempleado / Unemployed (282), Empleado / Employment (177)
# veces que ha viajado a la ciudad	ALTO (399), MEDIO (409), BAJO (333), MUY_ALTO (39), MUY_BAJO (85)
Fecha de viaje	ANO_2014 (198), ENTRE_2012_2013 (222), ANTES_2000 (204), ENTRE_2008_2011 (214), ENTRE_2004_2007 (208), ENTRE_2000_2003 (219)
Motivo de viaje	Deportes / Sports (227), Turismo / Tourism (205), Educación / Education (204), Aventura / Adventure tourism (205), Comercio / Buy in stores (200), Cultural / Culture (206), Docencia / Teaching (18)
Días de permanencia	MEDIO (487), BAJO (76), BAJO (477), ALTO (302), MUY_ALTO (13)
Compañeros de viaje	MEDIO (194), BAJO (320), MUY_BAJO (170), ALTO (588)
Como se enteró de la ciudad	Agencia de viajes / Travel agency (254), Empresa donde labora / Your employer (273), Internet / Internet (245), Televisión / TV (236), Recomendado / Friend (257)
Lugar alojamiento	Amigos / Friends (320), Familia / Family (312), Hotel / Hotel (318), Hostal / Hostel (315)
Actividad planeada que realizó	Cultural / Culture (231), Comercio / Buy in stores (191), Aventura / Adventure tourism (208), Docencia / Teaching (207), Educación / Education (219), Turismo / Tourism (209)
Medio de transporte para llegar a la ciudad	Transporte público / Public transport (0), Avión / Traveling by plane (1265), Transporte Privado / private transport (0)
Transporte para desplazarse en la ciudad	Transporte público / Public transport (322), Transporte privado / Private transport (316), Bicicleta / Bicycle (318), Caminar / Walk (399)
Sitio de alimentación	Familia / Family (261), Amigos / Friends (251), Comida rápida en la calle / Fast food in the street (236), Comida rápida en restaurantes / Fast food in restaurants (254), Restaurantes / Restaurants (261)

Figura 59. Estadística del cluster dos.

Fuente: Resultado obtenido desde el software RapidMiner.

No.	Premises	Conclusion	Support	Confidence
1	Compañeros de viaje = ALTO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.458	1
2	Días de permanencia = MEDIO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.393	1
3	Días de permanencia = BAJO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.377	1
4	Estado civil = Soltero / Single	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.360	1
5	Estado civil = Casado / Married	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.350	1
6	# veces que ha viajado a la ciudad = MEDIO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.323	1
7	# veces que ha viajado a la ciudad = ALTO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.315	1
8	Ocupacion = Independiente / Independent	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.267	1
9	# veces que ha viajado a la ciudad = BAJO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.263	1
10	Compañeros de viaje = BAJO	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.258	1
11	Transporte para desplazarse en la ciudad = Transporte publico / Public transport	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.255	1
12	Lugar alojamiento = Amigos / Friends	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.253	1
13	Ocupacion = Estudiante / Student	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.252	1
14	Transporte para desplazarse en la ciudad = Bicicleta / Bicycle	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.251	1
15	Lugar alojamiento = Hotel / Hotel	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.251	1
16	Transporte para desplazarse en la ciudad = Transporte privado / Private transport	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.250	1
17	Lugar alojamiento = Hostal / Hostal	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.249	1
18	Lugar alojamiento = Familia / Family	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.247	1
19	Transporte para desplazarse en la ciudad = Caminar / Walk	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.244	1
20	Ocupacion = Empleado / Employee	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.243	1
21	Ocupacion = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.238	1
22	Fuente de ingreso = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.223	1
23	Fuente de ingreso = Familia / Family-Parents	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.217	1
24	Como se entero de la ciudad = Empresa donde labora / Your employer.	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.216	1
25	Fuente de ingreso = Estudiante / Student	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.211	1
26	Fuente de ingreso = Independiente / Independent	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.209	1
27	Sitio de alimentacion = Restaurantes / Restaurants.	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.206	1
28	Sitio de alimentacion = Familia / Family	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.206	1
29	Edad = ENTRE_50_60	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.206	1
30	Como se entero de la ciudad = Recomendado / Friend.	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.203	1
31	Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants.	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.201	1
32	Como se entero de la ciudad = Agencia de viajes / Travel agency	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.201	1
33	Sitio de alimentacion = Amigos / Friends	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.198	1
34	Edad = ENTRE_30_40	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.198	1
35	Como se entero de la ciudad = Internet / Internet	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.194	1
36	Sitio de alimentacion = Comida rápida en la calle. / Fast food in the street.	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.188	1
37	Como se entero de la ciudad = Television / TV	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.187	1
38	Actividad planeada que realizo = Cultural / Culture	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.183	1
39	Motivo de viaje = Deportes / Sports	Medio de transporte para llegar a la ciudad = Avión / Traveling by plane	0.179	1

Figura 60. Reglas de inducción del Cluster dos.

Fuente: Resultado obtenido desde el software RapidMiner.

```

Text View Annotations

RuleModel

if Edad = ENTRE_50_60 and Compañeros de viaje = ALTO then Amigos / Friends (49 / 39 / 23 / 20)
if Como se entero de la ciudad = Empresa donde labora / Your employer. and Estado civil = Casado / Married then Hostal / Hostal (11 / 26 / 20 / 34)
if Actividad planeada que realizo = Docencia / Teaching and Transporte para desplazarse en la ciudad = Transporte publico / Public transport then Hotel / Hotel (9 / 9 / 19 / 4)
if Transporte para desplazarse en la ciudad = Transporte publico / Public transport and Compañeros de viaje = BAJO then Hostal / Hostal (15 / 7 / 11 / 29)
if Transporte para desplazarse en la ciudad = Caminar / Walk and Compañeros de viaje = BAJO then Hotel / Hotel (16 / 13 / 27 / 15)
if # veces que ha viajado a la ciudad = BAJO and Sitio de alimentacion = Restaurantes / Restaurants. then Amigos / Friends (16 / 7 / 8 / 5)
if Compañeros de viaje = BAJO and Estado civil = Divorciado / Divorced then Familia / Family (2 / 10 / 0 / 1)
if Edad = ENTRE_40_50 and Fuente de ingreso = Estudiante / Student then Hotel / Hotel (5 / 2 / 15 / 10)
if Compañeros de viaje = BAJO and # veces que ha viajado a la ciudad = ALTO then Familia / Family (9 / 21 / 9 / 13)
if Ocupacion = Desempleado / Unemployed and Sitio de alimentacion = Comida rápida en la calle. / Fast food in the street. then Amigos / Friends (17 / 6 / 7 / 4)
if Dias de permanencia = MEDIO and Transporte para desplazarse en la ciudad = Transporte publico / Public transport then Hostal / Hostal (13 / 15 / 12 / 23)
if Ocupacion = Estudiante / Student and Fuente de ingreso = Estudiante / Student then Hotel / Hotel (5 / 6 / 17 / 5)
if Transporte para desplazarse en la ciudad = Bicicleta / Bicycle and Nivel estudios = Ninguno / None then Amigos / Friends (11 / 3 / 6 / 3)
if Compañeros de viaje = ALTO and Ocupacion = Independiente / Independent then Hostal / Hostal (12 / 15 / 22 / 30)
if Fuente de ingreso = Independiente / Independent and Dias de permanencia = MEDIO then Familia / Family (6 / 15 / 7 / 10)
if Dias de permanencia = MEDIO and Como se entero de la ciudad = Agencia de viajes / Travel agency then Amigos / Friends (12 / 4 / 4 / 6)
if Como se entero de la ciudad = Agencia de viajes / Travel agency and Dias de permanencia = BAJO then Familia / Family (6 / 18 / 9 / 7)
if Edad = ENTRE_30_40 and Sitio de alimentacion = Amigos / Friends then Amigos / Friends (9 / 3 / 4 / 1)
if Edad = ENTRE_22_30 and Como se entero de la ciudad = Television / TV then Hotel / Hotel (4 / 6 / 11 / 4)
if Nivel estudios = Secundaria / High school and Dias de permanencia = MEDIO then Amigos / Friends (7 / 1 / 0 / 1)
if Nivel estudios = Universitario / College degree. and # veces que ha viajado a la ciudad = ALTO then Hostal / Hostal (3 / 1 / 2 / 6)
if Estado civil = Casado / Married and Nivel estudios = Especialización / Specialization then Familia / Family (1 / 7 / 3 / 1)
if # veces que ha viajado a la ciudad = BAJO and Fuente de ingreso = Independiente / Independent then Hostal / Hostal (4 / 2 / 2 / 7)
if Compañeros de viaje = MUY_BAJO and Dias de permanencia = BAJO then Hotel / Hotel (7 / 9 / 16 / 10)
if Fuente de ingreso = Independiente / Independent and Estado civil = Soltero / Single then Amigos / Friends (10 / 2 / 2 / 3)
if Edad = ENTRE_40_50 and Como se entero de la ciudad = Internet / Internet then Familia / Family (2 / 8 / 2 / 1)
if Estado civil = Soltero / Single and Edad = ENTRE_40_50 then Hostal / Hostal (2 / 5 / 0 / 6)
if Como se entero de la ciudad = Internet / Internet and Nivel estudios = Universitario / College degree. then Hotel / Hotel (0 / 1 / 5 / 1)
if Actividad planeada que realizo = Cultural / Culture and Motivo de viaje = Comercio / Buy in stores then Amigos / Friends (5 / 0 / 2 / 0)
if Fuente de ingreso = Empleado / Employment and Dias de permanencia = BAJO then Hostal / Hostal (4 / 0 / 1 / 7)
if Edad = ENTRE_30_40 and Ocupacion = Independiente / Independent then Hotel / Hotel (0 / 1 / 5 / 2)
if Nivel estudios = Ninguno / None and Edad = ENTRE_18_21 then Familia / Family (0 / 4 / 0 / 1)
if Como se entero de la ciudad = Television / TV and # veces que ha viajado a la ciudad = MEDIO then Amigos / Friends (10 / 3 / 1 / 3)
if # veces que ha viajado a la ciudad = MEDIO and Como se entero de la ciudad = Internet / Internet then Hotel / Hotel (1 / 2 / 8 / 1)

```

Figura 61. Datos de validación de las reglas del clúster dos.

Fuente: Resultado obtenido desde el software RapidMiner.

Reglas de Asociación: Se describen las 5 primeras reglas de asociación generadas.

Association Rules [Compañeros de viaje = ALTO] --> [Medio de transporte para llegar a la ciudad = Avión / Traveling by plane] (confidence: 1.000).

[Dias de permanencia = MEDIO] --> [Medio de transporte para llegar a la ciudad = Avión / Traveling by plane] (confidence: 1.000).

[Dias de permanencia = BAJO] --> [Medio de transporte para llegar a la ciudad = Avión / Traveling by plane] (confidence: 1.000).

[Estado civil = Soltero / Single] --> [Medio de transporte para llegar a la ciudad = Avión / Traveling by plane] (confidence: 1.000).

[Estado civil = Casado / Married] --> [Medio de transporte para llegar a la ciudad = Avión / Traveling by plane] (confidence: 1.000).

Análisis sobre los resultados del clúster dos:

- Este grupo está conformado por personas entre menores de 18, entre 18-21, 22-30, 30-40, 40-50, 50-60 y 60-70.
- El clúster está conformado por su mayoría hombres.
- El estado civil que tiene más viajeros es casado y soltero, los que le siguen son divorciado y viudo.
- Tenemos viajeros con todos los niveles de estudio y todas las ocupaciones, los principales niveles de estudio del clúster son: Técnicos y le siguen los que no tienen profesión.
- Las principales fuentes de ingresos en orden de cantidad son: Familia, desempleado, independiente, empleado.
- El nivel de viajes es medio y alto, siendo bajo el tercero con una gran diferencia de cantidad.
- El número de días de permanencia en la ciudad es: medio y bajo principalmente.
- Los sitios de hospedaje son muy parejos, el orden por diferencias mínimas es: amigos, hotel, hostel y familia.
- Principalmente cuando el motivo de viaje es docencia, utilizan transporte público y se hospedan en hotel.
- Cuando no tienen estudios, utilizan la bicicleta para desplazarse y se hospedan con los amigos.

- Si se enteró de la ciudad con una agencia de viajes, los días de permanencia son bajos y se hospeda en familia.
- Si la edad esta entre 30 y 40 años el sitio de alimentación es con los amigos.

En este clúster se observa una gran dependencia de la familia, en cuanto al hospedaje, alimentación, transporte y se asume que los viajeros visitan la ciudad con familiares cercanos.

Ya que la principal fuente de ingresos y la mayoría son hombres, se puede deducir que hay una fuerte dependencia económica de su familia, quizás culturalmente es normal para los viajeros ya que en muchos países los hijos de más de 30 años aún viven con sus padres, dependen económicamente de ellos y viajan con sus padres aunque se encuentren casados (Press, 2015).

Este grupo es interesante, ya que de acuerdo con una gran clasificación que realizaron para este tipo de nuevos jóvenes adultos el comercio tendría un potencial para la venta de artículos como juegos, videojuegos, actividades recreativas, es decir, actividades que podríamos tomar como de niños (Kindsein, 2015).

Análisis sobre las reglas de asociación del clúster dos: De acuerdo con la tabla de la Figura 55. Datos de validación de las reglas del clúster cero., podemos observar que tenemos reglas con un soporte entre 0.201 al 0.458 que indica entre un 20% y un 46% de las transacciones o viajeros que cumplen dicha regla en el clúster dos y un nivel de confianza de 1 lo que indica que la credibilidad de que ocurra es del 100%.

La descripción de las actividades planeadas tiene un 100% de confianza:

El 17% de los viajeros realizo turismo de aventura.

El 15% realizo actividades de comercio.

El 16% de los viajeros realizo actividades de docencia.

El 17% de los viajeros realizo actividades de educación.

El 17% de los viajeros realizo actividades turismo.

El 18% de los viajeros realizo actividades de cultura.

Como se enteraron de la ciudad con un nivel de confianza del 100%:

Agencia de viajes: 20%

Empresa donde labora: 22%

Internet: 19%

Recomendado por un amigo: 20%

Televisión: 19%

Entre los viajeros que utilizaron medios de transporte para llegar a la ciudad con un nivel de confianza del 100%:

Avión: 100%

Transporte privado: 0%

Transporte público: 0%

Medios de transporte para desplazarse dentro de la ciudad con un nivel de confianza del 100%:

Bicicleta: 25%

Caminar: 25%

Transporte privado: 25%

Transporte público: 25%

Los diferentes sitios de alimentación con un nivel de confianza del 100%:

Amigos: 20%

Comida rápida en la calle: 20%

Comida rápida en restaurantes: 20%

Familia: 20%

Restaurantes: 20%

En este clúster todos utilizaron avión para llegar a la ciudad.

Entre los estados civiles más representativos tenemos dos: solteros y casados.

Si el estado civil es soltero o casado, tienen unos días de permanencia bajos y por lo tanto número de veces que han viajado medio y alto, cuando los días de permanencia son medios el número de veces que ha viajado son alto, medio y bajo.

En este grupo se clasificarían viajeros con grandes acompañantes o familiares, con un buen número de días que han visitado la ciudad, posiblemente como de paso o como ciudad desde la cual visitan otras partes del país, debido a su tranquilidad, al paisaje cafetero, bajo costo, poca contaminación, fincas cafeteras, grandes montañas verdes (Vanguardia, 2015).

Con las características descritas del clúster 2 se deduce que disponen de recursos económicos para viajar en avión, utilizar servicios de hotelería más exclusivos, destinos gastronómicos más exquisitos.

Análisis del Clúster 3: El Clúster tres tiene un total de 1226 viajeros.

Name	Range
Nombre	{5.000; 5122.000}
outlier	false (1226), true (0)
cluster	cluster_0 (0), cluster_1 (0), cluster_2 (0), cluster_3 (1226)
ubicación territorial	ASIA_DEL_NORTE (64), ASIA_MERIDIONAL (71), SURAMERICA (89), ASIA_ORIENTAL (72), AFRICA_ORIENTAL (80), EUROPA (90), NORTEAMERICA (80), OCEANIA (83), CENTROAMERICA (70), ASIA_SURORIENTAL (79), AFRICA_CENTRAL (71), AFRICA_DEL_NORTE (74), AFRICA...
Edad	ENTRE_40_70 (56), ENTRE_70_80 (0), ENTRE_30_40 (233), ENTRE_50_60 (225), MENOR_18 (139), ENTRE_18_21 (89), ENTRE_40_50 (254), ENTRE_22_30 (228), ENTRE_80_90 (0)
Sexo	M / Male (626); F / Female (600)
Estado civil	Divorciado / Divorced (148), Viudo / Widowed (169), Soltero / Single (465), Casado / Married (444)
Nivel estudios	Especialización / Specialization (135), Secundaria / High school (138), Doctorado / Doctorate (93), Ninguno / None (167), Primaria / Elementary School (157), Universitario / College degree (152), Maestría / Masters (84), Técnico / Technician (137), Tecnología / Technologist (153)
Ocupacion	Pensionado-Jubilado / Pensioner-Retired (1), Desempleado / Unemployed (285), Empleado / Employee (318), Estudiante / Student (315), Independiente / Independent (307)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (1), Familia / Family-Parents (252), Estudiante / Student (274), Independiente / Independent (253), Desempleado / Unemployed (245), Empleado / Employment (200)
# veces que ha viajado a la ciudad	ALTO (386), MEDIO (399), BAJO (324), MUY_ALTO (37), MUY_BAJO (80)
Fecha de viaje	ANO_2014 (210), ENTRE_2012_2013 (214), ANTES_2000 (207), ENTRE_2008_2011 (191), ENTRE_2004_2007 (189), ENTRE_2000_2003 (215)
Motivo de viaje	Deportes / Sports (211), Turismo / Tourism (206), Educación / Education (191), Aventura / Adventure tourism (200), Comercio / Buy in stores (217), Cultural / Culture (198), Docencia / Teaching (13)
Días de permanencia	MEDIO (480), MUY_BAJO (84), BAJO (438), ALTO (200), MUY_ALTO (24)
Compañeros de viaje	MEDIO (183), BAJO (334), MUY_BAJO (168), ALTO (541)
Como se entero de la ciudad	Agencia de viajes / Travel agency (254), Empresa donde labora / Your employer (243), Internet / Internet (250), Television / TV (225), Recomendado / Friend (254)
Lugar alojamiento	Amigos / Friends (339), Familia / Family (235), Hotel / Hotel (298), Hostal / Hostel (294)
Actividad planeada que realizo	Cultural / Culture (209), Comercio / Buy in stores (226), Aventura / Adventure tourism (211), Docencia / Teaching (189), Educación / Education (207), Turismo / Tourism (193)
Medio de transporte para llegar a la ciudad	Transporte publico / Public transport (0), Avión / Traveling by plane (0), Transporte Privado / private transport (1226)
Transporte para desplazarse en la ciudad	Transporte publico / Public transport (288), Transporte privado / Private transport (313), Bicicleta / Bicycle (311), Caminar / Walk (304)
Sitio de alimentacion	Familia / Family (248), Amigos / Friends (256), Comida rápida en la calle / Fast food in the street (229), Comida rápida en restaurantes / Fast food in restaurants (240), Restaurantes / Restaurants (253)

Figura 62. Estadística del cluster tres.

Fuente: Resultado obtenido desde el software RapidMiner.

RuleModel

```

if Transporte para desplazarse en la ciudad = Transporte publico / Public transport and Fuente de ingreso = Estudiante / Student then Amigos / Friends (36 / 15 / 14 / 6)
if Dias de permanencia = BAJO and Nivel estudios = Ninguno / None then Amigos / Friends (25 / 12 / 12 / 13)
if Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants, and Motivo de viaje = Comercio / Buy in stores then Amigos / Friends (18 / 8 / 7 / 9)
if Motivo de viaje = Turismo / Tourism and # veces que ha viajado a la ciudad = BAJO then Hostal / Hostel (11 / 12 / 11 / 24)
if Fuente de ingreso = Familia / Family-Parents and Actividad planeada que realizo = Cultural / Culture then Hotel / Hotel (6 / 6 / 17 / 7)
if Fuente de ingreso = Empleado / Employment and Ocupacion = Desempleado / Unemployed then Amigos / Friends (21 / 5 / 7 / 13)
if Estado civil = Soltero / Single and Fuente de ingreso = Desempleado / Unemployed then Familia / Family (16 / 27 / 10 / 14)
if # veces que ha viajado a la ciudad = MEDIO and Fuente de ingreso = Independiente / Independent then Hotel / Hotel (8 / 11 / 29 / 19)
if Como se entero de la ciudad = Internet / Internet and Dias de permanencia = BAJO then Familia / Family (9 / 19 / 9 / 8)
if Compañeros de viaje = ALTO and Como se entero de la ciudad = Empresa donde labora / Your employer, then Amigos / Friends (28 / 16 / 13 / 13)
if Dias de permanencia = BAJO and Ocupacion = Empleado / Employee then Hotel / Hotel (13 / 2 / 20 / 9)
if Motivo de viaje = Deportes / Sports and Actividad planeada que realizo = Turismo / Tourism then Familia / Family (0 / 8 / 1 / 3)
if Dias de permanencia = BAJO and Transporte para desplazarse en la ciudad = Caminar / Walk then Hostal / Hostel (5 / 7 / 5 / 14)
if Actividad planeada que realizo = Educación / Education and Edad = ENTRE_40_50 then Familia / Family (2 / 11 / 2 / 3)
if Como se entero de la ciudad = Agencia de viajes / Travel agency and Actividad planeada que realizo = Aventura / Adventure tourism then Amigos / Friends (11 / 2 / 5 / 3)
if Transporte para desplazarse en la ciudad = Bicicleta / Bicycle and Motivo de viaje = Aventura / Adventure tourism then Hotel / Hotel (8 / 3 / 15 / 5)
if Estado civil = Soltero / Single and Motivo de viaje = Deportes / Sports then Familia / Family (5 / 13 / 1 / 9)
if Edad = ENTRE_40_50 and Transporte para desplazarse en la ciudad = Bicicleta / Bicycle then Hostal / Hostel (2 / 4 / 5 / 13)
if Sitio de alimentacion = Comida rápida en la calle, / Fast food in the street, and Actividad planeada que realizo = Comercio / Buy in stores then Amigos / Friends (7 / 0 / 3 / 2)
if Nivel estudios = Especialización / Specialization and Ocupacion = Independiente / Independent then Familia / Family (4 / 10 / 3 / 2)
if Transporte para desplazarse en la ciudad = Transporte privado / Private transport and Ocupacion = Independiente / Independent then Hotel / Hotel (6 / 5 / 15 / 8)
if Compañeros de viaje = ALTO and Motivo de viaje = Cultural / Culture then Familia / Family (10 / 15 / 3 / 5)
if Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants, and # veces que ha viajado a la ciudad = BAJO then Hostal / Hostel (4 / 0 / 2 / 10)
if Motivo de viaje = Cultural / Culture and Transporte para desplazarse en la ciudad = Transporte privado / Private transport then Hotel / Hotel (1 / 2 / 8 / 3)
if Compañeros de viaje = BAJO and Motivo de viaje = Educación / Education then Amigos / Friends (8 / 2 / 1 / 0)
if Ocupacion = Independiente / Independent and Transporte para desplazarse en la ciudad = Transporte publico / Public transport then Familia / Family (6 / 10 / 4 / 1)
if Actividad planeada que realizo = Turismo / Tourism and Motivo de viaje = Comercio / Buy in stores then Hostal / Hostel (1 / 2 / 3 / 8)
if Transporte para desplazarse en la ciudad = Bicicleta / Bicycle and Como se entero de la ciudad = Agencia de viajes / Travel agency then Hotel / Hotel (1 / 1 / 10 / 3)
if Motivo de viaje = Educación / Education and Actividad planeada que realizo = Aventura / Adventure tourism then Hostal / Hostel (0 / 2 / 0 / 6)
if Sitio de alimentacion = Familia / Family and Ocupacion = Estudiante / Student then Amigos / Friends (9 / 3 / 5 / 0)
if Sitio de alimentacion = Comida rápida en la calle, / Fast food in the street, and Fuente de ingreso = Estudiante / Student then Familia / Family (2 / 6 / 1 / 0)
if Transporte para desplazarse en la ciudad = Transporte privado / Private transport and Fuente de ingreso = Estudiante / Student then Hostal / Hostel (2 / 2 / 0 / 9)
if Dias de permanencia = BAJO and # veces que ha viajado a la ciudad = MEDIO then Hotel / Hotel (1 / 1 / 5 / 1)
if Ocupacion = Empleado / Employee and Compañeros de viaje = MEDIO then Amigos / Friends (6 / 1 / 1 / 1)
if Fuente de ingreso = Desempleado / Unemployed and Transporte para desplazarse en la ciudad = Bicicleta / Bicycle then Familia / Family (2 / 6 / 0 / 2)
if Actividad planeada que realizo = Comercio / Buy in stores and Ocupacion = Estudiante / Student then Hotel / Hotel (0 / 1 / 5 / 0)
if Motivo de viaje = Turismo / Tourism and Sitio de alimentacion = Familia / Family then Hostal / Hostel (0 / 0 / 0 / 5)

```

Figura 63. Reglas de inducción del Cluster tres.

Fuente: Resultado obtenido desde el software RapidMiner.

No.	Premises	Conclusion	Support	Confidence
1	Compañeros de viaje = ALTO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.441	1
2	Dias de permanencia = MEDIO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.392	1
3	Estado civil = Soltero / Single	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.379	1
4	Estado civil = Casado / Married	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.362	1
5	Dias de permanencia = BAJO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.357	1
6	# veces que ha viajado a la ciudad = MEDIO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.325	1
7	# veces que ha viajado a la ciudad = ALTO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.315	1
8	Lugar alojamiento = Amigos / Friends	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.277	1
9	Compañeros de viaje = BAJO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.272	1
10	# veces que ha viajado a la ciudad = BAJO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.264	1
11	Ocupacion = Empleado / Employee	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.259	1
12	Ocupacion = Estudiante / Student	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.257	1
13	Transporte para desplazarse en la ciudad = Transporte privado / Private transport	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.255	1
14	Transporte para desplazarse en la ciudad = Bicicleta / Bicycle	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.254	1
15	Ocupacion = Independiente / Independent	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.250	1
16	Transporte para desplazarse en la ciudad = Caminar / Walk	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.248	1
17	Transporte para desplazarse en la ciudad = Transporte publico / Public transport	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.243	1
18	Lugar alojamiento = Hotel / Hotel	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.243	1
19	Lugar alojamiento = Familia / Family	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.241	1
20	Lugar alojamiento = Hostal / Hostal	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.240	1
21	Ocupacion = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.232	1
22	Fuente de ingreso = Estudiante / Student	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.223	1
23	Sitio de alimentacion = Amigos / Friends	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.209	1
24	Edad = ENTRE_40_50	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.207	1
25	Como se entero de la ciudad = Recomendado / Friend.	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.207	1
26	Como se entero de la ciudad = Agencia de viajes / Travel agency	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.207	1
27	Sitio de alimentacion = Restaurantes / Restaurants.	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.206	1
28	Fuente de ingreso = Independiente / Independent	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.206	1
29	Fuente de ingreso = Familia / Family-Parents	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.206	1
30	Como se entero de la ciudad = Internet / Internet	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.204	1
31	Fuente de ingreso = Familia / Family-Parents	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.202	1
32	Fuente de ingreso = Desempleado / Unemployed	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.201	1
33	Compañeros de viaje = ALTO, Dias de permanencia = MEDIO	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.199	1
34	Como se entero de la ciudad = Empresa donde labora / Your employer.	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.198	1
35	Sitio de alimentacion = Comida rápida en restaurantes / Fast food in restaurants.	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.196	1
36	Edad = ENTRE_30_40	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.190	1
37	Sitio de alimentacion = Comida rápida en la calle. / Fast food in the street	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.187	1
38	Edad = ENTRE_22_30	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.186	1
39	Actividad planeada que realizo = Comercio / Buy in stores	Medio de transporte para llegar a la ciudad = Transporte Privado / private transport	0.184	1

Figura 64. Datos de validación de las reglas del clúster tres.

Fuente: Resultado obtenido desde el software RapidMiner.

Reglas de Asociación: Se describen las 5 primeras reglas de asociación generadas.

[Compañeros de viaje = ALTO] --> [Medio de transporte para llegar a la ciudad = Transporte Privado / private transport] (confidence: 1.000).

[Dias de permanencia = MEDIO] --> [Medio de transporte para llegar a la ciudad = Transporte Privado / private transport] (confidence: 1.000).

[Estado civil = Soltero / Single] --> [Medio de transporte para llegar a la ciudad = Transporte Privado / private transport] (confidence: 1.000).

[Estado civil = Casado / Married] --> [Medio de transporte para llegar a la ciudad = Transporte Privado / private transport] (confidence: 1.000).

[Días de permanencia = BAJO] --> [Medio de transporte para llegar a la ciudad = Transporte Privado / private transport] (confidence: 1.000)

Análisis sobre los resultados del clúster tres:

- La mayor cantidad de viajeros se encuentran en el rango de edades de forma descendente: 40-50, 30-40, 22-30, 50-60.
- El clúster está conformado por su mayoría hombres, pero la diferencia con las mujeres es mínima.
- El estado civil que tiene más viajeros es soltero y casado, los que siguen en forma descendente son viudo y divorciado pero con una diferencia muy marcada.
- Tenemos viajeros con todos los niveles de estudio y todas las ocupaciones, siendo los niveles de estudios con un número muy marcado: ninguno, primaria, tecnólogo, universitario.
- La principal fuente de ingreso es “Estudiante”, continúan con una diferencia marcada: Independiente, desempleado, empleado.
- El nivel de viajes es alto y medio.
- Cuando tienen bajo nivel de permanencia coincide que no tiene estudios y se aloja con amigos.
- Si la fuente de ingresos es la familia, entonces realiza actividades culturales.
- Hospedaje principalmente en orden descendente: amigos, hotel, familia.
- Cuando el motivo de viaje es deportes se hospedan con la familia, cuando es turismo se hospedan en hostal.
- Cuando la ocupación son empleados, el nivel de viaje es medio, y se hospedan donde los

amigos, cuando son independientes se alojan donde familiares y en la ciudad se desplazan en transporte público.

- Cuando tienen un nivel de estudios elemental como primaria, se hospedan en hoteles.
- En este clúster utilizan mucho el alojamiento con amigos.

Análisis sobre las reglas de asociación del clúster tres: De acuerdo con la tabla de la Figura 55. Datos de validación de las reglas del clúster cero., podemos observar que tenemos reglas con un soporte intermedio, del 0.201 al 0.441 que indica entre un 20% y un 44% de las transacciones o viajeros que cumplen dicha regla en el clúster uno y un nivel de confianza de 1 lo que indica que la credibilidad de que ocurra es del 100%.

La descripción de las actividades planeadas tiene un 100% de confianza:

El 17% de los viajeros realizo turismo de aventura.

El 19% realizo actividades de comercio.

El 15% de los viajeros realizo actividades de docencia.

El 17% de los viajeros realizo actividades de educación.

El 16% de los viajeros realizo actividades turismo.

El 16% de los viajeros realizo actividades de cultura.

Como se enteraron de la ciudad con un nivel de confianza del 100%:

Agencia de viajes: 21%

Empresa donde labora: 19%

Internet: 21%

Recomendado por un amigo: 21%

Televisión: 18%

Entre los viajeros que utilizaron medios de transporte para llegar a la ciudad con un nivel de confianza del 100%:

Avión:	0%
Transporte privado:	100%
Transporte público:	0%

Medios de transporte para desplazarse dentro de la ciudad con un nivel de confianza del 100%:

Bicicleta:	26%
Caminar:	25%
Transporte privado:	26%
Transporte público:	23%

Los diferentes sitios de alimentación con un nivel de confianza del 100%:

Amigos:	21%
Comida rápida en la calle:	18%
Comida rápida en restaurantes:	20%
Familia:	20%
Restaurantes:	21%

En este clúster todos utilizaron el transporte privado para llegar a la ciudad.

Entre los estados civiles más representativos tenemos dos: casados y solteros.

Los solteros tienen un nivel alto de compañeros de viaje.

Podemos comentar que este clúster está muy marcado por las siguientes características:

Se encuentran principalmente en etapa productiva y en etapa en la cual tienen muy claro el

rumbo de su vida (A., 2000).

Los viajeros tienen muchos amigos en la región, utilizan mucho las agencias de viajes, referencias de los amigos y el internet, es posible que este grupo utilice muy bien la tecnología, que utilicen plataformas de viajes y hospedaje con personas de todo el mundo los cuales alquilan sus cuartos (Airbnb, s.f.).

El uso del transporte privado para llegar a la ciudad nos puede dar un porcentaje mayor de claridad en cuanto al uso de plataformas, ya que para un viajero que no conoce un país o una ciudad, tiene más sentido alquilar un transporte privado o que algún amigo lo transporte durante su estadía.

Internamente el uso del vehículo privado es muy marcado, seguido por el uso de la Bicicleta.

Los mayores sitios de alimentación de los viajeros del clúster son en restaurantes normales y con sus amigos.

En este clúster podríamos tener en cuenta que los viajeros son personas de mundo, ahorradores, conocedores de culturas, personas que se sienten más a gusto al hospedarse con familias alrededor del mundo, que utilizan las plataformas y la tecnología para compartir y alquilar recursos a precios más económicos y con un calor de familia.

Es posible que este grupo de personas utilice dichas herramientas por el mal servicio de los hoteles, costos elevados, impuestos, inseguridad, para sentir la cultura directamente desde un núcleo familiar (Quintero,2015).

Esta sería la oportunidad para las agencias de viajes y los hoteles de revisar sus políticas, sus costos, la calidad de sus servicios y de esta manera permitir que las personas inconformes o

que quieren ahorrar dinero puedan alojarse en hoteles diseñados específicamente para esto.

A continuación analizaremos los Outliers:

ExampleSet (10 examples, 2 special attributes, 15 regular attributes)	
Name	Range
Nombre	[36.000 ; 2936.000]
outlier	false (0), true (10)
Edad	ENTRE_50_70 (0), ENTRE_70_80 (0), ENTRE_30_40 (2), ENTRE_50_60 (4), MENOR_18 (1), ENTRE_18_21 (0), ENTRE_40_50 (2), ENTRE_22_30 (1), ENTRE_60_90 (0)
Estado civil	Divorciado / Divorced (1), Viudo / Widowed (7), Soltero / Single (0), Casado / Married (2)
Nivel estudios	Especialización / Specialization (0), Secundaria / High school (1), Doctorado / Doctorate (2), Ninguno / None (2), Primaria / Elementary School (1), Universitario / College degree (1), Maestría / Masters (0), Técnico / Technician (0), Tecnología / Technologist (3)
Ocupación	Pensionado-Jubilado / Pensioner-Retired (0), Desempleado / Unemployed (2), Empleado / Employee (2), Estudiante / Student (4), Independiente / Independent (2)
Fuente de ingreso	Pensionado-Jubilado / Pensioner-Retired (0), Familia / Family-Parents (1), Estudiante / Student (1), Independiente / Independent (1), Desempleado / Unemployed (2), Empleado / Employment (5)
# veces que ha viajado a la ciudad	ALTO (1), MEDIO (0), BAJO (0), MUY_ALTO (5), MUY_BAJO (4)
Motivo de viaje	Deportes / Sports (5), Turismo / Tourism (0), Educación / Education (3), Aventura / Adventure tourism (0), Comercio / Buy in stores (2), Cultural / Culture (0), Docencia / Teaching (0)
Días de permanencia	MEDIO (0), MUY_BAJO (6), BAJO (1), ALTO (2), MUY_ALTO (1)
Compañeros de viaje	MEDIO (9), BAJO (1), MUY_BAJO (0), ALTO (0)
Como se enteró de la ciudad	Agencia de viajes / Travel agency (4), Empresa donde labora / Your employer (2), Internet / Internet (3), Televisión / TV (1), Recomendado / Friend (0)
Lugar alojamiento	Amigos / Friends (3), Familia / Family (2), Hotel / Hotel (2), Hostal / Hostel (3)
Actividad planeada que realizo	Cultural / Culture (0), Comercio / Buy in stores (2), Aventura / Adventure tourism (2), Docencia / Teaching (2), Educación / Education (3), Turismo / Tourism (1)
Medio de transporte para llegar a la ciudad	Transporte público / Public transport (3), Avión / Traveling by plane (2), Transporte Privado / private transport (5)
Transporte para desplazarse en la ciudad	Transporte público / Public transport (1), Transporte privado / Private transport (4), Bicicleta / Bicycle (1), Caminar / Walk (4)
Sitio de alimentación	Familia / Family (2), Amigos / Friends (1), Comida rápida en la calle / Fast food in the street (3), Comida rápida en restaurantes / Fast food in restaurants (0), Restaurantes / Restaurants (4)

Figura 65. Distribución de Outliers

Fuente: Resultado obtenido desde el software RapidMiner.

Este grupo de 10 personas tiene gran cantidad de información, el hecho de que los hayamos filtrado para poder ejecutar K-Means es debido a las restricciones del algoritmo. En varios tipos de investigación los Outliers son el objetivo propuesto a analizar ya que determinan los casos poco comunes como por ejemplo: determinar transacciones fraudulentas, en este caso las transacciones poco comunes u Outliers darían las pistas para detectar los fraudes.

Al examinar los Outliers se tienen las siguientes observaciones:

- Cuando es estado civil es viudo, el motivo de viajes es educación y con un número medio de compañeros de viaje.
- Viajero se enteró de la ciudad por internet, con una edad entre 50 y 60 años y días de permanencia bajos.
- Número de veces que han viajado muy alto con un número medio de compañeros de viaje y se enteró por agencia de viaje.

Este grupo tiene gran oportunidad de ser explotado, el rango de edad esta entre 50 y 60

años, empleados la mayoría, con días de permanencia muy bajos, se informan en agencias de viaje e internet, la mayoría son viudos y la mayoría utiliza un medio de transporte privado.

El sitio de alojamiento de este clúster está distribuido en las 4 categorías, llama la atención que se informan en agencias de viaje e internet pero utilizan transporte privado para llegar a la ciudad y una vez en la ciudad caminan o utilizan nuevamente el transporte privado. El motivo de viaje está dividido en: Deportes educación y comercio.

Como ejemplo de campaña en las agencias se puede ofrecer un paquete turístico de varios días con vehículo incluido, paseos de aventura y visitas guiadas, de esta manera lograr que los viajeros Outliers regresen y realicen un gasto mayor en la ciudad.

Nombraremos los 5 grupos para identificar los viajeros, de esta manera el lector podrá apuntar diferentes campañas publicitarias, mejorar el servicio o realizar un nuevo análisis a partir del conocimiento que tenga de la ciudad y de la temática tratada.

Grupo 0 (Abuelitos): Adultos mayores, pensionados, viudos y casados, realizan muchos viajes.

Grupo 1 (Familias): Viajeros familiares, llegan a la ciudad en transporte privado y se quedan muchos días, alto número de acompañantes, solteros y casados que tienen ingresos.

Grupo 2 (AdultosEnCasa): Viajeros familiares, grandes grupos, llegan a la ciudad en avión, tiene la mayor tasa de desempleo.

Grupo 3 (Aventureros): Viajeros que se hospedan con los amigos, para viajar, utilizan el transporte privado.

Grupo 4 (Outliers): Casos especiales, en edad cercana al retiro, potenciales clientes para incentivar la visita a la ciudad.

Las agencias de viaje y las entidades interesadas en el turismo y comercio deberían diseñar una campaña para que no solamente los grupos clasificados sino también los Outliers puedan regresar a la ciudad.

Desde este punto en adelante se desarrolla el Objetivo 3: **Determinar en los viajeros que visitan la Ciudad de Pereira el nivel de gastos, percepción y satisfacción luego de visitar la Ciudad por medio de un análisis de clúster.**

Con el desarrollo del tercer objetivo segmentaremos en sub clústers cada uno de los cuatro grupos o clústers generados en el *Desarrollo del Objetivo 1 : Segmentar y describir hábitos de viajeros que visitan la ciudad de Pereira por medio de un Análisis de Clúster para la industria Turística de Pereira*”, para obtener y analizar la percepción del viaje a la ciudad.

El proceso utilizado para filtrar los datos se puede ver en la siguiente imagen:

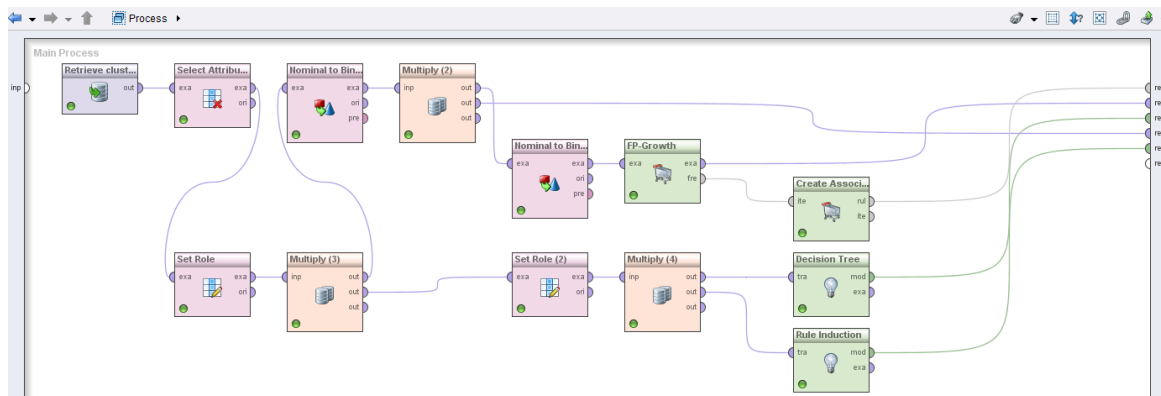


Figura 66. Proceso para Análisis de la Percepción.

Fuente: Resultado obtenido desde el software RapidMiner.

La explicación de cada parte del proceso se omite ya que los componentes se encuentran explicados y definidos en procesos anteriores, lo que si se aclara es en general el objetivo del

proceso y las salidas:

- Se realiza la lectura de cada clúster (clúster cero al clúster tres). No se generan nuevos clústers porque requerimos conocer los grupos creados, es decir los 4 clústers más el clúster de Outliers y analizar cómo fue la percepción que tuvieron de su visita, de esta manera mejorar servicios, crear nuevos o eliminar servicios.
- Se seleccionan los atributos relacionados con la percepción, gasto y satisfacción del viajero (Gasto estimado antes de la visita, calificación del gasto, aspecto que lo impactó, motivación para regresar, oferta de ocio y entretenimiento, oferta comercial, oferta cultural, conservación monumentos, señalización turística, cultura ciudadana, oferta de alojamiento, Disponibilidad de sitios, Calidad del alojamiento, Precio del alojamiento, Información en la Oficina de Turismo, información en el hotel, información en la calle, información por parte de la policía, información por parte del Tránsito, información por parte de la Policía de Carreteras, información en vallas, avisos, letreros.
- Se define un rol con la etiqueta “Label”, de esta manera las reglas de inducción y el árbol de decisión pueden generar resultados y obtendremos datos para su análisis.
- Analizando las reglas de inducción Figura 67. Reglas del Clúster 0 sobre percepción., Figura 68. Reglas del Clúster 1 sobre percepción., Figura 69. Reglas del Clúster 2 sobre percepción., Figura 70. Reglas del Clúster 3 sobre percepción. y la estadística de los valores más altos se llenan la **¡Error! No se encuentra el origen de la referencia.**, Tabla 40. Tabla dos de la percepción del gasto de los viajeros,

Clúster	Precio del alojamiento	Info Oficina Turismo	Info en el hotel.	Info en la calle	Info la policía	Info Transit o	Info Policía Carreteras	Info en vallas
0	Bueno, E	E, Bueno	Bajo, E, B	Bajo, Bueno, E	Bueno, E	Bueno	E, Bajo	Bueno, E, Bajo.
1	Bueno, E	Excelente	Bajo	Bajo, Bueno	Bueno, E	Bueno	E, Bajo	Bueno, E, Bajo.
2	E	Bajo, E, Bueno	Bueno	Bajo	Excelente	Excelente	Bueno	Excelente, Bueno
3	Bueno	Bueno, Bajo	Bueno	Bueno, Excelente	Bajo, Excelente	Bajo, Excelente	Bueno	Excelente
Outliers	Bajo	Bajo	Excelente	Bueno	Bueno	Bueno	Excelente	Excelente

- .
- La última tarea es la del analista, determinar el conocimiento en cada clúster.

A continuación se pueden observar las imágenes de las reglas y las tablas consolidadas en donde se tiene en cuenta la moda para el análisis:

```

RuleModel
if Información por parte de la policía de carreteras = EXCELENTE and Calificación del gasto de la visita = Muy alto / Very high then Mucha indigencia / Many indigent people. (11 / 13 / 7 / 10 / 10 / 21 / 6 / 7 / 7)
if Información por parte de la policía de tránsito = EXCELENTE and Oferta Cultural = BAJO then Mucha indigencia / Many indigent people. (19 / 15 / 7 / 15 / 23 / 32 / 16 / 18 / 17)
if Disponibilidad de sitios de alojamiento = BUENO and Cultura de la Ciudadanía = BUENO then Ciudad insegura / Unsafe city. (11 / 11 / 14 / 7 / 11 / 20 / 15 / 8 / 21)
if Conservación de los Monumentos = BUENO and Calificación del gasto de la visita = Alto / High then Clima agradable / Very nice weather. (6 / 4 / 3 / 16 / 9 / 5 / 5 / 7 / 7)
if Señalización Turística = BUENO and Gasto estimado antes de visita = $100 - $300 Dollars then Ambiente agradable / Very nice city. (5 / 4 / 12 / 1 / 1 / 4 / 2 / 1 / 4)
if Oferta de Alojamiento = EXCELENTE and Información por parte de la policía de tránsito = BAJO then Clima agradable / Very nice weather. (10 / 15 / 13 / 22 / 13 / 11 / 15 / 12 / 11)
if Calidad del alojamiento = EXCELENTE and Gasto estimado antes de visita = $700 - $900 Dollars then Gente amable / Very nice people. (2 / 2 / 3 / 2 / 2 / 1 / 10 / 3 / 5)
if Oferta de Ocio y Entrenamiento = EXCELENTE and Conservación de los Monumentos = BAJO then Mucha indigencia / Many indigent people. (8 / 6 / 10 / 7 / 10 / 19 / 13 / 11 / 11)
if Oferta de Ocio y Entrenamiento = BAJO and Información en el hotel = BAJO then Clima agradable / Very nice weather. (9 / 11 / 10 / 18 / 10 / 9 / 11 / 10 / 14)
if Disponibilidad de sitios de alojamiento = EXCELENTE and Oferta de Ocio y Entrenamiento = BAJO then Ambiente agradable / Very nice city. (8 / 4 / 17 / 9 / 8 / 5 / 6 / 6 / 3)
if Gasto estimado antes de visita = $900 - $1100 Dollars and Señalización Turística = EXCELENTE then Ciudad segura / Safe city. (2 / 2 / 2 / 1 / 7 / 2 / 1 / 1 / 0)
if Conservación de los Monumentos = EXCELENTE and Información en la oficina de turismo = BAJO then Ciudad insegura / Unsafe city. (9 / 9 / 6 / 10 / 5 / 10 / 8 / 11 / 19)
if Información por personas en la calle = EXCELENTE and Información en el hotel = EXCELENTE then Ciudad segura / Safe city. (6 / 3 / 5 / 4 / 11 / 4 / 5 / 2 / 2)
if Disponibilidad de sitios de alojamiento = EXCELENTE and Cultura de la Ciudadanía = BAJO then Gente poco amable / Rude people. (1 / 10 / 3 / 1 / 4 / 4 / 3 / 2 / 5)
if Precio del alojamiento = BUENO and Información por personas en la calle = BAJO then Ciudad Ruidosa / Noisy city. (5 / 4 / 3 / 3 / 2 / 7 / 3 / 14 / 2)
if Disponibilidad de sitios de alojamiento = BUENO and Información en la oficina de turismo = BUENO then Gente amable / Very nice people. (5 / 4 / 4 / 8 / 6 / 6 / 13 / 4 / 5)
if Gasto estimado antes de visita = $700 - $900 Dollars and Calidad del alojamiento = BUENO then Ambiente agradable / Very nice city. (1 / 1 / 6 / 2 / 0 / 0 / 1 / 2 / 0)
if Información por parte de la policía de tránsito = BUENO and Gasto estimado antes de visita = $700 - $900 Dollars then Alta contaminación. / High pollution. (4 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 2)
if Cultura de la Ciudadanía = EXCELENTE and Motivación para regresar = No volvería nunca / Never return then Gente poco amable / Rude people. (1 / 4 / 0 / 1 / 0 / 0 / 0 / 0 / 3)
if Oferta de Alojamiento = BAJO and Señalización Turística = BUENO then Ciudad segura / Safe city. (5 / 1 / 4 / 2 / 9 / 2 / 3 / 2 / 2)
if Información por parte de la policía = BAJO and Información turística en vayas, carteles, volantes, folletos, guías, etc = BUENO then Mucha indigencia / Many indigent people. (1 / 1 / 1 / 1 / 3 / 6 / 2 / 1 / 2)
if Información turística en vayas, carteles, volantes, folletos, guías, etc = BUENO and Información por parte de la policía de carreteras = BUENO then Gente poco amable / Rude people. (0 / 6 / 4 / 1 / 0 / 0 / 0 / 1 / 2)
if Información turística en vayas, carteles, volantes, folletos, guías, etc = EXCELENTE and Calificación del gasto de la visita = Muy bajo / Very low then Clima agradable / Very nice weather. (1 / 2 / 2 / 6 / 0 / 0 / 0 / 1 / 1)
if Gasto estimado antes de visita = $300 - $500 Dollars and Conservación de los Monumentos = EXCELENTE then Mucha indigencia / Many indigent people. (0 / 0 / 1 / 0 / 0 / 4 / 0 / 0 / 0)
if Señalización Turística = BUENO and Cultura de la Ciudadanía = BUENO then Alta contaminación. / High pollution. (5 / 2 / 0 / 0 / 2 / 2 / 0 / 1 / 0)
if Precio del alojamiento = EXCELENTE and Calificación del gasto de la visita = Bajo / Low then Gente amable / Very nice people. (1 / 1 / 2 / 1 / 0 / 3 / 7 / 1 / 0)
if Oferta de Ocio y Entrenamiento = BAJO and Calidad del alojamiento = BUENO then Ciudad insegura / Unsafe city. (0 / 0 / 0 / 1 / 0 / 0 / 0 / 2 / 3)
if Información en la oficina de turismo = BAJO and Conservación de los Monumentos = BAJO then Ciudad segura / Safe city. (1 / 1 / 1 / 0 / 5 / 1 / 0 / 2 / 3)
if Motivación para regresar = Vida nocturna / city nightlife and Señalización Turística = EXCELENTE then Clima agradable / Very nice weather. (0 / 0 / 0 / 3 / 0 / 0 / 0 / 0 / 0)
if Motivación para regresar = La comida / Food and Cultura de la Ciudadanía = EXCELENTE then Gente poco amable / Rude people. (0 / 3 / 1 / 0 / 0 / 0 / 0 / 0 / 0)
if Información en el hotel = BUENO and Calificación del gasto de la visita = Medio / Medium then Alta contaminación. / High pollution. (4 / 1 / 1 / 0 / 0 / 0 / 0 / 1 / 1)
if Oferta de Alojamiento = EXCELENTE and Oferta de Ocio y Entrenamiento = BUENO then Clima agradable / Very nice weather. (0 / 2 / 3 / 6 / 0 / 2 / 0 / 1 / 0)
if Información por parte de la policía de carreteras = EXCELENTE and Información por parte de la policía de tránsito = BAJO then Gente amable / Very nice people. (1 / 0 / 0 / 0 / 0 / 0 / 5 / 0 / 0)
if Gasto estimado antes de visita = $5,000 or more dollars and Oferta Cultural = BUENO then Ciudad segura / Safe city. (0 / 0 / 0 / 0 / 4 / 0 / 0 / 0 / 0)
if Oferta de Ocio y Entrenamiento = BUENO and Motivación para regresar = Paisaje cafetero / Coffee landscape then Ciudad Ruidosa / Noisy city. (0 / 0 / 0 / 0 / 0 / 0 / 0 / 2 / 0)
if Información por parte de la policía de tránsito = BAJO and Información turística en vayas, carteles, volantes, folletos, guías, etc = BUENO then Ciudad insegura / Unsafe city. (0 / 1 / 0 / 0 / 1 / 0 / 1 / 0 / 4)
if Motivación para regresar = Gente amable / Nice people. and Información turística en vayas, carteles, volantes, folletos, guías, etc = BAJO then Mucha indigencia / Many indigent people. (0 / 0 / 0 / 0 / 0 / 0 / 3 / 0 / 0 / 1)

```

Figura 67. Reglas del Clúster 0 sobre percepción.

Fuente: Resultado obtenido desde el software RapidMiner.

```

RuleModel
if Conservación de los Monumentos = BAJO and Calificación del gasto de la visita = Alto / High then Alta contaminación. / High pollution. (9 / 16 / 5 / 3 / 5 / 4 / 5 / 5 / 10)
if Motivación para regresar = No volvería nunca / Never return and Información en la oficina de turismo = BUENO then Mucha indigencia / Many indigent people. (13 / 4 / 5 / 1 / 5 / 4 / 0 / 5 / 3)
if Conservación de los Monumentos = EXCELENTE and Calificación del gasto de la visita = Medio / Medium then Clima agradable / Very nice weather. (9 / 9 / 7 / 3 / 10 / 7 / 16 / 4 / 8)
if Información por personas en la calle = BAJO and Motivación para regresar = Paisaje cafetero / Coffee landscape then Ambiente agradable / Very nice city. (4 / 6 / 4 / 11 / 2 / 3 / 1 / 6 / 3)
if Información en el hotel = BUENO and Calificación del gasto de la visita = Medio alto / Medium high then Gente amable / Very nice people. (1 / 6 / 19 / 4 / 7 / 3 / 9 / 4 / 8)
if Oferta de Alojamiento = BAJO and Información turística en vayas, carteles, volantes, folletos, guías, etc = BAJO then Ciudad segura / Safe city. (8 / 12 / 7 / 10 / 10 / 20 / 10 / 7 / 12)
if Oferta Comercial = BAJO and Cultura de la Ciudadanía = BAJO then Mucha indigencia / Many indigent people. (19 / 6 / 10 / 10 / 6 / 8 / 6 / 11 / 7)
if Disponibilidad de sitios de alojamiento = BAJO and Cultura de la Ciudadanía = BUENO then Gente poco amable / Rude people. (7 / 7 / 8 / 9 / 19 / 5 / 11 / 10 / 10)
if Conservación de los Monumentos = BAJO and Oferta de Ocio y Entrenamiento = BUENO then Alta contaminación. / High pollution. (3 / 13 / 9 / 4 / 6 / 4 / 3 / 4 / 3)
if Motivación para regresar = Bajo costo de vida / low cost of living and Información por personas en la calle = BAJO then Clima agradable / Very nice weather. (0 / 0 / 0 / 1 / 1 / 1 / 6 / 2 / 1)
if Información por parte de la policía = BAJO and Oferta Cultural = BUENO then Ciudad segura / Safe city. (6 / 4 / 7 / 10 / 5 / 15 / 3 / 8 / 9)
if Información por parte de la policía = EXCELENTE and Motivación para regresar = Clima agradable / Nice weather then Clima agradable / Very nice weather. (3 / 2 / 0 / 2 / 1 / 2 / 9 / 2 / 1)
if Gasto estimado antes de visita = $300 - $500 Dollars and Información por parte de la policía de carreteras = BAJO then Mucha indigencia / Many indigent people. (7 / 2 / 1 / 3 / 1 / 3 / 2 / 1 / 0)
if Calificación del gasto de la visita = Muy alto / Very high and Oferta de Ocio y Entrenamiento = BUENO then Ciudad Ruidosa / Noisy city. (0 / 2 / 0 / 4 / 1 / 2 / 2 / 2 / 10)
if Cultura de la Ciudadanía = BUENO and Oferta Cultural = EXCELENTE then Alta contaminación. / High pollution. (5 / 14 / 4 / 8 / 5 / 4 / 3 / 8 / 4)
if Gasto estimado antes de visita = $5,000 or more dollars and Oferta Cultural = EXCELENTE then Gente amable / Very nice people. (1 / 1 / 8 / 1 / 1 / 0 / 2 / 0 / 2)
if Motivación para regresar = No volvería nunca / Never return and Calificación del gasto de la visita = Alto / High then Gente poco amable / Rude people. (0 / 1 / 0 / 1 / 6 / 3 / 0 / 0 / 1)
if Señalización Turística = BUENO and Calidad del alojamiento = EXCELENTE then Mucha indigencia / Many indigent people. (10 / 6 / 7 / 6 / 2 / 2 / 2 / 2 / 0)
if Precio del alojamiento = BUENO and Gasto estimado antes de visita = $1,100 - $1,500 Dollars then Clima agradable / Very nice weather. (2 / 1 / 0 / 1 / 0 / 0 / 6 / 1 / 1)
if Cultura de la Ciudadanía = BUENO and Información por personas en la calle = BUENO then Ciudad Ruidosa / Noisy city. (4 / 1 / 3 / 2 / 1 / 3 / 3 / 3 / 8)
if Cultura de la Ciudadanía = EXCELENTE and Motivación para regresar = Ninguno / None then Gente poco amable / Rude people. (1 / 0 / 4 / 0 / 6 / 1 / 1 / 5 / 1)
if Precio del alojamiento = BAJO and Gasto estimado antes de visita = $900 - $1,100 Dollars then Ciudad segura / Safe city. (2 / 0 / 0 / 1 / 2 / 7 / 0 / 3 / 1)
if Cultura de la Ciudadanía = BAJO and Precio del alojamiento = EXCELENTE then Clima agradable / Very nice weather. (1 / 3 / 5 / 2 / 5 / 4 / 10 / 5 / 4)
if Información por parte de la policía de carreteras = BAJO and Precio del alojamiento = BUENO then Ambiente agradable / Very nice city. (1 / 6 / 1 / 9 / 0 / 1 / 3 / 6 / 1)
if Oferta de Alojamiento = EXCELENTE and Información por personas en la calle = BAJO then Mucha indigencia / Many indigent people. (11 / 1 / 1 / 1 / 3 / 3 / 3 / 2 / 3)
if Oferta de Alojamiento = BUENO and Información por parte de la policía de tránsito = BAJO then Gente poco amable / Rude people. (4 / 2 / 4 / 3 / 8 / 1 / 3 / 0 / 2)
if Oferta Cultural = EXCELENTE and Información en la oficina de turismo = BAJO then Ciudad Ruidosa / Noisy city. (0 / 1 / 0 / 0 / 1 / 2 / 1 / 0 / 6)
if Información turística en vayas, carteles, volantes, folletos, guías, etc = BUENO and Oferta de Ocio y Entrenamiento = BAJO then Alta contaminación. / High pollution. (2 / 7 / 1 / 3 / 2 / 0 / 3 / 1 / 2)
if Disponibilidad de sitios de alojamiento = BAJO and Información por personas en la calle = BUENO then Ciudad segura / Safe city. (0 / 1 / 2 / 0 / 1 / 6 / 1 / 0 / 2)
if Información por personas en la calle = BUENO and Motivación para regresar = Gente amable / Nice people. then Ambiente agradable / Very nice city. (0 / 0 / 0 / 3 / 0 / 0 / 0 / 0 / 0)
if Motivación para regresar = Todos los anteriores / All above and Oferta Cultural = EXCELENTE then Gente amable / Very nice people. (0 / 0 / 3 / 0 / 0 / 0 / 0 / 0 / 0)
if Calificación del gasto de la visita = Medio alto / Medium high and Calidad del alojamiento = BUENO then Ciudad insegura / Unsafe city. (0 / 0 / 0 / 0 / 0 / 1 / 0 / 4 / 0)
if Información por parte de la policía de tránsito = BUENO and Clima agradable / Nice weather then Gente poco amable / Rude people. (0 / 0 / 1 / 0 / 2 / 0 / 0 / 1 / 0)
if Disponibilidad de sitios de alojamiento = BAJO and Señalización Turística = BUENO then Clima agradable / Very nice weather. (1 / 0 / 0 / 0 / 0 / 0 / 4 / 0 / 1)
if Motivación para regresar = Cultura / Culture and Información por parte de la policía = EXCELENTE then Ambiente agradable / Very nice city. (1 / 0 / 1 / 3 / 0 / 0 / 0 / 0 / 0)
if Gasto estimado antes de visita = $750 - $900 Dollars and Calificación del gasto de la visita = Muy alto / Very high then Alta contaminación. / High pollution. (0 / 2 / 0 / 0 / 0 / 0 / 0 / 1 / 0)
if Información por parte de la policía = BUENO and Gasto estimado antes de visita = $100 - $300 Dollars then Ciudad segura / Safe city. (0 / 0 / 0 / 0 / 0 / 3 / 0 / 0 / 0)
if Información por parte de la policía = BAJO and Conservación de los Monumentos = EXCELENTE then Gente poco amable / Rude people. (0 / 1 / 1 / 0 / 3 / 0 / 0 / 0 / 0)
if Oferta de Ocio y Entrenamiento = BUENO and Conservación de los Monumentos = BUENO then Ciudad Ruidosa / Noisy city. (0 / 1 / 0 / 0 / 0 / 1 / 1 / 1 / 4)

```

Figura 68. Reglas del Clúster 1 sobre percepción.

Fuente: Resultado obtenido desde el software RapidMiner.

```

RuleModel
if Oferta de Ocio y Entretenimiento = BUENO and Informaci3n en la oficina de turismo = BAJO then Mucha indigencia / Many indigent people. (15 / 11 / 16 / 13 / 8 / 11 / 29 / 13 / 11)
if Cultura de la Ciudadan1a = BAJO and Informaci3n por parte de la policia de transito = BAJO then Gente amable / Very nice people (12 / 13 / 13 / 26 / 13 / 8 / 22 / 12 / 10)
if Calidad del alojamiento = EXCELENTE and Disponibilidad de sitios de alojamiento = EXCELENTE then Ciudad Ruidosa / Noisy city (26 / 10 / 9 / 6 / 14 / 15 / 17 / 16 / 16)
if Precio del alojamiento = EXCELENTE and Oferta Comercial = EXCELENTE then Ambiente agradable / Very nice city. (11 / 3 / 21 / 5 / 10 / 13 / 3 / 15 / 6)
if [Informaci3n por personas en la calle = EXCELENTE and Cultura de la Ciudadan1a = EXCELENTE then Ciudad segura / Safe city (15 / 12 / 7 / 12 / 23 / 9 / 7 / 5 / 6)
if Oferta Comercial = BAJO and Informaci3n por parte de la policia = BUENO then Gente amable / Very nice people (9 / 12 / 8 / 19 / 8 / 8 / 7 / 8 / 8)
if Precio del alojamiento = EXCELENTE and Calificaci3n del gasto de la visita = Medio alto / Medium high then Mucha indigencia / Many indigent people. (1 / 5 / 2 / 3 / 1 / 2 / 8 / 2 / 1)
if Informaci3n en el hotel = EXCELENTE and Oferta de Alojamiento = BUENO then Ambiente agradable / Very nice city. (7 / 11 / 15 / 6 / 7 / 7 / 6 / 6 / 4)
if Precio del alojamiento = BUENO and Gasto estimado antes de visita = $100 - $700 Dollars then Alta contaminaci3n. / High pollution. (0 / 3 / 2 / 3 / 3 / 1 / 2 / 7 / 0)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BUENO and Oferta Cultural = BUENO then Mucha indigencia / Many indigent people. (2 / 4 / 6 / 3 / 4 / 4 / 13 / 8 / 5)
if Oferta de Ocio y Entretenimiento = EXCELENTE and Informaci3n por parte de la policia = BAJO then Alta contaminaci3n. / High pollution (7 / 9 / 4 / 8 / 3 / 8 / 4 / 13 / 6)
if Motivaci3n para regresar = La comida / Food and Oferta de Ocio y Entretenimiento = BUENO then Ambiente agradable / Very nice city. (2 / 0 / 6 / 0 / 0 / 0 / 1 / 1 / 0)
if Gasto estimado antes de visita = $5.000 or more dollars and Oferta de Ocio y Entretenimiento = EXCELENTE then Ciudad segura / Safe city (1 / 1 / 0 / 0 / 6 / 0 / 1 / 0 / 1)
if Gasto estimado antes de visita = $900 - $1.100 Dollars and Calificaci3n del gasto de la visita = Alto / High then Ciudad insegura / Unsafe city (0 / 1 / 0 / 1 / 3 / 0 / 0 / 0 / 4)
if Gasto estimado antes de visita = $1 - $100 Dollars and Informaci3n en la oficina de turismo = BUENO then Gente poco amable / Rude people (0 / 2 / 4 / 1 / 0 / 9 / 0 / 0 / 1)
if Calificaci3n del gasto de la visita = Alto / High and Oferta de Ocio y Entretenimiento = EXCELENTE then Mucha indigencia / Many indigent people. (1 / 0 / 0 / 1 / 2 / 1 / 6 / 0 / 1)
if Disponibilidad de sitios de alojamiento = EXCELENTE and [Informaci3n por personas en la calle = BUENO then Gente amable / Very nice people (1 / 4 / 1 / 9 / 3 / 1 / 3 / 2 / 1)
if Gasto estimado antes de visita = $500 - $700 Dollars and Motivaci3n para regresar = Bajo costo de vida / Low cost of living then Ciudad Ruidosa / Noisy city (4 / 0 / 1 / 0 / 0 / 0 / 0 / 0 / 0)
if Gasto estimado antes de visita = $2.000 - $3.000 Dollars and Oferta de Alojamiento = BUENO then Ciudad insegura / Unsafe city (1 / 0 / 0 / 0 / 0 / 0 / 0 / 0 / 3)
if Calificaci3n del gasto de la visita = Bajo / Low and Conservaci3n de los Monumentos = EXCELENTE then Alta contaminaci3n. / High pollution (0 / 1 / 0 / 2 / 1 / 0 / 3 / 2 / 6 / 2)
if Calificaci3n del gasto de la visita = Medio / Medium and Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO then Ambiente agradable / Very nice city. (0 / 1 / 5 / 0 / 2 / 2 / 1 / 1 / 2)
if Informaci3n en la oficina de turismo = BUENO and Precio del alojamiento = BAJO then Ciudad Ruidosa / Noisy city (10 / 1 / 2 / 1 / 3 / 8 / 5 / 2 / 4)
if Oferta Comercial = BUENO and Conservaci3n de los Monumentos = BAJO then Ciudad segura / Safe city (5 / 4 / 5 / 3 / 10 / 0 / 5 / 4 / 3)
if Informaci3n por parte de la policia de transito = EXCELENTE and Informaci3n en el hotel = BAJO then Gente amable / Very nice people (2 / 2 / 3 / 8 / 2 / 1 / 2 / 2 / 2)
if Precio del alojamiento = BUENO and Calidad del alojamiento = EXCELENTE then Ciudad insegura / Unsafe city (2 / 0 / 3 / 1 / 1 / 1 / 0 / 2 / 7)
if Disponibilidad de sitios de alojamiento = BAJO and Gasto estimado antes de visita = $100 - $300 Dollars then Gente poco amable / Rude people (0 / 1 / 0 / 0 / 0 / 3 / 0 / 1 / 0)
if Calidad del alojamiento = BAJO and Gasto estimado antes de visita = $100 - $300 Dollars then Mucha indigencia / Many indigent people. (0 / 0 / 0 / 0 / 0 / 0 / 3 / 1 / 0)
if Oferta de Alojamiento = BUENO and Se1alizacion Turistica = BUENO then Ambiente agradable / Very nice weather (0 / 5 / 1 / 0 / 1 / 0 / 0 / 1 / 0)
if Informaci3n por parte de la policia de transito = BUENO and Calificaci3n del gasto de la visita = Alto / High then Ambiente agradable / Very nice city. (0 / 1 / 3 / 1 / 0 / 1 / 0 / 0 / 0)
if Oferta de Ocio y Entretenimiento = EXCELENTE and Informaci3n por parte de la policia = BUENO then Gente poco amable / Rude people (0 / 0 / 1 / 0 / 2 / 4 / 2 / 2 / 1)
if [Informaci3n por personas en la calle = BAJO and Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO then Ciudad Ruidosa / Noisy city (4 / 1 / 0 / 1 / 1 / 0 / 1 / 1 / 2)
if Gasto estimado antes de visita = $200 - $500 Dollars and Calificaci3n del gasto de la visita = Muy bajo / Very low then Gente amable / Very nice people (1 / 3 / 0 / 2 / 0 / 0 / 0 / 0 / 0)
if Informaci3n por parte de la policia de transito = BUENO and Oferta Comercial = BAJO then Ambiente agradable / Very nice city. (0 / 5 / 4 / 2 / 2 / 0 / 1 / 0 / 0)
if Oferta de Ocio y Entretenimiento = BUENO and Disponibilidad de sitios de alojamiento = BUENO then Alta contaminaci3n. / High pollution (1 / 1 / 1 / 1 / 2 / 1 / 1 / 6 / 3)
if Informaci3n por personas en la calle = BUENO and Motivaci3n para regresar = Gente amable / Nice people. then Gente poco amable / Rude people (0 / 0 / 0 / 0 / 0 / 2 / 0 / 0 / 0)
if Informaci3n en la oficina de turismo = EXCELENTE and Informaci3n por parte de la policia de transito = EXCELENTE then Ambiente agradable / Very nice weather (1 / 5 / 0 / 0 / 0 / 0 / 0 / 0 / 2)
if Oferta de Alojamiento = EXCELENTE and Disponibilidad de sitios de alojamiento = EXCELENTE then Mucha indigencia / Many indigent people. (1 / 0 / 0 / 1 / 0 / 0 / 3 / 0 / 0)
if Calidad del alojamiento = EXCELENTE and Oferta Comercial = BUENO then Gente amable / Very nice people (0 / 0 / 0 / 2 / 0 / 0 / 3 / 0 / 0)
if Oferta Comercial = EXCELENTE and Oferta Cultural = EXCELENTE then Gente poco amable / Rude people (0 / 0 / 1 / 0 / 0 / 3 / 1 / 1 / 2)

```

Figura 69. Reglas del Cl1ster 2 sobre percepci3n.

Fuente: Resultado obtenido desde el software RapidMiner.

```

RuleModel
if Oferta de Alojamiento = EXCELENTE and Calificaci3n del gasto de la visita = Muy bajo / Very low then Ciudad Ruidosa / Noisy city (4 / 9 / 6 / 10 / 3 / 15 / 7 / 11 / 7)
if [Informaci3n por personas en la calle = EXCELENTE and Informaci3n por parte de la policia de carreteras = BUENO then Ciudad insegura / Unsafe city (17 / 14 / 29 / 10 / 16 / 13 / 19 / 15 / 15)
if Oferta Comercial = BAJO and Oferta Cultural = EXCELENTE then Ciudad Ruidosa / Noisy city (8 / 8 / 19 / 8 / 13 / 27 / 19 / 11 / 10)
if Informaci3n en el hotel = BUENO and Cultura de la Ciudadan1a = EXCELENTE then Alta contaminaci3n. / High pollution (7 / 10 / 7 / 21 / 12 / 15 / 10 / 6 / 7)
if Oferta de Alojamiento = BAJO and Disponibilidad de sitios de alojamiento = BUENO then Clima agradable / Very nice weather (9 / 10 / 9 / 12 / 12 / 7 / 13 / 22 / 6)
if Conservaci3n de los Monumentos = BAJO and Motivaci3n para regresar = Vida nocturna / City nightlife then Gente poco amable / Rude people (9 / 4 / 1 / 2 / 3 / 2 / 1 / 3 / 0)
if Calificaci3n del gasto de la visita = Medio / Medium and Oferta Cultural = BUENO then Gente amable / Very nice people (4 / 4 / 3 / 5 / 3 / 4 / 2 / 3 / 11)
if Oferta de Alojamiento = BUENO and Se1alizacion Turistica = BUENO then Mucha indigencia / Many indigent people. (12 / 10 / 7 / 5 / 8 / 7 / 17 / 7 / 8)
if Precio del alojamiento = BUENO and Calificaci3n del gasto de la visita = Muy bajo / Very low then Ciudad insegura / Unsafe city (2 / 0 / 7 / 4 / 2 / 0 / 4 / 0 / 3)
if Informaci3n por parte de la policia de carreteras = EXCELENTE and Precio del alojamiento = EXCELENTE then Clima agradable / Very nice weather (2 / 5 / 8 / 3 / 5 / 12 / 6 / 15 / 7)
if [Informaci3n por personas en la calle = BAJO and Informaci3n por parte de la policia = BAJO then Gente poco amable / Rude people (15 / 8 / 4 / 4 / 8 / 4 / 3 / 5 / 4)
if Oferta de Ocio y Entretenimiento = BAJO and Informaci3n en la oficina de turismo = EXCELENTE then Alta contaminaci3n. / High pollution (6 / 3 / 4 / 12 / 6 / 2 / 5 / 6 / 5)
if Motivaci3n para regresar = Ninguno / None and Informaci3n por parte de la policia de transito = BUENO then Gente amable / Very nice people (0 / 2 / 0 / 1 / 0 / 0 / 1 / 0 / 1 / 6)
if Oferta Cultural = EXCELENTE and Informaci3n por parte de la policia de carreteras = BUENO then Ciudad insegura / Unsafe city (2 / 3 / 11 / 0 / 5 / 1 / 2 / 2 / 1)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO and Oferta de Ocio y Entretenimiento = EXCELENTE then Ciudad Ruidosa / Noisy city (7 / 3 / 5 / 2 / 2 / 13 / 1 / 6 / 3)
if Calificaci3n del gasto de la visita = Bajo / Low and Cultura de la Ciudadan1a = BAJO then Mucha indigencia / Many indigent people. (0 / 2 / 0 / 0 / 1 / 1 / 2 / 7 / 3 / 1)
if Informaci3n por parte de la policia de transito = EXCELENTE and Precio del alojamiento = EXCELENTE then Alta contaminaci3n. / High pollution (3 / 2 / 1 / 8 / 3 / 3 / 4 / 2)
if Oferta Comercial = EXCELENTE and Conservaci3n de los Monumentos = BUENO then Gente amable / Very nice people (3 / 3 / 4 / 4 / 4 / 3 / 3 / 6 / 9)
if Informaci3n por parte de la policia de transito = BAJO and Precio del alojamiento = BAJO then Mucha indigencia / Many indigent people. (5 / 4 / 5 / 1 / 1 / 5 / 11 / 1 / 1)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO and Oferta Cultural = BUENO then Alta contaminaci3n. / High pollution (2 / 0 / 0 / 5 / 2 / 3 / 1 / 0 / 0)
if Calificaci3n del gasto de la visita = Muy bajo / Very low and Disponibilidad de sitios de alojamiento = EXCELENTE then Gente amable / Very nice people (0 / 0 / 1 / 0 / 0 / 0 / 0 / 0 / 4)
if Informaci3n en el hotel = BAJO and Oferta de Ocio y Entretenimiento = EXCELENTE then Ambiente agradable / Very nice city. (2 / 5 / 0 / 1 / 2 / 0 / 1 / 1 / 1)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BUENO and Alojamiento = BUENO then Ciudad insegura / Unsafe city (0 / 1 / 4 / 2 / 1 / 0 / 2 / 0 / 0)
if [Informaci3n por personas en la calle = EXCELENTE and Conservaci3n de los Monumentos = EXCELENTE then Gente amable / Very nice people (0 / 1 / 1 / 1 / 0 / 1 / 0 / 5)
if Oferta de Ocio y Entretenimiento = BUENO and Motivaci3n para regresar = Todos los anteriores / All above then Clima agradable / Very nice weather (0 / 0 / 0 / 0 / 0 / 0 / 0 / 3 / 0)
if Disponibilidad de sitios de alojamiento = BUENO and Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = EXCELENTE then Ciudad segura / Safe city (0 / 2 / 0 / 1 / 4 / 0 / 0 / 1 / 0)
if Calidad del alojamiento = BUENO and Informaci3n por parte de la policia de transito = BUENO then Gente poco amable / Rude people (5 / 1 / 1 / 2 / 0 / 0 / 1 / 1 / 0)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO and Informaci3n por personas en la calle = BAJO then Ciudad Ruidosa / Noisy city (0 / 0 / 1 / 0 / 0 / 3 / 0 / 1 / 0)
if Conservaci3n de los Monumentos = BAJO and Disponibilidad de sitios de alojamiento = BUENO then Gente amable / Very nice people (0 / 0 / 0 / 0 / 1 / 0 / 0 / 0 / 3)
if Informaci3n turistica en vayas, carteles, volantes, folletos, guias, etc = BAJO and Conservaci3n de los Monumentos = EXCELENTE then Alta contaminaci3n. / High pollution (0 / 0 / 1 / 3 / 0 / 0 / 1 / 0 / 0)
if Se1alizacion Turistica = BAJO and Oferta Comercial = EXCELENTE then Ambiente agradable / Very nice city. (1 / 4 / 0 / 0 / 0 / 0 / 0 / 1 / 1)
if Informaci3n por parte de la policia de transito = BUENO and Oferta de Ocio y Entretenimiento = EXCELENTE then Ciudad Ruidosa / Noisy city (0 / 0 / 1 / 1 / 0 / 3 / 0 / 0 / 1)
if Informaci3n por parte de la policia de carreteras = EXCELENTE and Informaci3n por parte de la policia de transito = EXCELENTE then Ciudad segura / Safe city (1 / 0 / 2 / 0 / 5 / 1 / 2 / 0 / 1)
else Clima agradable / Very nice weather (4 / 3 / 4 / 4 / 2 / 3 / 3 / 6 / 3)
correct: 327 out of 1217 training examples.

```

Figura 70. Reglas del Cl1ster 3 sobre percepci3n.

Fuente: Resultado obtenido desde el software RapidMiner.

Tabla 39. Tabla uno de la percepción del gasto de los viajeros.

Clúster	Gasto estimado	Calificación del gasto	Aspecto que lo impacta	Motivo para regresar	Oferta de ocio y entretenimiento	Oferta comercial	Oferta cultural
0	5.000 o mas	Alto Medio Medio-alto	Indigencia, ciudad insegura, clima	Nunca, Todos los anteriores	Excelente- Bueno	Bajo, Excelente	Bajo
1	1.100-1.500	Alto(213) Medio(212)	Contaminación, Indigencia	Gente, comida, clima, vida nocturna, paisaje, Nunca	Bajo	Excelente, Bueno	Bueno, Bajo, E
2	300-500 1.500-2.000 1-100	Alto Medio-Alto	Indigencia, Ruido, Ambiente Agradable	Comida, Ninguno, Bajo costo de vida	Excelente	Bajo	Bajo
3	900-1.100 500-700	Bajo Medio Medio-Alto	Ciudad Ruidosa, Insegura,	Comida, Todos, Bajo	Bajo	Bueno, E, Bajo	E, Bajo

			Clima agradable	costo de vida,			
				Ninguno			
Outliers	1.100-1.500	Medio alto. Alto	Contaminación, gente poco amable, ruido, inseguridad	Ninguno	Bueno, Bajo	Excelente, Bajo	Bajo

Fuente: Elaboración propia.

Tabla 40. Tabla dos de la percepción del gasto de los viajeros

Clúster	Conservación de los monumentos	Señalización turística	Cultura ciudadana	Oferta de alojamiento	Disponibilidad de sitios	Calidad del alojamiento
0	Excelente	Bueno, Bajo	E, Bueno	Bueno, E	Bueno, E	E, Bajo
1	Excelente	Bueno, Bajo, E	E, Bueno, Bajo	Bueno	Bajo	E, Bajo, Bueno
2	Bajo	Bajo	Bajo, E, Bueno	Bajo, Bueno, E	Bueno	Excelente
3	Excelente	Excelente, Bajo	Bajo, Bueno, E	Bajo, Bueno, E	Bueno	Bueno
Outliers	Bueno	Bueno, Excelente	Excelente	Bueno, Bajo	Bueno	Excelente

Fuente: Elaboración propia.

Tabla 41. Tabla tres de la percepción del gasto de los viajeros

Clúster	Precio del alojamiento	Info Oficina Turismo	Info en el hotel.	Info en la calle	Info la policía	Info Transito	Info Policía Carreteras	Info en vallas
0	Bueno, E	E, Bueno	Bajo, E, B	Bajo, Bueno, E	Bueno, E	Bueno	E, Bajo	Bueno, E, Bajo.
1	Bueno, E	Excelente	Bajo	Bajo, Bueno	Bueno, E	Bueno	E, Bajo	Bueno, E, Bajo.
2	E	Bajo, E, Bueno	Bueno	Bajo	Excelente	Excelente	Bueno	Excelente, Bueno
3	Bueno	Bueno, Bajo	Bueno	Bueno, Excelente	Bajo, Excelente	Bajo, Excelente	Bueno	Excelente
Outliers	Bajo	Bajo	Excelente	Bueno	Bueno	Bueno	Excelente	Excelente

Fuente: Elaboración propia.

Caracterización del gasto estimado y percepción del clúster cero: El clúster inicialmente tenía un gasto estimado de 5.000 mil o más dólares, después del viaje se considera que el gasto realizado es alto, en totales sigue medio y medio alto.

- Las ofertas que consideran más altas son: Ocio y entretenimiento.

- Las ofertas que consideran más bajas son: Comercial, cultural.
- Las ofertas que consideran intermedias, es decir buenas: alojamiento.

El valor estimado a comparación de lo considerado al final del viaje es elevado, el costo de vida de la ciudad ha subido y se considera alto respecto las demás ciudades del país, pero no tan elevado como otros países, se puede concluir que hubo exceso de gastos más que gastos en el viaje en si (comida, alojamiento, transporte), las ofertas que se consideran más altas son las de ocio, es posible que los viajeros hayan gastado bastante dinero en actividades como: casinos, apuestas, discotecas, bares, etc. (DANE, 2015).

Las ofertas que se consideran bajas son las comerciales, en la ciudad hay bastantes almacenes de cadena y la ciudad se caracteriza por el comercio (República, 2014), podríamos considerar que los viajeros tienen gustos diferentes, recordemos que el clúster cero está conformado principalmente por pensionados, por lo tanto los gustos comerciales pueden variar bastante respecto a lo que la ciudad ofrece.

El alojamiento está en un puesto intermedio, en la ciudad casi todos los hoteles se clasifican como 3 estrellas (TripAdvisor, 2015).

La principal característica de la ciudad respecto al impacto es: indigencia.

- Calificaciones positivas: Buen clima.
- Calificaciones negativas: Indigencia, ciudad insegura.

Analizando las reglas y el consolidado, los viajeros tienen una percepción de indigencia e inseguridad ya que la conservación de los monumentos es muy baja, la oferta cultural es muy baja y la información en vallas también es muy baja y se puede apreciar cuando visitamos un

parque, hay poca vigilancia y muchas personas indigentes consumiendo sustancias psicoactivas, pocos letrados, vallas, pancartas y personal de turismo.

Calificación respecto a la información en la ciudad tenemos:

- Calificaciones positivas (Excelentes): Ninguna.
- Calificaciones negativas (Bajas): información en el hotel y en la calle, policía de carreteras y en vallas.
- Calificaciones intermedias (Buenas): alojamiento, información en oficina de turismo, hotel, calle, tránsito, vallas.

Tenemos datos con bastantes viajeros en calificación baja e intermedia lo cual da una clara visión de inconformidad en dichos aspectos, el alojamiento es regular pero como vimos anteriormente, la ciudad se caracteriza por hoteles de 3 estrellas, la información en las oficinas de turismo no es muy buena ya que dichas oficinas funcionan únicamente en semana santa y son pequeñas, las personas encargadas solo entregan un volante informativo y no hablan varios idiomas.

Calificación principal de motivos para regresar a la ciudad:

- Calificaciones positivas: Todas las anteriores (clima, gente, paisaje, etc)
- Calificaciones negativas: Todas las negativas.
- Calificaciones intermedias: Ninguna.

Las calificaciones positivas y negativas están casi al mismo nivel, siendo superiores las negativas, en la estadística se observa que el clima es de los que más puntaje obtuvo.

Pereira es un eje por el cual circulan personas de muchas ciudades, tal vez por esto sea

difícil la cultura ciudadana.

Caracterización del gasto estimado y percepción del clúster uno: El clúster inicialmente tenía un gasto estimado de 1.100 a 1.500 dólares, después del viaje se considera que el gasto realizado es alto y medio.

- Las ofertas que consideran más altas son: Comercial.
- Las ofertas que consideran más bajas son: Ocio y entretenimiento.
- Las ofertas que consideran intermedias, es decir buenas: alojamiento, cultural.

El valor estimado a comparación de lo considerado al final del viaje es elevado e intermedio, el costo de vida de la ciudad se ha incrementado y se considera alto respecto las demás ciudades del país, recordemos que el clúster uno está conformado por familias que viajan en transporte público, puede ser una razón por la cual se califica como alta e intermedia, alta debido al costo de vida y cantidad de gastos que conlleva viajar con una familia numerosa y ahorrando dinero en transporte público más económico que transporte privado (DANE, 2015).

Las ofertas que se consideran bajas son las de ocio y entretenimiento, el viajar con una familia numerosa con hijos de diversas edades hace complicado satisfacer todas las necesidades, por ejemplo si un niño de 6 años desea ir al circo y un adolescente desea jugar video juegos, es difícil encontrar las dos actividades en un solo sitio, esto dificulta el transporte y la visita a dichos lugares.

La principal característica de la ciudad respecto al impacto es: Contaminación.

- Calificaciones positivas: gente, comida, vida nocturna, paisaje.
- Calificaciones negativas: todas.

Analizando las reglas y el consolidado, lo viajeros tienen una percepción de indigencia, inseguridad, gente poco amable. Hay actividades para adultos, tal vez para niños existan algunas pero no totalmente gratuitas como en otras ciudades.

Calificación respecto a la información en la ciudad tenemos:

- Calificaciones positivas (Excelentes): Oficina de Turismo.
- Calificaciones negativas (Bajas): información en el hotel y en la calle, policía de carreteras y en vallas.
- Calificaciones intermedias (Buenas): alojamiento, información en oficina de turismo, hotel, calle, tránsito, vallas.

Así como en el clúster cero, coinciden varios aspectos sobre la información, tenemos datos con bastantes viajeros en calificación baja e intermedia lo cual da una clara visión de inconformidad en dichos aspectos, el alojamiento es regular pero como vimos anteriormente, la ciudad se caracteriza por hoteles de 3 estrellas, la información en las oficinas de turismo no es muy buena ya que dichas oficinas funcionan únicamente en semana santa y son pequeñas, las personas encargadas solo entregan un papel informativo y no hablan varios idiomas.

Calificación principal de motivos para regresar a la ciudad:

- Calificaciones positivas: Todas las anteriores (clima, gente, paisaje, etc)
- Calificaciones negativas: Todas las negativas.
- Calificaciones intermedias: Ninguna.

Similar al clúster cero, las calificaciones positivas y negativas están casi al mismo nivel, siendo superiores las negativas, en la estadística se observa que el clima es de los que más

puntaje obtuvo.

Caracterización del gasto estimado y percepción del clúster dos: El clúster inicialmente tenía un gasto estimado de 300-500, 1.500-2.000, y de 1-1.000 dólares, después del viaje se considera que el gasto realizado es alto y medio alto.

- Las ofertas que consideran más altas son: Ocio y entretenimiento.
- Las ofertas que consideran más bajas son: Comercial, cultural, alojamiento.
- Las ofertas que consideran intermedias, es decir buenas: Alojamiento.

El valor estimado a comparación de lo considerado al final del viaje es elevado, hay que considerar que el clúster antes del viaje tenía diversos valores como gasto estimado, pero seguramente en el transcurso del viaje requirieron más dinero dependiendo de las actividades realizadas en la ciudad.

Las ofertas más bajas son las comerciales, es posible que lo que se consigue en la ciudad no es de gusto de los viajeros, lo mismo sucede con lo cultural o hay poca variedad de eventos culturales.

El alojamiento se encuentra entre medio y bajo puntaje, ya que en la ciudad casi todos los hospedajes son de 3 estrellas, es posible que los edificios se encuentren deteriorados, como poco diseño, cuartos poco iluminados o ventilados, televisores viejos, camas desajustadas, etc.

La principal característica de la ciudad respecto al impacto es: Contaminación.

- Calificaciones positivas: gente, comida, vida nocturna, paisaje.
- Calificaciones negativas: todas.

Analizando las reglas y el consolidado, los viajeros tienen una percepción de indigencia, inseguridad, gente poco amable.

La ciudad posee sitios para mayores de edad como bares, discotecas, casinos, etc. Ya que el clúster tres de viajeros asumíamos que estaba conformado por familias numerosas con posibilidades económicas, la buena y mala imagen está dividida en dicho clúster, es posible que al visitar el centro de la ciudad y ver la cantidad de vendedores ambulantes o al ver indigentes consumiendo sustancias en el centro de la ciudad o al intentar comunicarse con alguien y encontrar que pocas personas hablan diferentes idiomas.

Calificación respecto a la información en la ciudad tenemos:

- Calificaciones positivas (Excelentes): Oficina de Turismo.
- Calificaciones negativas (Bajas): información en el hotel y en la calle, policía de carreteras y en vallas.
- Calificaciones intermedias (Buenas): alojamiento, información en oficina de turismo, hotel, calle, tránsito, vallas.

A calificación del hotel aparece en intermedia y negativa, da motivos para pensar en el servicio al cliente que están ofreciendo o también dependería del tipo de hotel en el que están alojados los viajeros, en caso de que se alojen con sus familiares, es posible que hayan decidido buscar hoteles para poder pagar la comida y no utilizar restaurantes en la calle o comidas rápidas y el hotel no disponga de dicho servicio.

Calificación principal de motivos para regresar a la ciudad:

- Calificaciones positivas: Todas las anteriores (clima, gente, paisaje, etc)

- Calificaciones negativas: Todas las negativas.
- Calificaciones intermedias: Ninguna.

Similar al clúster cero, las calificaciones positivas y negativas están casi al mismo nivel, siendo superiores las negativas, en la estadística se observa que el clima es de los que más puntaje obtuvo.

Caracterización del gasto estimado y percepción del clúster tres: El clúster inicialmente tenía un gasto estimado de 900-1.100, 500-700 dólares, después del viaje se considera que el gasto realizado es bajo y medio.

- Las ofertas que consideran más altas son: Cultural.
- Las ofertas que consideran más bajas son: Ocio y entretenimiento, alojamiento.
- Las ofertas que consideran intermedias, es decir buenas: Comercial.

La oferta cultural se encuentra con la más alta calificación, los viajeros de dicho clúster se sienten a gusto con la sobras teatrales que se presentan en la ciudad, con las exposiciones de arte, exposiciones de pinturas.

Las ofertas más bajas son las de ocio y entretenimiento, es posible que para dicho viajeros el acceder a bares, discotecas, casinos no sea de su gusto, algún sitio que les agrade podría ser una cafetería biblioteca como las que se pueden aprecia en otras ciudades y países (Prisa., 2013).

La principal característica de la ciudad respecto al impacto es: Ruido.

- Calificaciones positivas: clima.
- Calificaciones negativas: inseguridad.

Analizando las reglas y el consolidado, los viajeros tienen una percepción de indigencia, inseguridad, gente poco amable.

Este clúster está conformado por personas que viajan regularmente solas, se hospedan en casas, posiblemente en casas alojadas en sitios muy transitados de la ciudad como por ejemplo el centro de la ciudad, la zona rosa en donde hay muchas viviendas pero hay problemas de ruido, etc. (otun, 2015).

Calificación respecto a la información en la ciudad tenemos:

- Calificaciones positivas (Excelentes): Información en vallas.
- Calificaciones negativas (Bajas): policía, tránsito.
- Calificaciones intermedias (Buenas): oficina turismo, hotel, calle, policía de carreteras.

La mejor información se encuentra en las vallas, esto es correcto únicamente para las personas que viven en la zona rosa de la ciudad o en el centro de Pereira, a medida que nos alejamos del centro vemos poca señalización y prácticamente es muy fácil no ubicarse adecuadamente.

Calificación principal de motivos para regresar a la ciudad: Comida.

- Calificaciones positivas: Todas las anteriores (clima, gente, paisaje, etc)
- Calificaciones negativas: Todas las negativas.
- Calificaciones intermedias: Ninguna.

Similar al clúster cero, las calificaciones positivas y negativas están casi al mismo nivel, siendo superiores las negativas, en la estadística se observa que el clima es de los que más puntaje obtuvo.

Evaluación

Evaluación de los resultados. Los resultados obtenidos en el proceso de minería de datos nos brindan una clasificación del grupo de viajeros de acuerdo con validaciones internas del algoritmo de clustering k-Means.

Las pruebas realizadas y combinaciones de variables se pueden apreciar en el **¡Error! No se encuentra el origen de la referencia.**, a medida que se realizaban las pruebas se realizaban las validaciones internas, la validación externa de los clústers no se tuvo en cuenta ya que esta requiere conocer personalmente los viajeros y conocer muy bien su clasificación lo cual era imposible dada la forma de recolección de los datos.

Las diferentes pruebas realizadas a los clústers mediante los procesos y los resultados del índice Silhouette el cual matemáticamente es el índice de validación recomendado para los clústers nos dan cuenta de cuatro clústers generados y recomendados.

Como se pudo apreciar, los cuatro clústers tienen características que los hacen únicos y con una tendencia muy marcada.

Cada clúster nos brinda información general y específica de perfiles de viajeros, los análisis realizados en la etapa anterior están enfocados a los resultados de los algoritmos de Clustering y argumentando con las respectivas referencias las posibles causas de dicho comportamiento de los viajeros.

Los análisis realizados se dan ya que el analista conoce el objeto del negocio el cual es “La Ciudad de Pereira”, conoce sus calles, su cultura, las dificultades diarias, las zonas de peligro, algunos de los comentarios que realizan las personas de otras regiones y de otros países.

Lo anterior no indica que este análisis es el único, por el contrario, brinda un marco de

referencia de la ciudad para que los lectores realicen su propio análisis a la luz de los datos.

Los resultados obtenidos son satisfactorios, se obtuvo 5 grupos en total incluyendo los Outliers, se caracterizó a la población, de acuerdo a los clústers se concluye lo siguiente:

- Los resultados son claros y dan una descripción de los viajeros.
- Se han descubierto 4 perfiles de viajeros sobre los cuales mejorar todos los procesos de la ciudad y los procesos de los entes encargados del turismo y promoción de la ciudad.
- Dentro de dichos perfiles llama la atención de la gran cantidad de viajeros con diversos niveles de preparación académica.
- Otro aspecto que llama la atención es que comparando los problemas de la ciudad y los aspectos negativos de la visita por parte de los viajeros, estos han perdurado en el tiempo ya que los encuestados llegan hasta el año 2014 y hasta la fecha 2015 existen las mismas dificultades que hace más de 6 años atrás.
- También llama la atención que a pesar de las dificultades, el eje cafetero y la ciudad de Pereira siguen siendo unos de los destinos preferidos en la región (DANE, 2015).
- El aspecto que se puede rescatar es la disminución de la indigencia desde el desalojo de las personas de la calle de la zona de la calle 14 en donde se encuentra en el momento la plazoleta ciudad victoria, el centro comercial victoria y las tiendas del almacén éxito.

Respecto a los datos, método de encuesta y las preguntas de la encuesta:

- Las preguntas seleccionadas para la encuesta fueron estudiadas cuidadosamente para dar una visión general de los viajeros.
- Como en toda encuesta, no podemos saber a ciencia cierta si los datos son reales o no.

- Algunas preguntas se pueden modificar de acuerdo a las características de la ciudad o la problemática en el momento de la recolección de los datos.
- Las preguntas fueron pertinentes para la ciudad debido a su idiosincrasia, problemas, vías de transporte, idioma, ubicación respecto a otras regiones y paisaje cafetero.

En términos comerciales los resultados son pertinentes porque:

- Da una perspectiva desde los viajeros a la ciudad.
- Nos muestra una clara tendencia de las problemáticas vividas diariamente las cuales se convierten el algo cotidiano para los que vivimos en la ciudad pero las cuales afectan el turismo.

Proceso de revisión:

Como parte de una autoevaluación se tienen en cuenta los siguientes aspectos:

- El proceso de minería de datos utilizado resultó efectivo para analizar los datos nominales y numéricos.
- El proceso puede ser mejorado en su herramienta RapidMiner integrando Loops y ciclos para realizar el análisis en un solo momento, es decir, el proceso ETL y minería en una sola herramienta.
- La recolección de los datos y la limpieza se debería mejorar con el apoyo de los entes de control encargados del turismo en la región.
- El proceso realizado en la presente investigación se puede aplicar a los datos socioeconómicos o datos descriptivos de características en otras regiones o ciudades ya que es genérico para datos nominales, discretizados y normalizados.
- Se deberían aplicar las encuestas a grupos familiares y crear un nuevo campo que los

identifique, de esta manera evitar segmentar la muestra, esto ya que muchas veces la persona que paga disfruta pero el viaje tiene una percepción diferente del aspecto económico con respecto a la persona que simplemente viaja, disfruta pero no gasta su propio dinero.

Despliegue

El despliegue de la aplicación puede ser realizado por cualquier persona interesada en analizar los datos de la ciudad o implementar dichos análisis a sus propias ciudades.

Debería ser parte del plan de cada alcaldía la cual aportaría bastante al mejoramiento de la ciudad y también en la innovación de productos y/o servicios para atraer a los turistas y mantener a los actuales que nos visitan.

El proyecto plantea una arquitectura inicial de análisis de datos para los viajeros, esta queda abierta para modificaciones, ejecuciones y análisis, ya que la gran cantidad de información recopilada es un insumo para cualquier analista de datos que quiera tratar la información y de esta manera poder contribuir al mejoramiento de la región.

Debe ser un propósito para el lector interesado, el continuar haciendo pruebas, análisis, conclusiones y argumentaciones.

La presente investigación da una referencia de cómo abordar un problema de minería de datos, para etapas posteriores se debería aplicar otros algoritmos de minería de datos los cuales permitan predecir resultados de los viajeros o predecir si un visitante regresará o no.

Análisis de Resultados

- En la ciudad tenemos muchos visitantes de un solo día, los cuales no pernoctan pero hacen diversas actividades, de acuerdo con los análisis realizados a los datos, se debe mejorar en la oferta cultural y así posiblemente el viajero se quedaría en la ciudad, dicha oferta debe estar a

cargo de las personas encargadas de turismo o de los comerciantes de la región.

- La ciudad de Pereira recibe diversidad de visitantes, hombres, mujeres, adultos, jóvenes y niños, para cada grupo de personas hay algunos intereses en común como se pudo apreciar en los clústers, como por ejemplo, el clúster cero (Análisis sobre las reglas de asociación del clúster cero:) que está conformado por personas de la tercera edad utilizan la bicicleta y caminan para movilizarse en su mayoría por la ciudad, adicional a esto el clúster (Análisis sobre *las reglas de asociación del clúster uno:*) en cuanto a los sitios de alimentación se puede observar que utilizan restaurantes, comida rápida y comida rápida en restaurantes, estos intereses son comunes para ellos y coinciden con los viajeros del segundo clúster, los comerciantes o la entidades de la ciudad podrían aprovechar estos segmentos y desarrollar actividades de recreación y/o culturales.
- Llama la atención que a la ciudad llegan personas de diversos niveles educativos con intereses en común, la región podría aprovechar dichos conocimientos al utilizar el centro de convenciones o al crear una base de datos de los viajeros, determinar sus intereses, conocimientos y posibilidades de que realicen conferencias, charlas o eventos y de esta manera aportar al desarrollo de la región
- Ya que nos visitan tantas personas independientes, se podrían generar mejoras en la ciudad para que dichas personas puedan hacer parte de los proyectos de la alcaldía, como por ejemplo, generar un portal para que cualquier persona pueda participar de licitaciones de proyectos privados de la región.
- Según el análisis realizado, de todos los Clústers, en el que había jóvenes involucrados, el porcentaje de ingresos familiares era del 80% en la población del clúster, según el análisis dichos

jóvenes visitan la ciudad para compras, esto indica un poder adquisitivo por parte de sus padres, ya que la mayoría tienen como ingreso el dinero familiar.

- En la ciudad, de acuerdo con los análisis de datos realizados, un buen porcentaje de visitantes de la ciudad concuerdan en que hay mucha indigencia, esta imagen se puede dar muy fácilmente ya que en muchas regiones y países del mundo no se ven las personas sin hogar y/o indigentes y por lo tanto la cifra que maneja Pereira puede ser muy escandalosa.
- El nivel de gasto estimado es bajo para las personas que visitan un solo día y vienen a hacer deporte, para los almacenes deportivos es una oportunidad de negocio para hacer ofertas, promociones o eventos programados con anterioridad.
- Al analizar los datos se apreció que el nivel de ingresos y el nivel educativo es mandatorio al momento de definir si el costo de algún bien o servicio en la ciudad es alto o bajo, como es el caso del Clúster 2 (**¡Error! No se encuentra el origen de la referencia.**) en el cual el nivel de gastos inicialmente era muy bajo y luego del viaje se clasifico como Alto, además de que dicho clúster está formado en su mayoría por personas desempleadas con formación como Técnicos.
- En la región nos visitan muchas personas de edad pensionadas, la ciudad debe brindar seguridad y educar a los conductores del servicio público para mejorar las relaciones humanas, de esta manera las personas se sentirían más seguras y utilizarían dicho transporte, ya que los datos que arroja el análisis indica que ellos utilizan un servicio privado.
- El uso de la bicicleta se incrementó en la ciudad, las entidades encargadas de la seguridad deben tener en cuenta que este es un medio de transporte que no contamina y ayuda a mejorar la

movilidad, según los datos arrojados, muchas personas que vienen en transporte privado a la ciudad utilizan la bicicleta, son hombres y laboran, teniendo un poder adquisitivo razonable.

- Los 4 clústers coinciden en datos que indican a la ciudad como ruidosa, insegura y con mucha indigencia, la alcaldía y las fundaciones encargadas de las personas de la calle o de rehabilitar a las personas con problemas de alcohol y/o drogas deben trabajar de la mano de la policía y de esta manera tratar de que estas personas se encuentren en un proceso de rehabilitación y no deambulando por la ciudad.
- En el caso del ruido en la ciudad, hay zonas identificadas con esta problemática, la política debe ser más estricta o dar una solución de forma tecnológica al ruido.

Aun así parece que muchas personas regresan a la región y dan buenas referencias, parece ser que la amabilidad de la gran mayoría de personas y el clima tienen mayor peso que las cosas malas.

Conclusiones

Los modelos de minería de datos aplicados a los datos propios de proceso y actividades que son intrínsecos a las ciudades, sirven para todos los visitantes y personas de la región, el trabajo realizado es tan importante para la ciudad de Pereira como para el área metropolitana y las ciudades que hacen parte del triángulo del café, ya que de esta manera, los lectores visualizarán como los datos nos generan información relevante que podemos utilizar para mejorar varios aspectos que no veíamos a simple vista.

Las entidades encargadas del turismo en la región tienen toda la infraestructura para enfocarse realmente en el turista que es el principal actor del negocio: viaje-región-agencia de viaje- turismo. Un turista contento recomienda la región, un turista que no fue bien recibido es un turista que no ve potencial en la región y lo más razonable es que no la recomiende.

Los análisis realizados en el presente trabajo no son los únicos posibles, los datos recopilados todavía tienen gran cantidad de información por deducir, por extraer y por experimentar; La presente investigación es solo una parte del potencial de la minería de datos enfocada a los viajeros.

La minería de datos tiene muchas herramientas para realizar diferentes tipos de análisis, por ejemplo: se podría utilizar minería web en los portales públicos de las entidades de la región para hacer análisis de texto e identificar quejas y/o aspectos positivos que publican los usuarios en los blogs.

Haciendo un contraste con la realidad de la región y con los datos recopilados, podemos

concluir varias cosas:

- El número de personas en las calles ha reducido, pero para un viajero siempre resultara molesto.
- Según el DANE, en la actualidad la ciudad ya no es la numero uno en desempleo, pero en los datos locales se puede ver un alto número de desempleados, es posible que la región prepare a los ciudadanos en idiomas y de esta manera se generaría empleo y mejoraría la imagen de la región (DANE, 2015).
- Un alto número de viajeros no pernoctan en la ciudad, en ese caso se puede plantear la hipótesis de la cultura popular, que Pereira es una ciudad de paso, siendo así y por la ubicación geográfica, esta región debe promover la gestión por y para los ciudadanos, por y para el turista.
- La percepción del gasto es marcada, para algunos encuestados es alta, para otros es baja, pero según el análisis realizado, todo va de acuerdo con el número de días de visita y la cantidad de dinero gastada en la región, seguramente en los almacenes de cadena ya que hay un alto consumo en las diversas opciones planteadas en la encuesta como: actividades culturales, turismo, ocio, aventura, educación, según los análisis de la minería de datos.
- Sería interesante estudiar que actividades de docencia y educación realizaron los viajeros de la ciudad, es posible que sean actividades en áreas del deporte, salud, arte, esto ya que en la ciudad tenemos eventos de pintura, arte, eventos deportivos de ciclismo, patinaje, halterofilia. Esto se plantea para que las entidades interesadas realicen un estudio sobre dichos aspectos.

Recomendaciones

Las recomendaciones para los lectores son varias:

- Tener en cuenta que la minería de datos es un proceso que involucra conocimiento del tema por parte del analista, de esta manera las variables nos darán mucha o poca información.
- Una recomendación para las entidades públicas y privadas es que le brinden la importancia necesaria a los datos, a las encuestas y a los visitantes, ya que como se pudo observar en el presente trabajo, los datos recopilados de manera rigurosa brindan información valiosa para ser analizada y posteriormente sacar conclusiones, recomendaciones y cartas de navegación en los sectores interesados.

A los funcionarios de Fomento al Turismo para que le presten atención y publiciten más la ciudad y sus puntos turísticos y poder recopilar datos del turismo para su análisis y así poder mejorar sus gestiones.

- A los analistas de datos para que tomen este ejemplo de minería de datos por medio de técnicas no supervisadas y lo repliquen a sus necesidades, estudien el algoritmo, estudien los datos, conozcan muy bien el objeto de negocio y validen las conclusiones con expertos en los temas.
- Para los analistas de datos y programadores, respecto a la herramienta es la de profundizar en la extensión de la herramienta, es decir, poder programar la herramienta y aprovechar las potencialidades y algoritmos matemáticos de minería de datos.

Recomendaciones para continuar la línea de investigaciones de minería de datos aplicada al turismo y los viajeros, entre algunas tenemos:

- Posteriormente al clustering se puede generar una segmentación con los mismos datos, de esta manera se van creando subgrupos de viajeros y a cada subgrupo aplicarle nuevamente clustering o segmentación, en ese caso el analista debe ir dividiendo los grupos de acuerdo con los requerimientos.
- Encuestar y segmentar los viajeros únicamente por edades, así realizar clustering únicamente por edades, este tipo de tarea se denomina clustering con segmentación.
- Crear más intereses en la lista de actividades planeadas, actividades realizadas y crear dos nuevas categorías: una que se denomine actividades de interés y actividades que le gustaría realizar que no existen en la ciudad.
- Con las modificaciones anteriores se puede hacer un nuevo estudio de reglas de asociación para determinar si una variable influye a otra y validar.
- Otro estudio que se puede realizar es el de un análisis de regresión el cual utiliza un DataSet de entrenamiento como puede ser el DataSet de la presente investigación y generar modelos para diferentes preguntas, entre algunas podríamos predecir las características de un viajero y determinar si dicho viajero regresaría a la ciudad o no.
- Luego de predecir si un viajero regresaría a la ciudad o no, hacer un seguimiento al modelo para determinar si dicho viajero regreso o no, el motivo y margen de error.
- Para trabajos posteriores, se puede plantear una herramienta automatizada dentro de los portales de alcaldías, gobernaciones, páginas de periódicos de la región, perfiles de Facebook y

cuentas de twitter de entidades para que un personal especializado analice los comentarios quejas y reclamos de las personas que opinan por dichos medios.

- En caso de que las gestiones de las entidades encargadas del turismo no permitan la herramienta planteada anteriormente, los analistas de datos pueden aprovechar las potencialidades de la minería web, así, generar una herramienta para filtrar los comentarios y observaciones sobre la ciudad, de diversas fuentes y determinar los comentario positivos, negativos, neutrales y posteriormente hacer Clustering o clasificación para obtener el conocimiento necesario con el propósito de mejorar y/o crear nuevos productos y servicios en la región.

Bibliografía

- A. K. Jain, M. M. (1999). Data clustering: a review. En M. M. A. K. Jain, *ACM Computing Surveys* (págs. 264-323).
- A., M. E. (1 de 12 de 2000). *Revista de Investigación en Psicología*. Obtenido de <http://sisbib.unmsm.edu.pe>:
http://sisbib.unmsm.edu.pe/bvrevistas/investigacion_psicologia/v03_n2/pdf/a08v3n2.pdf
- ABS. (2002). Measuring a Knowledge Based Economy and Society. *Australian Bureau of Statistics*.
- Acharya., S. M. (2003). *Data Mining: Multimedia, Soft Computing and Bioinformatics*. Wiley Inter-Science.
- Agrawal, R. e. (1993). *Mining Association Rules between Sets of Items in Large Database*. Recuperado el Octubre 15 de 2014, de <http://rakesh.agrawal-family.com>:
<http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>.
- Aguilar, L. J. (2013). Big Data. En L. J. Aguilar, *Análisis de grandes volúmenes de datos en organizaciones* (págs. 153-210). Mexico: Alfaomega.
- Ahmed, P. K. (2002). *Learning Through Knowledge Management*. Oxford: Butterworth.
- Airbnb. (s.f.). *Aairbnb*. Obtenido de <https://www.airbnb.es/>: <https://www.airbnb.es/>
- Ayala, H. (2000). Surprising Partners: Hotel Firms and Scientists Working Together to. *Cornell Hotel and Restaurant Administration Quarterly*, 42-57.

- Bahra, N. (2001). *Competitive Knowledge Management*. Basingstoke: Palgrave.
- Bernal, F. (2012). *Departamento de Ciencias de la Computacion e I.A.* Obtenido de Departamento de Ciencias de la Computacion e I.A.:
<http://elvex.ugr.es/decsai/intelligent/slides/dm/D2%20Association.pdf>
- Bornhorst, T. R. (2010). Determinants for DMO & destination success: an empirical examination. *Tourism Management*, 572-589.
- Bouncken, R. (2002). Knowledge Management for Quality Improvement in Hotels. *The Haworth Hospitality Press*, págs. 25-29.
- C, P., & Santín, D. (2007). *Mineria de Datos. Técnicas y Herramientas*. Madrid: Paraninfo.
- CAMPBELL, M. (2009). *base IV Guía de Autoenseñanza*. España. : McGraw Hill – Interamericana.
- Carson, D. a. (2004). Strategic Knowledge Management in Tourism. *paper presented to the CAUTHE conference*.
- Catolica, U. (1 de MAyo de 2014). www.ucp.edu.co, 1. Recuperado el 10 de Septiembre de 2014, de www.ucp.edu.co: http://www.ucp.edu.co/ucp_docs/ppr/documentos-generales/PLAN_DESARROLLO-pereira.pdf
- Chapman, P. e. (2007). *CRISP-DM 1.0 - Step-by-step data mining guide*. Recuperado el 15 de Octubre de 2014, de <http://www.crisp-dm.org/CRISPWP-0800.pdf>: <http://www.crisp-dm.org/CRISPWP-0800.pdf>
- Cho, V. &. (2002). Knowledge discovery techniques in database marketing for the tourism industry. *Journal* , 109-131.
- Cho, V. &. (2002). Knowledge discovery techniques in database marketing for the tourism

- industry. *Journal of Quality Assurance in Hospitality & Tourism*, 109-131.
- Choi, T. a. (2000). Towards a Knowledge Discovery Framework for Yield Management in the. *Hospitality Management*, 17-31.
- Chu, F. L. (2004). Forecasting tourism demand: a cubic polynomial approach. *Tourism Management*, 209-218.
- Comercio, M. d. (1 de 1 de 2007). *Ministerio de comercio*, 2008. Recuperado el 1 de Septiembre de 2014, de Ministerio de comercio:
<http://www.mincomercio.gov.co/econtent/documentos/turismo/2008/PlanSectoria2007-2010->
- ComputerWorld. (10 de 04 de 2014). *ComputerWorld*. Obtenido de ComputerWorld:
<http://www.computerworld.es/tendencias/el-universo-digital-se-expande-acelerado-por-el-crecimiento-de-los-datos>
- Cooper, & Ruhanen. (2003). Developing a Knowledge Management Approach to Tourism Research.
- Cooper, C. (5 de Abril de 2005). *www.elsevier.com/locate/atoures*. Recuperado el 24 de Septiembre de 2014, de *www.elsevier.com/locate/atoures*:
<http://www.sciencedirect.com/science/article/pii/S0160738305000915>
- Cooper, C. (4 de Enero de 2006). *Knowledge management and tourism*. Recuperado el 25 de Septiembre de 2014, de <http://www.sciencedirect.com>:
<http://www.sciencedirect.com/science/article/pii/S0160738305000915>
- Council, W. T. (1 de Noviembre de 2012). *wttc.org*, 1. Recuperado el 1 de Septiembre de 2014, de *wttc.org*: <http://www.wttc.org/~ /media/files/reports/benchmark>

Crompton, J. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, 408-424.

Crompton, J. y. (1997). Motives attending festival events. *Annals of Tourism Research*, 425-439.

DANE. (1 de 08 de 2015). <http://www.dane.gov.co>. Obtenido de <http://www.dane.gov.co>:

http://www.dane.gov.co/files/investigaciones/boletines/ipc/cp_ipc_feb15.pdf

Daniel, C. K. (4 Mar 2009). *Sistemas de información para la toma de decisiones*. México. :

McGraw-Hill.

Davidson, C. a. (2002). *Knowledge Management*. Auckland: Tandem.

Dean, J. (2014). *Big data, Data mining and Machine Learning - Value Creation for Business*

Leaders ans Practitioners. Wiley & SAS Business Series.

Devesa Fernández María, L. G. (1997). Motivación, satisfacción y lealtad en el turismo: el caso

de un destino de interior. *Revista Electronica de Motivación y Emoción*.

Devesa, M. y. (2006). Determinantes de la satisfacción percibida en el turismo rural. En M. y.

Devesa, *Determinantes de la satisfacción percibida en el turismo rural*. Valencia:

Tirant Lo Blanch.

Dolnicar, S. &. (2008). Selective marketing for environmentally sustainable tourism. *Tourism*

Management, 672-680.

Dolnicar, S. &. (2008). Selective marketing for environmentally sustainable tourism. *Tourism*

Management, 672-680.

Dolnicar, S. C. (2008). Environment-friendly tourists: what do we really know about them?

Journal of Sustainable Tourism, 197-210.

Duque Méndez, N. D. (2005). *Universidad Nacional de Colombia*. Obtenido de Universidad

Nacional de Colombia:

<http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060029/lecciones/cap8-1.html>

E, B. G. (2003). *Lic. en sistemas de información. Almacenes de datos y Minería de Datos.*

Obtenido de Lic. en sistemas de información. Almacenes de datos y Minería de Datos:

<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm>

Esap. (21 de Mayo de 2008). <http://cdim.esap.edu.co>. Recuperado el 1 de Septiembre de 2014, de <http://cdim.esap.edu.co>:

<http://cdim.esap.edu.co/BancoMedios/Documentos%20PDF/pd%20-%20plan%20de%20desarrollo%20-%20pereira%20-%20risaralda%20-%202008%20-%202011.pdf>

Estadística, D. A. (27 de Diciembre de 2007). <https://www.dane.gov.co>, 1. Obtenido de <https://www.dane.gov.co>:

https://www.dane.gov.co/files/investigaciones/fichas/pib/ficha_sateli_turismo.pdf

Feng, Y. L.-x.-y. (2010). Design of the Low-cost Business Intelligence System Based on Multi-agent. . *International Conference of Information Science and Management Engineering.*

Franks, B. (2012). Taming the big data tidal wave. En *Finding opportunities in huge data streams with advanced analytics*. New Jersey: Wiley.

Fu, R. X. (2008). Research on the Organizational Decision Support System for Small & Medium-Size Enterprise Based on Agent. *International Symposiums on Information Processing*. Obtenido de Research on the Organizational Decision Support System for Small & Medium-Size Enterprise Based on Agent.

Fuchs, M. &. (2004). Destination benchmarking - an indicator-system's potential for exploring

- guest satisfaction. *Journal of Travel Research*, 212-225.
- Fuchs, M. &. (2009). Data mining in tourism (In German: „Data Mining im Tourismus"). *Praxis der Wirtschaftsinformatik*, 73-81.
- Fuchs, M. C. (2011). Destination Brand Equity Modelling and Measurement. *A case from Sweden* (págs. 95-116). Emerald.: Goldsmith.
- Fuchs, M., Abadzhiev, A., & Svensson, B. (2013). A knowledge destination framework. *Scientific Journals of Croatia*.
- Gamble, P. M. (2000). Customer Knowledge Management in the Travel Industry. *Journal of Vacation Marketing*, 83-91.
- Google. (24 de Julio de 2013). <https://code.google.com/p/google-refine/>. Recuperado el 11 de Octubre de 2014, de google-refine/: <https://code.google.com/p/google-refine/>
- Goossens, C. (2000). Tourism information and pleasure motivation. *Annals of Tourism Research.*, 301-321.
- Gössling, S. (2009). Carbon neutral destinations: a conceptual analysis. *Journal of Sustainable Tourism*, 17-37.
- Grabova, Darmont, Chauchat, & Zolotaryova. (Junio de 2010). *ACM Digital Library*.
Recuperado el 27 de Octubre de 2014, de ACM Digital Library:
<http://dl.acm.org/citation.cfm?id=1893180>
- Gretzel, U. &. (2004). Implementing a knowledge-based tourism marketing information system. *Journal of Information Technology & Tourism*, 245-255.
- Guptara, P. (2000). Why Knowledge Management Fails. *Knowledge Management Review*, 26-30.

- H Witten, I., Eibe, F., & A. Hall, M. (2011). *Data Mining: Practical Machine Learning Tool and Techniques*. Morgan Kaufmann.
- Hallin, C. A. (2008). Knowledge management in the hospitality industry: a review of empirical research. *Tourism Management*, 366-381.
- Hastie, T. T. (2009). *The Elements of Statistical Learning*. New York: Springer.
- Hawkins, D. (2000). Transferring Tourism Knowledge. *Madrid: WTO Tourcom Conference*.
- Hayes-Roth, F. a. (1994). The State of Knowledge-Based Systems. *Communications of the ACM*, 27-39.
- Höpken, W. F. (2011). The knowledge destination – a customer information-based . *Information and Communication Technologies in Tourism 2011*, 417-429.
- Höpken, W. F. (2014). The Knowledge destination- Applying methods of business intelligence to tourism. *Encyclopedia of business analytics and optimization*, 307-321.
- <http://www.mividafreelance.com>. (s.f.). <http://www.mividafreelance.com>. Obtenido de <http://www.mividafreelance.com>: <http://www.mividafreelance.com/intercambiar-tu-casa-con-viajeros-tendencia-vacacional-auge/>
- Hunday, D. (2011). *Performance comparison of Apriori and FP-Growth algorithms in generating association rules*. Paris: Department of Computer Science.
- IBM. (18 de 06 de 2012). *Ibm Developers*. Recuperado el 2014, de Ibm Developers: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- IDC. (1 de Diciembre de 2012). <http://www.emc.com>. Recuperado el 1 de Octubre de 2014, de IDC EMC: <http://www.emc.com/about/news/press/2012/20121211-01.htm>
- IDC. (2014). <http://www.emc.com>. Recuperado el Octubre de 2014, de IDC EMC:

<http://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf>

Intermediaries, E. T. (23 de Febrero de 2012). *Enhancing Tourism Intermediaries*. Recuperado el 15 de Septiembre de 2014, de Enhancing Tourism Intermediaries: www.ipcsit.com/vol45/025-ICIKM2012-M0061.pdf

Jared, D. (2014). *Big Data, Data Mining and Machine Learning, value creation for business leaders and practitioners*. Wiley.

Jiawei han, M. K. (2012). 3rd edition data mining concepts and techniques. En M. K. Jiawei han, *3rd edition data mining concepts and techniques*. Morgan Kaufmann.

Jiawei, H., Micheline, K., & Jian, P. (2012). *Data Mining : Concepts and Techniques*. Morgan Kaufmann.

Jones, A. (2001). Knowledge Capitalism—Business. *Oxford University Press*.

Kahle, E. (s.f.). Implications of “New Economy” Traits for the Tourism Industry. *Knowledge Management in Hospitality and Tourism* (págs. 5-23). The Haworth Hospitality Press.

Kimball, R., Caserta, J., & 2004. (s.f.). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. En R. & Kimball. Indianapolis: Wiley Publishing, Inc.

Kindsein. (1 de 8 de 2015). <http://www.kindsein.com/es/29/1/666/>. Obtenido de <http://www.kindsein.com/es/29/1/666/>: <http://www.kindsein.com/es/29/1/666/>

Kuo, N.-W. (15 de Enero de 2011). *A Holistic Customer Experien*, 1. Recuperado el 15 de Septiembre de 2014, de A Holistic Customer Experien: www.aicit.org/AISS/ppl/AISS103PPL.pdf

Lang., K. N. (2 de Febrero de 2010). *The increasing focus on managing relationships and*

customer retention, 1. (Technol, Editor, Entrep, Productor, & Afr) Recuperado el 15 de Septiembre de 2014, de The increasing focus on managing relationships and customer retention: www.ajol.info/index.php/jolte/article/download/51992/40627

Larose, D. T. (2005). *Discovering Knowledge in Data*. John Wiley & Sons.

Larose, D. T. (2005). *Discovering Knowledge in Data an Introduction to Data Mining*. . New Jersey: Jhon Wiley & Sons, Inc Publication.

Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics*. Canada: Wiley.

Law, R. (1998). Room occupancy rate forecasting – A neural network approach. *International Journal of Contemporary Hospitality Management*, 234-239.

Li, X. Z. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses. *International Conference on Computational Science*.

Lin, T.-C. (25 de Enero de 2012). *International Conference on Information and Knowledge Management*. Recuperado el 2 de Septiembre de 2014, de International Conference on Information and Knowledge Management: <http://www.ipcsit.com/vol45/025-ICIKM2012-M0061.pdf>

Liu, B. (2008). *Web Data Mining* (2. New York: Springer.

LUCENA, M. A. (01 de 01 de 2005). Análisis comparativo entre métodos estadísticos y de minería de datos. Madrid. Obtenido de <http://www.iit.upcomillas.es/pfc/resumenes/42b98a68079d0.pdf>

Magnini, V. H. (2003). Data mining for hotel firms: use and limitations. *Cornell Hotel and Restaurant Administration Quarterly*, 94-105.

Mallo, P. E., Artola, M. A., Zanfrillo, A. I., Morettini, M., & Galante, M. (2007). *Sistema de*

Información Científica. Recuperado el 1 de Octubre de 2014, de www.redalyc.org:
www.redalyc.org/pdf/276/27611106.pdf

Manizales, A. d. (15 de Enero de 2012). www.manizales.gov.co, 1. Recuperado el 1 de Septiembre de 2014, de www.manizales.gov.co:
<http://www.manizales.gov.co/dmd/pd/doc1.pdf>

Mario, P., Adoración, M., & Esperanza, M. (2000). *Diseño de bases de datos relacionales*. Alfaomega.

McDonough, W. &. (2002). *Cradle to Cradle. Remaking the Way We Make Things*. Toronto: Douglas & McIntyre.

Merv Adrian. (2011). <http://www.teradatamagazine.com>. Recuperado el 23 de Octubre de 2014, de <http://www.teradatamagazine.com>:
<http://www.teradatamagazine.com/v11n01/Features/Big-Data/>

Microsoft. (1 de Febrero de 2014). <http://msdn.microsoft.com>. Recuperado el 1 de Octubre de 2014, de Microsoft: <http://msdn.microsoft.com/es-es/library/ms174949.aspx>

Min, H. M. (2002). A data mining approach to develop the profile of hotel customers. *International Journal of Contemporary Hospitality Management*, 274-285.

Min, W. (22 de Marzo de 2010). www.aicit.org, 5. Recuperado el 15 de Septiembre de 2014, de www.aicit.org: www.aicit.org/JCIT/paper_detail.html?q=269

Mining, J. H. (2000). *Concepts and Techniques*. Morgan Kaufmann Publishers Inc.

Moore. (1 de 11 de 2014). *Wikipedia*. Obtenido de Wikipedia:
http://es.wikipedia.org/wiki/Ley_de_Moore

Morales, D. R. (2008). Passenger name record data mining based cancellation forecasting for

revenue management. *Innovative Applications of O.R.*, 554-562.

Olmeda, I. &. (2002). Data mining techniques and applications for tourism internet marketing.

Journal of Travel & Tourism Marketing, 1-20.

Oracle. (15 de Enero de 2014). *Oracle NoSQL Database*. Recuperado el 23 de Octubre de 2014,

de Oracle NoSQL Database: <http://www.oracle.com/technetwork/database/database-technologies/nosqldb/overview/index.html>

otun, E. d. (28 de 7 de 2015). <http://www.eldiario.com.co>. Obtenido de

<http://www.eldiario.com.co>: <http://www.eldiario.com.co/seccion/LOCAL/c-mo-suena-pereira-de-noche-1507.html>

P, J. R. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster

analysis. *Journal of Computational and Applied Mathematics*, 20:53-65.

Pereira, A. d. (15 de Enero de 2014). www.pereira.gov.co, 1. Recuperado el 1 de Septiembre de

2014, de www.pereira.gov.co: www.pereira.gov.co/es/inicio.html

Pereira, F. a. (11 de 03 de 2014). <http://pereiraculturayturismo.gov.co>. Obtenido de

<http://pereiraculturayturismo.gov.co>:

http://pereiraculturayturismo.gov.co/index.php?option=com_content&view=article&id=350:puntos-de-informacion-turistica-reforzaran-la-atencion-de-propios-y-visitantes-durante-la-semana-santa&catid=89:noticiasinicio

PEREIRA, I. D. (15 de Octubre de 2014). *INSTITUTO DE CULTURA DE PEREIRA*.

Recuperado el 15 de Octubre de 2014, de <http://www.pereiraculturayturismo.gov.co>:

http://www.pereiraculturayturismo.gov.co/index.php?option=com_content&view=article&id=91&Itemid=631

- Pitman, A. Z. (2010). Web usage mining in tourism – a query term analysis and. *Information and Communication Technologies in* , 393-403.
- Press, E. (1 de 8 de 2015). <http://www.europapress.es>. Obtenido de <http://www.europapress.es>:
<http://www.europapress.es/sociedad/noticia-60-hijos-viven-padres-estan-parados-son-hombres-20150602142815.html>
- Prisa., E. P. (24 de 9 de 2013). *elpais.com*. Obtenido de elpais.com:
http://elviajero.elpais.com/elviajero/2013/09/19/actualidad/1379610088_258672.html
- Pujari, A. K. (2001). *Data Mining Techniques*. Universities Press.
- Pyo. (2002). Knowledge Discovery Database for Tourist Destinations. *Journal of*, (págs. 396–403).
- Pyo, S. (2005). Knowledge map for tourist destinations. *Tourism Management*, 583-594.
- Quindio, C. d. (1 de Mayo de 2014). www.eam.edu.co, 28. Recuperado el 1 de Septiembre de 2014, de www.eam.edu.co:
<http://www.eam.edu.co/centrodeinvestigaciones/politicaspUBLICAS/Observatorio%20Tur%20C3%ADstico>
- Quindio, O. T. (1 de Febrero de 2014). <http://www.eam.edu.co>. Recuperado el 15 de Octubre de 2014, de <http://www.eam.edu.co>:
<http://www.eam.edu.co/centrodeinvestigaciones/politicaspUBLICAS/Observatorio%20Tur%20C3%ADstico%20del%20Quind%20C3%ADo%202014.pdf>
- Quindio., T. (1 de 5 de 2015). <http://www.turismoquindio.com>. Obtenido de <http://www.turismoquindio.com>:
http://www.turismoquindio.com/fincaspanaca_herreria14.php

- Quintero K. (2015). Ideafreelance1: Alquilar tu casa o habitación a viajeros. Obtenido de:
<http://www.mividafreelance.com/intercambiar-tu-casa-con-viajeros-tendencia-vacacional-auge/>
- Rainardi, V. (2007). Building a data warehouse with examples in SQL Server. En V. Rainardi.
Nueva York: Apress.
- Redalyc. (05 de 10 de 2011). *Red de Revistas Científicas de América Latina y el Caribe, España y Portugal*. Obtenido de Universidad Autónoma de México:
<http://www.redalyc.org/home.oa>
- Reinsberg, C. &.-T. (2010). Characterizing environment friendly tourists. *Master Thesis*.
University of Stavanger. The Norwegian school of Hotel Management.
- República, B. d. (1 de 9 de 2014). <http://www.banrep.gov.co>. Obtenido de
<http://www.banrep.gov.co>:
http://www.banrep.gov.co/docum/Lectura_finanzas/pdf/dtser_208.pdf
- Risaralda., A. d. (1 de 06 de 2015). *Invest In Pereira*. Obtenido de Invest In Pereira:
<http://www.investinpereira.org/es/ipaginas/ver/98/ubicacion-geoestrategica-privilegiada/>
- Ritchie, R. J. (2002). A framework for an industry supported destination marketing information system. *Tourism Management*, 439-454.
- Roberto, N., & Marina, D. L. (abril de 2009). <http://www.scielo.org.co/>. Recuperado el 27 de Septiembre de 2014, de <http://www.scielo.org.co/>:
<http://www.scielo.org.co/pdf/iei/v29n1/v29n1a08.pdf>
- Sambamurthy, V. &. (2005). Information technologies and knowledge management.

Management , 1-7.

Santos, M. &. (2005). *Data Mining – Descuberta de Conhecimento em Bases de Dados. FCA Publisher.*

Schianetz, K. &. (2007). The learning tourism destination: the potential of a learning organization approach for improving the sustainability of tourism destinations. *Tourism Management*, 1485-1496.

Shawn, G., & Williams, A. (2009). Knowledge transfer and management in tourism organisations. *Tourism Management*, 325-335.

Sinnexus. (1999). <http://www.sinnexus.com>. Obtenido de <http://www.sinnexus.com>:
http://www.sinnexus.com/business_intelligence/

Soares, S. (2012). *Big data governance. An Emerging Imperative*. MC Press Online.

Sveiby, E. (15 de Marzo de 2001). *What is Knowledge Management?* Recuperado el 15 de Septiembre de 2014, de *What is Knowledge Management?*:
<http://www.sveiby.com/articles/>

Thierstein, A. &. (2000). The Sustainable Region: A Model for Action. *Series of works published by the Institute for Public Services and Tourism Contributions to the Regional Economy.*

Tiempo, E. (27 de 09 de 2015). *El Tiempo Noticias*. Obtenido de *El Tiempo Noticias*:
<http://www.eltiempo.com/politica/justicia/trafico-de-personas-en-colombia/16388090>

TripAdvisor. (1 de 9 de 2015). <https://www.tripadvisor.co>. Obtenido de
<https://www.tripadvisor.co>: <https://www.tripadvisor.co/HotelsList-Pereira-Hoteles-Baratos-zfp8016993.html>

Turismo, M. d. (1 de Septiembre de 2010). */www.mincomercio.gov.co*, 1. Recuperado el 1 de Septiembre de 2014, de www.mincomercio.gov.co:

<http://www.mincomercio.gov.co/eContent/documentos/turismo/2010/2010-TurismoNoviembre.pdf>

Turismo, M. y. (8 de Abril de 2014). *www.mincit.gov.co*, 1. Recuperado el 2014, de www.mincit.gov.co: <http://www.mincit.gov.co/descargar.php?id=71370>

Turismo, O. M. (22 de Enero de 2007). *www.unwto.org*, 1. Recuperado el 8 de Septiembre de 2014, de www.unwto.org: <http://media.unwto.org/es/content/entender-el-turismo-glosario-basico>

Turismo., M. I. (1 de 10 de 2015). <http://colombia.co>. Obtenido de <http://colombia.co>: <http://www.colombia.co/turismo/turismo-por-regiones/bogota-se-recorre-en-bicicleta.html>

Uber. (1 de 5 de 2015). <https://www.uber.com/es/cities/bogota>. Obtenido de <https://www.uber.com/es/cities/bogota>: <http://www.sociosbogota.com/vehiculos-aceptados.html>

unstats. (2 de Mayo de 1998). <http://unstats.un.org>, 1. Recuperado el 2 de Septiembre de 2014, de <http://unstats.un.org>: http://unstats.un.org/unsd/publication/SeriesM/SeriesM_67rev1S.pdf

Valhondo, D. (2003). *Gestión del conocimiento: Del mito a la realidad*. Madrid: Ediciones Díaz de Santos.

Vanguardia, P. (27 de 9 de 2015). <http://www.vanguardia.com>. Obtenido de <http://www.vanguardia.com>: <http://www.vanguardia.com/colombia/329543-eje->

cafetero-un-destino-de-clase-mundial

Velazco, R. H. (2007). *Almacenes de datos (Datawarehouse)*. Obtenido de Almacenes de datos (Datawarehouse).: <http://www.rhernando.net/modules/tutorials/doc/bd/dw.html>

Vlahogianni, E. I. (2010). Advanced computational approaches for predicting tourist arrivals. In T. *Advanced computational approaches for predicting tourist arrivals*. (págs. 309-324). Vienna: InTech.

Wang, Y. &. (2007). Conceptualizing and evaluating the functions of destination marketing systems. . *Journal of Vacation Marketing*, 187-203.

Watson, H. J., & Wixom, B. H. (Septiembre de 2007). *Ieee Computer Society*, 4. Recuperado el 23 de Octubre de 2014, de The community for Techology.

wikipedia.org. (11 de Mayo de 2014). *MongoDB*. Recuperado el 23 de Octubre de 2014, de MongoDB: <http://en.wikipedia.org/wiki/MongoDB>

Wöber, K. W. (1998). Global statistical sources- TourMIS. *Pacific Tourism Review*, 273-286.

Xie, H. (2009). The Research on Knowledge Management of Small and Medium-Sized IT Enterprises Based on Data Mining. *International Conference on Information Management, Innovation Management and Industrial Engineering*.

Yu, L. y. (2006). A comparative analysis of international tourists' satisfaction in Mongolia. *Tourism Management*, 1331-1342.

Yüksel, A. y. (2002). Measurement of tourist satisfaction with restaurant services A segment based approach. . *Journal of Vacation Marketing*., 52-68. .

Zaki, M., & Meira, Jr, W. (2014). *Data Mining and Analysis*. Cambridge.

Anexos

Anexo 1. Presupuesto Global Y Discriminado

PRESUPUESTO GLOBAL

Rubros	Valor (\$)
Personal	\$2.100.000
Materiales	\$1.420.000
Viajes socialización	\$0
Salidas de campo	\$0
Total	\$3.520.000

DESCRIPCION DE LOS GASTOS DE PERSONAL

Investigador / experto / auxiliar	Formación académica	Dedicación	Valor Unidad (\$)	Total (\$)
Investigadores principales:	Especialización	2 Horas Diarias	\$0	\$0
Co-Investigadores:	Ninguno			

Asesor estadístico	1	1 Semestre	\$2.100.000	\$2.100.000
Total				\$2.100.000

MATERIALES, SUMINISTROS Y BIBLIOGRAFIA

Materiales	Cantidad	Valor Unitario (\$)	Total (\$)
Fotocopias	5.000	\$50	\$ 250.000
Hojas digitadas	300	500	\$150.000
Uso computador (horas)	1	0	\$0
Uso Internet (horas)	2.000	\$6.000	\$720.000
Empastados	0	0	0
Argollados	0	0	0
Papelería y útiles de oficina	0	0	0
Discos compactos	0	0	0
Software licenciado	1	\$300.000	\$300.000
Total			\$1.420.000

DESCRIPCION Y JUSTIFICACION DE LOS VIAJES

Lugar / No. De viajes	Justificación	Pasajes (\$)	Estadía	Total (\$)
0	Ninguna	\$0	\$0	\$0
Total				\$0

SALIDAS DE CAMPO

Aspecto	Valor unitario (\$)	Cantidad (\$)	Proyectos	Total (\$)
0	Ninguna	\$0	\$0	\$0
Total				\$0

Anexo 2. Resultados, productos esperados y potenciales beneficiarios

Aspecto	Resultado / Producto	Indicador	Beneficiario
Generación de nuevo conocimiento	<p>- Documento de segmentación y descripción de hábitos de los viajeros de la ciudad de Pereira.</p> <p>-Documento con análisis de tendencias y relaciones entre los viajeros y los aspectos de su visita.</p> <p>-Documento con informe de percepción, satisfacción y nivel de gastos.</p>	<p>Inclusión en planes de desarrollo.</p> <p>Inclusión en planes de innovación.</p>	<p>Alcaldía.</p> <p>Gobernación.</p> <p>Fomento al turismo.</p> <p>Restaurantes.</p> <p>Hoteles.</p> <p>Comerciantes.</p> <p>Universidades.</p> <p>Periodistas.</p>
Fortalecimiento de la comunidad científica	<p>Cursos de minería e datos aplicados a encuestas.</p> <p>Ejemplos de Minería de datos aplicados a la planeación de las ciudades.</p>	<p>Número de personas capacitadas en minería de datos</p>	<p>Sociólogos.</p> <p>Grupos de investigación.</p> <p>Psicólogos.</p> <p>Servidores públicos.</p> <p>Comerciantes.</p>

	<p>Herramienta automática para segmentación de grupos de personas con x característica.</p> <p>Uso de herramientas de minería de datos para análisis y generación de informes ejecutivos a partir de encuestas.</p>	<p>aplicadas a sus áreas de desempeño y profesiones.</p>	<p>Alcaldes.</p> <p>Gobernadores.</p> <p>Departamentos de planeación.</p>
<p>Apropiación social del conocimiento</p>	<p>- Resumen ejecutivo de segmentación y descripción de hábitos de los viajeros de la ciudad de Pereira.</p> <p>-Resumen ejecutivo con análisis de tendencias y relaciones entre los viajeros y los aspectos de su visita.</p> <p>-Resumen ejecutivo con informe de</p>	<p>Número de asistentes a ponencias sobre el tema.</p> <p>Numero de emprendedores resultantes de</p>	<p>Alcaldía.</p> <p>Gobernación.</p> <p>Fomento al turismo.</p> <p>Restaurantes.</p> <p>Hoteles.</p> <p>Comerciantes.</p> <p>Universidades.</p> <p>Periodistas.</p> <p>Público en general.</p>

	percepción, satisfacción y nivel de gastos. -Presentación de resultados del estudio en ponencias.	ponencias.	
--	--	------------	--

Anexo 3. Cuadro de impactos esperados

Impacto esperado	Plazo (años)	Indicador verificable	Supuestos
Económicos: Estudio de mercado de la ciudad	4	Aumento del turismo no orientado a compras.	La ciudad está proyectada como comercial, pero podría tener más potenciales.
Ambiental: Planes para el turismo	3	Aumento en el uso de medios de transporte alternativos. Disminución de contaminación vehicular por uso de transporte alternativo. Aumento de turismo ecológico sostenible con responsabilidad social empresarial.	La ciudad tiene transporte público como: buses, busetas, taxis, transporte no legal como moto taxis y vehículos piratas, se podrían utilizar otros medios no contaminantes.

<p>Productividad:</p> <p>Planes para los viajeros de la ciudad</p>	5	<p>Número de eventos de convenciones nacionales.</p> <p>Número de eventos de convenciones internacionales.</p>	<p>En la ciudad hay un centro de exposiciones como Expofuturo y se están construyendo un centro de convenciones.</p> <p>La ciudad se podría proyectar como centro de convenciones importantes, no solo de comercio.</p> <p>Esto con un estudio para determinar la percepción del viajero con la ciudad.</p>
<p>Social:</p> <p>Implementar medidas de seguridad</p>	1	<p>Aumento de la seguridad por medio de la disminución de robos.</p>	<p>En las capitales de los departamentos de todo el país se vive la violencia</p>

Centros de apoyo a los indigentes			<p>diariamente, Pereira no es la excepción.</p> <p>Teniendo en cuenta la percepción de los viajeros sobre la ciudad se podrían crear planes de choque adicionales a los actuales para mejorar la convivencia.</p>
<p>Competitividad:</p> <p>Identificar opciones de negocio</p>	3	Aumento de las opciones de negocio de I+D+I y menos comercio.	<p>La ciudad ha aumentado el número de negocios formales.</p> <p>Ideal fuera que aumentaran los negocios del conocimiento.</p>

Social: Mejorar su servicio al cliente	1	Aumento de la percepción positiva por parte de viajeros.	Teniendo en cuenta la percepción de los viajeros sobre la ciudad. Mejorar muchos aspectos relacionados con las relaciones humanas, educación.

Anexo 4. Modelo inicial de la encuesta

Instituto de Cultura y Fomento al Turismo de Pereira**Datos Personales:**1. **Correo electrónico:**

2. **Edad:** 15-20 21-30 31-40 41-50 51-60 61-70 71-80 81-
903. **Sexo:** M F4. **Ciudad/ País de Nacimiento:**

5. **Ciudad/ País de Origen:**

6. **Ocupación:**

7. **Estado Civil:** Soltero(a) Casado(a) Divorciado(a)

Viudo(a)

8. **Número de veces que ha viajado a la Ciudad en un periodo de 3 meses: (Valor Numérico): _____**9. **Motivos de Viaje:** Turismo Comercio Aventura Educación Docencia

Cultural Otras:

10. **Número de Días de Permanencia: (Valor Numérico):**

11. **Compañeros de Viaje: (Valor Numérico):**

12. **Fuentes de Ingreso:** Empleado(a) Independiente

Otro: _____

13. **Como contactó la Agencia de Viajes con la cual viajó a la Ciudad de Pereira?**

 Personalmente Recomendado por un Tercero Empresa donde Labora

Otro: _____

14. **Lugar de Alojamiento durante el viaje:** Hotel Familiar Amigos

Otro: _____

15. **Primera vez que visita la Ciudad de Pereira?** Si No_# de veces:

16. **Que Actividades tenía planeadas en la Ciudad?**

 Turismo Comercio Aventura Educación Docencia Cultural Otras:

17. **Que actividades realizó de las que tenía planeadas?**

Turismo Comercio Aventura Educación Docencia Cultural Otras:

18. **Qué medio de transporte utilizo para llegar a la ciudad?**

Avión Transporte privado Transporte Publico

Otro(s): _____

19. Qué medio de transporte utilizo en la ciudad?

Avión Transporte privado Transporte Publico

Otro(s): _____

20. Qué lugares de alimentación utilizo en la Ciudad?

Restaurantes Comida Rápida en la Calle Familiar Amigos

Otro(s): _____

21.Cuál es su comida Preferida:

22.Cuál es el estimado del gasto de su visita (Valor en

Pesos)?: _____

23.Cuál es su calificación del gasto por motivo de su visita?:

24. Muy Alto Alto Medio Medio alto Bajo

Muy Bajo

25. Podría describir algún(os) aspecto(s) general(es) que lo impactaron cuando llegó a la Ciudad?

26. **Ej.** Alta contaminación, gente amable, mucho ruido, ambiente agradable, mucha indigencia.

27. Podría describir algún(os) aspecto(s) general(es) los cuales lo motivarían a regresar a la Ciudad?

Ej. Gente amable, clima, café, indigencia, motivos personales.

Anexo 5. Modelo encuesta aplicada

1.- Por favor ingrese su nombre / Please enter your name *

Su nombre no se utilizara por ningún motivo, permanecerá oculto de forma definitiva, es solo para validar que ya haya realizado la encuesta, muchas gracias por su comprensión. / Your Name will not be used, hidden will remain permanently , is just to validate that have already made the survey, thank you very much for your understanding.

2.- Edad / Age *

Por favor seleccione su rango de edad / Please select your age.

3.- Por favor seleccione su sexo / Please select your gender. *

- M / Male
- F / Female
- Other:

4.- Seleccione su país / Select your Country. *

Por favor seleccione su país de origen / Select your country of origin.

5.- Estado civil / Marital status. *

Por favor seleccione su estado civil / Please select your marital status.

- Soltero / Single
- Casado / Married
- Viudo / Widowed
- Divorciado / Divorced
- Other:

6.- Nivel de estudios / Level of education *

Seleccione una opción. / Select an option.

7.- Ocupación / Occupation *

Por favor ingrese su ocupación / Please select your occupation.

- Empleado / Employee
- Independiente / Independent
- Pensionado-Jubilado / Pensioner-Retired
- Desempleado / Unemployed
- Estudiante / Student
- Other:

8.- Fuentes de ingreso / Source of income *

Por favor ingrese su fuente de ingresos / Please input your source of income

- Empleado / Employment
- Independiente / Independent
- Pensionado-Jubilado / Pensioner-Retired
- Desempleado / Unemployed
- Estudiante / Student
- Familia / Family-Parents
- Other:

9.- Número de veces que ha viajado a la ciudad de Pereira / Number of times it has traveled to the city *

Si es la primera vez por favor seleccione cero (0) / If this is the first time please select zero (0).

10.- Fecha de Viaje a la Ciudad / Date of travel *

Seleccione la fecha de su ultimo viaje. / Select the date of your last trip to the city.

11.- Motivo de viaje a la ciudad / Purpose of travel to the city. *

Seleccione la opción que mas haya planeado / Select the option that more is planned.

- Turismo / Tourism
- Comercio / Buy in stores
- Aventura / Adventure tourism

- Educación / Education
- Docencia / Teaching
- Cultural / Culture
- Deportes / Sports
- Other:

12.- Numero de días de permanencia / Number of days of stay *

Seleccione un número, si tu viaje fue de paso por favor selecciona cero (0) / Please select a number. If you didn't sleep in the city please select zero (0).

13.- Numero de compañeros de viaje / Number of companions *

Seleccione un numero / Please select a number.

14.- Como se enteró de la ciudad para realizar el viaje / How did you hear of the city for the trip *

Seleccione una opción / Select an option.

- Agencia de viajes / Travel agency
- Television / TV
- Internet / Internet
- Recomendado / Friend.

- Empresa donde labora / Your employer.
- Other:

15.- Lugar de alojamiento durante el viaje / Travel accommodations *

Seleccione una opción / Select an option

- Hotel / Hotel
- Familia / Family
- Amigos / Friends
- Hostal / Hostal
- Other:

16.- Que actividad planeada realizo a satisfaccion ? / What activities performed from which you had planned? *

Seleccione una sola opción / Select one option.

- Turismo / Tourism
- Comercio / Buy in stores
- Aventura / Adventure tourism
- Educación / Education
- Docencia / Teaching
- Cultural / Culture

- Other:

17.- Que medio de transporte utilizo para llegar a la ciudad / Transport used to get into town. *

Seleccione una opción / Select an answer.

- Avión / Traveling by plane
- Transporte Privado / private transport
- Transporte publico / Public transport
- Other:

18.- Que medio de transporte utilizo para desplazarse en su estadía en la ciudad ? / Which transport did you use in the city ? *

Seleccione una opción / Select an option.

- Transporte publico / Public transport
- Transporte privado / Private transport
- Caminar / Walk
- Bicicleta / Bicycle
- Other:

19.- Que tipo de sitio de alimentación utilizo más ? / Which kind of site did you use to take meals ? *

Seleccione una opción. / Select an option.

- Restaurantes / Restaurants.
- Comida rápida en restaurantes / Fast food in restaurants.
- Comida rápida en la calle. / Fast food in the street.
- Familia / Family
- Amigos / Friends
- Other:

20.- Cual es su comida preferida ? / What is your favorite meal ? *

Seleccione una opción / Select an option.

- Pastas o pizzas / Italian food.
- Carnes Rojas / Meat
- Pollo / Chicken
- Mariscos / Seafood
- Comida picante o condimentada / Spicy food
- Other:

21.- Cual fue el gasto estimado antes de su visita / What was the estimated cost of your visit ? *

El costo estimado antes de la visita en dólares.: 1 dolar = \$2.000 pesos colombianos /

Estimated cost of your visit before traveling.

22.- Cual es su calificación del gasto de su visita ? / What is your rating of spending in your visit? *

Seleccione una opción, el costo esta dado en dólares en donde 1 dolar = \$2.000 pesos colombianos. / Select an option

- Muy bajo / Very low
- Bajo / Low
- Medio / Medium
- Medio alto / Medium high
- Alto / High
- Muy alto / Very high

23.- Seleccione el aspecto que más lo impacto cuando llego a la ciudad. / Select the aspect that most impact when it came to town. *

Seleccione una opción. / Select an option.

- Ambiente agradable / Very nice city.
- Clima agradable / Very nice weather
- Ciudad Ruidosa / Noisy city
- Alta contaminación. / High pollution
- Gente amable / Very nice people

- Gente poco amable / Rude people
- Ciudad segura / Safe city
- Ciudad insegura / Unsafe city
- Mucha indigencia / Many indigent people.
- Other:

24.- Que aspecto lo motivaría a regresar a la ciudad ? / Select the aspect would be caused to return to the city *

Seleccione una opción. / Select an option.

- Gente amable / Nice people.
- Clima agradable / Nice weather
- Bajo costo de vida / low cost of living
- La comida / Food
- Paisaje cafetero / Coffee landscape
- Cultura / Culture
- Vida nocturna / City nighthlife
- Todos los anteriores / All above
- Ninguno / None
- No volvería nunca / Never return

25.- Grado de Satisfacción- Aspectos Generales sobre la ciudad. *

Calificación de 1 a 5, siendo 1 la más baja y 5 la mayor.

	1	2	3	4	5
Oferta de Ocio y Entretenimiento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Oferta Comercial	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Oferta Cultural	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conservación de los Monumentos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Señalización Turística	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cultura de la Ciudadanía	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

26.- Grado de Satisfacción sobre la ciudad. Alojamiento *

Calificación de 1 a 5, siendo 1 la más baja y 5 la mayor.

	1	2	3	4	5
Oferta de Alojamiento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Calidad en la	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
atención recibida					
Disponibilidad					
de sitios de alojamiento	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Calidad del alojamiento					
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Precio del alojamiento					
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

27.- Grado de Satisfacción sobre la ciudad. Información turística brindada por diferentes sectores. *

Calificación de 1 a 5, siendo 1 la más baja y 5 la mayor.

	1	2	3	4	5
Información en la oficina de turismo					
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Información en el hotel					
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Información por personas en la calle					
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
Información					
por parte de la policía.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Información					
por parte de la policía de tránsito	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Información					
por parte de la policía de carreteras.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Información					
turística en vayas, carteles, volantes, folletos, guías, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		

Anexo 6. Tablas de pruebas

DATOS	ITERACION								
	1	2	3	4	5	6	7	8	9
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X	X	X	X	X	X	X	X
Fecha viaje	X	X	X	X				X	X
Sexo	X	X	X	X	X			X	X
Estado civil	X	X	X	X	X			X	X
Nivel estudio	X	X	X	X	X	X	X	X	X
Ocupación		X		X	X	X	X	X	X
Fuente ingreso		X	X	X	X	X	X		X
#veces que ha viajado		X	X	X	X	X	X		X
Motivo viaje		X		X	X	X	X		X
Días permanencia		X	X	X	X	X	X		X
Compañeros de viaje			X					X	X
Como se enteró de la ciudad				X				X	X
Lugar alojamiento							X	X	
Actividad que realizo							X	X	
Medio de transporte							X	X	

numero(unique integers)									
-------------------------	--	--	--	--	--	--	--	--	--

DATOS	ITERACION								
	10	11	12	13	14	15	16	17	18
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X	X	X	X	X	X	X	X
Fecha viaje	X	X	X	X				X	X
Sexo	X	X	X	X	X			X	X
Estado civil	X	X	X	X	X			X	X
Nivel estudio	X	X	X	X	X	X	X	X	X
Ocupación		X		X	X	X	X	X	X
Fuente ingreso		X	X	X	X	X	X		X
#veces que ha viajado		X	X	X	X	X	X		X
Motivo viaje		X		X	X	X	X		X
Días permanencia		X	X	X	X	X	X		X
Compañeros de viaje			X					X	X
Como se enteró de la ciudad				X				X	X
Lugar alojamiento							X	X	

)									
DATOS	ITERACION								
	19	20	21	22	23	24	25	26	27
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X	X	X	X	X	X	X	X
Fecha viaje								X	X
Sexo	X							X	X
Estado civil	X	X						X	X
Nivel estudio	X	X	X					X	X
Ocupación	X	X	X					X	X
Fuente ingreso	X	X	X	X					X
#veces que ha viajado	X	X	X	X				X	X
Motivo viaje	X	X	X	X	X			X	X
Días permanencia	X	X	X	X	X	X		X	X
Compañeros de viaje	X	X	X	X	X	X	X	X	X
Como se enteró de la ciudad	X	X	X	X	X	X	X	X	X
Lugar alojamiento							X	X	X
Actividad que realizo							X	X	X
Medio de transporte							X	X	X

nominal									
numeric(Range)									
Nominal a binomial, nominal numeric(Inter cuartil)									

DATOS	ITERACION								
	37	38	39	40	41	42	43	44	45
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X	X	X	X	X	X	X	X
Fecha viaje								X	X
Sexo	X							X	X
Estado civil	X	X						X	X
Nivel estudio	X	X	X					X	X
Ocupación	X	X	X					X	X
Fuente ingreso	X	X	X	X					X
#veces que ha viajado	X	X	X	X				X	X

	NORMALIZACION								
Nominal a binomial									
Nominal a binomial, nominal numeric(z- transformation)									
Nominal a binomial, nominal numeric(Range)									
Nominal a binomial, nominal numeric(Inter cuartil)	X	X	X	X	X	X	X	X	X

DATOS	ITERACION								
	46	47	48	49	50	51	52	53	54
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X	X	X					
Fecha viaje									
Sexo	X	X	X	X					
Estado civil	X	X	X	X					

itemDistribution	X	X	X	X	X	X	X	X	X
Gini	X	X	X	X	X	X	X	X	X
Silhouette	X	X	X	X	X	X	X	X	X
Jackard									
	NORMALIZACION								
Nominal a binomial									
Nominal a binomial, nominal numeric(z- transformation)									
Nominal a binomial, nominal numeric(Range)	X	X	X	X	X	X	X	X	X
Nominal a binomial, nominal numeric(Inter cuartil)									

DATOS	ITERACION								
	56	56	57						
VARIABLES									
<i>Objetivo 1 y 2</i>									
Edad	X	X							

Fecha viaje	X								
Sexo	X								
Estado civil	X								
Nivel estudio	X	X							
Ocupación	X	X							
Fuente ingreso	X	X	X						
#veces que ha viajado	X	X	X						
Motivo viaje	X	X	X						
Días permanencia	X	X	X						
Compañeros de viaje	X	X	X						
Como se enteró de la ciudad	X	X	X						
Lugar alojamiento	X	X	X						
Actividad que realizo	X	X	X						
Medio de transporte para llegar	X	X	X						
Medio de transporte utilizado	X	X	X						
Sitio alimentación	X	X	X						
Comida preferida	X	X							
INDICES									

DaviesBouldin	X	X	X						
avgWithin-Distance	X	X	X						
clusterDensity	X	X	X						
itemDistribution	X	X	X						
Gini	X	X	X						
Silhouette	X	X	X						
Jackard									
	NORMALIZACION								
Nominal a binomial, nominal numeric(Unique integers)	X	X	X						

DATOS	ITERACION								
	1	2	3						
VARIABLES									
<i>Objetivo 3</i>									
Gasto estimado	X	X	X						
Calificación del gasto	X		X						

Aspecto que lo impacto	X		X						
Oferta de ocio y entretenimiento	X		X						
Oferta comercial			X						
Oferta cultural			X						
Conservación de los monumentos			X						
Señalización turística	X		X						
Cultura ciudadana	X		X						
Oferta de alojamiento	X		X						
Disponibilidad de sitios	X		X						
Calidad del alojamiento			X						
Precio del alojamiento		X	X						
Info Oficina Turismo		X	X						
Info en el hotel		X	X						
Info en la calle	X	X	X						
Info policía	X	X	X						

Info Transito	X	X	X						
Info Policía Carreteras	X	X	X						
Info en vallas	X	X	X						
INDICES									
DaviesBouldin	X	X	X						
avgWithin-Distance	X	X	X						
clusterDensity	X	X	X						
itemDistribution	X	X	X						
Gini	X	X	X						
Silhouette	X	X	X						
Jackard									
NORMALIZACION									
Nominal a binomial, nominal numeric(Unique integers)	X	X	X						