

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

Trends and characteristics of travelers visiting the city of Pereira through data mining techniques.

Diego Fernando Salcedo Toro

Departamento de Ciencias Computacionales, Universidad Autónoma de Manizales, Manizales, Colombia
bayala@autonoma.edu.co

Resumen—El presente trabajo se realizó sobre los viajeros de la ciudad de Pereira hasta el año 2014.

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.

En los resultados arrojaron 5 clústers con los cuales se realizaron los análisis a la luz de los algoritmos y la realidad de la ciudad.

Palabras clave—Descubrimiento de conocimiento, aprendizaje no supervisado, el análisis asistido por ordenador, gestión del conocimiento, análisis de datos.

Abstract—This work was carried out on travelers Pereira city until 2014.

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 eventually were grouped by the level of expenditure, perception and satisfaction after visiting the city through cluster analysis.

In the results they yielded 5 clusters from which one it was realized the analysis.

Key Word —Knowledge Discovery, Unsupervised learning, Computer aided analysis, Knowledge management, Data analysis.

I. INTRODUCCIÓN

En el triángulo del café [1], luego de realizar un estudio del sector del turismo se pudo observar que las entidades públicas y privadas de la región recolectan datos sobre los viajeros, 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.

Según las estadísticas del ministerio de comercio del año 2011. [2] 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 [3] se mantiene dicha tendencia

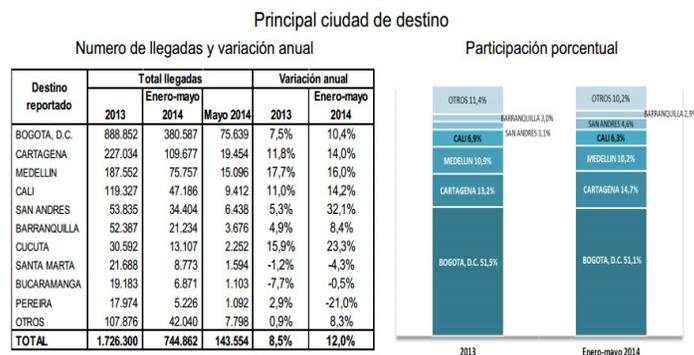


Figura 1. Principal ciudad de destino. Extranjeros no residentes.

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 [4].

En el plan de desarrollo de la ciudad de Pereira 2012-2015 “Por una Pereira mejor” [5] se tiene que 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% 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.[5]

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 [6].

Hasta el momento la información real de los turistas no se encuentra recopilada, solo se obtiene información de las cuentas satélites de turismo [7]. Dichas cuentas cada vez abarcan más indicadores de turismo para determinar satisfacción [6].

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 [8].

Por todo lo anterior es clara la necesidad de contar con datos estadísticos de Pereira y sus visitantes esto para que todas las

empresas de la ciudad de Pereira, la alcaldía, la cámara de comercio y demás entidades interesadas mejoren aspectos desconocidos para todos o conocidos para unos pocos a través de la caracterización de los visitantes.

II. DISEÑO DE LA METODOLOGIA

A. Selección Metodología CRISP-DM

Para esta investigación se requirió 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 [9].

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

Figura 2. Tareas genéricas y salidas del modelo CRISP-DM.

III. DESARROLLO DE LA METODOLOGÍA CRISP-DM

La metodología CRISP-DM se desarrolla por etapas, en total seis etapas.

A. Comprensión del Dominio

En esta etapa tenemos: objetivos institucionales de los encargados de turismo, evaluación de la situación, recursos hardware y software y los objetivos de la minería.

1). 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.

2). *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 general es:
- 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).

3). *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.

B. *Comprensión de los datos*

Esta etapa permite comprender la información con la que contamos para el análisis.

1). *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 [10], pero forman parte de un grupo a analizar, otro caso son los campos en los cuales se puede ingresar un dato escrito (campo Otros).

2). *Descripción de los datos*

Nuestra tabla consolidada contiene todos los datos de cada uno de los viajeros en un solo registro, estos fueron normalizados:

- 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 motivaría a regresar.
- Información sobre el grado de satisfacción: Sobre los aspectos de la ciudad, sobre el alojamiento, sobre información turística brindada a los viajeros.

3). *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 [11].

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 modificó 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 de los 50 mil recolectados inicialmente 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.

C. Preparación de los datos

En esta etapa se realiza la selección de los datos, la limpieza de los datos, discretización de los datos

1). Discretización

Todos los datos fueron seleccionados, la discretización permitió manipular los datos más fácilmente ya que reduce el rango de valores a utilizar [12].

Edad: La edad se definió en rangos: menos de 12, entre 18-21, entre 22-30, entre 31-40, entre 41-50, entre 51-60, entre 61-70, entre 71-80, entre 81-90, mayor a 91.

Los valores numéricos se discretizaron con rangos de: bajo, muy bajo, alto, medio, etc.

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

Tabla 1. Campos discretizados del número de viajes a la ciudad

Los datos para el grado de satisfacción se discretizaron en 3 descripciones:

Grado de Satisfacción.	Valor	Peso.
Entre 1 y 2	BAJO	1
Entre 3 y 4	BUENO	3
El valor de 5	EXCELENTE	5

Tabla 2. Campos discretizados del grado de satisfacción

2). Carga de datos

Luego de modificar los datos, limpiarlos y discretizarlos, se cargaron en la herramienta seleccionada para el análisis y la minería "RapidMiner".

clusterDensity										X		
itemDistribution												
Gini												
Silhouette												
NORMALIZACION												
Nominal número(dummy)	a	X	X	X	X	X	X	X	X	X	X	X
Nominal a número(effect)												
Nominal número(uniqueintegers)	a											
Nominal número(uniqueintegers)	a											

Tabla 3. Tablas de pruebas

El número de iteraciones para determinar el número óptimo de clústers para el objetivo uno y el objetivo 2 fue de 57, para el objetivo 3 fue de 3, esto teniendo en cuenta:

- Combinación de variables (edad, número de viajes, compañeros de viaje, etc).
- Indicador de calidad de los clústers (Davies Bouldin, avgWhithin, clusterdensity, ítem distribution, gini, silhouette)

El índice más valioso para determinar el número de clústers óptimo fue el Silhouette, este se determinó al tener un valor cercano a 1, indicando el número óptimo de clústers = 4.

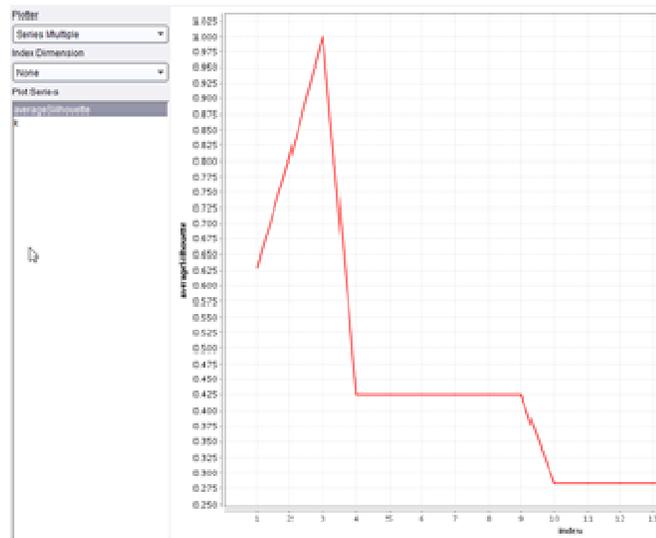


Figura 3. Plotter del índice Silhouette.

Para evaluar el proceso de reglas de asociación se tiene en cuenta las reglas que tengan mejor soporte y confianza, luego de esto se redacta la regla para ser comprendida más eficientemente.

Ejemplo de regla de asociación sobre el clúster 0:

Regla:

[Estado civil = Viudo / Widowed, Días de permanencia = BAJO] --> [Fuente de ingreso = Pensionado-Jubilado / Pensioner-Retired, Ocupación = Pensionado-Jubilado / Pensioner-Retired] (confidence: 0.801) (support: 0.348)

Esta regla indica que en el 35% de viajeros la condición de ser viudos y tener bajos días de permanencia (support) ocurre y el nivel de confianza de 80% indica que tenemos la probabilidad y seguridad de que dicha condición del 35% se cumplirá el 80% de las veces. En caso de que la confianza sea 1, esta equivale a que en el 100% de las veces siempre ocurrirá la condición.

3). Construcción del modelo

En esta fase se obtuvo 4 clústers y un quinto clúster al agrupar los outliers. El análisis se realizó teniendo en cuenta los resultados numéricos, el análisis objetivo y el análisis subjetivo, a la luz de los valores numéricos, los clústers utilizados para desarrollar el objetivo uno, dos y tres de la investigación son:

- 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.

Entre las estimaciones del gasto y percepción que hacen parte del objetivo tres tenemos que para el clúster uno:

El clúster inicialmente tenía un gasto estimado de 5.000 mil dólares o más, 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. [15].

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 [15], 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 [16].

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, lo 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, pocas vallas informativas, pocas pancartas y poco o nada de 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.

E. Evaluación del Modelo

En esta fase desarrollamos la evaluación de los resultados, el despliegue y el análisis de los resultados.

1). Evaluación de los Resultados

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 [16].

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.
- 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.

2). *Proceso de Revisión*

- 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.

3). *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.

4). *Análisis de los 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 (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.
- 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, muchos 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 valor que los asuntos negativos.

IV. 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 visualizaran como los datos nos generan información relevante que podemos utilizar para mejorar varios aspectos que no veamos 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

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 número 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 [16]
- 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.

V. RECOMENDACIONES

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.

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.
- 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.
- 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.

Los autores expresan sus más grandes agradecimientos a la facultad de Ingeniería de la Universidad Autónoma de Manizales, a su coordinadora Msc Beatriz Ayala, su decana MSc Alba Patricia Arias, al Dr. Mauricio F. Alba Castro Coordinador de la Maestría en gestión y desarrollo de proyectos de software, los docentes de la maestría: MSc. Sandra Victoria Hurtado Gil, Msc Javier Hernández Cáceres, al asesor de la tesis el ingeniero Msc Jairo Iván Vélez Bedoya, a la rectoría de la UAM, todos sus integrantes y al rector el Dr Javier Cadena, sin el apoyo de ninguna de las personas anteriormente mencionadas no hubiera sido posible el desarrollo y finalización de esta investigación.

VI. REFERENCIAS

- [1]. Observatorio Turístico del Quindío. (2014) Versión 28, Disponible en: <http://camaraarmenia.org.co/files/XXVIII%20Observatorio%20Tur%C3%ADstico%20del%20Quind%C3%ADo.pdf>
- [2]. Ministerio de Comercio Industria y Turismo. (2010) Informe Turismo Septiembre 2010. [En línea] Disponible en: <http://www.mincit.gov.co/descargar.php?id=56915>.
- [3]. Ministerio de Comercio Industria y Turismo. (2014). Informe Turismo Mayo 2014. [En Línea], Disponible en: <http://www.mincit.gov.co/descargar.php?id=71370>.
- [4]. (2014). Sitio web del Instituto Municipal de Cultura y Fomento al Turismo. [En Línea]. Disponible en: http://www.pereiraculturayturismo.gov.co/index.php?option=com_content&view=article&id=120&Itemid=672.
- [5]. Alcaldía de Pereira. (2012) Plan de desarrollo de Pereira 2012-2015 “Por una Pereira mejor”. [En Línea] Disponible en: http://temporal.ucp.edu.co/ucp_docs/ppr/documentos-generales/PLAN_DESARROLLO-pereira.pdf
- [6]. Departamento Nacional de Estadística DANE, Ficha metodológica cuenta satélite de turismo, Disponible en: http://www.dane.gov.co/files/investigaciones/fichas/pib/ficha_sateli_turismo.pdf 2007.
- [7]. BL. Arias Esquivel, C. Olaya Prieto. “Estudio de turismo para el departamento del Quindío”. Sinapsis Vol. 6, núm. 6 2014. [En Línea]. Disponible en: <http://www.eam.edu.co/ojs-2.4.2/index.php/prueba/article/view/71>
- [8]. N. Kefah. “The increasing focus on managing relationships and customer retention”, The Journal of language, technology and entrepreneurship in Africa, vol. 2, No 1, pp.1-8 Apr. 2010.
- [9]. M. Santos. *Descoberta de Conhecimento em bases de dados*. 1ra ed. Data Mining, Fca Publisher. 2005.
- [10]. DT: Larose, CD. Larose. *Data Mining and Predictive Analytics*. 1ra ed. Canada. Wiley. 2005.
- [11]. Google Refine, Google Refiner. Free-Open Edition Version 2.0. 2011-2015. Inc. Google.
- [12]. Facultad de ingeniera, Universidad del Valle. (2009). preprocesamiento. [En Línea]. Disponible en: <http://ocw.univalle.edu.co/ocw/ingenieria-de-sistemas-telematica-y-afines/descubrimiento-de-conocimiento-en-bases-de-datos/material-1/Preprocesamiento.pdf>
- [13]. J. Han., K. Micheline., J. Pei *Concepts and Techniques*, 3ra ed. Morgan Kaufmann Publishers Inc 2000.
- [14]. M.S. Mythili, A.R. Shanavas, "Performance Evaluation of Apriori and FP-Growth Algorithms atmosphere," International Journal of Computer Applications. (0975-8887) Vol 79-No 10, pp. 1-4 Oct. 2013.
- [15]. Departamento Nacional de Estadística DANE. (2015). Índice de Precios al Consumidor, Marzo 2015. [En Línea]. Disponible en: http://www.dane.gov.co/files/investigaciones/boletines/ipc/cp_ipc_feb15.pdf
- [16]. (2015). Sitio Web de Tripadvisor. [Online] Disponible en: <https://www.tripadvisor.co/HotelsList-Pereira-Hoteles-Baratos-zfp8016993.html>