



Universidad Cenfotec

Maestría en Tecnología de Bases de Datos.

Documento final de Proyecto de Investigación Aplicada 2

Metodología para identificar los puntos geográficos con mayor cantidad de incidentes de tránsito en las carreteras de Costa Rica, aplicando el algoritmo de agrupamiento DBSCAN, a la información recolectada por Ruta Alternativa, durante el primer semestre del año 2013.

Marín Umaña Greivin Alonso

Setiembre – 2015

Declaratoria de Derechos de Autor

- Ninguna parte de esta tesis puede reproducirse o transmitirse bajo ninguna forma o por ningún medio o procedimiento, sin permiso por escrito del autor.
- Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.
- La presentación de este documento está sometida a procesos de confidencialidad. La vigencia de esta cláusula es por un periodo de: 3 años, a partir de la fecha de realizado el documento. Una vez pasado dicho periodo, el trabajo se pondrá a disposición de estudiantes, personal docente o administrativo, en la biblioteca de la Universidad, tanto para estudio, consulta o ejemplificación en clases relacionadas con el tema.

Dedicatoria

Le dedico primeramente mi trabajo a Dios fue el creador de todas las cosas, el que me ha dado fortaleza para continuar cuando a punto de caer he estado; por ello, con toda la humildad que de mi corazón puede emanar.

De igual forma, a mis Padres, a quien le debo toda mi vida, les agradezco el cariño y su comprensión, a ustedes quienes han sabido formarme con buenos sentimientos, hábitos y valores, lo cual me ha ayudado a salir adelante buscando siempre el mejor camino.

Agradecimiento

Primero y antes que nada, dar gracias a Dios, por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio.

Agradecer hoy y siempre a mi familia por el esfuerzo realizado por ellos. El apoyo en mis estudios, de ser así no hubiese sido posible. A mi Esposa e Hijas ya que me brindan el apoyo, la alegría y me dan la fortaleza necesaria para seguir adelante.

Tabla de contenido

Declaratoria de Derechos de Autor.....	ii
Dedicatoria	iii
Agradecimiento.....	iii
Tabla de contenido	v
Resumen ejecutivo	1
Capítulo 1 – Introducción	2
1.1 Identificación del proyecto.....	2
1.2 Justificación del proyecto	2
1.3 Antecedentes del proyecto.....	3
1.4 Beneficios esperados	5
1.5 Objetivos	6
1.5.1 Objetivo general	6
1.5.2 Objetivos específicos	6
1.6 Productos y resultados.....	7
1.6.1 Productos	7
1.6.2 Resultados.....	7
1.7 Alcances y Limitaciones.....	7
1.7.1 Alcances	7
1.7.2 Limitaciones.....	8
Capítulo 2 – Marco Teórico	9

2.1	Funcionamiento del Algoritmo DBSCAN	11
2.2	Funcionalidad de agrupamiento de Google Maps	12
2.3	Funcionalidad de SQL Server Spatial Tools	15
Capítulo 3 – Marco Metodológico.....		17
3.1	Modalidad de la investigación	17
3.2	Tipo de Investigación.....	17
3.3	Enfoque de la Investigación.....	18
3.4	Programa de acción.....	18
3.4.1	Fase I – Investigación de Estadísticas Previas	18
3.4.2	Fase II – Diagrama de Estructura de datos.	19
3.4.3	Fase III – Diseño “Desnormalizado”	19
3.4.4	Fase IV – Traslado de los datos	20
3.4.5	Fase V – Desarrollo de aplicación para algoritmo DBSCAN	20
3.4.6	Fase VI – Prueba de Ejecución del aplicativo.....	21
3.5	Unidades de Observación	21
3.5.1	Variables.....	22
3.5.2	Indicadores.....	22
3.6	Población y Muestra.....	23
3.6.1	Población	23
3.6.2	Muestra.....	24
3.7	Técnicas de análisis de los datos	24

3.7.1	Procesando la información	24
3.7.2	Preparando los resultados obtenidos	24
Capítulo 4 – Propuesta de Solución		26
4.1	Plan de Inversión	26
4.2	Análisis del Software de Recolección de datos	26
4.2.1	Diseño de la Base de Datos	26
4.2.2	Software de Registro de Incidentes	31
4.3	Tabla des-normalizada	34
4.4	Elección del Algoritmo de Agrupamiento	34
4.5	Elección de la herramienta.....	35
4.6	Desarrollo de Herramienta que aplica el algoritmo de agrupamiento	35
4.7	Aplicando el algoritmo DBSCAN sobre la información recolectada	39
4.7.1	Cluster No. 1	43
4.7.2	Cluster No. 2	44
4.7.3	Cluster No. 3	45
4.7.4	Cluster No. 4	46
4.7.5	Cluster No. 5	47
4.7.6	Cluster No. 6	48
4.7.7	Cluster No. 7	49
4.7.8	Cluster No. 8	50
4.7.9	Cluster No. 9	51

4.7.10	Cluster No. 10	52
4.8	Otros datos Importantes.....	53
4.8.1	Incidentes por Provincia.....	53
4.8.2	Incidentes por Día de Semana	54
4.8.3	Incidentes por Hora del Día.....	54
4.8.4	Incidentes por Cantón (San José).....	55
4.8.5	Cantidad de Incidentes por Tipo	56
	Capítulo 5 – Análisis y Discusión de Resultados.....	56
	Capítulo 6 – Conclusiones y trabajos futuros	58
6.1	Conclusiones	58
6.2	Reflexiones Finales	59
6.2.1	Reflexiones Documentales	59
6.2.2	Reflexiones Técnicas	60
6.3	Trabajos Futuros	61
	Referencias	63
	Bibliografía	64
	Apéndices	65

Resumen ejecutivo

Inicialmente este proyecto fue enfocado en el análisis de información sobre la problemática de los incidentes de tránsito que se presentan en las carreteras de Costa Rica, mediante la utilización de algoritmos de minería de datos, más específicamente, algoritmos de agrupamiento (*Clustering*). Sin embargo, en un intento por buscar información sobre este tema, encontramos que existe una enorme carencia de este tipo de datos en el país.

La poca información que se logra recolectar no está estructurada de ninguna manera, por lo que no permite realizar un análisis o aplicar algoritmos de minería de datos.

Debido a este problema, se hace necesario crear una base de datos que permita almacenar esta información y, a su vez, desarrollar un aplicativo que permita ingresar dicha información en la base datos.

Se realiza un convenio con Ruta Alternativa, entidad que se encarga de recolectar información sobre incidentes de tránsito desde hace varios años, para que sean ellos quienes registren y almacenen dicha información y, de esta forma, luego de varios meses de datos recolectados, se pueda aplicar el algoritmo, y así encontrar las ubicaciones geográficas que presentan mayores problemas en el país.

El algoritmo a utilizar en el análisis es el “Algoritmo de agrupamiento espacial basado en densidad con ruido” o DBSCAN por sus siglas en Inglés (Density-Based Spatial Clustering of Applications with Noise), el cual consiste en agrupar diferentes puntos basándose en la distancia existente entre ellos.

Una metodología que innova la forma de visualizar e identificar los incidentes de tránsito haciéndolo de una manera gráfica. Directamente en un mapa de Costa Rica se visualizan los resultados del algoritmo aplicado a la información, mostrando los diferentes clusters. Se espera que los resultados obtenidos en este proyecto sean de ayuda para el análisis y la toma de decisiones por parte del gobierno en la infraestructura vial de Costa Rica.

Palabras clave (Keywords): Ruta Alternativa, Incidentes de Tránsito, Ubicaciones Geográficas, Minería de Datos, Agrupamiento, Clustering, DBSCAN.

Capítulo 1 – Introducción

1.1 Identificación del proyecto.

Este proyecto se clasifica como un proyecto de investigación y análisis, en el cual se plantea analizar la información sobre incidentes de tránsito en nuestro país. Con dicho análisis se busca determinar cuáles puntos o ubicaciones geográficas son las más problemáticas.

A pesar de ser en esencia un proyecto de investigación y análisis, se requiere una parte de desarrollo de software, ya que el convenio con la empresa Ruta Alternativa radica en que ellos van a registrar la información si cuentan con el software para registrarla, por lo que se requiere desarrollar este pequeño módulo de registro.

1.2 Justificación del proyecto

Según las estadísticas existentes en nuestro país, los accidentes de tránsito son, desde hace varios años, una de las principales causas de muerte de manera violenta.

Las estadísticas de accidentes automovilísticos en el transcurso del año 2010 son preocupantes, los mismos han enlutado a muchos hogares costarricenses, y son una de las principales causas de muerte de este país. Es importante indicar que para el periodo en mención (2010-2011) las estadísticas reportaron un total de 62604 accidentes ocasionando 287 víctimas mortales generando en promedio 0.78 muertes al día. (Informe Final de Gestión, 2012)

A pesar de ser una situación de mucho impacto para el país, existe una carencia de datos estadísticos de los diferentes incidentes sucedidos en las carreteras: “Ante la ausencia de bases de datos adecuadas, que deriven en estadísticas de accidentes a nivel nacional, ha sido común adoptar como propias estadísticas elaboradas en otros países” (Colegio de ingenieros Civiles, s.f.).

Las principales estadísticas encontradas, en su mayoría se refieren al número de muertes categorizadas por tipo de accidentes, provincia, o periodos de tiempo. Sin

embargo, si se quisiera realizar un análisis sobre cuáles son las rutas más peligrosas, calles o cruces con más accidentes, no se podría llevar a cabo de manera sencilla, ya que no existe una base de datos con dicha información.

Ahora bien, la información sobre las ubicaciones de los incidentes de tránsito es manejada por muchos de los medios de comunicación, como lo son canales de televisión o emisoras radiales. Algunos medios tienen programas televisivos o de radio que aluden a este tipo de información con la finalidad de dar a conocer a sus televidentes o radioescuchas, respectivamente, donde están ubicados los incidentes.

No obstante, aunque esta información se maneja en estos medios, no existe hábito de almacenarla para análisis o estadísticas y, si en algún momento se almacena, no se guarda de una manera estructurada.

Ruta Alternativa, una entidad dedicada a recopilar esta información y brindarla a medios de difusión, ha ofrecido sus servicios para almacenar estos datos y, al cabo de unos meses, poder brindar información concisa y real que permita ayudar con el propósito de este proyecto.

Este proyecto busca demostrar la importancia del análisis de la información como método para la toma de decisiones con respecto a la infraestructura vial de Costa Rica. Además, y lo que es más importante aún, busca fomentar el hábito de crear una base de datos con información histórica que se encuentre de manera ordenada y estructurada, y que, a su vez, permita desarrollar fácilmente estos análisis y estadísticas.

1.3 Antecedentes del proyecto

Al tratarse el proyecto de una problemática generalizada en Costa Rica, los incidentes de tránsito, podemos encontrar diversos trabajos previos con respecto a estadísticas y análisis de información sobre este tema. Tenemos en nuestro país entidades gubernamentales involucradas, como lo son el MOPT, COSEVI, 911, y encontramos entidades no gubernamentales, como “Fundación S.O.S. Paz en las Carreteras”, Cruz Roja, entre otras.

A continuación, se muestran dos gráficas de estadísticas sobre incidentes de tránsito sucedidos en Costa Rica. La figura 1.1 tomada del sitio web de la “Fundación SOS Paz en las carreteras”, nos muestra la cantidad de muertes totales en sitio por accidente de tránsito (2000-2009).

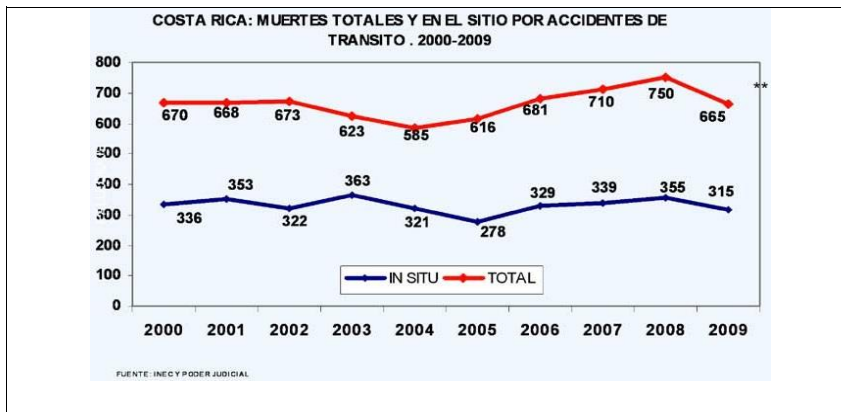


Figura 1.1 Muertes Totales en Sitio por accidentes de tránsito

Fuente: Estadística, s.f.

En la figura 1.2 apreciamos una gráfica con el número de accidentes con víctimas según tipo de lesión entre los periodos 2008 y 2009.

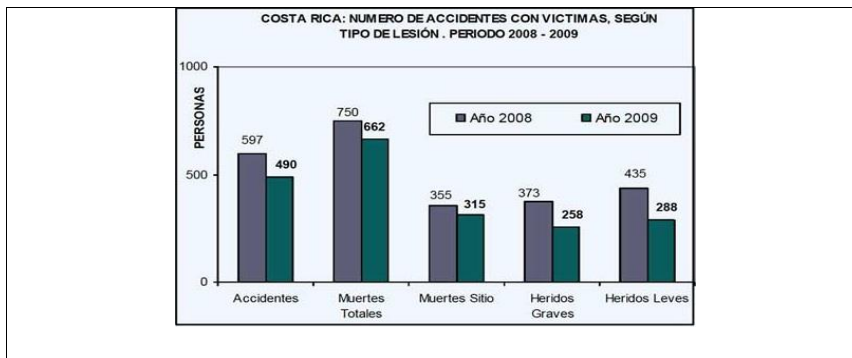


Figura 1.2 Número de accidentes con víctimas según tipo de lesión

Fuente: Estadística, s.f.

Adicionalmente, se encuentran estadísticas realizadas por la Dirección General de la Policía de Tránsito (DGPT), que muestran las cantidades de accidentes y muertes por mes. La figura 3 presenta un cuadro comparativo entre el año 2010 y el 2011 con respecto a los accidentes fatales y las “muertes in situ” por mes.

Meses	ACCIDENTES FATALES			MUERTES "IN SITU"		
	2010	2011	Diferencia	2010	2011	Diferencia
Enero	23	27	4	27	29	2
Febrero	25	25	0	28	27	-1
Marzo	16	20	4	16	21	5
Abril	22	24	2	24	26	2
Mayo	27	28	1	31	29	-2
Junio	16	26	10	20	32	12
Julio	18	22	4	19	31	12
Agosto	34	14	-20	37	15	-22
Septiembre	16	23	7	16	24	8
Octubre	18	13	-5	20	14	-6
Noviembre	18	10	-8	19	11	-8
Diciembre	28	25	-3	30	30	0
TOTAL	261	257	-1	287	289	2

Figura 1.3 Comparativo año 2010 y 2011 Accidentes fatales y “muertes in situ”

Fuente: Informe Final de Gestión, 16/07/2012

Podemos notar que, en su mayoría, las estadísticas existentes tienden a enumerar o contabilizar las personas lesionadas o fallecidas en incidentes de tránsito. Sin embargo, la información sobre las ubicaciones geográficas precisas de estos incidentes es escasa, por lo que la principal motivación de este proyecto es el poder complementar las estadísticas existentes con información de las ubicaciones geográficas más precisa.

1.4 Beneficios esperados

Se espera que este proyecto provea información útil que ayude a identificar los puntos más peligrosos de las carreteras de nuestro país y que, en algún grado, permita a las entidades competentes tomar mejores decisiones con respecto a la infraestructura vial.

1.5 Objetivos

1.5.1 Objetivo general

Proponer una metodología para identificar los puntos geográficos con mayor cantidad de incidentes de tránsito en las carreteras de Costa Rica, aplicando el algoritmo de agrupamiento DBSCAN a la información recolectada por Ruta Alternativa durante el primer semestre del año 2013.

1.5.2 Objetivos específicos

- Investigar las estadísticas existentes en Costa Rica sobre incidentes de tránsito que provean información, para plantear los antecedentes e información existentes en esta área.
- Desarrollar un proceso que permita trasladar la información sobre incidentes de tránsito desde el repositorio original de Ruta Alternativa hacia un área de trabajo en el que se pueda ordenar dicha información de una manera conveniente para el proceso de análisis.
- Desarrollar una aplicación informática que permita aplicar el algoritmo basado en densidad a la información recolectada por Ruta Alternativa sobre incidentes de tránsito, esto con la finalidad de identificar las ubicaciones geográficas con mayor cantidad de incidentes.
- Aplicar el algoritmo de agrupamiento a los datos recopilados por Ruta Alternativa durante el primer semestre del año 2013, utilizando la herramienta mencionada en el objetivo anterior.
- Realizar un informe final en el que se documente los hallazgos y conclusiones encontradas en el proceso de análisis aplicado a la información de incidentes de tránsito.

1.6 Productos y resultados.

1.6.1 Productos

- Software de aplicación del Algoritmo DBSCAN, ejecutables y código fuente realizado en Visual Studio 2010 (Web).
- Documento con el resultado del análisis y las conclusiones.

1.6.2 Resultados

Análisis de la información sobre los incidentes más comunes sucedidos en las carreteras de Costa Rica durante el primer semestre del año 2013.

Además, se obtendrá un software que aplica el algoritmo de agrupamiento DBSCAN, a la información de los incidentes de tránsito como metodología para identificar los puntos geográficos con más incidentes en las carreteras de Costa Rica.

1.7 Alcances y Limitaciones

1.7.1 Alcances

- a. **Herramienta de análisis para la información:** Fue necesario desarrollar el software que aplica el algoritmo de densidad a la información recolectada por Ruta Alterna, debido a que se dieron cambios que se podían explotar de una mejor manera con una herramienta hecha a la medida, en lugar de utilizar las herramientas que hay en el mercado actualmente. Dicho cambio obedece a que no se iba a poder lograr el nivel de granularidad y exactitud deseado con la estructura que se había planteado, la cual consistía en almacenar las direcciones con la estructura Provincia, Cantón, Distrito electoral y un texto de dirección, sino que hubo que agregar las coordenadas geográficas. Aunque estos datos adicionales nos proporcionan una mayor exactitud, se decidió que la herramienta a utilizar ya no sería Weka, puesto que se deseaba visualizar los *clusters* en un mapa de Google que permitiera

visualizarlos sobre el mapa de Costa Rica y no simplemente en la pantalla de exploración de Weka.

- b. **Aplicando algoritmo DBSCAN:** A la información recolectada se le aplicará el algoritmo DBSCAN mediante una aplicación desarrollada como parte de este proyecto, la cual calculará la distancia entre los incidentes a partir de las coordenadas geográficas provistas por Ruta Alterna. Además generará los *clusters* de los incidentes para que se muestren en un mapa de Google Maps y así determinar la ubicación geográfica, tanto del *cluster*, como de los incidentes en específico.

1.7.2 Limitaciones

En este apartado se describen las limitantes que se encontraron durante la realización de este proyecto. Cabe destacar que estas limitaciones no impidieron la realización del proyecto, sin embargo, son limitantes que afectan de una manera u otra el resultado.

- a. **Fidelidad de los datos:** Ruta Alterna se compromete a recolectar la información de la manera más exacta y eficiente posible. Sin embargo, siempre hay que tener en cuenta que la información podría no estar completa para algunos casos, ya que existe el riesgo de que en ocasiones la información no llegue a manos de Ruta Alterna, por ejemplo, debido a la lejanía de la ubicación del incidente, o a que ninguno de los corresponsales de Ruta Alterna reciba el dato.
- b. **Coordenadas Geográficas:** Se debe tomar en cuenta que las coordenadas geográficas son seleccionadas en un mapa de Google y que, tanto la Latitud como la longitud, se presentan en grados, los cuales son números muy largos y en ocasiones podrían presentar leves diferencias con respecto a su ubicación real.
- c. **Tiempo de recolección de datos:** Se tiene una limitante con respecto al tiempo. Aunque hubiera sido bueno contar con la información de un año completo, debido a la limitante del tiempo, solo contaremos para este proyecto con información con coordenadas geográficas de cinco meses, ya que se inició a recolectar el 2 de Febrero del 2013 hasta la fecha en que se hizo un corte el 14 de julio del 2013. De

manera que, para este proyecto, no contaremos con información para los demás meses (desde Julio 2013 en adelante).

1.8 Reseña de la empresa

Ruta Alternativa es una pequeña empresa que se dedica desde el 2003 a recolectar información sobre los diferentes incidentes sucedidos en carretera para brindarlos a los medios de comunicación, como canales televisivos y emisoras radiales.

Es manejada principalmente por su fundador, el cual es ayudado en ocasiones por su familia (esposa e hijo), por lo cual no es una empresa de mayores dimensiones y no han definido muchas características propias de una empresa, como lo son la cédula jurídica (está en proceso), misión, visión, y tampoco se podría obtener un organigrama de dicha empresa.

Actualmente, se cuenta con un aproximado de veinte corresponsales fijos que ayudan con la recolección de la información sobre los incidentes.

Capítulo 2 – Marco Teórico

El texto citado abajo deja claro desde hace cuánto tiempo aproximadamente los incidentes de tránsito comenzaron a ser un problema social para Costa Rica.

Un aspecto importante a rescatar en el desarrollo de esta investigación es el concerniente al origen de la problemática vial. En este plano, conviene descartar la expresión que afirma que los accidentes de tránsito son una problemática exclusiva de los llamados países industrializados, más bien se explicará el fenómeno a partir del crecimiento y expansión de la economía y del crecimiento demográfico que experimenta el país a partir del conflicto político y militar del año 1948.

Cabe recordar que antes de esa fecha el desarrollo nacional se apoyaba básicamente en la agricultura, los factores de producción no se caracterizaban por su desarrollo tecnológico. Así mismo la comunicación vía terrestre al interior del país presentaba

serias deficiencias (tanto por las escasas carreteras, su mala condición o por la escasez de vehículos), aspectos que de una u otra forma incidían en la existencia de una flota vehicular relativamente baja, así como de una red de carreteras poco extensa.

Ya para la década de los 50, el panorama social y económico comienza a reconfigurarse a partir de la implementación de un modelo de desarrollo que buscaba la sustitución de las importaciones por la promoción de las exportaciones...

[...] Para 1983 el INS registró un total de 210044 vehículos en circulación, ocho años después la cantidad ascendió a 332876 vehículos, lo que significa un aumento del 58.47%; no obstante, la población en ese mismo período crece un 26.04% (de 2403781 a 3029746 personas). Siempre basándose en estadísticas del INS, el siguiente dato resulta aún más alarmante: para 1983 se reportaron 16569 accidentes, en 1991 se reportaron 40092 accidentes, lo que implica que en ese período los accidentes crecen un 141.94%. (Chinchilla, 1992.)

Podemos encontrar en el texto citado anteriormente un punto de inicio en el año 1948, cuando los cambios políticos y económicos del país provocaron, entre otras cosas, un crecimiento en la flota vehicular. Se da así inicio a una gran problemática con los incidentes de tránsito, que aqueja al país aún hoy en día.

En el año 2012, con estadísticas más recientes por parte del Ministerio de Obras Públicas y Transportes (MOPT), se puede observar que aún existe esta situación con los incidentes de tránsito.

Los tipos de accidentes que generaron mayores víctimas mortales durante el año 2010 fueron las Colisiones (134 Colisiones de 261 accidentes) equivalente al 51.34% y ocasionando 134 fallecidos, los atropellamientos (77 Atropellos de 261 accidentes) para un 29.50% y 79 muertes, Salidas de Vehículos de la Vía (30 Salidas de Vía de 261 accidentes) para un 11.5% y 37 Muertes, Colisión con objeto fijo (18 Colisiones con Objeto Fijo de 261 accidentes) para un 6.9% ocasionando 20 muertes, vuelco (15 Vuelcos de 261 Accidentes) para un porcentaje de 5.74% ocasionando 15 fallecidos y finalmente Colisión con Semoviente (2 Colisión con

Semoviente de 261 accidentes) para 0.76% ocasionando 2 muertes. (Informe Final de Gestión, 2012.)

2.1 Funcionamiento del Algoritmo DBSCAN

Este algoritmo realiza agrupamientos basado en densidad, la cual sería la cantidad de elementos en un específico radio de distancia (eps). El parámetro EPS es la distancia que debe existir entre dos elementos. DBSCAN permite identificar cuáles elementos son simple ruido mediante la variable de Cantidad mínima de puntos (MinPts).

Para más detalle sobre la lógica de programación utilizada en el algoritmo DBSCAN, se puede visualizar el seudocódigo de dicho algoritmo (véase Apéndice A).

Los elementos se clasifican en tres:

1. Elementos centrales (Core Point): son los que están en el centro del *cluster*.
2. Elementos de borde o límite (Border Point): están un poco alejados del centro del *cluster*; sin embargo, se encuentran cercanos a un Elemento central o Core Point.
3. Elementos Ruido (Noise Point): Estos simplemente son los restantes después de clasificar los elementos centrales y de borde.

El cálculo de DBSCAN se basa en seis reglas:

1. **El vecino eps de un punto:** Para que un punto pertenezca a un *cluster* debe tener otro punto cercano a él según la distancia establecida (eps).
2. **Densidad alcanzable directa:** Existen dos tipos de Puntos, los centrales y los de Borde.
3. **Densidad alcanzable:** Un punto p es alcanzable en la densidad desde un punto q con respecto a la distancia (eps) y la cantidad mínima de puntos (MinPts).
4. **Densidad Conectada:** Existen casos en los que dos puntos borde pertenecientes a un mismo *cluster* no comparten el mismo punto central. En esta situación los puntos no van a tener Densidad alcanzable entre ellos mismos.

5. **Cluster:** Si un punto p pertenece a un *cluster* C y un punto q es de Densidad Alcanzable del punto p , dada la distancia y la cantidad mínima de puntos, entonces q es también parte del *cluster*.
6. **Ruido (Noise):** Son los puntos que no pertenecen a ningún *cluster*.

A continuación se muestran algunos puntos que ayudaron a determinar la elección del algoritmo DBSCAN:

- DBSCAN no requiere saber el número de *clusters* como se requiere en el K-Means.
- DBSCAN puede encontrar *clusters* de diferentes formas arbitrarias, lo que favorece este proyecto al tratarse de información de carreteras donde se está aplicando el algoritmo.
- DBSCAN tiene definida la noción de ruido (noise).
- DBSCAN requiere solo de dos parámetros (eps y MinPts) y no es afectado por el orden que tengan los puntos en la base de datos.
- No se puede dejar de mencionar el algoritmo SNN, el cual, al igual que el DBSCAN, es un algoritmo basado en densidad; sin embargo, no calcula la similitud de dos puntos a partir de la distancia entre ellos, sino que lo hace en términos de cuántos elementos vecinos comparten ambos elementos, por lo que se decide no utilizar este algoritmo.

2.2 Funcionalidad de agrupamiento de Google Maps

Google Maps se utilizó en este proyecto para visualizar los incidentes en el mapa de Costa Rica. Se investigó sobre la capacidad de agrupar las marcas que se pintaban en el mapa y se encontró que Google Maps V3 sí cuenta con una funcionalidad para agrupar marcas (imágenes que se colocan en el mapa de Google Maps utilizando las coordenadas geográficas). Sin embargo, esto es solo un efecto visual para disminuir la gran cantidad de marcas simultáneas en el mapa. Se encontró que no existe forma de configurar la distancia

que debería existir entre puntos; además de esto, Google Maps agrupa de manera diferente según el zoom (aumento) que tenga el mapa, así que se descarta la posibilidad de utilizarlo para que aplique algún algoritmo de agrupamiento.

A pesar de que la utilización de esta herramienta como medio para aplicar el algoritmo de agrupamiento no fue funcional para el fin requerido en este proyecto, se documenta la experiencia obtenida en este documento porque esta agrupación que hace Google Maps permite visualizar de una manera ordenada la cantidad de marcas o incidentes registrados por Ruta Alternativa.

La Figura 3.1 y la Figura 3.2 permiten visualizar cómo se aprecian los resultados de agrupar la información de los incidentes de tránsito recolectados por Ruta Alternativa haciendo uso de la funcionalidad de agrupamiento que tiene Google Maps.



Figura 3.1 Incidentes de tránsito primer semestre 2013, Agrupación Google Maps V3

Fuente: Ruta Alternativa

En la Figura 3.1 se pudo apreciar la cantidad de incidentes que han sido registrados por Ruta Alternativa durante el primer semestre del año 2013: son alrededor de ocho mil registros.

Aunque se trata de un agrupamiento solo visual de Google Maps V3, la imagen muestra claramente que la gran mayoría de incidentes se presentan en el área metropolitana (marca color morado con 6774 marcas agrupadas), principalmente en la zona de San José, Cartago y Alajuela.

En la Figura 3.2 podemos apreciar el mismo mapa con la misma cantidad de puntos, solo que con un aumento en el tamaño del mapa. Seguimos viendo que es el área metropolitana la más afectada. Observamos que los *clusters* con mayor cantidad de incidentes (Color Rojo) están en toda el área metropolitana. En la parte sur podemos encontrar estos *clusters* en las cercanías de San Miguel y Desamparados; en el sector norte podemos observar estos *clusters* en Heredia, cerca de Santo Domingo y San Pablo; en el borde este podemos ver las marcas rojas en comunidades como Curridabat, Guadalupe y Moravia; mientras que hacia el límite oeste podemos ver afectada la provincia de Alajuela.

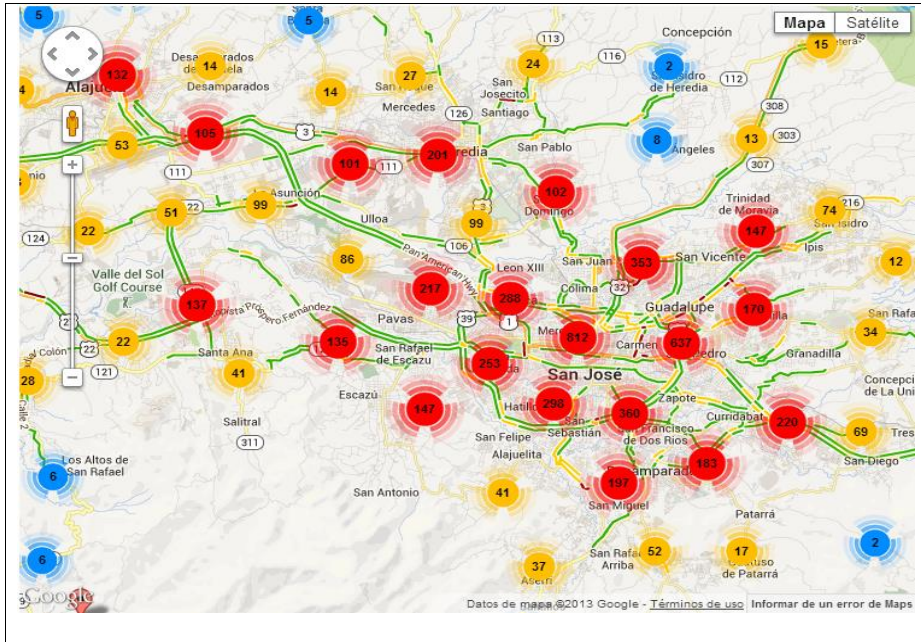


Figura 3.2 Incidentes de tránsito primer semestre 2013, Agrupación Google Maps V3

Fuente: Ruta AlternA

2.3 Funcionalidad de SQL Server Spatial Tools

Esta poderosa herramienta que provee Sql Server a partir de su versión 2008, a pesar de ser una muy buena opción para hacer cálculos con datos geoespaciales como calcular distancias entre puntos geográficos y trazar figuras geométricas en un mapa, no fue utilizada para los resultados finales del proyecto, debido más que todo a los efectos finales de visualización, ya que no permitía apreciar con el detalle que permite hacerlo la herramienta de Google Maps V3. Sin embargo, también se decide documentar brevemente los resultados obtenidos de esta investigación.

En la Figura 3.3 podemos apreciar, aunque no hay un mapa de fondo, que las marcas que producen los incidentes registrados por Ruta AlternA prácticamente forman la silueta de Costa Rica.

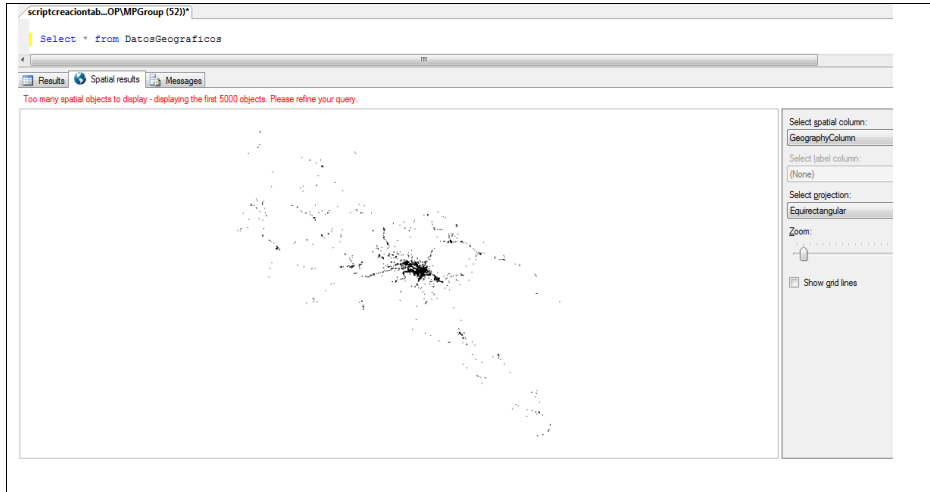


Figura 3.3 Incidentes de tránsito primer semestre 2013, SQL Server Spatial

Fuente: SQL Server Spatial Tool

Además de las limitantes de visualización obvias (detalles de mapa) de esta herramienta, también encontramos una limitante en la cantidad máxima de objetos espaciales a mostrar: solo nos permitió visualizar los primeros cinco mil registros, y se cuentan con más de ocho mil registros recolectados por Ruta Alterna.

Capítulo 3 – Marco Metodológico

3.1 Modalidad de la investigación

La modalidad de investigación a la cual se adapta mejor este proyecto es a la Investigación de Campo, dado que se estudiará una problemática de la sociedad como lo son los incidentes de tránsito en Costa Rica. La información es recopilada directamente de la realidad por Ruta Alterna, lo que proporcionará datos primarios a los cuales mediante la aplicación de algoritmos de agrupación (DBSCAN) permita identificar las ubicaciones geográficas con mayor cantidad de incidentes de tránsito en el país, realizando una interpretación de los resultados obtenidos.

Al ser datos reales no hay espacio para la manipulación de variables sino que se mantienen tal como fueron recopilados originalmente, Ruta Alterna recopilará diversos datos sobre los incidentes de tránsito como lo son: Tipo de Incidente, Fecha, Hora y la Coordenadas Geográficas de la ubicación.

Se entiende por Investigación de Campo, el análisis sistemático de problemas en la realidad, con el propósito bien sea de describirlos, interpretarlos, entender su naturaleza y factores constituyentes, explicar sus causas y efectos, o predecir su ocurrencia, haciendo uso de métodos característicos de cualquiera de los paradigmas o enfoques de investigación conocidos o en desarrollo. Los datos de interés son recogidos en forma directa de la realidad; en este sentido se trata de investigaciones a partir de datos originales o primarios. (UPEL, 2006)

3.2 Tipo de Investigación

Al estar enfocado este proyecto en recopilar, observar, procesar la información y finalmente realizar conclusiones sobre los resultados obtenidos, además de esto manteniendo la premisa de que el proyecto servirá como una herramienta para toma de decisiones, se puede discernir que el tipo de Investigación es Evaluativa.

3.3 Enfoque de la Investigación

El enfoque abordado en esta investigación es cuantitativo, la principal variable a tomar en cuenta es el incidente de tránsito, es evidente que al tratarse de identificar las ubicaciones geográficas con mayor cantidad de incidentes de tránsito se está hablando de variables medibles, es un proceso básicamente de conteo de incidentes por regiones geográficas del país. La siguiente cita describe el enfoque cuantitativo.

utiliza la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecidas previamente y confía en la medición numérica, el conteo y frecuentemente en el uso de la estadística para establecer con exactitud patrones de comportamiento de una población (Hernández et al, 2003).

3.4 Programa de acción

En el programa de acción encontraremos la lista de actividades que deben cumplirse para satisfacer los objetivos de este proyecto, a continuación se mostrarán las diferentes etapas o fases en las que se detallan las actividades del programa.

En este contexto, el programa de acción es un diseño prospectivo que orienta la ejecución de unas acciones intencionadas para cambiar o transformar una realidad diagnosticada. Es un programa de acción porque contiene una serie de elementos o insumos que al ser puestos en marcha deben producir unos resultados relativamente predecibles.

El programa es un acto intencionado o deliberado que debe tener una coherencia interna o sea una lógica de relaciones entre sus componentes. El programa establece el deber ser y simula el orden de los acontecimientos. (Correa, Puerta y Restrepo, 1996)

3.4.1 Fase I – Investigación de Estadísticas Previas

Es requerido iniciar una investigación sobre las estadísticas existentes en Costa Rica sobre el tema de los incidentes de tránsito, para obtener una visión de lo que existe actualmente e identificar alguna carencia que permita impulsar este proyecto, presentando

características innovadoras y valores agregados en la problemática de los incidentes de tránsito.

El proceso de investigación sobre estadísticas se basará en su totalidad en una búsqueda en internet, mediante las diferentes herramientas como lo son los motores de búsqueda web. Se investigarán los Sitios Web de las entidades tanto gubernamentales como no gubernamentales con información de Incidentes de tránsito.

3.4.2 Fase II – Diagrama de Estructura de datos.

Es necesario conocer la base de datos que almacena la información sobre los incidentes de tránsito recopilada por Ruta Alternativa, para ello se debe realizar un diagrama de las tablas junto con sus campos, tipos de dato, llaves primarias, llaves foráneas y además las referencias entre las tablas.

Para esto se utilizará el SQL Server Management Studio, el cual en el apartado de diagramas permite crear un diagrama, seleccionar las tablas de interés y ordenarlas según se necesite dando la facilidad de visualizarlas de una manera gráfica y amigable.

Se iniciará con las tablas de datos Maestros, por ejemplo Tipos de Incidente, Tipos de eventos, entre otras, después se procederá a visualizar las tablas transaccionales como los incidentes y los datos asociados con dichos incidentes como lo son la fecha, hora, dirección, coordenadas geográficas, entre otros.

3.4.3 Fase III – Diseño “Desnormalizado”

Es necesario reorganizar la información que tiene Ruta Alternativa sobre los incidentes de tránsito de una manera más óptima para la aplicación de algoritmos o análisis de los datos, para este objetivo se requiere diseñar una estructura “Desnormalizada”.

Tomando en cuenta el diseño creado en la fase anterior, se contará con las tablas, campos y tipos de datos que tiene la base de datos transaccional de Ruta Alternativa, de esta forma se podrá realizar el diseño compatible en lo que a tipos de datos respecta o en su defecto considerar si se requiere alguna transformación de los datos.

3.4.4 Fase IV – Traslado de los datos

En las dos fases anteriores se consideró tanto conocer la estructura anterior de la base de datos como el crear un diseño de una nueva estructura donde se albergará la información sobre incidentes de tránsito a la cual se le aplicará el algoritmo de agrupamiento. Esta fase contemplará el traslado de la información desde la estructura anterior a la que se diseñó.

Al completarse las dos fases anteriores, se tendrá un panorama claro de cual herramienta o proceso es el más óptimo para realizar el traslado de la información. Ha de tomarse en consideración la herramienta *Microsoft Integration Services* con la cual se puede desarrollar ETL (Extracción, Transformación y Carga por su siglas en Inglés).

3.4.5 Fase V – Desarrollo de aplicación para algoritmo DBSCAN

En diferentes apartados de este documento se mencionan las razones por las cuales se decide desarrollar un software para aplicar el algoritmo de agrupamiento DBSCAN y no utilizar uno ya existente, en esta fase se definirá las diferentes actividades necesarias para el desarrollo de este aplicativo.

- Preparación de ambiente: Es requerido un ambiente de desarrollo, que tenga las herramientas necesarias para el desarrollo de la aplicación, dichas herramientas son: SQL Server 2008 R2, Visual Studio .Net 2010, DevExpress Web UI Control.
- Como paso siguiente está el proceso de codificar o transcribir el algoritmo DBSCAN en el lenguaje C# de .Net para que pueda ser adaptado y utilizado en este proceso.
- Realizar el método encargado de calcular la distancia entre dos ubicaciones a partir de las coordenadas geográficas, esto implica hacer la conversión de Grados (de las coordenadas geográficas) a una unidad de medida longitudinal como metros.
- Desarrollar el proceso que realice la conexión a la base de datos y cargue toda la información sobre incidentes de tránsito y permita cargarla en memoria para el proceso de agrupamiento.

- Desarrollar el proceso que aplique el algoritmo de agrupamiento, este deberá de recorrer uno a uno los datos de incidentes y compararlos con los demás elementos según la distancia entre ellos, determinando así cuales *clusters* se crearán y cuales elementos pertenecen a cada uno de esos *clusters*.
- Desarrollar la interfaz gráfica que permita visualizar la información obtenida por el proceso de agrupamiento, se debe tomar en cuenta que esta interfaz debe ser mediante la utilización de Google Maps API V3, con la finalidad de mostrar los datos obtenidos sobre un mapa de Costa Rica, permitiendo al usuario ubicar geográficamente de una manera gráfica los *Clusters*.

3.4.6 Fase VI – Prueba de Ejecución del aplicativo

Como última fase se considera la ejecución propia del aplicativo desarrollado, esta ejecución debe toda la información suministrada por Ruta Alterna, según los rangos de fecha establecidos para la muestra de data utilizada, con la finalidad de plasmar los resultados en este documento.

Para esta fase se considerará un máximo de diez *Clusters*, se tomarán los que mayor cantidad de incidentes tengan, y se registrará la ubicación exacta, al ser utilizados mapas se puede fácilmente tomar imágenes que permitan identificar estos puntos y de esta manera poder realizar las conclusiones para este proyecto.

3.5 Unidades de Observación

“Las unidades de observación son las variables, categorías, asuntos o componentes que deben ser medidos.” (Correa, Puerta y Restrepo, 1996)

Teniendo en consideración los aspectos mencionados en la cita anterior se listarán las diferentes variables importantes para este proyecto. Se debe tener en consideración que existe una limitación en cuanto a la información recopilada por Ruta Alterna, es decir si bien es cierto que sería de gran utilidad contar con datos como por ejemplo cantidad de decesos en los incidentes, un dato que ayudaría mucho en temas de estadísticas, es información que no siempre puede ser recopilada, de manera que junto con el personal de

Ruta Alternativa, se deja claro cuáles serán los datos a recopilar y por ende poder determinar cuáles serán los datos a utilizar en este proyecto.

3.5.1 Variables

La variable principal o el tema central de esta investigación es el Incidente de tránsito, a partir de esta variable se captura la demás información complementaria, como lo es la Dirección, Fecha, Hora, Tipo de Incidentes, Coordenadas Geográficas entre otros.

- Incidente: Es la variable principal, todas las demás variables describirán o complementarán esta variable.
- Fecha y Hora: Esta variable nos permita ubicar el Incidente en el tiempo
- Variables de Ubicación (Provincia, Cantón, Distrito): Estas variables definen la ubicación del incidente, organizándolo por Provincia, Cantón y Distrito.
- Dirección: Adicional a los datos de Provincia, Cantón y Distrito, se agrega una dirección, que es un simple texto digitado por el funcionario de Ruta Alternativa, en la que se anotan puntos de referencia, puntos cardinales y cualquier otra información que permita ubicar más precisamente el Incidente.
- Coordenadas Geográficas: A pesar de tener datos de ubicación bastante detallados con las variables anteriores, para efecto de una ubicación aún más exacta se agregan las coordenadas geográficas, tanto la Latitud como la Longitud.

3.5.2 Indicadores

El comentario citado a continuación nos deja una idea clara de lo que es un indicador, en este apartado se indicarán cuáles son los indicadores a utilizar en este proyecto.

Los indicadores son medidas específicas (explícitas y objetivamente verificables de los cambios o resultados de una actividad). Los indicadores sirven de patrón para medir, evaluar o mostrar el progreso de una actividad respecto a las medidas

establecidas, en cuanto a la entrega de sus insumos (indicadores de insumo), la obtención de los productos (indicadores de producto) y el logro de sus objetivos (indicadores de efecto e impacto) (Quintero y Velásquez, 1993).

A continuación se listan los indicadores utilizados y mostrados en este proyecto:

- Ubicación geográfica de los *Clusters* con mayor cantidad de incidentes: Es el principal indicador de este proyecto, nos permite visualizar en que zonas geográficas del país se concentra la mayor cantidad de incidentes, por ende permite identificar cuáles son las zonas más conflictivas y peligrosas del país en materia de incidentes de tránsito.
- Incidentes por Provincia: Permite realizar un conteo de la cantidad de incidentes que se presentan por provincia, permitiendo clasificar cuál de ellas es la que cuenta con la mayor cantidad de incidentes.
- Incidentes por Día de Semana: Con este indicador podremos medir en cuál día de la semana suceden más incidentes de tránsito.
- Incidentes por Hora del Día: Este indicador nos muestra en que franja horaria se da la mayor cantidad de incidentes en el país.
- Incidentes por Cantón (San José): Al igual que los demás indicadores, este nos permite medir la cantidad de incidentes, solo que para este en específico se medirán agruparán por cantón, es importante mencionar que no se tomarán en cuenta todos los cantones sino que solamente los de la provincia de San José, la cual presenta una notable mayoría en la cantidad de incidentes.

3.6 Población y Muestra

3.6.1 Población

Con respecto al tema de la población debemos clasificarla como una población Infinita ya que este proyecto abarca la población toda Costa Rica

3.6.2 Muestra

Refiriéndose a la muestra, es un tema algo complejo, ya que no hay forma de seleccionar solo un porcentaje de la población, ya que para este proyecto se trata los incidentes de tránsito a nivel nacional, sin embargo Ruta Alterna tiene claro que no se logra abarcar el cien por ciento de los incidentes a nivel nacional.

Ruta Alterna estima que se maneja entre un 80% y 85% de los incidentes sucedidos a nivel nacional, de manera que estimaremos esto como la muestra.

3.7 Técnicas de análisis de los datos

En la sección de análisis y procesamiento de la información, se hará referencia a como se va a procesar la información recolectada por Ruta Alterna, además de esto se dictarán los pasos a seguir para mostrar los resultados obtenidos luego de procesada la información.

3.7.1 Procesando la información

El primer paso por tener la información estructurada y cargada en el orden requerido es aplicar el algoritmo DBSCAN, el cual será aplicado por el aplicativo desarrollado en este proyecto, los datos a mostrar son: El identificador del incidente, la descripción o tipo de incidente, Provincia, Cantón, Distrito, Dirección y las Coordenadas geográficas. Al aplicar el algoritmo se agregará una columna más en la cual se agregará el Identificador del *Cluster* que el algoritmo le asignó.

3.7.2 Preparando los resultados obtenidos

Por haber aplicado el algoritmo y saber en cuál Cluster está ubicado cada Incidente, el paso siguiente será pasar a la visualización, se debe escoger un máximo de diez Clusters y mediante el identificador asignado colocarlos en la pantalla de visualización, la cual según el orden en que son digitados le asignará un color diferente a cada uno y los mostrará en el mapa de Costa Rica utilizando la tecnología de Google Maps.

- Tomar imagen de cada uno de los diez *Clusters* seleccionados, en cada uno se debe describir la ubicación geográfica de este y una breve explicación del tipo de carretera, es decir, si es un cruce, una rotonda, un paso a desnivel u otro.
- Tomar los registros resultantes y mediante una función de Microsoft Excel agruparlos por provincia, crear gráfica.
- Crear una gráfica en Microsoft Excel con la información resultante, haciendo agrupación por día de la semana.
- Crear un gráfico en Microsoft Excel agrupada por hora del día, esto determinará cual es la hora en la que suceden más incidentes.
- Utilizando la información resultante diseñar un gráfico que muestre la cantidad de incidentes por Cantón (Solamente San José).

3.8 Cronograma

Fase	Actividad	Semana														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
I	Investigación de Estadísticas Previas	■														
II	Diagrama de Estructura de datos		■													
III	Diseño "Des normalizado"			■												
IV	Traslado de los datos				■											
V	Preparación Ambiente					■	■									
	Código DBSCAN						■	■								
	Cálculo Distancias							■	■							
	Conexión y Carga BD								■	■						
	Procedimiento Aplica Algoritmo									■	■					
	Interfaz gráfica										■	■	■	■		
VI	Piloto ejecución aplicativo													■	■	■

Capítulo 4 – Propuesta de Solución

4.1 Plan de Inversión

Antes de realizar cualquier proyecto, es importante tener presente cuantificar la inversión económica que este va a representar. Debido a esto, se coloca como punto inicial la propuesta de la solución del Plan de Inversión.

Recurso Material:	
Recurso	Costo
Computador Diseño y Desarrollo	\$ 600.00
Computador Desarrollo aplicativo	\$ 600.00
Computador Mantenimiento Datos	\$ 450.00
Datacard (Utilización 6 meses)	\$ 162.72
Dispositivo Móvil (GPS)	\$ 338.00
TOTAL Recursos Materiales	\$ 2,150.72
Recurso Humano:	
Recurso	Costo
Analista/Desarrollador (\$16x8Hx15Sem)	\$ 1,920.00
Diseñador DB (\$30x8Hx6Sem)	\$ 1,400.00
Programador (\$16x8Hx5Sem)	\$ 640.00
Experto Datos Ruta Alternativa (\$25x20Hx25Sem)	\$ 12,500.00
TOTAL Recurso Humano	\$ 16,460.00
TOTAL	\$18,610.72

4.2 Análisis del Software de Recolección de datos

En este proceso se muestra la estructura de la base de datos que almacena la información recolectada por Ruta Alternativa, así como el aplicativo con que se registra.

4.2.1 Diseño de la Base de Datos

Para este apartado se mostrará el diseño de la base de datos. En primera instancia, se mostrarán las tablas con datos maestros como son las provincias, cantones, distritos, entre

otros, y posteriormente, las tablas transaccionales que almacenan la información como tal sobre los incidentes.

La tabla 4.1 muestra el listado de las tablas de la base de datos correspondiente a los maestros o catálogos.

Tabla	Descripción
Provincia	Almacena el catálogo de provincias de Costa Rica.
Cantón	Almacena el catálogo de los cantones de Costa Rica.
Distrito	Almacena el catálogo de Distritos Electorales (no los administrativos) de cada cantón de Costa Rica.
Rutas_Cat	Contiene las rutas del País, con la numeración oficial que tienen estas rutas.
Rutas_Segmento_Cat	Almacena segmentos de ruta, nombres de lugares estratégicos los que pasa la ruta y el kilómetro en el cual se encuentra dicho lugar.
Eventos_Cat	En esta tabla se encuentran los diferentes eventos a registrar, por ejemplo: Colisión, Atropello, Vuelco, entre otros.
Estados_Cat	Nos indica el estado del lugar en que sucedió el incidente (Paso normal, Paso Cerrado, Paso Regulado, entre otros).
Motivos_Cat	En esta tabla se almacena el motivo del incidente (si se tiene, no siempre se logra obtener este dato.), por ejemplo: Falso adelantamiento, exceso de velocidad, entre otros.

Tabla 4.1 Listado de tablas Maestras o Catálogos de la base de datos de Incidentes

Fuente: Ruta Alterna

En la Figura 4.1 se puede apreciar el diagrama de las tablas de datos maestros o catálogos correspondientes a la base de datos de incidentes de Ruta Alternativa.

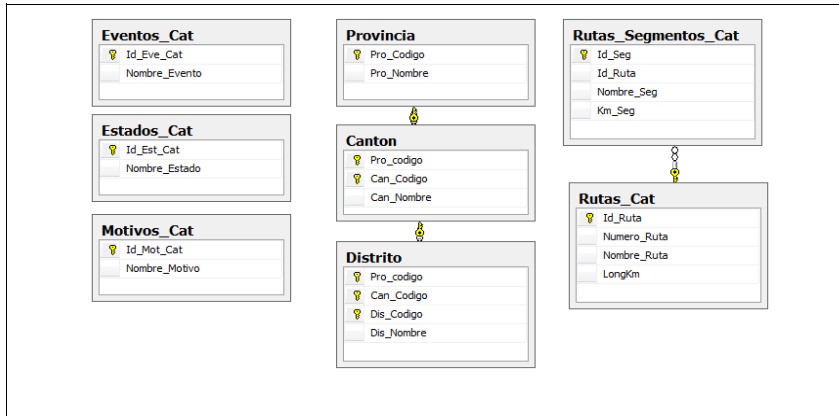


Figura 4.1 Diagrama de las tablas de catálogos de la base de datos de Incidentes

Fuente: Ruta Alternativa

La Tabla 4.2 muestra el listado de tablas transaccionales de la base de datos de registro de incidentes de Ruta Alternativa.

Tabla	Descripción
Incidentes	Almacena el listado de incidentes
Incidente_x_ubicación	Almacena la dirección del incidente, compuesta de Provincia, Cantón, Distrito y un texto de la dirección.
Incidente_x_Ruta	Almacena la ruta en la que sucedió el incidente (si aplica, no en todos los casos se obtiene este dato).
Incidente_x_Motivos	Almacena el motivo por el cual se dio el incidente (si se tiene el dato, no todo el tiempo se obtiene).
Incidentes_x_Fallecidos	Permite almacenar la información de la cantidad de fallecidos que tuvo el incidente si es que los hay (no todo el tiempo se obtiene este dato).
Incidentes_x_Coordenadas	Esta tabla contiene las coordenadas geográficas de la ubicación del incidente, obtenidas mediante Google Maps.

Tabla 4.2 Listado de tablas Transaccionales de la base de datos de Incidentes

Fuente: Ruta Alternativa

La Figura 4.2 muestra el diagrama de las tablas transaccionales de la base de datos de incidentes de Ruta Alternativa.

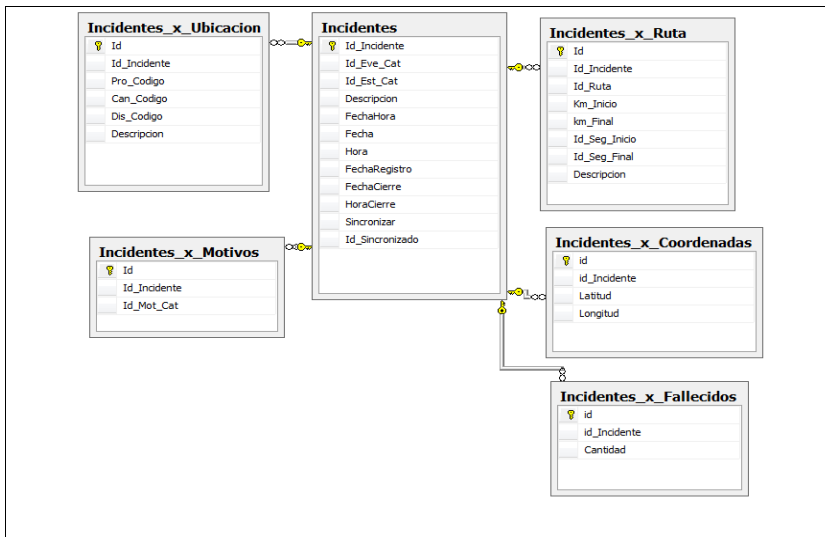


Figura 4.2 Diagrama de tablas transaccionales de la base de datos de incidentes

Fuente: Ruta Alternativa

4.2.2 Software de Registro de Incidentes

En este apartado se muestra el software utilizado por Ruta Alternativa, para realizar el registro de los incidentes de tránsito. A continuación se detallan las pantallas del software.

Pantalla Principal: En esta pantalla se agrega toda la información relacionada con el incidente como tal. Por ejemplo: Tipo de Incidente (Evento), Estado, Fecha, Hora, Dirección, Coordenadas Geográficas, entre otros. En la Figura 4.3 se detalla la pantalla principal del software de registro de incidentes.

The screenshot shows a web-based application window titled "Control Incidentes". The interface is divided into several sections. At the top, there is a menu bar with "Administración", "Reporte", and "Histórico y Estadísticas". Below this, the "Incidentes" section is active, showing a form for incident registration. The form includes fields for "Evento" (set to "Colisión"), "Estado" (set to "Abierto"), "Fecha Hora Incidente" (Date: 14/07/2013, Hora: 03:42:51 p.m.), and "Fecha Hora Cierre" (Date: 14/07/2013, Hora: 03:42:51 p.m.). There is also a checkbox labeled "Habilitado Para Sincronizar" which is checked. Below these fields are "Nuevo" and "Guardar" buttons. The "Ubicación" section is also visible, with tabs for "Coordenadas Geográficas", "Rutas", and "Otros Datos". It contains dropdown menus for "Provincia" (SAN JOSE), "Cantón" (CENTRAL), and "Distrito" (HOSPITAL), along with "Dirección" and "Dirección Corta" fields.

Figura 4.3 Pantalla principal del software de registro de incidentes

Fuente: Ruta Alternativa

Pantalla de Reporte: Esta pantalla muestra los incidentes del día. Por defecto se carga con la fecha del día actual y se refresca conforme se van ingresando los incidentes. Muestra en un Grid de datos la información resumida como el id del Incidente, provincia, cantón, distrito, tipo de incidentes, entre otros. En la Figura 4.4 se puede apreciar la pantalla de Reporte.

IdIncidente	Provincia	Canton	Distrito	Direccion	Tipo	Estado	Fecha	Hora	FechaCierre	HoraCierre	Sincronizar	Id_Sincroniz
10547	GUANAC...	TILARAN	TILARAN	PROYEC...	Vuelco	Abierto	14/07/2013	03:00:02			<input checked="" type="checkbox"/>	0
10548	ALAJUELA	GRECIA	SAN JOSE	SANTA G...	Vehículo s...	Abierto	14/07/2013	06:30:02			<input checked="" type="checkbox"/>	0
10549	ALAJUELA	CENTRAL	SAN RAF...	DEL PUE...	Colisión	Abierto	14/07/2013	11:30:02			<input checked="" type="checkbox"/>	0
10550	SAN JOSE	CURRIDA...	CURRID...	DETRAS ...	Colisión	Abierto	14/07/2013	11:30:02			<input checked="" type="checkbox"/>	0
10551	LIMON	SIQUIRRES	ALEGRIA	FRENTE ...	Colisión	Abierto	14/07/2013	12:00:02			<input checked="" type="checkbox"/>	0
10552	HEREDIA	SANTA B...	SAN JUAN	CINCO ES...	Colisión	Abierto	14/07/2013	12:00:02			<input checked="" type="checkbox"/>	0
10553	ALAJUELA	GRECIA	GRECIA	300 MET...	Colisión	Abierto	14/07/2013	12:00:02			<input checked="" type="checkbox"/>	0
10554	SAN JOSE	CENTRAL	HOSPITAL	BARRIO ...	Colisión	Abierto	14/07/2013	12:00:02			<input checked="" type="checkbox"/>	0
10555	ALAJUELA	GRECIA	PUENTE ...	INICIA RE...	Otros	Abierto	14/07/2013	14:00:00			<input checked="" type="checkbox"/>	0
10556	CARTAGO	CENTRAL	SAN NIC...	INICIA RE...	Otros	Abierto	14/07/2013	14:00:00			<input checked="" type="checkbox"/>	0
10557	HEREDIA	CENTRAL	HEREDIA	INICIA RE...	Otros	Abierto	14/07/2013	14:00:00			<input checked="" type="checkbox"/>	0
10558	SAN JOSE	VAZQUE...	JESUS O ...	RUTA 32 ...	Otros	Abierto	14/07/2013	14:00:00			<input checked="" type="checkbox"/>	0
10559	ALAJUELA	OROTINA	COYOLAR	RUTA 27 ...	Otros	Abierto	14/07/2013	14:00:00			<input checked="" type="checkbox"/>	0

Figura 4.4 Pantalla reporte actual de software de registro de incidentes

Fuente: Ruta Alternativa

Pantalla de Histórico y Estadísticas: Finalmente, se muestra la pantalla de Histórico y Estadísticas, en la que se cargan datos similares a la pantalla de Reporte Actual, con la salvedad de que aquí sí se puede seleccionar un rango de fechas (Inicio y Final) para ver un reporte en un determinado lapso de tiempo; además de esto presenta un pequeño Grid de datos con un conteo de los incidentes por tipo de incidente que se presenta en el rango de fechas seleccionadas. En la Figura 4.5 se aprecia la pantalla de Histórico y Estadística.

IdIncidente	Provincia	Canton	Distrito	Direccion
9633	SAN JOSE	MONTES DE OCA	SAN PEDRO	FRENTE A LA FARMACIA LA BOMBA
9634	ALAJUELA	NARANJO	SAN MIGUEL	EN EL CRUCE
9635	CARTAGO	LA UNION	SAN RAFAEL	AUTOPISTA FLORENCIO DEL CASTILLO DEL CRUCE DE SAN RAFAEL 800
9636	SAN JOSE	PEREZ ZELEDON	BARU	PLATANILLO EN EL HOTEL NATURAL
9637	GUANACASTE	LIBERIA	LIBERIA	FRENTE A LA PANADERIA SANCHEZ
9638	SAN JOSE	CENTRAL	HOSPITAL	BARRIO CORAZON DE JESUS EN LA ENTRADA A YAMBER
9639	SAN JOSE	PURISCAL	MERCEDES SUR	SALITRALES FRENTE AL SALON BELLA VISTA
9641	LIMON	CENTRAL	LIMON	RUTA 32 BARRIO LA COLINA RETEN POLICIAL Y BALACERA NO HAY PASC
9642	SAN JOSE	CENTRAL	SAN SEBASTIAN	CIRCUNVALACION EN LA ROTONDA DE PASO ANCHO
9643	CARTAGO	OREAMUNO	SAN RAFAEL	EL BOSQUE RESIDENCIAL VILLA LAURA AL COSTADO ESTE DE LA SUBES
9644	SAN JOSE	DESAMPARADOS	DESAMPARADOS	FRENTE A LA BOMBA GLOBAL
9645	SAN JOSE	MONTES DE OCA	SABANILLA	DE UNED UN KILOMETRO ESTE

Figura 4.5 Pantalla de Reporte Histórico y Estadísticas, software de registro de incidentes

Fuente: Ruta Alternativa

4.3 Tabla des-normalizada

Esta tabla será a la cual se trasladará el historial de la información recolectada de incidentes de tránsito en el periodo establecido, y estará de manera des-normalizada para efecto de aplicar el algoritmo de agrupamiento o *clustering*.

La Figura 4.6 muestra la estructura que se diseña para dicha tabla. Se pueden apreciar, tanto los nombres de los campos, como el tipo de dato que alberga.

DatosGeograficos	
Column Name	Condensed Type
Id	bigint
Incidente	varchar(200)
provincia	varchar(100)
Canton	varchar(100)
Distrito	varchar(100)
Direccion	varchar(MAX)
FechaHora	datetime
Latitud	varchar(MAX)
Longitud	varchar(MAX)
GeographyColumn	geography
DiaSemana	varchar(50)
HoraDia	int

Figura 4.6 Diseño Tabla des normalizada datos de Incidentes de tránsito

Fuente propia

4.4 Elección del Algoritmo de Agrupamiento

Existen diversos algoritmos para agrupamiento, algunos de ellos son basados en densidad. Sin embargo, con una breve investigación se puede determinar que el algoritmo que encaja mejor en este proyecto es el DBSCAN (Density Based Spatial Clustering of Applications with Noise).

El algoritmo DBSCAN puede identificar clusters en grandes conjuntos de datos espaciales buscando en la densidad de los elementos, usando tan solo un parámetro. Además, el usuario obtiene una sugerencia de cual valor sería adecuado para el parámetro. Por lo tanto se requiere un mínimo conocimiento del dominio. El

DBSCAN puede también determinar cuál información debe ser clasificada como ruido o valores atípicos. A pesar de esto el proceso de trabajo es rápido y escala muy bien con el tamaño de la base de datos. (Bäcklund, Hedblom & Neijman, 2011.)

Con formato: Español (Costa Rica)

4.5 Elección de la herramienta

Inicialmente se había planteado realizar este proyecto utilizando el software Weka, el cual es un software que permite aplicar algoritmos de Minería de Datos, como lo es el algoritmo de agrupamiento DBSCAN. Sin embargo, con el ideal de presentar los *clusters* en un mapa que permita al usuario ubicar geográficamente los puntos o incidentes, se realiza una búsqueda de otras herramientas que permitan lograr la meta de este proyecto.

A pesar de que con las herramientas Google Maps y *SQL Server Spatial Tools*, no se logra la meta propuesta, estas se mencionan como parte de la investigación y el marco teórico de este proyecto. También se explica por qué se descartaron como las herramientas para aplicar el algoritmo a la información sobre los incidentes recolectados por Ruta Alterna.

4.6 Desarrollo de Herramienta que aplica el algoritmo de agrupamiento

A pesar de que ya algunas herramientas de minería de datos soportan el algoritmo DBSCAN, incluyendo la herramienta sugerida en el anteproyecto “Weka”, se deseaba visualizar, tanto los puntos de los incidentes, como los *clusters* sobre un mapa de Costa Rica, y no simplemente sobre la pantalla de exploración de Weka, de manera que se decidió desarrollar un software que aplique el algoritmo de agrupamiento y a la vez utilice las bondades de visualización de Google Maps para mostrar los *clusters* con sus elementos.

La aplicación cuenta con tres pantallas, Data, Clusters y Visualización, las cuales se explican a continuación:

1. **Pantalla Datos (Data):** Muestra la lista de registros con la información de cada incidente con la estructura tal como se diseñó en la tabla des-normalizada.

La Figura 4.7 muestra esta pantalla. Se puede observar un Grid de datos con toda la información que se migró desde la base de datos de Ruta Alternativa a la tabla des-normalizada.

DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE) [Log In]									
TÉCNICAS DE AGRUPAMIENTO SOBRE INFORMACIÓN DE INCIDENTES MÁS COMUNES OCURRIDOS EN LAS CARRETERAS DE COSTA RICA.									
Data Clusters Visualización									
Arrastre una columna aquí para agrupar por dicha columna									
Incidente	provincia	Cantón	Distrito	Dirección	Fecha	Día	Hora	Lat	Lon
Fuego En Vehículo	ALAJUELA	CENTRAL	RIO SEGUNDO	EN LA ENTRADA A EL BARRIO EL CARMEN FRENTE AL HOTEL CASA ANTIGUA	25/02/2013 06:00:00 a.m.	Monday	6	10.0021	-84.18988466262817
Atrapelo	ALAJUELA	CENTRAL	SAN ANTONIO DEL TEJAR	VILLA BONITA EN EL PUENTE SOBRE LA PISTA	24/02/2013 05:30:00 p.m.	Sunday	17	9.99931	-84.21977519989013
Colisión	ALAJUELA	CENTRAL	SAN ANTONIO DEL TEJAR	VILLA BONITA EN EL CRUCE DE SENAFOROS	25/02/2013 02:00:00 p.m.	Monday	14	10.0031	-84.21898126602173
Árbol cae en la vía	ALAJUELA	OROTINA	OROTINA	DEL COLEGIO SANTA FE 150 METROS OESTE	24/02/2013 10:00:00 a.m.	Sunday	10	9.91444	-84.53307867050171
Colisión	ALAJUELA	CENTRAL	RIO SEGUNDO	EN LA ENTRADA A LA CHANCERA	25/02/2013 06:00:00 a.m.	Monday	6	10.0021	-84.18986320495605
Colisión	SAN JOSE	ESCAZU	SAN RAFAEL	UN KILOMETRO SUR DE LA ENTRADA A BELLO HORIZONTE	24/02/2013 09:00:00 p.m.	Sunday	21	9.91616	-84.12742137908935
Colisión	SAN JOSE	ASERRI	ASERRI	DE LA FABRICA JOKEY 200 METROS SUR	24/02/2013 03:00:00 a.m.	Sunday	3	9.86696	-84.09478425979614

Figura 4.7 Diseño Pantalla de Datos, Software para aplicar algoritmo de Agrupamiento

Fuente propia

2. **Pantalla Algoritmo (Clusters):** Es la pantalla que me permite elegir los parámetros del algoritmo (eps y MinPts) y a su vez aplicar el algoritmo a los registros. La Figura 4.8 muestra la pantalla mencionada. En ella se puede apreciar los espacios para digitar Distancia mínima entre los elementos (eps) y la cantidad mínima de elementos que se permite tener en un *cluster*.

DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE)

TÉCNICAS DE AGRUPAMIENTO SOBRE INFORMACIÓN DE INCIDENTES MÁS COMUNES OCURRIDOS EN LAS CARRETERAS DE COSTA RICA.

Data Clusters Visualización

Distancia (km) Cantidad Mínima Puntos

Cluster

Incidente	provincia	Cantón	Distrito	Dirección	Fecha	Día	Hora	Lat
Cluster: 45 (Count=215)								
Cluster: 28 (Count=180)								
Cluster: 548 (Count=76)								
Cluster: 92 (Count=71)								
Cluster: 57 (Count=61)								
Cluster: 42 (Count=59) (Continuar a la página siguiente)								
Colisión	SAN JOSE	MONTES DE OCA	SAN PEDRO	FRENTE AL MALL SAN PEDRO	12/07/2013 12:30:00 p.m.	Friday	12	9.93275
Colisión	SAN JOSE	MONTES DE OCA	SAN PEDRO	CIRCUVALACION EN LA ROTONDA DE LA FUENTE DE LA HISPANIDAD	05/07/2013 07:30:00 p.m.	Friday	19	9.93287
Colisión	SAN JOSE	MONTES DE OCA	SAN PEDRO	EN LA ROTONDA DE LA FUENTE DE LA HISPANIDAD	10/07/2013 12:30:00 p.m.	Wednesday	12	9.93292

Figura 4.8 Diseño Pantalla Algoritmo (Clusters), Software para aplicar algoritmo de Agrupamiento.

Fuente propia

3. **Pantalla de Visualización:** Permite visualizar en un mapa de Google Maps los *clusters* con sus elementos, y se asignan colores distintos para cada *cluster*. Actualmente solo permite visualizar un máximo de diez *clusters* a la vez, ya que se considera que más de esta cantidad no permitiría analizar y entender correctamente los datos.

La Figura 4.9 muestra una imagen de cómo se aprecia la información en la pantalla de Visualización del software, cada *cluster* se marca con un color diferente.

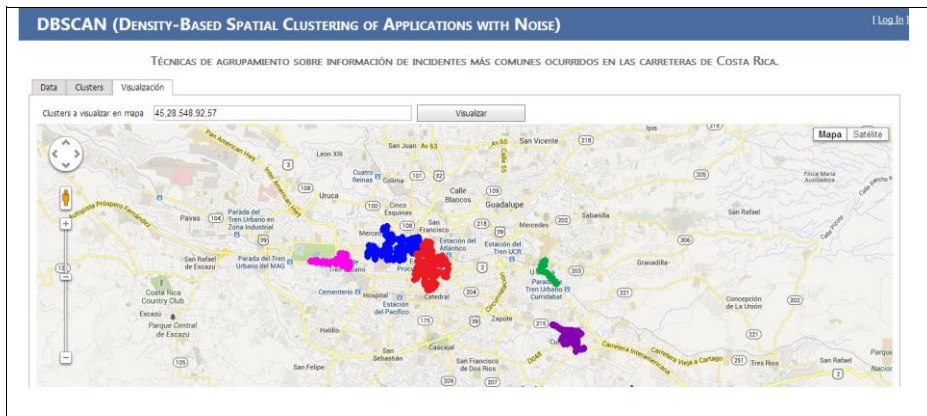


Figura 4.9 Diseño Pantalla Visualización, Software para aplicar algoritmo de Agrupamiento, mostrando cada cluster en colores diferentes.

Fuente propia

4.7 Aplicando el algoritmo DBSCAN sobre la información recolectada

Anteriormente se mostró la herramienta que se desarrolló en este proyecto para aplicar el algoritmo de agrupamiento a la información recolectada sobre incidentes de tránsito. En este apartado se mostrará cómo se utilizará dicha herramienta y los resultados que arroje para poder analizar y tomar conclusiones a partir de dichos resultados.

Para la elección de los valores que se le asignarán a los parámetros requeridos por el algoritmo DBSCAN, se determinó que la distancia más aceptable (eps) para aplicar el algoritmo es entre cincuenta (50) y cien (100) metros, ya que una distancia mayor produce *clusters* muy grandes, situación que dificultaría el análisis. Se elige una cantidad mínima de tres (3) puntos por *cluster*, debido a que se considera que una cantidad menor que esta sería ruido (Noise).

Iniciamos aplicando el algoritmo con una distancia (eps) de cincuenta (50) metros y un mínimo de tres (3) puntos por *cluster*. Luego de esto, analizaremos los primeros diez *clusters* que nos arroje el resultado.

La Figura 4.10 muestra los datos ya agrupados con un respectivo número de *cluster* asignado a los puntos. Para este ejemplo podemos ver que el primer *cluster* es el número 289, el cual contiene una cantidad de 51 puntos pertenecientes a él.

Cluster	Count
Cluster: 289	51
Cluster: 28	43
Cluster: 188	42
Cluster: 18	36
Cluster: 273	32
Cluster: 175	30
Cluster: 96	28
Cluster: 437	26
Cluster: 78	25
Cluster: 127	25
Cluster: 23	23
Cluster: 636	22
Cluster: 76	21
Cluster: 70	20
Cluster: 131	18
Cluster: 12	17
Cluster: 45	17
Cluster: 498	17
Cluster: 591	17
Cluster: 54	16

Figura 4.10 Pantalla Clusters, Software para aplicar algoritmo de Agrupamiento, mostrando el listado de cluster resultante al aplicar algoritmo.

Fuente propia

Pasamos a la pantalla de visualización y elegimos los primeros diez *clusters* para que se muestren en el mapa. La visualización se encuentra en un nivel de zoom bastante bajo. Solamente se observan pequeñas marcas de los respectivos colores correspondientes a cada *cluster*.

Debemos tener presente que Google Maps permite ampliar el nivel de zoom para poder observar en más detalle la ubicación de un elemento, incluso su información específica (Provincia, Cantón, Distrito, Dirección, entre otros), colocando el cursor del mouse sobre el elemento.

La Figura 4.11 permite visualizar en el mapa de Google Maps los *clusters* resultantes del proceso de aplicar el algoritmo DBSCAN sobre la información de incidentes de tránsito.

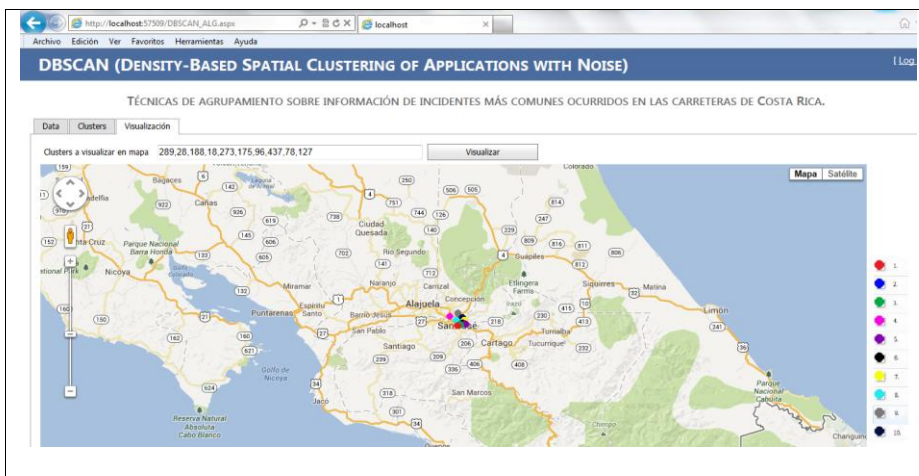


Figura 4.11 Pantalla Visualización, Software para aplicar algoritmo de Agrupamiento, mostrando cada cluster en un color diferente.

Fuente propia

Se incrementa el nivel del zoom del mapa a un nivel que nos permita visualizar los diez *clusters* en el mismo cuadro. La figura 4.12 muestra una imagen editada para agregar el orden de los *clusters* según su color. Se asigna el rojo al que mayor cantidad de incidentes tiene y el azul oscuro al que menos incidentes tiene de los *clusters* seleccionados.



Figura 4.12 Pantalla Visualización, Software para aplicar algoritmo de Agrupamiento, mostrando cada cluster en un color diferente junto con su número en la posición de la lista empezando con “1” el cual tiene mayor cantidad de incidentes.

Fuente propia

Se hará un análisis de cada uno de los diez *clusters*, incrementando el nivel máximo de zoom del mapa de Google que nos permite ver todos los puntos de cada *cluster*, e identificaremos cuál es la ubicación geográfica de este *cluster* en el mapa de Costa Rica.

4.7.1 Cluster No. 1

En la Figura 4.13 se puede apreciar el primer *cluster* al cual se le asignó la marca de color rojo. Está ubicado en la Rotonda de la Guacamaya sobre la Circunvalación en San José (San Sebastián). Cuenta con cincuenta y un (51) elementos.



*Figura 4.13 Pantalla Visualización, Cluster número 1 (eps = 50 metros)
Software para aplicar algoritmo de Agrupamiento.
Fuente propia*

4.7.2 Cluster No. 2

El segundo cluster se ubica en la Rotonda de la Fuente de la Hispanidad ubicada en San Pedro de Montes de Oca, San José. Este cuenta con cuarenta y tres (43) elementos. Figura 4.14 nos muestra este *cluster*, al cual se le asignó la marca de color azul.

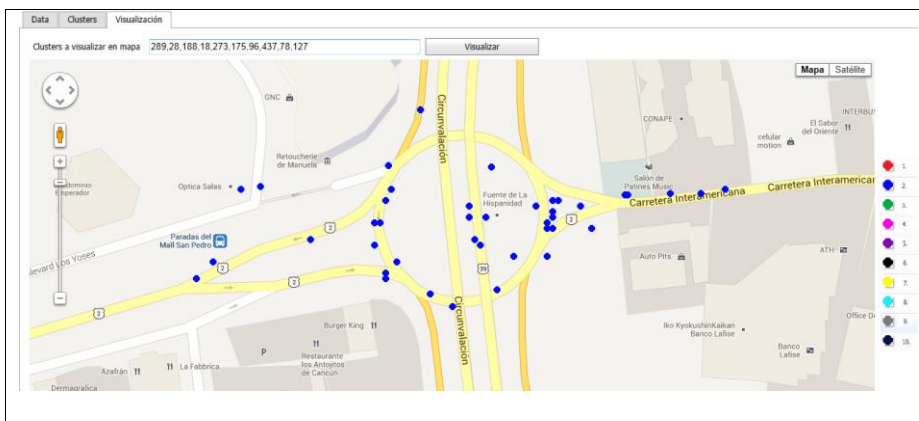


Figura 4.14 Pantalla Visualización, Cluster número 2 ($eps = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.3 Cluster No. 3

En la Rotonda de la I Griega, ubicada en San Francisco de Dos Ríos, San José, tiene lugar el tercer *cluster* resultante del algoritmo DBSCAN. A este se le asignó la marca de color verde y cuenta con cuarenta y dos (42) puntos.

En la Figura 4.15 se puede apreciar la imagen correspondiente al *cluster* número tres.

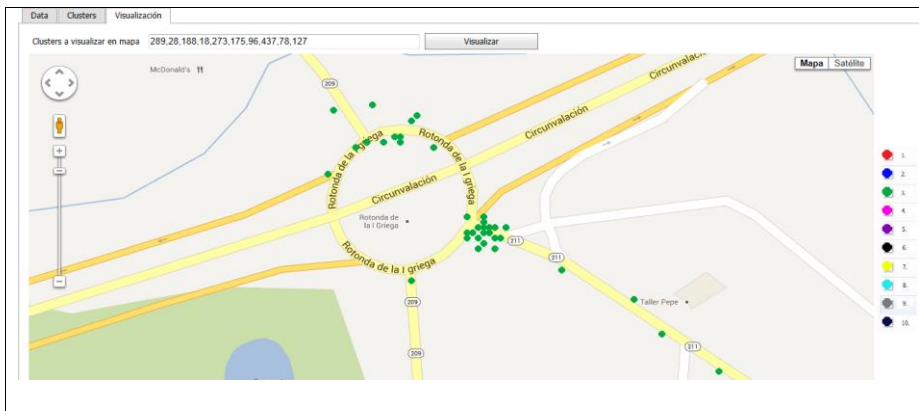


Figura 4.15 Pantalla Visualización, Cluster número 3 ($eps = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.4 Cluster No. 4

Se continúa mostrando el cuarto *cluster*, el cual está ubicado en el Puente Juan Pablo II, La Uruca, San José. Como podemos observar en la imagen 2.9, este lugar también es una rotonda, aunque cuenta con puente o paso a desnivel. A este *cluster* se le asigna la marca de color rosado y cuenta con treinta y seis (36) puntos.

La Figura 4.16 muestra la imagen del *cluster* número cuatro.

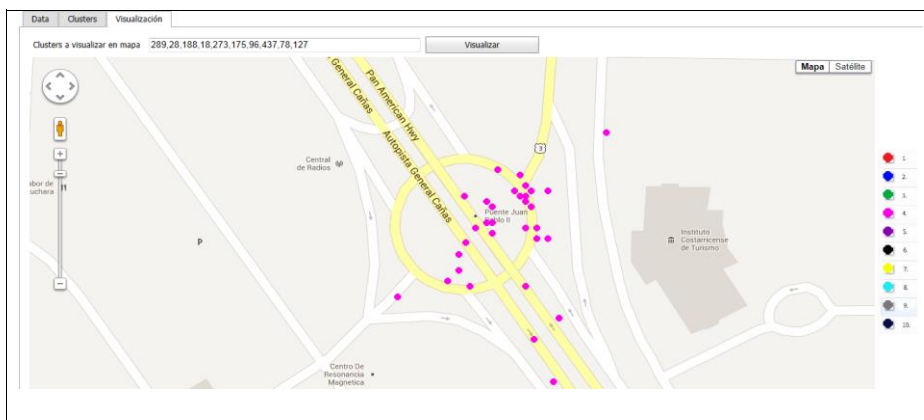


Figura 4.16 Pantalla Visualización, Cluster número 4 ($eps = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.5 Cluster No. 5

En el quinto *cluster* podemos observar que, a diferencia de los anteriores cuatro, no es una rotonda, sino más bien un pequeño segmento de aproximadamente 400 metros, prácticamente en línea recta. Está ubicado en Curridabat, San José, y va desde la parte frontal del Registro Nacional hasta la entrada principal del centro comercial Multiplaza. Este *cluster* se identificó con la marca de color morado y tiene una cantidad de treinta y dos (32) puntos. Se aprecia en la Figura 4.17.

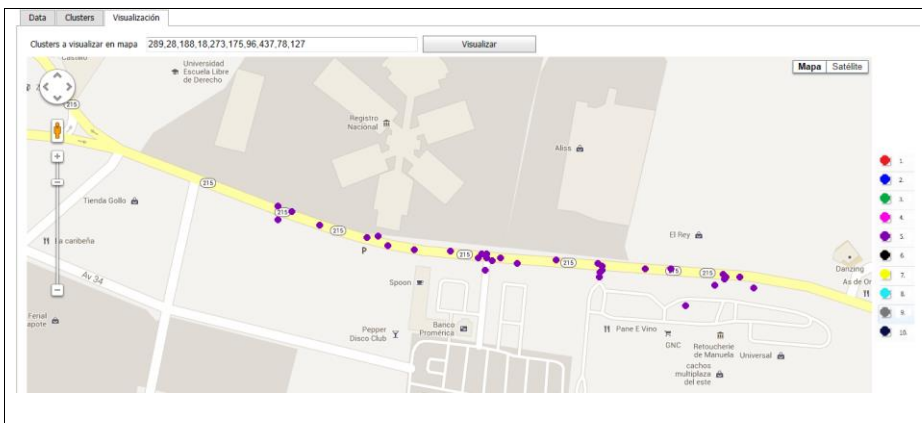


Figura 4.17 Pantalla Visualización, Cluster número 5 ($\text{eps} = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.6 Cluster No. 6

La Figura 4.18 muestra el quinto *cluster* identificado con la marca de color negro. Tiene lugar en la Rotonda de las Garantías Sociales, San José, Zapote, y cuenta con una cantidad de treinta (30) incidentes.



Figura 4.18 Pantalla Visualización, Cluster número 6 ($eps = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.7 Cluster No. 7

Identificado con el color amarillo podemos apreciar el *cluster* número siete, que tiene lugar en San Pedro de Montes de Oca, San José, en un segmento de la Carretera Interamericana de aproximadamente unos trescientos (300) metros, comprendidos entre la esquina suroeste del parque de la localidad hasta el cruce con la Calle 5 (Frente a la farmacia Fischel del lugar). En la Figura 4.19 se muestra este *cluster*.

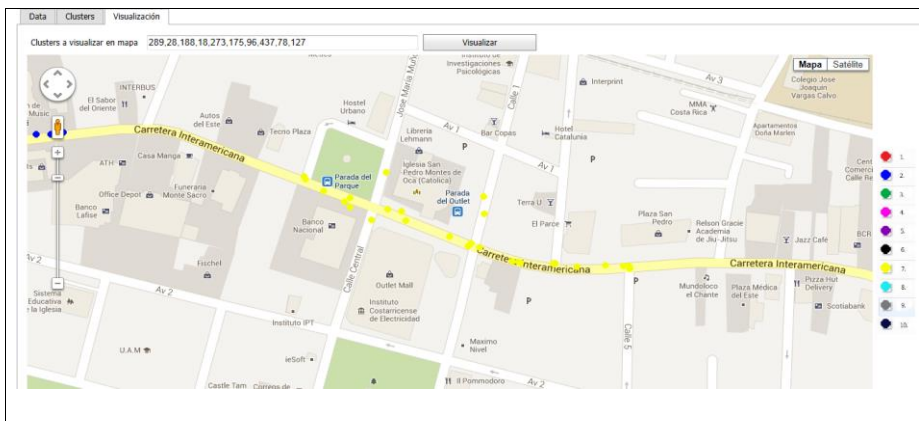
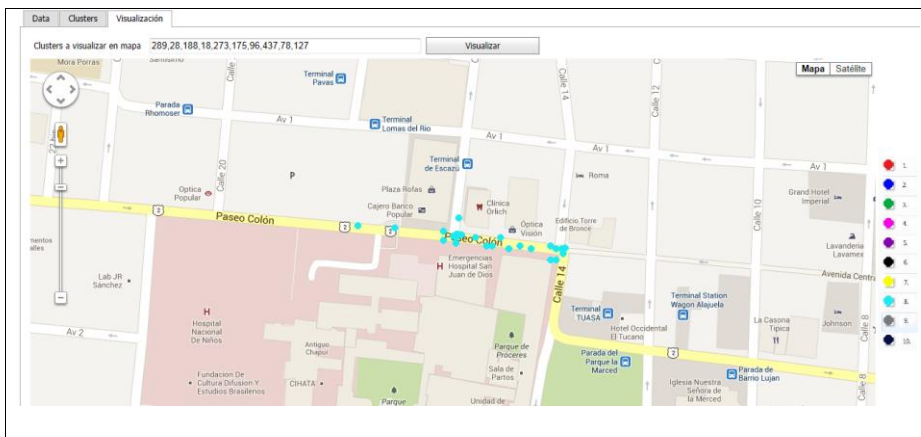


Figura 4.19 Pantalla Visualización, Cluster número 7 ($\epsilon = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.8 Cluster No. 8

Con una cantidad de veintiséis (26) elementos, visualizamos el octavo *cluster*, demarcado con el color verde claro y ubicado en el centro de San José, Paseo Colón, justo frente al Hospital San Juan de Dios, en un segmento de aproximadamente ciento setenta (170) metros. Se puede apreciar en la Figura 4.20.



*Figura 4.20 Pantalla Visualización, Cluster número 8 (eps = 50 metros)
Software para aplicar algoritmo de Agrupamiento.*

Fuente propia

4.7.9 Cluster No. 9

En la novena posición encontramos el *cluster* demarcado con el color gris, el cual se ubica en San Juan de Tibás, San José, en las cercanías del parque de la localidad. Cuenta con una cantidad de veinticinco (25) elementos y se puede observar en la Figura 4.21.

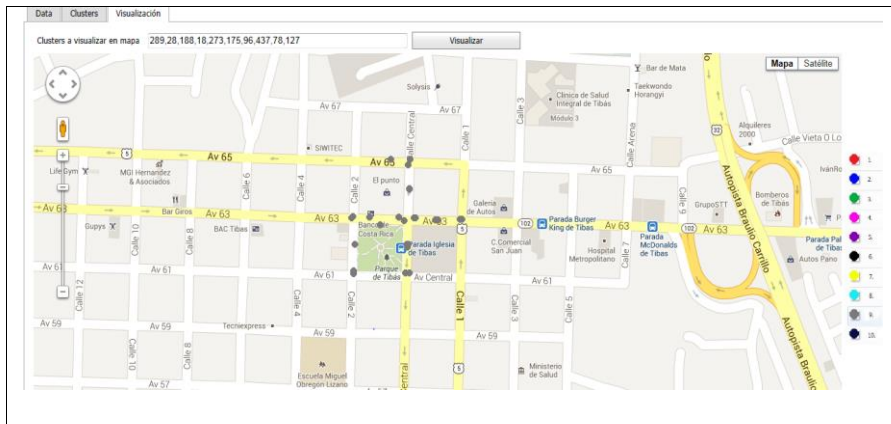


Figura 4.21 Pantalla Visualización, Cluster número 9 ($\text{eps} = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.7.10 Cluster No. 10

Finalmente, encontramos el décimo y último *cluster* de los seleccionados para este análisis, demarcado con los puntos de color azul oscuro. Cuenta con 25 puntos y se encuentra ubicado en Guadalupe de Goicoechea, San José. Notamos que es la misma tendencia de circunvalación; aunque este sector no tiene un rotonda, sí es un cruce bastante concurrido y complicado.

La Figura 4.22 muestra la ubicación en el mapa de este *cluster*.

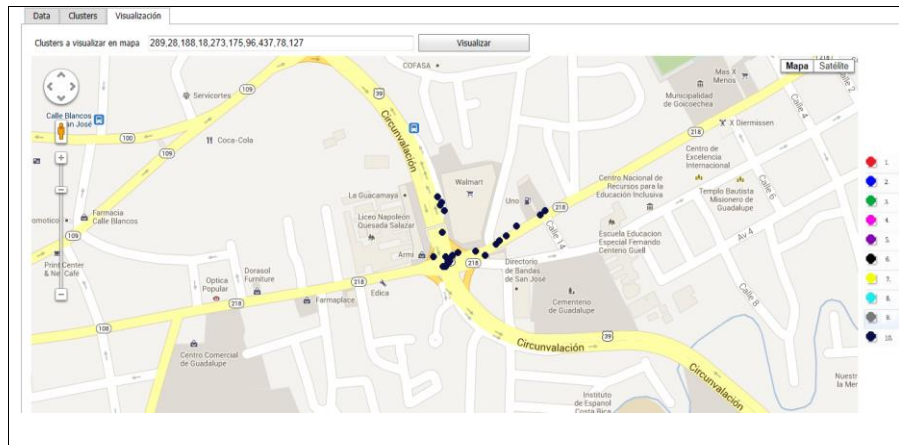


Figura 4.22 Pantalla Visualización, Cluster número 10 ($eps = 50$ metros)
Software para aplicar algoritmo de Agrupamiento.

Fuente propia

4.8 Otros datos Importantes

A pesar de que el principal proceso de este proyecto estaba orientado a utilizar el algoritmo de agrupamiento DBSCAN sobre la información de los incidentes de tránsito recolectada por Ruta AlternA, también se pudo descubrir otros datos que son bastante relevantes y que se pueden ver como un agregado producto de este trabajo. Son datos que, aunque no fueron resultados de aplicar la técnica de *clustering*, pueden resultar importantes en términos de los incidentes de tránsito.

Se presentan a continuación una serie de gráficos que se generaron en Microsoft Excel 2010 a partir de la información recopilada por Ruta AlternA.

4.8.1 Incidentes por Provincia

La Figura 4.23 muestra un gráfico circular con la cantidad de incidentes por Provincia.

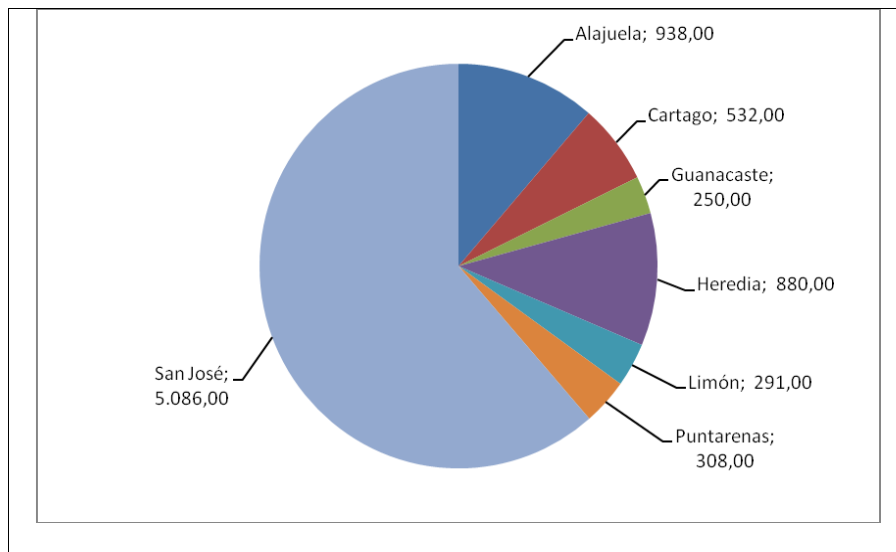


Figura 4.23 Gráfica Circular, Cantidad de Incidentes por Provincia

Fuente: Ruta AlternA

4.8.2 Incidentes por Día de Semana

En la Figura 4.24 se puede apreciar el gráfico de barras que muestra la cantidad de incidentes por día de semana.

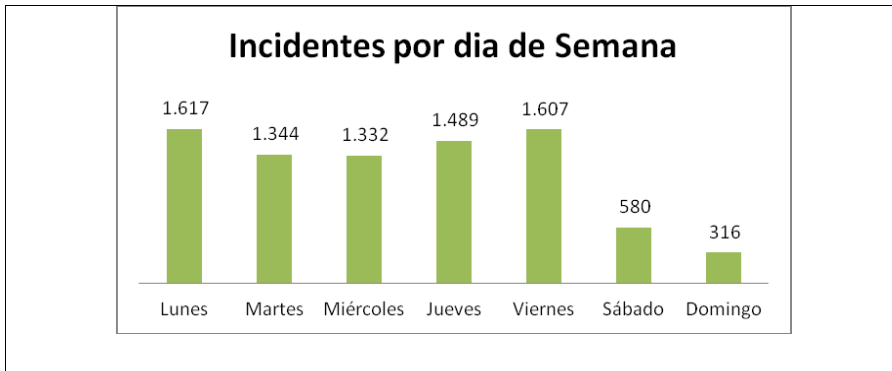


Figura 4.24 Gráfica de Barras, Cantidad de Incidentes por Días de Semana

Fuente: Ruta Alternativa

4.8.3 Incidentes por Hora del Día

La Figura 4.25 presenta un gráfico con la cantidad de incidentes por hora del día, en formato de 24 horas desde la hora Cero (0) hasta la hora veintitrés (23).

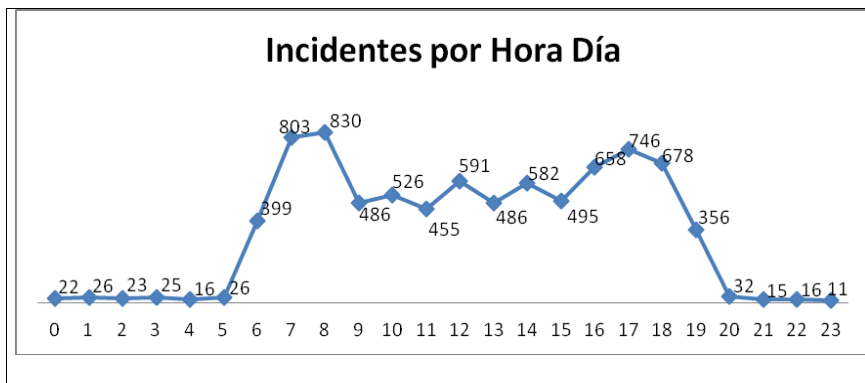


Figura 4.25 Gráfica de Líneas, Cantidad de Incidentes por Hora del día

Fuente: Ruta Alternativa

4.8.4 Incidentes por Cantón (San José)

En la figura 4.26 se muestra el gráfico de barras con los datos de incidentes por cantón solamente de la provincia de San José.

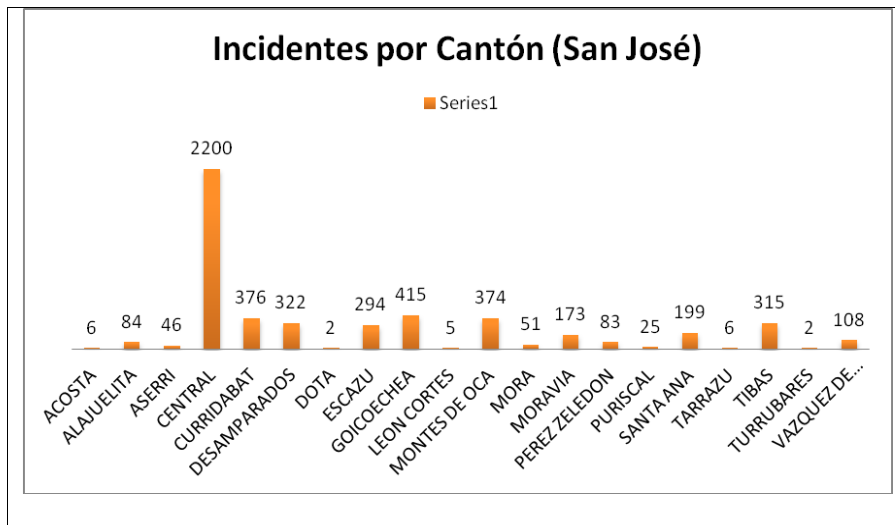


Figura 4.26 Gráfica de Barras, Cantidad de Incidentes por Cantón (Solo San José)

Fuente: Ruta Alternativa

4.8.5 Cantidad de Incidentes por Tipo

Se puede apreciar en la Figura 4.27 un gráfico de barras que presenta la cantidad de incidentes clasificados por Tipo de Incidentes.

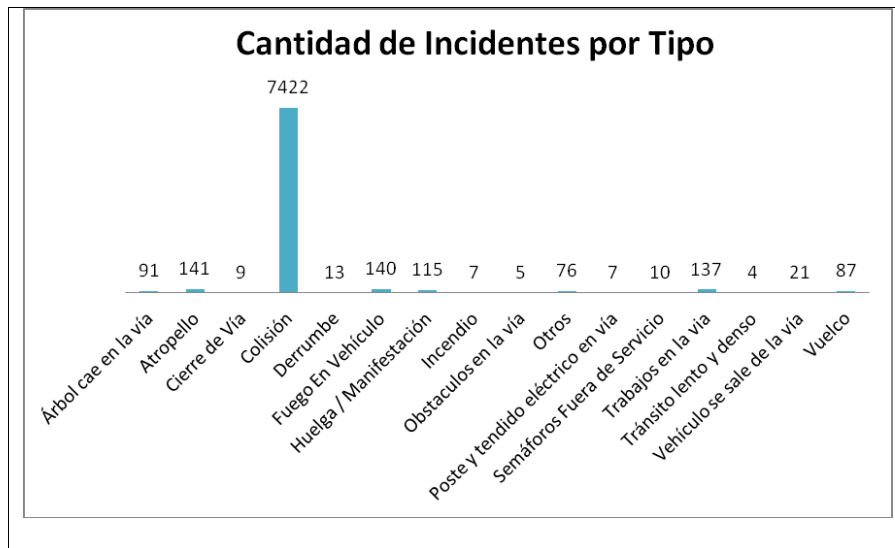


Figura 4.27 Gráfica de Barras, Cantidad de Incidentes por Tipo de Incidente

Fuente: Ruta Alternativa

Capítulo 5 – Análisis y Discusión de Resultados

Al hacer un análisis de los datos resultantes de este proyecto, se puede determinar los siguientes puntos o afirmaciones:

- La ubicación Geográfica con mayor cantidad de incidentes de tránsito está ubicada en la Rotonda de La Guacamaya en San Sebastián, San José.
- Se nota claramente que la mayor problemática con respecto a incidentes de tránsito se encuentra en las rotondas: cinco de los diez *clusters* seleccionados, es decir, un cincuenta por ciento, son rotondas.

- Adicional a esto se puede observar que cinco de los diez clusters seleccionados, o sea, también el cincuenta por ciento, se presentan en la carretera de la Circunvalación en el casco metropolitano. Como resultado de este análisis identificamos la ruta de circunvalación como la que mayor cantidad de incidentes de tránsito tiene.
- Puede inferirse por las ubicaciones geográficas definidas por los elementos de cada *cluster* que el flujo que más afectado se encuentra, es en el sector sur de la capital, es decir, el tráfico vehicular desde San José centro hacia el sur (Curridabat, Zapote, Cartago) y viceversa.

Cabe destacar que el segundo *cluster* con mayor cantidad de incidentes es en la Fuente de la Hispanidad, la cual es una de las principales salidas y entradas a la capital, el tercer *cluster* está ubicado en la Rotonda de la I Griega en San Francisco de Dos Ríos, la cual es también una salida importante hacia el sur de la capital, y lo mismo ocurre con la Rotonda de las Garantías Sociales, la cual es una salida importante hacia Zapote y Curridabat. Finalmente, tenemos los *clusters* cinco y siete, que están ubicados en las carreteras principales que salen de las rotondas de las Garantías Sociales y de la Fuente de la Hispanidad, respectivamente.

- Se determina que el día en que suceden más incidentes es el día lunes con 1617 incidentes, seguido por el día viernes con 1607 incidentes. Además, la hora en la que más suceden incidentes es a las 8 de la mañana con 830 incidentes.

Capítulo 6 – Conclusiones y trabajos futuros

6.1 Conclusiones

- Se logró realizar una investigación sobre las estadísticas existentes en Costa Rica respecto a los incidentes de tránsito. A pesar de que se encuentran diferentes sitios en Internet con este tipo de publicaciones, así como con información reciente y muy importante de entidades gubernamentales como lo es el MOPT o la DGPT, cabe rescatar que no se logra encontrar una estadística similar a la de este proyecto, en la cual se muestran las ubicaciones geográficas con más incidencias directamente en un mapa de Costa Rica.
- Habiendo contado con un diagrama de la estructura de la base de datos transaccional de Ruta Alternativa, se logra conocer dicha base de datos lo cual tener una noción clara para toda el resto del proyecto tanto en la parte del diseño “desnormalizado”, como el proceso de traslado de la información a la nueva estructura.
- Al hacer un refrescamiento de los diferentes procesos y metodologías para el diseño y estructuración de bases de datos, se logra realizar un diseño que, aunque simple, es bastante efectivo para almacenar la información sobre incidentes de tránsito, la cual permitió aplicar el algoritmo de agrupamiento de una manera ágil y rápida.
- Se logra crear un *Script* de SQL que permitió realizar el traslado de la información desde la base de datos transaccional a la nueva estructura “desnormalizada” de una manera exitosa teniendo la información necesaria y la estructura adecuada que permitió aplicar el algoritmo de agrupamiento
- Se concluyó con el desarrollo del software que realiza el proceso de aplicar el algoritmo de agrupamiento DBSCAN a la información recolectada por Ruta Alternativa, se logra determinar la distancia entre los elementos a partir de las coordenadas geográficas, además de esto el aplicativo permite visualizar los resultados del algoritmo en el mapa de Costa Rica, separando los diferentes *Clusters* por colores.

- Al aplicar el algoritmo de agrupamiento elegido, el DBSCAN, se logró realizar la agrupación según la distancia entre los incidentes, creando *clusters* con la información de incidentes de tránsito, los cuales permitieron una visualización de la información muy sencilla.
- Finalmente, de manera general, se logró presentar una metodología capaz de identificar los puntos geográficos del país con mayor cantidad de incidentes de tránsito. Haciendo uso de diferentes tecnologías se logró realizar una visualización sencilla, pero bastante representativa, la cual permite interpretar la información de manera ágil y acertada.

6.2 Reflexiones Finales

6.2.1 Reflexiones Documentales

Finalmente, una de las principales conclusiones de este proyecto es que se identifica cuál es punto geográfico con más incidentes: la Rotonda de la Guacamaya, en la vía de la Circunvalación, San José.

Además de esto, se identifica como la ruta con más incidentes de tránsito la ruta de la Circunvalación en San José, que rodea la capital y es una de las principales vías para entrar y salir de esta provincia.

Es importante recalcar que este proyecto permitió innovar la forma en que se identifican las ubicaciones geográficas con mayor cantidad de incidentes, de una manera gráfica, al ser mostrados directamente en un mapa de Costa Rica (Google Maps) se tiene las bondades de poder ubicar las carreteras, cruces, rutas cercanas a los puntos con mayor cantidad de incidentes, además el poder elegir diferentes escalas en el mapa para una mejor visualización.

6.2.2 Reflexiones Técnicas

Google Maps: Se debe mencionar la gran facilidad que representa trabajar con Google Maps, para un desarrollador acostumbrado a trabajar en JavaScript o JQuery le será de gran facilidad adaptar esta herramienta a sus proyectos programados. Es importante recalcar que el uso de esta herramienta es gratuito hasta un cierto límite de uso de consulta de los mapas (un límite bastante amplio)

SQL Server Spatial Tools: Esta herramienta se analizó pero no se utiliza para el proyecto, se considera a criterio personal que es una buena herramienta para efectos de cálculos geoespaciales incluso dibujar diferentes figuras geométricas, sin embargo no provee una aplicación o software de interface de usuario sino que provee una colección de funciones para SQL Server que se pueden reutilizar, de manera que con respecto a visualización esta herramienta no tiene mayor valor agregado. Es importante mencionar que al momento de la confección de este documento SQL Server Spatial Tools está en una versión beta y que la última versión liberada fue en 2011.

6.3 Trabajos Futuros

Sin duda alguna, a pesar del gran esfuerzo que realiza Ruta Alternativa por proveer una base de datos pública con información de incidentes de tránsito, aún quedan puntos de mejora, se considera que algunas de estas mejoras o valores agregados a este proyecto podrían darse cumpliendo con las siguientes ideas:

- Desarrollar una plataforma para dispositivos móviles que permita la interacción con terceras personas o usuarios externos a Ruta Alternativa que puedan registrar información sobre los incidentes de tránsito, de esta manera se podrá recolectar la información de incidentes de tránsito con un mayor porcentaje de exactitud, ya que actualmente Ruta Alternativa no logra manejar la información total sobre estos incidentes, principalmente porque no se recolecta información las veinticuatro horas del día, o simplemente no se tiene acceso a dicha información. Es importante mencionar que la plataforma debe ser capaz de permitirle a usuarios internos de Ruta Alternativa aprobar o desaprobar la información brindada por estos usuarios externos, con la finalidad de que sean datos verídicos y no algún usuario dando información falsa.
- Crear una plataforma que permita unificar los datos de las diferentes instituciones que tienen contacto en su día a día con información de los incidentes de tránsito ya sean gubernamentales o privadas, esto debido a que en la actualidad dicha información se encuentra descentralizada o en algunos casos no se almacena. Evidentemente el éxito de la plataforma a desarrollar dependerá de que la o las instituciones tengan un interés real en apoyar este proyecto por lo cual implicará el negociar con estas instituciones y lograr adaptarse a las necesidades que estas tengan en términos de negocio, confidencialidad entre otros.
- Recolectar información complementaria de los incidentes de tránsito. Este proyecto se enfocó en las carreteras (sus ubicaciones geográficas) que es una de las tres variables que involucran los incidentes de tránsito, hasta el momento la principal información que se ha logrado recolectar es la ubicación, fecha, hora y el tipo de incidente; sin embargo, sería muy útil para efecto de análisis y estadísticas poder tener acceso a datos de las otras dos variables relacionadas en los incidentes de

tránsito que son conductor y el vehículo, sería muy interesante poder realizar análisis sobre esta información y por ejemplo identificar perfiles de los conductores que tienen mayor cantidad de incidentes para poder tomar decisiones y acciones antes que estos conductores puedan originar futuros accidentes. Algo similar se podría aplicar a la información de los vehículos, identificar el o los perfiles de los vehículos que ocasionan mayor cantidad de accidentes y de esta manera poder tomar alguna acción preventiva antes que ocurran estos accidentes.

- Crear un proceso de depuración de la información recolectada actualmente por Ruta Alterna, al momento de la creación de este documento la base de datos contaba con más de treinta mil (30,000) registros, número que es completamente manejable para el algoritmo de agrupamiento que se aplica en este proyecto, sin embargo es importante prever para un mediano o largo plazo un proceso de depuración que permita trasladar la información a una base de datos histórica con la finalidad de optimizar el procesamiento de datos nuevos. Se aclara que el proceso que ejecuta el algoritmo de agrupamiento debe tener acceso tanto a la base de datos actual como a la histórica en caso de que se desee realizar un análisis de datos históricos.

Referencias

- A Density-Based Algorithm, (s.f.). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Recuperado el 12/11/2013 de <http://dns2.icar.cnr.it/manco/Teaching/2005/datamining/articoli/KDD-96.final.frame.pdf>
- Bäcklund, H., Hedblom, A. & Neijman, N. (2011). DBSCAN A Density-Based Spatial Clustering of Application with Noise. Recuperado el 11/11/2013 de [http://staffwww.itn.liu.se/~aidvi/courses/06/dm/Seminars2011/DBSCAN\(4\).pdf](http://staffwww.itn.liu.se/~aidvi/courses/06/dm/Seminars2011/DBSCAN(4).pdf)
- Chinchilla (1992). Diagnóstico Situacional de los accidentes de tránsito en Costa Rica. Recuperado el 14/07/2013 de <http://www.ts.ucr.ac.cr/binarios/pela/pl-000015.pdf>
- Colegio de Ingenieros Civiles de Costa Rica, (s.f.). Seguridad Vial. Recuperado el 22/04/2013 de <http://www.civiles.org/publi/articulos/seguridadvial.pdf>
- Estadística, (s.f.). Fundacion paz en las carreteras. Recupero el 22/04/2013 de <http://fundacionsospazenlascarreteras.org/estadisticas.html>
- Hernández et al, (2003). Metodología de la Investigación, Mc Graw Hill, México
- Informe Final de Gestión, MOPT, (16/07/2012). Recuperado el 22/04/2013 de <http://mopt.go.cr/%5Cdocumentos%5Cpoliciatransito%5Cinformecesar.pdf>
- Quintero, Víctor Manuel y Velásquez, Mario, (1993). Los indicadores sociales
- Ruta Alternativa, (s.f.). Información del estado de las carreteras en Costa Rica, Recuperado el 22/04/2013 de <http://www.rutaalternativa.org>

SQL Server Spatial Tool, (31/03/2011). Colección de herramientas para ser utilizadas con los tipos de datos espaciales que tiene el nuevo SQL Server 2008, Recuperado el 16/07/2013 de <http://sqlspatialtools.codeplex.com/releases/view/63588>

UPEL, (2006). Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales, Recuperado el 11/11/2013 de <http://neutron.ing.ucv.ve/NormasUPEL2006.pdf>

Bibliografía

Bouckaert, R. et al. (2010). WEKA Manual for Version 3-6-2. Recuperado de <http://kent.dl.sourceforge.net/project/weka/documentation/3.6.x/WekaManual-3-6-2.pdf>

Microsoft Developer Network (2012). Algoritmos de minería de datos (Analysis Services: Minería de datos). Recuperado de <http://msdn.microsoft.com/es-es/library/ms175595.aspx>

Moreira, A., Santos M., & Carneiro, S. (2005). [Density-based clustering algorithms – DBSCAN and SNN](#). Recuperado de <http://denis.kraynov.2009.homepage.auditory.ru/2006/Ivan.Ignatyev/AD/snn%26dbscan.pdf>

Vulpeş (2011). [Implementing the DBSCAN Algorithm using C#](#). Recuperado de <http://www.c-sharpcorner.com/uploadfile/b942f9/implementing-the-dbscan-algorithm-using-C-Sharp/>

Con formato: Fuente: Sin Cursiva, Inglés (Reino Unido)

Con formato: Fuente: Sin Cursiva, Inglés (Reino Unido)

Con formato: Inglés (Reino Unido)

Apéndices

1. Seudocódigo DBSCAN

```

DBSCAN (SetOfPoints, Eps, MinPts)

// SetOfPoints is UNCLASSIFIED
ClusterId := nextId(NOISE);
FOR i FROM 1 TO SetOfPoints.size DO
  Point := SetOfPoints.get(i);
  IF Point.ClId = UNCLASSIFIED THEN
    IF ExpandCluster(SetOfPoints, Point,
      ClusterId, Eps, MinPts) THEN
      ClusterId := nextId(ClusterId)
    END IF
  END IF
END FOR
END; // DBSCAN

ExpandCluster(SetOfPoints, Point, ClId, Eps,
  MinPts) : Boolean;
seeds:=SetOfPoints.regionQuery(Point,Eps);
IF seeds.size<MinPts THEN // no core point
  SetOfPoint.changeClId(Point,NOISE);
  RETURN False;
ELSE // all points in seeds are density-
  // reachable from Point
  SetOfPoints.changeClIds(seeds,ClId);
  seeds.delete(Point);
  WHILE seeds <> Empty DO
    currentP := seeds.first();
    result := SetOfPoints.regionQuery(currentP,
      Eps);
    IF result.size >= MinPts THEN
      FOR i FROM 1 TO result.size DO
        resultP := result.get(i);
        IF resultP.ClId
          IN {UNCLASSIFIED, NOISE} THEN
          IF resultP.ClId = UNCLASSIFIED THEN
            seeds.append(resultP);
          END IF;
          SetOfPoints.changeClId(resultP,ClId);
        END IF; // UNCLASSIFIED or NOISE
      END FOR;
    END IF; // result.size >= MinPts
    seeds.delete(currentP);
  END WHILE; // seeds <> Empty
  RETURN True;
END IF
END; // ExpandCluster

```

Tabla 8.1 Algoritmo DBSCAN

Fuente: A Density-Based Algorithm

