

División de Posgrado

Trabajo de Titulación

Aplicación de Herramientas de Inteligencia de Negocios para la Construcción de un Cuadro de Mando de Alertas de Anomalías de Transporte Público.

Proyecto de Desarrollo que se presenta como opción al título de Magíster en:

Máster en Sistemas de Información con Mención en Inteligencia de Negocios

Autor:

Geancarlos Molina González

Tutor:

Ing. Luis Llanganate

Guayaquil, 16 de junio de 2019

Agradecimiento

Le doy gracias a Dios y a mi familia, por ser fuente de apoyo e inspiración constante, por ser mi fortaleza y motor para mejorar cada día, ya que sin su estímulo no hubiese sido posible equiparar el esfuerzo de llevar consecutivamente el desarrollo profesional, combinado con el estudio, enriquecido por los docentes de la prestigiosa institución que hará posible que ostente este título, y a quienes también dejo, en adición, mi más profundo sentir de gratitud por los conocimientos impartidos.

Certificación del tutor

Tutor: Ing. Luis Llanganate

Certifica

Que el presente trabajo de titulación: **"Aplicación de Herramientas de Inteligencia de Negocios para la Construcción de un Cuadro de Mando de alertas de anomalías de transporte público."** Fue revisado en su totalidad y se autoriza para su presentación ante el Tribunal de Grado, siendo su contenido original en su totalidad.

DE ACUERDO A LOS ANTECEDENTES MENCIONADOS, AUTORIZO AL SR. GEANCARLOS MOLINA GONZÁLEZ PARA SU IMPRESIÓN Y PRESENTACIÓN.

Guayaquil, 2 de octubre 2019.

Firmado electrónicamente por:
LUIS FERNANDO LLANGANATE MORENO

Ing. Luis Llanganate

Certificado del porcentaje de coincidencias

Habiendo sido nombrado, MSc. Luis Fernando Llanganate Moreno, tutor del trabajo de titulación **"Aplicación de Herramientas de Inteligencia de Negocios para la Construcción de un Cuadro de Mando de alertas de anomalías de transporte público."** elaborado por Ingeniero Geancarlos Molina González, con mi respectiva supervisión como requerimiento parcial para la obtención del título de Magister en Sistemas de Información con mención en Inteligencia de Negocios.

Se informa que el mismo ha resultado tener un porcentaje de coincidencias

5% mismo que se puede verificar en el siguiente link:

https://secure.urkund.com/account/account/submitter/11224057-609105-779663

Adicional se adjunta la captura de pantalla de dicho resultado.



Índice General

TABLA DE CONTENIDOS
Agradecimiento 2
Índice General 5
Resumen 11
Introducción 13
Tema 15
Detalle de la Necesidad 15
Antecedentes 15
Planteamiento del Problema 16
Pregunta problémica 16
Idea para defender 16
Variables 17
Justificación 17
Alcance o Tipo de Investigación 18
Objetivo General 19
Objetivos Específicos. 19
Aspecto Innovador del proyecto. 19
1 Marco Teórico 21
1.1 Sistema de posicionamiento global 21
1.2 Arquitectura 21
1.3 Principios de funcionamiento 22
1.4 Sistemas de control de transporte público. 23
1.4.1 Proyecto SITU 23
1.5 Inteligencia de Negocios 23

1.6 Metodología CRISP-DM 24
1.6.1 Comprensión del negocio 26
1.6.2 Determinar los objetivos del negocio. 28
1.6.2.1 Evaluación de la situación 28
1.6.2.2 Determinar los objetivos de la minería de datos 28
1.6.2.3 Realizar el plan del proyecto 29
1.6.3 Comprensión de los datos 29
1.6.3.1 Recolectar los datos iniciales 30
1.6.3.2 Descripción de los datos. 31
1.6.3.3 Exploración de los datos 31
1.6.3.4 Verificar la calidad de los datos 31
1.6.4 Preparación de los datos. 31
1.6.4.1 Seleccionar los datos 33
1.6.4.2 Limpiar los datos. 34
1.6.4.3 Construir los datos. 34
1.6.4.4 Integrar los datos. 34
1.6.4.5 Formateo de los datos 34
1.6.5 Modelado 35
1.6.5.1 Escoger la técnica de modelado 36
1.6.5.2 Generar el plan de prueba. 37
1.6.5.3 Construir el modelo 37
1.6.5.4 Evaluar el modelo 37
1.6.6 Evaluación 38
1.6.6.1 Evaluar los resultados 39
1.6.6.2 Revisar el proceso. 39
1.6.6.3 Determinar los próximos pasos 39
1.6.7 Despliegue o implantación 39
1.6.7.1 Planear la implantación 41
1.6.7.2 Planear la monitorización y mantenimiento 41
1.6.7.3 Producir el informe final 41

1.6.7.4 Revisar el proyecto 41
1.7 **Herramientas** 42
1.7.1 Power BI 43
1.7.1.1 Componentes de Power BI 45
1.7.1.2 Perfiles del usuario 46
1.7.1.3 El flujo de trabajo en Power BI 47
1.7.1.4 Servidor local de informes de Power BI 48
2 Marco Metodológico y Análisis de Resultados 49
2.1 Metodología CRISP-DM 49
2.1.1 Comprensión del negocio. 49
2.1.1.1 Determinar los objetivos del negocio. 50
2.1.1.2 Contexto 50
2.1.1.3 Objetivos del negocio 51
2.1.1.4 Evaluación de la situación. 51
2.1.1.4.1 Inventario de recursos 52
2.1.1.4.2 Riesgos y contingencias 52
2.1.1.4.3 Costes y beneficios 53
2.1.1.4.4 Determinar los objetivos de la minería de datos 54
2.1.1.4.5 Realizar el plan del proyecto 54
2.1.1.4.6 Evaluación inicial de herramientas y técnicas 55
2.1.1.5 Comprensión de los datos 55
2.1.1.6 Recolectar los datos iniciales 56
2.1.1.7 Descripción de los datos. 56
2.1.1.7.1 Tabla de Vehículos – tbl_vehiculos 57
2.1.1.7.2 Tabla de proveedores de servicio - tbl_proveedor 57
2.1.1.7.3 Tabla de alertas de tránsito - tbl_alertas 58
2.1.2 Exploración de los datos 58
2.1.2.1 Verificar la calidad de los datos 62
2.1.2.2 Preparación de los datos. 62
2.1.2.3 Seleccionar los datos 62

2.1.2.4 Limpiar los datos. 63
2.1.2.5 calidad de los datos. 64
2.1.2.5.1 Completitud 64
2.1.2.5.2 Conformidad 64
2.1.2.5.3 Consistencia 64
2.1.2.5.4 Precisión / Exactitud 64
2.1.2.5.5 Duplicación 64
2.1.2.5.6 Integridad 65
2.1.2.6 Construir los datos. 65
2.1.2.7 Integrar los datos. 65
2.1.2.8 Formateo de los datos 66
2.1.3 Modelado 67
2.1.3.1 Escoger la técnica de modelado. 67
2.1.3.2 Generar el plan de prueba. 68
2.1.3.3 Construir el modelo. 68
2.1.3.3.1 Indicadores de situación 68
2.1.3.3.2 Indicadores de infracción 68
2.1.3.3.3 Producción de Multas de Velocidad. 68
2.1.3.3.4 Multas de Velocidad por cooperativas 68
2.1.3.3.5 Multas de Velocidad por Vehículo 69
2.1.3.3.6 Mapa de calor de concentración de Multas de Velocidad 69
2.1.3.4 Evaluar el modelo 69
2.1.4 Evaluación 69
2.1.4.1 Evaluar los resultados 70
2.1.4.2 Revisar el proceso. 70
2.1.4.3 Determinar los próximos pasos 70
2.1.5 Despliegue o implantación 70
2.1.5.1 Planear la implantación. 70
2.1.5.2 Instalación de del entorno. 71
2.1.5.2.1 Instalar Power BI Desktop 71

2.1.5.2.2 Instalar Power BI Gateway. 71
2.1.5.2.3 Acceder desde la nube. 71
2.1.5.2.4 Planear la monitorización y mantenimiento. 72
2.1.5.3 Producir el informe final 73
2.1.5.4 Revisar el proyecto 73
2.2 Metodología de los 3 Ciclos 73
2.2.1 Entendimiento del Negocio 73
2.2.1.1 Instrucción a la metodología 73
¿Quién**?** 73
2.2.1.2 Desarrollo 74
2.2.2 Entendimiento de los objetivos: 75
2.2.2.1 Introducción 75
2.2.3 Preparación de datos 77
2.2.3.1 la Matriz Origen – Destino 77
2.2.3.2 Modelado 79
2.2.3.3 ETL 79
2.2.3.4 Despliegue 80
3 CAPÍTULO 3. VALIDACIÓN 81
3.1 Validación de los resultados. 81
3.2 Esquema de validación 81
Aplicación y descripción de los resultados 83
3.2.1 Indicadores de situación 83
3.2.2 Producción de Multas de Velocidad. 84
3.2.3 Multas de Velocidad por cooperativas 85
3.2.4 Multas de Velocidad por Vehículo 86
3.2.5 Mapa de calor de concentración de Multas de Velocidad 87
3.3 Impacto 87
Conclusiones y Recomendaciones 88
4 Bibliografía 90

INDICE DE ILUSTRACIONES

Ilustración 1 Fases de la Metodología CRISP-DM....26
Ilustración 2 Fases de Compresión del negocio....27
Ilustración 3 Fase de comprensión de los datos....30
Ilustración 4 Fase de preparación de los datos....33
Ilustración 5 Fase de Modelado....36
Ilustración 6 Fase de Evaluación....38
Ilustración 7 Fase de Implantación....40
Ilustración 8 Cuadrante Mágico de para analítica e inteligencia de negocios Fuente Gartner Febrero 2019....43
Ilustración 9 Esquema de conexiones de Power BI....44
Ilustración 10 Componentes de Power BI....46
Ilustración 11Comparación de Power BI Desktop y el servicio Power BI....47
Ilustración 12 Diccionario de datos....56
Ilustración 13 Exploración de Unidades de transporte publico Registradas por proveedor de sistemas de flotas....59
Ilustración 14 Unidades por proveedor y tipo....60
Ilustración 15 Unidades por cooperativa....61
Ilustración 16 Diagrama de entidades involucradas....63
Ilustración 17 Modelo de Datos en Power BI....66
Ilustración 18 Modelo de datos implementado en Power BI....79
Ilustración 19 Listado de operadoras de transporte....80
Ilustración 20 Listado de proveedores de servicios....80
Ilustración 21 Listado de Vehículos registrados....81
Ilustración 22 Cuadro de mando de estado de Registro....83
Ilustración 23 Producción de multas de velocidad....84
Ilustración 24 Multas de velocidad por cooperativa....85
Ilustración 25 Multas de velocidad por vehículo....86
Ilustración 26 Mapa de calor de Infracciones por velocidad....87
Ilustración 27 Carta de autorización de uso de datos....91

Resumen

En julio del 2017 la Autoridad de Tránsito de Guayaquil (*ATM*) firma el contrato para iniciar el proceso de implementación de la plataforma de monitoreo del proyecto SITU, con el Consorcio Aurora Transporte Público. A partir de esa fecha se comienzan a generar los datos telemétricos que a su vez generan multas electrónicas basadas en reglas asociadas a normas de transporte público relativas a velocidad, permisos de operación y reglas de parada.

En este trabajo, se describe una propuesta de mejora al proceso de generación de multas electrónicas con base en telemetría de datos GPS de transporte público, realizada en la ciudad de Guayaquil, a través del uso de herramientas de inteligencia de negocios, de tal forma que se use una perspectiva proactiva para analizar y tomar decisiones acerca de los comportamientos que causan estas infracciones.

Habiendo obtenido los permisos adecuados ante la autoridad de tránsito de Guayaquil, se trabajó con el total de la información. Se analizó la clasificación cronológica y agrupación geográfica de los datos obtenidos de multas electrónicas a través de diferentes escenarios asociados a un indicador específico, con el fin de encontrar patrones y tendencias en estos comportamientos que sirvan para poder establecer decisiones relacionadas al mejoramiento de los procesos de transportación, señalética o algún otro aspecto que la autoridad del tránsito considere relevante para cumplir el objetivo buscado.

Se utilizó CRISP-DM como metodología de implementación de inteligencia de negocios y Power BI como herramienta de desarrollo de los escenarios abordados en este estudio, pudiendo mostrar así que los hallazgos encontrados en términos de comportamiento de transporte público sirven para mejorar de manera proactiva el proceso de generación de multas electrónicas.

Palabras clave: Multas electrónicas de tránsito, telemetría de GPS, cuadros de mando

Abstract

On July 2017, the Guayaquil Transit Authority (ATM) signs the contract to start implementing the monitoring platform of the SITU project, with the Aurora Transporte Publico Consortium. From that date onwards, the telemetric data is generated, from which electronic fines are established based on rules associated with public transport regulations regarding to speed, operating permits and stop rules.

This paper describes a propossal of improving the process of generating electronic fines based on GPS data telemetry of public transport carried out in the city of Guayaquil, through the use of business intelligence tools, in such a way that a proactive perspective is used to analyze and make decisions about the behaviors that cause these infractions. After obtaining the appropriate permits from the Guayaquil transit authority, the total information was worked on. The chronological classification and geographic grouping of the data obtained from electronic fines was analyzed through different scenarios associated with a specific indicator, in order to find patterns and trends in these behaviors that serve to establish decisions related to the improvement of transportation processes, signage or any other aspect that the traffic authority considers relevant to fulfill the objective sought.

Crisp-dm was used as a methodology for implementing business intelligence and Power BI as a tool for developing the scenarios addressed in this study, thus being able to show that the findings found in terms of public transport behavior serve to proactively improve the process of generating electronic fines

Keywords: Electronic traffic fines, GPS telemetry, dashboards

Introducción

La movilidad de las ciudades contribuye a su consolidación y es un efecto directamente proporcional al crecimiento de la densidad poblacional, creando un sistema económico cada vez más globalizado y competitivo en busca de mejores condiciones de vida y facilidades para la ejecución de actividades productivas. La comprensión de los sistemas de movilidad podría tener efectos sobre la eficiencia de la economía urbana.

El transporte público colectivo en la ciudad de Guayaquil representa una de las diferentes alternativas de desplazamiento, entre los cuales se encuentran los buses de transporte urbano, buses de transporte rápido como la Metrovía, el taxi autorizado, los denominados taxis ejecutivos, el automóvil particular, la bicicleta o la marcha a pie, entre otros. La calidad de los medios de transporte en el contexto urbano indica las condiciones a las que se enfrenta el ciudadano común en el uso y acceso a los servicios urbanos.

En el Ecuador, existe una cultura muy arraigada de parte de los choferes de transportación pública, de no tener responsabilidad al momento de conducir los buses; por lo tanto, el sistema de control y monitoreo, de la ciudad de Guayaquil, planifica la monitorización del recorrido de los buses de transporte público mediante tecnología GPS, usando el posicionamiento global en su recorrido, evitando notificaciones de exceso de velocidad en el área urbana, controlando la parada en el paradero, la circulación fuera de rutas autorizadas, asegurar el cumplimiento de las frecuencias de salida y el control de circulación con puertas abiertas.

El uso de sistemas de posicionamiento global (*GPS*, por sus siglas en inglés) para el control de unidades de transporte público constituye una tendencia ampliamente utilizada en el control de flotas de vehículos. Con la instalación de este tipo de equipo en las unidades en

operación, se puede obtener información exacta en tiempo real relacionada al posicionamiento, sentido, velocidad y. estado de diferentes sensores telemétricos.
Desde el punto de vista del operador de transporte, permite mejorar sus rendimientos de servicio, llevar un control eficiente de sus recorridos, un estimado de los pasajeros que suben y bajan de la unidad. Además, permite brindar información útil al usuario en el momento, y como efecto proveer servicios operados de manera confiable, eficiente y al menor costo posible.

Tanto los operadores de transporte como los entes reguladores deben tener un adecuado control del cumplimiento de las regulaciones existentes. Auditando reglas específicas de operación que corresponden a velocidad, recorrido, paradas y tiempos. Así como la evaluación del servicio a través del monitoreo constante de los datos obtenidos de la telemetría.

A pesar de que ya se han implantado soluciones innovadoras que apoyan el control y cumplimiento de las normas establecidas, muchos de los incidentes terminan en multas o procesos penales por accidentes y otras contravenciones que pudieran ser evitadas con un análisis proactivo basado en el comportamiento histórico del tránsito público.

Tema

Aplicación de Herramientas de Inteligencia de Negocios para la Construcción de un Cuadro de Mando de alertas de anomalías de transporte público.

Detalle de la Necesidad

Antecedentes

El rápido crecimiento urbano que han experimentado las principales ciudades del Ecuador ha hecho necesaria la intervención de los gobiernos locales con el fin de mejorar las condiciones de movilidad urbana, mediante políticas tendientes a la creación de sistemas de transporte público, eficientes y accesibles.

En la ciudad de Guayaquil los buses urbanos están dotados de unidades de GPS que reportan telemetría de ubicación en tiempo real basado en la obligatoriedad definida legalmente a través de una ordenanza municipal, con el fin de poder controlar usando un medio electrónico, el incumplimiento de las normas de tránsito relacionada a velocidad, desvíos de rutas, respeto a las paradas, circulación de puertas abiertas, etc.

Desde enero del año 2018, la autoridad de tránsito viene generando multas automáticas basadas en los datos obtenidos de geoposicionamiento, con buenos resultados en términos del mejoramiento de acatar las reglas estipuladas en la normativa del transporte público, sin embargo, se busca siempre el mejoramiento continuo a través de la implementación de más y mejores formas de análisis de los datos generados.

Una de estas formas es el uso de dispositivos GPS, sensores de apertura de puerta, botones de pánico contadores de pasajeros, etc. que alimentan una base de datos geoespacial centralizada, que nos ha servido como fuente de datos principal de esta investigación

Planteamiento del Problema

Pregunta problémica

¿Cómo mejorar la identificación de incumplimientos de normas de tránsito a partir de la telemetría que arrojan los dispositivos GPS de buses urbanos?

Idea para defender

Si se aplican técnicas de inteligencia de negocios basadas en indicadores y tableros de control para la identificación de anomalías en el transporte público, se mejorará la identificación de incumplimientos de normas de tránsito y por ende la toma de decisiones de la autoridad competente.

Variables

Tomando en cuenta que la unidad de análisis es el proyecto SITU de la ciudad de Guayaquil se definen las siguientes variables:

Independientes:

Telemetría de Ubicación Satelital, Multa de tránsito electrónica, Concentración geográfica.

Dependientes:

Identificación de incumplimiento de normas de transporte público

Justificación

El control de tránsito en las ciudades constituye un factor importante en relación con el ordenamiento y control que ejercen las autoridades encargadas de este tema en vías de normar y reducir los índices de aglomeración de vehículos, siniestros y violaciones a las leyes asociadas en este contexto. Uno de los actores causantes de dichas anomalías son los buses de transporte urbano, los cuales al momento ya están siendo controlados de acuerdo la obligatoriedad del marco regulatorio de competencias de los Gobiernos Autónomos descentralizados. En ellas se especifica a través de ordenanzas el uso de dispositivos de telemetría de posicionamiento global (GPS) para el monitoreo y control del transporte público, como se menciona en (Telégrafo, 2017)

Esta normativa ya está en ejecución en los gobiernos locales de “Categoría A” de Ecuador, y es usada para el control y monitoreo de las flotas de transporte público de varias ciudades, específicamente en la ciudad de Guayaquil se implementa desde julio del 2017 la

plataforma de monitoreo del SITU (Sistema Integrado de Transporte Urbano), con efectos altamente positivos, como se menciona en (elcomercio.com, 2018)

El alcance actual de estos controles reactivos está en el monitoreo y la generación de multas automáticas por exceso de velocidad, desvíos de rutas, irrespeto de paradas y circulación con puertas abiertas. Estas medidas no son suficientes para disminuir los índices de siniestros que se originan por el incumplimiento de las leyes de tránsito y seguridad vial.

Por esta razón se propone la implementación de tableros de control que permitan llevar análisis de las distorsiones que se encuentren en la operación de la transportación pública a un nivel mayor de efectividad, transformando estos controles en una herramienta proactiva que disminuya los índices de infracción aún más de lo conseguido actualmente, a través, del uso y estudio de diferentes escenarios de inteligencia de negocios, que permitan determinar si existen acciones preventivas que las autoridades de tránsito puedan ejecutar, como por ejemplo: mejoramiento de la señalética, mayor inversión en capacitación de conductores, adecuación de vías o la necesidad de inspecciones físicas de los lugares donde ocurren concentración de eventos.

Alcance o Tipo de Investigación

Este es un proyecto de investigación de tipo exploratorio-diagnóstico se concluye con en el ámbito explicativo con una propuesta tecnología aplicada en un escenario real, tomando datos de la plataforma SITU de la ciudad de Guayaquil, usando la herramienta de inteligencia de negocios con la cual se proyecta resolver la mejora de identificación de anomalías en transporte público

Objetivo General

Desarrollar un cuadro de mando de anomalías en el transporte público utilizando una herramienta de inteligencia de negocios, para mejorar el proceso identificación de patrones inadecuados e incumplimientos de normas de tránsito .

Objetivos Específicos.

- Contextualizar el marco teórico relacionado con la telemetría de transporte público, la inteligencia de negocios y los cuadros de mando.
- Establecer los indicadores que se desplegaran en el cuadro de mando de comportamiento de anomalías en el transporte público de en la ciudad de Guayaquil.
- Visualizar en una herramienta de inteligencia de negocios que permitan analizar efectivamente los indicadores de comportamiento de las anomalías en el transporte público.
- Aplicar la propuesta en un escenario real.

Aspecto Innovador del proyecto.

Aplicación de técnicas de Inteligencia de negocios sobre las anomalías de transporte público para reducir comportamientos que generen contravenciones y comportamientos inadecuados a través de la toma de decisiones proactivas. Este aspecto innovador permitirá determinar a través de herramientas de inteligencia de negocios, como mapas de calor, por ejemplo, la posible concentración geográfica de anomalías existentes en la operación diaria de la transportación pública, las cuales reflejarán diferentes posibles causas como falta de

señalética, posibles daños en vías, o aspectos que requerirán de una revisión física de las zonas de concentración detectadas.

1 Marco Teórico

1.1 Sistema de posicionamiento global

El sistema global de posicionamiento satelital o GPS, del acrónimo de la expresión en idioma inglés "*Global Positioning System*", pone a disposición pública uno de los más importantes avances tecnológicos de las últimas décadas. Su diseño original fue como herramienta de carácter militar para la estimación precisa de posición, velocidad y tiempo, se ha utilizado ampliamente también en múltiples aplicaciones civiles que involucran rastreo y ubicación (U.S.government, 2016),.

Sin embargo, por razones de seguridad, las señales GPS generadas para uso civil se someten a una degradación deliberada, y a la vez su emisión se restringe a una determinada frecuencia. aun así, las aplicaciones civiles siguen proliferando a un ritmo acelerado ya que es común que los sistemas GPS estén integrados desde teléfonos inteligentes hasta vehículos y demás dispositivos comerciales de rastreo satelital.

1.2 Arquitectura

Según la referencia (Pozo-Ruz, 2000) el sistema se descompone en tres segmentos básicos, los dos primeros de uso militar:

1 Segmento espacio, formado por 24 satélites GPS con una órbita de 26560 Km. de radio y un período de 12h.;
2 Segmento control, que consta de cinco estaciones monitoras encargadas de mantener en órbita los satélites y supervisar su correcto funcionamiento, tres antenas terrestres que envían a los satélites las señales que deben transmitir y una estación experta de supervisión de todas las operaciones

3 Segmento usuario, formado por las antenas y los receptores pasivos situados en tierra. Los receptores, a partir de los mensajes que provienen de cada satélite visible, calculan distancias y proporcionan una estimación de posición y tiempo. Situación de las estaciones monitoras y de la estación de control maestra

1.3 Principios de funcionamiento

El sistema GPS tiene por objetivo calcular la posición de un punto cualquiera en un espacio de coordenadas, partiendo del cálculo de las distancias del punto a un mínimo de tres satélites cuya localización es conocida. La distancia entre el usuario (receptor GPS) y un satélite se mide multiplicando el tiempo de vuelo de la señal emitida desde el satélite por su velocidad de propagación. Para medir el tiempo de vuelo de la señal de radio es necesario que los relojes de los satélites y de los receptores estén sincronizados, pues deben generar simultáneamente el mismo código. Ahora bien, mientras los relojes de los satélites son muy precisos los receptores son osciladores de cuarzo de bajo coste y por tanto imprecisos.

1.4 Sistemas de control de transporte público.

1.4.1 Proyecto SITU

Sistema Integrado de Transporte Urbano (SITU). Proyecto para el mejoramiento del transporte público de la ciudad de Guayaquil que bajo una Plataforma Tecnológica permite el control de 6 ejes:

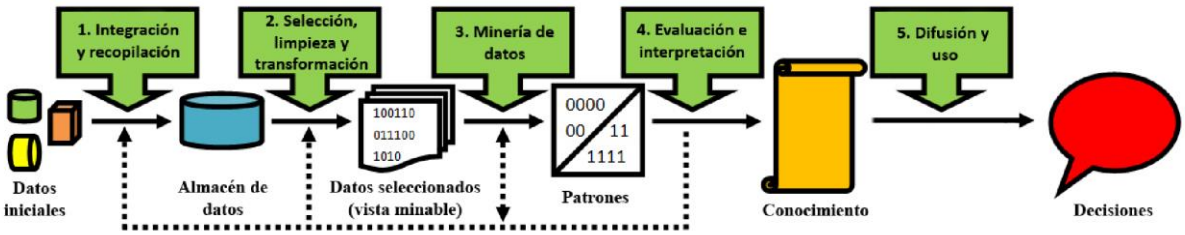
1. Exceso de velocidad.
2. Desvíos de ruta.
3. Incumplimiento de frecuencia.
4. Apertura de puertas en lugares no autorizados.
5. Incumplimiento de parada en paradero.
6. Botón de Pánico

Tiene como objetivo principal generar seguridad vial para los ciudadanos a través de la reducción de la siniestralidad en el corto plazo y el cambio cultural en el mediano y largo plazo.

1.5 Inteligencia de Negocios

Para hallar patrones significativos en grandes volúmenes de datos se ha usado inicialmente la estadística y, más recientemente, el aprendizaje automático (Coria, 2016), un área de la inteligencia artificial. La conjunción de estas disciplinas con la teoría y práctica de las bases de datos ha dado origen a la minería de datos, conocida también como descubrimiento de conocimiento en bases de datos. Los patrones hallados constituyen modelos descriptivos, predictivos o clasificadores que posteriormente pueden servir para implementar software de aplicación especializada o para guiar la revisión de políticas o procedimientos de las organizaciones. La necesidad de preservar y organizar datos para facilitar su consulta y

análisis ha dado origen a los data warehouse y data marts. Un data warehouse es una base de datos sumariados, organizados en tablas de hechos y dimensiones que facilita el procesamiento analítico en línea (OLAP), favoreciendo el desempeño en consultas masivas.



1.6 Metodología CRISP-DM

El Modelo *CRISP-DM* (Cross Industry Standard Process for Data Mining), (crisp-dm.eu, 2015), fue concebido a fines de 1996 por DaimlerChrysler, SPSS y NCR. Este es un marco de estructuración de procesos de Analítica y minería de datos que describe la forma en la que gran parte los expertos en esta materia aborda el problema. Para implementar una tecnología en un negocio es necesaria una metodología. Estos métodos suelen venir de las experiencias propias y también de los procedimientos estándar más conocidos.

En el caso de los proyectos de implementación de analítica y minería de datos una de las metodologías que ha tenido más apoyo de las empresas privadas y organismos públicos es *CRISP-DM*, como se puede observar en la siguiente gráfica a continuación, publicada en *kdnuggets.com*, que es un sitio líder en IA, Analytics, Big Data, Data Mining, Data Science y Machine Learning , editado por Gregory Piatetsky-Shapiro y Matthew Mayo, (Piatetsky, 2014), según esta referencia *CRISP-DM*, sigue siendo la metodología vigente e importante

como modelo de estructuración de proyectos de analítica y minería de datos y siendo el proyecto SITU de corte público es una de las razones de porque usar esta metodología .

¿Qué metodología principal está utilizando para sus proyectos de análisis, minería de datos o ciencia de datos? [200 votos total]

2014 2007

CRISP-DM (86)	43% 42%
My Own (55)	27.5% 19%
SEMMA (17)	8.5% 13%
Other, not domain-specific (16)	8% 4%
KDD Process (15)	7.5% 7.3%
My Organizations' (7)	3.5% 5.3%
A Domain-specific Methodology (4)	2% 4.7%
None (0)	0% 4.7%

CRISP-DM incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior

para ser revisada, por lo que la sucesión de fases no tiene porqué ser ordenada desde la primera hasta la última. En la figura siguiente se puede observar las fases en las que se divide *CRISP-DM* y las posibles secuencias a seguir entre ellas.

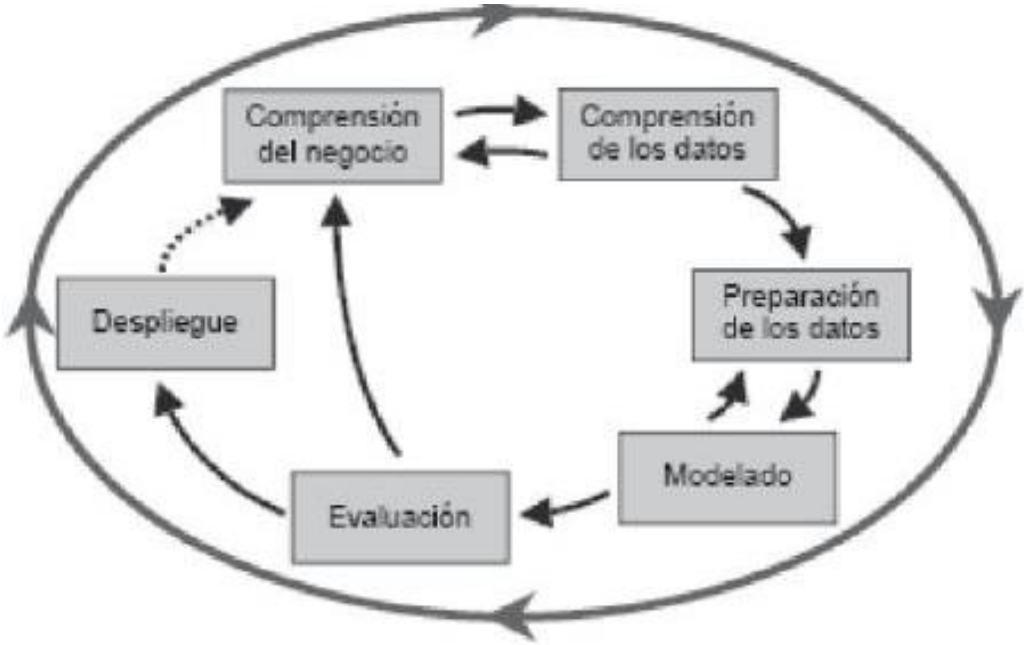


Ilustración 1 Fases de la Metodología CRISP-DM

1.6.1 Comprensión del negocio

Esta primera fase es probablemente la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva de negocio, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea permitirá obtener resultados fiables. Para obtener el mejor provecho de la minería de datos, es necesario entender de la manera más completa el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados. En esta fase, es muy importante la

capacidad de poder convertir el conocimiento adquirido del negocio en un problema de minería de datos y en un plan preliminar cuya meta sea el alcanzar los objetivos del negocio. A continuación, vemos una descripción de cada una de las principales tareas que componen esta fase.

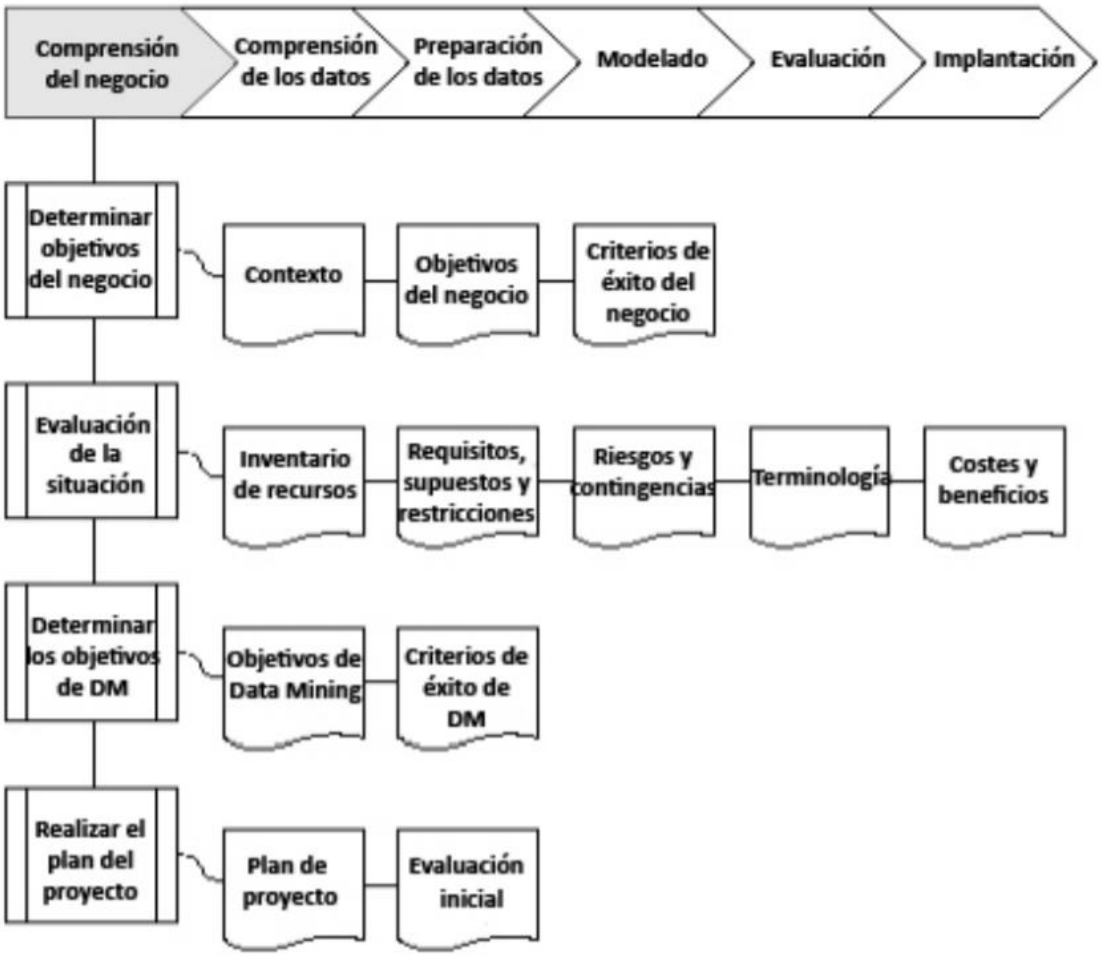


Ilustración 2 Fases de Compresión del negocio

1.6.2 Determinar los objetivos del negocio.

Esta es la primera tarea para desarrollar y tiene como metas determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar la minería de datos y definir los criterios de éxito. Los problemas pueden ser diversos, como, por ejemplo, detectar fraude en el uso de tarjetas de crédito, detección de intentos de ingreso indebido a un sistema, asegurar el éxito de una determinada campaña publicitaria, etc. En cuanto a los criterios de éxito, estos pueden ser de tipo cualitativo, en cuyo caso un experto en el área de dominio califica el resultado del proceso de minería de datos, o bien de tipo cuantitativo, por ejemplo, el número de detecciones de fraude o la respuesta de clientes ante una campaña publicitaria.

1.6.2.1 Evaluación de la situación

En esta tarea se debe calificar el estado de la situación antes de iniciar el proceso de minería de datos, considerando aspectos tales como: ¿Cuál es el conocimiento previo disponible acerca del problema?, ¿Se cuenta con la cantidad de datos requerida para resolver el problema?, ¿Cuál es la relación coste beneficio de la aplicación de minería de datos?, etc. En esta fase se definen los requisitos del problema, tanto en términos de negocio como en términos de minería de datos.

1.6.2.2 Determinar los objetivos de la minería de datos

Esta tarea tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de minería de datos, como, por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar la asignación de créditos hipotecarios, la meta de minería de datos será por ejemplo determinar el perfil de los clientes respecto de su capacidad de endeudamiento.

1.6.2.3 Realizar el plan del proyecto

Esta última tarea de la primera fase de *CRISP-DM* tiene como meta desarrollar un plan para el proyecto, que describa los pasos a seguir y las técnicas a emplear en cada uno de ellos.

1.6.3 Comprensión de los datos

Esta segunda fase comprende la recolección inicial de los datos con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis. Esta fase junto a las dos siguientes fases son las que demandan el mayor esfuerzo y tiempo en un proyecto de minería de datos. Por lo general si la organización cuenta con una base de datos corporativa, es deseable crear una nueva base de datos específica para el proyecto de DM (*Data Mining*), ya que durante el desarrollo del proyecto es posible que se generen frecuentes y abundantes accesos a la base de datos con el fin de realizar consultas y probablemente se produzcan modificaciones, lo cual podría generar muchos problemas. Vemos las tareas que componen esta fase

1.6.3.1 Recolectar los datos iniciales

La primera tarea en esta segunda fase del proceso de *CRISP-DM* es la recolección de los datos iniciales y su adecuación para el futuro procesamiento. Esta tarea tiene como objetivo elaborar informes con una lista de los datos adquiridos, su localización, las técnicas utilizadas en su recolección y los problemas y soluciones inherentes a este proceso.

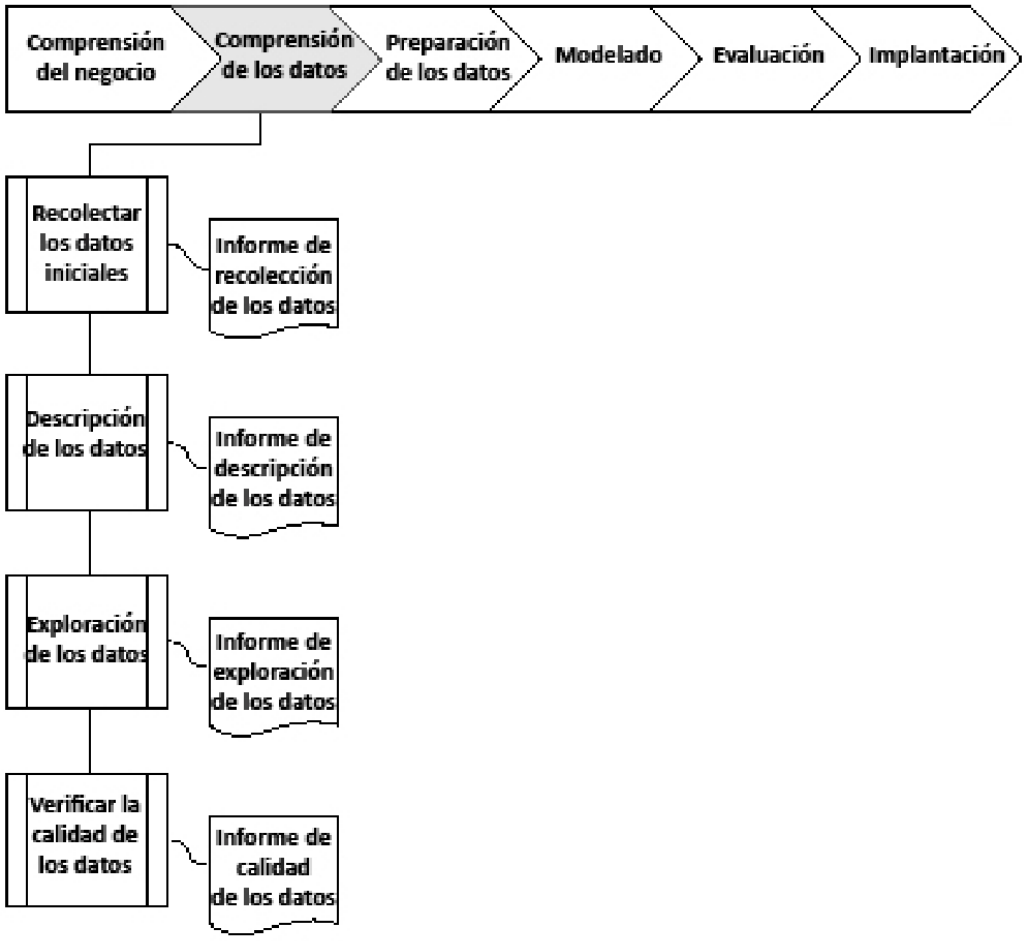


Ilustración 3 Fase de comprensión de los datos

1.6.3.2 Descripción de los datos.

Después de adquiridos los datos iniciales, estos deben ser descritos. Este proceso implica establecer volúmenes de datos (número de registros y campos por registro), su identificación, el significado de cada campo y la descripción del formato inicial.

1.6.3.3 Exploración de los datos

Una vez realizada la descripción de los datos, se procede a su exploración, con el fin de encontrar una estructura general para los datos. Esto implica la aplicación de pruebas estadísticas básicas que revelen propiedades en los datos recién adquiridos, se crean tablas de frecuencia y se construyen gráficos de distribución. La salida de esta tarea es un informe de exploración de los datos.

1.6.3.4 Verificar la calidad de los datos

En esta tarea se efectúan verificaciones sobre los datos para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso. La idea una vez llegados a este punto es poder garantizar la completitud y corrección de los datos.

1.6.4 Preparación de los datos.

En esta fase y una vez efectuada la recolección inicial de los datos, se procede a su preparación para adaptarlos a las técnicas de minería de datos que se van a utilizar posteriormente, éstas pueden ser técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para explotación de los datos. La preparación de

los datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato.

Esta sección se encuentra relacionada con la fase de modelado, ya que, en función de la técnica de modelado elegida, los datos requieren ser procesados de una manera o de otra, por esta razón las fases de preparación y de modelado interactúan de forma permanente.

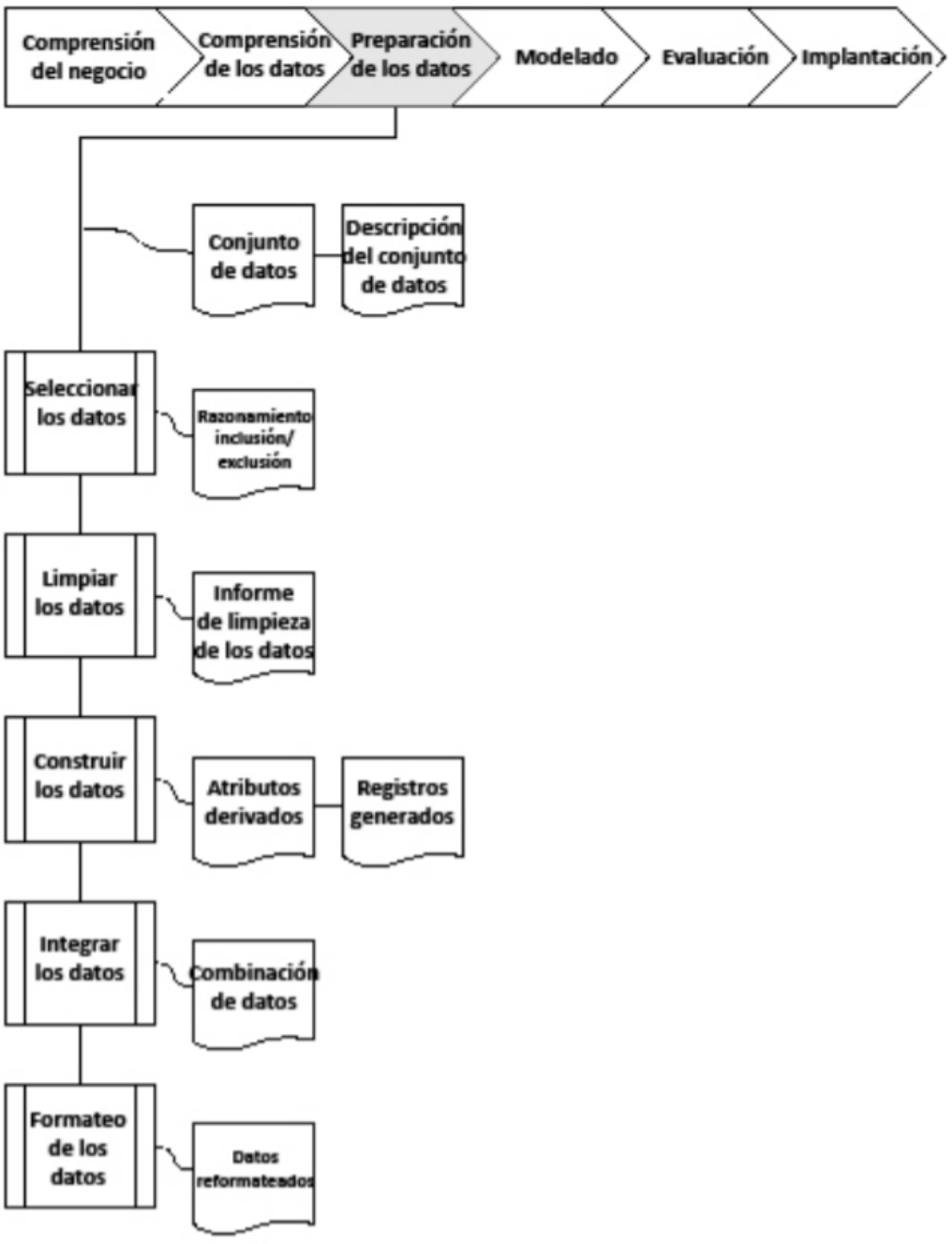


Ilustración 4 Fase de preparación de los datos

1.6.4.1 Seleccionar los datos

En esta etapa se selecciona un subconjunto de los datos adquiridos anteriormente apoyándose en criterios previamente definidos en las fases anteriores como la calidad de los datos en cuanto a su completitud, corrección de los datos y limitaciones en el volumen o en los tipos de datos que están relacionados con las técnicas de minería de datos seleccionadas.

1.6.4.2 Limpiar los datos.

Esta tarea complementa a la anterior y es una de las que más tiempo y esfuerzo consume debido a la diversidad de técnicas que pueden aplicarse para optimizar la calidad de los datos con el objeto de prepararlos para la fase de modelación. Algunas de las técnicas a utilizar para este propósito son la normalización de los datos, discretización de campos numéricos, tratamiento de valores faltantes, reducción del volumen de datos, etc.

1.6.4.3 Construir los datos.

Esta tarea incluye las operaciones de preparación de los datos tales como la generación de nuevos atributos a partir de atributos ya existentes, integración de nuevos registros o transformación de valores para atributos existentes.

1.6.4.4 Integrar los datos.

La integración de los datos implica la creación de nuevas estructuras a partir de los datos seleccionados, por ejemplo, generación de nuevos campos a partir de otros existentes, creación de nuevos registros, fusión de tablas campos o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen.

1.6.4.5 Formateo de los datos

Esta tarea consiste principalmente en la realización de transformaciones sintácticas de los datos sin modificar su significado de tal forma que se permita y se facilite utilizar alguna técnica de minería de datos en concreto, como por ejemplo el reordenamiento de los campos y/o de los registros de la tabla o el ajuste de los valores de los campos a las limitaciones de las herramientas de modelación (eliminar comas, tabuladores, caracteres especiales, máximos y mínimos para las cadenas de caracteres, etc.)

1.6.5 Modelado

En esta fase de *CRISP-DM* se seleccionan las técnicas de modelado más apropiadas para el proyecto de minería de datos específico. Las técnicas por utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada para el problema.
- Disponer de los datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica.

Previamente al modelado de los datos se debe determinar un método de evaluación de los modelos que permita establecer el grado de adecuación de cada uno de ellos. Después de concluir estas tareas genéricas se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo. La figura a continuación muestra las tareas y las salidas que se obtienen en esta fase.

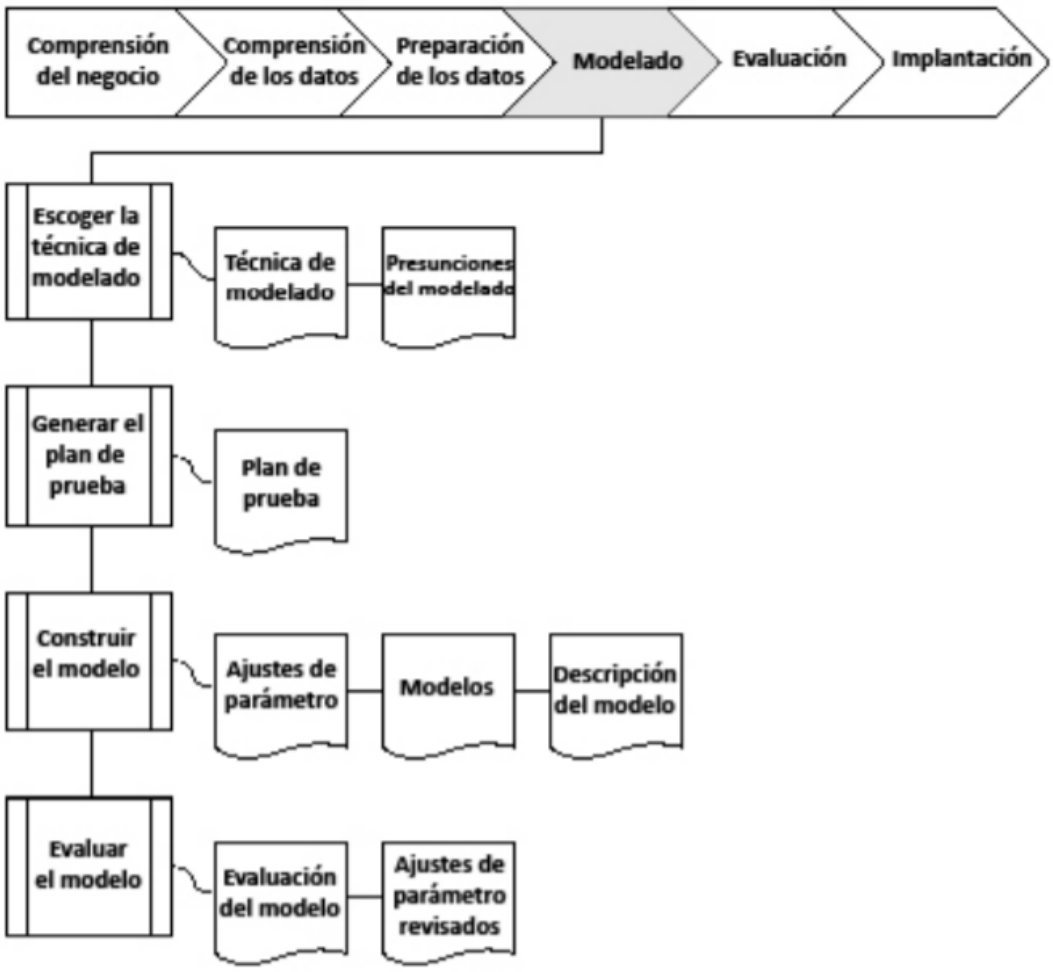


Ilustración 5 Fase de Modelado

1.6.5.1 Escoger la técnica de modelado

Esta tarea consiste en la selección de la técnica de minería de datos más apropiada al tipo de problema que se quiere resolver. Para esta selección, se debe considerar el objetivo principal del proyecto y la relación con las herramientas de minería de datos existentes. Por ejemplo, si el problema es de clasificación, se podrá elegir de entre árboles de decisión, k-nearest neighbors o razonamiento basado en casos (CBR), si el problema es de predicción,

análisis de regresión o redes neuronales, o si el problema es de segmentación, redes neuronales, técnicas de visualización, etc.

1.6.5.2 Generar el plan de prueba.

Se debe generar un procedimiento destinado a probar la calidad y validez del modelo elegido una vez que éste esté construido. Por ejemplo, en una tarea supervisada de minería de datos como la clasificación, es común usar la razón de error como medida de la calidad. Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba.

1.6.5.3 Construir el modelo

A continuación, se ejecuta la técnica seleccionada sobre los datos previamente preparados para generar uno o más modelos. Todas las técnicas de modelado tienen un conjunto de parámetros que determinan las características del modelo a generar. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.

1.6.5.4 Evaluar el modelo

En esta última tarea de esta fase de modelado los ingenieros de DM interpretan los modelos de acuerdo con el conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Expertos en el dominio del problema juzgan los modelos dentro del contexto del dominio y expertos en minería de datos aplican sus propios criterios (seguridad del conjunto de prueba, pérdida o ganancia de tablas, etc.).

1.6.6 Evaluación

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se pueda haber cometido algún error. Hay que considerar que se pueden emplear múltiples herramientas para la interpretación de los resultados. Si el modelo generado es válido en función de los criterios de éxito establecidos en la fase anterior, se procede a la explotación del modelo.

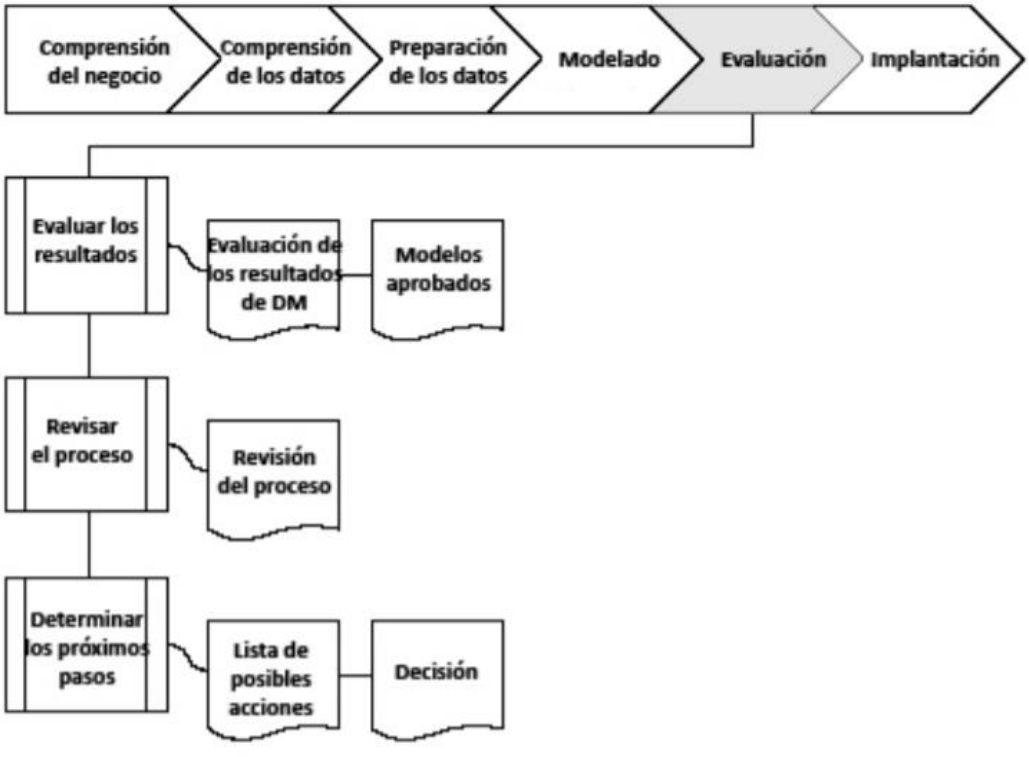


Ilustración 6 Fase de Evaluación

1.6.6.1 Evaluar los resultados

En los pasos de evaluación anteriores se trataron factores tales como la exactitud y generalidad del modelo generado. Esta tarea involucra la evaluación del modelo con relación a los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual el modelo sea deficiente, o si es aconsejable probar el modelo en un problema real si el tiempo y las restricciones lo permiten. Además de los resultados directamente relacionados con el objetivo del proyecto, ¿Es aconsejable evaluar el modelo en relación con otros objetivos distintos a los originales?, esto podría revelar información adicional.

1.6.6.2 Revisar el proceso.

Este proceso se refiere a calificar al proceso entero de minería de datos a objeto de identificar elementos que pudieran ser mejorados.

1.6.6.3 Determinar los próximos pasos

Si se ha determinado que las fases hasta este momento han generado resultados satisfactorios podría pasarse a la siguiente fase, en caso contrario podría decidirse por hacer otra iteración desde la fase de preparación de los datos o de modelado con distintos parámetros. Podría incluso darse el caso de que en esta fase se decida empezar desde cero con un nuevo proyecto de minería de datos.

1.6.7 Despliegue o implantación

En esta fase, y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, esto puede hacerse por

ejemplo cuando el analista recomienda acciones basadas en la observación del modelo y sus resultados, o por ejemplo aplicando el modelo a diferentes conjuntos de datos o como parte del proceso (en análisis de riesgo de créditos, detección de fraudes, etc.). Generalmente un proyecto de minería de datos no concluye en la implantación del modelo, ya que se deben documentar y presentar los resultados de manera comprensible para el usuario con el objetivo de lograr un incremento del conocimiento. Por otra parte, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados. Las tareas que componen esta fase son:

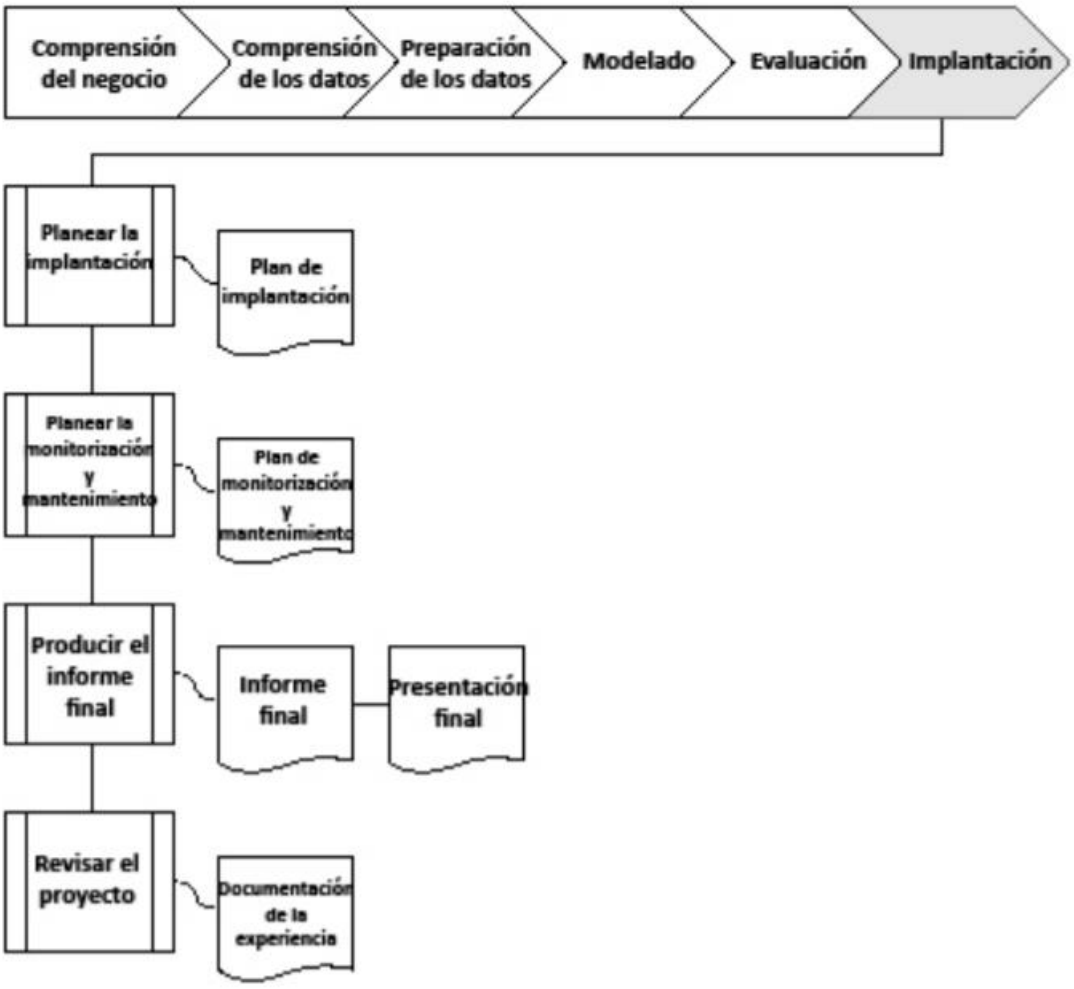


Ilustración 7 Fase de Implantación

1.6.7.1 Planear la implantación

Para implementar el resultado de la minería de datos en la organización, esta tarea tomó los resultados de la evaluación y concluye una estrategia para su implementación. Si un procedimiento general se ha identificado para crear el modelo, este procedimiento debe ser documentado para su posterior implementación.

1.6.7.2 Planear la monitorización y mantenimiento

Si los modelos resultantes del proceso de minería de datos son implementados en el dominio del problema como parte de la rutina diaria, es aconsejable preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos. La retroalimentación generada por la monitorización y mantenimiento pueden indicar si el modelo está siendo utilizado apropiadamente.

1.6.7.3 Producir el informe final

En este paso se debe presentar un informe resumiendo los puntos importantes del proyecto y la experiencia adquirida durante su desarrollo

1.6.7.4 Revisar el proyecto

En esta última etapa de la metodología se debe hacer una evaluación de aquellas cosas que se hicieron correctamente y aquellas que no, así como posibles mejoras para ir puliendo los fallos y se obtengan mejores resultados.

1.7 Herramientas

Parte del reto de cumplir con los objetivos de este estudio es elegir la herramienta con la cual se va a aplicar el diseño del modelo de datos y los escenarios de cuadro de mando conectados a un escenario real

Las empresas líderes en el mercado según el "Cuadrante Mágico para Plataformas de Analítica y de Inteligencia de Negocios de Gartner" (Microsoft, 2019), son en orden de importancia Microsoft, Tableau y Qlik, pero debido a la facilidad de uso para poder evaluar la herramienta se escogió Power BI como la plataforma para probar este estudio basado en los siguientes criterios:

- Microsoft ofrece sin costo todas las herramientas para poder desarrollar y publicar los escenarios de cuadro de mando a diseñar, inclusive con una cuenta personal se tiene 10 GB de espacio en la nube para almacenar los datos de los indicadores generados.
- Se integra perfectamente con las aplicaciones existentes, lo que le permite adoptar capacidades de análisis y generación de informes.
- Proporciona consultas y análisis avanzados sin necesidad de soporte técnico especializado. Soporta lenguaje natural y el uso de herramientas intuitivas de diseño gráfico, y gran cantidad de documentación en línea.
- Cuadros de mando personalizados, flexibles y dinámicos para satisfacer las necesidades de este estudio.
- Extracción de datos de forma fácil e intuitiva.
- Permite publicar informes de forma segura y es fácil configurar la actualización automática de datos, publicar informes y poderlos visualizar en diferentes plataformas de sistema operativo.



Ilustración 8 Cuadrante Mágico de para analítica e inteligencia de negocios Fuente Gartner Febrero 2019

1.7.1 Power BI

Es una herramienta orientada a la inteligencia de negocios y la minería de datos que permite conectarse fácilmente a múltiples fuentes de datos por medio de una serie de servicios de

software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos en información coherente, interactiva y atractiva visualmente.

Es de carácter sencillo y rápido capaz de crear información a partir de una hoja de cálculo de Excel o una base de datos local inclusive teniendo modestos conocimientos en manejo de bases de datos. Tiene una funcionalidad variable ,útil e intuitivo para los usuarios y las empresas interesados en análisis en tiempo real.



Ilustración 9 Esquema de conexiones de Power BI

En la hoja de ruta de Power BI tal como se menciona en (Russo, 2016), se describe que Microsoft introdujo la idea de Self-Service Business Intelligence (BI) en 2009, anunciando Power Pivot para Microsoft Excel 2010. Curiosamente, en ese momento, no hizo grandes anuncios, ni hizo una gran campaña de marketing para ello. Todo comenzó lentamente, con alguna cantidad de usuarios entusiastas que adoptan la nueva tecnología, pero la gran mayoría de las personas ni siquiera sabían sobre su existencia, sin embargo. En ese momento, se pudo ver claramente las ventajas para los usuarios de comenzar a adoptar Power Pivot como una herramienta para recopilar información de los datos, por lo que esta falta total de marketing fue algo decepcionante.

Fue la demora en la promoción de BI de autoservicio para la mayor audiencia de analistas de datos, científicos de datos, tomadores de decisiones y entusiastas de BI en todo el planeta. Con el primer lanzamiento de Power BI, una experiencia que aún no era completamente satisfactoria, mientras se esperaba para que Microsoft solucione los problemas con las versiones anteriores y comience a anunciar la versión actual

Power BI es una evolución de los complementos disponibles anteriormente en Excel: Power Pivot, Power Query y Power View. Actualmente se puede usar Power BI con o sin Excel: ya que no depende de una versión de Microsoft Office instalada. A la gente no le gustaba compartir informes usando solo SharePoint, y Microsoft se alejó de él. Los usuarios querían una experiencia móvil, y el equipo de desarrollo lo creó. Los analistas de datos querían poder, simplicidad, nuevas visualizaciones y todo está disponible en Power BI.

Además, se realizó un gran esfuerzo para crear una experiencia en la carga de datos de muchas fuentes de nube diferentes y en la construcción de la infraestructura necesaria para proporcionar a todos los entusiastas de BI un marco con el que puedan hacer crecer sus informes, compartirlos con sus equipos y actualice los datos de una manera simple pero efectiva

1.7.1.1 Componentes de Power BI

Según (microsoft, 2019), Power BI Consta de varios componentes complementarios entre sí que están diseñados para permitir a los usuarios crear, compartir y utilizar información de la forma que les resulte más eficaz, que listan a continuación:

- **Power BI Desktop**: Aplicación de escritorio de Windows para diseño de modelos.
- **Servicio Power BI**: Un servicio *SaaS* (*software como servicio*) en línea.
- **Aplicaciones móviles**: Para dispositivos Windows, iOS y Android.

- **Power BI Report Server**: Permite publicar informes de Power BI en un servidor de informes local, tras crearlos en Power BI Desktop.

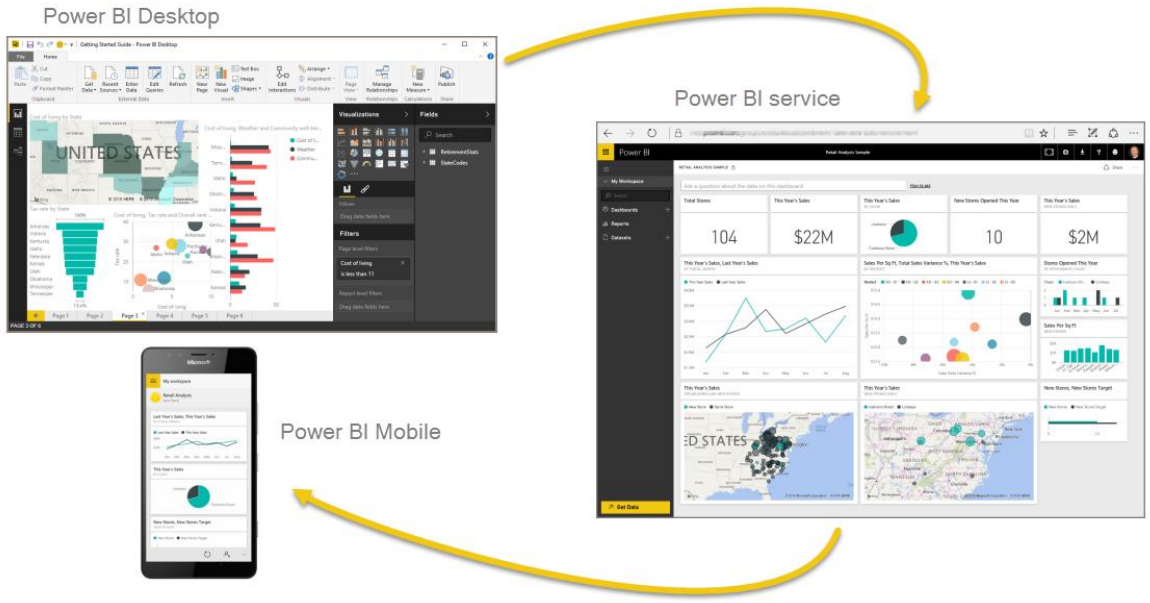


Ilustración 10 Componentes de Power BI

,

1.7.1.2 Perfiles del usuario

Es posible que la forma en que utilice Power BI dependa de su rol en un proyecto o equipo. Por tanto, otras personas con otros roles podrían utilizar esta herramienta de un modo distinto, lo cual es normal.

Se puede usar *Power BI Desktop* para crear informes para su propio equipo sobre estadísticas de involucración del cliente o ver el progreso de fabricación e inventario en un panel en tiempo real del servicio.

Según el perfil de cada usuario una clasificación usual es:

- Diseñadores
- Consumidores
- Desarrolladores
- Administradores

1.7.1.3 El flujo de trabajo en Power BI

Un flujo de trabajo común en Power BI comienza con la conexión a orígenes de datos y el diseño u generación de un informe en Power BI Desktop. A continuación, publicar dicho informe y lo comparte para que los usuarios finales del servicio y los dispositivos móviles puedan ver e interactuar con el informe. Este flujo de trabajo es habitual y muestra cómo los tres elementos principales de Power BI se complementan entre sí.

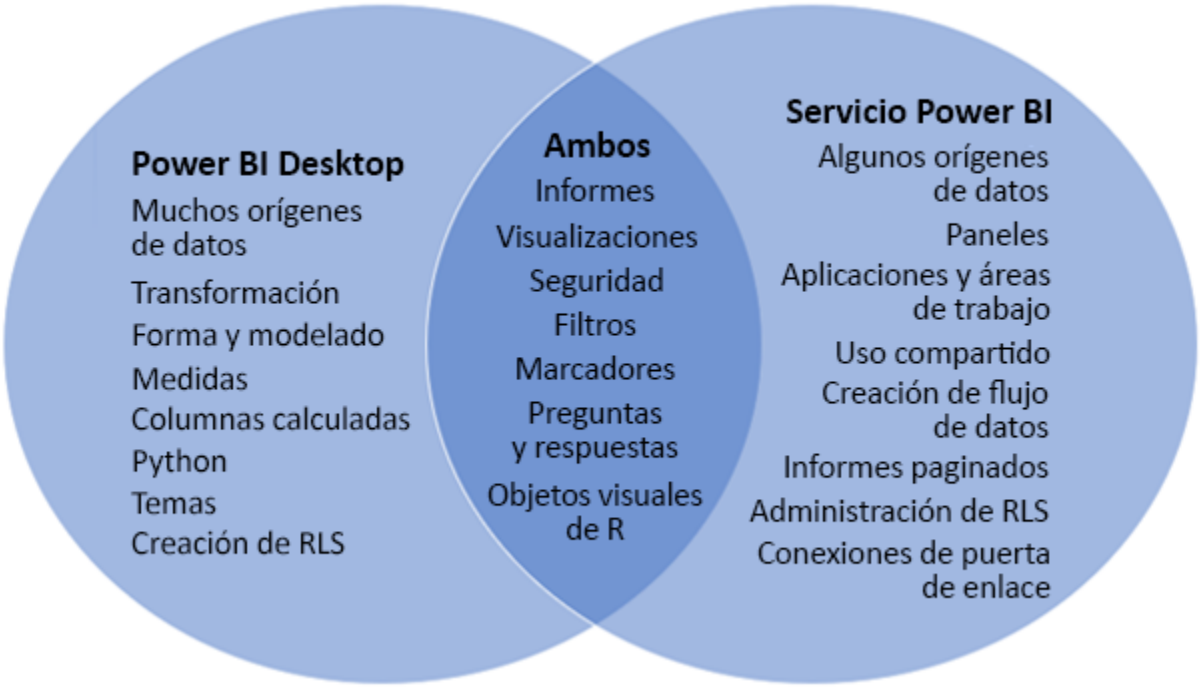


Ilustración 11Comparación de Power BI Desktop y el servicio Power BI

1.7.1.4 Servidor local de informes de Power BI

Power BI Report Server es una solución de instalación local que se implemente detrás del firewall en servidores propios de tal manera que provee una infraestructura dentro del ecosistema de servidores en la empresa que entrega los informes a los usuarios deseados de maneras diferentes, ya sea para visualizarlos en un explorador web, en un dispositivo móvil o como correo electrónico. Además, es compatible con Power BI en la nube, puede migrar a la nube cuando se desee.

2 Marco Metodológico y Análisis de Resultados

2.1 Metodología CRISP-DM

A continuación, se aplicará el desarrollo de los objetivos de esta tesis enfocándose en las fases en el marco metodológico de *CRISP-DM* según se menciona en el marco metodológico, se enumeran las diferentes fases según el documento original (crisp-dm.eu, 2015)

2.1.1 Comprensión del negocio.

A continuación, se describe cada tarea contenidas en la fase en el proceso de la minería de datos, con la finalidad de determinar los objetivos y requisitos del proyecto desde un punto de vista de negocios, de tal manera que más adelante pasen a ser en objetivos desde una perspectiva técnica y de planeación de proyectos.

Actores del proceso:

- Director de transporte público, dueño del proceso
- Departamento de transporte público, encargado de definiciones en términos de vialidad y reglas que rigen el sistema.
- Operadores de C.C.I.T.T. (Centro de Control Integrado de Tránsito Terrestre), encargados del monitoreo en tiempo real del sistema.
- Cooperativas de transporte. Agrupaciones de transportistas que poseen unidades de buses, con permisos de operación legalizados, son quien están obligados incorporar un GPS en cada una de sus unidades.
- Proveedores de Gestión de Flota, empresas que ofrecen el servicio de monitoreos de flotas y quienes dan el servicio de monitores a las cooperativas.

2.1.1.1 Determinar los objetivos del negocio.

- El objetivo de la inteligencia de negocios que se va a aplicar en este estudio es el de mejorar la perspectiva reactiva de la detección de Infracciones de tránsito basadas en telemetría de datos de transporte.
- El objetivo ulterior es proporcionar análisis fundado en inteligencia de negocios, que permita una mejor toma de decisiones basadas en indicadores y gráficos de concentración que permitan identificar proactivamente qué acciones deben tomar en zonas determinadas en situaciones específicas donde se concentren anomalías.

2.1.1.2 Contexto

La problemática de este proyecto está enmarcada en la ejecución del proyecto SITU de transporte público en la ciudad de Guayaquil, que controla 2600 buses transmitiendo en tiempo real cada 30 segundos. Con el propósito de determinar mediante reglas estructuradas y algoritmos diseñados para el cálculo de multas electrónicas de tránsito como herramienta para normar el ordenamiento vial de tal manera que debido al costo de la multa el transportista se abstenga de cometer una irregularidad.

Este proyecto está en ejecución desde junio del 2017, los primeros meses se realizó el registro de las unidades de transporte a la plataforma tecnológica llamada AURORA TP, que recibe los datos de ubicación permitiendo empezar el estudio preliminar de la información, lo cual permitió determinar los primeros umbrales de violación de excesos de velocidad para diciembre del 2017 que dieron cuenta del estado situacional en ese entonces.

Se detectó en ese momento como valores iniciales alrededor de 15000 infracciones de velocidad de por día, razón que fue comunicada al gremio del transporte, a través de un proceso de socialización que buscaba reducir estos umbrales exagerados de violaciones de

tránsito, paralelamente se diseñan algoritmos de detección de infracciones cuyo propósito es generar el insumo necesario para la generación de multas electrónicas basadas en las reglas dispuestas por la autoridad de tránsito.

2.1.1.3 Objetivos del negocio

Los objetivos del negocio están alineados a el mejoramiento en la detección de multas de tránsito a través del uso y explotación de herramientas de inteligencia de negocio.
Se basan en los objetivos específicos de esta tesis descritos al inicio de este documento

El alcanzar estos objetivos hará posible pasar de un control reactivo de multas de tránsito a una visión más proactiva del proceso, dando la posibilidad que, en base a patrones de concentración y agrupamiento, más las herramientas visuales de explotación de datos de la inteligencia de negocios, tomar mejores decisiones y en relación las causas y posibles regulaciones que se puedan hacer por mejorar el proceso en su conjunto.

2.1.1.4 Evaluación de la situación.

La infraestructura actual está basada en tecnologías de Microsoft, la fuente primaria de datos es SQL server 2017, donde residen los datos de alertas e infracciones detectadas por el sistema en tiempo real desde agosto del año 2017 hasta la fecha, los datos de movimiento de las unidades de transporte también están guardados en estos servidores y dependiendo de ciertos análisis serán también accesados de ser necesario, corroborando con esto que se posee acceso autorizados a todos los datos

2.1.1.4.1 Inventario de recursos

Los recursos de Software que posee la organización

- Servidores Dell de Power Edge R430 Basados en Windows Server 2016
- Base de datos SQL Server 2017 Enterprise

2.1.1.4.2 Riesgos y contingencias

Riesgo	Probabilidad	Impacto	Contingencia
Falta de flujo	Baja	Medio	Línea de crédito corporativa
Fallas de hardware	Baja	Alta	Revisiones y mantenimiento constantes para evitar fallas de hardware, software y servidores de contingencia
Fallas en enlaces de datos	Baja	Alta	Triple enlace redundante de fibra óptica
Falla de suministro de energía	Baja	Alta	Ups redundantes, planta de energía
No disponibilidad de fuentes de datos	Baja	Alta	Realizar una copia de seguridad de la base de datos de manera local y otra en la

			nube para prevenir la falta de acceso
Retraso en planificación de entregables	Baja	Media	Realizar un Project / Excel y determinar tiempos de desarrollo considerando el tiempo de los entregables
Desconocimiento del equipo en el desarrollo de proyectos	Baja	Baja	Realizar las reuniones con las personas que tengan mayor experiencia en las áreas involucradas del proyecto

2.1.1.4.3 Costes y beneficios

Los costes son directamente proporcionales al número de licencias a adquirir, en el modelo de nube seleccionado cada licencia tiene un costo de $10 mensuales se planifica tener al menos 20 usuarios. Se cuenta con servidores que harán el trabajo de puerta de enlace así que no habrá inversión en equipamiento.

Los beneficios están relacionados con alcanzar los objetivos generales de este estudio, esto es el poder establecer un mejorar proactiva a través del uso de inteligencia de negocios basadas en Power BI

2.1.1.4.4 Determinar los objetivos de la minería de datos

- Establecer los indicadores que se desplegaran en el cuadro de mando
- Diseñar los componentes de la Implementación del cuadro de mando
- Aplicar el Diseño de cuadros de mando con Power BI conectado al escenario real de base de datos.

2.1.1.4.5 Realizar el plan del proyecto

El proyecto se dividirá en las siguientes etapas para facilitar su organización y estimar el tiempo de realización de este:

- Etapa 1: Análisis de la estructura de los datos y la información de la base de datos.

Tiempo estimado: 40 horas.

- Etapa 2: Ejecución de consultas para tener muestras representativas de los datos.

Tiempo estimado: 40 horas.

- Etapa 3: Preparación de los datos (selección, limpieza, conversión y formateo, si fuera necesario) para facilitar la minería de datos sobre ellos.

Tiempo estimado: 120 horas.

- Etapa 4: Elección de las técnicas de modelado y ejecución de estas sobre los datos.

Tiempo estimado 40 horas.

- Etapa 5: Análisis de los resultados obtenidos en y posible retroalimentación.
  Tiempo estimado: 40 horas.
- Etapa 6: Producción de informes con los resultados obtenidos en función de los objetivos de negocio y los criterios de éxito establecidos.

Tiempo estimado: 40 horas.

- Etapa 7: Presentación de los resultados finales.
  Tiempo estimado: 40 horas.

2.1.1.4.6 Evaluación inicial de herramientas y técnicas

La herramienta escogida para el desarrollo de los objetivos de este proyecto es Power BI cuya descripción está documentada en este estudio, esta herramienta se alinea con las tecnologías existentes en el entorno de aplicación del proyecto, dado que la empresa tiene estandarizada la plataforma Microsoft. Sin embargo, otros factores como la posibilidad de validar el modelo sin restricciones a través de una cuenta personal en contraposición con otras tecnologías similares como Tableu o Qlick Sence las cuales requieren licenciamiento para su evaluación.

Además, factores como la simplicidad en la integración de datos, la disponibilidad de documentación publicada y la fácil implementación de los escenarios de visualización hacen de esta una herramienta deseable para este estudio, sin dejar de ser una opción más, debido a que la real importancia radica en los datos y el modelo a implementarse, lo cual es independiente de la tecnología a usarse en este estudio

2.1.1.5 Comprensión de los datos

En esta fase se realizará el aspecto metodológico concerniente a la recolección de la datos e información como paso inicial para alcanzar los objetivos de la inteligencia de negocios, este paso permite familiarizarse con la información su estado, calidad establecer los pasos primarios para ir construyendo la estructura técnica para alcanzar el propósito buscado.

2.1.1.6 Recolectar los datos iniciales

Los datos utilizados en este proyecto están relacionados a los datos de detección de multas de tránsito y su relación entre las unidades de transporte público que las generan en tiempo real y las cooperativas y socios a las cuales pertenecen.

2.1.1.7 Descripción de los datos.

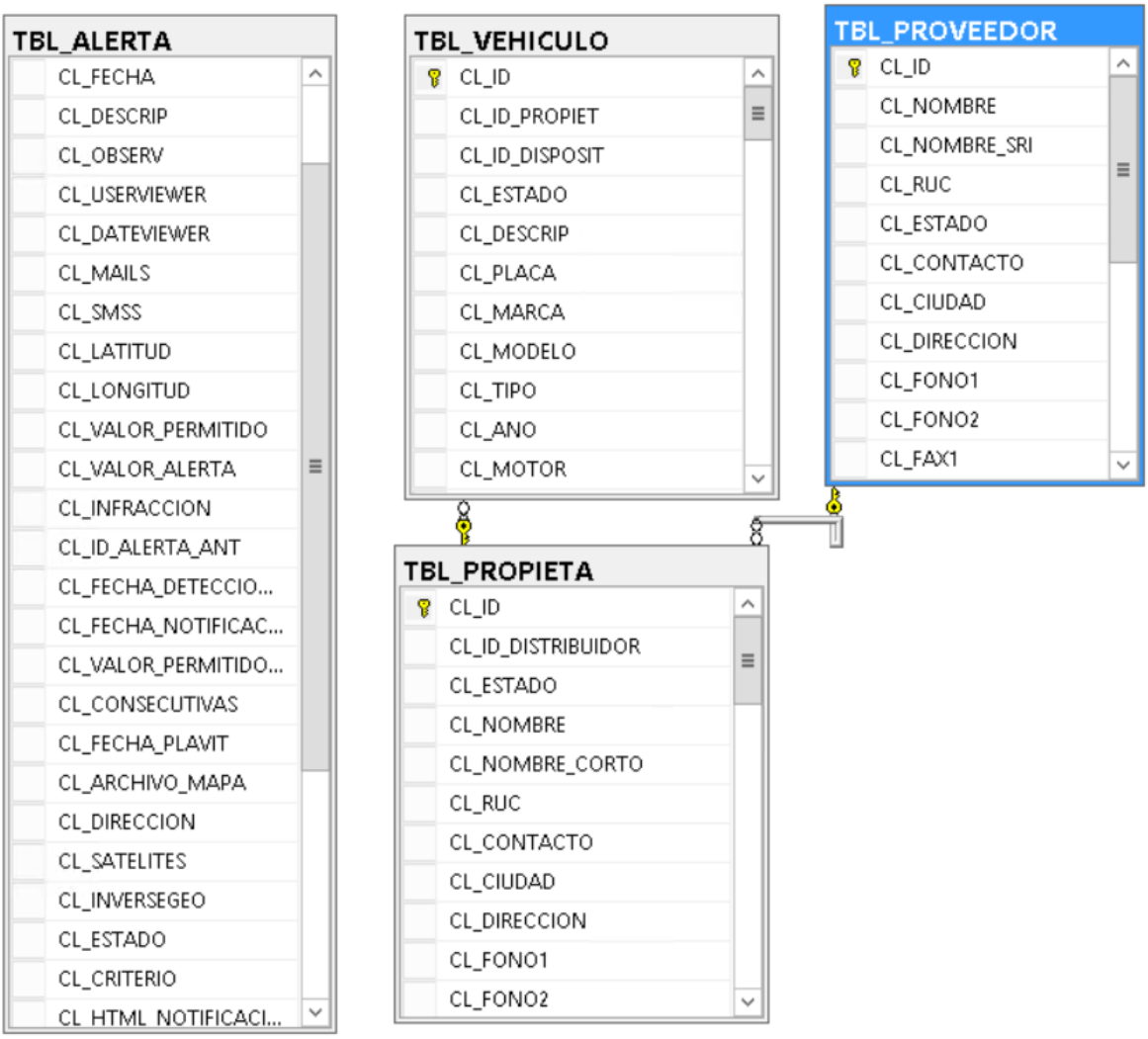


Ilustración 12 Diccionario de datos

2.1.1.7.1 Tabla de Vehículos – tbl_vehiculos

- CL_ID Tipo numérico, Este campo es un número que identifica a cada vehículo.
- Cl_ID_PROPIET, Tipo numérico, Campo identificador de la cooperativa a la que pertenece el vehículo.
- CL_ID_DISPOSITIVO, Campo Alfanumérico, identifica el IMEI o identificador del dispositivo de rastreo.
- CL_PLACA Tipo alfanumérico, placa del vehículo.
- CL_TIPO, Tipo alfanumérico, contiene el valor de tipo de vehículos ejemplo 'BUS, MOTO, BUS ELECTRICO, etc.'

2.1.1.1.1 Tabla de Cooperativas - tbl_propieta

- CL_ID Tipo numérico. Este campo es un número que identifica a cada cooperativa.
- CL_ID_DISTRIBUIDOR ID Tipo numérico, Campo identificador del proveedor de servicios de GPS
- CL_ESTADO Tipo Varchar , a = Activos,
- CL_NOMBRE Tipo Varchar , descripción de ott
- CL_NOMBRE_CORTO Tipo Varchar , descripción de corta
- ,CL_RUC Tipo Varchar ,,Registro único del contribuyente. ,

2.1.1.7.2 Tabla de proveedores de servicio - tbl_proveedor

- CL_ID Tipo numérico. Este campo es un número que identifica a cada proveedor.
- ,CL_NOMBRE Tipo Varchar , descripción de proveedor
- ,CL_NOMBRE_SRI Tipo Varchar , Nombre en el SRI de proveedor

- ,CL_RUC Tipo Varchar ,Registro único del contribuyente
- ,CL_ESTADO Tipo Varchar A=Activo

2.1.1.7.3 Tabla de alertas de tránsito - tbl_alertas

- CL_ID Tipo numérico. Este campo es un número que identifica a cada alerta.
- CL_ID_DISPOSITIVO Tipo numerico, código del dispositivo telemétrico
- CL_TIPO Tipo numérico, tipo de multa
- CL_FECHA,Tipo Fecha, fecha de la alerta
- CL_DESCRIP,Tipo Vqrchar, descripción de la alerta
- ,CL_OBSERV Tipo Vqrchar, Observación de la alerta
- CL_LATITUD, Tipo Numérico, longitud del evento
- CL_LONGITUD, Tipo Numérico, latitud del evento
- CL_VALOR_PERMITIDO Tipo Numérico, Valor de tolerancia de la alerta
- CL_VALOR_ALERTA Tipo Numérico, Valor real de la alerta
- CL_INFRACCION Tipo Numérico, Es o no Infracción

2.1.2 Exploración de los datos

Una vez que se han descrito los datos, se procede a explorarlos, esto implica aplicar pruebas estadísticas básicas que revelarán propiedades de los datos, y crear tablas de frecuencia y gráficos de distribución de los datos. Este informe sirve principalmente para determinar la consistencia y completitud de los datos. Mediante consultas SQL, cabe recalcar que esto no corresponde a un indicador específico sino a una herramienta de exploración simple, se han obtenido los datos necesarios para realizar las gráficas que se explican a continuación:

Exploración de datos de proveedores de consultado a través de un histograma simple con el fin de validar la consistencia y hacer un primer acercamiento s los datos estructurados

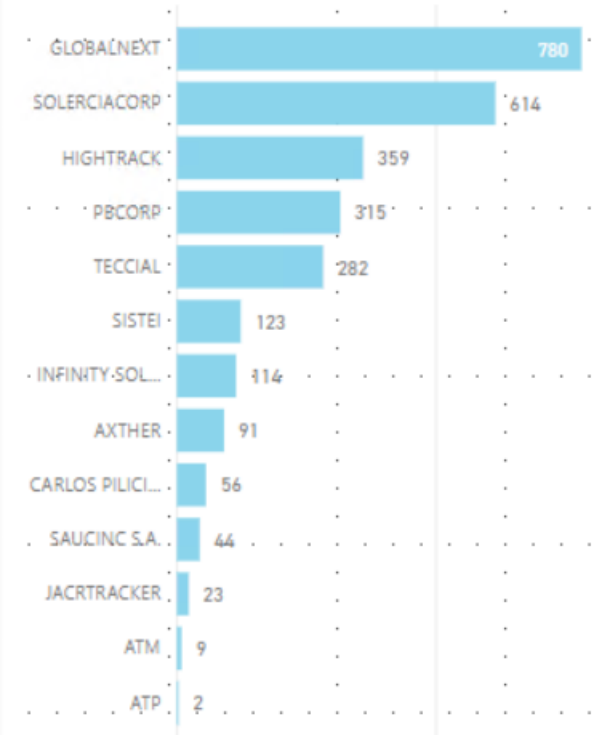


Ilustración 13 Exploración de Unidades de transporte publico Registradas por proveedor de sistemas de flotas

Consulta de numero unidades, asociadas a un proveedor segmentado por el tipo de vehículo asociado, usando un histograma segmentado.

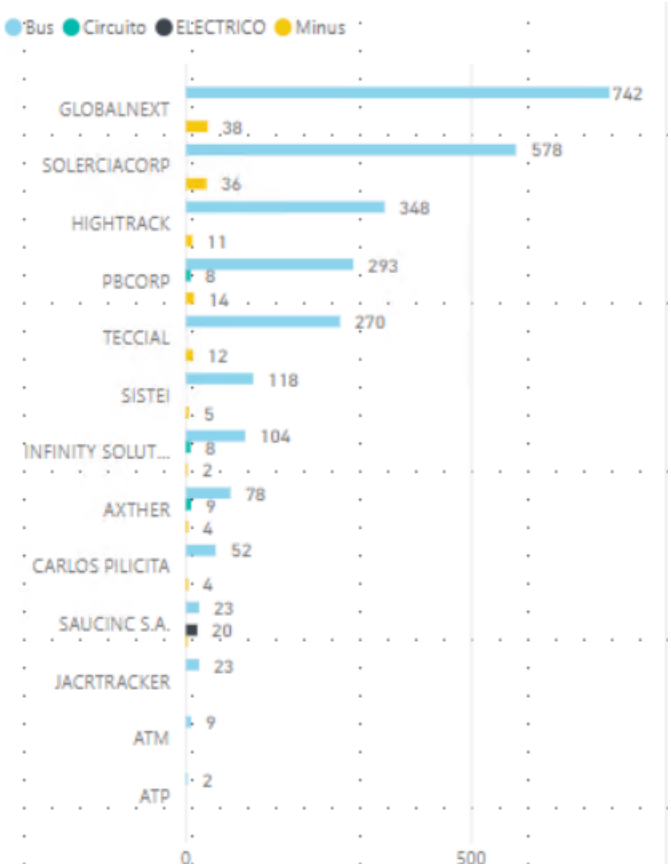


Ilustración 14 Unidades por proveedor y tipo

Consulta de número de vehículos asociados a una cooperativa de transporte usando un gráfico de pastel

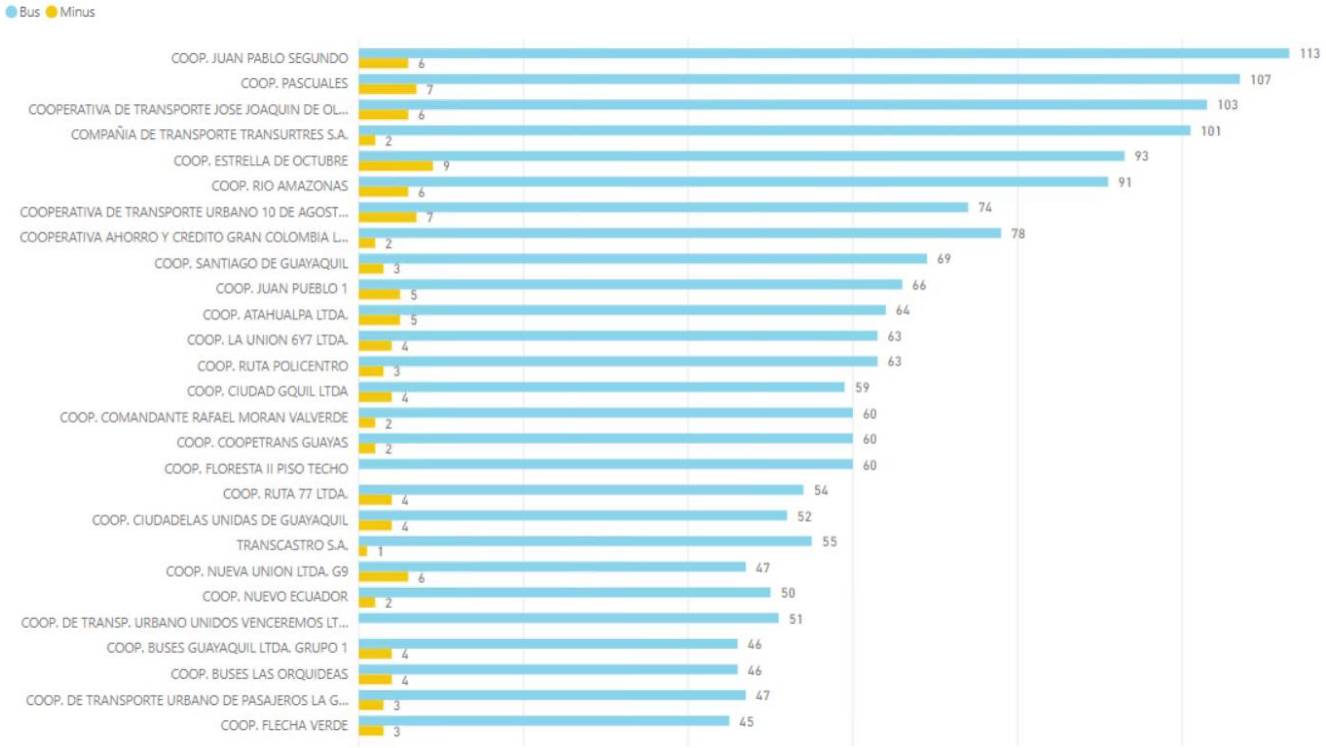


Ilustración 15 Unidades por cooperativa

2.1.2.1 Verificar la calidad de los datos

Después de realizar la revisión inicial de los datos podemos confirmar la consistencia de la información debido a que estos mismos datos son el insumo que se envía al sistema de multas de la autoridad de tránsito, razón por la cual pasa por proceso previos de verificación exhaustivo, se puede afirmar que estos son consistentes y válidos. Desde la fuente original de los datos de telemetría de GPS. si se encuentran datos con valores aberrantes como velocidades exageradas o fechas inconsistentes, pero toda esta información se vuelve irrelevante debido a que previamente esta información ya ha sido filtrada y el proceso se diseñará para usar la información ya procesada de alertas y multas de tránsito. .

2.1.2.2 Preparación de los datos.

En esta fase de la metodología se trata de preparar los datos para adecuarlos a las técnicas de minería de datos que se van a emplear sobre ellos. Esto implica seleccionar el subconjunto de datos que se va a utilizar, limpiarlos para mejorar su calidad, añadir nuevos datos a partir de los existentes y darles el formato requerido por la herramienta a usarse en este caso Power BI.

2.1.2.3 Seleccionar los datos

El insumo principal para usarse en este estudio son los maestros de operación de cooperativas y buses juntamente con la información correspondiente a multas de tránsito

Las entidades para usarse son las antes descritas y serán incluidas en Power Bi Desktop

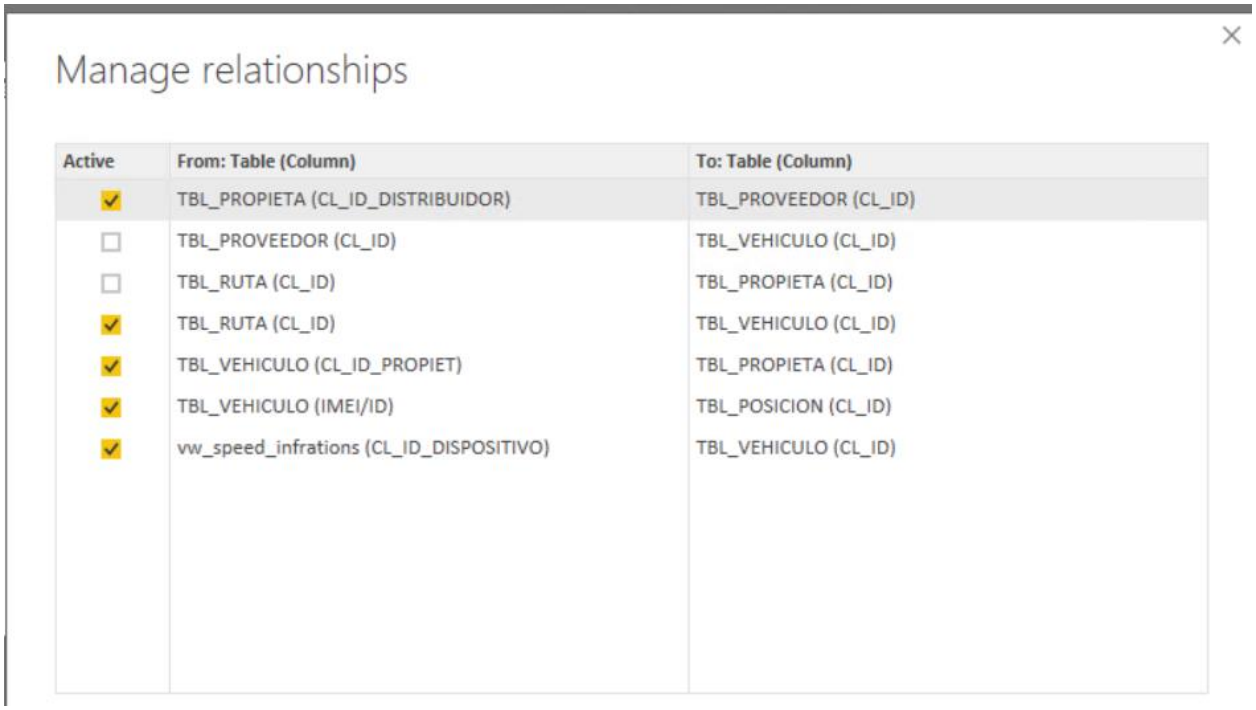


Ilustración 16 Diagrama de entidades involucradas

2.1.2.4 Limpiar los datos.

Los datos se encuentran verificados y limpios ya que se les ha dado tratamiento previamente, sin embargo, ciertas técnicas de unión y filtrado en lo que se refieren a infracciones se aplicarán debido a él volumen de datos se pasan a archivos históricos, sin embargo, se unen y se integran como una vista única para que el modelo las vea como tal.

Sin embargo, una de las bondades del uso de Power BI es que se puede limpiar y transformar los datos usando el editor de consultas integrado. Con el editor de consultas puede realizar modificaciones en los datos, como cambiar un tipo de datos, quitar columnas o combinar datos de varios orígenes, de esta manera se importaron todas las tablas involucradas en el proceso y se irá redujo el modelo hasta solo dejar lo útil y necesario, así como ir transformando otras columnas para diferentes propósitos de filtrado.

2.1.2.5 Calidad de los datos.

2.1.2.5.1 Completitud

Se valida que todos los datos estén completos antes de que, entre el proceso de evaluación de multas, previamente los datos han sido validados antes de enviarse al sistema de gestión de multas de la autoridad de tránsito.

2.1.2.5.2 Conformidad

Los datos que están en los campos de la tabla están en un formato estándar y legible y han sido validados antes de enviarse al sistema de gestión de multas de la autoridad de tránsito.

2.1.2.5.3 Consistencia

Se ha verificados que, al hacer el cruce de información con los registros, no existe información contradictoria.

2.1.2.5.4 Precisión / Exactitud

Los datos no son precisos, estos no pueden ser utilizados. En este sentido, para detectar si estos son precisos, se compara el dato con una fuente de referencia.

2.1.2.5.5 Duplicación

Se valida que cada dato se único en su valor y existencia en el tiempo, cada dato es único e irrepetible

2.1.2.5.6 Integridad

Se comprueba que lo datos tiene integridad referencial y pueden ser utilizados con nivel de confianza elevado.

2.1.2.6 Construir los datos.

La generación primaria de los datos se da en el proceso de generación de multas, sin embargo, se usa varias técnicas para segmentar y filtrar la información. Debido a que se tiene información transaccional e histórica Se construye una entidad de alertas e infracciones que,
una la tabla histórica con la tabla en uso para usar el modelo.

2.1.2.7 Integrar los datos.

La integración para utilizarse en la ingesta de datos mediante Power BI que admite una amplia gama de fuentes de datos. Puede hacer clic en Obtener datos y le muestra todas las conexiones de datos disponibles. Le permite conectarse a diferentes archivos planos, bases de datos SQL y la nube de Azure o incluso plataformas web como Facebook, Google Analytics y objetos de Salesforce. También incluye la conexión ODBC para conectarse a otras fuentes de datos ODBC, que no están en la lista.

Así conectamos directamente usando el conector a SQL Server, e integramos las tablas usarse directamente en el modelo.

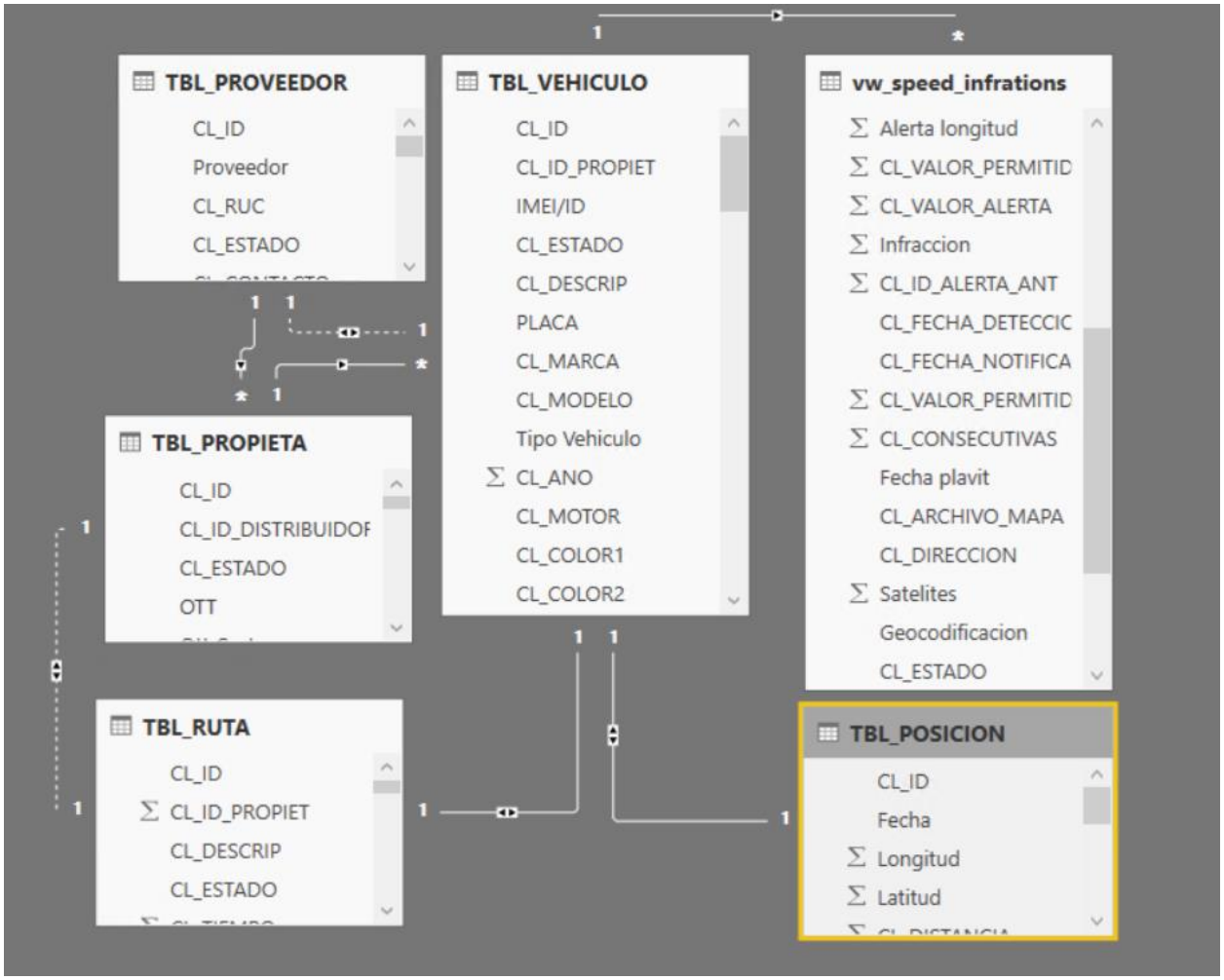


Ilustración 17 Modelo de Datos en Power BI

2.1.2.8 Formateo de los datos

Una característica de la herramienta Power BI Desktop es que puede conectarse a muchos tipos diferentes de orígenes de datos y después darles forma a los datos para satisfacer las necesidades del proyecto, lo que permite crear informes visuales que se pueden compartir con los demás. Formatear los datos significan aplicar técnicas de transformación, por ejemplo, al cambiar el nombre de columnas o tablas, convertir texto en números, quitar filas, configurar una primera fila como encabezado, etcétera. Además, Combinación de datos significa conectarse a dos o más orígenes de datos, darles forma según sea necesario y después consolidarlos en una consulta usable a los propósitos del problema en cuestión.

2.1.3 Modelado

En esta fase de la metodología se escogerá las técnicas apropiadas para los objetivos definidos en este estudio de Inteligencia de negocios. Con las funcionalidades de modelado de Power BI, puede conectarse a varios orígenes de datos y combinarlos de tal forma que se cree un modelo de datos único que se adapte a estas necesidades.

2.1.3.1 Escoger la técnica de modelado.

Una vez importados y transformados los datos, se empieza a modelar según las características que ofrece la herramienta para dicho propósito.

Uno de los puntos fuertes de Power BI radica en que no hace falta aplanar los datos para convertirlos en una tabla. En su lugar, puede utilizar varias tablas de diversos orígenes y definir la relación entre ellas. También puede crear sus propios cálculos personalizados y asignar nuevas métricas para ver segmentos concretos de los datos, así como usar estas nuevas medidas en las visualizaciones a fin de lograr un modelado sencillo, sin embargo, el modelo utilizado es el conocido como “Copo de nieve”, debido a que las tablas de hechos tienen múltiples relaciones.

2.1.3.2 Generar el plan de prueba.

Una vez conectados los datos al escenario real, se procederá a generar la carga y medir la afectación contra las bases trasnacionales.

2.1.3.3 Construir el modelo.

2.1.3.3.1 Indicadores de situación

Este indicador permitirá ver el estado de registro de cooperativas, proveedores y vehículos y sus diferentes combinaciones, dando una visualización de cómo está la operación y el registro del transporte público en un solo conjunto

2.1.3.3.2 Indicadores de infracción

El diseño de estos indicadores debe converger en el análisis de la situación de las tendencias en las anomalías del transporte público, de tal manera que se pueda identificar de manera más eficiente las concentraciones y dispersiones de las situaciones anómalas en el proceso en estudio, a continuación, se definirán varios escenarios basados en este objetivo.

2.1.3.3.3 Producción de Multas de Velocidad.

Este indicador tiene como propósito hacer un comparativo de las multas producidas en un periodo de tiempo anual y ver el resultado del impacto del control en el tiempo.

2.1.3.3.4 Multas de Velocidad por cooperativas

El poder analizar la tendencia de las cooperativas más infractoras visualizadas con una dimensión de tiempo anual y mensual, data la posibilidad de usar este escenario para generar controles específicos enfocados en estos operadores de transporte público.

2.1.3.3.5 Multas de Velocidad por Vehículo

Ese indicador deberá mostrar las diferentes dimensiones de tiempo anuales y mensuales de las unidades más infractoras y sus tendencias, de tal manera que permita identificar los vehículos que debe ser controlados específicamente por tener un indicador de recurrencia muy alto en el tiempo

2.1.3.3.6 Mapa de calor de concentración de Multas de Velocidad

Usando las diferentes dimensiones de tiempo, este indicador deberá dar una visualización geográfica de tendencia, esclareciendo las zonas en que se concentran las multas, de tal manera que, facilitar la detención de las áreas geográficas que deben ponerse atención, para determinar entre otras cosas si hubiese una falta de señalética o algún otro factor que estimule el incumplimiento de las normas de tránsito en estos sectores.

2.1.3.4 Evaluar el modelo

La evaluación en terminas de inteligencia de negocios demuestra su utilidad al mostrar varios escenarios que denotan agrupación de comportamientos y la efectividad en el tiempo que tiene la ejecución de los controles aplicados.

2.1.4 Evaluación

En esta fase de la metodología se intentan evaluar los modelos generados, pero en esta ocasión la evaluación se hace desde el punto de vista de los objetivos de negocio en lugar de los objetivos de minería de datos. Una vez realizada esta evaluación, se debe decidir si los objetivos han sido cumplidos y de ser así se puede avanzar a la fase de implantación, de lo contrario se tendría que identificar cualquier factor que se haya podido pasar por alto y hacer una revisión del proceso.

2.1.4.1 Evaluar los resultados

El proceso será evaluado con el personal de transporte público de la autoridad de tránsito

2.1.4.2 Revisar el proceso.

Se pondrá en consideración del departamento de transporte público la revisión de los indicadores.

2.1.4.3 Determinar los próximos pasos

Una vez validado el modelo el siguiente paso es la implementación del proyecto a una escala práctica, cómo se describirán en la sección siguiente.

2.1.5 Despliegue o implantación

2.1.5.1 Planear la implantación.

El plan de implementación involucra la serie de pasos necesarios para llevar a cabo la instalación de la plataforma de BI sobre la que se monta este proyecto, Power BI como la plataforma escogida será instalada al menos en dos instancias una instancia de servidor donde se dejará configurados los escenarios preconstruidos con acceso a la red donde está la base de datos de donde se toma la información, el diseñador de los cuadros de mando tendrá instalado Power BI, como herramienta de construcción de escenarios, y de manera calendarizada se subirá la información a la nube de Power BI, donde los usuarios funcionales obtendrán la información.

Este modelo de integración es posible debido a que Power BI permite usar el entorno de manera gratuita y otorga un espacio personal de 10 gigas para uso y validación del

modelo, sin embargo, esto llevado a términos de ambiente en producción involucra adquirir las licencias correspondientes, según el número de usuarios requeridos.

2.1.5.2 Instalación de del entorno.

A continuación, se describe los procesos necesarios para la instalación de Power BI dentro del contexto a utilizarse en la validación de este modelo:

2.1.5.2.1 Instalar Power BI Desktop

Se debe descargar e instalar Power BI Desktop, no hay nada que configurar, más allá marcar que aceptar el acuerdo de licencia y elegir en la carpeta se va a realizar la instalación. Pero hay ciertos aspectos a tener en cuenta, como la elección de la versión de 32 bits o 64 bits.

2.1.5.2.2 Instalar Power BI Gateway.

En este inciso mostrarán los pasos a seguir para la correcta configuración de Gateway en Power BI, y la actualización calendarizada de información de las fuentes de datos utilizando este método.

2.1.5.2.3 Acceder desde la nube.

Para los usuarios clave de este proceso se proveerá un acceso en línea en el sitio de Power BI al cual se accede con las credenciales correspondientes y accesos de acuerdo a los perfiles que posean cada uno.

2.1.5.2.4 Planear la monitorización y mantenimiento.

El plan de mantenimiento persigue garantizar la correcta evolución del actual sistema de BI para seguir en un esquema de mejoramiento continuo y alta disponibilidad:

Desde el Punto de Vista del uso de la Tecnología:

- Evolucionar y mantener la herramienta de explotación y análisis de la información de las áreas funcionales que forman parte de la organización y sus órganos de gestión.
- Fomentar el uso de Power BI y los cuadros de mando como complemento de ayuda a la toma de decisiones.
- Incrementar y mantener el sistema de BI, para integrar toda la información generada en los diferentes sistemas de información disponibles y permitir así el análisis de dicha información a las personas clave de la organización.
- Conseguir hacer del BI la herramienta de referencia para el análisis de la información, por encima de otras herramientas de uso personal, a través de la capacitación y el soporte.
- Conseguir la difusión del conocimiento generado del BI a toda la organización.

Desde el punto de Vista de los Servicios:

- Monitoreo constante de los servicios de puerta de enlace para la réplica de información en la nube, que estén activos.

- Monitoreo de capacidad de los servidores de BI y si disponibilidad 24/7.

- Administración de los servicios en la nube

2.1.5.3 Producir el informe final

En este paso se presentará un informe resumiendo los puntos importantes del proyecto y la experiencia adquirida durante su desarrollo. El público al que va dirigido este informe sería el personal de transporte público de la autoridad de tránsito de tal manera que se pueda estudiar la situación actual y tomar medidas correctivas para la mejora del servicio. Cabe decir que parte de este informe final será presentado de manera oral con una presentación, por lo que en este apartado solamente documentamos este breve resumen.

2.1.5.4 Revisar el proyecto

En esta última etapa de la metodología se debe hacer una evaluación de aquellas cosas que se hicieron correctamente y aquellas que no, así como posibles mejoras para que en las futuras ejecuciones de inteligencia de negocios se vayan retroalimentando el modelo y se obtengan mejores resultados.

2.2 Metodología de los 3 Ciclos

2.2.1 Entendimiento del Negocio

2.2.1.1 Instrucción a la metodología

¿Quién**?**

ATM, Empresa Pública Municipal de Tránsito de Guayaquil, EP

¿Qué**?**

Creada en Julio del 2012 por el Municipio de Guayaquil con la Misión de establecer y ejecutar políticas para implementar un sistema integrado de regulación, control, seguridad vial y la

preservación del medio ambiente acorde a la Constitución de la República y al Código Orgánico de Organización Territorial, Autonomía y Descentralización. (ATM, n.d.)

¿Cómo**?**

Estableciendo y ejecutando políticas para implementar un sistema integrado de regulación control y gestión del tránsito, transporte terrestre y seguridad vial, a través del uso de tecnología (sistema inteligente de manejo de tráfico), que permita la integración armoniosa de todos los entes demandantes de movilidad que redunde en mejorar la calidad de vida y la preservación del medio ambiente.

¿Dónde**?**

Ciudad de Guayaquil

¿Cuándo**?**

El proyecto tiene como alcance Agosto del Año 2017, donde se empezó la implementación del proyecto de la Plataforma de Monitoreo del sistema SITU.

¿Por qué?

Desarrollar un cuadro de mando, para la identificación de anomalías en el transporte público utilizando la herramienta de inteligencia de negocios Power BI, para mejorar la identificación de incumplimientos de las normas de tránsito.

2.2.1.2 Desarrollo

El consorcio Aurora Transporte Público creado para la implementación del proyecto de la plataforma de monitoreo de SITU de transporte público en la ciudad de Guayaquil, que controla 2600 buses transmitiendo en tiempo real cada 30 segundos. Con el propósito de determinar mediante reglas estructuradas y algoritmos diseñados para el cálculo de multas electrónicas de tránsito como herramienta para normar el ordenamiento vial de tal manera que debió a costo de la multa el transportista se abstenga de cometer una irregularidad.

Este proyecto está en ejecución desde junio del 2017, los primeros meses se realizó el registro de las unidades de transporte a la plataforma tecnológica llamada AURORA TP,

que recibe los datos de ubicación permitiendo empezar el estudio preliminar de la información, lo cual permitió determinar los primeros umbrales de violación de excesos de velocidad para diciembre del 2017 que dieron cuenta del estado situacional en ese entonces.

Los objetivos del negocio están alineados a el mejoramiento en la detección de multas de tránsito a través del uso y explotación de herramientas de inteligencia de negocio, estos objetivos se basan en los objetivos específicos de esta tesis:

- Establecer los indicadores que se desplegaran en el cuadro de mando
- Desarrollar la interfaz de visualización en Power BI aplicando reglas e indicadores de comportamiento.
- Aplicar la propuesta en un escenario real.

.

2.2.2 Entendimiento de los objetivos:

2.2.2.1 Introducción

¿Quién**?**

Departamento de transporte público de ATM

¿Qué**?**

Los indicadores que se presentan en este estudio son basados en dos aspectos:

Indicadores de situación

Indicadores de Infracción

Infracciones de velocidad General

Infracciones de velocidad por Operador de Transporte

Infracciones de velocidad por Unidades

Infracciones de velocidad por Proveedores

¿Cómo**?**

Aplicando herramientas de inteligencia de negocios y cálculo de indicadores, con herramienta Microsoft Power BI.

¿Dónde?

Unidad de sistemas de Consorcio Aurora Transporte Público

¿Cuándo**?**

Se implementará en 30 días.

¿Por qué**?**

Se requiere una solución de inteligencia de Negocios para la visualización de los indicadores y toma de decisiones orientadas a disminuir proactivamente los índices de violación de normas de tránsito.

Establecer los indicadores que se desplegaran en el cuadro de mando de comportamiento del transporte público en la ciudad de Guayaquil

El propósito del cuadro es contar con Indicadores sobre:

- Estado de situación del proyecto
- Infracciones de velocidad en General
- Infracciones de velocidad por Operador de Transporte
- Infracciones de velocidad por Unidades
- Infracciones de velocidad por Proveedores

La información será obtenida del Sistema denominado AURORATP, el cual es la plataforma tecnológica del proyecto público SITU de la Autoridad de tránsito Municipal de Guayaquil. Se cuenta con información desde el agosto del 2017 al 2019 para efectos de la construcción del Cuadro de Mando.

Para construir el cuadro de mando se tendrá que realizar el entendimiento de los siguientes datos:

- VEHÍCULOS
- COOPERATIVAS DE TRANSPORTE
- PROVEEDORES DE SERVICIOS FLOTA
- MULTAS DE TRÁNSITO

Visualizar en una herramienta de inteligencia de negocios (Power BI) aplicando reglas e indicadores de comportamiento de transporte público

Para la correcta explotación de la información se utilizarán los diferentes objetos de Power Bi combinados

Aplicar la propuesta en un escenario real.

Los datos están conectados a los datos en tiempo real en la infraestructura del proyecto de la plataforma de monitoreo del proyecto SITU.

2.2.3 Preparación de datos

2.2.3.1 la Matriz Origen – Destino

ORIGEN	TRANSFORMACIÓN	DESTINO

Vehículos de transporte Público	id, Placa, propietario, Operador de transporte público,	se filtra solo vehículos activos	dim_vehiculos
Operadora de transporte	id, nombre, ruc	se filtra solo operadores de transporte activos	dim_ott
Proveedores de servicio Meses	id, nombre	se filtra solo proveedores activos	dim_proveedores dim_meses
Infracciones Velocidad	id, fecha, descripción, valor permitido, valor, infracción		hec_velocidad
Alertas botón de pánico			hec_panico

2.2.3.2 Modelado

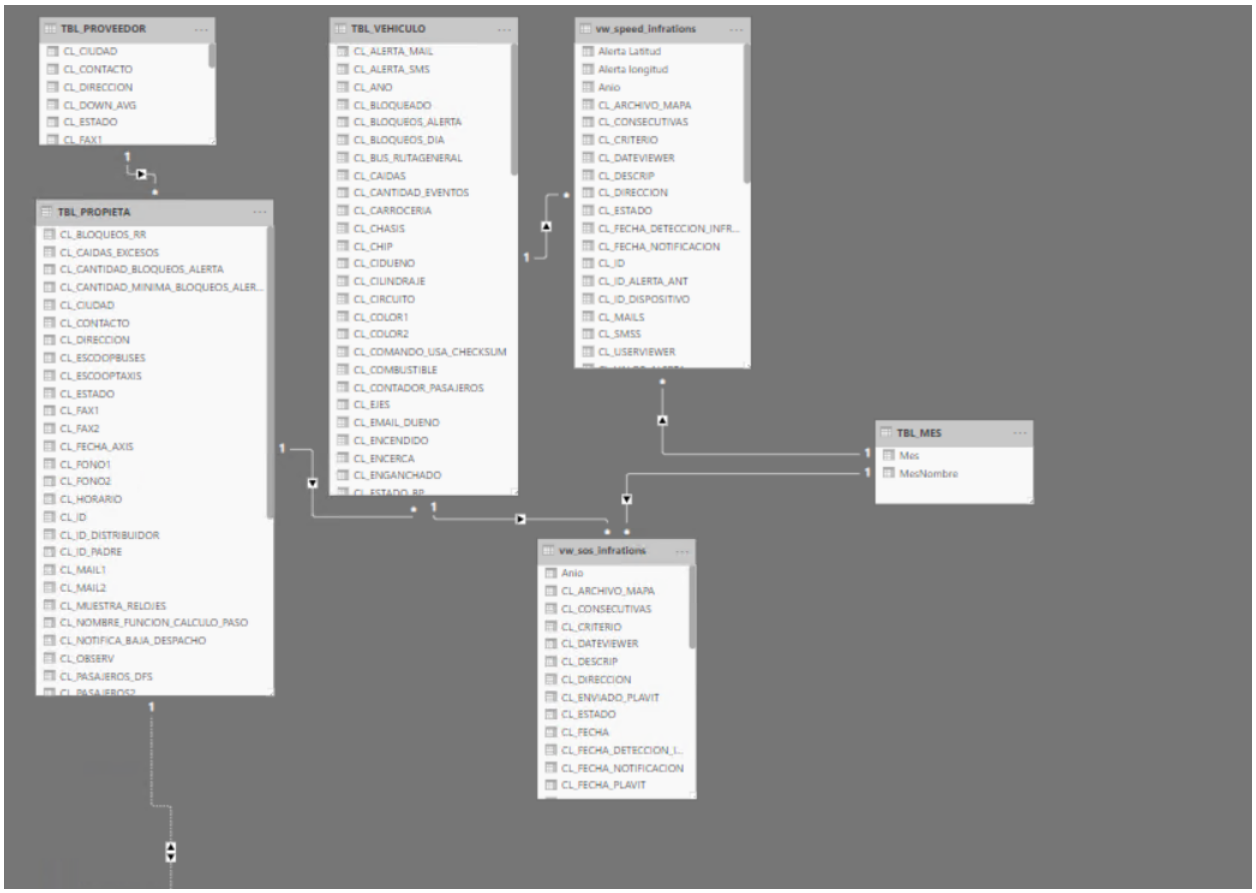


Ilustración 18 Modelo de datos implementado en Power BI

2.2.3.3 ETL

Se ha utilizados la conexión directa de a SQL server implementada en Power BI, para incluir el modelo y extraer los datos.

2.2.3.4 Despliegue

Operadores de transporte

CL_ID	CL_ID_DISTRIBUIDOR	CL_ESTADO	OTT	Ott Corto	CL_RUC	
1	1	A	CONSISTEI AURORA (pruebas)		1	
2	1	A	ATM		2	
3	2	A	COMPAÑIA DE SERVICIO DE TRANSPORTE ALTERNATIVO EN TRICIM	TRICIALTER S.A	0992760222001	L
4	2	A	COOPERATIVA DE AHORRO Y CREDITO GRAN COLOMBIA LTDA	COOPERATIVA AHORRO Y CREDITO GRAN COLOMBIA LTDA	0990162158001	
9	2	A	COOP. TRANS. EN TRICIMOTOS TRANSVERNORTE	COOPTRANSVERNOR	0993026158001	M
10	2	A	COMPAÑIA DE TRANSPORTE DE TRICIMOTO BASTION POPULAR S./	COMTRIBAP S.A.	0992790326001	R
15	2	A	COOP. TRANSPORTE EN TRICIMOTOS 28 DE FEBRERO	COOPTRIFEBRE	0993024392001	L
16	2	A	COOP. DE TRANSP. TRICIMOTOS 15 DE MAYO	COOPTRATRIMAY	0993028878001	M
17	2	A	COMPAÑIA DE TRICIMOTEROS ALFA Y OMEGA	COMTRIALFOME	0992905999001	M
19	2	A	COOP DE TRANSP EN TRICIMOTOS COSTA Y MAR COOPTRANSCOSI	COOPTRANSCOSMA	0993026646001	A
21	2	A	COOP. TRANS. EN TRICIMOTOS SHADDAY TOUR	COOPTRANSHATOUR	0993021636001	P
22	2	A	COOP DE TRANSP EN TRICIMOTO 3 DE DICIEMBRE COOPTRIDIC	COOPTRIDIC	0993028495001	U
23	2	A	COOPERATIVA TRANSPORTE PUBLICO DE PASAJEROS URBANO Y TU	COOP. TARQUI LTDA.	0990677948001	T
24	2	A	CALMORE S.A. CARLOS ALVEAR MOREIRA	CALMORE S.A.	0992459700001	

Ilustración 19 Listado de operadoras de transporte

Proveedores de Servicios de flota

CL_ID	Proveedor	CL_RUC	CL_ESTADO	CL_CONTACTO	
1	ATP	0993026654001	A	giancarlo Molina	
2	SISTEI	0992142758001	A	Sistei Solutions	
3	TECCIAL	0992154284001	A	TECHNOLOGY EQUINOCCIAL TECCIAL S.A.	
4	HIGHTRACK	0992773588001	A	NAVIATICORP S.A.	
5	INFINITY SOLUTIONS	0992918446001	A	Infinity Solutions	
6	JACRTRACKER	0912107836001	A	COCHEA REGALADO JOSE ANTONIO -JACRTRACKER	
7	SEAMTRACK	0992405112001	A	SEAMTRACK S.A.	
9	CARLOS PILICITA	1706261581001	A	CARLOS PILICITA	
10	AXTHER	0992638109001	A	Sistema de control AXTHER S.A.	
11	SOLERCIACORP	0992288590001	A	SOLERCIACORP	
12	ARTECHNOLOGIES S.A.	0992358009001	A	Antonio Ricaurte	
14	PBCORP	0910358076001	A	Carlos Paredes	
16	HOUNDTRACK	1	A		
18	GLOBALNEXT	0993035726001	A	Harry Potes	
20	SALICINC S.A.	0992155582001	A	Bryan Silva	

Ilustración 20 Listado de proveedores de servicios

Vehículos de Transporte público

CL_ID	CL_ID_PROPIET	IMEI/ID	EstadoVehiculo	Ruat/alias	PLACA	CL_MARCA	CL_MODELO	Tipo Vehiculo
2799	107	458403	A	1353	GBO2543	MERCEDES BENZ	OF 1721-59 6.0 4)	Bus
2264	103	458404	A	1008	GBN2648	VOLKSWAGEN	17210 OD	Bus
2266	103	458405	A	1023	GBN2400	VOLKSWAGEN	17.210 OD	Bus
2279	103	458406	A	0996	GBM0452	VOLKSWAGEN	17.210 OD	Bus
2278	103	458407	A	1036	GBN2595	VOLKSWAGEN	17210 OD	Bus
2281	103	458408	A	0986	GBN1016	VOLKSWAGEN	17.210 OD	Bus
2261	103	458409	A	1953	GAY0853	MERCEDES BENZ	OF 1620/60	Bus
2285	103	458410	A	1050	GBN2519	VOLKSWAGEN	17210 OD	Bus
2287	103	458411	A	0983	GBN1406	VOLKSWAGEN	17.210 OD	Bus
2294	104	458413	A	2374	TAU0892	MERCEDES BENZ	OF 1721-52	Bus
2277	103	458414	A	0994	GBN6339	MERCEDES BENZ	OF 1721-59	Bus
2272	103	458415	A	1129	GBN1323	MERCEDES BENZ	OF 1721-59	Bus
1637	89	458417	A	0537	GBG0631	VOLKSWAGEN	17210 OD BUS UR	Bus
1885	84	469246	A	1285	GBH0057	MERCEDES BENZ	OF 1721-59	Bus
2204	102	469247	A	0214	GBH0379	KING	DHZ1113KTD25	Bus
1827	91	469249	A	0961	GBN1077	MERCEDES BENZ	OF 1721-59	Bus
2762	112	469254	A	1742	GBM0516	MERCEDES BENZ	OF 1721-59 CHAS	Bus
1666	89	469341	A	2463	GBO2032	MERCEDES BENZ	OF 1721-59 6.0 3F	Bus

Ilustración 21 Listado de Vehículos registrados

3 CAPÍTULO 3. VALIDACIÓN

3.1 Validación de los resultados.

Una vez diseñados los escenarios de paneles de control se procedió a presentar el esquema para validación de los interesados para su respectiva revisión

3.2 Esquema de validación

Se propuso indicadores base de acuerdo con un análisis conjunto con el personal de desarrollo de la plataforma SITU tales como estado de registro, indicadores de Inflación desde la perspectiva

de los proveedores, operadoras de transporte y vehículos, así como mapas de calor de infracciones basados en agrupación geográfica.

Una vez integrado el modelo de datos, diseñados e implementada la herramienta de inteligencia de negocios, los cuadros de mando implementados muestran la información de forma íntegra, clara, precisa, de manera adecuada y funcional, facilitando la visualización, permitiendo manejar indicadores adecuados para la toma de decisiones proactiva.

Con una metodología apropiada y accesible como Crisp-dm cuyo nivel de detalle y formalidad es adecuado en proyectos públicos, en conjugación con Power BI y su modelo de datos asociativo que flexibiliza la construcción de aplicaciones y tableros de control, por su simplicidad, rapidez y efectividad.

Aplicación y descripción de los resultados

3.2.1 Indicadores de situación

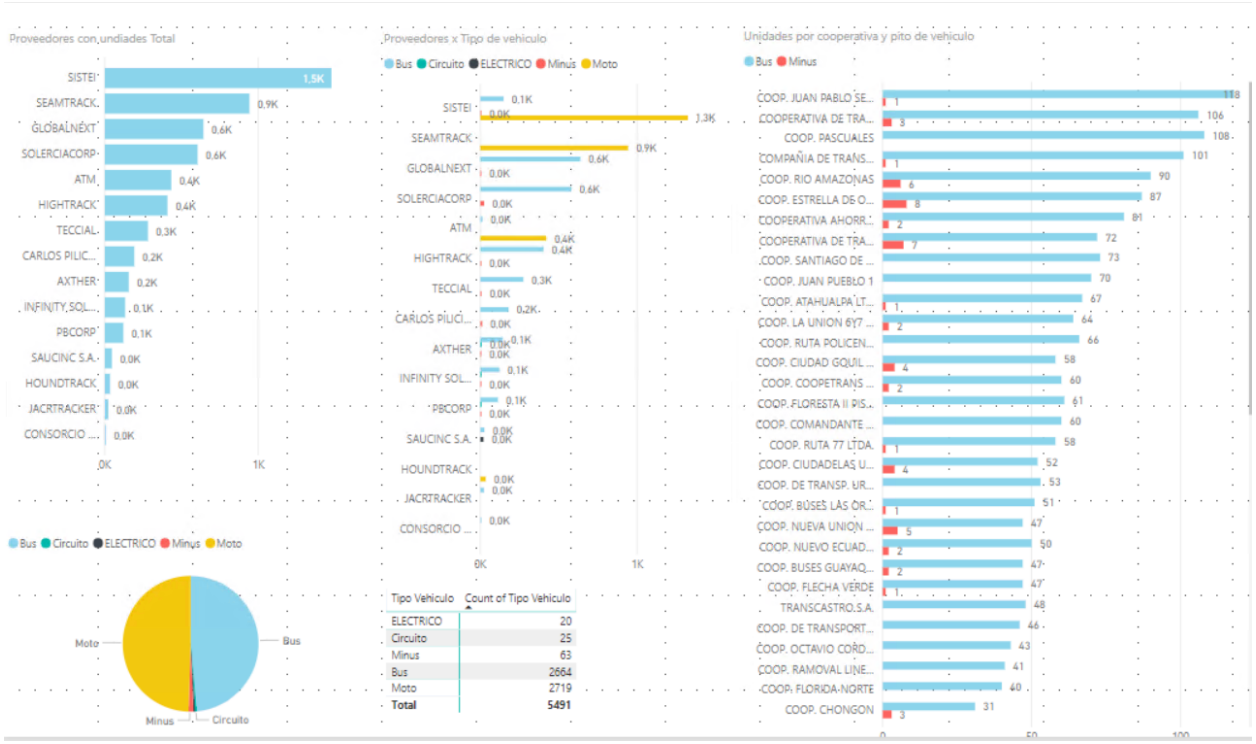


Ilustración 22 Cuadro de mando de estado de Registro

3.2.2 Producción de Multas de Velocidad.

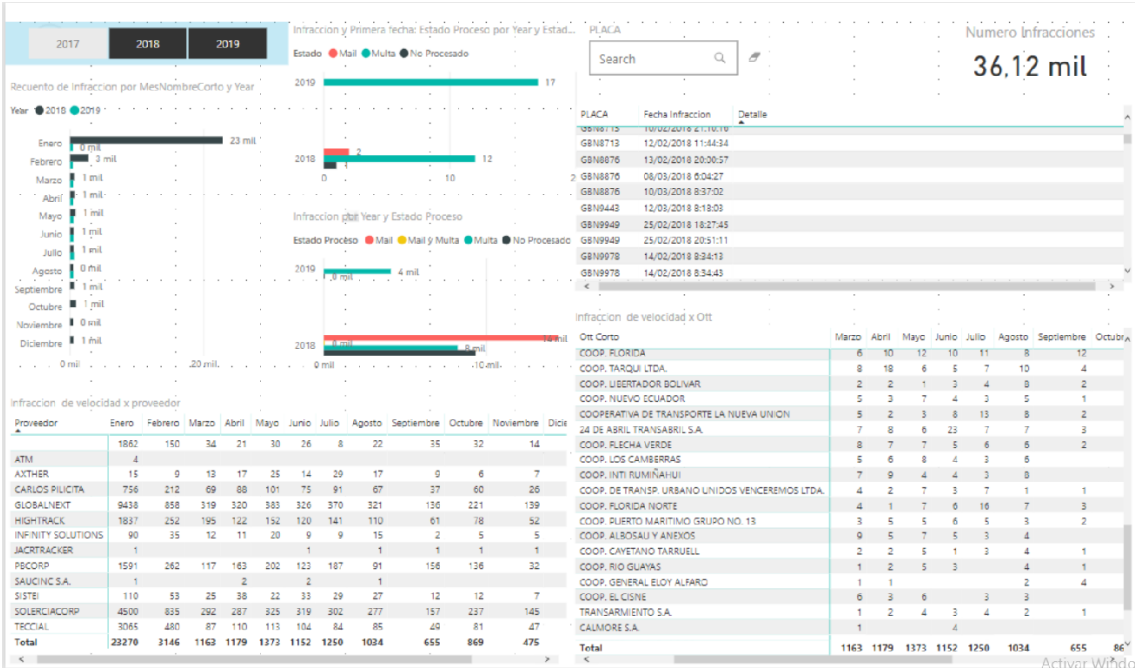


Ilustración 23 Producción de multas de velocidad

3.2.3 Multas de Velocidad por cooperativas

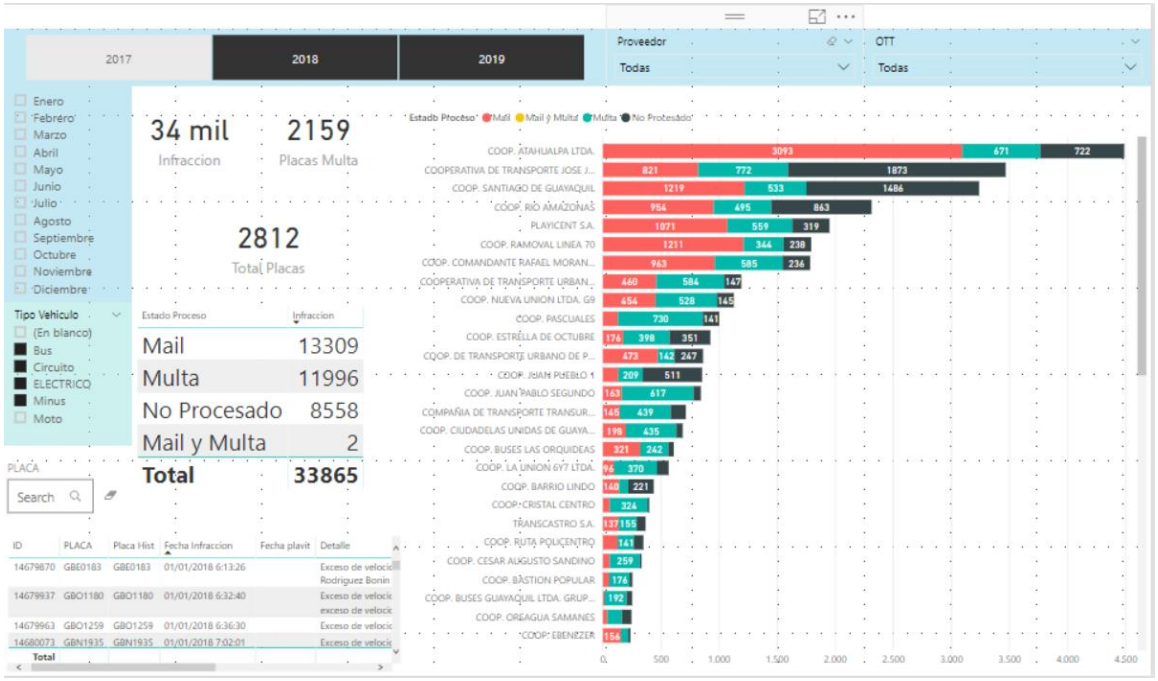


Ilustración 24 Multas de velocidad por cooperativa

3.2.4 Multas de Velocidad por Vehículo

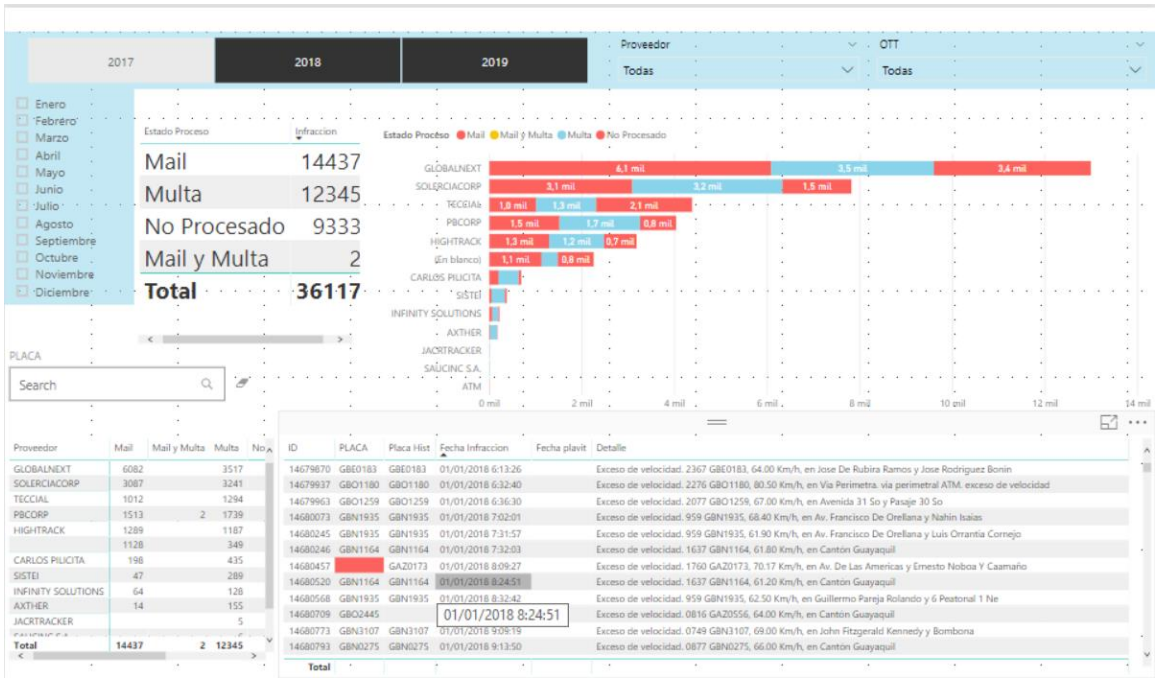


Ilustración 25 Multas de velocidad por vehículo

3.2.5 Mapa de calor de concentración de Multas de Velocidad

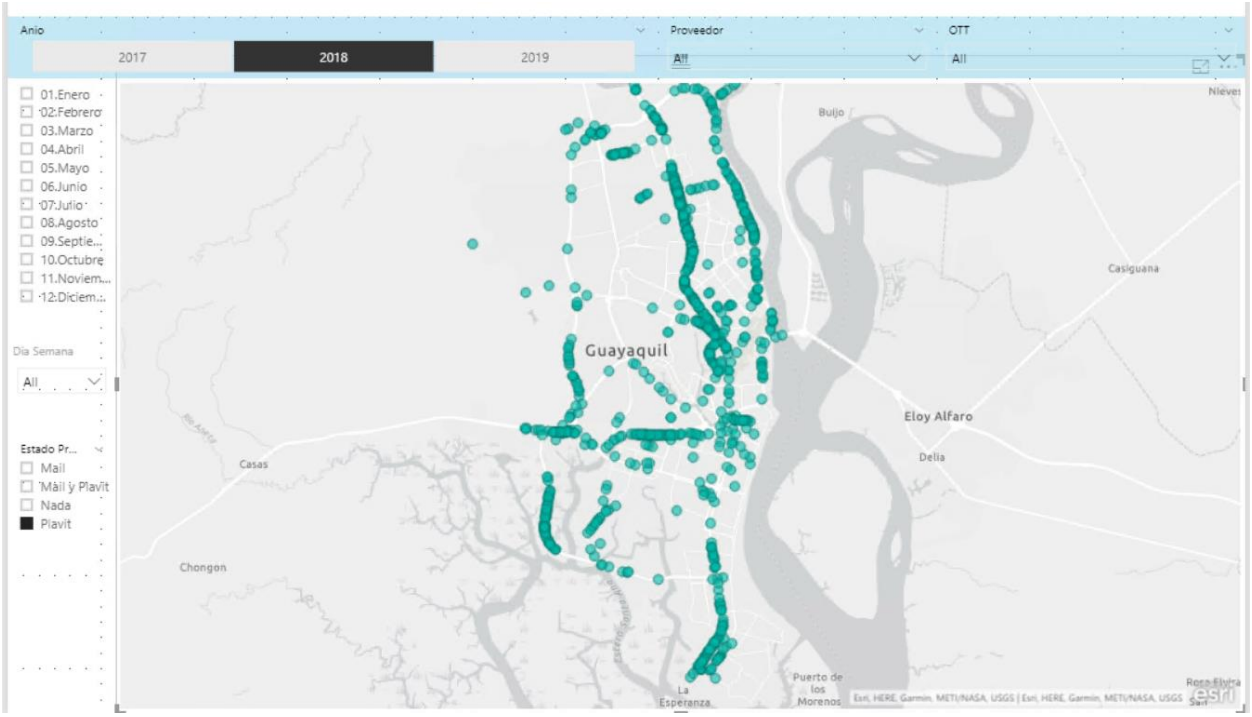


Ilustración 26 Mapa de calor de Infracciones por velocidad

3.3 Impacto

Al inicio de este proyecto no se contaba con indicadores, tampoco con una herramienta de inteligencia de negocios que visibilice los comportamientos anómalos y agrupaciones geográficas desde un punto de vista analítico por lo tanto es las multas electrónicas estaban siendo analizadas desde una perspectiva reactiva

Con la mejora en la identificación y visualización a través del uso de los cuadros de mando de concentraciones de multas se obtendrá un esquema para identificar las posibles mejoras en la trasportación publica en la urbe guayaquileña

Conclusiones y Recomendaciones

Conclusiones

1. Se contextualizó el marco teórico, exponiendo conceptos relacionados a la investigación y las tecnologías a utilizarse dentro de este estudio, definiendo un esquema adecuado a los requerimientos de un proyecto público, y la base necesaria para el análisis de datos deseado, de tal manera que se definió una metodología de implementación y una herramienta adecuada para la situación como CRISP-DM y Power BI.

2. Se estableció los indicadores del cuadro de mando orientados a mejorar el análisis de comportamiento del transporte público en guayaquil con el personal interesado del departamento de transporte público de la autoridad de tránsito municipal, pudiendo esclarecer los puntos de interés y las necesidades en términos del propósito de este estudio,

3. Se diseñó los diferentes escenarios base de cuadros de mando de anomalías y concentración de infracciones basados en la herramienta de inteligencia elegida y se conectó el modelo de datos a la nube de Power BI, permitiendo demostrar la facilidad y versatilidad del modelo así como su disponibilidad en tiempo real sin afectar el desempeño del origen datos.

4. Mediante la utilización los cuadros de mando de anomalías de transporte público basados en Power BI, conectados al modelo real de datos del proyecto de la plataforma tecnológica de SITU, se pudo validar el modelo en su completitud y en términos totalmente reales, obteniendo la visualización esperada para establecer

los parámetros de análisis requeridos y comprobando la efectividad del y la mejora esperada.

Recomendaciones

1. Usar los resultados obtenidos de este estudio para realizar adecuaciones, obras civiles, controles y reglamentaciones, orientadas a disminuir las concentraciones de infracciones en ciertos lugares.

2. Definir nuevos escenarios que ayuden a la organización en la toma de decisiones de forma rápida y oportuna, según se van generando nuevas necesidades control e integrarlas a los cuadros de mando definidos de tal manera que el proceso se vuelva dinámico.

4 Bibliografía

ATM. (2019). *Página web de ATM*. Obtenido de Sitio Web Oficial de ATM: https://www.atm.gob.ec

Coria, S. R. (2016). *Introducción a la Minería de Datos y el Data Warehousing.* Universidad de la Sierra Sur.

crisp-dm.eu. (2015). *crisp-dm.eu*. Retrieved from crisp-dm.eu: http://crisp-dm.eu/home/about-crisp-dm/

elcomercio.com. (2018, octubre 02). Retrieved from El comercio: https://www.elcomercio.com/actualidad/control-buses-gps-redujo-infracciones.html

Microsoft. (2019, febrero 11). *info.microsoft.com*. Retrieved from info.microsoft.com: https://info.microsoft.com/ww-landing-gartner-mq-bi-analytics-2019.html?ls=Website

microsoft. (2019, 09 03). *microsoft.com power bi overview*. Retrieved from https://docs.microsoft.com/es-es/power-bi/power-bi-overview

Pete Chapman (NCR), J. C. (2000). *CRISP-DM 1.0 Step-by-step data mining guide.*

Piatetsky, G. (2014). *kdnugget.* Retrieved from https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

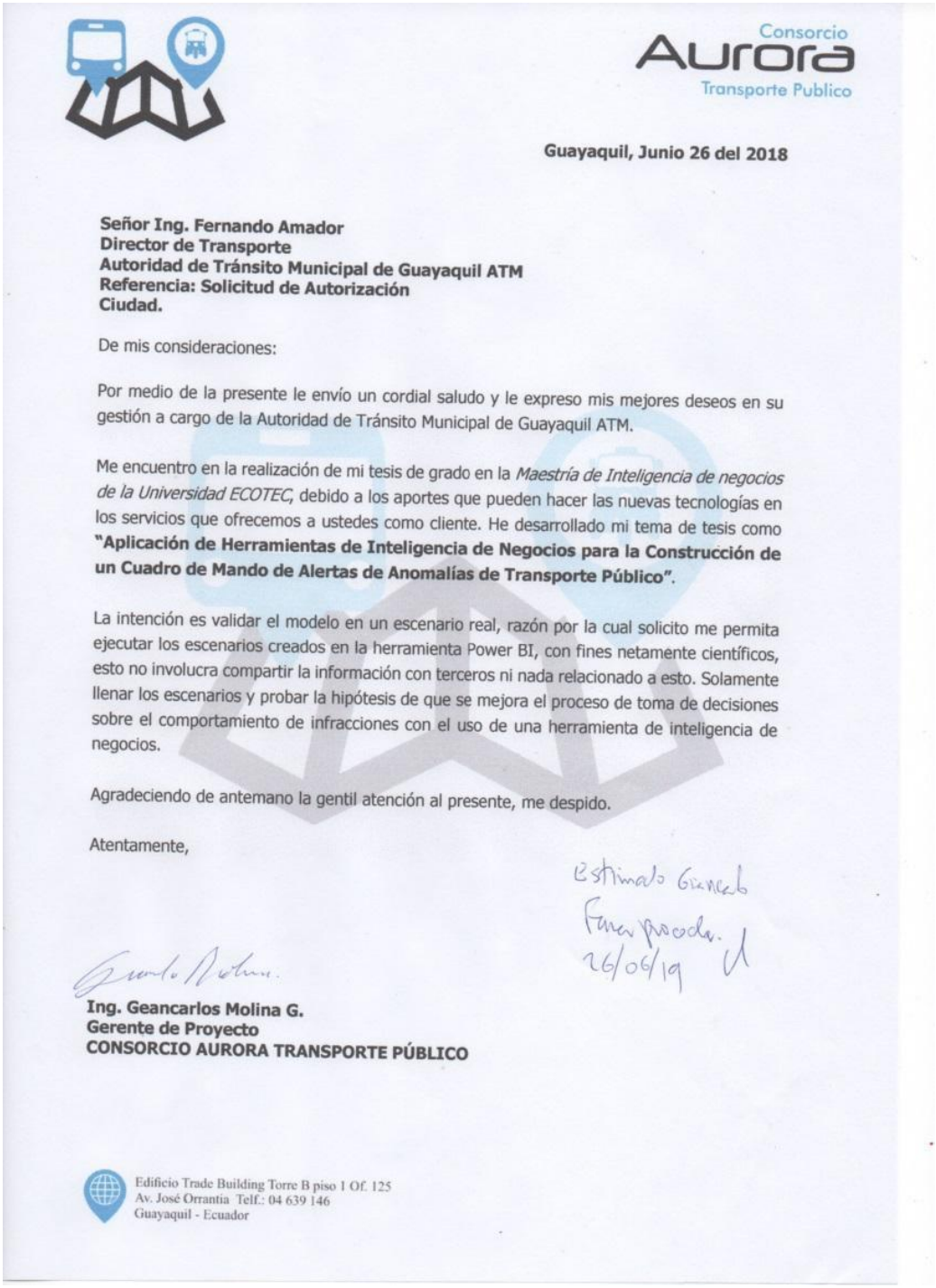
Pozo-Ruz, A. R.-A. (2000). Sistema de posicionamiento global (GPS): Descripción, análisis de errores, aplicaciones y futuro. ETS ingenieros de Telecomunicaciones. Universidad de Malaga. *ETS ingenieros de Telecomunicaciones. Universidad de Malaga*.

Russo, A. F. (2016). *Introducing Microsoft Power BI.* Microsoft Press.

Telegrafo, E. (2017, febrero 06). Retrieved from El Telegrafo: https://www.eltelegrafo.com.ec/noticias/guayaquil/1/ordenanza-obliga-a-uso-del-gps-en-transportes

U.S.government, g. (2016, enero 14). *Official U.S. government information about the Global Positioning System (GPS) and related topics*. Retrieved from gps.gov: https://www.gps.gov/systems/gps/

ANEXOS



Consorcio Aurora Transporte Publico

Guayaquil, Junio 26 del 2018

Señor Ing. Fernando Amador
Director de Transporte
Autoridad de Tránsito Municipal de Guayaquil ATM
Referencia: Solicitud de Autorización
Ciudad.

De mis consideraciones:

Por medio de la presente le envío un cordial saludo y le expreso mis mejores deseos en su gestión a cargo de la Autoridad de Tránsito Municipal de Guayaquil ATM.

Me encuentro en la realización de mi tesis de grado en la *Maestría de Inteligencia de negocios de la Universidad ECOTEC*, debido a los aportes que pueden hacer las nuevas tecnologías en los servicios que ofrecemos a ustedes como cliente. He desarrollado mi tema de tesis como **"Aplicación de Herramientas de Inteligencia de Negocios para la Construcción de un Cuadro de Mando de Alertas de Anomalías de Transporte Público"**.

La intención es validar el modelo en un escenario real, razón por la cual solicito me permita ejecutar los escenarios creados en la herramienta Power BI, con fines netamente científicos, esto no involucra compartir la información con terceros ni nada relacionado a esto. Solamente llenar los escenarios y probar la hipótesis de que se mejora el proceso de toma de decisiones sobre el comportamiento de infracciones con el uso de una herramienta de inteligencia de negocios.

Agradeciendo de antemano la gentil atención al presente, me despido.

Atentamente,

Estimado Giancarlo
Favor proceda.
26/06/19

Ing. Geancarlos Molina G.
Gerente de Proyecto
CONSORCIO AURORA TRANSPORTE PÚBLICO

Edificio Trade Building Torre B piso 1 Of. 125
Av. José Orrantia Telf.: 04 639 146
Guayaquil - Ecuador

Ilustración 27 Carta de autorización de uso de datos