



Universidad Don Bosco, El Salvador

PROYECTO FASE 1

Materia: Datawarehouse y Minería de Datos

Integrantes:

Apellido	Nombre	Carnet
Guerrero Zelaya	Diego Benjamin	GZ210369
Novoa Velásquez	Yesenia Nicole	NV210134

Docente: Karen Medrano

Fecha de entrega: Domingo 13 de noviembre del 2022

ÍNDICE

Contenido

OBJETIVOS	3
MARCO TEÓRICO	4
ANTECEDENTES.....	6
SITUACIÓN ACTUAL	6
METODOLOGÍA	7
FORMULACIÓN DE PROBLEMA.....	13
LIMITACIONES.....	14
PLANIFICACIÓN DE RECURSO	15
RECOMENDACIONES	15

OBJETIVOS

Objetivo general

Desarrollar un análisis sobre las esquelas de infracción de tránsito y parque vehicular.

Objetivos específicos

- Analizar los datos para los informes del gobierno central.
- Demostrar por medio de análisis y gráficas la información brindada para un mejor entendimiento de los datos.
- Determinar dos estrategias con la información brindada para un mejor análisis y mayor entendimiento de la información.

MARCO TEÓRICO

El ministerio de obras públicas (MOP) ha recolectado información sobre las esquelas de infracción de tránsito y parque vehicular, para poder presentar un informe al gobierno central, contrataron a un grupo de consultores para que analicen la información y proporcione un documento con el análisis que han realizaron acerca de los datos recolectados. Tenemos una restricción, el análisis debe de contener dos estrategias por tema de minería de datos.

En los datos recolectados por el MOP, podemos observar el tipo de falta cometida, el estado de la escala, el departamento donde se puso la esquila y por qué se puso la esquila, tenemos dos tipos de esquelas las de tránsito y transporte, nosotros buscamos presentar un análisis donde se pueda comprender de la mejor manera y sintetizado los datos, por medio de gráficas y representaciones.

El ministerio de obras públicas nos puso una restricción para el análisis la cual es ocupar dos estrategias por tema de minería de datos, lo realizaremos con Cubos OLAP y Power BI.

Cubos OLAP

Un cubo OLAP es una estructura de datos que supera las limitaciones de las bases de datos relacionales de esa manera poder proporcionar un análisis rápido de datos. Los cubos pueden mostrar y sumar grandes cantidades de datos. Los cubos de procesamiento analítico en línea (OLAP) son una característica de Service Manager que usan la infraestructura de almacenamiento de datos existente para proporcionar funcionalidades de inteligencia empresarial de autoservicio a los usuarios finales.

Los cubos OLAP se pueden considerar como la última pieza del rompecabezas para una solución de almacenamiento de datos. Un cubo OLAP, también conocido como cubo multidimensional o hipercubo, es una estructura de datos en SQL Server Analysis Services (SSAS) que se genera mediante bases de datos OLAP para permitir

el análisis casi instantáneo de datos. La característica útil de un cubo OLAP es que los datos del cubo pueden estar contenidos en un formulario agregado. Los cubos siempre tienen al menos una medida del negocio, que será quién responda las preguntas de los usuarios. Por ejemplo, en el caso de las ventas teníamos el importe y la cantidad vendida, y sabíamos que aunque no se almacenará era posible determinar el precio unitario. Podemos añadir también el número de ventas realizadas, y otras métricas derivadas según nuestro problema particular.

Power BI

Power BI es una solución de análisis empresarial basado en la nube, que permite unir diferentes fuentes de datos, analizarlos y presentar un análisis de estos a través de informes y paneles. Con Power BI se tiene de manera fácil acceso a datos dentro y fuera de la organización casi en cualquier dispositivo. Estos análisis pueden ser compartidos por diferentes usuarios de la misma organización; por lo que directivos, financieros, comerciales, etc., pueden disponer de la información del negocio en tiempo real.

Se conforma fundamentalmente de estos componentes:

- Power BI Desktop: aplicación gratuita de escritorio para transformar, visualizar datos y crear informes de los mismos.
- Power BI Service: servicio online (SaaS) con funcionalidad similar a la aplicación desktop y permite publicar informes y configurar la actualización de datos automáticamente para que el personal de la organización tenga los datos actualizados.
- Power BI Mobile: aplicación móvil disponible para Windows, iOS y Android para visualizar informes y que se actualiza automáticamente con los cambios de los datos.

Power BI permite conectar a cientos de orígenes de datos en la nube o entorno local, creando informes con objetos integrados o creando objetos personalizados. El acceso a los datos puede ser desde una tabla Excel, Salesforce, Dynamic CRM, Google Analytics, hasta complejas bases de datos (on-premise o en la nube), información de servicios de Azure, etc., lo cual facilita tener toda la información en una única visualización. Con el análisis de los datos, se pueden evaluar los crecimientos, debilidades, fortalezas de una organización, y permitir la toma de acertadas decisiones a corto, mediano y largo plazo, a través de los paneles, informes y gráficos interactivos.

Cubos OLAP y Power BI, no serán demasiado útiles en nuestro análisis ya que será parte esencial, porque podremos observar los datos de una mejor manera y más organizados, un ejemplo de ello es las esquelas en los departamentos, podremos visualizar en qué departamento se puso mas esquelas y en cual se puso menos esquelas, asi se podra llevar un orden de los datos.

ANTECEDENTES

El desarrollo del análisis de datos consiste en llevar un control y un orden en los datos que el MOP nos brindó, para que la información pueda entenderse de la mejor manera, por esa razón habrá gráficas, para hacer una mejor representación, más adelante en el documento se explicara los análisis detalladamente, también encontraremos para que los datos sean presentados de la mejor manera al gobierno central.

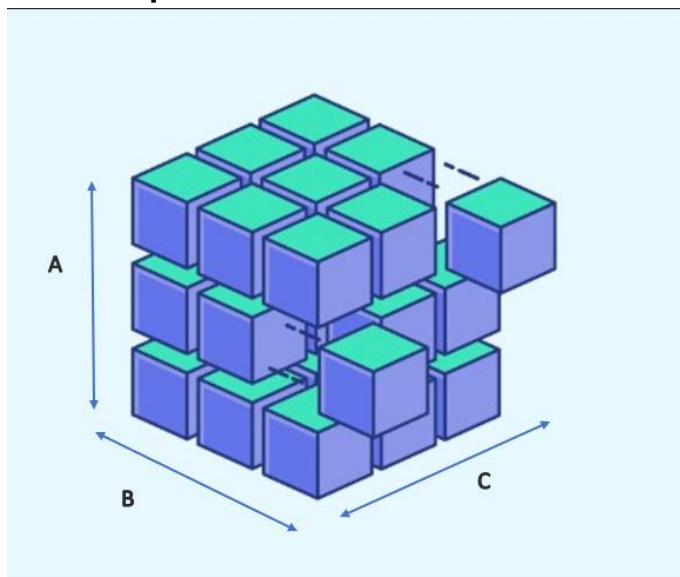
En los análisis que hemos realizado sabemos que el departamento donde más se colocan esquelas es San Salvador, y así podemos ir observando los datos por medio de representaciones, podrán ir viendo todo detalladamente.

SITUACIÓN ACTUAL

Según los datos del ministerio de obras públicas podemos observar que el departamento con más esquelas es en San Salvador y en la actualidad podemos observar si investigamos qué San Salvador sigue siendo el que más esquelas contiene con 4,602, y luego le sigue el departamento de La Libertad con 1,792 y podemos observar una pequeña variación, ya que después de San Salvador seguía el departamento de la Paz, ahora sigue el de La Libertad.

METODOLOGÍA

Cubo Olap



Los cubos Olap están constituidos por dos partes, el cubo mayor y cubo menor, en el mayor podemos ver la cara superficial del cubo mostrándonos el resultado que nosotros estamos solicitando, mientras que los cubos menores son los que conforman el cubo mayor y poseen todos los datos almacenados

En el caso de las esquelas podemos ordenar el cubo de la siguiente forma, que la cara A tenga todas las fechas de las que se tienen registro, La cara B tendrá los departamentos para conocer la ubicación de la escuela, y la cara C tendrá el motivo de la esquila.

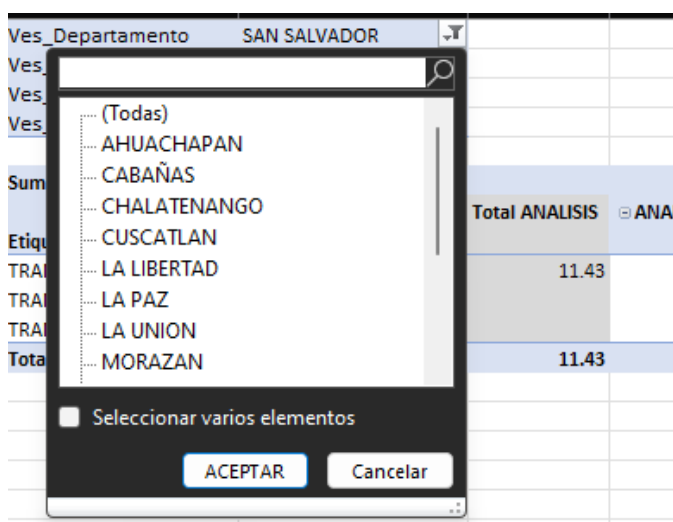
Pero esto es únicamente si vemos las caras superficiales, porque en un cubo Olap al estar formado de muchos cubitos cada cubito tiene más información para poder entrar mas a detalle, por lo que en las caras interiores podremos tener el valor de la escuela, si ya la cancelo o sigue pendiente de pago, el iva que se le realizo en caso de exceder caer en mora y el No. de esquila, un dato esencial a la hora de hacer búsquedas en base de datos.

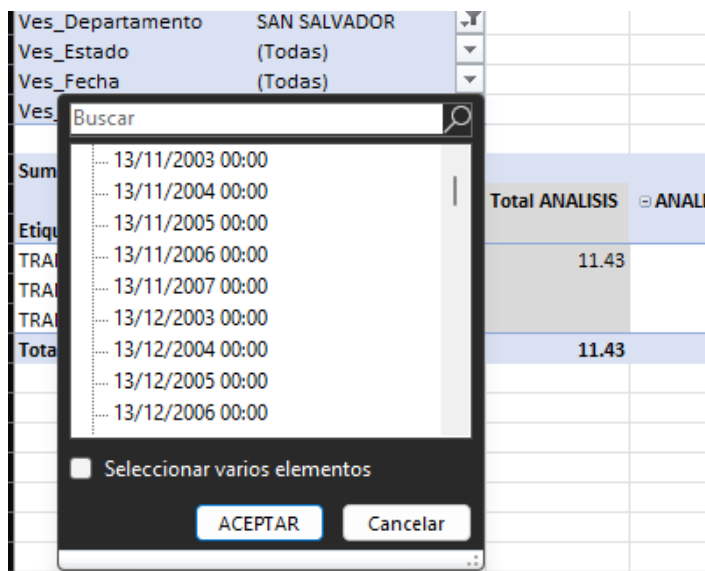
Si aplicamos el cubo con el reporte de esquelas del MOP lo podremos ver tal que así:

Ves_Departamento	SAN SALVADOR					
Ves_Estado	(Todas)					
Ves_Fecha	(Todas)					
Ves_Nro_Esquila	(Todas)					
Suma de Ves_valor	Etiquetas de columna					
	ANALISIS	Total ANALISIS	ANALISIS COBRADA	Total ANALISIS COBRADA	ANULADA	Total ANULADA
Etiquetas de fila		0	0	0	0	0
TRANSITO		11.43	11.43	125.72	788.56	788.56
TRANSPORTE CARGA						
TRANSPORTE TERRESTRE				102.86	34.29	34.29
Total general		11.43	11.43	228.58	822.85	822.85

Donde los filtros Departamentos, estado, fecha, no esquila serán como nuestras caras del cubo con lo que ordenaremos el cubo y en la tabla de abajo sera nuestro cubo con toda la información, pero de forma ordenada siguiendo los filtros que ingresamos anteriormente mostrándonos como resultado únicamente la “cara” del cubo que estamos solicitando.

Debido a la gran facilidad que tiene este cubo de buscar y ordenar los datos que se le están solicitando además de su eficacia, hemos decidido seleccionarla como una de las metodologías a utilizar.





Así se vería el cubo Olap si ingresamos ciertos filtros como lo pueden ser:

Departamento La Libertad con fecha 12/5/2005

Ves_Departamento	LA LIBERTAD				
Ves_Estado	(Todas)				
Ves_Fecha	12/5/2005				
Ves_Nro_Esquila	(Todas)				
Suma de Ves_Valor	Etiquetas de columna				
	<input type="checkbox"/> CANCELADA	Total CANCELADA	<input type="checkbox"/> INCONSISTENTE	Total INCONSISTENTE	
Etiquetas de fila		0	0		
TRANSITO		342.87	342.87	68.57	68.57
TRANSPORTE TERRESTRE		114.28	114.28		
Total general		457.15	457.15	68.57	68.57

Departamento San Miguel con fecha 10/4/2004

Ves_Departamento	SAN MIGUEL				
Ves_Estado	(Todas)				
Ves_Fecha	10/4/2004				
Ves_Nro_Esquila	(Todas)				
Suma de Ves_Valor	Etiquetas de columna				
	<input type="checkbox"/> CANCELADA	Total CANCELADA	<input type="checkbox"/> INCONSISTENTE	Total INCONSISTENTE	Total general
Etiquetas de fila		0	0		
TRANSITO		537.14	537.14	68.57	605.71
Total general		537.14	537.14	68.57	605.71

Power BI

Con los datos que el ministerio de obras públicas nos brindó, pudimos hacer diferentes análisis, en los datos que nos brindaron pudimos observar que existen dos tipos de esquila la de tránsito y transporte, también podemos encontrar los motivos por el cual le colocaron la esquelas, el valor, fecha y estado de la esquila, todo estaba mezclado y no se encontraba organizado, era muy difícil llevar un control de las esquelas que habían colocado ya que no podíamos observar cuál fue el departamento con mas esquelas, no podíamos llevar el control, ni ver cuántas están pendientes. Por esta razón decidimos utilizar Power Bi, ya que nos demostrara de forma gráfica y detallada todos los datos que el ministerio nos brindó.

- En la figura uno y dos podemos observar que el departamento que obtuvo más esquelas fue el departamento de San Salvador.

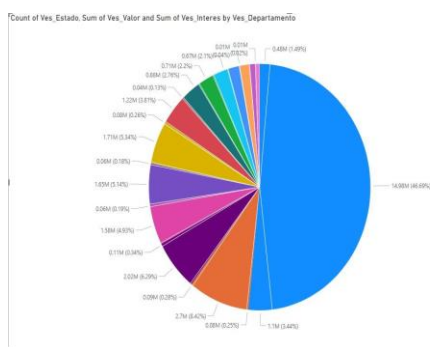


Figura 1

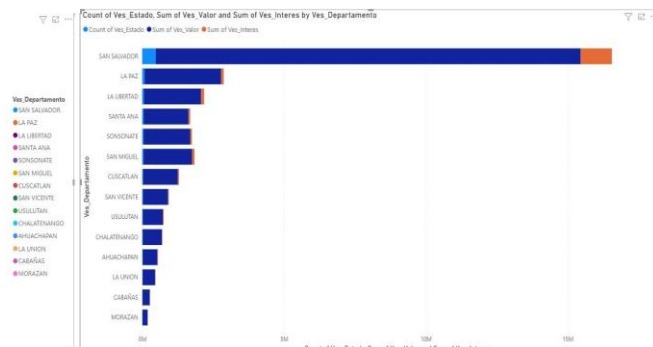


Figura 2

- En la figura tres, podemos observar el gráfico del valor , departamentos e intereses de las esquelas colocadas, San Salvador tiene el 48.02%.

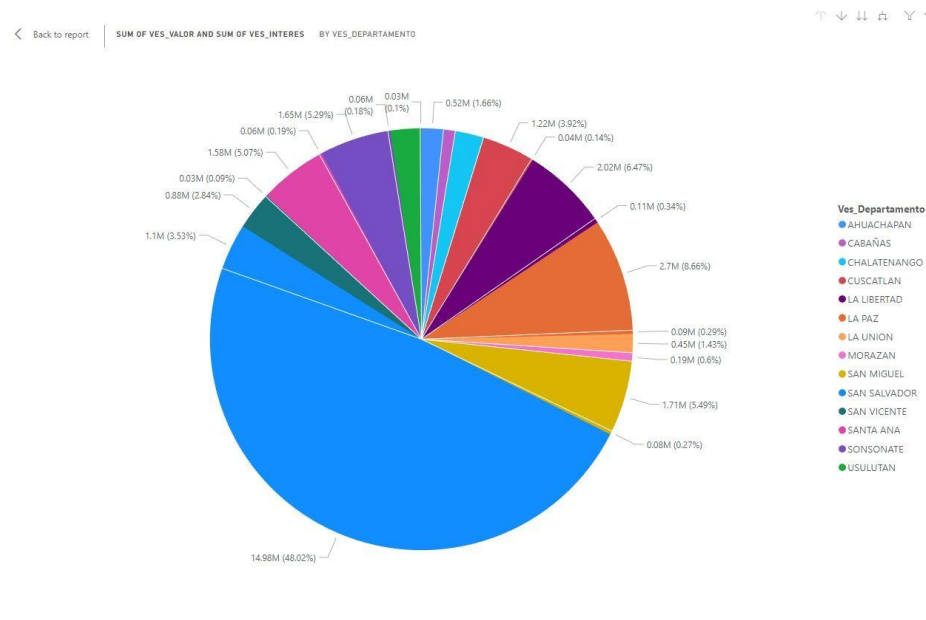


Figura 3

- En la figura cuatro están los datos del departamento y el tipo de falta, podemos ver que la falta que más se repite es la de tránsito.

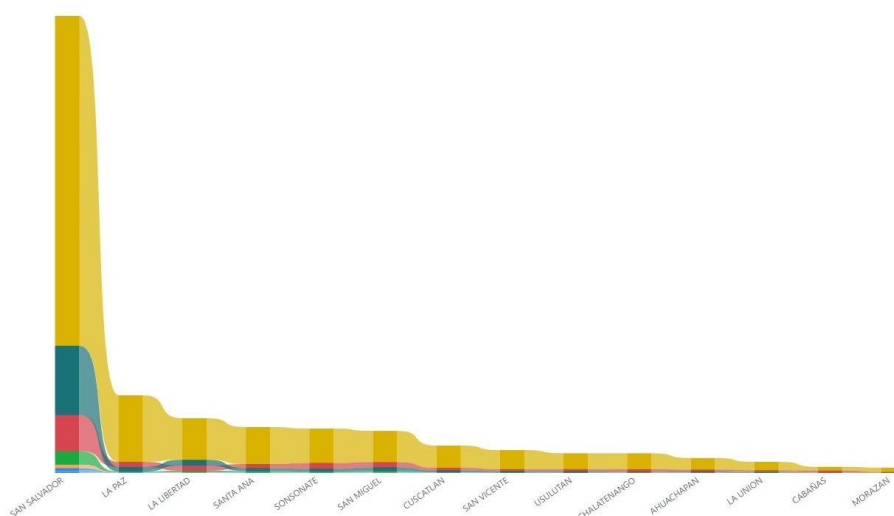


Figura 4

- En la figura cinco podemos observar el estado y departamento.

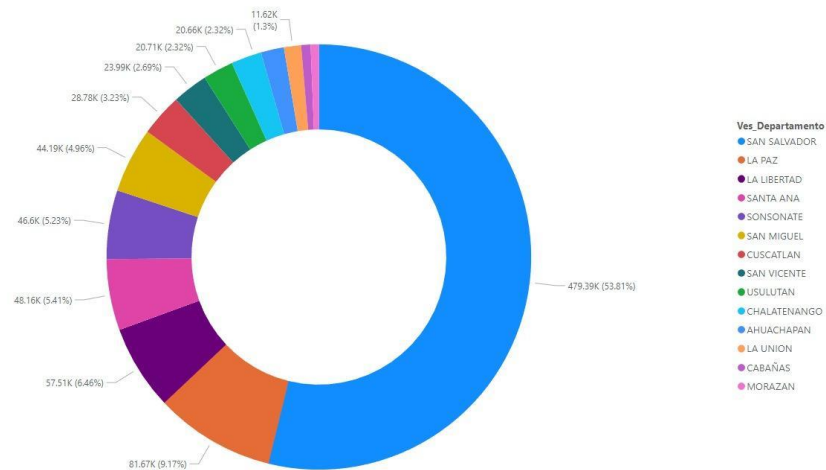


Figura 5

Pueden ir sacando los datos que ellos desean y pueden combinarlos de diferentes maneras un ejemplo podría ser poder el departamento, con el tipo de escuela y el estado en el que se encuentra la escuela, es una manera muy factible y eficaz de poder obtener los datos de una mejor manera y más organizada, así cuando se lo presente al gobierno central lo entiendan de una maner más fácil.

FORMULACIÓN DE PROBLEMA

En el informe brindado por el MOP sobre las esquelas en el país podemos observar 2 datos bastante interesantes y que podemos observar en los gráficos y tablas en Power Bi.

El primero es cómo se disparan la cantidad de infracciones en San Salvador con respecto al resto de departamentos siendo también un dato que se podría esperar teniendo en cuenta que San Salvador es la zona con mayor desarrollo y flujo de movimiento del país.

El otro dato que sobresale del resto es cantidad de esquelas ya canceladas que se tiene registro, teniendo unos números un poco bajos considerando la cantidad total de esquelas impuestas en ese lapso de tiempo

Como punto extra se puede comentar la gran cantidad de esquelas en general que se ponen a diario y el dinero que esto supone para el bolsillo de los conductores, esto debido al mal manejo al volante, malas decisiones como lo es parquear en lugares indebidos, conducir bajo el efecto del alcohol o de cualquier sustancia, invadir carriles contrarios o conducción temeraria, que son las esquelas que más se pueden ver a diario.

Para resolver este problema tenemos que analizar lo primero, nos están solicitando, el ministerio de obras públicas nos solicita que les presentemos un informe con los datos que recogió el MOP, para poder presentarles al gobierno central a lo largo de todo el año sobre las esquelas de los diferentes departamentos, ahora sabiendo esa información podemos investigar cuales metodologías son las que cumplen con ese primer requisito, investigando pudimos observar que las mejores opciones son Cubo Olap y Power BI por lo que nos decantamos por ellas.

LIMITACIONES

Tamaño de la muestra

Al tener una muestra tan extensa con una cantidad de valores a considerar igual de grandes, se nos dificultó poder dimensionar correctamente los gráficos en el caso de Power BI.

Ausencia de investigación previa en el mismo campo

Las estrategias que utilizamos al ser herramientas de análisis nuevas nos tuvimos que familiarizar con ellas rápidamente para poder obtener los resultados en el menor tiempo posible por lo que no pudimos profundizar mucho más acerca de las funciones o de la capacidad real de las dos estrategias.

Falta de datos

la muestra que se nos brindó al tener tanta cantidad de datos por cada categoría como lo puede ser, fecha, No. de escuela, motivo exacto de escuela, etc. No nos brindó otros datos que igualmente pudieron ser importantes o por lo menos interesantes a la hora de mostrarlos en el informe final como lo puede ser la ausencia de la categoría o subcategoría municipios, la ausencia de tipos de vehículos como transporte pesado, liviano, dentro del área Transporte terrestre.

PLANIFICACIÓN DE RECURSO

Infraestructura tecnológica:

Computadora

Internet

Datos recopilados por el ministerio de obras públicas.

Planificación de equipo

El proceso lógico a utilizar como equipo será primero analizar la información que se nos brindó para poder crear el informe

Una vez analizada la información conocer las herramientas que tenemos a nuestra disposición para poder elegir las 2 que mejor se adapten a nuestras necesidades; teniendo las herramientas y la información previamente estudiada se ingresará la información en las estrategias hasta conseguir los resultados esperados.

Terminaremos creando el informe con todas las partes solicitadas para su correcta entrega dentro de la fecha establecida.

RECOMENDACIONES

Los docentes a cargo de esta materia podrían incluir en sus clases pequeños resúmenes sobre cada tipo de metodología para tener una idea previa antes de realizar el trabajo de investigación.

En el caso de Power BI es recomendable la instalación y uso de su propia aplicación, ya que su aplicación web tiene un peor rendimiento a la hora de presentar los gráficos mostrando imágenes en gris, e incluso ralentizando el rendimiento de la computadora.

En el documento Esquelas.xlsx hacer uso de la función **Delimitadores** dentro de Excel para separar mediante los puntos y coma todo el texto que está situado en una sola columna, facilitando enormemente el análisis de la información al tenerla más ordenada y de más fácil lectura