



Universidad Don Bosco, El Salvador

## PROYECTO FASE 2

**Materia:** Datawarehouse y Minería de Datos

**Integrantes:**

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

**Docente:** Karen Medrano

**Fecha de entrega:** Sábado 10 de diciembre del 2022

# ÍNDICE

## Contenido

<b>OBJETIVOS</b> .....	<b>3</b>
<b>MARCO TEÓRICO</b> .....	<b>4</b>
ANTECEDENTES .....	9
SITUACIÓN ACTUAL .....	9
METODOLOGÍA .....	10
FORMULACIÓN DE PROBLEMA .....	18
LIMITACIONES .....	20
PLANIFICACIÓN DE RECURSO .....	21
RECOMENDACIONES .....	21

# ***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, las esquelas de infracción de tránsito lo analizaremos por medio de Cubos OLAP y Power BI y para parque vehicular lo analizaremos por medio de reglas de asociación y árbol de decisión.

## 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, así se podrá llevar un orden de los datos.

## **Reglas de asociación**

Es un conjunto de técnicas que permiten establecer relaciones de interés con la finalidad de descubrir hechos que aporten valor dentro de las variables que facilitan los datos. Estas técnicas utilizan diferentes algoritmos que generan y testean distintas pautas.

El objetivo principal es encontrar correlaciones entre los diferentes elementos u objetos de las bases de datos relacionales, transaccionales o data-warehouses, es de importante describir el algoritmo, explicar sus fases y definir las medidas alternativas para el proceso de descubrimiento de estas asociaciones.

Estas reglas cuentan con diversas aplicaciones:

- El soporte para poder proceder a la toma de decisiones.
- Hay un diagnóstico en las telecomunicaciones.
- Se analiza la información en las ventas.
- Se distribuye la mercancía en las tiendas.
- Se segmentan los clientes en función a su patrón de compra.

## **Técnicas de una clasificación exitosa.**

### **Análisis cluster**

A través del análisis cluster o de conglomerados se pueden clasificar los objetos en grupos más pequeños con características similares que se han conocido previamente.

### **Técnica de data mining**

La técnica de data mining para extraer patrones de datasets mediante la combinación de métodos estadísticos. Se incluyen técnicas de aprendizaje de reglas de asociación, el análisis de agrupamiento y la clasificación. Si aplicamos la minería de datos a los clientes, podremos llegar a modelar el comportamiento realizado en la cesta de la compra.

### **Algoritmos genéticos**

Una técnica que ayuda a optimizar los datos y tiene en cuenta la supervivencia de los que mejor se adaptan. Estos algoritmos evolutivos son muy funcionales para resolver problemas no lineales. Se puede mejorar la planificación de las tareas en la industria manufacturera u optimizar el rendimiento de una cartera de inversión.

### **Aprendizaje automático**

Una especialidad de la inteligencia artificial que se ocupa del diseño y desarrollo de algoritmos que te ofrecen datos empíricos. El objetivo principal es aprender de forma autónoma a reconocer patrones complejos y tomar decisiones según los datos obtenidos.

## **Características principales de las reglas de asociación**




- Cumplir con los niveles mínimos de soporte y confianza y que sus subconjuntos también lo cumplan.
- Si algún ítem no se rige por el nivel mínimo, no debe considerarse como superconjunto.
- Generar reglas con un solo consecuente para construir dos o más, de forma sucesiva.
- El esfuerzo que se realiza depende de la cobertura mínima que se requiere.
- Si un conjunto de ítems no supera la prueba de soporte, ninguno de sus superconjuntos la pasará.

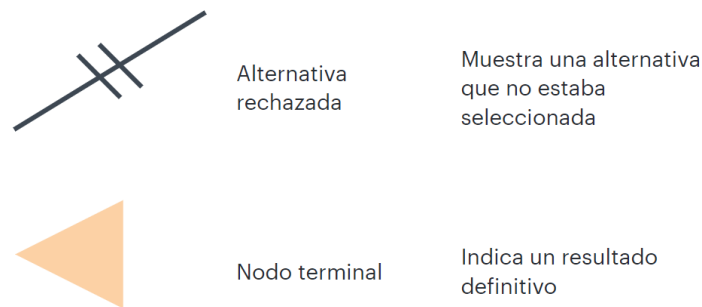
## Árbol de decisión

Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios. Se pueden usar para dirigir un intercambio de ideas informal o trazar un algoritmo que anticipe matemáticamente la mejor opción. Por lo general, comienza con un único nodo y luego se ramifica en resultados posibles. Cada uno de esos resultados crea nodos adicionales, que se ramifican en otras posibilidades. Esto le da una forma similar a la de un árbol.

Existen tres tipos de nodos: nodos de probabilidad, nodos de decisión y nodos terminales. Un nodo de probabilidad, representado con un círculo, muestra las probabilidades de ciertos resultados. Un nodo de decisión, representado con un cuadrado, muestra una decisión que se tomará, y un nodo terminal muestra el resultado definitivo de una ruta de decisión. Los árboles de decisión también se pueden dibujar con símbolos de diagramas de flujo, que a algunas personas les parecen más fáciles de leer y comprender.

### Símbolos de los árboles de decisión

Figura	Nombre	Significado
	Nodo de decisión	Indica una decisión que se tomará
	Nodo de probabilidad	Muestra múltiples resultados inciertos
	Ramificaciones alternativas	Cada ramificación indica un posible resultado o acción



## Cómo dibujar un árbol de decisión

Para dibujar un árbol de decisión, primero debes elegir un medio. Puedes dibujarlo a mano en un papel o una pizarra, o puedes usar un software de árboles de decisión. En cualquier caso, te mostramos los pasos que debes seguir:

1. Comienza con la decisión principal. Dibuja un pequeño recuadro para representar este punto, luego dibuja una línea desde el recuadro hacia la derecha para cada posible solución o acción. Etiquétalas correctamente.
2. Agrega nodos de decisión y probabilidad para expandir el árbol del siguiente modo:

- Si otra decisión es necesaria, dibuja otro recuadro.
- Si el resultado es incierto, dibuja un círculo (los círculos representan nodos de probabilidad).
- Si el problema está resuelto, déjalo en blanco (por ahora).

Desde cada nodo de decisión, dibuja soluciones posibles. Desde cada nodo de probabilidad, dibuja líneas que representen los resultados posibles. Si deseas analizar tus opciones de forma numérica, incluye la probabilidad de cada resultado y el costo de cada acción.

3. Continúa con la expansión hasta que cada línea alcance un extremo, lo que significa que no hay más decisiones que tomar o resultados probables que considerar. Luego, asigna un valor a cada resultado posible. Puede ser una puntuación abstracta o un valor financiero. Agrega triángulos para indicar los extremos.

## Ventajas y desventajas

- Son muy fáciles de entender
- Pueden ser útiles con o sin datos fehacientes, y cualquier dato requiere una preparación mínima
- Se pueden agregar nuevas opciones a los árboles existentes
- Su valor al seleccionar la mejor de numerosas opciones
- Se combinan fácilmente con otras herramientas de toma de decisiones
- las entradas y las decisiones fundamentales.

Para una mayor precisión, se suelen combinar múltiples árboles con métodos de ensamblado:



- Bagging (Ensamblado) crea múltiples árboles de decisión haciendo un remuestreo de los datos de origen, luego hace que los árboles voten para llegar a un consenso.
- Un clasificador de selva aleatoria consiste en múltiples árboles diseñados para aumentar la tasa de clasificación.
- Los árboles ampliados se pueden usar en árboles de clasificación y de regresión.
- Los árboles en una Selva aleatoria (Rotation Forest) se entrenan mediante el uso del análisis de componentes principales (ACP) en una porción aleatoria de los datos.

Un árbol de decisión se considera ideal cuando representa la mayor cantidad de datos con el menor número de niveles o preguntas. Los algoritmos diseñados para crear árboles de decisión optimizados incluyen CART, ASSISTANT, CLS y ID3/4/5. Un árbol de decisión también se puede generar mediante la creación de reglas de asociación, ubicando la variable objetivo a la derecha.

## ANTECEDENTES

El desarrollo del análisis de datos consiste en llevar un control y un orden en los datos que el ministerio de obras públicas nos brindó, para que la información pueda entenderse de la mejor manera posible, por esa razón habrá gráficas, cubos, mapas, etc. Para hacer una mejor representación de todos los datos de las esquelas y del parque vehicular, más adelante en el documento se explicarán los cuatro análisis realizados detalladamente, también podremos encontrar los datos separados y organizados para su mayor entendimiento.

Un ejemplo de nuestro análisis con respecto a las esquelas pudimos observar 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 y en el análisis de parque vehicular los que más se repiten son alquiler y particular.

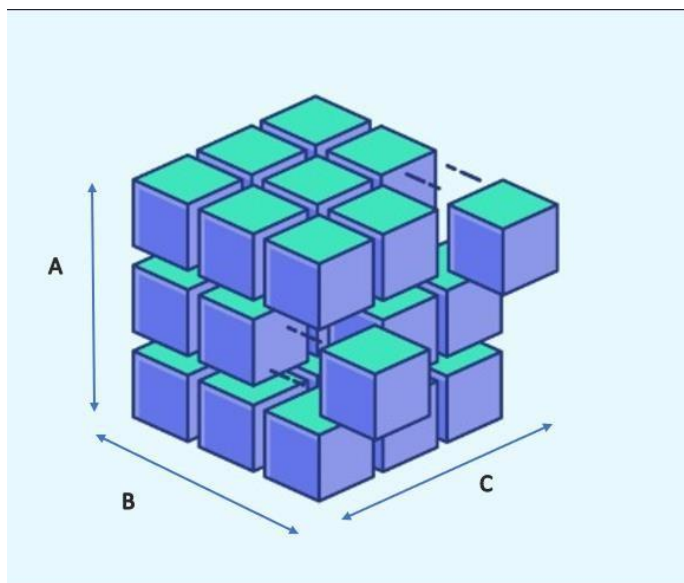
## SITUACIÓN ACTUAL

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 que 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.

El crecimiento del parque vehicular es del 7 % anual, hay aproximadamente 1.6 millones de vehículos en circulación, de los cuales 1.2 millones son de cuatro ruedas”, afirmó el viceministro de Transporte, Saúl Castelar, con base al Registro Público de Vehículos Automotores. Significa que cada año ingresan, en promedio, 112 mil automotores al país, el aumento del parque vehicular influye en el índice de siniestralidad vial. Del 1 al 5 de septiembre de 2022 registran 11,582 accidentes de tránsito en el país, según el Observatorio Nacional de Seguridad Vial.

## 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.

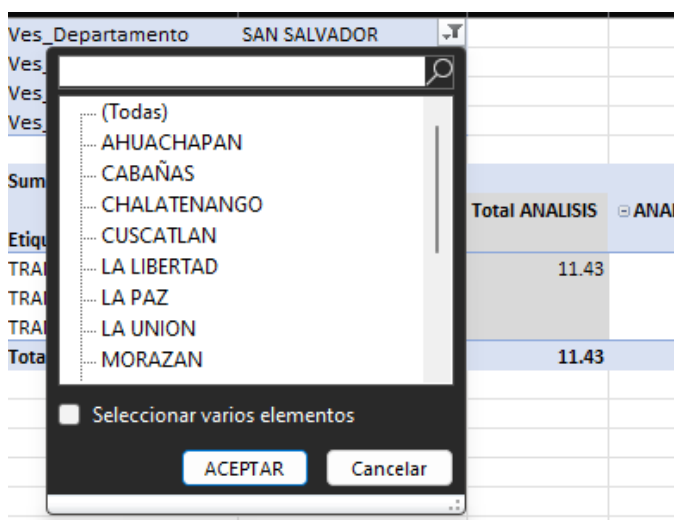
Pero esto es únicamente si vemos las caras superficiales, porque en un cubo Olap al estar formado de muchos cubitos cada cubito tiene más información para poder entrar mas a detalle, por lo que en las caras interiores podremos tener el valor de la escuela, si ya la cancelo o sigue pendiente de pago,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

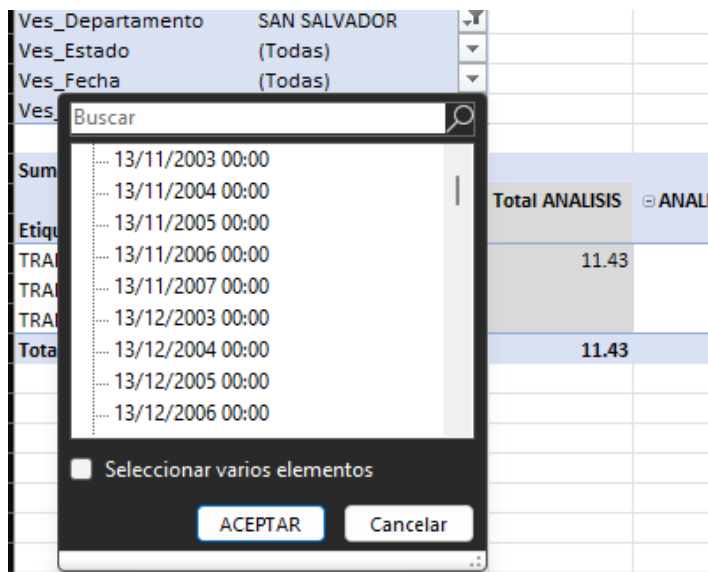
Si aplicamos el cubo con el reporte de esquelas del MOP lo podremos ver tal que así:

Ves_Departamento	SAN SALVADOR					
Ves_Estado	(Todas)					
Ves_Fecha	(Todas)					
Ves_Nro_Esuela	(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 escuela 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.





Asi se veria 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	68.57	605.71
Total general	537.14	537.14	68.57	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 podiamos 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.

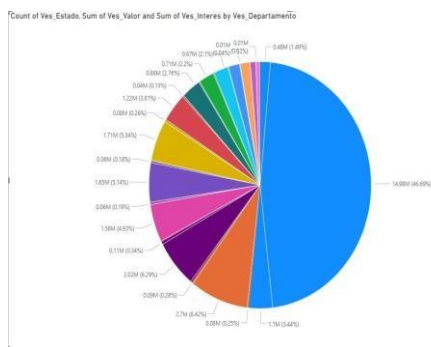


Figura 1

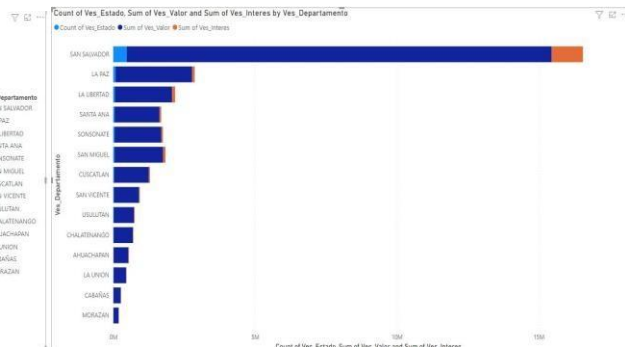
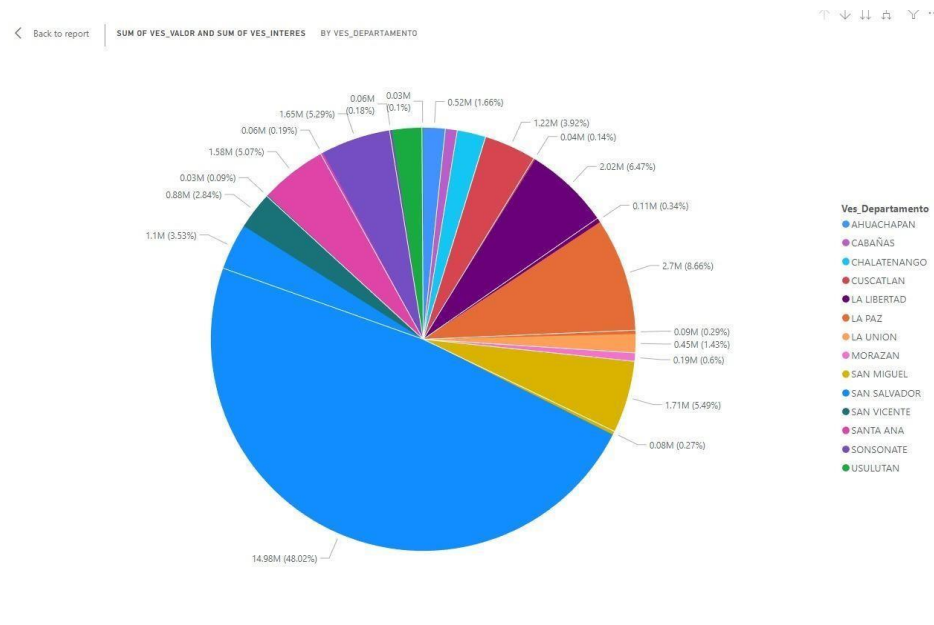
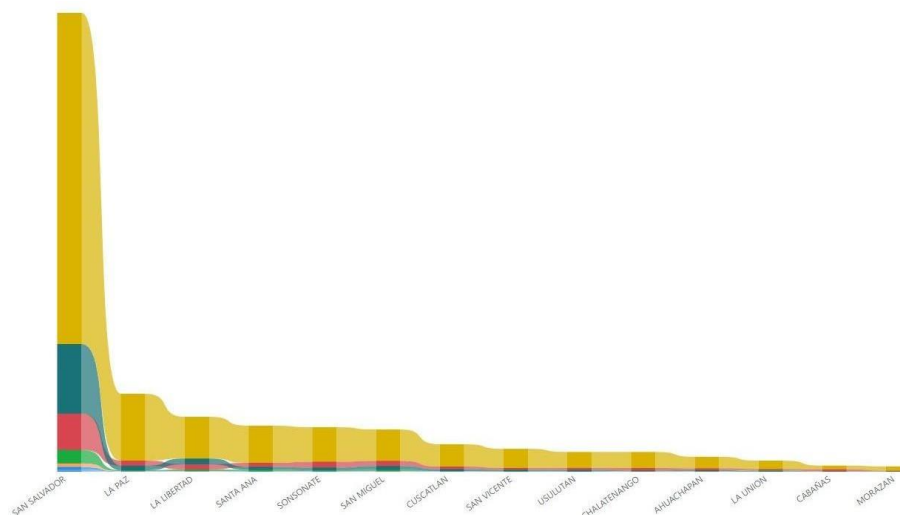


Figura 2

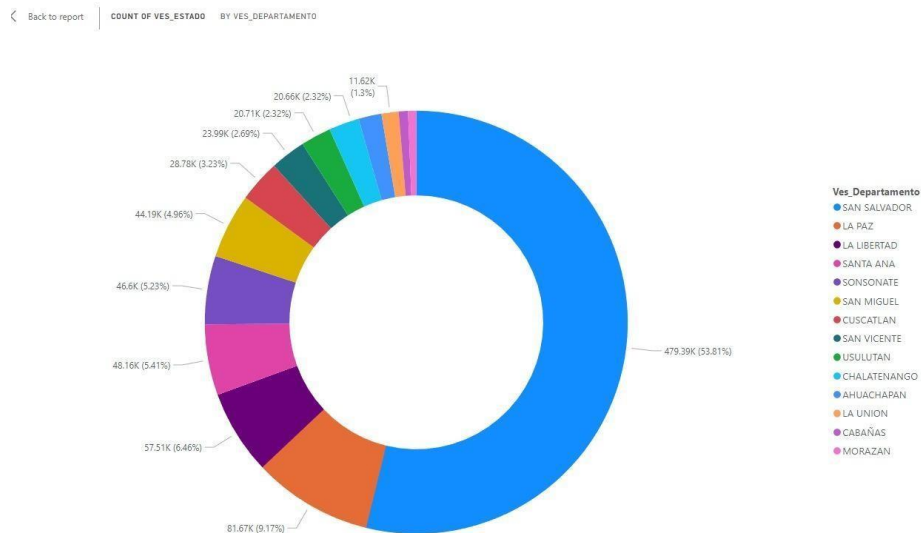
- En la figura tres, podemos observar el gráfico del valor , departamentos e intereses de las esquelas colocadas, San Salvador tiene el 48.02%.



- 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.



- 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 manera más fácil.

## Árbol de decisión

El árbol de decisión es un esquema bastante automatizado donde lo primero que hicimos fue hacer un planteamiento donde sacamos los datos de la base de datos parque vehicular y los colocamos en una tabla dinámica que podremos ver en las figuras 2, 3 y 4 estas tablas nos sirvieron de referencia para poder darle forma al árbol.

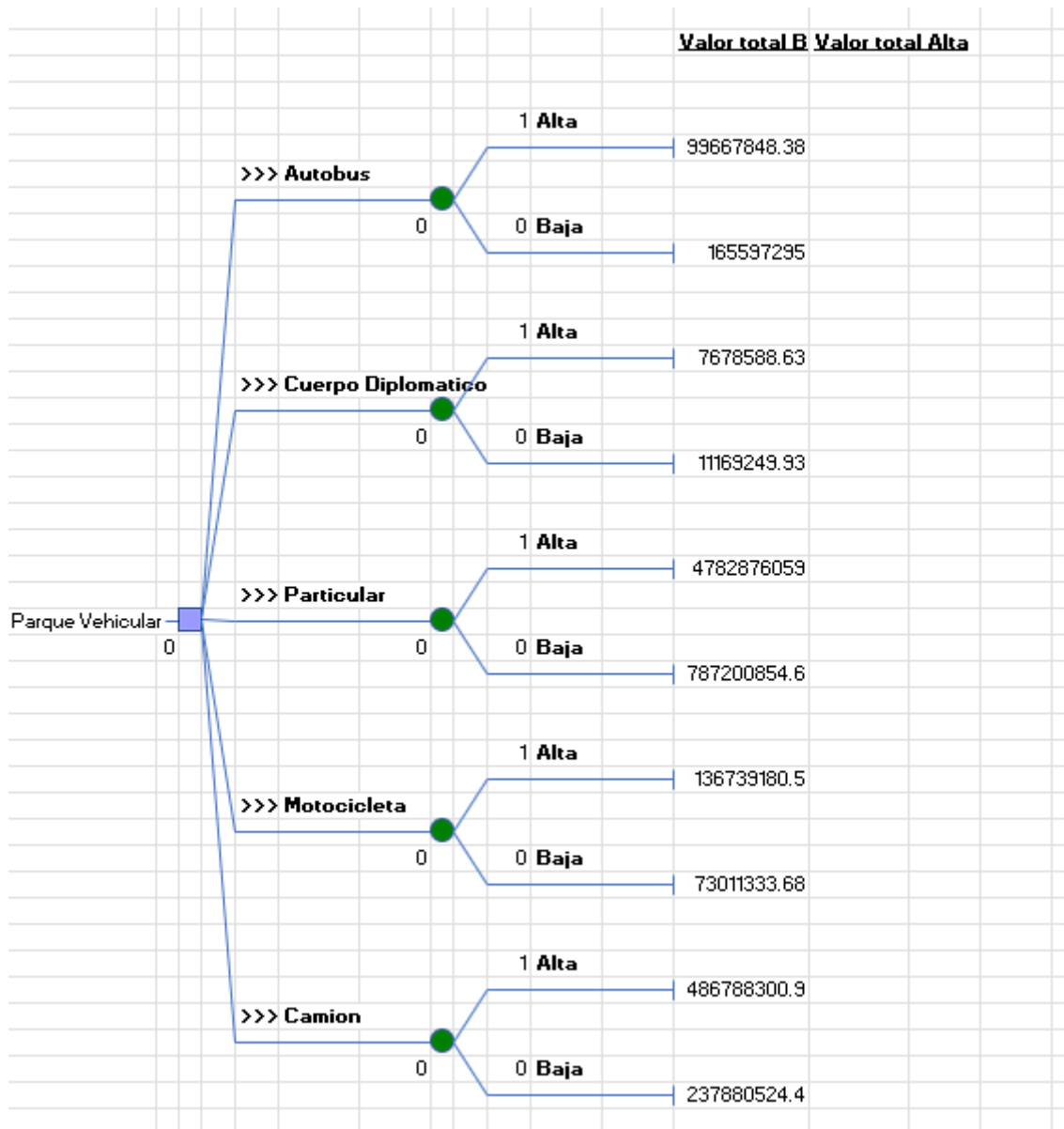


Figura 1.

Este es el árbol que hicimos donde pusimos las 5 principales categorías de vehículos que tenemos en el país, y estos a su vez los separamos si los vehículos siguen en funcionamiento (de alta) o si por el contrario ya termino su vida útil (de baja) y el total gastado en esas categorías.





TIPO_PLACA	PARTICULAR																	
Suma de IMP_VALOR_DEL_VEHICULO	Etiquetas de columna																	
Etiquetas de fila	AHUACHAPAN		CABAÑAS	CHALATENANGO	CUSCATLAN	LA LIBERTAD	LA PAZ	LA UNION	MORAZAN	SAN MIGUEL	SAN SALVADOR	SAN VICENTE	SANTA ANA	SONSONATE	USulutlan	(en blanco)	Total general	
ALTA	132747575.6	82741215.97	133641511.6	93387967.8	762699481.9	152754308.7	184667357.5	102320938.8	363460654	1998086029	76288966.96	337903800.3	194893321.4	167282930			4782876059	
ACURA	169129.77	41949.97	132571.24	129439.37	754142.08	225207.88	134699.79	19601.08	344642.48	1659156.32	112606.33	417342	285023.45	128657.96			4554169.72	
ALDA										0							0	
ALFA ROMEO	60188.59	48231.67	0	49222.39	425101.44	47779.3			9714.29	901698.1	16198.31	95811.55	83955.72	0			1737901.36	
AM GENERAL					3955.83					27172.49		135178.6					166306.92	
AMERICAN										0							0	
AMERICAN MOTORS	0							0		0		0					0	
APPLESTONE										8814							8814	
ARO					0					0							0	
ASIA	162647.81	54221.65	129672.85	71261.94	448606.47	185215.04	16263.05	26831.41	156930.26	1331200.22	64860.59	266843.8	215981.45	155921.82			3286458.36	
ASSM										5770.06							5770.06	
AUDI	131577.55				5031076.43	24240.18	98485.4	63568.95	224621.35	8495953.58	55018.95	229011.97	67969	45435.02			14466958.38	
AUSTIN				0	0				0	9142.86		0	0	0			9142.86	
AUSTIN HEALEY										114.29							114.29	
AUTOBIANCHI	0				0					0		0	0	0			0	
BAM		12714.43								7795.6							20510.03	
BAW					14118.98	6813.32				140507.3		6813.32					168252.92	
BEDFORD										0		0					0	
BLUE BIRD	19459.15		6105.95	37704.7	115287.93	43318.01	20520.57		12784.55	183175.76	3252.7	25753.41	14323.3				481686.03	
BMW	827097.44	518591.43	422671.1	394118.59	23134521.15	646825.67	766816.66	137664.35	2141823.53	47922563.56	427522.48	3578465.02	1155867.29	979575.83			83054124.1	
BTV										0							0	
BUGATTI										2153.72							2153.72	
BUICK		2262.26	16393.66		22371.18					43245.1		0	12501.7	8783.86	1692.72		364422.9	
BURHLEHEM					4832.45					257172.42							4832.45	
BYD										22879.11		8065.05					80944.16	
CADILLAC	25212.87		90770.15	30614.4	241397.54	54867.26	17648.56	10723.71	104758.92	842540.45	156600.97	108360.35	24073.11				1707568.29	
CAN AM					26788.79					38114.8							64903.59	
CARTER				2292.65	14610.57					3733.11				9059.84			29696.17	
CATERHAM										49544.26							49544.26	
CFMOTO					19839.69					13897.09		6226.83					39963.61	
CHANA	23708.29	11673.37	45137.85	20526.45	147561.61	39985.05			31679.55	702742.33	96596.84	50503.56	38907.65				1209022.55	
CHANGAN	9036.6		8791.08	12940.34					7474.95	110814.19	7856.63	9776.32					166690.11	
CHANGHE										7378.92							7378.92	
CHERY	45249.6	6798.02	20260.06	76804.97	402682.19	48481.3	24705.77	13501.07	83765.61	1088972.64	67642.76	153816.59	76596.56	100694.78			2211971.92	
CHEVROLET	4928250.57	2310225.22	4200583.68	3817325.02	36541658.63	5751871.1	3495281.93	2208070.04	10611834.29	103220437	2604463.26	12519146.56	7535675.14	5114502.78			204983333.2	
CHRYSLER	76795.46	75917.34	36430.28	198063.73	1067998.7	203013.52	100773.38	31820.23	148306.88	2243945.3	45211.23	301079.11	238297.69	208569.78			4976220.63	
CHYSLER				2159.43					2636.85	27688.08		28645.56					61129.92	

Figura4.

Tabla que utilizamos para conocer los datos en la categoría particular

## FORMULACIÓN DE PROBLEMA

En el informe brindado por el ministerio de obras públicas (MOP) sobre las esquelas y el parque vehicular, en el documento que se nos brindó de esquelas podemos ver que hay diferente tipo de esquelas y cada una tiene un estado diferente, como pendiente, cancelado, etc. En el documento de parque vehicular podemos observar que se divide en camión, motocicletas, particular, etc y cada uno tiene diferente color como rojo, azul, blanco, etc.

El primero documento que analizaremos será el de las esquelas, se disparan la cantidad de infracciones en San Salvador con respeto 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, estodebido 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 carrilescontrarios 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.

El segundo documento que analizaremos será el de parque vehicular, podemos observar que la mayoría de carros son particular y motocicletas, y sabemos que el próximo año aumentara un 7% las cifras, también sabemos que en San Salvador esta la mayor carga vehicular y la marcas que más hay son: Toyota y Nissan, el próximo año pueden cambiar las cifras, pero según el documento que nos brindó el ministerio de obras públicas esos son los datos que han obtenido ese año.

# 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. Con respecto al árbol de decisiones se nos dificultó un poco ya que el contenido es demasiado extenso, y al principio por el mismo tamaño del documento no se nos descargaba completo para poder pasar hacer el respectivo árbol de decisiones.

## **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. Tuvimos que observar videos de YouTube para poder hacer uso de las distintas plataformas y así poner en marcha nuestras cuatro estrategias de ambos temas.

## **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 observar y analizar la información que se nos a brindó para realizar el informe. Una vez analizada el documento de esquelas y parque vehicular, podemos comenzar a investigar las herramientas que nuestra docente nos a brindad para poder elegir las cuatro que mejor se adapten a nuestras necesidades, para poder entregar un informe de calidad y entendible; teniendo las herramientas y la información previamente estudiada se ingresará toda la información en las estrategias hasta conseguir los resultados esperados, y tener una mejor claridad de los datos que obtuvieron el ministerio de obras públicas. Terminaremos creando el informe con todas las partes solicitadas para su correcta entrega dentro de la fecha establecida.

# RECOMENDACIONES

Una recomendación para el ministerio de obras públicas seria que, al momento de sacar los datos y análisis, lo puedan hacer por medio de estas cuatro plataformas utilizadas anteriormente, ya que así podrán llevar un mejor control de todos los datos y se le facilitara saber cuándo a creció el parque vehicular en la actualidad, y podrán saber el promedio de las esquelas puestas, así podrán presentar cada año los datos de una manera muy entendible.

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.

Rapidminer es una aplicación que requiere de cierto estudio previo antes de su utilización por lo que se recomienda dedicarle un poco mas de horas en su uso

Si la computadora con la que se realice el trabajo es de especificaciones un poco sencillas se recomienda trabajar lo menos posible con el apartado fechas en ambos documentos, ya que estos consumen gran parte de los recursos de su pc.

En el documento Esquelas.xlsx y parque.vehicular.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.