# Cubo OLAP

Para el cubo Olap se hace uso de POWER BI. Inicialmente se cargan los datos desde un .xlsx

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Para analizar los datos desde este cubo se han realizado 3 ventas, cada una representando sus comparaciones: Esquelas Vs. Departamentos, Periodo Vs. Tipo, Falta Vs. Valor/Interés.

Quedando de la siguiente manera:

**Esquelas Vs. Departamentos**, la cual nos permite analizar la cantidad de esquelas por cada uno de los departamentos, distribuidos por el estado de la esquela.

Gráfico, Histograma

Descripción generada automáticamente

**Periodo Vs. Tipo**, el cual nos permite analizar los tipos de faltas por periodos (año, cuarto y mes). Esto permite analizar específicamente etapas del año en las que puede haber eventos y comparar si esto produce un alza o baja de esquelas.

Interfaz de usuario gráfica, Aplicación, Tabla, Correo electrónico

Descripción generada automáticamente

**Falta Vs. Valor/Interés**, permite analizar cada una de las faltas con su valor, así como su acumulado e interés.

Interfaz de usuario gráfica, Aplicación

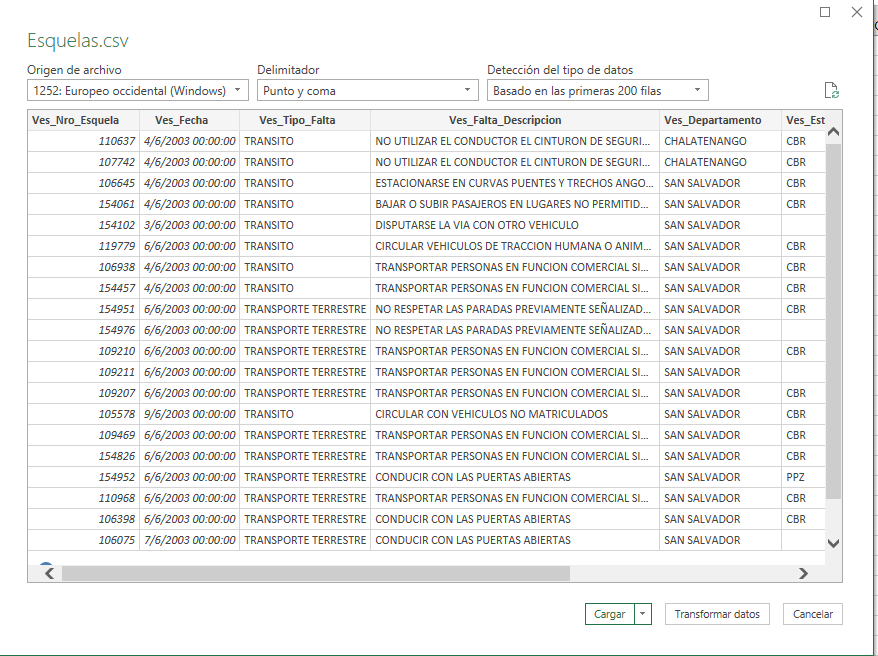
Descripción generada automáticamente con confianza media

# Árbol de decisión

Para el árbol de decisión se hace uso de RapidMiner se convertirá el .csv a .xlsx para revisar los datos inicialmente



Vista previa de datos



Al importar, descubrimos algunos errores: Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

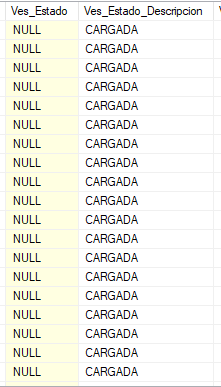
Observamos que algunos campos se encuentran vacíos, vamos a determinar un valor para que no queden null

Tabla

Descripción generada automáticamente

Para realizar las modificaciones vamos a apoyarnos de SQL y la BD que ya hemos montado. (Estas misma modificaciones se realizaron para el Cubo Olap)

Identificamos que el estado cargado, no tiene un valor

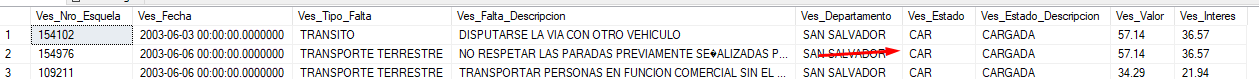


Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamenteImagen que contiene Texto

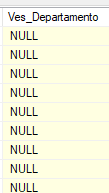
Descripción generada automáticamente

Se modificaron las columnas



Vemos departamentos que no tienen datos





Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamenteImagen que contiene Forma

Descripción generada automáticamente

Observamos que algunas no tienen valor, o tienen error

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza baja

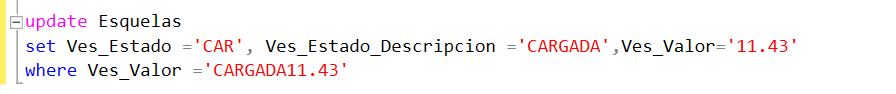
Tabla

Descripción generada automáticamente

En este caso vemos que hay un desplazamiento de los datos, que hay que corregir

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente





Una vez realizadas todas las modificaciones, se ha generado un BD de Backup, en la cual se sustituyen el nombre de los campos por sus equivalentes en número, quedando de la siguiente manera.

Tipo de falta

|  |  |
| --- | --- |
| 1 | TRANSITO |
| 2 | TRANSPORTE TERRESTRE |
| 3 | TRANSPORTE CARGA |

Departamentos

|  |  |
| --- | --- |
| SAN SALVADOR | 1 |
| CHALATENANGO | 2 |
| MORAZAN | 3 |
| LA PAZ | 4 |
| SONSONATE | 5 |
| SAN MIGUEL | 6 |
| LA LIBERTAD | 7 |
| CUSCATLAN | 8 |
| SAN VICENTE | 9 |
| SANTA ANA | 10 |
| AHUACHAPAN | 11 |
| LA UNION | 12 |
| USULUTAN | 13 |
| CABAÑAS | 14 |
| Sin Determinar | 15 |

Estado:

Tabla

Descripción generada automáticamente

Con estas modificaciones, se exporta un nuevo archivo .csv con el cual trabajaremos en rapid miner



Una vez en rapid miner, se establece el siguiente método:

Icono

Descripción generada automáticamente

Dando como resultado el siguiente árbol de decisiones:

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

## Análisis Cubo OLAP y Árbol de Decisiones

Dado el cubo OLAP, árbol y los datos reflejados, se puede realizar el siguiente análisis.

* La esquela mas antigua registrada data de 1968
* El tipo de falta “Transito” contiene la mayor cantidad de faltas: 853,088. Mientras que “Transporte Carga” la menor con: 9134.
* Para el tipo de faltas “Transito”, las 4 faltas repetidas son las siguientes:

|  |  |
| --- | --- |
| Cantidad | Descripción de falta |
| 53,517 | “No portar licencia de conducir” |
| 45,583 | “Estacionarse en Zona Prohibida o Eje Preferencial” |
| 45,117 | “No utilizar el cinturón de seguridad” |
| 44,559 | “No portar triangulo reflectivo” |

* Para el tipo de faltas “Transporte Carga”, las 4 faltas repetidas son las siguientes:

|  |  |
| --- | --- |
| Cantidad | Descripción de falta |
| 2,599 | “Circular vehículos que transportan carga con un peso mayor al autorizado” |
| 1,424 | “Carecer de cubierta protectora sobre la carga” |
| 779 | “Carecer de cinta Adhesiva reflejante u ojos de gato reglamentarios” |
| 517 | “Llevar sobre la carga del camión o rastrados personas” |

* Para el tipo de faltas “Transporte terrestre”, las 4 faltas repetidas son las siguientes:

|  |  |
| --- | --- |
| Cantidad | Descripción de falta |
| 42,600 | “Transportar personas en función comercial sin el permiso correspondiente” |
| 36,130 | “Alterar las tarifas autorizadas por las autoridades” |
| 34,944 | “Estacionarse más tiempo del necesario para subir o bajar” |
| 31,624 | “Conducir con las puertas abiertas” |

* Los tres Departamentos con mayor cantidad de esquelas son:

|  |  |
| --- | --- |
| Cantidad | Descripción de falta |
| 485,201 | San Salvador |
| 146,140 | Sin Determinar |
| 82,395 | La Libertad |

* La esquela que mas se impone tiene un valor de $57.14, siendo “Muy Grave”.

# Reglas de asociación

Para este apartado, se utiliza el archivo compartido, el cual se ha cargado en SQL Server para realizar limpieza de datos.

Reglas de asociación establece que los datos deben ser binomiales, por lo cual, como primer punto se identifican los datos que pueden ser tratados como binomiales.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Para estos datos vamos a establecer el Tipo de Placa como el ID. Se procede a limpiar los datos, estableciendo los datos como 0 y 1.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Se crea una nueva tabla, en la que haremos la carga de los datos que utilizaremos para consulta

Tabla

Descripción generada automáticamente

Una vez trabajados los datos, procedemos a Rapid Miner a crear la estructura de reglas de asociación:

Diagrama

Descripción generada automáticamente

Con un Fp-Growth inicial, tenemos una imagen de los datos de la siguiente manera:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Ahora se procede a crear las reglas de asociación, quedando la siguiente estructura:

Gráfico, Diagrama

Descripción generada automáticamente

Resultando lo siguiente:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Como resultado podemos ver también los siguientes gráficos

Gráfico

Descripción generada automáticamente

Texto

Descripción generada automáticamente con confianza media

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

# Agrupamiento con K-Means

Para los K-MEANS haremos uso de una tercer tabla creada

Tabla

Descripción generada automáticamente

Se establece la siguiente estructura en rapid miner

Diagrama

Descripción generada automáticamente

Texto

Descripción generada automáticamente con confianza baja

Gráfico

Descripción generada automáticamente

# Fase 2

Para esta fase, se realiza un archivo global en PowerBI, el cual contiene todos los datos necesarios para realizar el análisis.

Reporte PowerBI: https://app.powerbi.com/groups/me/reports/436e1639-aec1-4e7c-adfb-2d6aebc6e220?ctid=f9afe020-14e8-4555-b638-b98f896aa94b