

Universidad Don Bosco



Docente: Ing. Alexander Alberto Sigüenza

Materia: Datawarehouse y Minería de Datos

Presentado por:

Ruiz Saz, Carlos José RS181977

Rosales Mendoza, Patrick Ernesto RM181976

Porcentaje realizado:100%

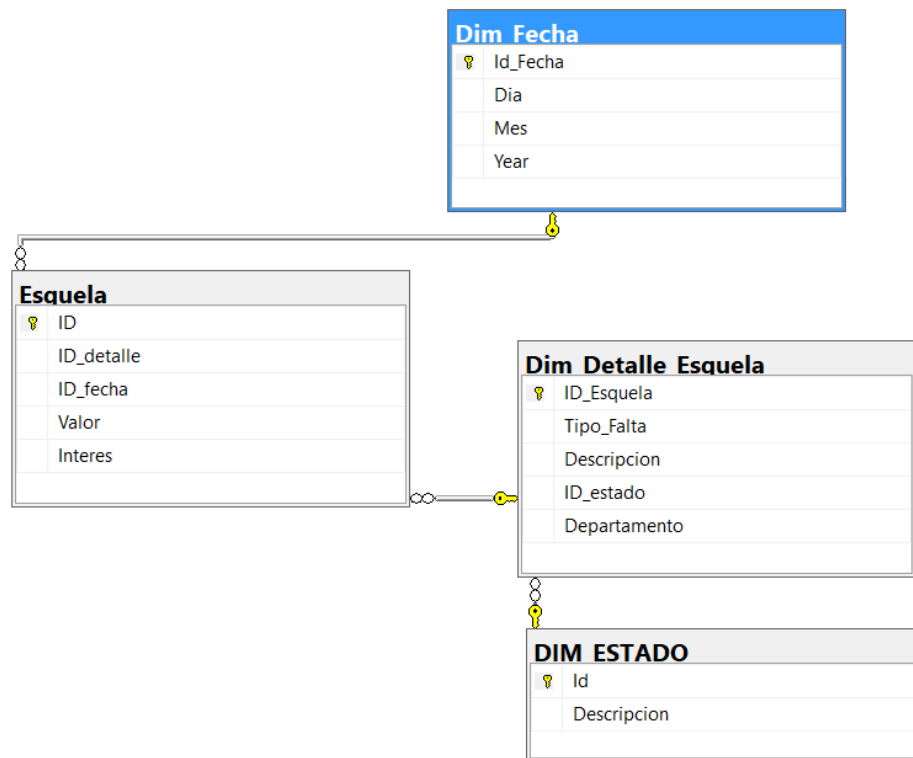
Ciclo: 02-2020

Soyapango, Domingo 15 de noviembre del 2020

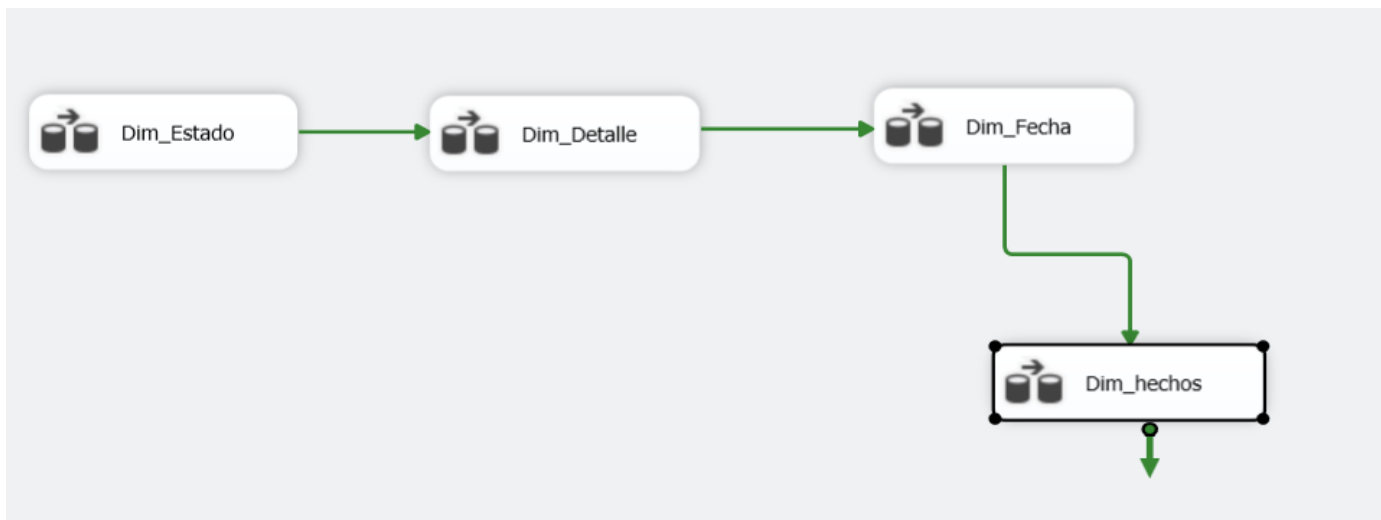
Capturas desafio III

Esquemas de infracción de tránsito

Diagrama de la base de datos a utilizar:



Se utilizará lo siguiente para el etl:



En cada una de las tareas se realizará para llenar cada una de las tablas.

Una vez tenemos nuestra base de datos con los registros cargados correctamente, haremos la creación de un cubo, para ello crearemos un nuevo proyecto de análisis y minería de datos.

- Creamos el origen de datos:

Administrador de conexiones

Proveedor: OLE DB nativo\SQL Server Native Client 11.0

Nombre del servidor: DESKTOP-TBIH0UF\MSSQLSERVER2 Actualizar

Conexión con el servidor

Autenticación: Autenticación de SQL Server

Nombre de usuario: sa

Contraseña: •••••

☐ Guardar mi contraseña

Establecer conexión con una base de datos

☒ Seleccionar o escribir el nombre de la base de datos:

Esquelas

☐ Adjuntar un archivo de base de datos:

Examinar...

Nombre lógico:

Probar conexión Aceptar Cancelar Ayuda

- En vista del origen de datos, colocamos los siguientes elementos:

Asistente para vistas del origen de datos

Seleccionar tablas y vistas

Seleccione los objetos de la base de datos relacional que deben incluirse en la vista del origen de datos.

Objetos disponibles:

Nombre	Tipo
sysdiagrams (dbo)	Tabla

Objetos incluidos:

Nombre	Tipo
Dim_Detalle_Esquila (dbo)	Tabla
DIM_ESTADO (dbo)	Tabla
Dim_Fecha (dbo)	Tabla
Esquila (dbo)	Tabla

Filtrar:

☐ Mostrar objetos del sistema

Agregar tablas relacionadas

< Atrás Siguiente > Finalizar >> Cancelar

- Creamos el cubo

Nuestra tabla de hechos será esquila:

Asistente para cubos

Seleccionar tablas de grupo de medida

Seleccione una vista o diagrama de origen de datos y, a continuación, seleccione las tablas que se usarán para los grupos de medida.

Vista del origen de datos:
Esquelas

Tablas de grupo de medida:

- ☐ Dim_Detalle_Esquila
- ☐ DIM_ESTADO
- ☐ Dim_Fecha
- ☒ Esquila

Sugerir

< Atrás **Siguiente >** Finalizar >> Cancelar

Y estas serán nuestras medidas:

Asistente para cubos

Seleccionar medidas

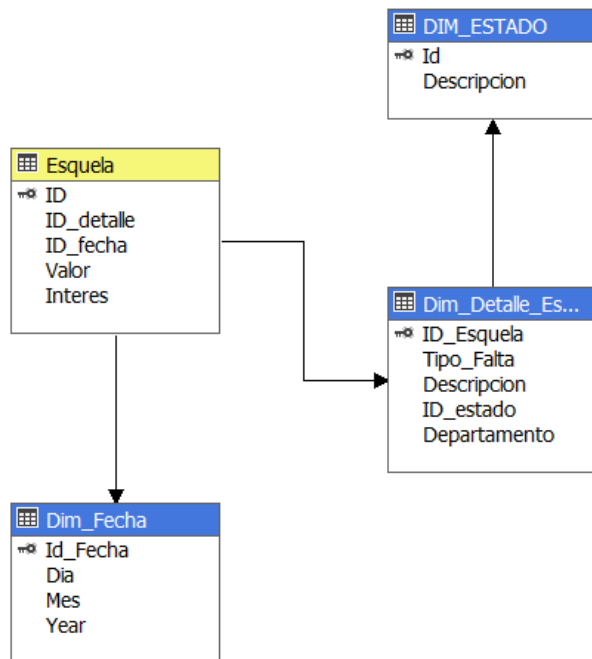
Seleccione las medidas que desea incluir en el cubo.

☒ Medida

- ☒ Esquila
 - ☒ Valor
 - ☒ Interes
 - ☒ Recuento Esquila

< Atrás **Siguiente >** Finalizar >> Cancelar

Nuestro cubo deberá verse así:



Procesamos el cubo:

Progreso del proceso

Comando

- Procesando Cubo, 'Esquelas' completados.
 - Hora de inicio: 6/11/2020 17:04:57; Hora de finalización: 6/11/2020 17:05:00; Duración: 0:00:02
- Procesando Grupo de medida, 'Esquela' completados.

Estado:

✓ Proceso finalizado correctamente.

Detener Reprocesar Ver detalles... Copiar

Cerrar Ayuda

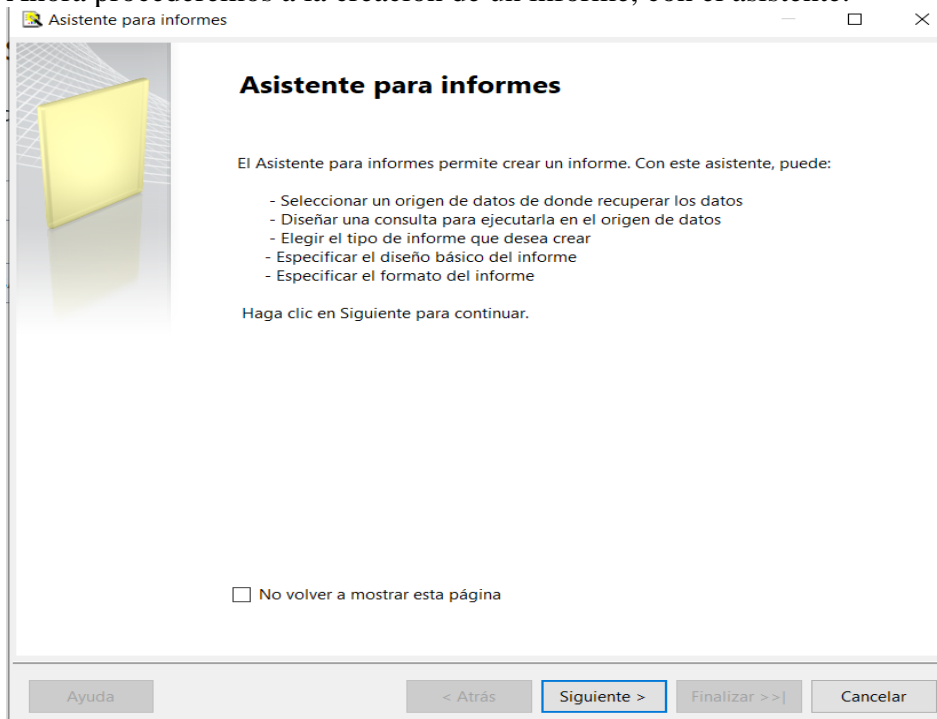
Análisis esquelas por departamento

Departamento	Valor	Recuento Esquela
AHUACHAPAN	520552.820000001	15728
CABAÑAS	260831.35	6379
CHALATENANGO	674227.860000002	20726
CUSCATLAN	1227088.969999998	28906
LA LIBERTAD	2042680.269999992	58345
LA PAZ	2721429.499999998	82359
LA UNION	446913.4	11680
MORAZAN	187340.960000001	5582
SAN MIGUEL	1730818.879999992	44844
SAN SALVADOR	15108792.98000068	484951
SAN VICENTE	888591.180000001	24113
SANTA ANA	1597661.089999991	48732
SONSONATE	1655718.469999992	46883
USULUTAN	719671.750000003	21164

Como podemos observar, el departamento con un mayor número de esquelas es San salvador, lo cual no sorprende mucho, ya que es la capital del país en donde la mayoría de sus habitantes ocupan sus vehículos diariamente, por el contrario podemos ver al departamento de Morazán como el que menos esquelas se han impuesto, sin embargo el departamento que ha pagado menos por esquelas ha sido cabañas, esto quiere decir que Morazán aunque cuenten con el menor número de esquelas, ha cometido faltas muy graves, en cambio cabañas ha cometido faltas más leves.

En el cuadro anterior podemos observar como los departamentos con mayor circulación se han cometido más faltas a comparación de departamentos con menor circulación, resaltando a Cabañas y Morazán

Ahora procederemos a la creación de un informe, con el asistente.



Asistente para informes

El Asistente para informes permite crear un informe. Con este asistente, puede:

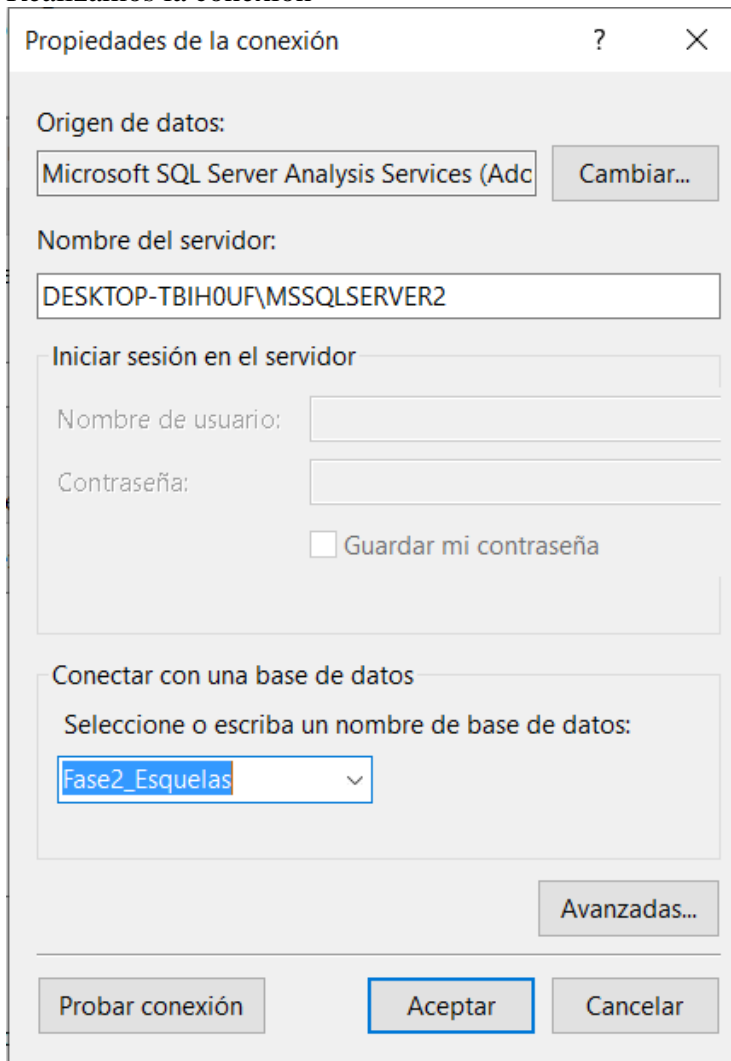
- Seleccionar un origen de datos de donde recuperar los datos
- Diseñar una consulta para ejecutarla en el origen de datos
- Elegir el tipo de informe que desea crear
- Especificar el diseño básico del informe
- Especificar el formato del informe

Haga clic en Siguiente para continuar.

☐ No volver a mostrar esta página

Ayuda < Atrás Siguiente > Finalizar >> Cancelar

Realizamos la conexión



Propiedades de la conexión

Origen de datos:
Microsoft SQL Server Analysis Services (Adc) Cambiar...

Nombre del servidor:
DESKTOP-TBIH0UF\MSSQLSERVER2

Iniciar sesión en el servidor

Nombre de usuario:

Contraseña:

☐ Guardar mi contraseña

Conectar con una base de datos

Seleccione o escriba un nombre de base de datos:
Fase2_Esqueletos

Avanzadas...

Probar conexión Aceptar Cancelar

En el diseñador de consultas, colocamos los siguientes campos:

Year	Valor
1968	34.29
1980	57.14
1994	11.43
2000	2319.97
2001	91.43
2002	240.01
2003	1923085.899999989
2004	3991543.11000015
2005	4312665.43000001
2006	6079723.08999998
2007	8248745.01000076
2008	5223299.76000001
2012	502.91

Diseñamos la tabla de la siguiente forma:

Diseñar la tabla

Elija cómo agrupar los datos en la tabla.

[illegible]

Al previsualizar el reporte, tenemos lo siguiente:

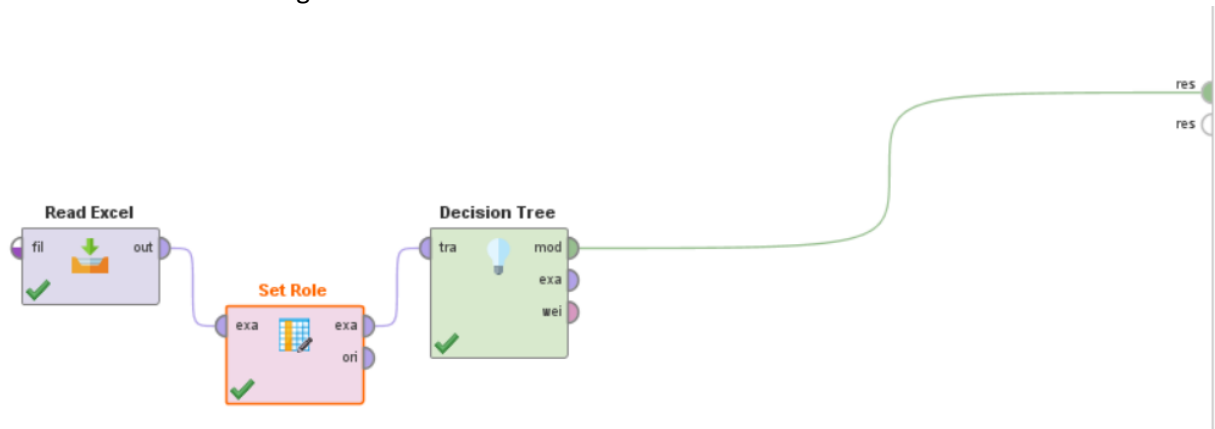
Reporte_por_año

Year	Valor
1968	
	34,29
1980	
	57,14
1994	
	11,43
2000	
	2319,97
2001	
	91,43
2002	
	240,01
2003	
	1923085,89999 989
2004	
	3991543,11000 015
2005	
	4312665,43000 001
2006	
	6079723,08999 98
2007	
	8248745,01000 076
2008	
	5223299,76000 001
2012	
	502,91

Según la información que se ha analizado, el año en el que se dio pago el mayor valor con respecto a esquilas fue el año 2007, y el menor valor fue el año 1994, hay que recordar que esto tiene sentido, ya que antes, las personas ocupaban menos el vehículo, se podría decir que antes tener un auto era un lujo, ya que hoy en día un auto se vuelve algo necesario, por lo que podemos ver el incremento de los años 90 al año 2007, sin embargo del entre los años 2003-2012 podemos observar como el valor pagado por las esquilas es muy superior con respecto a los de los años anteriores al 2003.

Parque Vehicular

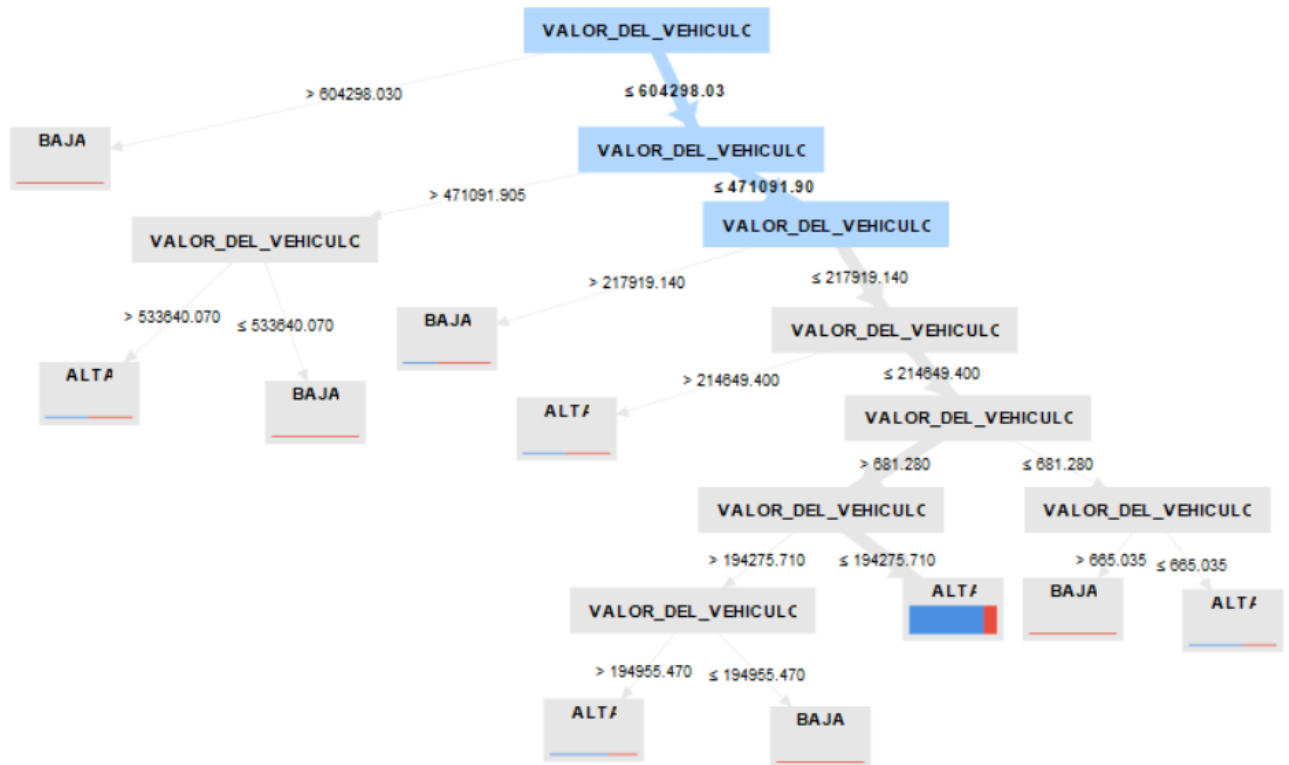
Para este análisis usaremos el software rapid miner.
Para ello colocaremos los siguientes elementos:



Haremos 2 ejemplos sencillos con árbol de decisión, ya que un ejemplo muy complejo sería complicado colocar la captura.

Árbol de decisión comparando valor de vehículo para clasificar el vehículo en estado alto o bajo.

Resultado del árbol de decisión:



En este diagrama de árbol solamente debemos interpretar cada una de las condiciones.
En caso de que el valor de el vehículo sea mayor a 604298.030 se va a clasificar en estado bajo.

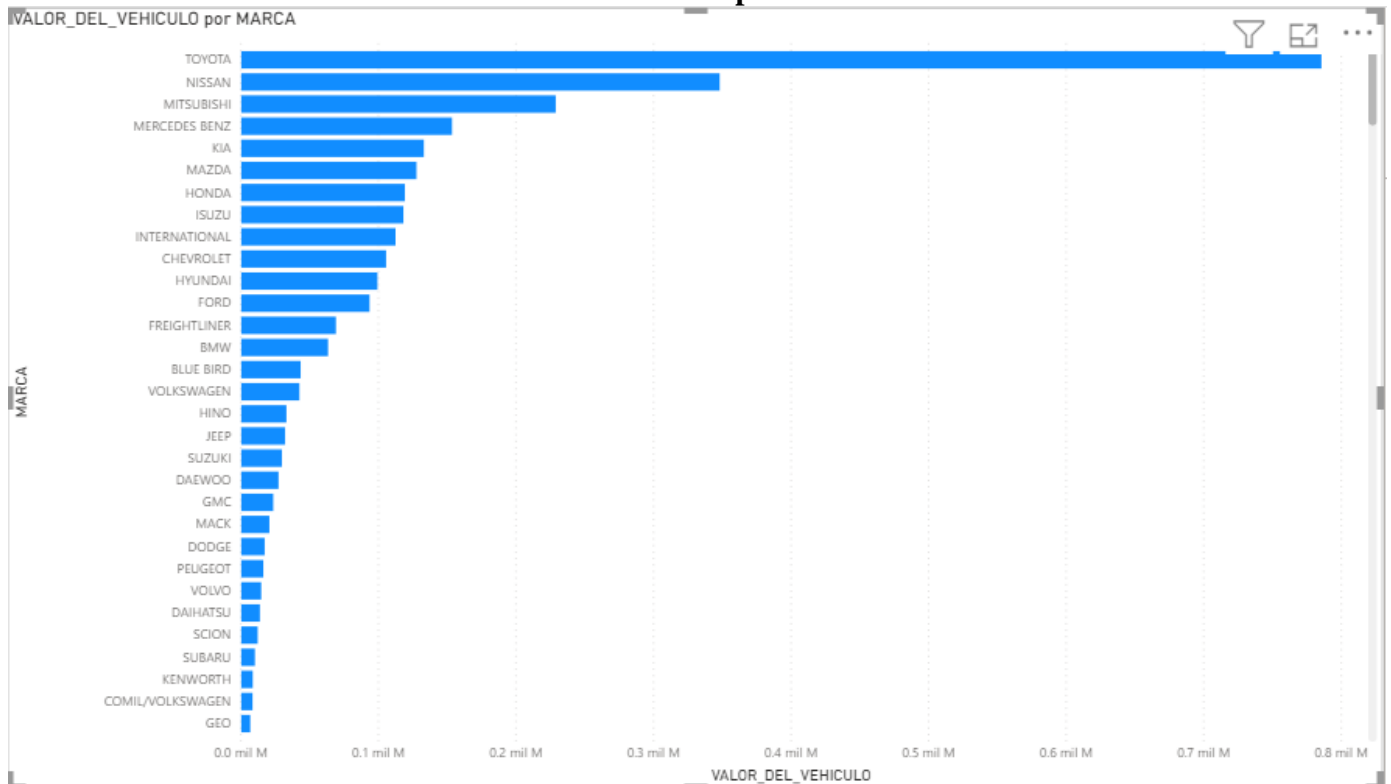
Árbol de decisión comparando año de fabricación para clasificar el vehículo según su condición de ingreso.



En este diagrama de árbol podemos observar que según el año de fabricación hay X probabilidad de que el auto sea usado o nuevo, en lo que se ve en el diagrama hay más probabilidad de que el auto sea usado, esto quiere decir que El Salvador la mayoría de los autos que hay son usados.

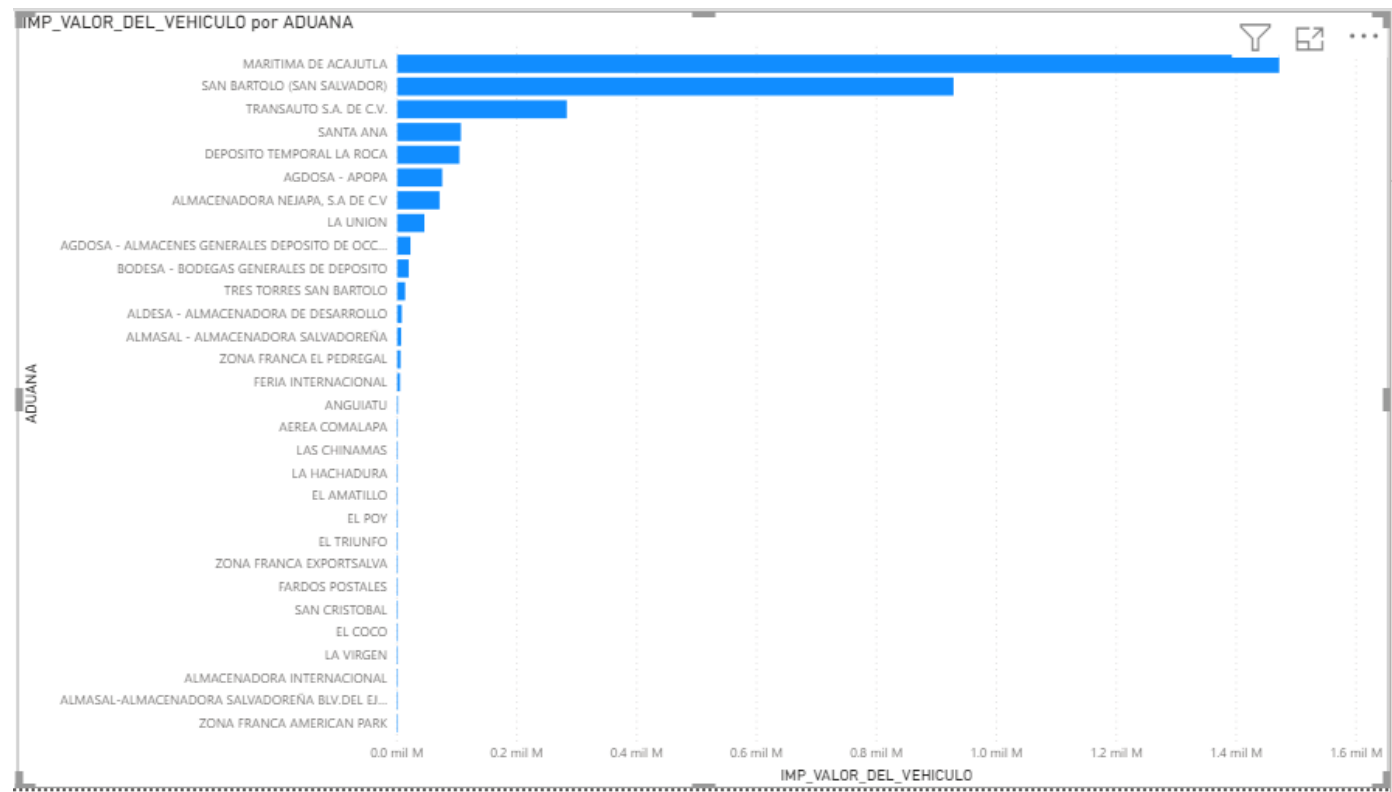
Análisis con power bi

Valor del vehículo por marca



Mediante el gráfico anterior, podemos observar que la mayoría de los autos del parque vehicular son de la marca toyota.

Impuesto del valor del vehículo por aduana.



Esto quiere decir que la aduana marítima Acajutla es la aduana que más impuestos obtiene de los vehículos, por lo que la mayoría de los vehículos vienen de dicha aduana.