UDB VIRTUAL

Universidad Don Bosco, El Salvador

Datawarehouse y Minería de Datos

Proyecto de clase

Docente: Inga. Karens Medrano

Nombre completo de integrantes del equipo de trabajo:	Número de carné:
José Ernesto Sorto González	**SG202883**

Formulación del Problema

Se cuentan con dos bases de datos en formato “csv” que se deberá transformar para la interpretación de la información.

La primera base para analizar trata de datos reunidos de esquelas de tránsito en El Salvador, se deberá transformar la información de tal manera que se puedan presentar informes para su posterior análisis.

La segunda base de datos a tratar consiste en información acumulada del parque vehicular en El Salvador el cual muestra la información de número de unidades vehiculares registradas por el gobierno.

En este contexto, se necesita conocer a través de un sistema de alta gerencia que les permita el ahorro de tiempo al consultar información histórica para la toma de decisiones.

Objetivo General

Identificar características y patrones de comportamiento relacionadas con las bases de datos propuestas, a través del proceso de descubrimiento de conocimiento en bases de datos y herramientas de Minería de Datos.
Los modelos identificados se propondrán como contribución a la toma de decisiones en el ámbito de la gestión de información

Objetivos Específicos

Utilizar herramientas de minería de datos para detectar patrones y relaciones entre los datos propuestos.

Elaborar recomendaciones sobre los posibles usos de los resultados obtenidos y las características de las fuentes de información que sirvan como estrategias innovadoras y apropiadas para la gestión de información.

Justificación

Dentro de la industria de desarrollo de software existen los sistemas de inteligencia de negocios que permite a los usuarios disponer de información de una manera rápida para poder tomar decisiones.

En este trabajo se presentará una propuesta para diseñar un cubo OLAP para la base de datos de esquelas vehiculares que permita obtener
resultados oportunos en tiempo real a la alta gerencia para tomar correctas decisiones. La puesta en marcha de este sistema, conllevará diferentes beneficios y así generar buenos resultados en el área de información.

MARCO TEÓRICO

Inteligencia Empresarial

Es una mezcla de tecnologías, herramientas y técnicas que acceden a convertir los datos almacenados en información y la información en conocimiento, destinado a mejorar el paso de toma de decisiones en los negocios.

Con esto dentro de la inteligencia de negocios encontramos como una estrategia el manejo, control y asimismo gestión de información. Las herramientas de inteligencia de negocios le resultan beneficioso para una empresa, así como a aumentar la eficiencia de ella, por ende, se obtienen mejores resultados.

Beneficios de la integración de Inteligencia Empresarial

- Se toman mejores estrategias de alto nivel con respecto a las ventas por ejemplo de productos.
- Se genera la reducción de los valores de costes y aumentan los ingresos, puede disminuir actividades.
- Muchos usuarios podrán obtener información oportuna y disponible para el logro de una buena toma de decisión.

Datamart.

Es considerado como una parte de la información de la empresa, esto quiere decir que tiene información de un solo departamento en específica se puede realizar un análisis a partir de varias perspectivas.

Datamart son subconjuntos de DataWarehouse diseñados para satisfacer necesidades determinadas de un área de la organización. Ya que su función es especificar la necesidad de datos seleccionados, destacando el fácil acceso a una información relevante.

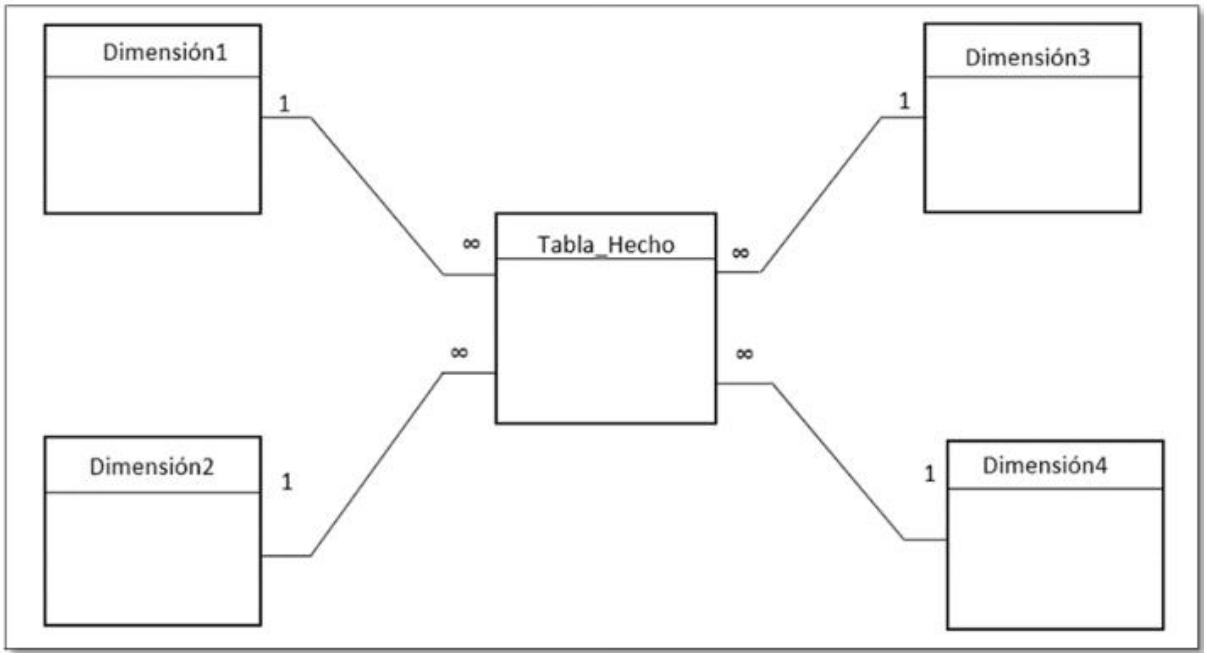
Características:
Algunas de las características de integración de los Datamart son

- No volátiles: la información no se modifica, ni se elimina.
- Se actualizan constantemente.
- Contiene información detallada.

Metodologías Multidimensionales

Un esquema multidimensional puede tener dos esquemas uno es el copo de nieve y la otra estrella, la información es agrupada por dimensiones.

Este modelo logra que los datos multidimensional sean representados en una base relacional, las dimensiones contienen información que describen las relaciones, y la tabla de hechos contiene datos numéricos.



Cubos OLAP

Los cubos OLAP pueden generar información para un análisis desde las dimensiones, así como varios niveles.

Existen dos ventajas para la utilización de los cubos:

- Facilidad de manipulación de la información: una vez ejecutado el cubo el cliente podrá manipularlo con facilidad así no tenga conocimientos técnicos. Esta estructura es muy fácil de comprender y manipular.
- Respuesta Ágil: una vez que el cubo haya sido creado de una manera correcta, al momento de hacer las consultas necesarias, la información podrá estar disponible en tiempo real.
-

Herramientas para construir soluciones de BI

Las herramientas de BI son aplicaciones que están creadas para apoyar durante la presentación y el análisis de los datos. En la actualidad las herramientas son mucho más modernas y potentes, poseen la capacidad de poder procesar y analizar grandes cantidades de datos y esto es de ayuda para las empresas para llegar al logro de conclusiones que ayuden a la empresa a estar en un mejor nivel competitivo. A continuación, se presentan las

siguientes herramientas: Microsoft SQL Server, Microsoft Integration server, Microsoft Analysis Services, Power BI y Excel.

Microsoft SQL Server

Es un servidor de base de datos creado por Microsoft. En el centro de SQL server se encuentran localizados lo motores SQL Server los cuales son los encargados de procesar los comandos de bases de datos se lo utiliza como herramienta para el análisis de la información. Ofrece escalabilidad, seguridad y fiabilidad suficiente para lograr poner en ejecución cualquier aplicación en poco tiempo, Microsoft SQL Server se caracteriza por la capacidad que tiene para el análisis de información y por sus tareas sencillas de administración, establece una solución completa, productiva y confiable para BI.

Características de Microsoft SQL Server

- Contiene procedimientos almacenados
- Posee seguridad, escalabilidad y estabilidad.
- Soporta transacciones.
- Posee un entorno grafico permite uso de comando DDL y DML.
- Se puede trabajar de manera cliente-servidor.
- Puede administrar la información de cualquier otro servidor de datos.

Power BI

Power BI es un servicio de análisis para las empresas, ayuda a generar información de una manera detallada para lograr conceder la toma de decisiones y estas puedan ser rápidas y con informes como lo podemos observar en la figura 5, ayuda con la supervisión del estado de la empresa, entre los beneficios de utilizar Power BI (Microsoft, 2019) menciona los siguientes:

- Creación de informes interactivos
- Posee una interfaz intuitiva que cualquier persona puede utilizar
- Transformar los datos en objetos visuales
- Explorar y analizar los datos de manera local y en la nube
- Se obtiene información de predicciones y tendencias.

Association Rules

Búsqueda de patrones frecuentes, asociaciones, correlaciones o estructuras causales entre conjuntos de elementos u objetos en bases de datos de transacciones, bases de datos relacionales y otros repositorios de información disponibles.

Aplicaciones:

- Análisis de datos de la banca.

- Cross-marketing (poner la crema batida junto a las fresas).
- Diseño de catálogos.

Arboles de Decisiones

En minería de datos, un árbol de decisión sirve para abordar problemas tales como la clasificación, la predicción y la segmentación de datos con la finalidad de obtener información que pueda ser analizada para tomar decisiones futuras.

Si trasladamos el concepto al área de Business Analytics, los árboles de decisión se utilizan mayoritariamente para predecir las probabilidades de alcanzar un resultado en función de unas variables de entrada tales como edad, sexo, demografía o ingresos que indicarán, por ejemplo, si el cliente es apto o no para recibir un préstamo.

Antecedentes

Según (Tello & Velasco, 2016), en la actualidad en el entorno empresarial es normal ver que cada vez es mayor la cantidad de información y las bases de datos de las empresas poseen grandes cantidades de información, tanto el análisis como los requisitos de información para la toma de decisiones ha aumentado la demanda de software y soluciones de análisis de BI.

METODOLOGÍA

El primer paso que se realizó para aplicar la metodología fue recolectar información, una vez que se ha recolectado la información necesaria para el desarrollo de nuestra propuesta, se dio el paso al procesamiento de los datos para el proyecto técnico planteado que consistió en consolidar toda la información en una Datamart

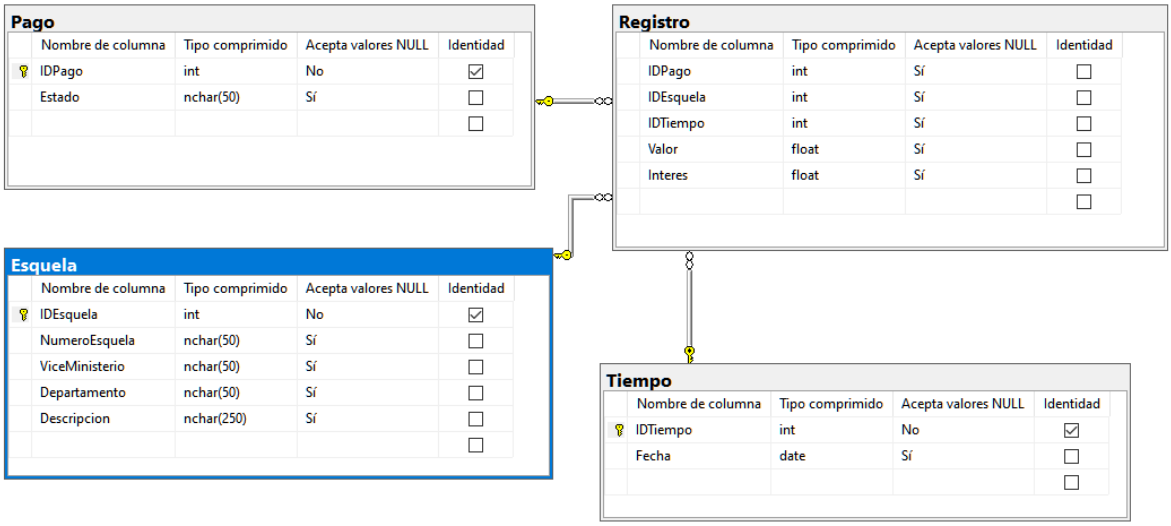
PROPUESTA DE SOLUCIÓN

Diseño de un cubo OLAP para el análisis de la base de datos de Esquelas vehiculares.

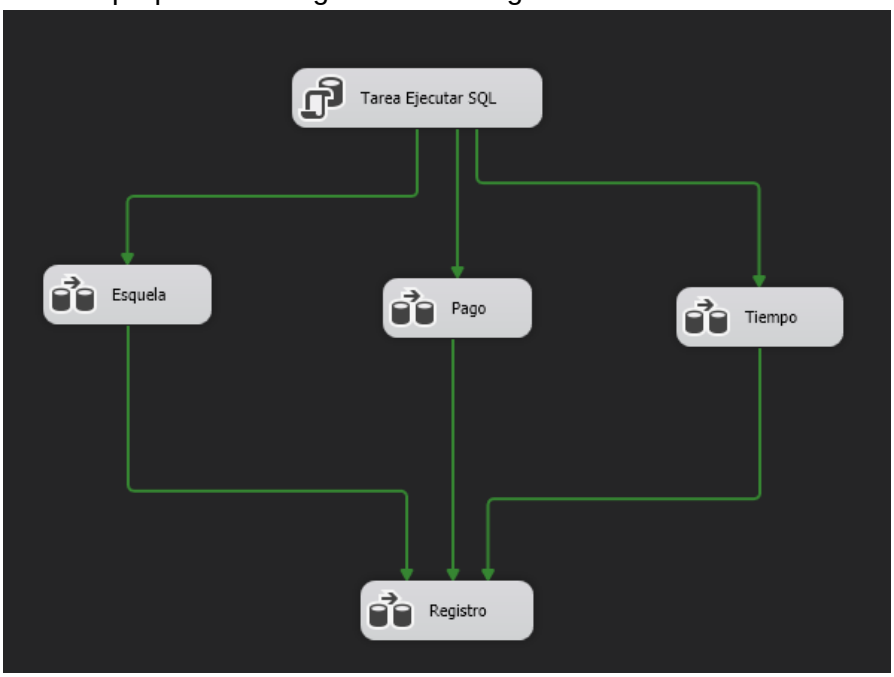
Descripción de la propuesta de solución 1

Una vez que se han establecido los requisitos que el sistema va a requerir, procedemos a realizar los análisis respectivos para llegar al logro de la solución. Se hizo uso de la base transaccional de la base de datos original.

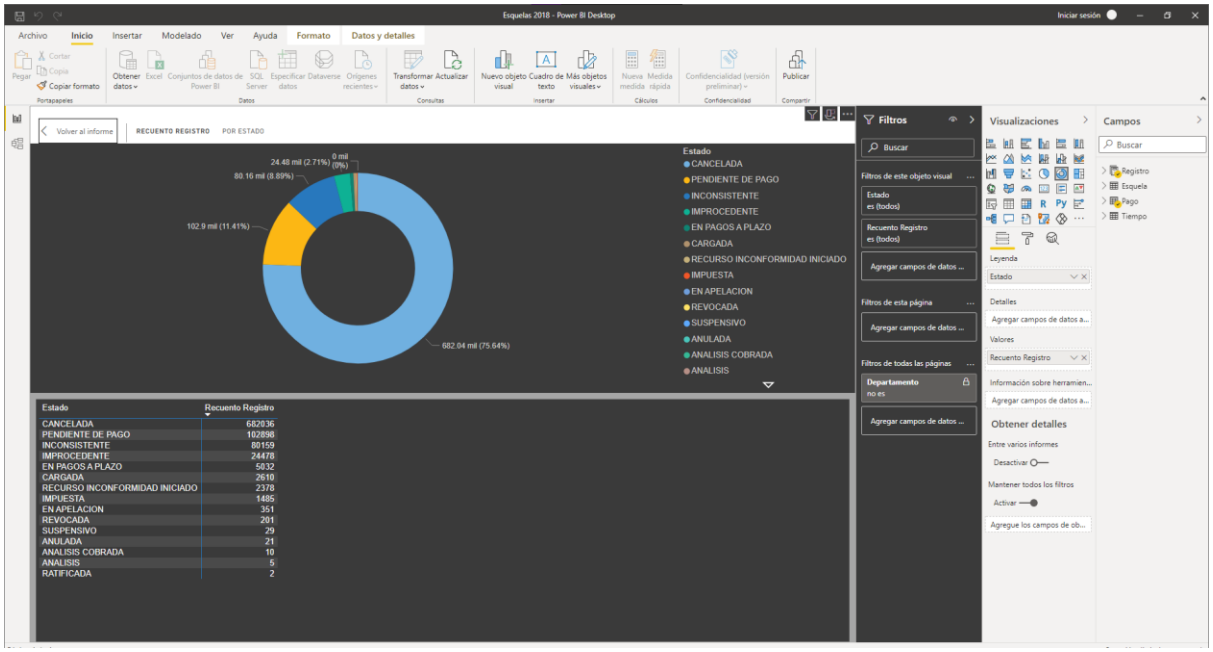
Esquema estrella utilizado.



Construcción de los paquetes de migración en Integration Services



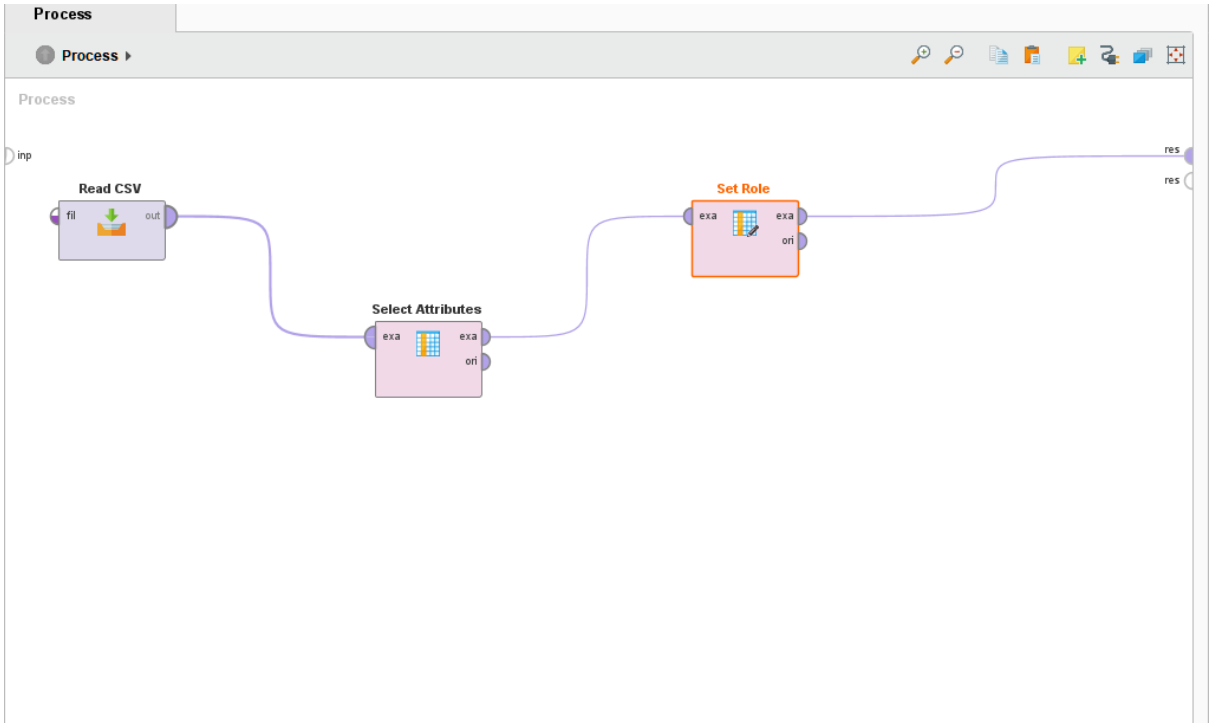
Presentación de informes en Power BI de los datos del cubo OLAP



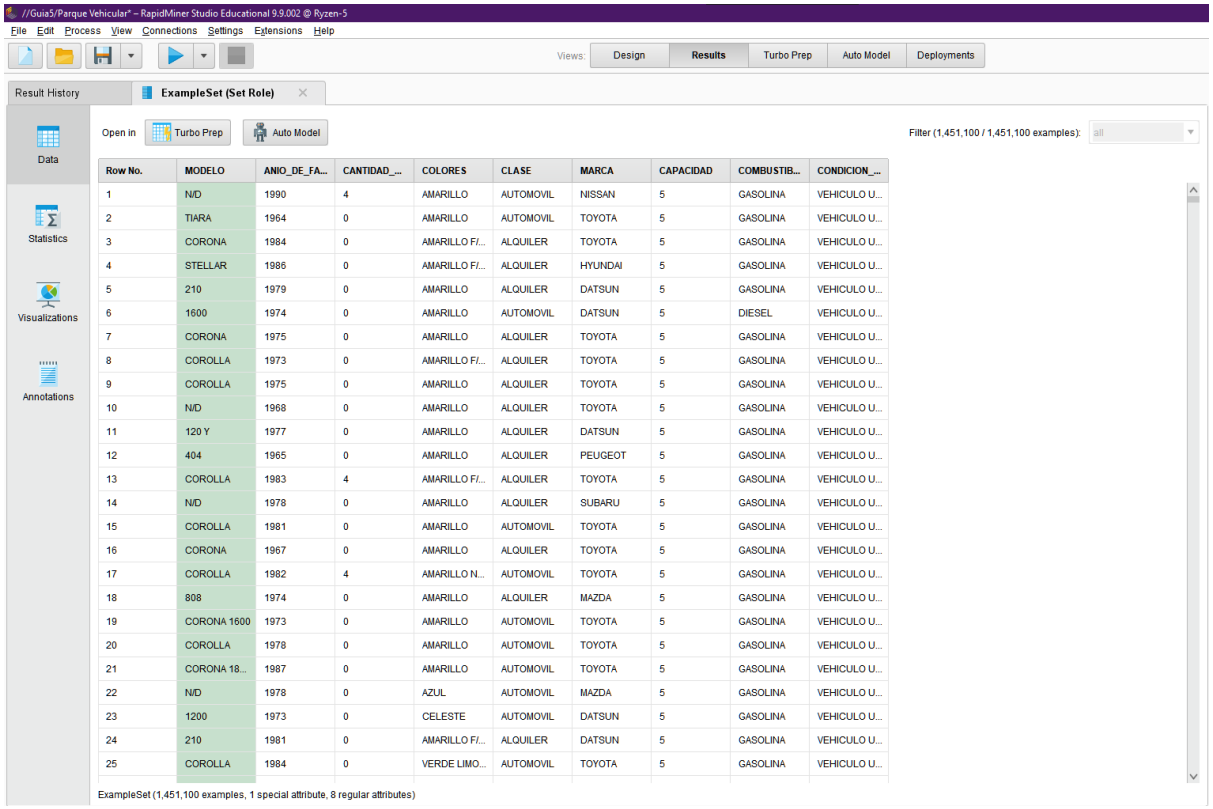
Descripción de la propuesta de solución 2

Diseño de reglas de asociación y árbol de decisión para la base de datos parque vehicular.

Reglas de asociación.

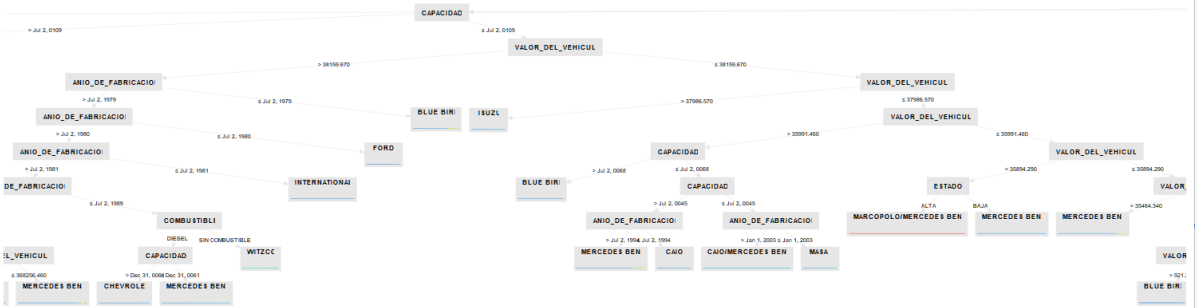


Resultados.



Visualización de los datos generados por el modelo creado.

Creación del árbol de decisión.


CAPACIDAD
> Jul 2, 0109
≤ Jul 2, 0109
VALOR_DEL_VEHICUL
> 38159.670
≤ 38159.670
ANIO_DE_FABRICACIO
VALOR_DEL_VEHICUL
> Jul 2, 1979
≤ Jul 2, 1979
≤ 37986.570
> 37986.570
BLUE BIR
ISUZU
ANIO_DE_FABRICACIO
VALOR_DEL_VEHICUL
> Jul 2, 1980
≤ Jul 2, 1980
> 35991.460
≤ 35991.460
FORD
ANIO_DE_FABRICACIO
CAPACIDAD
VALOR_DEL_VEHICUL
> Jul 2, 1981
≤ Jul 2, 1981
> Jul 2, 0068
≤ Jul 2, 0068
> 35894.290
≤ 35894.290
DE_FABRICACIO
INTERNATIONAI
BLUE BIR
CAPACIDAD
ESTADO
VALOR
≤ Jul 2, 1989
> Jul 2, 0045
≤ Jul 2, 0045
ALTA
BAJA
COMBUSTIBLI
ANIO_DE_FABRICACIO
ANIO_DE_FABRICACIO
MARCOPOLO/MERCEDES BEN
MERCEDES BEN
MERCEDES BEN
> 35464.340
DIESEL
SIN COMBUSTIBLE
> Jul 2, 1994
≤ Jul 2, 1994
> Jan 1, 2003
≤ Jan 1, 2003
L_VEHICUL
CAPACIDAD
WITZCC
MERCEDES BEN
CAIO
CAIO/MERCEDES BEN
MASA
VALOR
≤ 368256.460
> Dec 31, 0068
≤ Dec 31, 0081
> 921.
MERCEDES BEN
CHEVROLE
MERCEDES BEN
BLUE BIR