



Universidad Nacional Autónoma de México
Facultad de Ingeniería
División de Ingeniería Eléctrica
Ingeniería en Computación



Proyecto Final
**“Aplicación de Minería de Datos para la Generación de
Promociones Oportunas en la Venta de Autos de Segunda Mano”**

Integrantes:

- Castelan Ramos Carlos

Materia: Minería de Datos

Grupo: 03

Semestre: 2024-1

Fecha de entrega: 09 de noviembre 2023



Proyecto Final “Aplicación de Minería de Datos para la Generación de Promociones Oportunas en la Venta de Autos de Segunda Mano”

Descripción de la Empresa.

Estamos trabajando con una empresa llamada "AutoSelect", que se dedica a la compra y venta de coches de segunda mano. AutoSelect opera a nivel nacional en España y tiene una plataforma en línea donde los usuarios pueden buscar autos de acuerdo con sus necesidades y presupuesto, a su vez los usuarios pueden enlistar sus propios autos para venta siempre y cuando la empresa verifique las características y los puntos de calidad del auto que beneficien su posterior venta.

Descripción del Caso de Negocio.

El objetivo de AutoSelect es optimizar su criterio de compra de autos y estrategia de precios. Para lograrlo, la empresa quiere utilizar la minería de datos para entender mejor cómo varios factores afectan los precios de los coches usados y en última instancia, ajustar sus precios para maximizar las ventas y los beneficios.

Proceso de Minería de Datos (CRISP-DM).

Este proceso aplicado se puede observar en: https://colab.research.google.com/drive/1yp-bZgeS4DtDgp9_iaiOpqCVoSrZLf6o?usp=sharing

a) Comprensión del Caso de Negocio.

Como ya se mencionó AutoSelect es una empresa que se dedica a comprar y vender autos de segunda mano, donde la demanda de estos y la competencia del rubro han ido aumentando, es por eso por lo que es importante comprender el comportamiento de los compradores para así manejar la selección de autos de una mejor forma, buscando reducir costos en reparaciones y estadías de los mismos.

Es entonces que debemos establecer objetivos en conjunto con Autoselect para definir las actividades a realizar posteriormente en la Minería de Datos.

Definición de objetivos:

- Establecer criterios puntuales para la elección de la compra de autos (año, marca, modelo, kilometraje, color, número de puertas, transmisión, etc.)
- Predecir precios futuros para la compra y venta de los autos.
- Establecer promociones oportunas para la venta y compra de autos.

b) Comprensión de los Datos.

En esta etapa, se han obtenido datos de un repositorio de datos sobre ventas de autos de segunda mano en las principales plataformas de venta en línea (Facebook Marketplace, Ebay, Mercado Libre, etc.).

Datos obtenidos de: <https://datamarket.es/#vehiculos-de-segunda-mano-dataset>

Las características de este dataset son las siguientes:

- Frecuencia de actualización: actualizado cada 24h.
- Volumen estimado: 300.000 registros cada día.
- Histórico: desde noviembre de 2020.



Dentro de los datos se considera una gran variedad de características como: Marca, modelo, versión, precio, precio financiado, combustible, año, kilómetros, potencia, puertas, transmisión, color, concesionario, provincia y país.

c) Preparación de los Datos.

Es importante entender la calidad de los datos, identificar si hay valores faltantes o datos erróneos, y definir las variables clave que pueden influir en los precios de los autos. Es por ello por lo que se lleva a la limpieza de los datos, donde se han eliminado columnas no necesarias y que generan confusión en su análisis. Columnas eliminadas:

- company
- url
- publish_date
- insert_date

Teniendo listos los datos para manejar, es necesario identificar los datos principales a observar y analizar, siendo estos los más comunes para la elección de un auto.

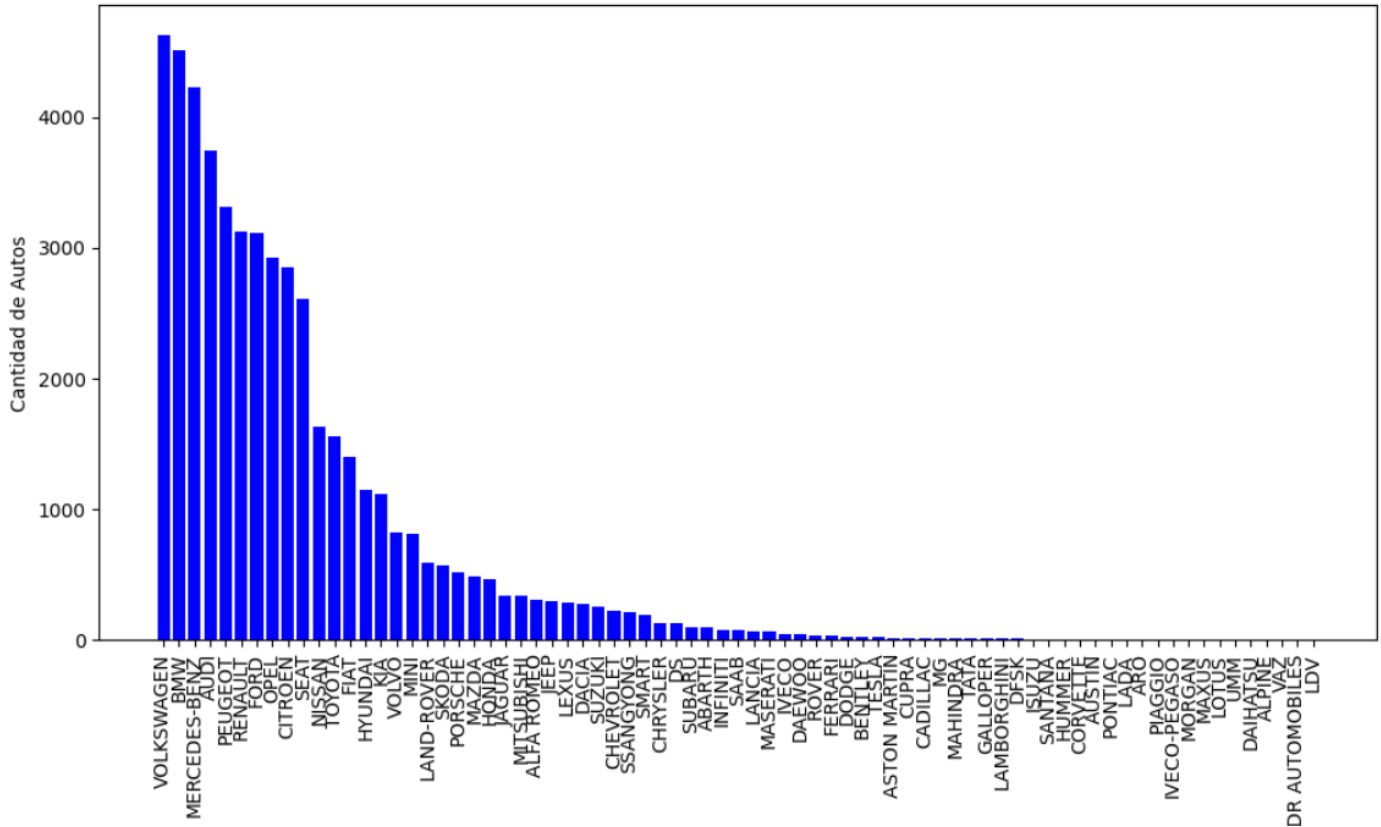
- Marca
- Modelo
- Precio
- Combustible
- Año
- Puertas
- Transmisión
- Color
- Provincia de venta

d) Identificación de Tareas de Minería de datos.

Para esta etapa vamos a definir la aplicación de diversos métodos para comprender de diferentes maneras el comportamiento de los datos.

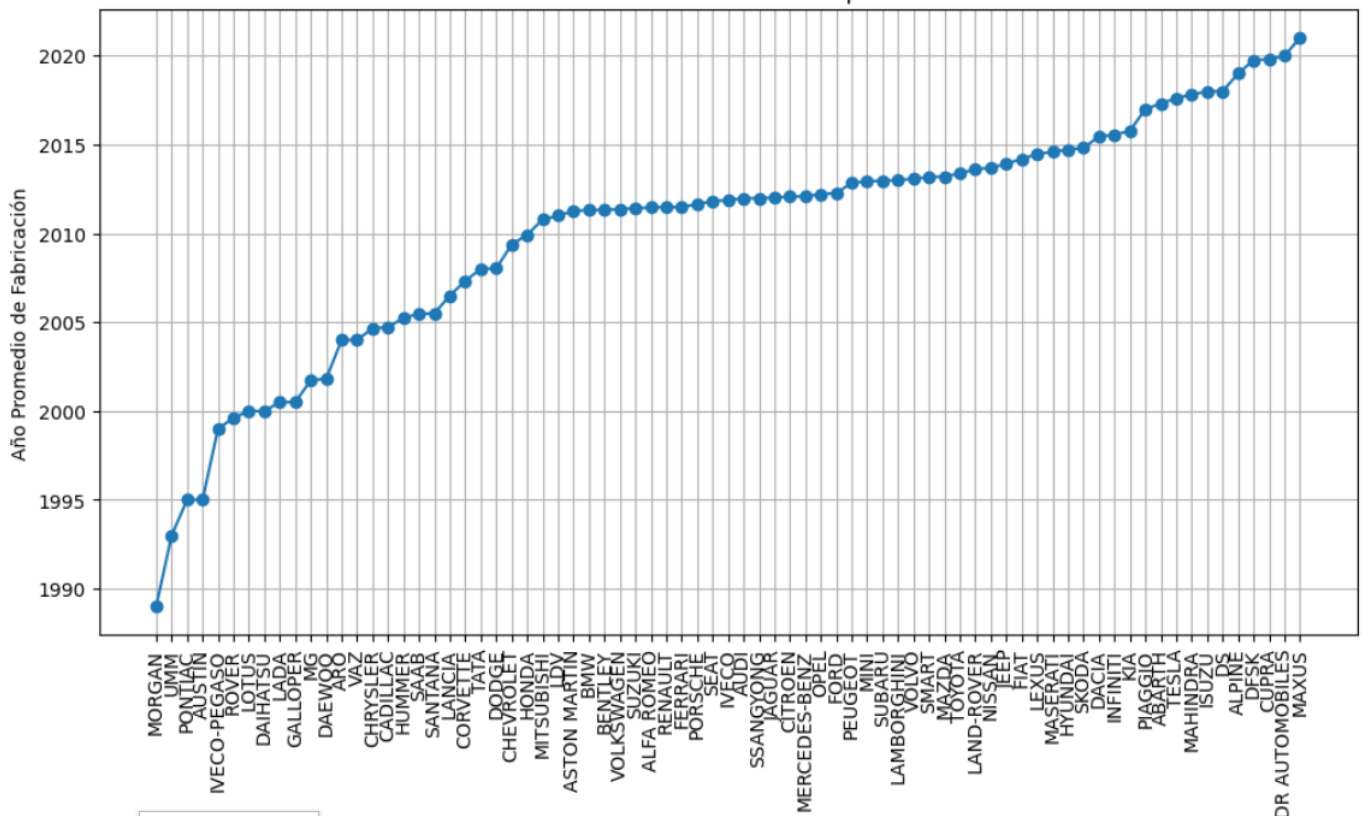
- Distribución de Marcas de Autos Vendidos desde noviembre del 2020 a la actualidad.
En primera instancia, observaremos las marcas que más se han vendido de segunda mano en este lapso de tiempo, dado que gran parte de las marcas de autos comparten grupos familiares, es necesario identificar aquellas marcas que pueden compartir piezas para el ahorro de costos en reparaciones pequeñas.

Distribución de Marcas de Autos



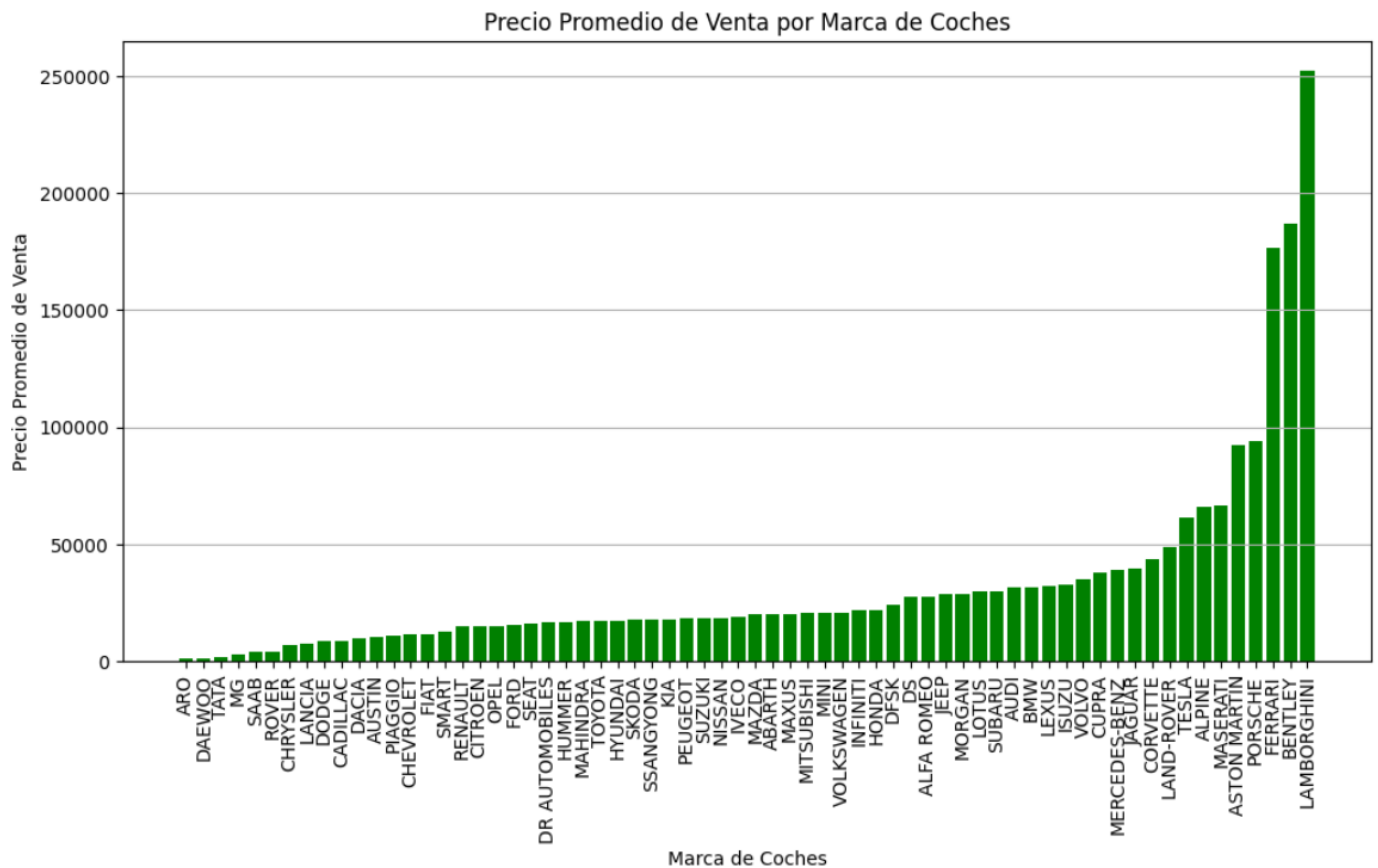
- Tendencia de preferencia de marcas de acuerdo con su año de fabricación.
Se realizó la media de la fecha de fabricación de los autos por marca. Observamos que la tendencia de compra de autos de segunda mano suele ser en la mayoría de los casos con un año de fabricación posterior al 2010.

Tendencia de Preferencia de Marca de Coches por su año de Fabricación

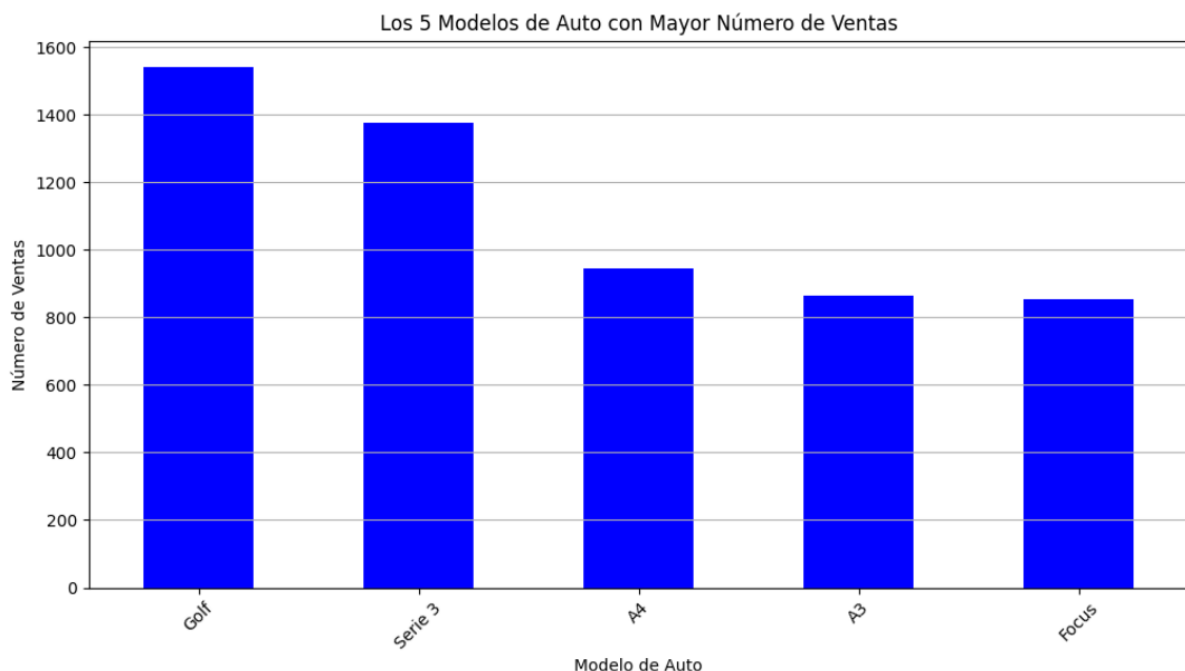


- Precio promedio de venta para autos de segunda mano con kilometraje menor a 80,000 km. Se calculó el promedio de kilometraje de los autos vendidos en el lapso siendo el siguiente:
Promedio de Kilometraje de autos vendidos= 116,359.09 km

En el algoritmo, implementado se consideró como autos de segunda mano de buena calidad a aquellos con un kilometraje de 80,000 kilómetros, variando la cantidad de kilómetros no se observa gran cambio en los precios promedios por marca de coches sin rebasar el promedio de kilometraje calculado.



- Los 5 modelos más vendidos.





Evaluación del modelo:

Para evaluar el modelo se realizó uso de el cálculo del error cuadrático medio, donde los resultados fueron los siguientes.

Error cuadrático medio de la validación: 4454.905€

Error cuadrático medio del entrenamiento: 1964.275€

- El RMSE del conjunto de validación es de 4454.905€, lo cual indica la raíz cuadrada de la varianza de los errores de predicción en el conjunto de validación. Cuanto menor sea el RMSE, mejor será el rendimiento del modelo. En este caso, el valor parece relativamente bajo en comparación a los precios que se manejan para los autos más vendidos.
- El RMSE del conjunto de entrenamiento es de 1964.275€, lo cual es más bajo que el del conjunto de validación. Esto era de esperar, ya que el modelo se entrenó precisamente en estos datos. Sin embargo, la gran diferencia entre los errores de entrenamiento y validación podría indicar cierto sobreajuste (overfitting), donde el modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos.

Coeficiente de Determinación (R^2):

Coeficiente de determinación (R^2) del conjunto de validación: 0.91

Coeficiente de determinación (R^2) del conjunto de entrenamiento: 0.985

- El R^2 del conjunto de validación es 0.91, lo cual es bastante alto y sugiere que el modelo explica aproximadamente el 91% de la variabilidad en los datos de validación. Esto indica una buena capacidad predictiva del modelo en el conjunto de validación.
- El R^2 del conjunto de entrenamiento es aún más alto, 0.987, lo cual es esperado ya que el modelo se entrenó con estos datos. Sin embargo, la gran diferencia entre los valores de R^2 de entrenamiento y validación nuevamente sugiere la posibilidad de sobreajuste.

f) Interpretación y evaluación de los datos.

A partir del análisis básico de autos observamos que las marcas más vendidas son:

- Volkswagen
- BMW
- Mercedes Benz
- Audi
- Peugeot
- Renault

Por otro lado, los 5 modelos más vendidos son:

- Golf
- Serie 3
- A4
- A3



- Focus

El costo de los autos más vendidos está en un rango de 40,000 € a 75,000 €.

Así entonces estos autos suelen ser de una fecha de fabricación de los años 2010 al 2015.

Teniendo estas premisas podemos realizar las siguientes conclusiones:

- Comprar autos Volkswagen y Audi tienen un buen rendimiento en compra y venta, en primera instancia podemos decir que comparten grupo financiero por lo que la compra y venta de piezas es compatibles para ambas marcas, donde la contratación de especialistas de la marca sería una buena opción, desde mecánicos hasta proveedores.
- Comprar autos Renault y Peugeot es rentable también, ya que son autos baratos que se venden rápido y que además de ser franceses, comparten grupo financiero y la obtención de piezas lo hace muy factible. La cantidad de especialistas es grande por la cercanía del país de origen.
- Marcas de prestigio como BMW, Mercedes Benz y Audi se venden continuamente, sin embargo, se recomienda únicamente comprar automóviles de estas marcas de años del 2010 – 2013 ya que si se compran más recientes los precios de compra y venta se elevan demasiado lo que hace difícil sacarlos del stock.

Promociones recomendadas.

- Se recomienda establecer un centro de refacciones oficial de las marcas Volkswagen y Audi.
- Se recomienda establecer promociones para la compra y venta de autos Volkswagen y Audi, tales como servicios gratis en la compra de insumos.
- Descuentos por Temporada: Ofrecer descuentos especiales o promociones específicas para temporadas clave, como eventos de cambio de estación o promociones de fin de año.
- Programa de Referidos: Recompensar a los clientes actuales que refieran a nuevos clientes con descuentos, servicios gratuitos o regalos.
- Financiamiento Especial: Colaborar con instituciones financieras para ofrecer tasas de interés preferenciales o planes de financiamiento especiales para la compra de autos.
- Inspección y Mantenimiento Gratis: Ofrecer servicios de inspección y mantenimiento gratuitos para los autos comprados en el centro durante un período específico.
- Paquetes de Accesorios Gratuitos: Incluir paquetes de accesorios gratuitos, como kits de entretenimiento, sistemas de navegación, o servicios de mantenimiento, con la compra de un auto.