

Equipo de trabajo

Roles para el desarrollo del Proyecto

Para efectos de este trabajo los roles se definen de la siguiente forma:

- Líderes: Jorge Guerrero / Paulo Gonzalez
- Analistas de datos = Esteban Sanchez / Gonzalo Rojas
- Control de calidad y validación de datos = Esteban Sanchez / Gonzalo Rojas

Sepan con que estan trabajando por ello es que son los mismos.

- Ingeniero/a de modelamiento = Daniel Mardones / Jorge Guerrero
- Visualización de datos = Daniel Mardones / Jorge Guerrero
- Documentador/a = Daniel Mardoones / Paulo Gonzalez

Selección del tema y motivación

Se selecciona el tema propuesto Automotora, debido a que es una necesidad replicable, universal y contemporanea a cualquier epoca. La motivación del proyecto es lograr generar una herramienta que proporcione un apoyo a la toma de decisions de compra y venta, para obtener las mejores opciones presentes en el mercado, mejorando la rentabilidad del negocio.

Planificación de la investigación

Requerimientos:

- # 1. Generar una herramienta que ayude al equipo de compras a encontrar las mejores oportunidades.

Atraves de un modelo de clasificación que nos permita encontrar la mejor rentabilidad, usando técnicas de ML.

2. Definir una propuesta de inversión (lista de automóviles convenientes) para la Automotora.

Por medio de las metricas obtenidos podremos definir una propuesta para inversion, mediante vehiculos que esten por debajo del precio de venta promedio, lo cual permite un mejor margen de venta.

3. Definir una estrategia de venta por gama de vehículos.

Se incorpora una columna la cual clasifica la marca del vehiculo en generalista o premium. Con esto podremos generar un analisis del comportamiento en la venta de vehiculos y con esto desarrollar una estrategia.

4. Encontrar oportunidades de compra y venta entre ciudades/estados.

Atraves de un EDA, podemos determinar en que ciudades / estados se encuentran las mejores oportunidades de compra y determinar las mejores ciudades / estados para vender estos vehiculos, maximizando la utilidad.

Hipotesis

Nuestra hipótesis es poder determinar, mediante un modelo clasificatorio si es rentable la compra de un vehículo, de acuerdo a sus especificaciones, para una posterior venta.

Definición del Vector Objetivo

El vector objetivo es la variable Price.

Estrategias analíticas a nivel descriptivo

Se desarrollará un EDA, el cual nos visualizara la distribución de vehículos de acuerdo a la marca, modelos, año, ciudad, estado, millas y valores de vehículos.

Modelación y predicción de trabajo

Como primera estrategia utilizaremos la librería Lazy Predict, utilizando su implementación estándar, para conocer de forma preliminar los mejores clasificadores para esta solución. Para almacenar los resultados, utilizaremos una base de datos Postgres, lo que permite obtener mediante una query realizada por web, si es conveniente la compra del vehículo.

Consideraciones

- Los datos no serán encriptados.
- Los set de datos no contienen datos faltantes.
- Existen 3 dataset, los cuales por origen se encuentran divididos en: muestra total = true_car_listings.csv muestra entrenamiento = true_cars_train.csv muestra validación = true_cars_test.csv Debido a esto, no se realiza divisiones de muestra, solo de vector objetivo y matriz de atributos.

