



DATA SCIENCE – CODER HOUSE

Manuel Muñoz – Venta de vehículos en USA



Introducción:

I.I.- Descripción Temática

En el competitivo mercado de automóviles usados de los Estados Unidos, tanto compradores como vendedores buscan obtener la mejor oferta posible. ¿Qué factores realmente influyen en el precio de un vehículo? ¿Es el kilometraje? ¿El año de fabricación? ¿O tal vez la marca tiene un peso mayor en la decisión final de compra? Este análisis surge con el objetivo de responder a esas preguntas y brindar claridad a los actores involucrados en la compra-venta de autos.

Partimos de un amplio conjunto de datos que incluye diversas características de los vehículos vendidos. Este estudio busca desentrañar las relaciones clave entre estos atributos —como la marca, el modelo, el kilometraje, el estado del vehículo, entre otros— y su impacto en el precio de venta. Al final de este análisis, no solo conoceremos qué factores son más determinantes, sino que también seremos capaces de predecir con precisión el precio de un automóvil basado en sus características particulares.

Este enfoque analítico no se queda en lo teórico. Pretendemos que este modelo predictivo sea una herramienta práctica que permita a concesionarios, vendedores particulares y plataformas de venta online estimar precios de venta con mayor exactitud. Para ello, utilizamos técnicas avanzadas como la regresión lineal con múltiples variables, que nos permitirá captar cómo estas interacciones influyen en el valor final del vehículo.

I.II.- Hipótesis:

La hipótesis nula que guía este análisis establece que no existe una relación significativa entre las variables del vehículo —como marca, modelo o año— y el precio final de venta. En otras palabras, se asume que estas características no influyen de manera relevante en el valor del automóvil.

No obstante, en nuestro análisis intentaremos demostrar lo contrario: que las características del vehículo tienen un impacto directo y discernible en su precio. Este conocimiento no solo nos permitirá comprender mejor el comportamiento del mercado, sino que también nos permitirá ofrecer predicciones más precisas para futuros precios de venta.

I.III.- Objetivo:

En el dinámico mundo de la compra y venta de automóviles, entender qué factores determinan el precio de un vehículo puede marcar la diferencia entre una compra inteligente y una inversión fallida. Este proyecto tiene como propósito crear un modelo predictivo que ofrezca una estimación precisa del valor de venta de un automóvil, considerando atributos como la marca, el kilometraje, el estado general, el año de fabricación y otras características clave.



A través del análisis de grandes volúmenes de datos sobre ventas de autos en Estados Unidos, identificaremos patrones y relaciones que ayuden a comprender mejor el mercado automotriz. Nos proponemos responder preguntas clave como:

- ¿Existen algunas marcas de autos que tienden a ser más valiosas en el mercado?
- ¿Qué características influyen más en el precio de venta de un vehículo?
- ¿Las ventas varían dependiendo del estado o la región?
- ¿Qué factores tienden a repetirse con mayor frecuencia en las transacciones de vehículos?
- ¿Cómo afecta el tiempo de uso o la antigüedad de un vehículo al momento de su venta?

Además de predecir el valor de un vehículo, este análisis también profundiza en la comprensión del comportamiento de las ventas, buscando identificar patrones valiosos en las bases de datos de múltiples transacciones. Queremos explorar si ciertos atributos del vehículo son más atractivos para los compradores, si algunas zonas geográficas tienen más transacciones, y si factores como el tipo de transmisión o el año de fabricación inciden significativamente en el precio.

Este enfoque busca ofrecer respuestas claras y accionables para los distintos actores del mercado automotriz, brindando herramientas para optimizar la toma de decisiones y mejorar la comprensión de las dinámicas detrás de los precios de venta de los automóviles.

I.V.- Fuente del dataset:

<https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data>

I.VI.- Metadata:

El dataset seleccionado cuenta con las siguientes características:

- Número de filas: 558.837
- Número de columnas: 16
- Detalle de columnas:
 1. Year: Año de fabricación del vehículo (por ejemplo, 2015)
 2. Make: Marca o fabricante del vehículo (por ejemplo, Kia, BMW, Volvo)
 3. Model: Modelo específico del vehículo (por ejemplo, Sorento, Serie 3, S60, serie 6 Gran Coupé)
 4. Trim: Designación adicional para una versión o paquete de opciones particular del modelo (por ejemplo, LX, 328i SULEV, T5, 650i)
 5. Body: Tipo de carrocería del vehículo (por ejemplo, SUV, Sedán)
 6. Transmission: Tipo de transmisión en el vehículo (por ejemplo, automática)
 7. VIN: Número de Identificación del Vehículo, un código único utilizado para identificar vehículos individuales



8. State: Estado en el que se encuentra o está registrado el vehículo (por ejemplo, CA para California)
 9. Condition: Representación numérica de la condición del vehículo (por ejemplo, 5.0)
 10. Odometer: Millaje o distancia recorrida por el vehículo
 11. Color: Color exterior del vehículo
 12. Interior: Color interior del vehículo
 13. Seller: Entidad o empresa que vende el vehículo (por ejemplo, Kia Motors America Inc, Financial Services Remarketing)
 14. MMR: Manheim Market Report, una herramienta de precios utilizada en la industria automotriz norteamericana. Es el equivalente a Acara, en Argentina. Se usa para tener precios de referencia del mercado de los automóviles.
 15. Selling Price: Precio al que se vendió el vehículo
 16. Sale Date: Fecha y hora en que se vendió el vehículo
- Tipos de variables:
- Categóricas: Make, Model, trim, body, transmission, state, color, interior, seller,
 - Numéricas: Año, condition (flotante), odometro, mmr, sellingprice.
 - Fecha: Saledate

VI.- : Colab

https://colab.research.google.com/drive/1HsHTZM0spVQodb0WpPG7ja7e_eyW_LYn#scrollTo=6AwuleMS1C80