

INSTITUTO SUPERIOR POLITÉCNICO DE CÓRDOBA

**TECNICATURA SUPERIOR EN CIENCIA DE DATOS E
INTELIGENCIA ARTIFICIAL**

Módulo de Práctica Profesionalizante

Análisis Exploratorio de Datos de Flota de Autos

Entrega 2 - Exploración y Análisis de Datos (EDA)

16/09/2024

Docente:

- Charletti, Carlos

Integrantes:

- López, Erick
- Nüesch, Christian
- Zurita Rojo, Debora
- Galeano, Agustín

Índice

Índice.....	2
Resumen Ejecutivo.....	3
Introducción.....	3
Metodología.....	3
Proceso de Análisis.....	3
Datos Utilizados.....	4
Hallazgos Clave.....	4
Precio.....	4
Año.....	4
Kilometraje.....	5
Puertas.....	5
Motor.....	5
Carrocería y Kilometraje:.....	5
Tipo de Combustible:.....	5
Visualizaciones y Tendencias.....	5
Conclusión.....	6

Resumen Ejecutivo

En esta fase del proyecto, se llevó a cabo un **Análisis Exploratorio de Datos (EDA)** sobre un dataset de una flota de automóviles con el objetivo de entender la estructura y características de los datos, identificar patrones iniciales y detectar problemas como valores atípicos o datos faltantes. Los análisis proporcionaron insights preliminares que serán útiles para la toma de decisiones informadas en futuras etapas del proyecto.

Introducción

El **objetivo del EDA** es explorar los datos, generar visualizaciones descriptivas y descubrir patrones que permitan mejorar la comprensión de las variables. Esta fase es esencial para el análisis posterior, ya que establece las bases para modelado predictivo o recomendaciones estratégicas para la flota automotriz.

El dataset incluye información detallada sobre automóviles, como el precio, marca, modelo, año de fabricación, tipo de combustible, kilometraje acumulado, y otras características relevantes.

Metodología

Proceso de Análisis

El análisis se llevó a cabo utilizando **librerías de Python** como **Pandas** y **Seaborn** para la manipulación y visualización de datos. Los pasos realizados fueron:

1. **Carga y limpieza del dataset:** Identificación y tratamiento de valores faltantes o inconsistentes.
2. **Análisis univariado:** Exploración de las variables individuales mediante estadísticas descriptivas.
3. **Visualización:** Histogramas, boxplots y gráficos de barras para representar distribuciones y relaciones.
4. **Análisis bivariado:** Exploración de relaciones entre pares de variables, utilizando gráficos de dispersión y mapas de calor de correlaciones.

Datos Utilizados

El dataset con el que se trabajó en este análisis contiene múltiples características relacionadas con automóviles de una flota. A continuación, se describen las columnas del dataset:

- **Marca:** marca del auto. **Tipo de dato:** Cadena de caracteres.
- **Modelo:** modelo específico del auto. **Tipo de dato:** Cadena de caracteres.
- **Año:** año de fabricación del auto. **Tipo de dato:** Entero.
- **Color:** color del exterior del auto. **Tipo de dato:** Cadena de caracteres.
- **Combustible:** tipo de combustible que utiliza el vehículo. **Tipo de dato:** Cadena de caracteres.
- **Puertas:** cantidad de puertas que tiene el vehículo. **Tipo de dato:** Entero.
- **Caja:** tipo de caja de cambios del vehículo. **Tipo de dato:** Cadena de caracteres.
- **Motor:** tamaño del motor del auto, expresado en litros. **Tipo de dato:** Flotante con un decimal.
- **Carrocería:** tipo de carrocería del vehículo. **Tipo de dato:** Cadena de caracteres.
- **Kilómetros:** kilometraje acumulado por el auto. **Tipo de dato:** Entero.
- **Moneda:** moneda en la que se cotiza el precio del auto. **Tipo de dato:** Cadena de caracteres.
- **Precio:** precio del auto, expresado en la moneda especificada en la columna "Moneda". **Tipo de dato:** Entero.

Hallazgos Clave

Año

- La mayoría de los autos tienen menos de 10 años. El promedio de fabricación es 2016, lo que sugiere que el **parque automotor es relativamente moderno**.

Kilometraje

- El kilometraje promedio es de 74.732 km. La mayoría de los autos **han recorrido distancias moderadas**, con un 75% por debajo de 99.100 km, indicando que en general los vehículos están en buen estado.

Puertas

- La mayoría de los autos tiene entre **4 y 5 puertas**, con un promedio de **4.47 puertas**. Esto sugiere que el parque está compuesto principalmente por **vehículos familiares o utilitarios**, que suelen tener más puertas para comodidad de los pasajeros o para uso comercial.

Motor

- El **motor promedio** es de **1.88** litros, lo que indica que el parque está compuesto principalmente por **autos de cilindrada moderada**, comúnmente usados para tareas estándar.
- Hay autos con motores pequeños (**1.0 litros**), y el valor máximo de **6.4 litros** probablemente corresponda a un auto de alta gama o de características muy específicas.

Carrocería y Kilometraje:

- Las **minivans presentan los kilometrajes más altos**, lo cual es coherente con su uso en trayectos largos. Los Coupés, por otro lado, tienen el kilometraje más bajo, posiblemente debido a su naturaleza recreativa.

Tipo de Combustible:

- Los vehículos con **nafta y diésel son los más comunes**, mientras que los híbridos son modelos más recientes en el mercado, con menores rangos de antigüedad.

Precio

- De momento tiene poco sentido analizar esta variable, porque hay autos nominados en pesos, y autos nominados en dólares.

Visualizaciones y Tendencias

Se utilizaron gráficos de barras para identificar tendencias en la **distribución por marca y año de fabricación**, mostrando que la mayor parte del parque se comercializó entre 2012 y 2020.

Un **gráfico de caja (boxplot)** reveló la gran variabilidad en kilometraje por marcas como **BMW, Renault y Volkswagen**, mientras que marcas como **DS y Suzuki** mostraron kilometrajes más uniformes y bajos.

En la **distribución de cajas** (manual vs automática), el 58.8% de los autos en el mercado tienen caja manual, lo que sugiere una preferencia por vehículos más económicos en cuanto a mantenimiento.

Conclusión

El análisis exploratorio de datos permitió identificar patrones relevantes en el parque automotor, destacando que los vehículos con mayor kilometraje, como las minivans, requieren una **atención especial en mantenimiento preventivo**. Esto es fundamental para reducir costos operativos a largo plazo y mejorar la eficiencia del uso de estos autos.

Además, se evidenció que ciertos tipos de vehículos, según su **año de fabricación y tipo de uso**, presentan una mayor durabilidad, lo que puede ayudar a los clientes a **optimizar sus decisiones de compra y mantenimiento**. La segmentación por **tipo de combustible y caja de cambios** también proporciona una herramienta valiosa para elegir autos más eficientes y adecuados a las necesidades específicas de los usuarios.

Finalmente, el análisis sugiere que las **estrategias de renovación del parque** deben tener en cuenta las características específicas de los vehículos, como el kilometraje y el tipo de carrocería, para maximizar el retorno de inversión y garantizar un uso más prolongado y efectivo de los mismos.