

# Análisis de Datos de Autos Usados.

## Fuente de los datos:

### Descripción de las columnas:

- Breve explicación de cada columna:
  - **Car\_Name:** Nombre del modelo del auto.
  - **Year:** Año de fabricación.
  - **Selling\_Price:** Precio de venta en lakhs (1 lakh = 100,000).
  - **Kms\_Driven:** Kilómetros recorridos.
  - **Fuel\_Type, Transmission, Owner:** Tipo de combustible, transmisión y número de propietarios.

## explicación

En este proyecto, trabajé con un dataset sobre autos usados, originalmente en formato Excel, y seguí un flujo estructurado para analizar los datos y generar insights visuales:

### 1. Importación de Datos en SQL Server:

- Comencé importando la tabla desde Excel a una base de datos en SQL Server.
- Durante este paso, me aseguré de que los datos se cargaran correctamente, revisando la estructura y validando que no hubiera errores o valores inconsistentes.

### 2. Conexión de Python con SQL Server:

- Establecí una conexión entre Python y SQL Server utilizando la biblioteca sqlalchemy.
- Esto permitió extraer los datos directamente desde SQL para su análisis en Python.

### 3. Análisis Exploratorio y Limpieza de Datos:

- Revisé la estructura del dataset (info y describe) para comprender las columnas y su contenido.

- Validé valores únicos en columnas clave como Fuel\_Type, Seller\_Type, y Transmission para confirmar la consistencia de los datos.

#### **4. Generación de Gráficos y Análisis Visuales:**

- Utilicé matplotlib y seaborn para crear gráficos que facilitaron el análisis visual de las principales tendencias en los datos:
  - Distribuciones: Precios de venta y kilómetros recorridos.
  - Relaciones: Impacto del año del modelo, tipo de combustible, y número de propietarios en el precio de venta.
  - Composiciones: Proporción de autos recientes vs. antiguos, y distribución por tipo de transmisión.

#### **5. Resultados e Insights:**

- Los gráficos generados me permitieron identificar patrones clave, como:
  - Rango de precios predominantes.
  - Relación entre kilometraje, antigüedad y precio.
  - Preferencias del mercado en cuanto a tipo de combustible y transmisión.

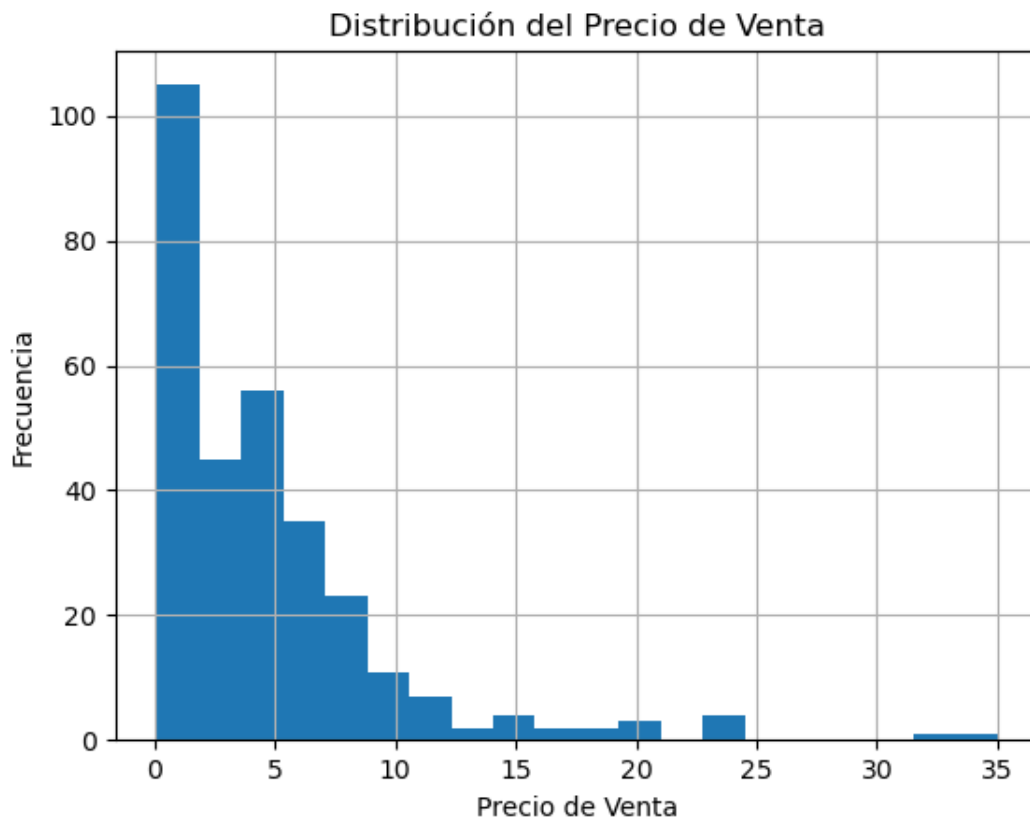
## **Herramientas y Tecnologías Utilizadas**

- Python (pandas, matplotlib, seaborn).
- SQL Server.
- Matplotlib para visualizaciones.

## **Análisis**

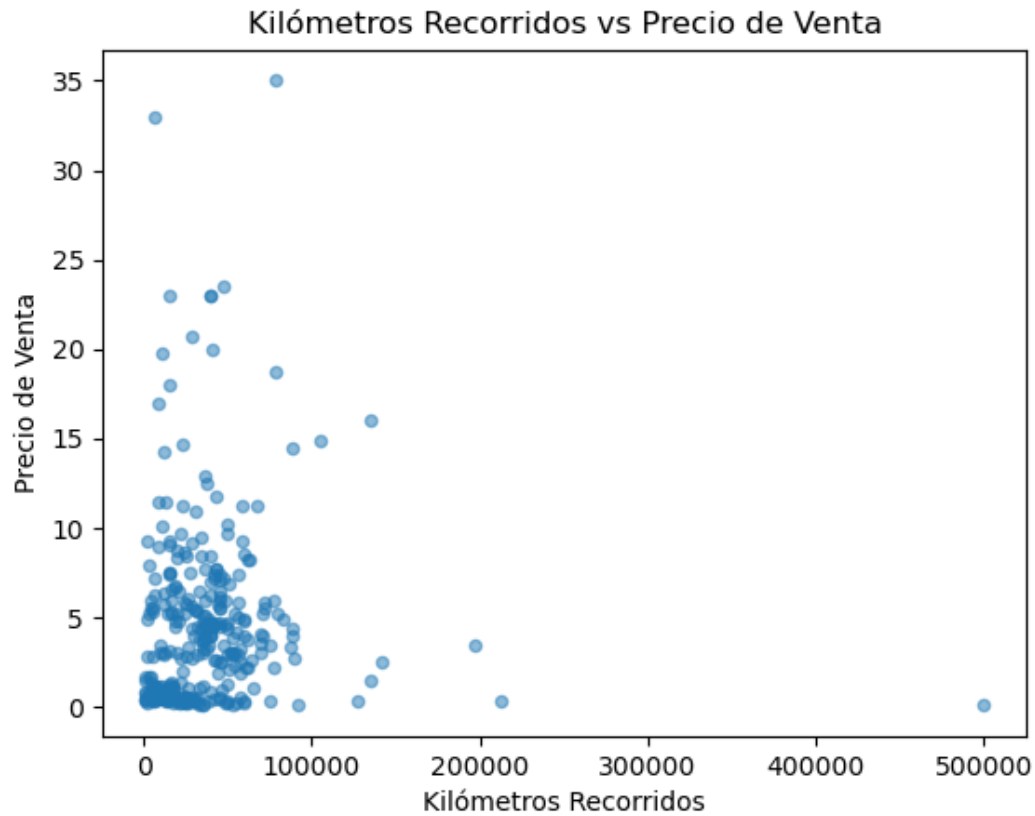
### **Distribución del Precio de Venta:**

El gráfico muestra cómo se distribuyen los precios de venta de los autos usados. Permite identificar los rangos de precios más comunes y analizar si hay una concentración de vehículos en precios bajos, medios o altos. Este análisis es útil para comprender las tendencias generales del mercado y determinar en qué segmentos de precios se encuentra la mayoría del inventario.



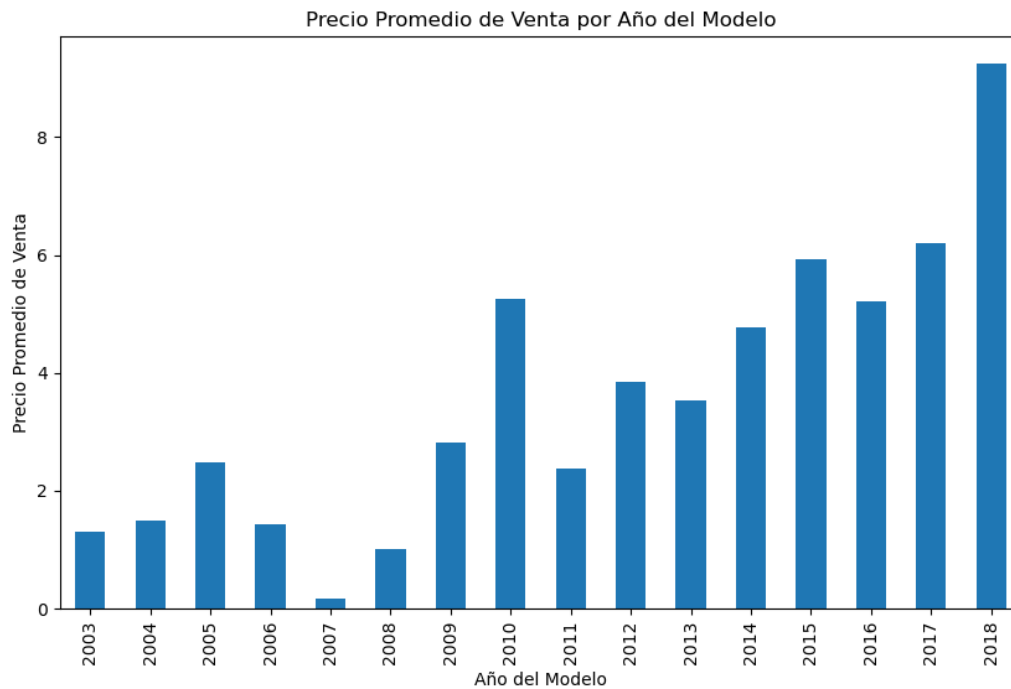
**Relación entre Kilómetros Recorridos y Precio de Venta:**

El gráfico de dispersión muestra la relación entre los kilómetros recorridos por los autos y su precio de venta. Cada punto representa un vehículo, donde el eje X indica los kilómetros recorridos y el eje Y el precio de venta. Este análisis ayuda a identificar patrones, como si los vehículos con más kilometraje tienden a tener precios más bajos, y permite evaluar cómo afecta el uso del vehículo a su valor en el mercado.



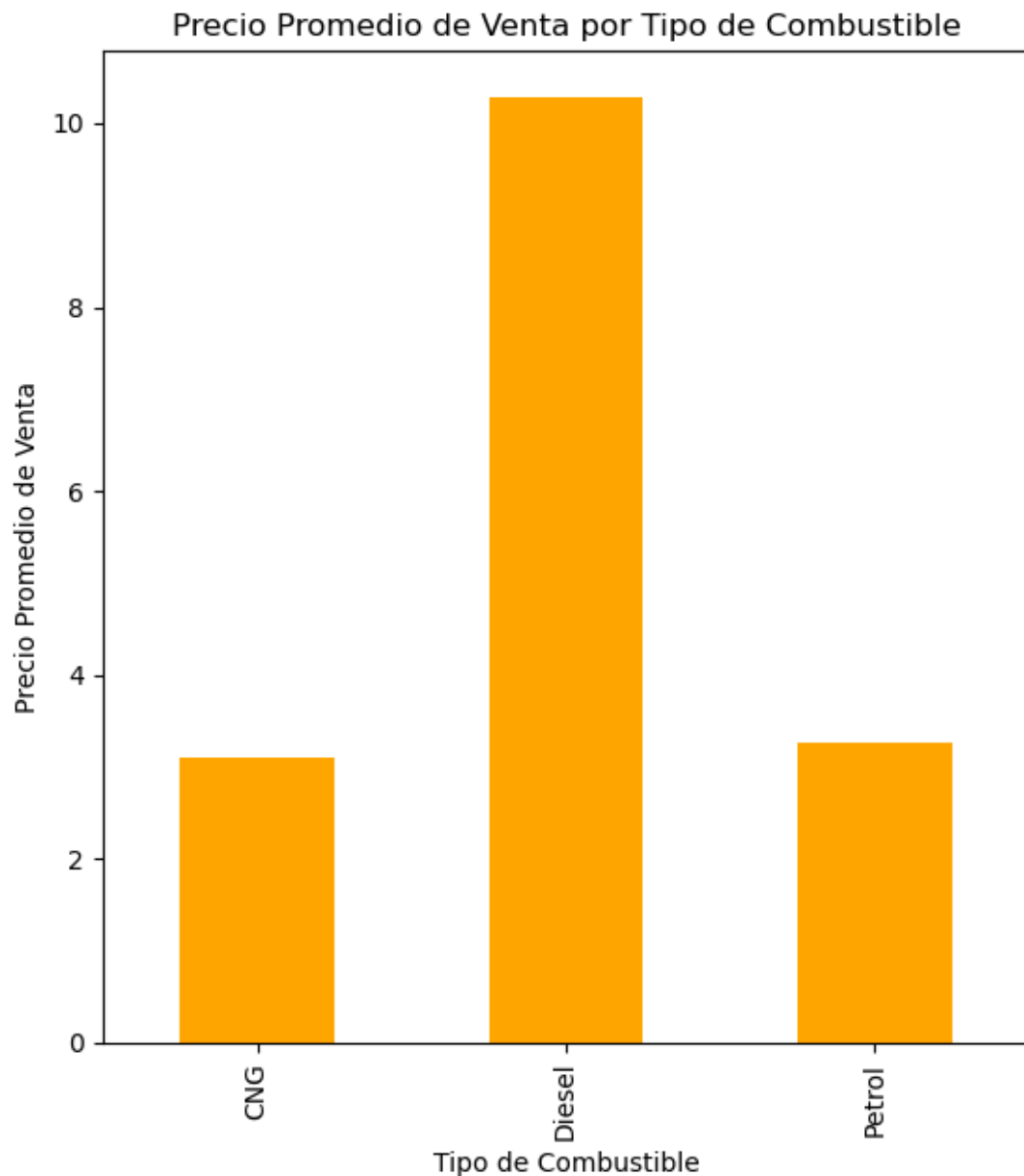
#### **Precio Promedio de Venta por Año del Modelo:**

El gráfico de barras muestra el precio promedio de venta de los autos según su año de modelo. En el eje X se representan los años del modelo, mientras que en el eje Y se indican los precios promedios correspondientes. Este análisis permite identificar tendencias, como si los modelos más recientes tienen precios significativamente más altos que los modelos más antiguos, proporcionando información clave para la valoración de vehículos según su antigüedad.



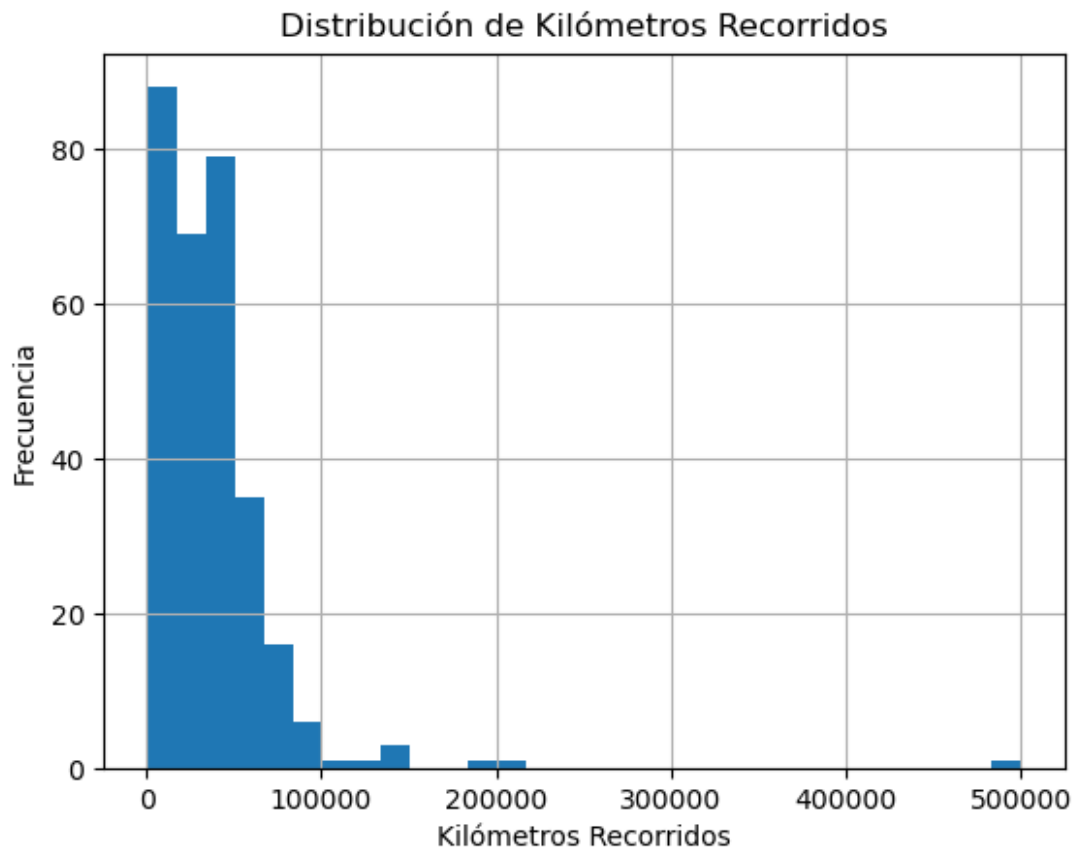
### **Precio Promedio de Venta por Tipo de Combustible:**

El gráfico de barras muestra el precio promedio de venta de los autos según su tipo de combustible. En el eje X se representan los tipos de combustible (por ejemplo, gasolina, diésel, etc.), y en el eje Y se indican los precios promedio correspondientes. Este análisis permite identificar si ciertos tipos de combustible están asociados con precios más altos, lo que puede reflejar preferencias del mercado o diferencias en el valor percibido de los vehículos según su eficiencia o costos operativos.



### Distribución de Kilómetros Recorridos:

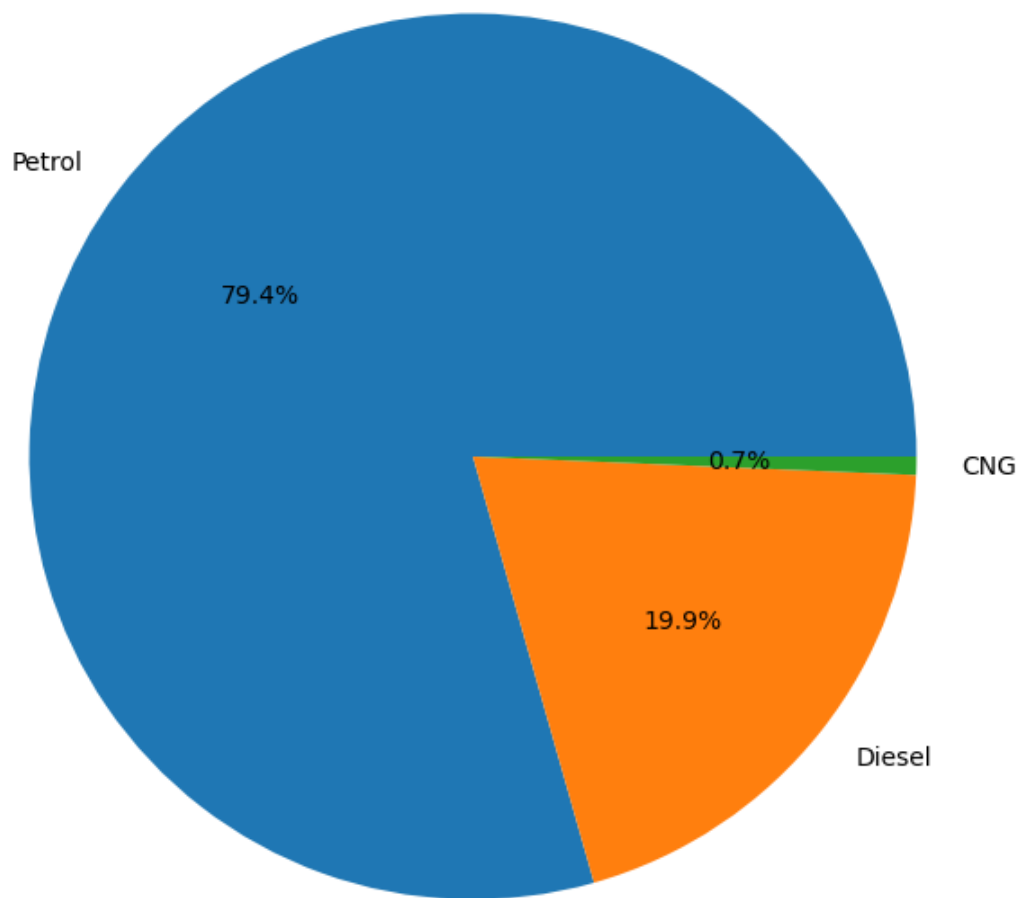
El histograma muestra cómo se distribuyen los kilómetros recorridos por los autos en el inventario. En el eje X se representan los rangos de kilometraje, mientras que en el eje Y se muestra la frecuencia de autos en cada rango. Este análisis permite identificar si la mayoría de los vehículos son de bajo, medio o alto kilometraje, lo que puede ser útil para evaluar el estado general del inventario y las expectativas de mantenimiento o uso futuro.



### **Distribución por Tipo de Combustible:**

El gráfico de pastel muestra la proporción de autos disponibles según su tipo de combustible. Cada segmento representa un tipo de combustible (por ejemplo, gasolina, diésel, etc.), con el porcentaje correspondiente indicado dentro del gráfico. Este análisis permite identificar qué tipo de combustible es predominante en el inventario, proporcionando información útil para entender las preferencias del mercado o para planificar estrategias de oferta y demanda.

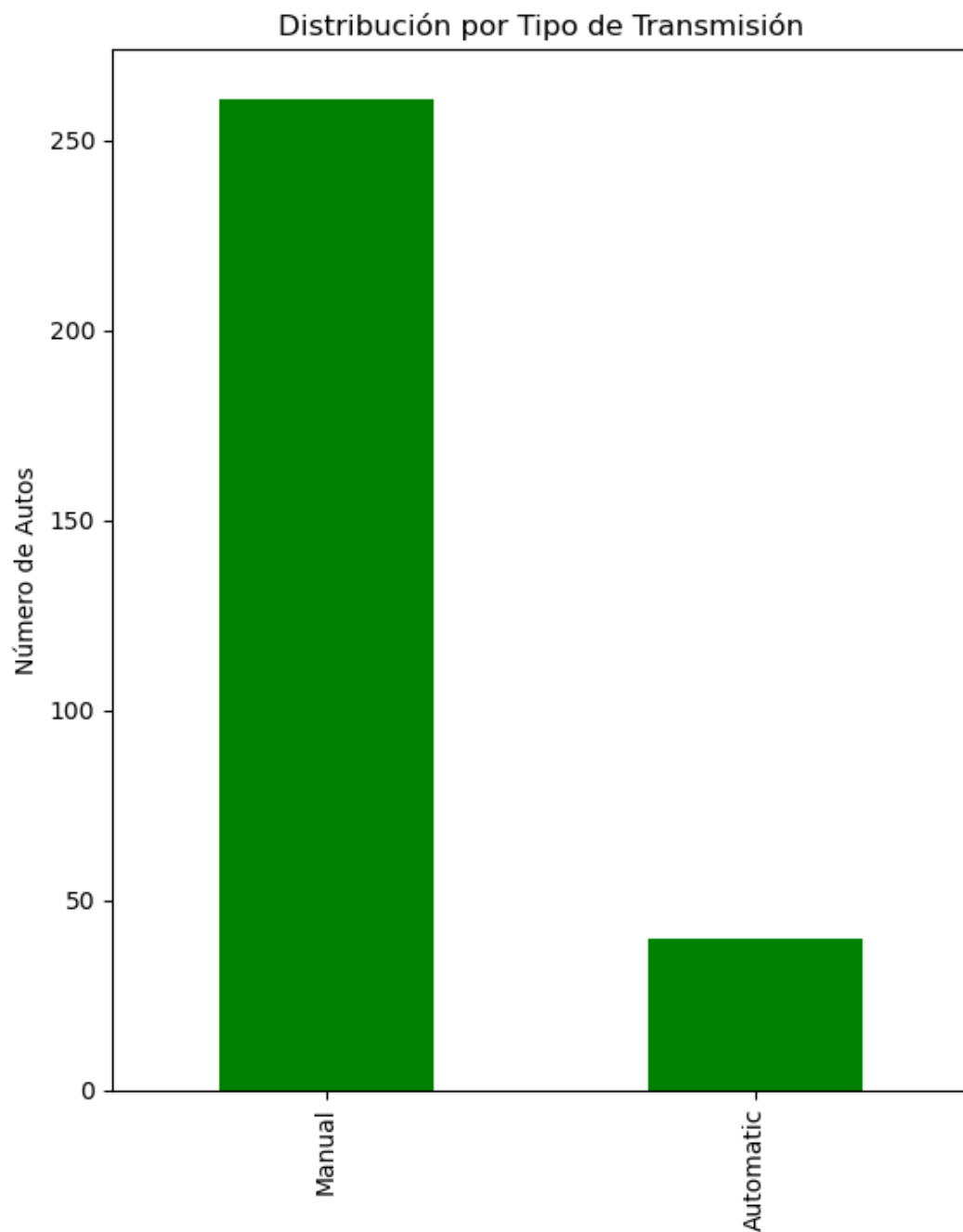
Distribución por Tipo de Combustible



**Distribución por Tipo de Transmisión:**

El gráfico de barras muestra la cantidad de autos según su tipo de transmisión (manual o automática). En el eje X se representan los tipos de transmisión, y en el eje Y se indica el número de vehículos en cada categoría. Este análisis ayuda a identificar cuál es el tipo de transmisión más común en el inventario, lo que puede ser útil para entender las preferencias del mercado y orientar decisiones de inventario o ventas.





### Matriz de Correlación:

La matriz de correlación muestra las relaciones entre las variables numéricas del dataset. Cada celda representa el coeficiente de correlación entre dos variables, con valores que oscilan entre -1 y 1:

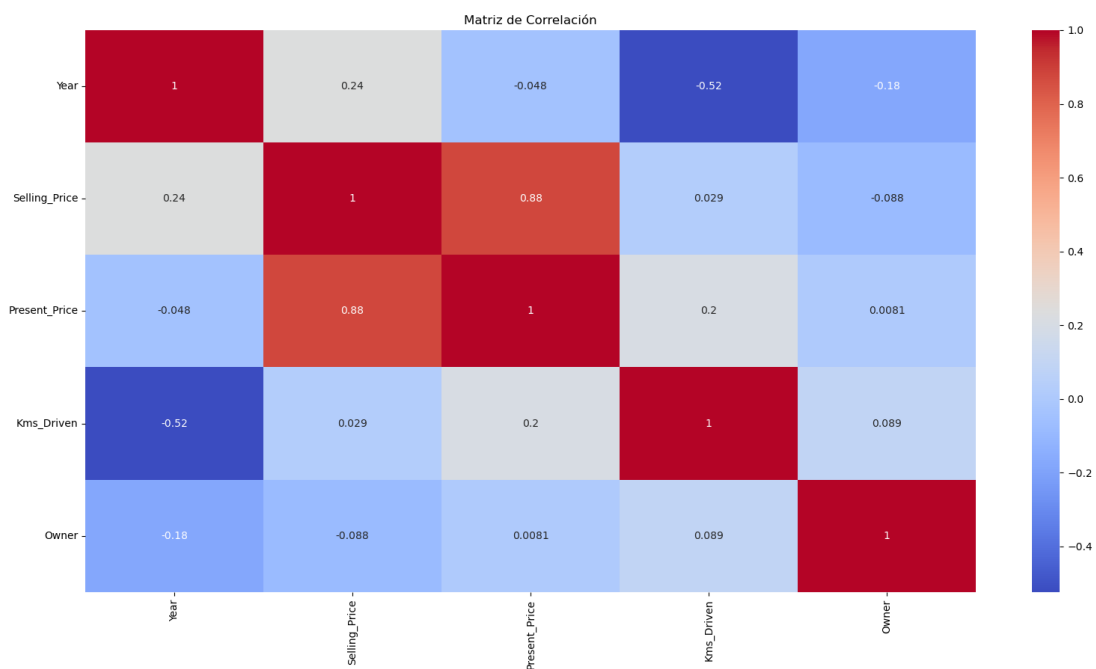
- **Cercano a 1:** Correlación positiva fuerte (cuando una variable aumenta, la otra también lo hace).
- **Cercano a -1:** Correlación negativa fuerte (cuando una variable aumenta, la otra disminuye).

- **Cercano a 0:** Sin relación significativa.

El gráfico utiliza un mapa de calor donde los colores indican la intensidad y dirección de la correlación:

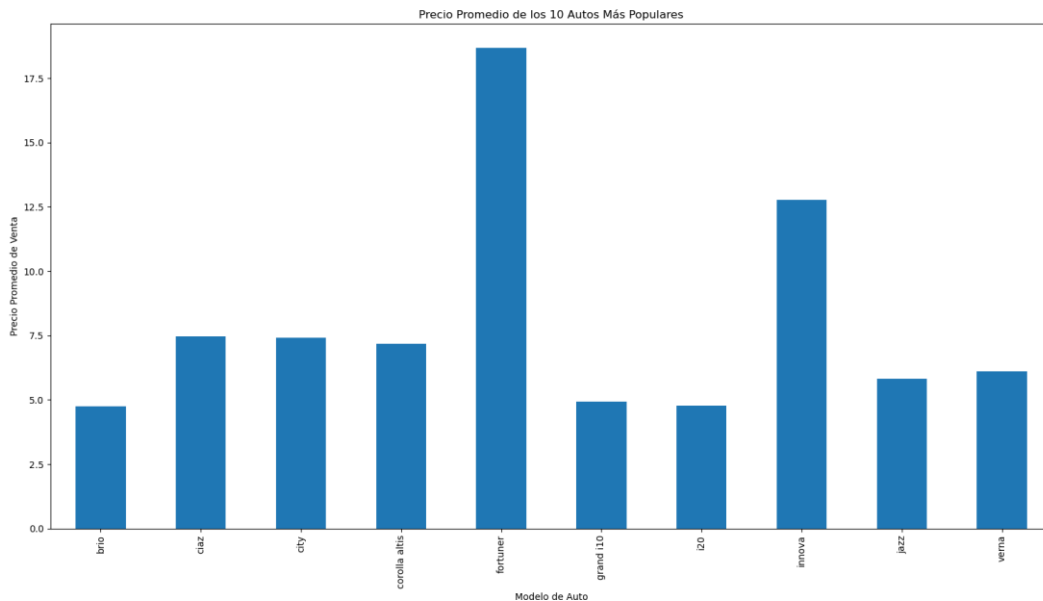
- **Colores cálidos (rojo):** Correlaciones positivas altas.
- **Colores fríos (azul):** Correlaciones negativas altas.

Este análisis es útil para identificar relaciones entre variables, como el impacto del precio actual o el kilometraje en el precio de venta.



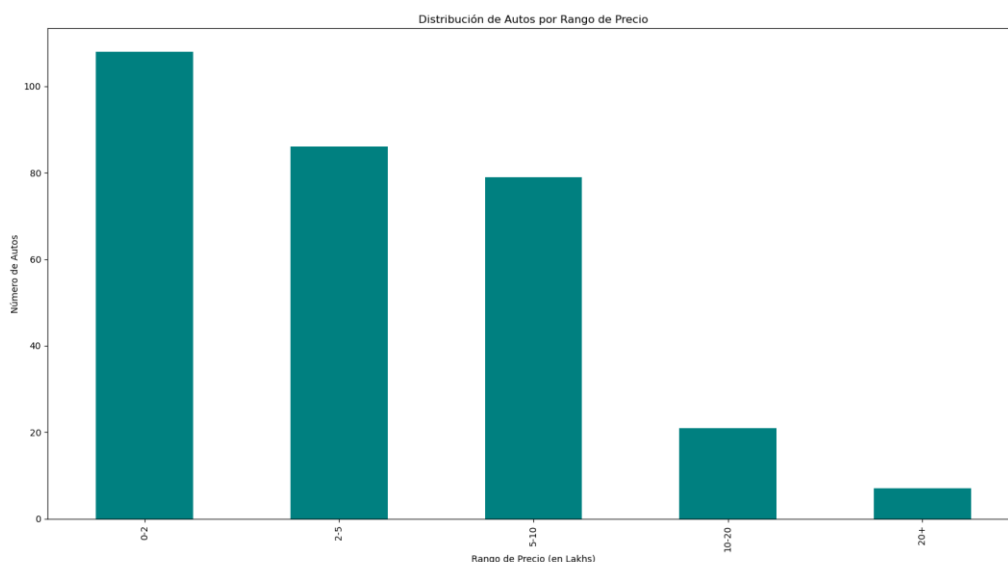
### Precio Promedio de los 10 Autos Más Populares:

El gráfico de barras muestra los 10 modelos de autos más comunes en el inventario y su precio promedio de venta. En el eje X se encuentran los nombres de los modelos de autos, y en el eje Y el precio promedio correspondiente. Este análisis permite identificar cuáles son los autos más populares y su valor en el mercado, proporcionando información clave para priorizar estrategias de venta o marketing enfocadas en estos modelos.



### Distribución de Autos por Rango de Precio:

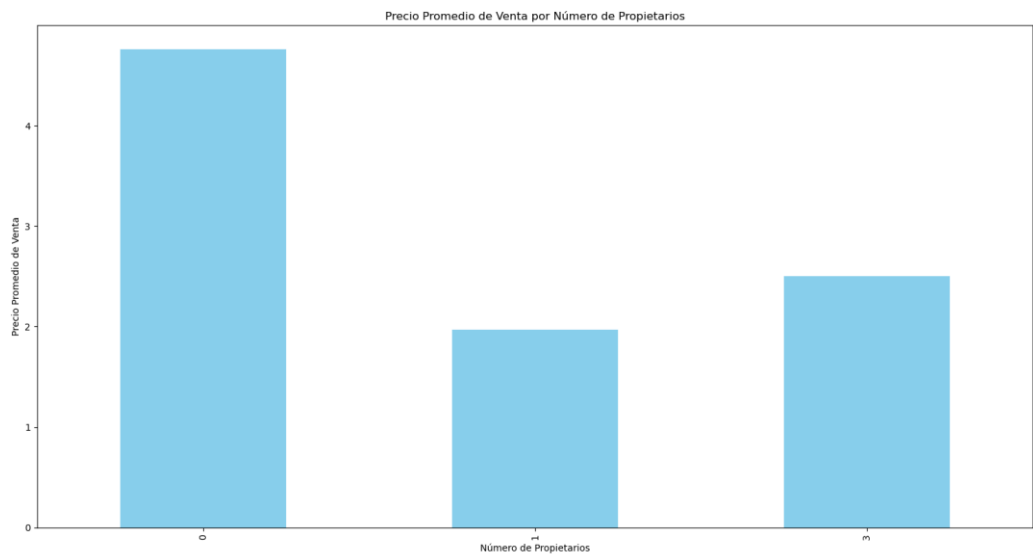
El gráfico de barras muestra la cantidad de autos en diferentes rangos de precios, definidos en Lakhs (1 Lakh = 100,000). En el eje X se representan los rangos de precios (por ejemplo, 0-2 Lakhs, 2-5 Lakhs, etc.), mientras que en el eje Y se muestra el número de autos en cada rango. Este análisis es útil para identificar en qué segmentos de precios se concentra la mayor parte del inventario, ayudando a planificar estrategias de precios y marketing para los diferentes niveles de clientes.



### Precio Promedio de Venta por Número de Propietarios:

El gráfico de barras muestra el precio promedio de venta de los autos según la cantidad de propietarios previos. En el eje X se encuentran las categorías de

propietarios (por ejemplo, 0, 1, 2, etc.), mientras que en el eje Y se representan los precios promedio correspondientes. Este análisis permite identificar cómo el historial de propietarios afecta el valor de los vehículos, destacando si los autos con menos propietarios tienen precios más altos debido a una mejor percepción de su estado o menor desgaste.

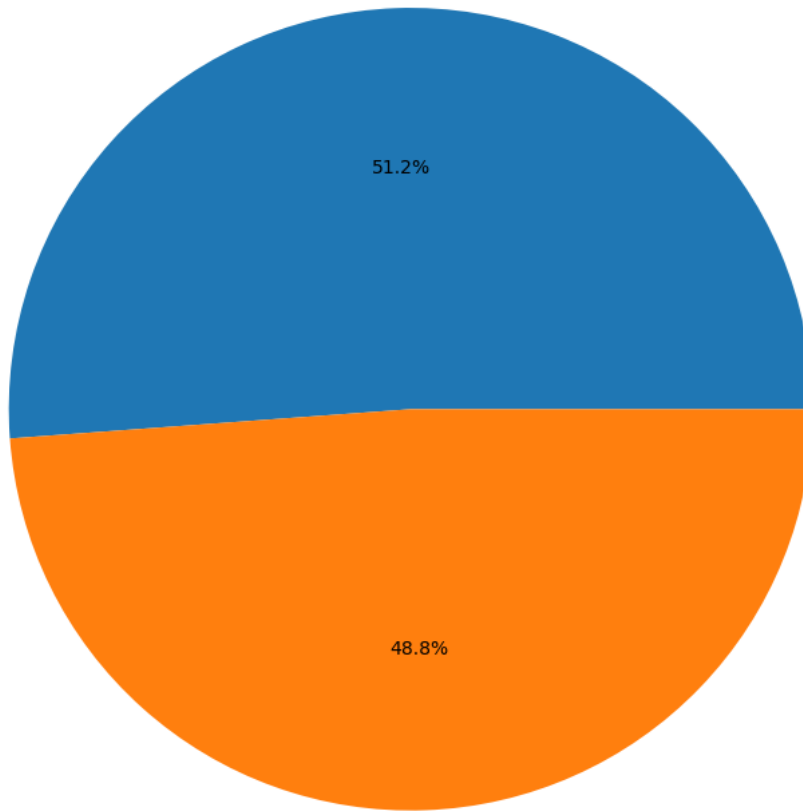


**Distribución del Inventario: Autos Recientes vs Antiguos:**

El gráfico de pastel muestra la proporción de autos en el inventario clasificados como recientes (modelos de 2015 o más nuevos) y antiguos (modelos anteriores a 2015). Cada segmento representa una categoría de antigüedad, y los porcentajes indican la participación relativa de cada grupo en el total del inventario. Este análisis es útil para evaluar si el inventario está mayormente compuesto por autos más modernos o por modelos más antiguos, lo que puede influir en estrategias de marketing y precios.

## Distribución del Inventario: Autos Recientes vs Antiguos

Antiguos (<2015)



Recientes (2015+)