**README PROYECTO DATAPROJECT: DASHBOARD & ANÁLISIS DE DATOS**

**📋 Estructura:**

**1. Título del proyecto**

🚗 Análisis de precios de coches usados en España (2000-2024).

Explorando tendencias y patrones en el mercado automovilístico.

**2.** 📖 **Descripción del proyecto**  
El objetivo de este proyecto es realizar un análisis exploratorio de las variables que influyen en el precio de los coches usados en España entre los años 2000 y 2024. A través de esta base de datos, se busca identificar patrones, tendencias y relaciones entre las características de los vehículos y su precio de mercado.

La base de datos incluye información detallada sobre diversas características de los coches, como la marca, modelo, especificaciones técnicas, tipo de combustible, dimensiones físicas, y métricas de rendimiento. Esto permite explorar cómo estas variables afectan al precio y qué factores tienen mayor impacto en el mercado de coches usados.

**Variables Incluidas**

* **Brand:** Marca del coche.
* **Name:** Nombre del modelo exacto.
* **Sticker:** Pegatina ambiental (C, B, Eco, etc.).
* **Year:** Año de compra del coche.
* **KM:** Kilómetros recorridos.
* **Fuel:** Tipo de combustible utilizado.
* **CV:** Caballos de potencia.
* **Transmission:** Tipo de transmisión (manual o automática).
* **One\_owner:** Indicador de si el coche ha tenido un solo dueño (True o False).
* **Location:** Ciudad donde se vende el coche.
* **Length:** Largo del coche (en metros).
* **Width:** Ancho del coche (en metros).
* **Height:** Alto del coche (en metros).
* **Weight:** Peso del coche (en kilogramos).
* **Trunk:** Capacidad del maletero (en litros).
* **Vmax:** Velocidad máxima (en km/h).
* **Oto100:** Tiempo de aceleración de 0 a 100 km/h (en segundos).
* **Consumption:** Consumo de combustible por cada 100 km (o autonomía para vehículos eléctricos).
* **Emissions:** Emisiones de CO2.
* **Keys\_num:** Número de llaves disponibles.
* **Price:** Precio del coche.

**Enfoque del Análisis**

Este análisis utilizará técnicas de exploración de datos (EDA) para identificar patrones y relaciones clave. Se emplearán visualizaciones, estadísticas descriptivas y correlaciones para entender cómo cada variable impacta en el precio. Los resultados podrán ser utilizados como base para futuros modelos predictivos del precio de coches usados.

**3.** 🗂 **Estructura del proyecto**

├README.md #Documento inicial del repositorio

├Archivo Google Sheets #Archivo de trabajo

├datos\_conjuntos/ # Datos crudos y procesados

├análisis\_variables\_categóricas/ # Tablas dinámicas y gráficos variables categóricas

├análisis\_variables\_numéricas/ # Tablas dinámicas y gráficos variables numéricas

├verificación\_suposiciones/ # Variables que afectan al precio de los coches usados

├dashboard/ #Dashboard con gráficos

├CSV #Base de datos

**4.** 🛠 **Instalación y Requisitos**

Google Sheets.

**5.** 📊 **Resultados y Conclusiones**

Pasos del proyecto:

* Transformación y limpieza de los datos.
* Análisis descriptivo de los datos.
* Dashboard.
* Informe explicativo del análisis.

● **Transformación y limpieza de los datos.**

Acciones que realizar:

* Tratamiento de duplicados
* Tratamiento de nulos
* Transformación de datos

**Duplicados:**

Elimino las dos últimas columnas vacías.

Formato – formato condicional

O fórmula filter.

Pero lo mejor es: Datos -borrado de datos – quitar duplicados – seleccionamos solo la columna en la que se encuentren los duplicados que queremos (columna A).

No hay datos duplicados.

**Nulos:**

Usar filtros y poner desconocido o la mediana. Uso la mediana para evitar que los valores extremos afecten el análisis. Al poner el valor central representamos mejor los datos sesgados. En las filas donde no hay valores pongo la mediana como valor escogido para analizar.

Transformación de datos: editar - buscar y reemplazar para cambiar los los datos decimales con . para escribirlos con coma.

Eliminación de la columna tank ya que no hay datos en ninguna fila sobre esa variable.

Creación de una nueva columna llamada intervalo de precios para separar los precios por intervalos y así facilitar y mejorar el análisis. Intervalos de precios:

* 0-1.000€
* 10.000-19.999€
* 20.000-29.999€
* 30.000-39.999€
* 40.000-49.999€
* 50.000-62.000€

Además, se ha creado un intervalo de valores para facilitar también el análisis de las variables: KM, CV, Consumo y Emisión.

Analizando las columnas de las 4 nuevas variables (intervalo de KM, CV, Consumo y Emisión) he observado que hay 132 celdas vacías pero al buscarlas en los datos no las encuentro, por ello, en todas las tablas dinámicas/gráficos he ocultado estos datos.

También, he realizado varios cambios: las columnas con números adaptadas al formato número, las de texto a texto y la de precio a moneda.

**● Análisis descriptivo de los datos.**

**Resumen de las columnas**

* **Brand, Name, Sticker, Fuel, Transmission, Location**: Variables categóricas que describen la marca, modelo, características ambientales y ubicación del coche.
* **Year, KM, CV, Length, Width, Height, Weight, Trunk, Vmax, 0to100, Consumption, Emissions**: Variables numéricas relacionadas con las especificaciones técnicas del coche.
* **Price**: Precio del coche (variable objetivo del análisis).

**Patrones e identificación de anomalías:**

En este análisis vamos a tratar de identificar que columnas y que valores de esas columnas y que valores de esas columnas afectan directamente a nuestro objetivo del análisis: el precio de los coches usados en España.

Queremos conocer que variables afectan más al precio de un coche usado.

Lo haremos a través de las siguientes herramientas:

* Tablas dinámicas
* Gráficos
* Formato condicional

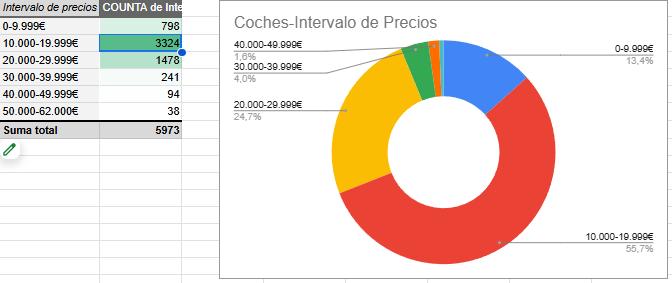
**Variables que pueden afectar más al precio de los coches usados:**

* **Brand:** Marcas premium suelen tener precios más altos.
* **Name:** Depende del modelo de coche será más caro o más barato.
* **Fuel:** Los coches eléctricos o híbridos pueden tener precios diferentes comparados con gasolina o diésel.
* **Transmission:** Los automáticos suelen ser más caros en algunos mercados.
* **Location**: depende de la ciudad donde se encuentre, tendrá un precio u otro.
* **Year:** Los coches más nuevos tienden a ser más caros.
* **KM (kilometraje):** Un mayor kilometraje suele reducir el precio.
* **CV (potencia):** Coches con más potencia generalmente tienen precios más altos.
* **Consumption:** El consumo eficiente puede aumentar el precio en mercados donde esta característica sea valorada.
* **Emissions:** Las bajas emisiones pueden aumentar el precio en mercados donde esta característica sea valorada.

Se han creado controles de filtros para visualizar los datos que nos interesen y macros para cambiar rápidamente la vista de los datos de porcentajes a valores absolutos.

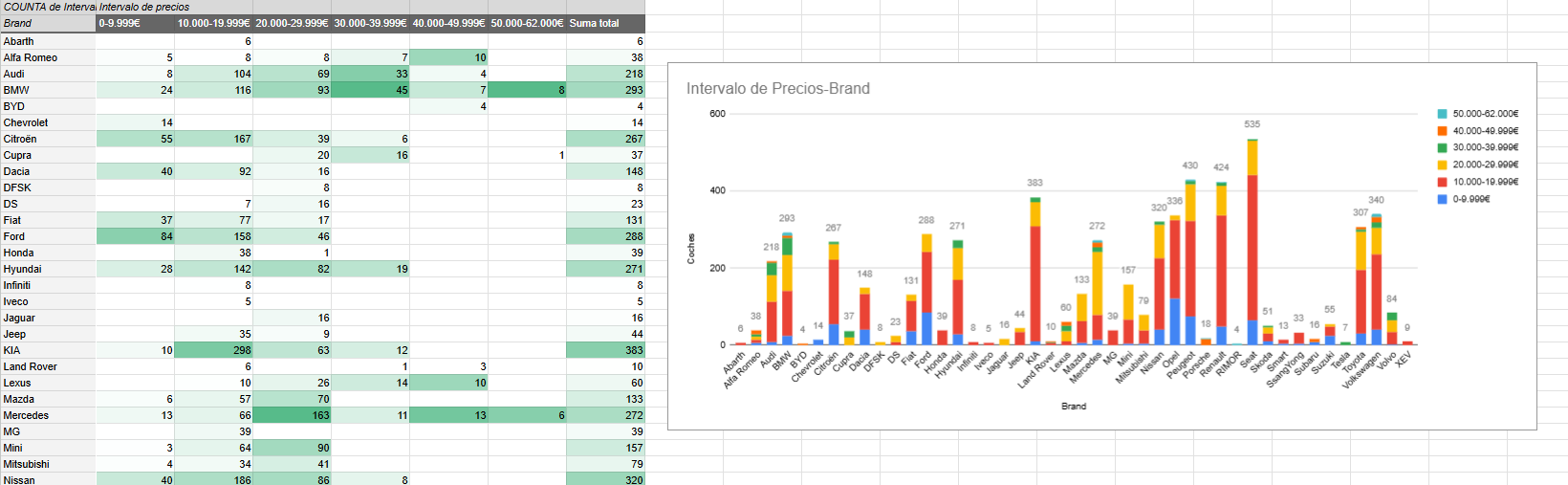
**Análisis tablas y gráficos (hoja: verificación\_suposiciones):**

**1. Coches-Precio**



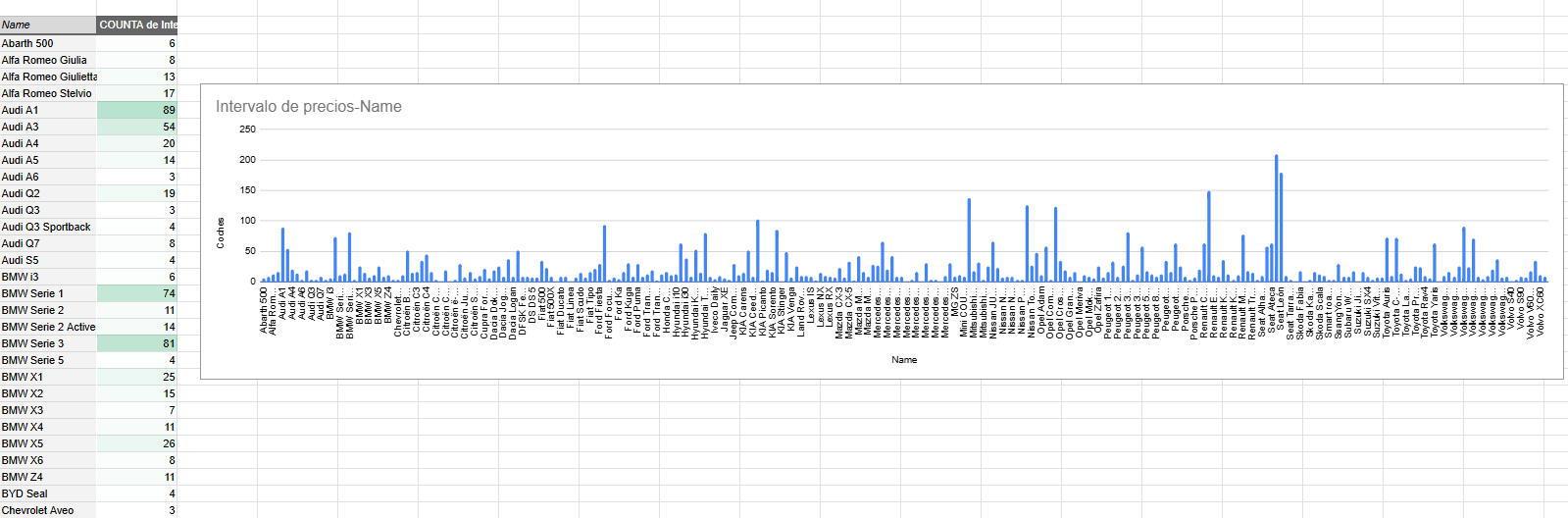
Podemos observar tanto en la tabla como en el gráfico que de la totalidad de coches (**5.973**), en el intervalo de precio de **10.000-19.999€** es donde tenemos la mayor muestra de coches (**3.324**).

**2. Precio-Brand**



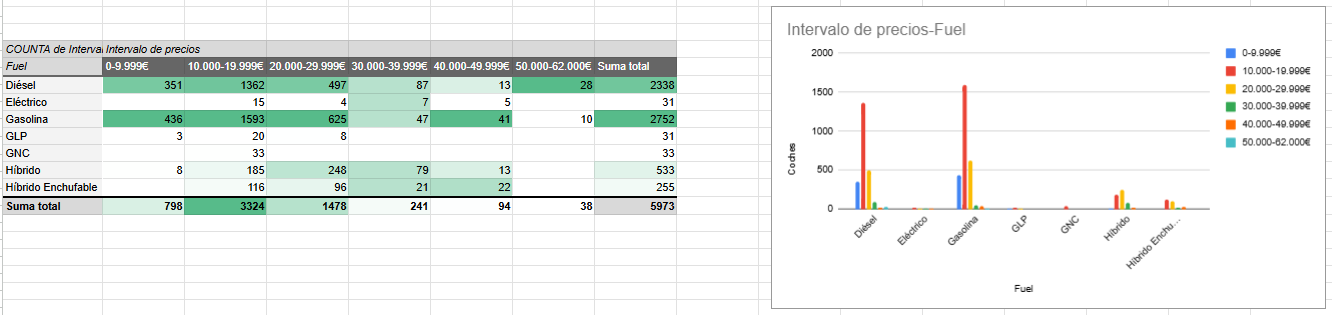
El gráfico muestra que Seat, Renault y Peugeot son las marcas más representadas en el mercado de coches usados en España, destacando en el rango de precios **10.000-19.999€**, que domina ampliamente el mercado. Mercedes domina el mercado del precio de **20.000-29.999€**. Las marcas premium como Audi y BMW concentran la mayoría de sus vehículos en intervalos superiores a **30.000€**, mientras que las marcas generalistas ofrecen una mayor diversidad en todos los rangos de precios. Esto sugiere que el mercado está liderado por vehículos asequibles y accesibles, mientras que los coches de lujo representan un nicho más pequeño (**40.000-62.000€**), aunque mantienen precios elevados.

3. **Coches-Nombre del coche**



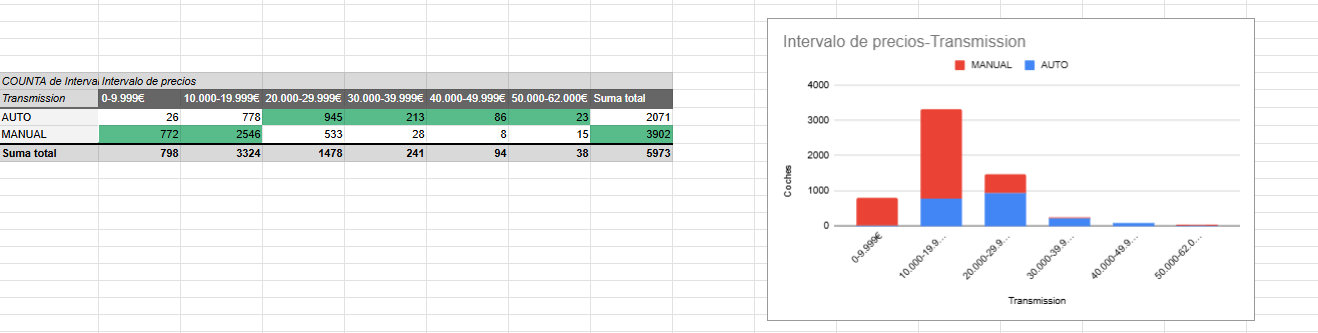
**Seat Ibiza** y **Seat León** son los dos modelos de los que hay más coches en el mercado de coches usados, reafirmando que el mercado está liderado por vehículos asequibles y accesibles. Depende del modelo, tendrá un precio u otro.

**4. Precio-fuel**



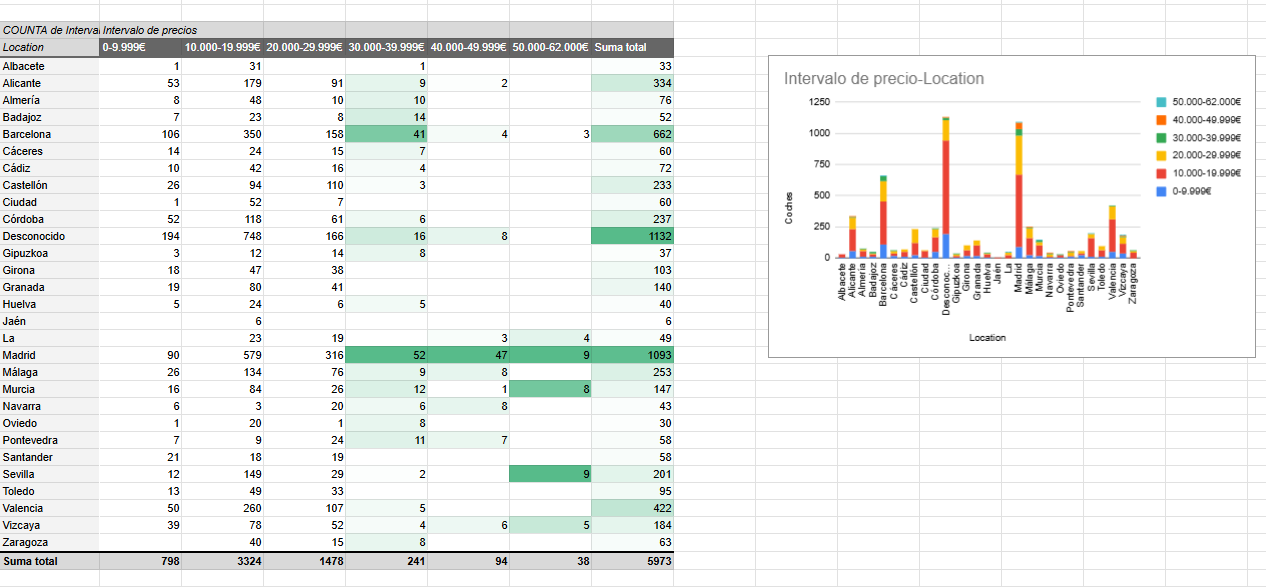
El **tipo de combustible influye significativamente en el precio de los coches usados**. Los coches de gasolina y diésel dominan en los rangos más asequibles (**10.000-19.999€**), con los de gasolina siendo los más representados en general. Los vehículos híbridos y eléctricos, al ser tecnologías más modernas y sostenibles, tienen una mayor proporción en los rangos superiores (**20.000€** o más), reflejando su valor tecnológico y menor depreciación. Los vehículos con combustibles menos comunes como GNC y GLP tienen poca representación, indicando que son opciones menos demandadas y posiblemente más económicas.

**5. Precio-Transmission**

****

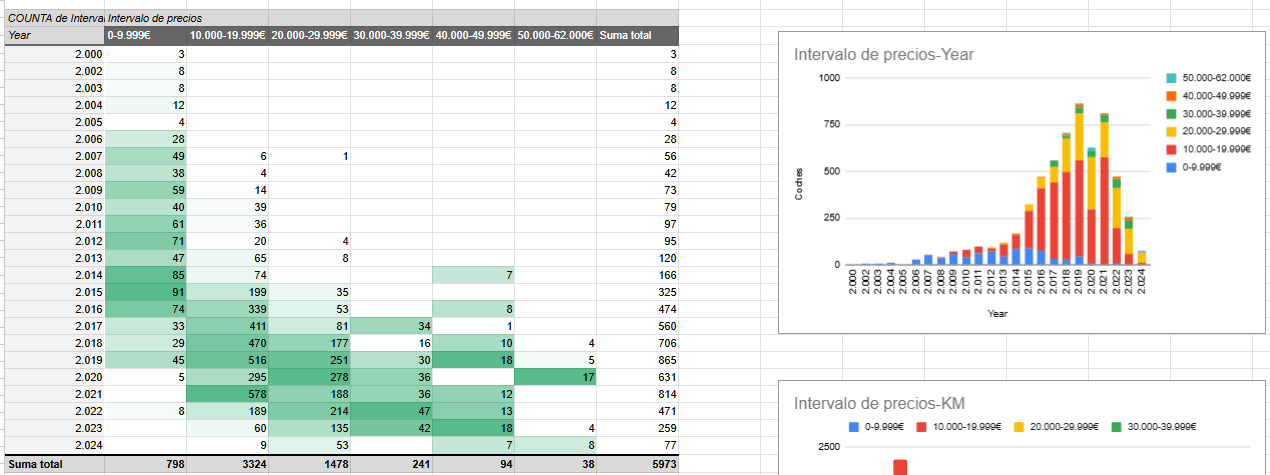
La transmisión influye notablemente en el precio de los coches usados. Los vehículos con transmisión manual predominan en los rangos de precios más bajos, especialmente en el rango **10.000-19.999€**, lo que sugiere que suelen ser más económicos y comunes en el mercado. Por otro lado, los coches con transmisión automática están más representados en los rangos superiores (**20.000-29.999€** y más), reflejando que suelen ser modelos más modernos, con mayor tecnología y, por ende, de mayor valor. Esto indica que **la transmisión automática puede ser un factor que incrementa el precio, mientras que la manual está asociada a vehículos más accesibles**.

**6. Precio-Location**



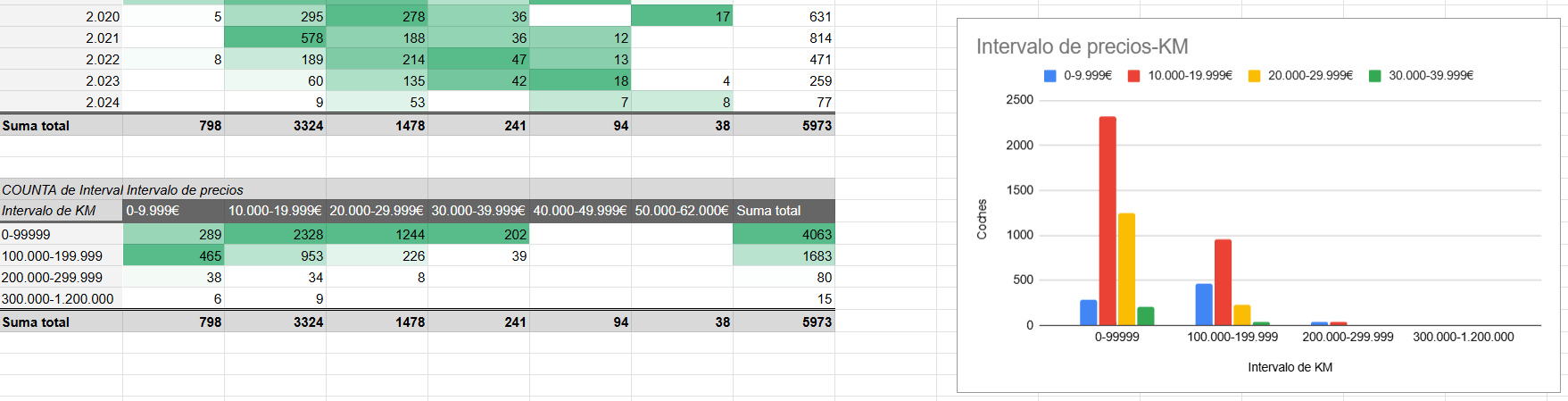
**La ubicación afecta de manera significativa al precio de los coches usados, reflejando tanto la demanda local como el poder adquisitivo de cada región**. Ciudades grandes como Madrid y Barcelona concentran la mayor cantidad de vehículos, incluyendo un número considerable en rangos de precios altos (**20.000 €** o más), lo que podría deberse a una mayor oferta de coches premium o recientes. En contraste, en localidades más pequeñas o menos urbanizadas, como Albacete o Almería, predominan los vehículos en rangos más bajos (**0-19.999 €**). Las categorías "Desconocido" también tienen una alta proporción de coches asequibles, posiblemente relacionados con vehículos no registrados en ubicaciones específicas.

**7. Precio-Año**



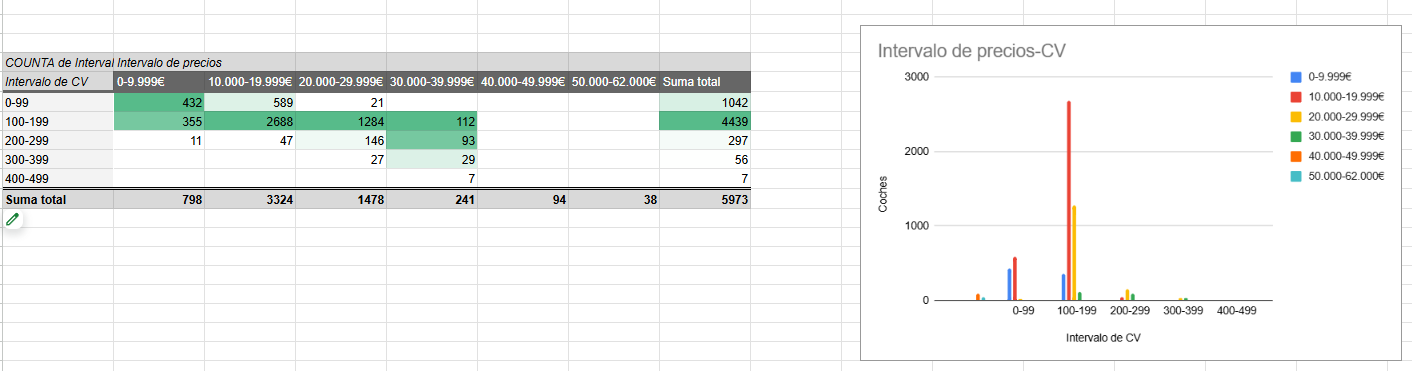
El año de fabricación afecta directamente al precio de los coches usados, reflejando la depreciación con el paso del tiempo. **Los vehículos más recientes (2020-2024) se concentran en rangos de precios altos (20.000 € o más),** debido a su menor antigüedad, tecnología moderna y menor desgaste. En cambio, **los coches más antiguos (2000-2010) predominan en los rangos más bajos (0-9.999 €),** lo que indica mayor depreciación y posiblemente mayores costes de mantenimiento. Los vehículos de años intermedios **(2015-2019)** tienden a distribuirse en rangos medios. En general, el año es un indicador clave del valor del coche en el mercado.

**8. Precio - KM**



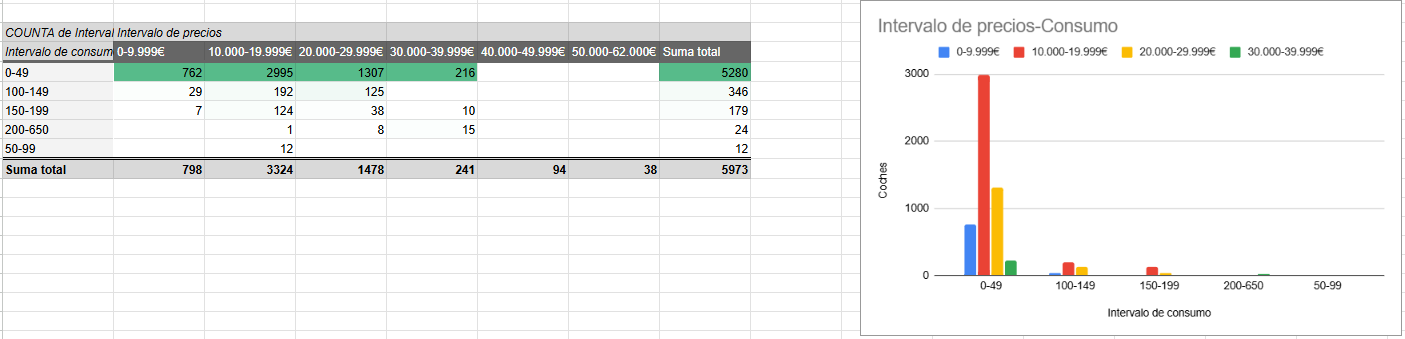
El kilometraje tiene un impacto significativo en el precio de los coches usados, ya que refleja su nivel de uso y desgaste. **Los coches con menor kilometraje (0-99.999 km) están predominantemente en los rangos de precios altos (20.000 € o más),** debido a su menor desgaste mecánico y a que tienen menos años. En contraste, **los vehículos con mayor kilometraje (200.000-299.999 km) se concentran en los rangos más bajos (0-9.999 €),** lo que indica una mayor depreciación y posibles gastos en mantenimiento. Los coches con kilometraje intermedio **(100.000-199.999 km)** presentan una distribución equilibrada entre los rangos de precios medios y bajos. En general, el kilometraje es un factor determinante que afecta directamente la percepción de calidad y valor del vehículo.

**10. Precio - CV**



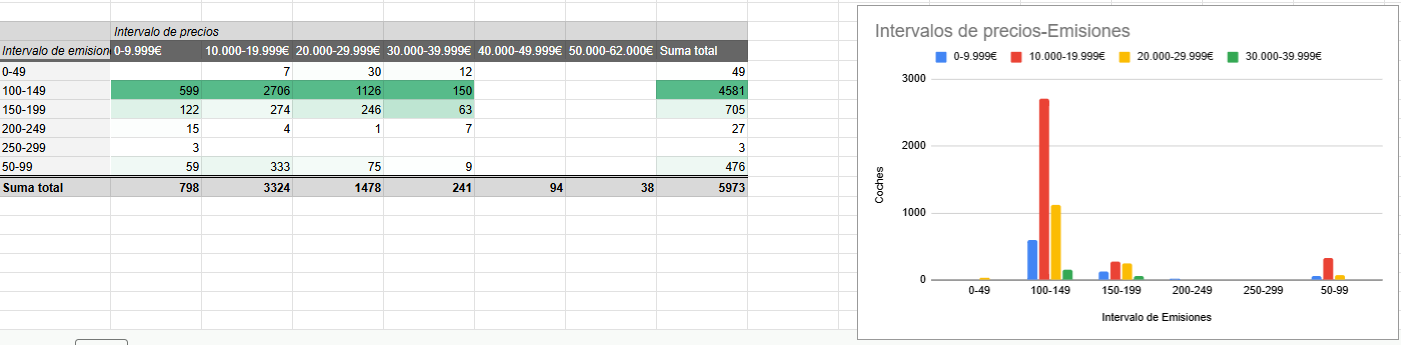
La variable de caballos de potencia (CV) influye directamente en el precio de los vehículos, ya que está relacionada con el rendimiento y la potencia del motor. **Los coches con menor potencia (0-99 CV) se concentran principalmente en los rangos de precios bajos (0-9.999 €)**,reflejando menor rendimiento y posiblemente modelos con menor motor o económicos. En comparación, **los vehículos con mayor potencia (200-299 CV y superiores) predominan en los rangos de precios más altos (20.000 € o más),** indicando modelos de alta gama o deportivos. La mayoría de los vehículos se encuentran en el intervalo de 100-199 CV, que domina los precios intermedios (10.000-19.999 €), siendo el rango más demandado por su equilibrio entre potencia y precio.

**11. Precio - Consumo**



La variable de consumo de combustible afecta significativamente al precio de los vehículos. **Los vehículos con un consumo bajo (0-49)** dominan todos los rangos de precios, especialmente en los intervalos de **10.000-19.999 €,** lo que indica que los consumidores priorizan la eficiencia en sus coches**.** En cambio, **los vehículos con consumos más altos (150-199 y 200-650) tienden a estar en los rangos de precios elevados (20.000 € o más),** suponiendo que son modelos más potentes o de mayor capacidad que implican un mayor gasto en combustible.

**12. Precio – Emisiones**



Por último, la variable emisiones también afecta directamente al precio de los vehículos, ya que está relacionada con las normativas ambientales y la eficiencia del motor. **La mayoría de los vehículos en el rango de 100-149 se concentran en precios de 10.000-19.999 €, lo que muestra que los coches con emisiones moderadas tienen mayor demanda en este rango de precios accesibles.** Por otra parte, **los vehículos con emisiones más altas (150-249) están asociados a precios más elevados (20.000 € o más),** probablemente debido a que corresponden a coches más potentes o de mayor categoría. Esto sugiere que las bajas emisiones son un factor clave para vehículos económicos, mientras que las altas emisiones tienen mayor cabida en modelos premium.

En definitiva, con estos análisis podemos confirmar que todas estas variables condicionan el precio de los coches usados.

**● Dashboard.**

Realizado en Google Sheets.

* Tablas dinámicas – gráficos de las variables que más afectan al precio de los coches usados en España.
* Incrustación de controles de filtros para visualizar los datos que nos interesen en cada momento.

**● Informe explicativo del análisis.**

**Resumen general del dashboard**

El **dashboard** proporciona una visión global del mercado de coches usados en España, con un total de **5.973 vehículos** analizados. Se presentan diferentes gráficos que muestran la distribución de precios en función de diversas variables como marca, tipo de combustible, tipo de transmisión, ciudad, año, kilómetros recorridos (KM), caballos de potencia (CV) y cantidad de consumo.

**1. % de precios de coches**

* **55.7%** de los coches están en el rango de **10.000 – 19.999€**, lo que indica que es el precio más común en el mercado.
* **24.7%** tienen un precio de **20.000 – 29.999€**, representando casi una cuarta parte del total.
* Menos del **10%** superan los **30.000€**, lo que sugiere que la mayoría de los coches usados están en precios accesibles.
* **(2,2% = 1.6% + 0,6%)** están en las categorías más altas (**40.000 - 62.000€)** lo que muestra una oferta limitada de coches de lujo.

**Conclusión**: La mayor parte del mercado se concentra en precios entre **10.000 y 29.999€**, con menos coches en segmentos más caros.

**2. Precios de coches por marca**

* **Peugeot, Seat y Renault** dominan la categoría de **10.000 - 19.999€**.
* Marcas premium como **BMW, Audi y Mercedes-Benz** tienen mayor cantidad de vehículos en los rangos de **20.000€ en adelante**.
* **Volkswagen y Porsche** son las marcas con precios más altos.

**Conclusión**: Las marcas generalistas tienen precios más bajos, mientras que las premium y eléctricas dominan los segmentos superiores.

**3. Precios de coches por kilómetros recorridos (KM)**

* Los coches con **0-99.999 km** están principalmente en el rango de **10.000 – 19.999€**.
* A medida que aumenta el kilometraje, disminuye el precio, con muy pocos coches con más de **300.000 km**.

**Conclusión**: El kilometraje es un factor clave en la depreciación del coche.

**4. Precios de coches por tipo de combustible**

* **Gasolina y diésel** dominan en todos los rangos de precios, con más presencia en los **10.000 – 19.999€**.
* **Eléctricos e híbridos** tienen precios más altos, con pocos vehículos en los rangos más bajos.
* **GLP y GNC** representan una minoría en el mercado

**Conclusión**: Los coches eléctricos siguen siendo más caros, lo que refleja su menor oferta en el mercado de segunda mano.

**5. Precios de coches por ciudad**

* Las ciudades con más coches usados son **Madrid y Barcelona**.
* En las **grandes ciudades** hay más diversidad en los rangos de precios, mientras que en ciudades pequeñas predominan los coches más baratos.
* En **ciudades pequeñas** hay más coches en los rangos más bajos.

**Conclusión**: La ubicación influye en el precio y la variedad de opciones disponibles.

**6. Precios de coches por número de caballos (CV)**

* Los coches con menos de **100 CV** tienen precios menores a **19.999€**.
* La mayoría de los coches en el mercado tienen entre **100 y 199 CV**, con precios mayoritarios en **10.000 – 19.999€**.
* Vehículos con más de **200 CV** son escasos y tienen precios más elevados.

**Conclusión**: La potencia del motor influye en el precio, con coches más potentes siendo más costosos.

**7. Precios de coches por tipo de transmisión**

* La mayoría de loscoches con **transmisión manual** están en el rango **10.000 – 19.999€**.
* Los coches **automáticos** tienen una mayor presencia en los rangos superiores de precios.

**Conclusión**: Los coches automáticos tienden a ser más caros, reflejando la mayor demanda y precios asociados.

**8. Precios de coches por año de compra**

* La mayoría de los coches usados son de los años **2015-2022**, con picos en **2019 y 2021**.
* Los modelos más nuevos tienen precios más altos.

**Conclusión**: Los coches más nuevos mantienen su valor, mientras que los modelos más antiguos se encuentran en los rangos de precios más bajos.

**9. Precios de coches por cantidad de consumo**

* La mayoría de los coches consumen entre **0-49 L/100km**, con predominancia en los **10.000 – 19.999€**.
* A medida que aumenta el consumo, hay menos coches en venta y los precios varían.

**Conclusión**: Los coches que menos consumen son los más comunes en el mercado de segunda mano.

**Conclusión final**

* **La mayoría de los coches usados en España tienen precios entre 10.000 y 29.999€.**
* **La marca, el número de kilómetros, el año de compra, el número de caballos y el tipo de transmisión son factores clave en el precio.**
* **Los coches eléctricos, automáticos y de marcas premium son los más caros.**
* **Las ciudades grandes tienen más oferta y variedad de precios.**
* **Los coches más eficientes en consumo dominan el mercado.**

**Este análisis permite identificar tendencias clave para compradores y vendedores, ayudando a tomar mejores decisiones en la compra-venta de coches usados**.

**6. 🔄 Próximos Pasos**

* Refinar el modelo predictivo usando más datos históricos.
* Implementar técnicas avanzadas de feature engineering para mejorar la precisión.
* Explorar el impacto de factores externos como campañas de marketing y eventos en las ventas.

**7.** **🤝 Contribuciones**

Las contribuciones son bienvenidas. Si deseas mejorar el proyecto, por favor abre un pull request o una issue.

**8. ✒ Autores y Agradecimientos**

Base de datos: <https://www.kaggle.com/datasets/gandpablo/cars-prices-for-machine-learning>

Créditos a GandPablo: Pablo Gandia (<https://github.com/gandpablo>)

**Enlace a la hoja de cálculo en Google Sheets:**

<https://docs.google.com/spreadsheets/d/1Fr2-dvqESFJ-wH4upJPTEDHWMfHkMytMiu23u7ffuTc/edit?usp=sharing>