# Proyecto de análisis de las ventas de coches en EE.UU

En la primera parte del proyecto, una vez descargado el fichero en formato csv, los transformamos a Excel en formato tabla para poder comenzar a trabajar con los datos.

Creamos la carpeta 'Data' que contendrá todos los ficheros. Dentro creamos la carpeta '1-Data\_raw' con dos subcarpetas más.

La primea es la carpeta ‘csv’ con el fichero original descargado, y la segunda, ‘excel’, con el fichero csv importado en Excel en formato tabla. Al pasar al formato tabla, corregimos el valor de columna Engine 'DoubleÂ' por 'Double'.

Subimos el proyecto a Git.

## Análisis de columnas

El siguiente paso que realizamos es el estudio de las columnas, indicando una descripción de cada columna y analizando que columnas son relevantes para el análisis.

* **Car\_id:** identificador único de la venta.
* **Date:** Fecha de la venta del vehículo.
* **Customer\_Name:** Nombre del cliente que ha comprado el coche.
* **Gender:** Género del cliente que ha comprado el coche.
* **Annual Income:** Ingresos anuales del comprador del coche.
* **Dealer\_Name:** Nombre del concesionario.
* **Company:** Marca del coche que se ha vendido.
* **Model:** Modelo del coche que se ha vendido
* **Engine:** Tipo de motor del coche vendido.
* **Transmission:** Tipo de transmisión del coche.
* **Color:** Color del coche.
* **Price ($):** Precio que ha costado el coche:
* **Dealer\_No:** Número del concesionario.
* **Body Style:** Estilo de la corrocería del coche.
* **Phone:** Télefono. Puede que sea el teléfono del comprador del coche. Es una columna irrevelente, así que no le vamos a dar importancia.
* **Dealer\_Region**: Región del concesionario.

### Columnas relevantes

Identificamos las columnas relevantes para el análisis.

Car\_id.

Date.

Gender.

Annual Income.

Dealer\_Name.

Company.

Model.

Engine.

Transmission.

Color.

Price ($).

Body Style.

Dealer\_Region.

## Transformación y limpieza

Creamos una subcarpeta nueva dentro de 'Data' con el nombre '2-Datos\_transformados'. Dentro creamos una nueva hoja Excel a partir de la anterior y la renombramos a 'car\_sales\_transformados'.

Creamos una columna nueva llamada 'Duplicados' para ver si tenemos algún identificador de ventas duplicado. No existen duplicados.

### Limpiamos los datos de la tabla

Eliminamos datos irrelevantes

Eliminamos las columnas customer\_name, dealer\_no, phone

#### Los tipos de datos correctos

Creamos una hoja nueva llamada 'Nulos' para ver cuantas columnas tienen valores nulos. Todos los datos están rellenos.

Vemos que todos los datos son correctos.

#### Normalizar los datos

Todas las columnas tienen los datos normalizados

## Análisis descriptivo de columnas numéricas

Creamos una subcarpeta nueva dentro de 'Data' con el nombre '3-Data\_analisis'. Dentro copiamos la Excel ‘car\_Sales\_transformados’ y la renombramos a ‘car\_sales\_analisis\_descriptivo’.

Creamos una nueva hoja Excel llamada 'analisis\_descriptivo\_numericas'

Borramos la hoja Nulos.

Eliminamos la columna 'Duplicado' de la hoja car\_sales.

Los insights de las columnas numéricas son:

**Annual Income:** Los datos muestran que el ingreso medio de los compradores es de 830.840,29 dólares. Hay un gran grupo de compradores que cobran 13.500 dólares tal y como detalla la moda. Lo que sugiere que hay valores atípicos elevados que elevan la media. Tenemos una curtosis muy alta, lo que sugiere colas muy pesadas con valores muy extremos. El coeficiente de asimetría, nos indica que tenemos muchos clientes con ingresos bajos y algunos con ingresos muy altos.

**Price ($):** Los datos muestran que el precio medio de los coches es de 28.090 dólares. El precio más común es de 22.000 dólares. El coeficiente de asimetría nos indica que hay muchos coches con precios bajos y algunos con precios muy elevados.

## Análisis descriptivo de columnas categóricas

Sobre la misma hoja Excel anterior, creamos una hoja nueva llamada 'analisis\_descriptivo\_categorica'.

Analizamos las siguientes columnas categóricas y obtenemos los siguientes insights:

**Gender:** El género parece ser un factor determinante en las ventas. Los hombres realizan muchas compras de coches que las mujeres.

**Dealer\_Name:**  Vemos que los datos se dividen en dos grupos. El primer grupo de concesionarios tienes unas ventas que rondan los 630 coches vendidos, mientras que el otro grupo está entre los 1250-1300 coches vendidos.

**Company:** Podemos ver que las marcas más vendidas son Chevrolet, Dodge, Ford con más de 1.500 unidades vendidas. Seguramente sean marcas de confianza para el público de este análisis. Mientras que las marcas menos vendidas son Saab, Jaguar, Infinity y Hyundai. Puedes que este menos número de ventas se deba al precio caro de los coches o la poca confianza en la marca.

**Engine:** Los datos muestras que el tipo de motor no es un factor determinante en las ventas. Si que tiene un número mayor de ventas el doble árbol de levas (Double overhead camshaft), pero no es muy significativo.

**Transmission:** Los datos muestran que la transmisión no es un factor determinante en las ventas. La transmisión automática tiene un mayor número de ventas. Si nos fijamos en los datos, podríamos decir que el tipo de motor de doble árbol de levas se vende con transmisión automática y el árbol de levas en modo manual.

**Color:** Los datos muestras solamente 3 tipo de colores. El que más se vende con diferencia es blanco pálido, seguido del negro. El rojo sería el tercer color más vendido.

**Body Style:** Los SUV y los hathcback son las carrocerías preferidas por los compradores. Mientras que los hardtop y Passenger son los menos deseados por los compradores.

**Dealer\_Region:** La región que más ventas de coches ha tenido es la de Austin con 4.135 seguida de Janesville con 3.821. Mientras que las que menos han tenido son las de Middletown, Greenville, Pasco y Aurora que casi han vendido lo mismo. Scottsdales se encontraría en medio con 3.433 ventas.

## Análisis descriptivo de columnas temporales

Sobre la misma hoja Excel anterior, creamos una hoja nueva llamada 'analisis\_descriptivo\_fechas' y analizamos las columnas de tipo fecha:

**Date**

* **Año**: Los datos muestran que en el año 2023 aumentaron un 10% las ventas de coche con respecto al año anterior.
* **Trimestre**: Podemos ver que el 4 trimestre es cuando las personas compran más coches. Puede ser porque a final de año sacan ofertas para que los comerciales lleguen a los objetivos anuales.
* **Mes**: Observamos que, las ventas de coches aumentan notablemente los dos últimos meses del año. Puede deberse a lo que comentamos en el punto anterior, están esperando a las ofertas de final de año.

## Análisis descriptivo de las numéricas con las ventas

En la carpeta ‘Data\_analisis’, duplicamos la Excel 'car\_sales\_analisis\_descriptivo' y la nombramos 'analisis\_descriptivo\_numericas\_ventas'.

Sobra la Excel nueva, eliminamos la hoja 'analisis\_descriptivo\_categorica' y 'analisis\_descriptivo\_fechas'. Creamos la hoja 'analisis\_ventas\_numericas'.

Realizamos el análisis de las columnas numéricas con las ventas de coches a través de tablas y gráficos.

Los insights de las columnas numéricas son:

**Anual Income**: El salario medio anual en estados unidos el año 2022 es de 77.000 y el año 2023 es de 80.115. Fuente salario medio: (https://datosmacro.expansion.com/mercado-laboral/salario-medio/usa)

Si realizamos una agrupación por 77.000, vemos que este rango de ingresos supone un 22,07% de las ventas. El resto de las ventas se reparten entre el resto de personas que tiene un salario superior a la media anual de estados unidos.

Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

Con los datos agrupados por 1.000.000 vemos que el rango de 10.080 a 1.010.079 tiene el 68.99% de las ventas.

Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.

Estos datos se pueden ver influenciados debido a que hay poca gente que llegue a cobrar más del salario medio.

**Price**: Realizamos una agrupación del precio por 10.000. Vemos que las mayores ventas de coches, un 38,49%, se realizan en los coches con un precio entre 11.200 y 21.199. Conforme el coche sube de precio, las ventas van disminuyendo. También vemos que los coches con el menor precio de mercado (1.200 a 11.199) tienen unas ventas de un 2,61%, unas ventas muy inferiores, equivalentes a los coches con un precio superior (61.200 a 71.199).

Gráfico

El contenido generado por IA puede ser incorrecto.

## Análisis descriptivo de las categóricas y temporales con las ventas.

En la carpeta ‘Data\_analisis’, duplicamos la Excel 'car\_sales\_analisis\_descriptivo' y la nombramos 'analisis\_descriptivo\_cat\_temp\_ventas'.

Eliminamos la hoja 'analisis\_descriptivo\_numericas'. Creamos la hoja 'vista\_categoricas' y 'vista\_fechas'.

Creamos una columna nueva 'Annual income type' para agrupar los tipos de ingresos en 'Annual income type' en función del valor de la mediana de la columna ‘Annual Income’. Si los ingresos son iguales o inferiores a 100.000, el valor será 'Low'. Si los ingresos están entre 101.001 y 735.000 (el doble de la mediana), el valor será 'High'. Y si están por encima de 1.470.001, el valor será 'Superior'.

Cogemos el valor de 100.000 ya que el salario medio anual en estados unidos ronda los 77.000$. El siguiente rango de ingresos lo establecemos en 734.000$, ya que es la mediana de la columna 'Annual income'.

Realizamos el análisis de las columnas categóricas con las ventas de coches a través de tablas y gráficos dinámicos.

Los insights de las columnas categóricas y temporales son:

**Gender:** Existe una gran diferencia en el género a la hora de comprar coches. Los hombres realizan el triple más de compras de coches que las mujeres. Podría ser debido a que el coche se pone a nombre del hombre, pero tiene un uso familiar.

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

**Color:** Los datos analizan únicamente tres colores, siendo el color ‘Pale white’ el más comprado con un 47,08%, casi la mitad de las ventas analizadas. El siguiente color es el negro con un 32,87% y en último lugar está el rojo con un 20,05%.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Body style:**Lo compradores prefieren estructuras tipo SUV (un coche grande – familiar) o Hatchback (un coche más pequeño). Ambos tienen una cuota casi del 26% de mercado cada una. El resto se reparten entre Sedan, Passenger y Hardtop.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Company:** Las marcas más compradas son Chevrolet con una cuota del 7,61%, seguidas de Dodge y Ford. Estas tres marcas son de origen americano, lo que puede transmitir más confianza a los compradores. Las marcas con menor cuota de mercado son Jaguar (de origen inglés), Infiniti (de origen japonés) y Saab (de origen sueco).

**Pantalla de juego de computadora

El contenido generado por IA puede ser incorrecto.**

## Análisis adicional de las categóricas.

**Annual Income type:** Vemos que las personas que tienen unos ingresos superiores (más de 735.000 dólares) realizan la mitad de las compras. El siguiente grupo serían los que tienen unos ingresos altos con un 27.78%, siendo las personas con unos ingresos bajos, los que menos coches compran.

**Gráfico

El contenido generado por IA puede ser incorrecto.**

Los insights de las columnas temporales son:

**Ventas por años:** Vemos que el año 2023 ha tenido casi 3.000 ventas más que el año anterior.

**Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.**

Esto supone un aumento del 10% con respecto al año anterior.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

**Ventas por trimestres:** Las ventas en el primer trimestre del año son muy bajas con un 12,97%. Conforme pasan los meses, las ventas van aumentando hasta llegar a un 35,94% en el último trimestre del año. Parece que la gente se espera a final de año a comprar un coche.

Interfaz de usuario gráfica, Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

**Ventas por meses:** Los meses con más ventas de coches son noviembre y diciembre con un 15% de las ventas totales. Los meses más flojos son enero y febrero con un 3%. Claramente los dos últimos meses del año son lo que concentran el mayor número de ventas.

**Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.**

**Ventas:** Podemos ver el mismo patrón de ventas en los dos años. Los primeros meses del año son flojos y luego van subiendo las ventas. Curiosamente en septiembre tenemos un pico de ventas, en octubre un bajón y luego los meses más fuertes con noviembre y diciembre. También vemos que en el año 2023 han crecido el número de ventas en todos los meses del año.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Elección de las columnas relevantes para el dashboard

Creamos una carpeta llamada 'Datos\_dashboard'. En esta carpeta copiamos el fichero 'analisis\_descriptivo\_cat\_temp\_ventas' y lo renombramos a 'dashboard'.

Eliminamos todas las hojas a excepción de la 'car sales'. Creamos una hoja nueva llamada 'Columnas\_relevantes\_dashboard'y otra llamada 'Kpi'.

Hemos analizados las columnas para decidir nuestros kpi, segmentadores y gráficas.

* KPI:
  + Total ventas.
  + Importe de las ventas.
  + Importe medio de las ventas.
* Segmentadores:
  + Year.
  + Month.
  + Gender.
  + Transmision.
  + Color.
* Gráficos:
  + Company.
  + Annual Income Type.
  + Body style.
* Gráfico temporal:
  + Date