# PROYECTO FINAL DATA ANALYTICS

**CURSO: DATA ANALYTICS** 

TEMA: VENTAS DE
VEHÍCULOS USADOS EN
ESPAÑA

Angel Lucero

Docente: Rebecca Figueroa

# INDICE

INTRODUCCIÓN	2
HIPÓTESIS	2
TIPO DE ANALISIS	3
ALCANCE	4
DATASET	4
TABLAS (Versión1)	5
TABLAS (Versión2)	6
DIAGRAMA ENTIDAD RELACIÓN (Versión 1)	7
DIAGRAMA ENTIDAD RELACIÓN (Versión 2)	8
MODELO RELACIÓNAL	9
DOCUMENTACIÓN DE TRANSFORMACIÓNES	10
TABLAS (PowerBI)	11
TABLA CALENDARIO	12
COLUMNAS Y MEDIDAS CALCULADAS	13
MOCKUP	14
STORYTELLING (Paleta de colores)	15
PORTADA PRINCIPAL DEL DASHBOARD	16
PORTADA VENTAS GLOABALES	16
SEGMENTACIÓNES ELEGIDAS	17
TARJETAS	18
GRAFICO VENTAS POR MODELOS Y MARCAS	18
TABLA PROMEDIO PRECIOS POR MARCAS	19
GRAFICO DE EVOLUCIÓN DE VENTAS POR MES	19
MAPAS Y PASOS APLICADOS	20, 21
PORTADA PREFERENCIAS DEL USUARIO	22
GRAFICO DE ANILLOS SOBRE CARACTERISTICAS VEHICULARES	22
ELECCIONES DE COLOR DE LOS USUARIOS	23
GRAFICO DE VENTAS POR MARCA Y TRANSMISIÓN	23
GRAFICO DE AREA DOBLE	24
TECNOLOGÍAS USADAS	24
FUTURAS LINEAS	25
CONCLUSIÓN	26
BIBLIOGRÁFIA	

# INTRODUCCIÓN

Para la elección del siguiente trabajo, obtuvimos información correspondiente a las ventas de automóviles usados en España, teniendo en cuenta datos particulares como la cantidad de marca, modelo, precio, kilometraje, años de fabricación, etc.

El análisis contiene información desde la fecha 15/07/2020 hasta el 26/01/2021, teniendo en cuenta los principales portales web de ventas de vehículos en el país.

# HIPÓTESIS

Obtener información completa y organizada, acerca de las preferencias al momento de elegir un vehículo usado, utilizando diferentes técnicas que nos permitirán determinar la elección preferida para el mercado de usuarios de España en el período anteriormente mencionado.

Basándonos en la información que obtengamos, podremos planificar y desarrollar un plan de acción para potenciar las ventas de elecciones menos deseadas por los usuarios, que tengan características similares a las más demandadas.

Para ello, utilizaremos análisis cruzados de tablas con información precisa acerca de los ítems de nuestra base de datos, y determinaremos la elección más y menos favorable para el usuario promedio.

# TIPO DE ANÁLISIS

**Descriptivo:** Reingeniería de bases de datos para hacer un reporte sobre el volumen de ventas en condiciones estándar, teniendo en cuenta las preferencias de los usuarios según determinadas características de los vehículos.

**Diagnóstico:** Implementación de Power BI con el objetivo de mostrar las relaciones existentes entre las características y preferencias de los usuarios.

**Predictivo:** Modelo predictivo para determinar las preferencias que tendrán los usuarios con base en los datos obtenidos anteriormente observando las características más solicitadas.

**Prescriptivo:** Modelo prescriptivo, con base en el modelo predictivo analizar planes de acción para lograr una mayor eficiencia en ventas de modelos de autos con poca demanda por los usuarios.



El Dashboard es a nivel **estratégico.** Se puede apreciar de manera macro para que la gerencia pueda tomar las mejores decisiones y aplicar los planes de acción correspondientes para aumentar su cantidad de ventas.

09/07/2022



A continuación, se adjunta el EXCEL con las bases de datos organizadas y estructuradas utilizadas en el presente trabajo.

Para poder ingresar al Excel, ir al archivo subido en la carpeta drive del siguiente link



# TABLAS

De acuerdo con la consigna de la segunda entrega del proyecto final, creamos una tabla en forma de guía, teniendo en cuenta la información solicitada de cada tabla, sus campos, el tipo de dato que se va a manejar, y también el tipo de clave en cada una de las tablas.

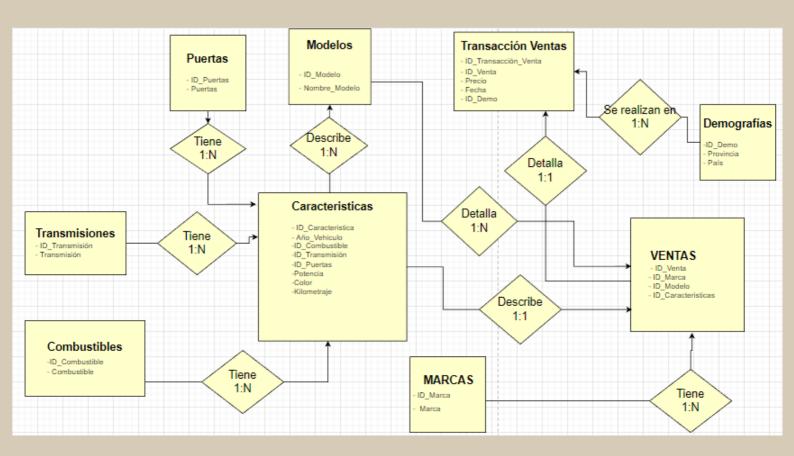
Table	Comme	Time de dete	Tine de desse
Tabla 🔻	Campo 🔻	Tipo de dato	Tipo de clave 🔻
Ventas	ID_Venta	INT NOT NULL	PK
Ventas	Ventas	INT	N/A
Ventas	ID_Marca	INT NOT NULL	FK
Ventas	ID_Característica	INT NOT NULL	FK
Ventas	ID_Modelo	INT NOT NULL	FK
Marcas	ID_Marca	INT NOT NULL	PK
Marcas	Marca	INT	N/A
Transacción Ventas	ID_Transacción_Venta	INT NOT NULL	PK
Transacción Ventas	ID_Venta	INT NOT NULL	FK
Transacción Ventas	Precio	FLOAT	N/A
Transacción Ventas	Fecha_venta	DATETIME	N/A
Transacción Ventas	ID_Demo	INT NOT NULL	FK
Ventas por modelo	ID_Modelo	INT NOT NULL	PK
Ventas por modelo	Nombre_Modelo	VARCHAR	N/A
Ventas por modelo	ID_Característica	INT NOT NULL	FK
Características	ID_Característica	INT NOT NULL	PK
Características	Color	TEXT	N/A
Características	ID_Cantidad_puertas	INT NOT NULL	FK
Características	Año_Vehículo	INT	N/A
Características	ID_Combustible	INT NOT NULL	FK
Características	Transmisión	INT NOT NULL	FK
Características	Potencia	VARCHAR	N/A
Usabilidad	ID_Modelo	INT NOT NULL	FK
Usabilidad	Precio	FLOAT	N/A
Usabilidad	Kilometraje	VARCHAR	N/A
Demografia	ID_Demo	INT NOT NULL	PK
Demografia	Provincia	TEXT	N/A
Demografia	Pais	TEXT	N/A
Puertas	ID_Cantidad_Puertas	INT NOT NULL	PK
Puertas	Puertas	INT	N/A
Combustible	ID_Combustible	INT NOT NULL	PK
Combustible	Combustible	VARCHAR	N/A

Teniendo en cuenta la normalización de el dataset se restructura la tabla, agregando y modificando tablas.

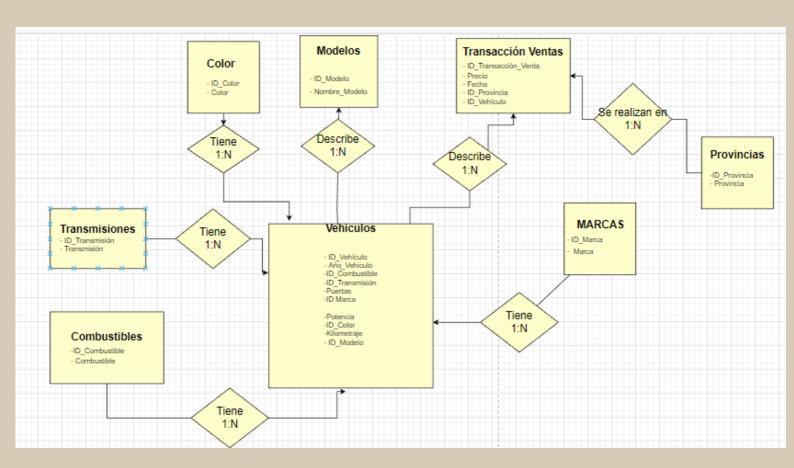
Tabla	Campo	Tipo de dato	Tipo de clave
Modelos	ID_Modelo	INT NOT NULL	PK
Modelos	Nombre_Modelo	INT NOT NULL	FK
Color	ID_Color	INT NOT NULL	PK
Color	Color	INT NOT NULL	FK
Transmisiones	ID_Transmisión	INT NOT NULL	PK
Transmisiones	Transmisión	INT NOT NULL	FK
Combustibles	ID_Combustible	INT NOT NULL	PK
Combustibles	Combustible	INT NOT NULL	FK
Marcas	ID_Marca	INT NOT NULL	PK
Marcas	Marca	INT NOT NULL	FK
Provincias	ID_Provincia	INT NOT NULL	PK
Provincias	Provincia	TEXT	FK
Transacción Ventas	ID_Transacción_Venta	INT NOT NULL	PK
Transacción Ventas	Precio	FLOAT	FK
Transacción Ventas	Fecha	DATETIME	FK
Transacción Ventas	ID_Provincia	INT NOT NULL	FK
Transacción Ventas	ID_Vehículo	INT NOT NULL	FK
Vehículos	Año	INT NOT NULL	FK
Vehículos	ID_Vehículo	INT NOT NULL	PK
Vehículos	ID_Combustible	INT NOT NULL	FK
Vehículos	ID_Transmisión	INT NOT NULL	FK
Vehículos	Puertas	VARCHAR	N/A
Vehículos	ID_Marca	INT NOT NULL	FK
Vehículos	Potencia	VARCHAR	N/A
Vehículos	ID_Color	INT NOT NULL	FK
Vehículos	Kilometraje	VARCHAR	N/A
Vehículos	ID_Modelo	INT NOT NULL	FK

# DIAGRAMA ENTIDAD RELACIÓN

Primer diagrama entidad relación usando la primera tabla planteada en este proyecto.



Después de haber realizado modificaciones en la tabla y haberla restructurado, se tomo la decisión de modificar el diagrama entidad relación optando por el esquema estrella.



23/07/2022

# DESCRIPCIÓN

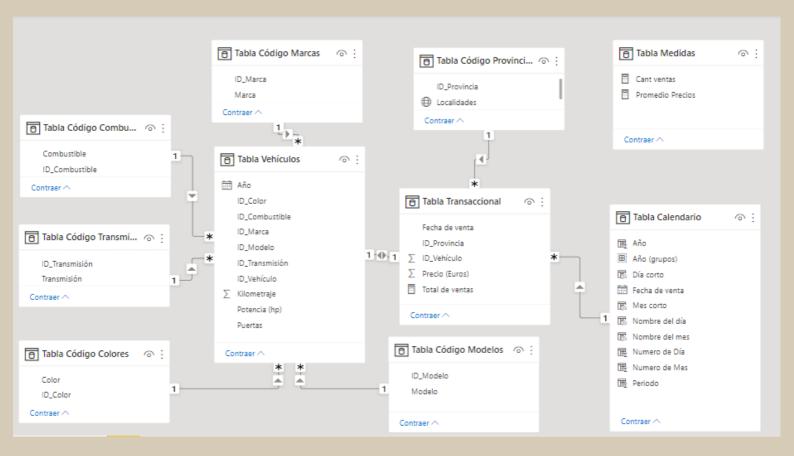
A continuación dejamos los links para poder acceder tanto al diagrama entidad relación como a la tabla de información, donde se pueden agregar comentarios en ambos documentos, en caso de tener oportunidades de mejora.

Para poder ingresar al Excel, ir al archivo subido en la carpeta drive del siguiente link





# MODELO RELACIONAL



# DOCUMENTACIÓN DE TRANSFORMACIONES

### CONEXIÓN A BASE DE DATOS

```
let
    Origen = Excel.Workbook(File.Contents(#"URL DATOS VENTAS DE VEHICULOS"), null, true),
    #"Tabla Código Colores_Sheet" = Origen{[Item="Tabla Código Colores",Kind="Sheet"]}[Data],
    #"Encabezados promovidos" = Table.PromoteHeaders(#"Tabla Código Colores_Sheet", [PromoteAllScalars=true]),
    #"Tipo cambiado" = Table.TransformColumnTypes(#"Encabezados promovidos",{{"Color", type text}, {"ID_Color", Int64.Type}})
in
    #"Tipo cambiado"
```

Efectuamos la conexión a la base de datos por medio de archivo excel, además a través del editor avanzado se logra hacer el cambio de URL para cada una de las tablas.

Agregamos pestañas correspondientes de inicio, tablero e información. (Mockup)

### Se procede a realizar:

- Normalizar base de datos.
- Modificar tipo de datos de los campos.
- Eliminar datos duplicados.
- Generar parámetros para el origen.

## **TABLAS**

Se cargan las tablas correspondientes, a fin de usarlas para la creación de los gráficos que se verán a continuación.

- > 🖫 Tabla Medidas
- > III Tabla Calendario
- > III Tabla Código Colores
- > III Tabla Código Combustibles
- > III Tabla Código Marcas
- > III Tabla Código Modelos
- > III Tabla Código Provincias
- > III Tabla Código Transmisiones
- > III Tabla Transaccional
- > III Tabla Vehículos

# TABLA CALENDARIO

00	× ✓								
╗	Fecha de venta	Año ▼	Numero de Mes	Nombre del mes 🔻	Mes corto 💌	Numero de Día 🔻	Nombre dle día 🔻	Día corto ▼	Periodo 💌
	21/12/2020 18:14:41	2020	12	diciembre	dic	21	lunes	lu.	4
	23/11/2020 17:25:39	2020	11	noviembre	nov	23	lunes	lu.	4
-8	24/10/2020 21:59:26	2020	10	octubre	oct	24	sábado	sá.	4
	5/1/2021 16:43:56	2021	1	enero	ene	5	martes	ma.	1
	26/11/2020 10:48:00	2020	11	noviembre	nov	26	jueves	ju.	4
	5/11/2020 16:37:22	2020	11	noviembre	nov	5	jueves	ju.	4
	12/1/2021 12:53:11	2021	1	enero	ene	12	martes	ma.	1
	10/1/2021 10:51:02	2021	1	enero	ene	10	domingo	do.	1
	21/9/2020 18:55:10	2020	9	septiembre	sep	21	lunes	lu.	3
	20/11/2020 12:12:06	2020	11	noviembre	nov	20	viernes	vi.	4
	19/1/2021 12:36:38	2021	1	enero	ene	19	martes	ma.	1
	3/1/2021 12:19:40	2021	1	enero	ene	3	domingo	do.	1
	20/1/2021 18:15:37	2021	1	enero	ene	20	miércoles	mi.	1
	9/1/2021 18:52:40	2021	1	enero	ene	9	sábado	sá.	1
	23/11/2020 23:34:33	2020	11	noviembre	nov	23	lunes	lu.	4
	14/11/2020 12:50:18	2020	11	noviembre	nov	14	sábado	sá.	4
	5/12/2020 13:24:45	2020	12	diciembre	dic	5	sábado	sá.	4
	4/11/2020 13:41:07	2020	11	noviembre	nov	4	miércoles	mi.	4
	14/10/2020 17:03:49	2020	10	octubre	oct	14	miércoles	mi.	4
	20/10/2020 17:01:57	2020	10	octubre	oct	20	martes	ma.	4
	29/12/2020 21:47:50	2020	12	diciembre	dic	29	martes	ma.	4
	7/12/2020 18:39:04	2020	12	diciembre	dic	7	lunes	lu.	4
	20/11/2020 09:18:09	2020	11	noviembre	nov	20	viernes	vi.	4

Pagina: 12

### COLUMNAS Y MEDIDAS CALCULADAS

#### **COLUMNAS CALCULADAS**

#### **Tabla Calendario**

- a. Año = YEAR('Tabla Calendario'[Fecha de venta])
- b. Numero de Mes = MONTH('Tabla Calendario'[Fecha de venta])
- c. Nombre del mes = 'Tabla Calendario'[Fecha de venta].[Mes]
- d. Mes corto = FORMAT('Tabla Calendario'[Fecha de venta], "mmm")
- e. Numero de Día = DAY('Tabla Calendario'[Fecha de venta])
- f. Nombre del día = FORMAT('Tabla Calendario'[Fecha de venta], "dddd")
- g. Día corto = FORMAT('Tabla Calendario'[Fecha de venta], "ddd")
- h. Periodo = QUARTER('Tabla Calendario'[Fecha de venta])

#### **MEDIDAS CALCULADAS**

Se detallan las medias utilizadas en el proyecto que se usaron en indicadores y gráficos.

- a. Cant ventas = COUNT('Tabla Transaccional'[ID\_Vehículo])
- b. Promedio Precios = AVERAGE('Tabla Transaccional'[Precio (Euros)])

# MOCKUP

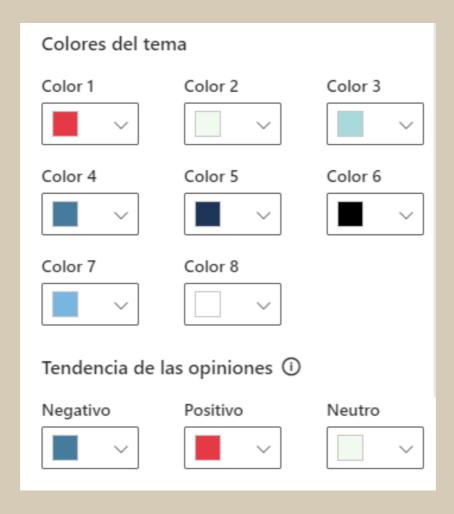
Se muestra a continuación el boceto de lo que se quiere llegar aproximadamente en la creación del tablero de control.



# STORYTELLING

### PALETAS DE COLORES

Se muestra los colores elegidos para usar en la visualización del dashboard



### PORTADA PRINCIPAL DEL DASHBOARD



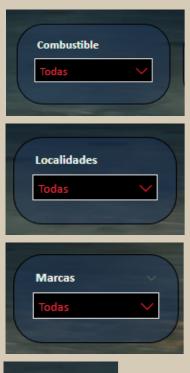
### PORTADA VENTAS GLOBALES



# SEGMENTACIONES ELEGIDAS

A continuación se muestra las segmentaciones elegidas para ser utilizadas en el dashboard.

Con estos botones de filtros, podemos obtener información más precisa, según la necesidad del usuario que utilice el dashboard. Teniendo en cuenta variaciones generales como lo son las marcas, el tipo de combustible utilizado y la ubicación geográfica de las transacciones.



#### Filtro combustible

#### Filtro localidades

#### Filtro marcas



**Ícono Borrar filtros:** Creado mediante marcadores, posee la funcionalidad de mostrar el tablero sin los filtros aplicados.

01/10/2022



**Botón portada** 

### **Botón ventas globales**



# Botón preferencias del usuario



# **TARJETAS**

En ellos, podemos ver las transacciones realizadas, la marca más elegida, y el promedio de precio abonado por cada transacción realizada. Al utilizar el panel de forma orgánica, podemos obtener información precisa sobre las ventas de cada marca, el precio promedio abonado por la misma, teniendo en cuenta las operaciones filtradas.







# GRAFICO VENTAS POR MODELO Y MARCAS

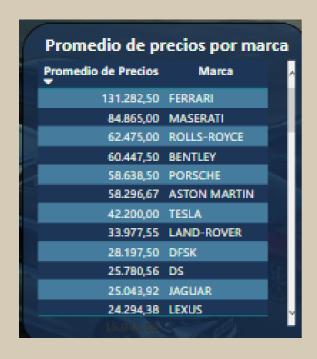
Ambos gráficos funcionan de manera similar, reflejan la cantidad de ventas generadas según las 10 marcas más transaccionadas, los modelos más elegidos, y las ventas obtenidas en cuanto al modelo más elegido.





### TABLA DE PROMEDIO DE PRECIOS POR MARCAS

Gracias a esta tabla, podemos obtener, en líneas generales, cuáles son las marcas más cotizadas en el mercado, según su promedio de precio, en las ventas efectuadas del período. Cabe destacar, que este gráfico no tiene en cuenta factores determinantes sobre las ventas, como lo son el año de los vehículos, o su devaluación de la marca con el paso del tiempo.



# GRAFICO DE EVOLUCIÓN DE VENTAS POR MES

En el mismo, podemos ver como las transacciones aumentan progresivamente hacia el último período del año



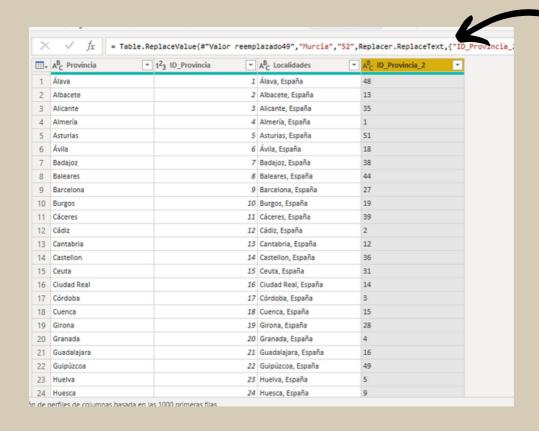
### MAPA Y PASOS APLICADOS



Acá se opto por el mapa de formas, se utilizaron datos como Total de ventas de la Tabla Transaccional y Provincia de la Tabla Provincias.

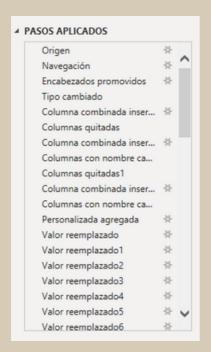
Con esto se quiere llegar a poder visualizar la cantidad de ventas en distintas provincias de mayor a menor con un color degradado.

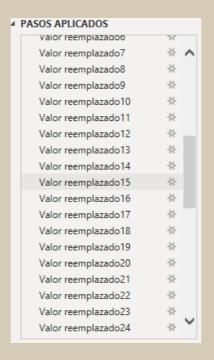
A continuación se detallara el proceso de ajusto que se realizo en el mapa de formas Provincias, debido a que el archivo de formato TopoJson de España ya contaba con un ID en su clave de tipo de mapa para identificar las provincias, y eso nos generaba un conflicto con nuestro ID ya definido en la Tabla Provincias. Donde el nombre de las provincias no coincidían con la locación geográfica de esa provincia.

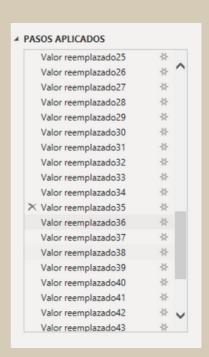


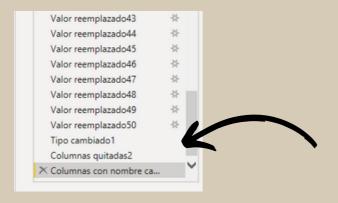
Paso 1: Acá se duplico una columna y se cambio los valores ID en cada registro, teniendo en cuenta los ID que venían ya definidos en el archivo de formato TopoJson del mapa de España, para que las provincias coincidan según el ID ya preestablecido.

**Paso 2:** Se muestra los paso de los valores remplazados, para hacer coincidir los ID que ya venían preestablecidos en el mapa de España.









Paso 3: En estos últimos pasos aplicados, se cambio el tipo de dato de la nueva columna creada y se elimino la anterior, para posteriormente cambiarle el nombre a la columna creada.

₩.	A <sup>B</sup> <sub>C</sub> Provincia  ▼	A <sup>B</sup> <sub>C</sub> Localidades	1 <sup>2</sup> <sub>3</sub> ID_Provincia $\checkmark$
1	Álava	Álava, España	48
2	Albacete	Albacete, España	13
3	Alicante	Alicante, España	35
4	Almería	Almería, España	1
5	Asturias	Asturias, España	51
6	Ávila	Ávila, España	18
7	Badajoz	Badajoz, España	38
8	Baleares	Baleares, España	44
9	Barcelona	Barcelona, España	27
10	Burgos	Burgos, España	19
11	Cáceres	Cáceres, España	39
12	Cádiz	Cádiz, España	2
13	Cantabria	Cantabria, España	12
14	Castellon	Castellon, España	36
15	Ceuta	Ceuta, España	31
16	Ciudad Real	Ciudad Real, España	14
17	Córdoba	Córdoba, España	3
18	Cuenca	Cuenca, España	15
19	Girona	Girona, España	28
20	Granada	Granada, España	4
21	Guadalajara	Guadalajara, España	16
22	Guipúzcoa	Guipúzcoa, España	49
23	Huelva	Huelva, España	5

Paso 4: Acá ya se puede visualizar el formato final de la Tabla Provincias, donde los ID de esta tabla ya coincide con el ID que venia prestablecido en el archivo de formato TopoJson del mapa España.

### PORTADA PREFERENCIAS DEL USUARIO



# GRAFICOS DE ANILLOS SOBRE CARACTERISTÍCAS VEHICULARES

En dichos gráficos podemos obtener información acerca de la cantidad de vehículos vendidos, según el número de puertas, así como también el tipo de combustible. Podemos corroborar que los de 5 puertas, así como los vehículos diésel, tienen una alta predominancia sobre las otras opciones de sus respectivas categorías.





### ELECCIONES DE COLOR DE LOS USUARIOS

Continuando con las características de los vehículos más transaccionados, podemos ver que en la elección de colores, se mantiene una elección hacia los tonos neutros, liderando la demanda el color gris.



# GRAFICO DE VENTAS POR MARCAS Y TRANSMISIÓN

Este gráfico de columnas agrupadas, nos brinda una imagen general de las marcas más vendidas, y el tipo de transmisión elegido según sus consumidores. Las marcas más demandadas con el tipo de transmisión automática, tienden a ofrecer vehículos de una gama superior, mientras que las más elegidas con el cambio manual, son del tipo generalistas



# GRAFICO DE DOBLE ÁREAS, DE CANTIDAD DE VENTAS SEGÚN KILOMETRAJE YA RECORRIDO

En este gráfico, podemos visualizar la relación de precio según el uso, en promedio de cada marca. Es correcto afirmar que la marca más valuada en relación a su uso, es BMW, ya que a pesar de tener un promedio de kilometraje superior al resto, mantiene un nivel de ventas muy competitivo e inclusive superior a la mayoría de su competencia.



# TECNOLOGIAS USADAS



**EXCEL:** Se utilizó para normalizar el dataset y crear las tablas.



**SQL:** Se utilizó para determinar los diferentes tipos de llaves en las tablas y su relacionamiento, logrando así diseñar un dataset orgánico y simple.



**POWER BI:** se utilizó para generar transformaciones de datos y así poder diseñar reportes para la visualización de forma correcta e interactiva.



**CANVA:** Se utilizó para poder archivar la documentación correspondiente de los pasos aplicados del proyecto.

# FUTURAS LINEAS

Contando con la información sobre las transacciones generadas del mercado automotor, podemos realizar diferentes tipos de análisis para potenciar y predecir el comportamiento del mismo, generando una maximización de ventas en el rubro, por lo que los próximos pasos del proyecto son:

- profundizar cuales son los vehículos más vendidos según su segmento, finalidad y utilización.
- Indagar sobre cuales son los factores que afectaron las transacciones, que generaron un incremento durante el último cuatrimestre del año.
- Obtener información sobre la cantidad de vehículos que se venden nuevos.
- Obtener información demográfica de las ventas ya realizadas.

# CONCLUSIÓN

Actualmente el mercado de vehículos usados, tiende a la compra de vehículos de gama media, obteniendo el mayor volumen de ventas el segmento C, entre los que se encuentran los modelos más vendidos: Golf, Serie 3 e Ibiza.

Teniendo esto en cuenta, el monto transaccional promedio, acompaña esta proyección, ya que a pesar de que la base de datos contiene vehículos de gama alta, el precio promedio oscila en los € 15.000.

También verificamos que la elección de los usuarios es mayor hacia los vehículos de procedencia europea, liderando este ranking, marcas provenientes de Alemania.

Con respecto a las características elegidas, el tipo de transmisión automática, tuvo mayor participación dentro de las cuatro marcas más vendidas, por lo que es posible que el mercado apunte a un mayor consumo de esta característica.

Es importante destacar, que a pesar de que la llegada de la electrificación vehicular, es inminente, en la preferencia de los usuarios, no destaca esta opción de propulsión al momento de comprar un vehículo usado, sino que son el diésel y la gasolina.

# **BIBLIOGRAFÍA**

#### Web del dataset:

https://www.kaggle.com/datasets/datamarket/venta-de-coches