Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

Proyecto Integrado III

Profesora: Ana María López

PREICA2402B020107

S20 - Evidencia 3 de Avance - Proyecto Integrado III

Grupo 10

Elaborado por: Juan Carlos Gutiérrez Riaño

Fecha 8 diciembre de 2024

**Índice**

1. Introducción
2. Definición del Problema de Negocio (Evidencia avance 1)
3. Fuentes de Datos y descripción
4. Hipótesis para la base de datos vehículos eléctricos en Colombia
5. Métricas de Éxito
6. ¿Qué datos tengo para lograr el objetivo?
7. Granularidad de la tabla
8. ¿Estos datos me sirven para lograr el objetivo?
9. Planeación con Trello
10. Exploración del Dataset y Conclusiones
11. Enlaces Links
12. Limpieza y Métodos de Limpieza (Evidencia avance 2) *Páginas 15 a la 32*
13. S20 - Evidencia 3 de Avance - Proyecto Integrado III (Evidencia avance 3) *Páginas 32 a la 48*
14. Dashboard
15. Creación de repositorio - Inicio de nuestro portafolio
16. Link video explicativo final
17. Conclusión
18. **Introducción**

En este trabajo de investigación se pretende dar repuestas al análisis de un Dataset tomado de los datos abiertos del gobierno de Colombia, que tiene contenido de los datos de carros eléctricos e híbridos que se han vendido en Colombia.

Este análisis nos permitirá saber cual es el tipo de carro que prefieren en Colombia, su tipo de carrocería, en que departamentos o municipios se vendió más y para que se están utilizando ya sean particulares o de trabajo, también podrían destinarse al gobierno y cuantos son, para lograr estos objetivos se realizara un análisis del Dataset con la herramienta COLAB con sus respectivas conclusiones, video y vinculo respectivo , también se planifican el proyecto y las tareas en una tabla de Trello, un análisis ,limpieza de los datos ,junto a un Dashboard sobre la investigación para llegar a afirmar las tesis o negarla , al final del ejercicio se dan las conclusiones.

1. **Definición del Problema de Negocio** (Evidencia avance 1)

**. Descripción del problema: Describa el problema de negocio que se abordará en el proyecto. Sea claro y conciso.**

Es conocido por todos que si logramos cambiar la movilidad de vehículos a gasolina por vehículos eléctricos en Colombia aportaríamos a disminuir el impacto negativo en el ambiente reduciendo las emisiones de gases a la atmosfera, sin embargo, la sustitución de esta tecnología esta pasando por varias dificultades, como lo son el precio del vehículos, soporte técnico y suministro de electricidad o estaciones de carga.

En este caso de estudio se pretende consultas algunos datos relevantes para determinar qué factores son favorables para la compra de estos vehículos en Colombia y además conocer las tendencias de crecimiento, utilidad y gustos de los compradores para a si tomar mejores decisiones en las estrategias de ventas y respaldo

En este caso de estudio se pretende analizar un Dataset de Vehículos eléctricos – Híbridos del Ministerio de Transporte de Colombia.

También se planifican las tareas en una table de Trello, junto a la generación del reporte del Dataset en COLAB con sus respectivas conclusiones y vinculo

**. Impacto empresarial: Explique el impacto potencial del problema en el negocio. Cuantifique el impacto si es posible.**

El gran impacto sobre la economía colombiana es considerable porque generaríamos perdidas con respecto a la competitividad en la región y se estancaría la infraestructura eléctrica del país, la venta se ve afectada por falta de incentivos gubernamentales y desconfianza en el respaldo de repuestos. actualmente se ofertan tres tipos de vehículos

Totalmente eléctrico

Hibrido a diésel

Hibrido a gasolina

Se requiere conocer mas datos sobre las tendencias de consumo de los compradores en Colombia para mejoras la oferta de estos, mejorar la innovación y a la vez la sostenibilidad de las empresas que puedan importar o fabricar estos vehículos, además de fortalecer las empresas del sector energético,

En la actualidad global la venta de estos vehículos esta prevista su crecimiento gradual en el 17 % hasta el 2030 Si Colombia alcanzara una tasa de compra similar, la industria automotriz local podría aumentar sus ingresos en hasta un 15% para 2030.

Su impacto seria cuantificado en:

Contaminación y Salud Pública: enfermedades respiratorias podría aumentar hasta en un 20%, debido a la contaminación en áreas urbanas.

Competitividad Automotriz: La industria automotriz perdería relevancia en el mercado regional, reduciendo sus ingresos en hasta un 15% al 2030

Emisiones y Compromisos Ambientales: Colombia podría no alcanzar su meta de reducción del 51% en emisiones de gases de efecto invernadero para 2030

Datos tomados de Ministerio de Ambiente y Desarrollo Sostenible de Colombia, Organización Mundial de la Salud (OMS) y el Banco Mundial, Bloomberg New Energy Finance (BNEF) y Banco Interamericano de Desarrollo (BID)

. **Pregunta de negocio: Formule una pregunta de negocio específica y medible que el proyecto busca responder.**

¿Qué factores influyen en el aumento de las ventas de automóviles eléctricos en Colombia?

1. **Fuentes de Datos y descripción**

**Fuente de los datos:**

[**https://www.datos.gov.co/**](https://www.datos.gov.co/)

[https://www.datos.gov.co/resource/7qfh-tkr3.json Creada el 17/3/2022](https://www.datos.gov.co/resource/7qfh-tkr3.json%20Creada%20el%2017/3/2022)

Número de Vehículos Eléctricos – Híbridos

Actualizado 20 de abril de 2024

Última actualización de los datos 14 de diciembre de 2022

Última actualización de metadatos 20 de abril de 2024

Fecha de creación 17 de marzo de 2022

Formato: CSV

Tamaño: 56545 filas y 22 columnas)

API en formato JSON Fuente: [www.datos.gov.co](http://www.datos.gov.co)

**4- Hipótesis para la base de datos vehículos eléctricos en Colombia**

**Hipótesis 1:**

Los automóviles eléctricos de más potencia son los que más se venden en Colombia

**Hipótesis 2:**

Las marcas chinas son las más vendidas en el mercado colombiano

**Hipótesis 3:**

El uso de los vehículos eléctricos comprados en Colombia, la mayoría se destinan para pasajeros, y la minoría para carga

**Hipótesis 4:**

El vehículo eléctrico preferido por el consumidor colombiano son de tipo SEDAN

**5- Métricas de Éxito**

**Número de ventas:**

Total, de registros de ventas por modelo, marca

**Promedio de cilindraje y potencia:**

Comparar estos promedios de ventas entre las diferentes modelos eléctricos y los híbridos (diésel o gasolina) en porcentaje de ventas.

**Distribución por servicio:**

Análisis del porcentaje de ventas por tipo de vehículo (sedan, carga)

**Tendencias de ventas por año:**

Analizar cómo han cambiado las ventas a lo largo de los años para identificar patrones.

**Análisis por municipio o departamento:**

Identificar en qué municipios se realizan más ventas y si hay diferencias significativas.

**6- ¿Qué datos tengo para lograr el objetivo?**

¿Cómo se llaman las columnas y qué significan? (Diccionario de datos), Tipos de Datos: Especificar los tipos de datos de cada columna (numérico, categórico, fecha, texto, etc.).

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de la columna** | **Descripción** | **Nombre del campo API** | **Tipo de Dato** |
| COMBUSTIBLE |  | combustible | **Texto** |
| ESTADO |  | estado | **Texto** |
| MODELO | MODELO DEL AUTOMOTOR | modelo | **Número** |
| FECHA\_REGISTRO |  | fecha\_registro | **Marca de tiempo variable, date** |
| AÑO\_REGISTRO |  | anio\_registro | **Número** |
| CLASIFICACION | CLASIFICACION VEHICULO | clasificacion | **Texto** |
| CLASE |  | clase | **Texto** |
| SERVICIO | TIPO SERVICIO DE AUTOMOTOR | servicio | **Texto** |
| MARCA |  | marca | **Texto** |
| LINEA |  | linea | **Texto** |
| CARROCERIA |  | carroceria | **Texto** |
| CILINDRAJE |  | cilindraje | **Texto** |
| MODALIDAD |  | modalidad | **Texto** |
| ORGANISMO\_TRANSITO |  | organismo\_transito | **Texto** |
| MUNICIPIO |  | municipio | **Texto** |
| DEPARTAMENTO |  | departamento | **Texto** |
| CAPACIDAD\_CARGA |  | capacidad\_carga | **Número** |
| CAPACIDAD\_PASAJEROS |  | capacidad\_pasajeros | **Número** |
| PESO |  | peso | **Número** |
| POTENCIA |  | potencia | **Número** |
| EJES |  | ejes | **Número** |
| CANTIDAD |  | cantidad | **Número** |

1. **Granularidad de la tabla**

La Granularidad es de tipo fino: Datos de ventas que se evidencia por las fechas de registros. por hora, incluyendo la cantidad vendida de cada producto en cada tienda. Por ejemplo:

**• Fecha: 2024-10-28**

**• Hora: 10:00**

**• Producto: clase de automóvil**

**• Cantidad: 1**

**• Municipio: localización del vehículo vendido**

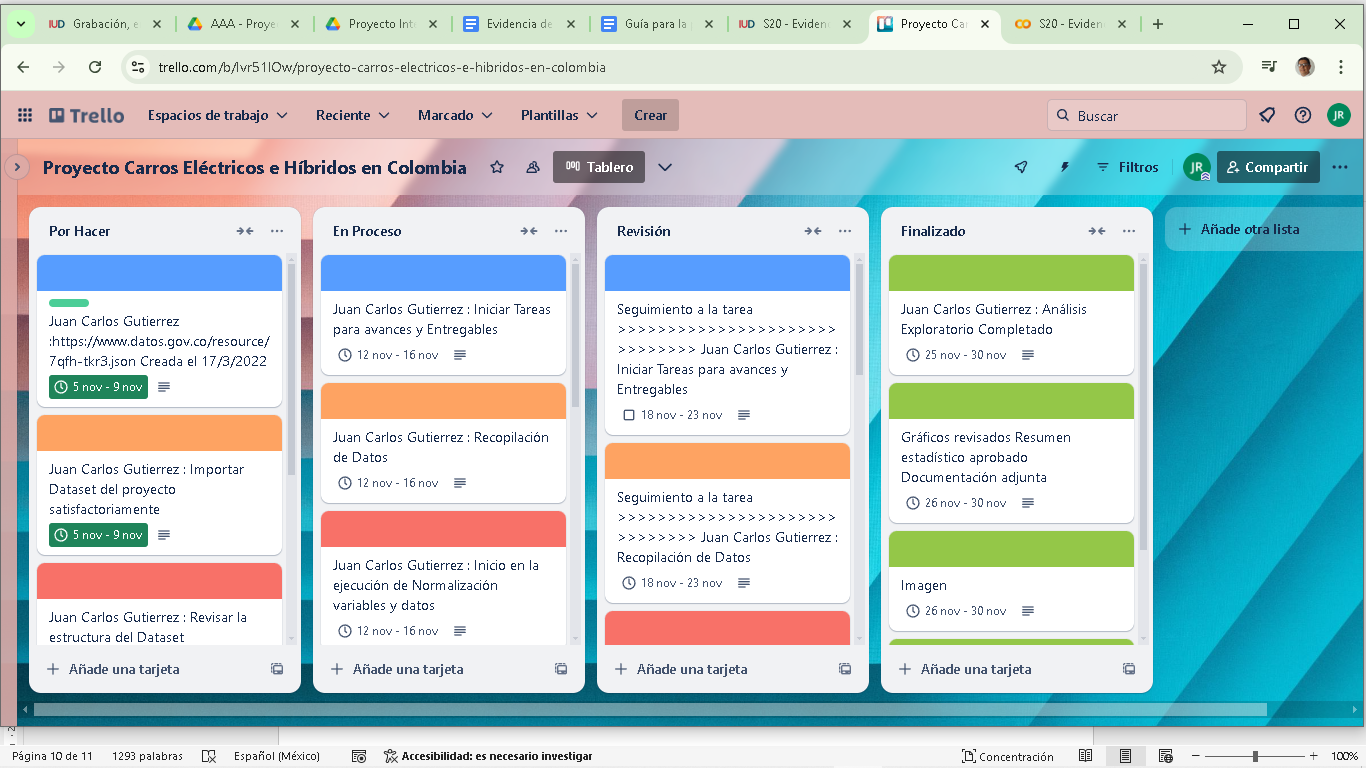
**• Marca del vehículo bien establecida**

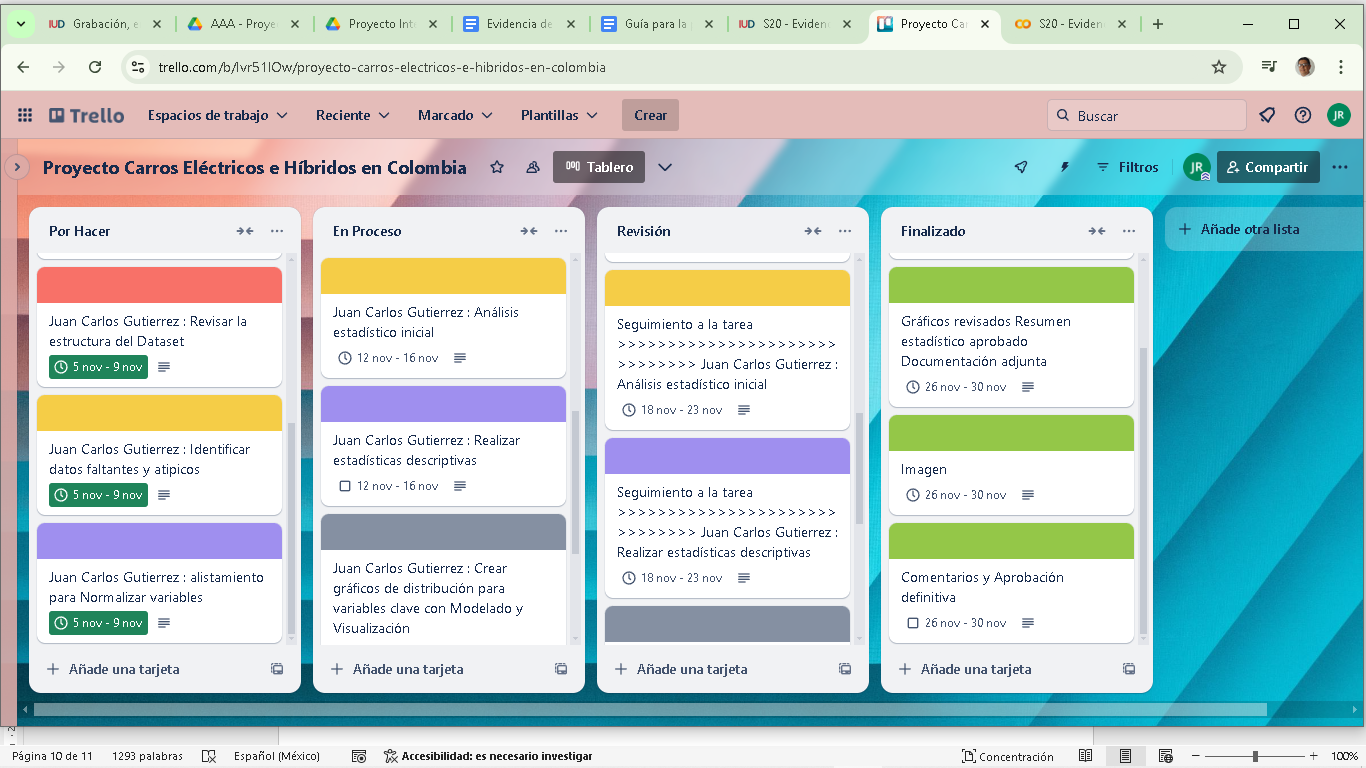
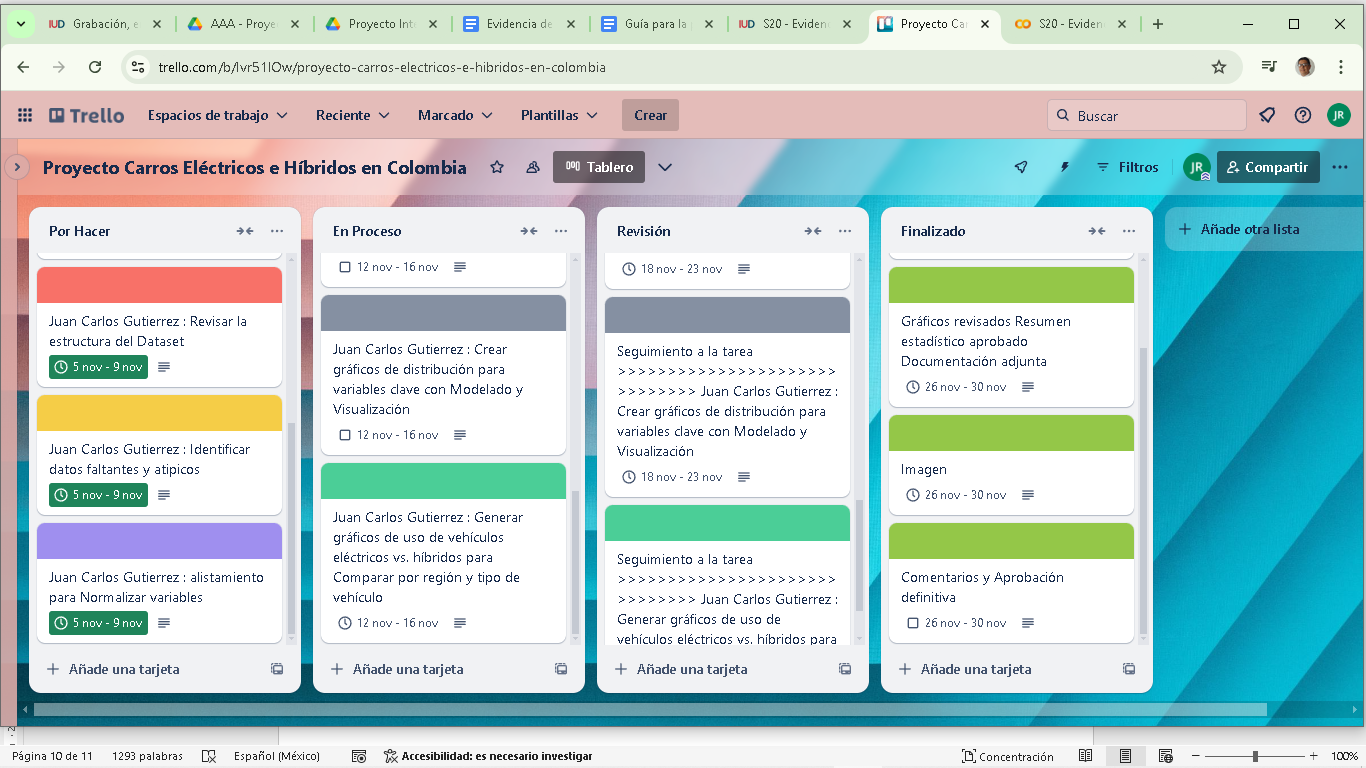
**• Eléctrico o hibrido (gasolina, diésel)**

**• etc**

1. **¿Estos datos me sirven para lograr el objetivo?**

Los datos Si me sirven para contestar las hipótesis y preguntas necesarias

1. **Planeación con Trello**

****

1. **Exploración del Dataset y Conclusiones**

**La exploración de datos se realizó en el COLAB**

* Los GASO ELEC representa la mayoría de los registros, con 45,550 vehículos, en comparación con 9,574 vehículos eléctricos (ELECTRICO) y 1,421 vehículos diésel-eléctricos (DIES ELEC).
* Los registros de vehículos eléctricos puros Con solo 9574, esta categoría aún tiene una baja representación en comparación con los híbridos
* Los 1421 vehículos diésel-eléctricos (DIES ELEC) son una minoría dentro de la muestra
* Los automóviles son mayoría de los registros, con 54682 vehículos en comparación con solo 1623 motocicletas y 240 motocarros
* La mayoría de los vehículos eléctricos e híbridos están registrados para uso particular con 52701 vehículos en comparación con 3575 vehículos de servicio público y otros 269 vehículos no determinados o del gobierno
* Se tiene una gran diversidad 203 marcas distintas de vehículos registradas
* La carrocería de tipo WAGON tiene la mayor representación 32800 vehículos
* Las carrocerías que le siguen son SEDAN (8935 vehículos) y **HATCH BACK** (7764 vehículos)
* El promedio de cilindraje de los carros híbridos son 1727.8 cc

1. **Enlaces Links**

Link: TRELLO

[**https://trello.com/invite/b/67269ed227e4415960a791cd/ATTI4104e2baf0770ec53bfcc9e02edd2d9fBA3E64B6/proyecto-carros-electricos-e-hibridos-en-colombia**](https://trello.com/invite/b/67269ed227e4415960a791cd/ATTI4104e2baf0770ec53bfcc9e02edd2d9fBA3E64B6/proyecto-carros-electricos-e-hibridos-en-colombia)

Link: Video

[**https://drive.google.com/file/d/1RPQPyBPMmw-Rm\_2\_3sG-oK-wE7\_6TtwJ/view?usp=sharing**](https://drive.google.com/file/d/1RPQPyBPMmw-Rm_2_3sG-oK-wE7_6TtwJ/view?usp=sharing)

1. **Limpieza y Métodos de Limpieza (Evidencia avance 2)**

**Link: Video**

<https://drive.google.com/file/d/1zBDF7TV7LhRU94JMO6qxehlHmCVfIIYy/view?usp=sharing>

Pasos de consulta y limpieza:

**. Carga de Datos y bibliotecas**

Se carga un Dataset desde un archivo CSV que contiene información sobre vehículos eléctricos e híbridos.

df = pd.read\_csv('Numero\_de\_Veh\_culos\_El\_ctricos\_-\_Hibridos\_20241027.csv', low\_memory=False)

. Bibliotecas

import numpy                 as np

import pandas                as pd

import matplotlib.pyplot     as plt

import seaborn               as sns

import folium  #needed for interactive map

from folium.plugins import HeatMap

from   collections           import Counter

from   datetime              import datetime

. **Exploración número de columnas**

Index(['COMBUSTIBLE', 'ESTADO', 'MODELO', 'FECHA\_REGISTRO', 'AÑO\_REGISTRO',

'CLASIFICACION', 'CLASE', 'SERVICIO', 'MARCA', 'LINEA', 'CARROCERIA',

'CILINDRAJE', 'MODALIDAD', 'ORGANISMO\_TRANSITO', 'MUNICIPIO',

'DEPARTAMENTO', 'CAPACIDAD\_CARGA', 'CAPACIDAD\_PASAJEROS', 'PESO',

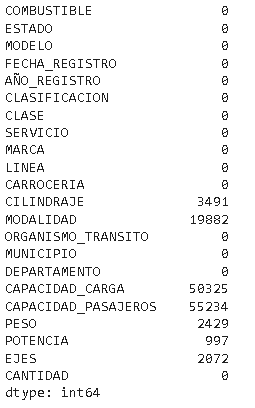
'POTENCIA', 'EJES', 'CANTIDAD'],

dtype='object')

. **Consulta para encontrar el número de valores nulos por columna**

valores\_nulos = df.isnull().sum()

print(valores\_nulos)



. **Identifico las columnas numéricas para datos nulos**

columnas\_numericas = ['CILINDRAJE', 'CAPACIDAD\_CARGA', 'CAPACIDAD\_PASAJEROS', 'PESO', 'POTENCIA', 'EJES', 'CANTIDAD']

. **Reemplazo los datos nulos en las columnas numéricas con 0 (cero)**

for col in columnas\_numericas:

    df[col] = df[col].fillna(0)

. **Identifico columnas de texto con los datos nulos**

columnas\_texto = ['COMBUSTIBLE', 'ESTADO', 'MODELO', 'CLASIFICACION', 'CLASE', 'SERVICIO', 'MARCA', 'LINEA', 'CARROCERIA', 'MODALIDAD', 'ORGANISMO\_TRANSITO', 'MUNICIPIO', 'DEPARTAMENTO']

. **Reemplazo los datos nulos de las columnas de texto con la palabra (Desconocido)**

for col in columnas\_texto:

    df[col] = df[col].fillna('Desconocido')

. **Verifico la limpieza de los datos nulos en todas las columnas tanto numéricas como de texto**

print(df.isnull().sum())



. **Consulto formatos de fecha** **y el resultado fue satisfactorio, Todas las fechas tienen formato correcto**

try:

    pd.to\_datetime(df['FECHA\_REGISTRO'])

    print("Todas las fechas tienen formato correcto")

except Exception as e:

    print("Errores en formato de fecha:")

    # Consulto registros con formato incorrecto

    mascara = pd.to\_datetime(df['FECHA\_REGISTRO'], errors='coerce').isna()

    print(df[mascara]['FECHA\_REGISTRO'])

<ipython-input-16-332fa8955c75>:2: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

pd.to\_datetime(df['FECHA\_REGISTRO'])

**Todas las fechas tienen formato correcto**

**. Identifico valores atípicos en columnas numéricas como 'CILINDRAJE', 'CAPACIDAD\_CARGA', 'CAPACIDAD\_PASAJEROS', 'PESO', 'POTENCIA', 'EJES', y 'CANTIDAD', con rango intercuartílico (IQR)**

Los valores atípicos parecen agruparse en dos extremos:

Extremadamente bajos como 0.0.

Extremadamente altos por encima de un límite específico

columnas = ['CILINDRAJE', 'CAPACIDAD\_CARGA', 'CAPACIDAD\_PASAJEROS', 'PESO', 'POTENCIA', 'EJES', 'CANTIDAD']

for columna in columnas:

    # Calcular Q1, Q3 e IQR

    Q1 = df[columna].quantile(0.25)

    Q3 = df[columna].quantile(0.75)

    IQR = Q3 - Q1

    # Definir límites para valores atípicos

    limite\_inferior = Q1 - 1.5 \* IQR

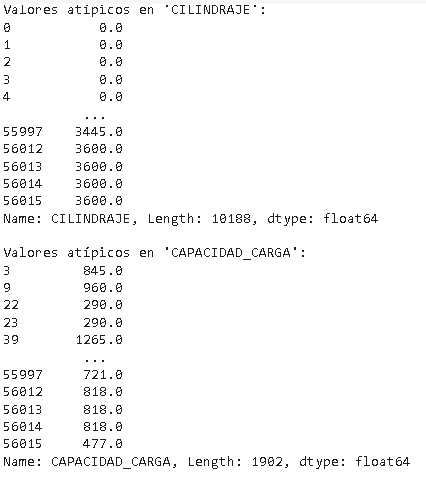
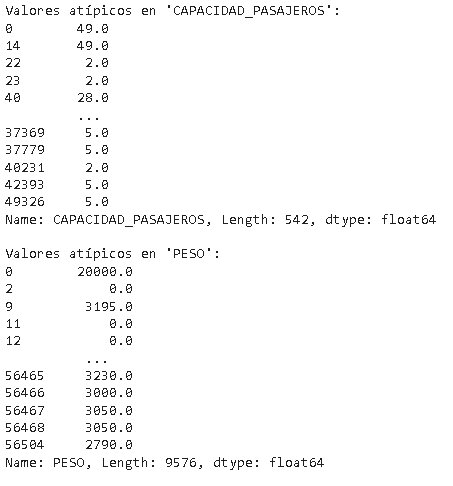
    limite\_superior = Q3 + 1.5 \* IQR

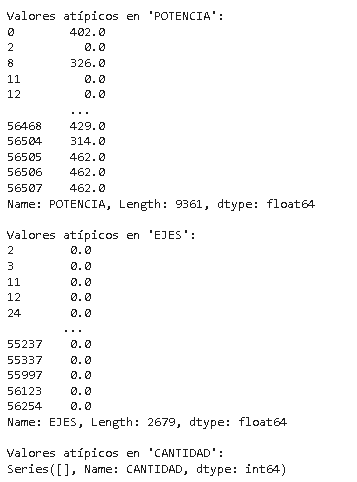
    # Filtrar valores atípicos

    valores\_atipicos = df[(df[columna] < limite\_inferior) | (df[columna] > limite\_superior)]

    print(f"\nValores atípicos en '{columna}':")

    print(valores\_atipicos[columna])



. **consistencia en categorías, valores únicos en columnas categóricas y valores poco comunes (frecuencia baja)**

**Datos encontrados**

45,550+9,574+1,421= 56,545 vehiculosen total

GASO ELEC con el 80.6% vehículos híbridos

ELECTRICO constituye el 16.9% O 100% eléctricos.

DIES ELEC tiene un menor porcentaje 2.5%diésel y electricidad.

Valores poco comunes en ESTADO: REGISTRADO 1

La columna 'CLASIFICACION' no tiene valores atipicos

AUTOMOVIL: 54,682 registros (96.7%)

MOTO: 1,623 registros (2.9%)

MOTOCARRO: 240 registros (0.4%)

La columna 'CLASE' CAMIONETA: 29,565 registros 52.3%

AUTOMOVIL: 17,061 registros 30.2%

CAMPERO: 4,541 registros 8.0%

**Valores atípicos identificados**

CUATRIMOTO: 3 registros

TRACTOCAMION: 1 registro

**Servicios:**

Particular: 52,701 registros 93.2%

Público: 3,575 registros 6.3%

Oficial: 269 registros 0.5%

La mayoría son de servicio particular como tendencia en la mayoría de regiones no se encontró valores atípicos

columnas\_categoricas = ['COMBUSTIBLE', 'ESTADO', 'CLASIFICACION', 'CLASE', 'SERVICIO']

for columna in columnas\_categoricas:

    print(f"\nValores únicos en {columna}:")

    print(df[columna].value\_counts())

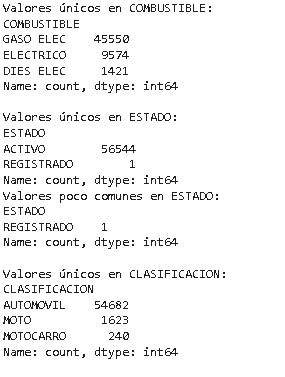
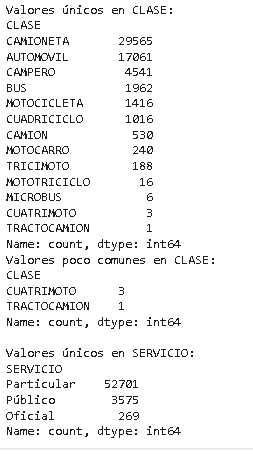
    # comunes de (frecuencia baja)

    valores\_raros = df[columna].value\_counts()[df[columna].value\_counts() < 5]

    if len(valores\_raros) > 0:

        print(f"Valores poco comunes en {columna}:")

        print(valores\_raros)

**. Consulto los valores atípicos en las columnas numéricas y también los datos alejados del resto de los datos aplicando método IQR**

**Inconsistencias encontradas**

**cilindraje:**

valor 0.00 y hasta 2 son muy bajos podrían ser indicar datos faltantes o fueron los que se rellenaron con 0

vehículos con cilindrada cero datos no se ingresaron correctamente

cilindrajes superiores a 10,000 podrían ser de vehículos como camiones o errores en los datos ya que la mayoría de los vehículos suelen tener una cilindrada mucho menor

**carga**: los datos con cero puede ser un error o carros que no tienen capacidad de carga, otros datos de 10.000 ,hasta 30.000 pueden ser error o carros como camiones

Promedio (mean): 55.16 Desviación estándar (std): 458.20 Valor mínimo (min): 0.00 Primer cuartil (25%): 0.00 Mediana (50%): 0.00 Tercer cuartil (75%): 0.00 Valor máximo (max): 30,000.00

**pasajeros:**

Como el primer cuartil (25%), la mediana (50%) y el tercer cuartil (75%) tienen el valor de 0 sugiere que la mayoría de los vehículos registrados no tienen capacidad de transportar pasajeros.

En las columnas como 49, 50, 53, 161. muestra una grancapacidad de transporte de pasajeros , esto podrian ser autobuses

**peso:**

como en las columnas anteriores los valores 0 indican datos faltantes o errores de digitacion

Tenemos valores como 20,000, 25,000, 50,000 y 60,000 estos valores son muy altos por que si la mayoria de vehiculos son de transporte particular no superan los 2,410

**potencia:**

como en las columnas anteriores los valores 0 indican datos faltantes o errores de digitacion

S aprecia valores altos como 8,000, 10,000, 80,000. los cuales son mayores que el promedio de 169.46 indicado valores exagerados o depronto son camiones y tambien se nota por valores alejados del promedio

**ejes:**

como en las columnas anteriores los valores 0 indican datos faltantes o errores de digitacion

valores atípicos son 0, 3, 21, 1, 5, y 4 encuanto a los valores de 2 al 3 es normal en carros pequenños o semicargueros , de 3 a 5 pueden ser camiones pero el dato mas atipico es 2

# columnas numéricas atipico

for columna in columnas\_numericas:

    print(f"\nEstadísticas para {columna}:")

    print(df[columna].describe())

    # atípicos usando el método IQR o datos alejados del resto de los datos en un conjunto

    Q1 = df[columna].quantile(0.25)

    Q3 = df[columna].quantile(0.75)

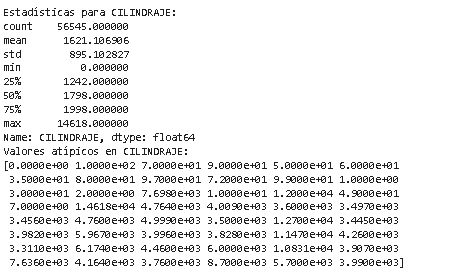
    IQR = Q3 - Q1

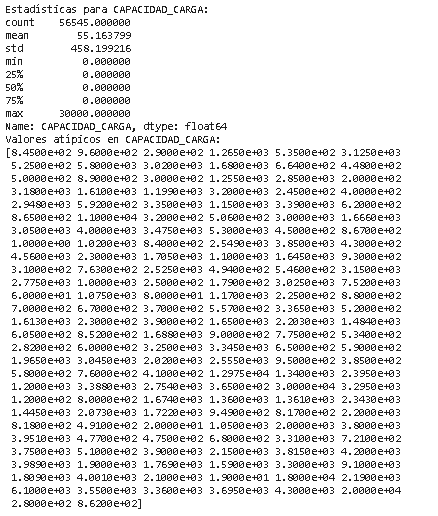
    atipicos = df[(df[columna] < (Q1 - 1.5 \* IQR)) | (df[columna] > (Q3 + 1.5 \* IQR))]

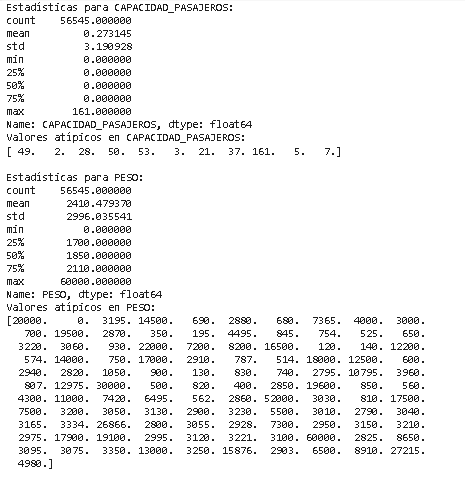
    if len(atipicos) > 0:

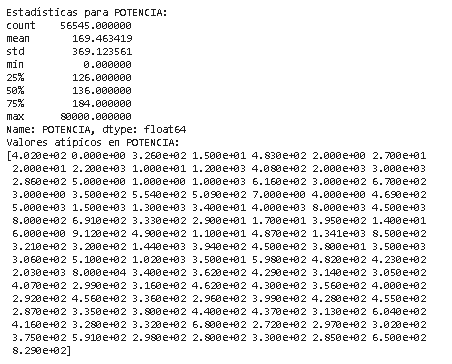
        print(f"Valores atípicos en {columna}:")

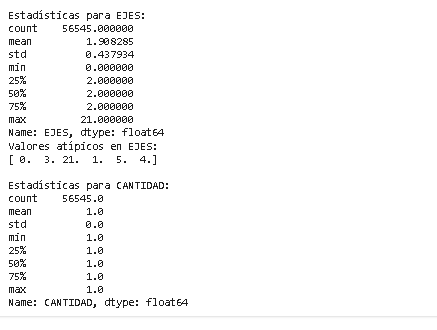
        print(atipicos[columna].unique())











. **Histplot**

# Todas las  columnas numéricas

plt.figure(figsize=(15, 10))

for i, columna in enumerate(columnas\_numericas):

    plt.subplot(3, 3, i + 1)

    sns.histplot(df[columna], bins=30, kde=True)  # kde  estimación de la densidad

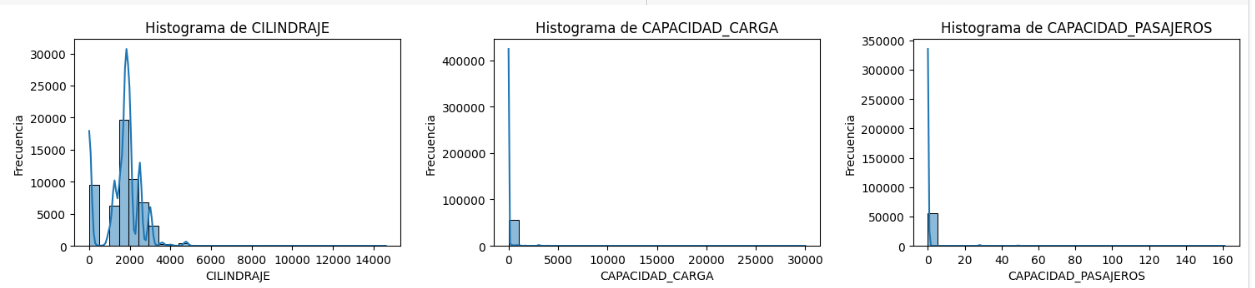
    plt.title(f'Histograma de {columna}')

    plt.xlabel(columna)

    plt.ylabel('Frecuencia')

plt.tight\_layout()

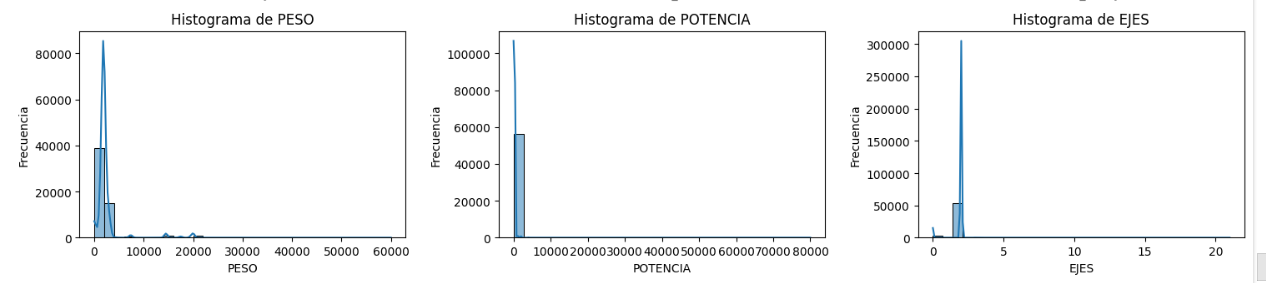
plt.show()



**La variable CILINDRAJE** demuestra que los carros pequeños que son la mayoria tienen un cilindraje bajo alrededor de 1000 y 2000 cm³, también tenemos valores cercanos a cero pueden ser datos cero o nulos, Los valores atípicos en la derecha pueden ser vehículos con cilindrajes más grandes

**La variable CAPACIDAD\_CARGA** muestra que la mayoría de los vehículos tienen una capacidad de carga baja pero también con valores cercanos a cero esto indica que son automóviles particulares no actos para carga en cambio los valores a la derecha indican camiones u otros especializados en cargar

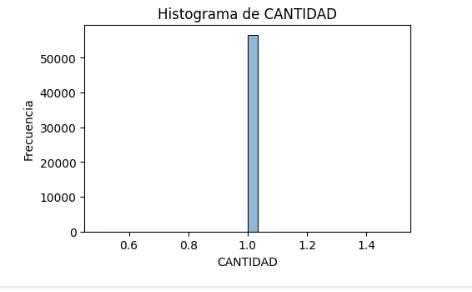
**La variable CAPACIDAD DE PASAJEROS** Al igual que los otros histogramas se destacan los carros de capacidad de pasajeros pequeña de 2 a 5 y estos datos tienden a datos cerca de cero lo cual serian errores o atípicos, los otros tipos de vehículos como autobuses y vehículos grandes no son frecuentes



**La variable PESO** tiene una distribución muy sesgada con valores bajos, esto indica que la mayoría de los vehículos son livianos o pequeños pueden ser la mayoría particulares, y los demás tienen un peso alto pueden ser los camiones

**La variable POTENCIA** distribución sesgada a la baja en este caso muestra la mayoría de los vehículos pequeños con potencias bajas y los camiones con altas

**La variable EJES** la mayoría de los vehículos son de pocos ejes y los demás como camiones más ejes



\*Esta columna no aporta variabilidad alguna y parece no ser útil para el análisis

1. **S20 - Evidencia 3 de Avance - Proyecto Integrado III**

**(Evidencia avance 3)**

**Proceso para identificar los datos relevantes y contestar las hipótesis planteadas**

**Link Video Explicativo:**

https://drive.google.com/file/d/1btBkdkqXvX4iF8G3JZgscR-RBN1U\_\_hr/view?usp=sharing

Después del filtrado se identifican los datos relevantes para contestar las hipótesis planteadas visualizar las métricas

**Hipótesis 1:**

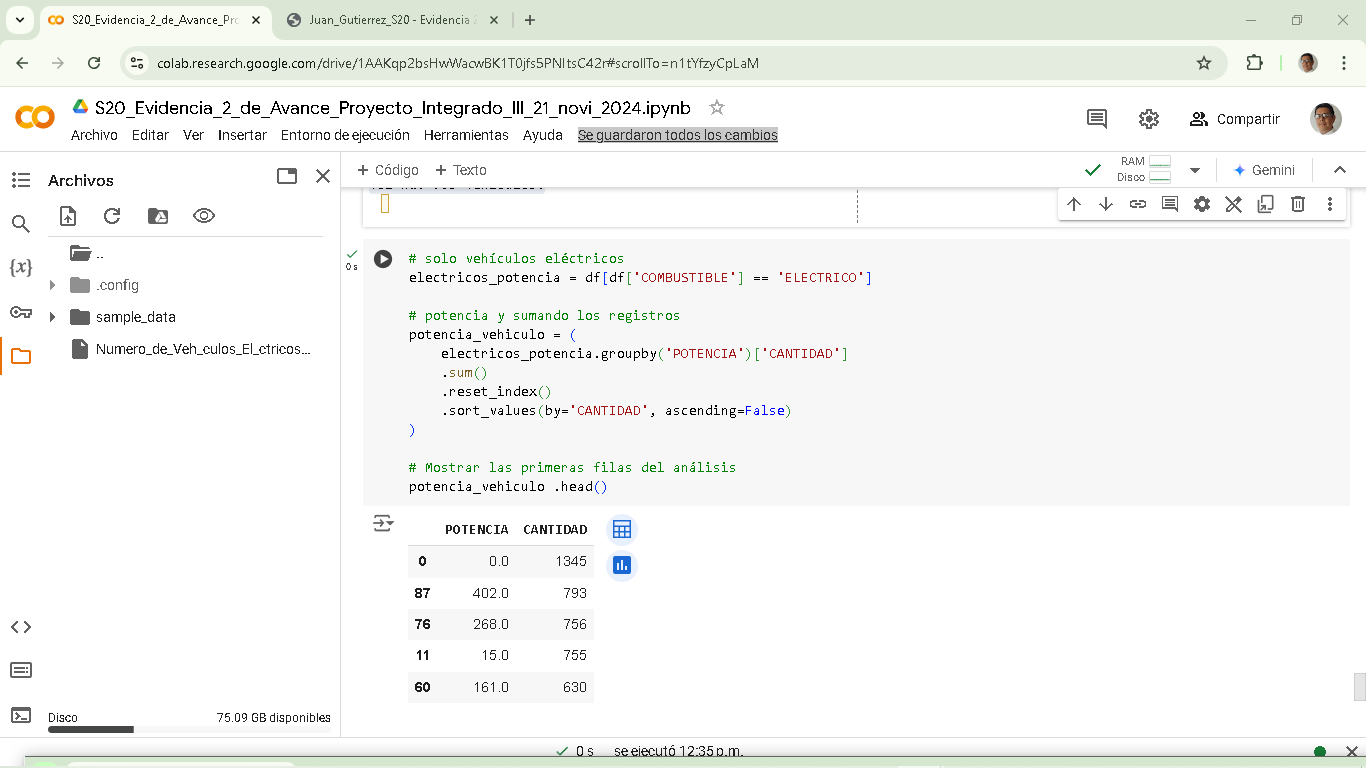
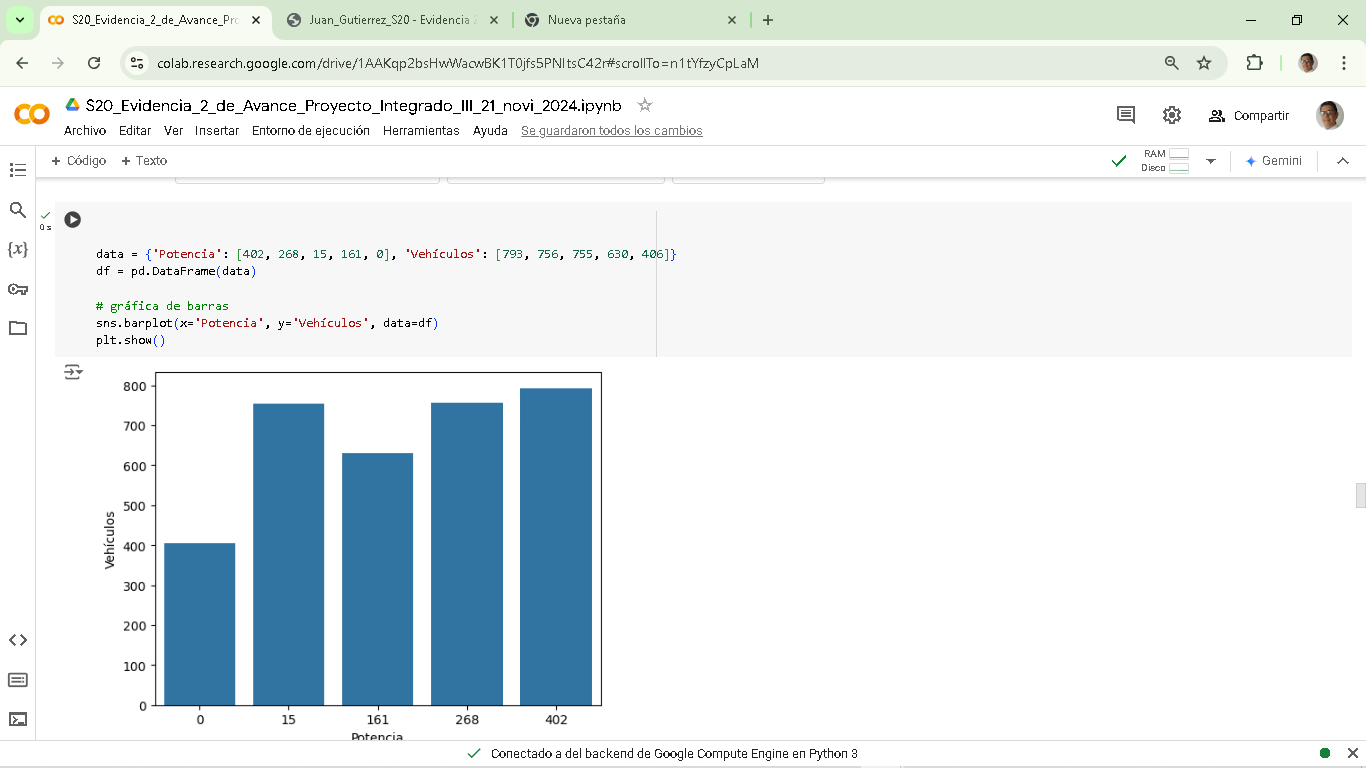
***Los automóviles eléctricos de más potencia son los que más se venden en Colombia***

Para esto, nos centramos en los vehículos solo eléctricos (COMBUSTIBLE = ELECTRICO) para tomar cuantos registros de venta según la potencia se venden en Colombia

El resultado de 1345 datos con 0 son resultados de los datos vacíos reemplazados por este 0.

La hipótesis resulta ser cierta por que la cantidad de vehículos solo eléctricos registrados muestra que el consumidor prefiere los carros de más potencia.

**402 kW: 793 vehículos.**



**Hipótesis 2:**

***Las marcas chinas son las más vendidas en el mercado colombiano***

Esta hipótesis es **parcialmente cierta** no la puedo responder con detalle porque la hipótesis no apunta a una marca china en específico, los datos que tengo no me permiten tener un Variable que de la descripción del ORIGEN DE FABRICACION del vehículo.

El único Resultado que puedo sacar con los datos, es el número de marcas de vehículos eléctricos e híbridos que se venden en nuestro país.

**Número de marcas únicas: 182**

Como el diagrama se presenta muy saturado al visualizar, la consulta de interpretar con un detalle de las 10 marcas más vendidas como parte de la presunción de venta de carros chinos que asumen el liderazgo

BYD 2506,

RENAULT 1409,

BMW 910,

ENERGY MOTION 461.

DONGFENG 410,

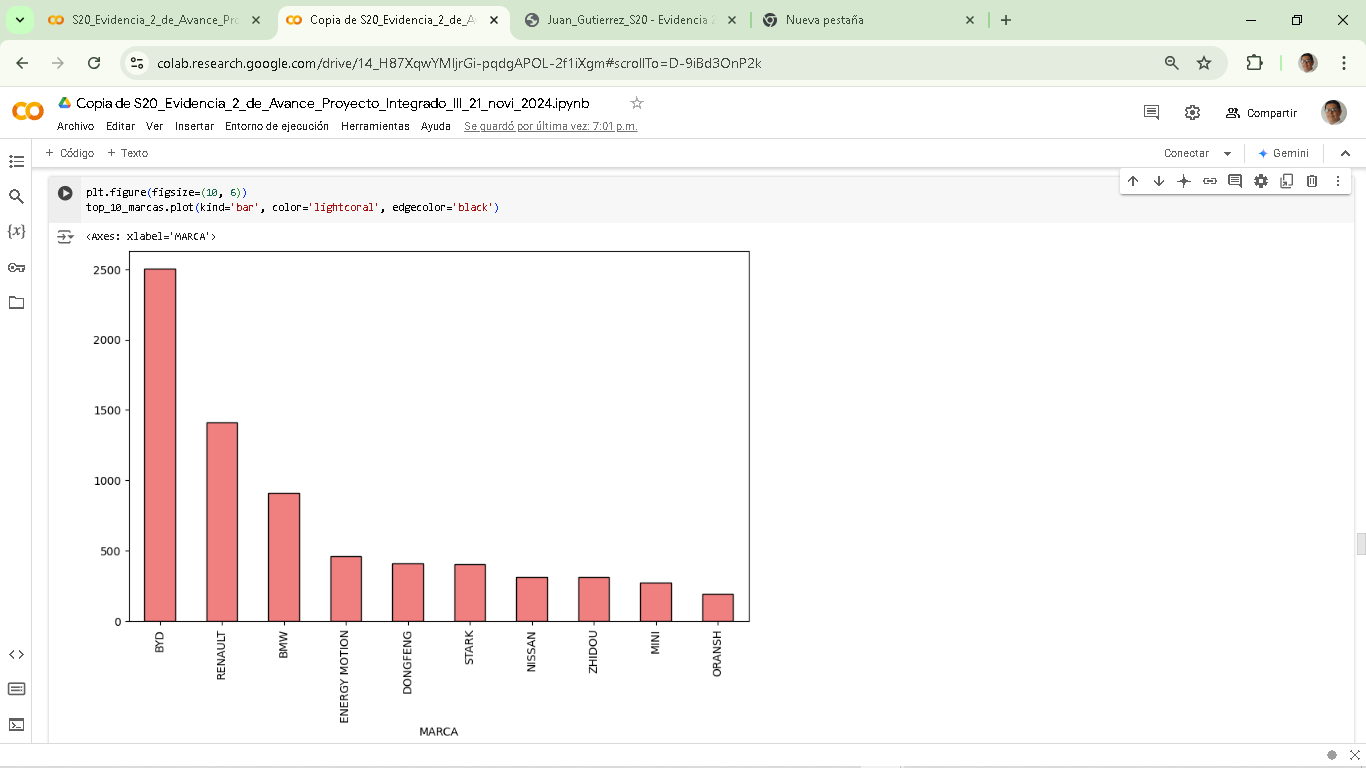
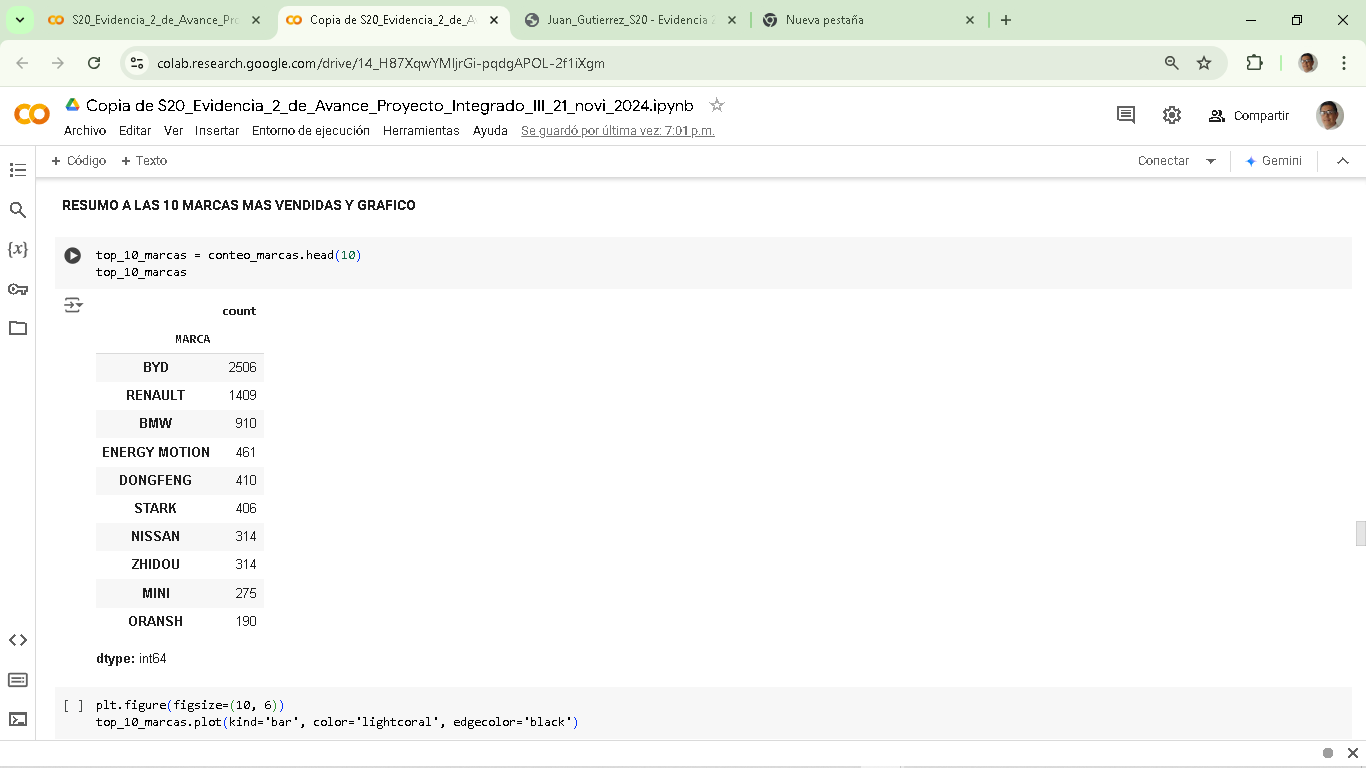
STARK 406,

NISSAN 314,

ZHIDOU 314,

MINI 275

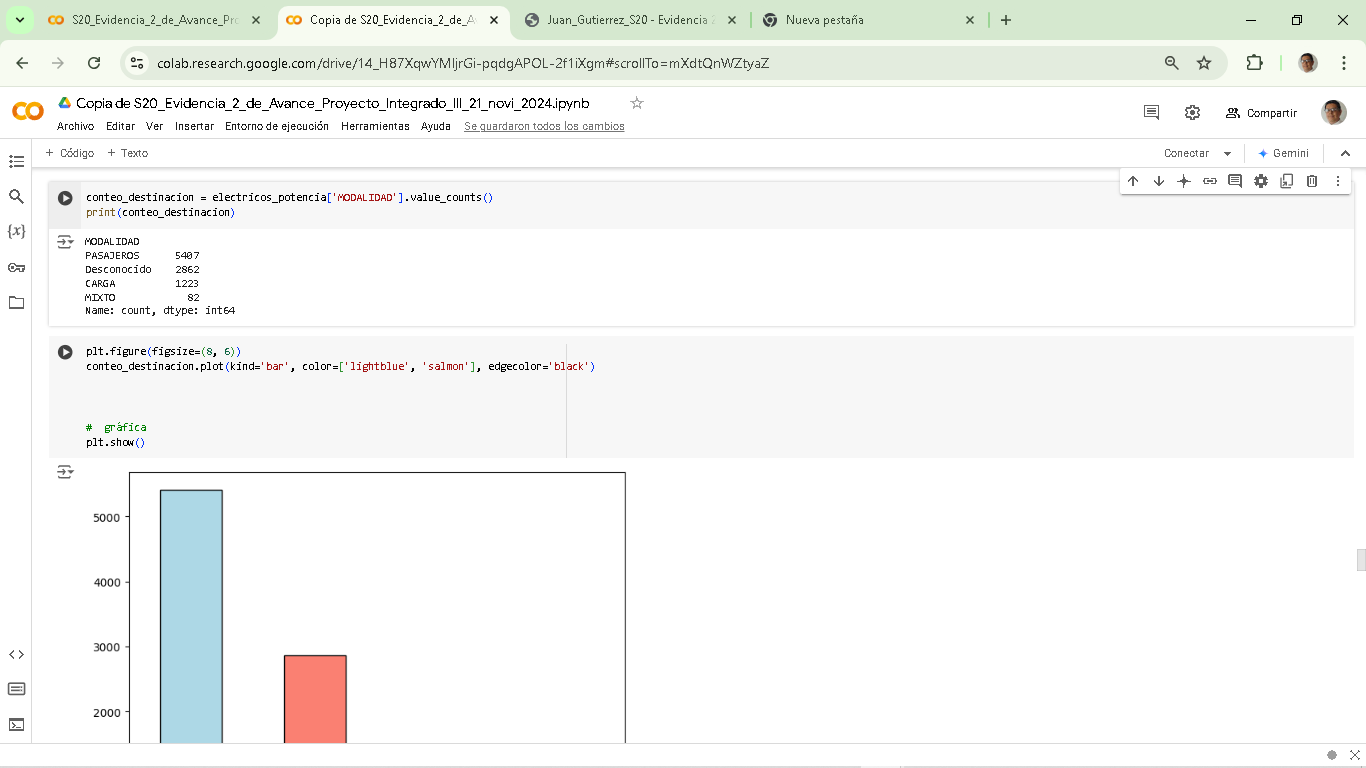
ORANSH 190

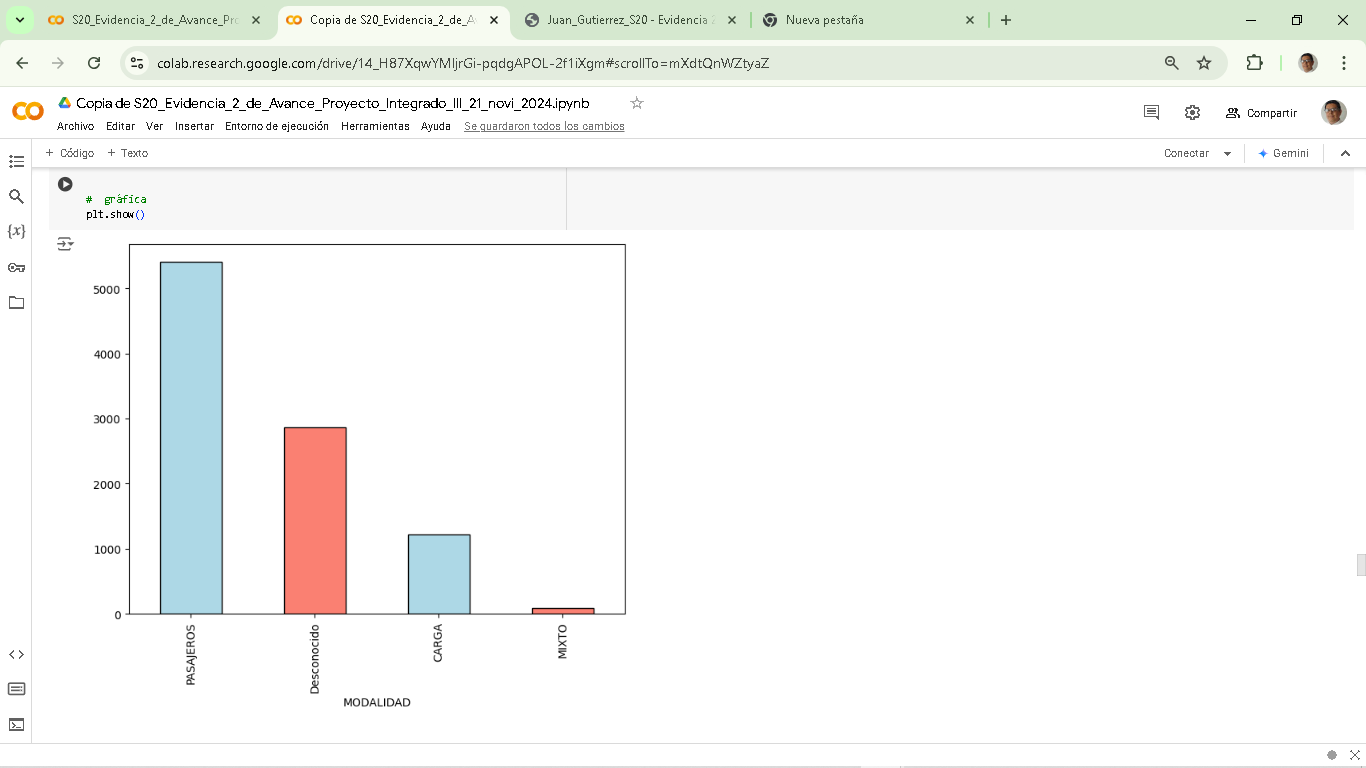


**Hipótesis 3:**

**El uso *de los vehículos eléctricos comprados en Colombia, la mayoría se destinan para pasajeros, y la minoría para carga***

La hipótesis resulta ser cierta según los datos filtrados el servicio mayoritario es para pasajeros con 5407 unidades registradas y los de carga son 1223 registros, también se obtuvo datos de vehículos de uso mixto y otros que no especificaron





**Hipótesis 4:**

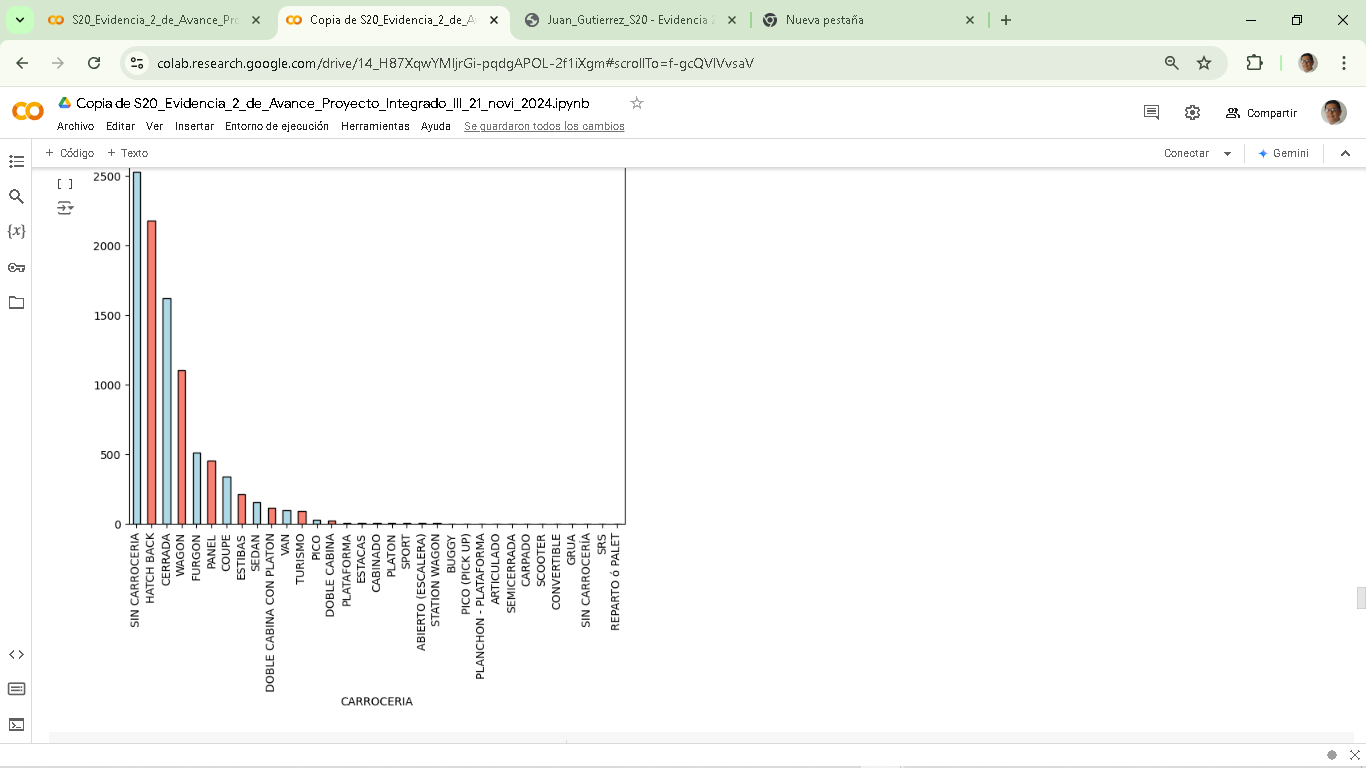
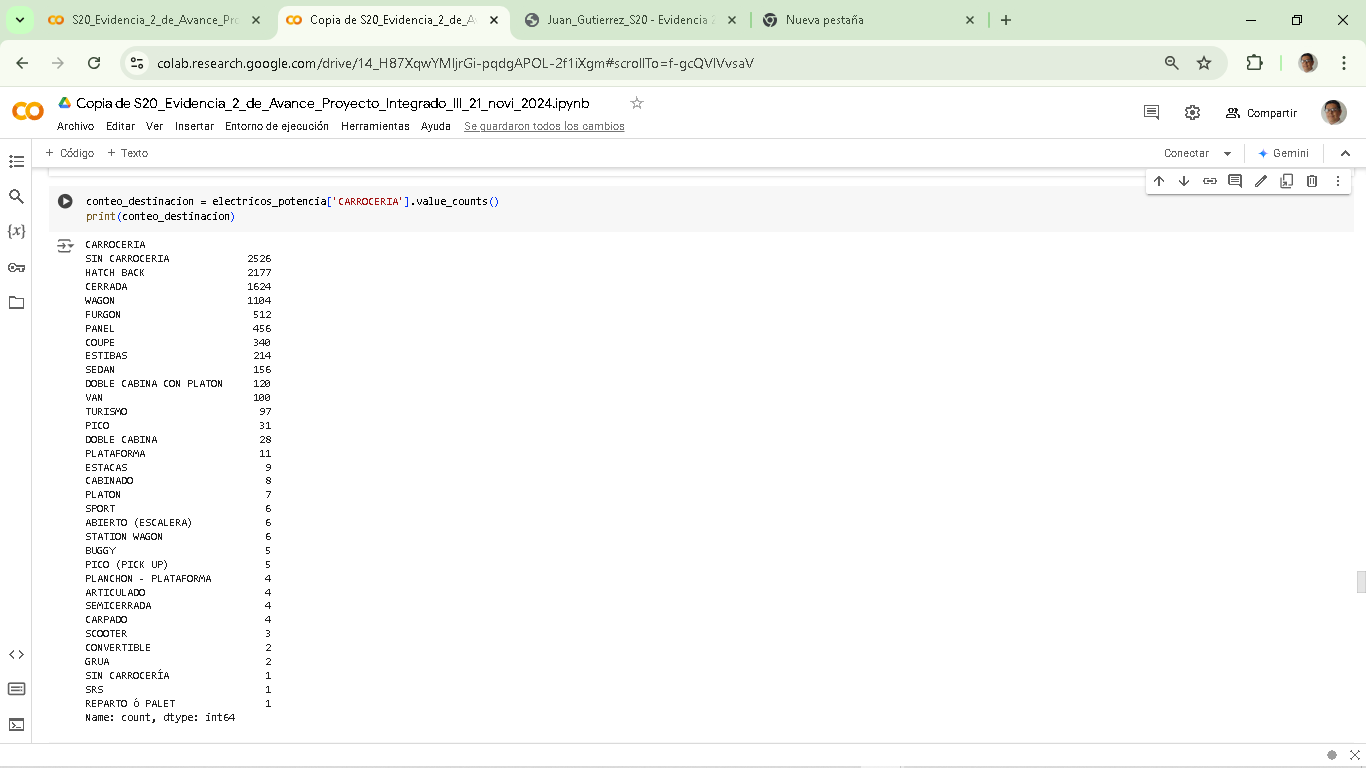
***El vehículo eléctrico preferido por el consumidor colombiano son de tipo SEDAN***

Esta hipótesis **no es correcta** el vehículo más vendido o de preferencia por los clientes es de tipo

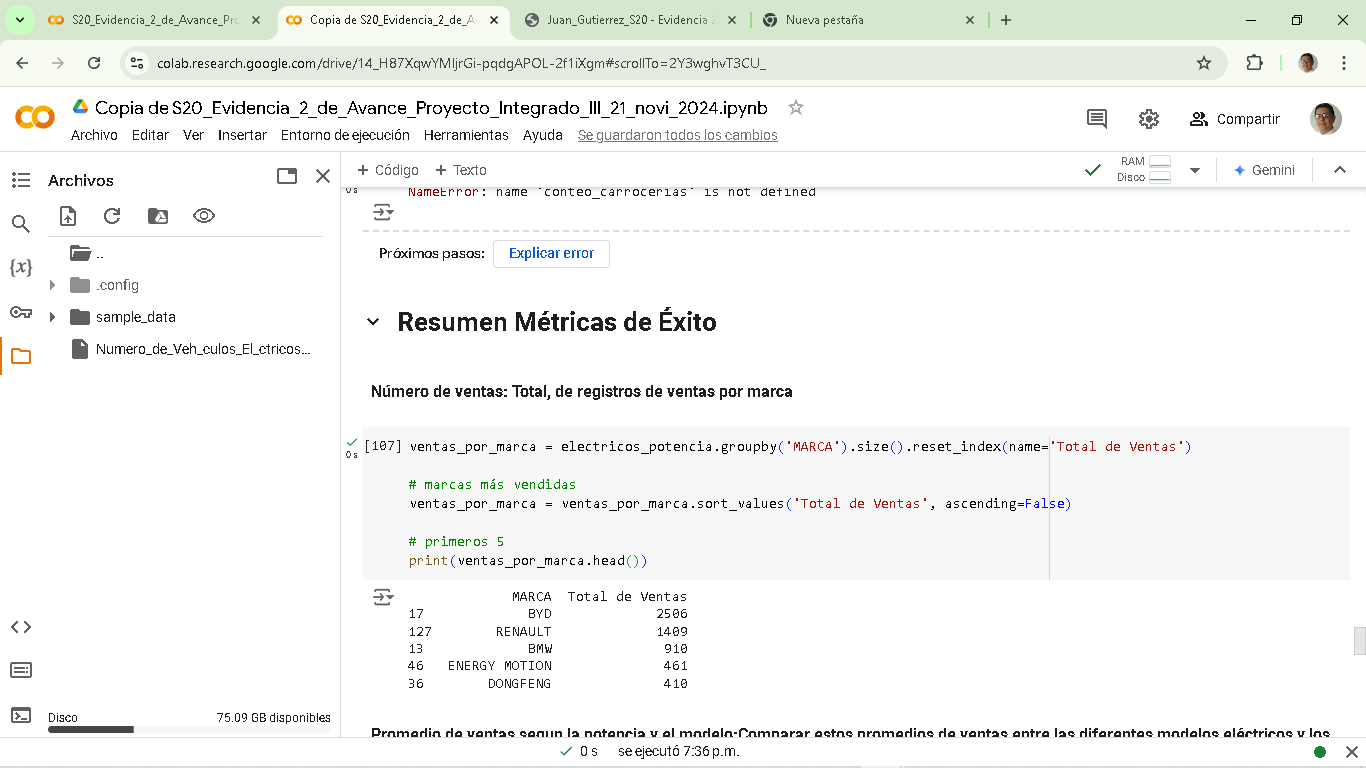
**HATCH BACK ------------- 2177**

seguido de la carrocería **CERRADA ------------- 1624** la cual presumo que es un carro de 2 puertas pequeño

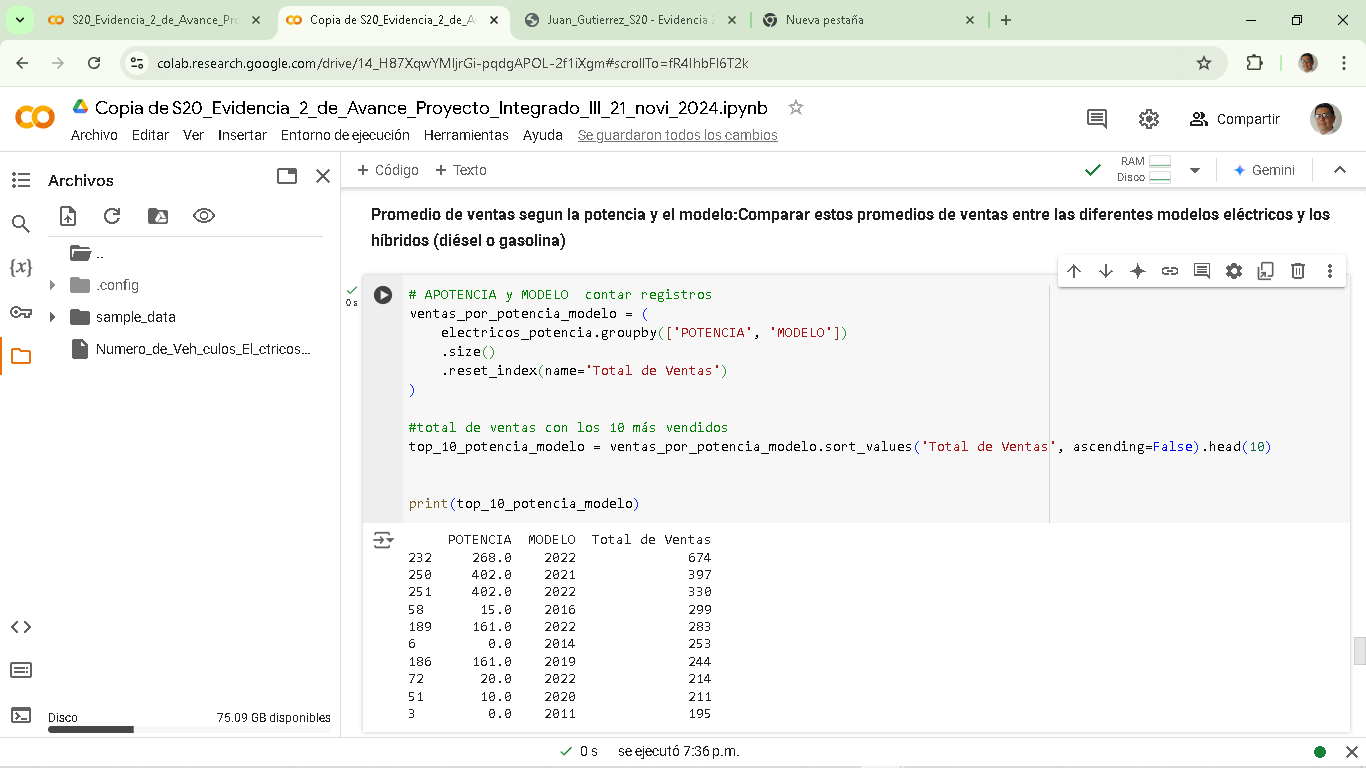
mientras los sedan están en **SEDAN-------------------156**



**Métricas de Éxito**

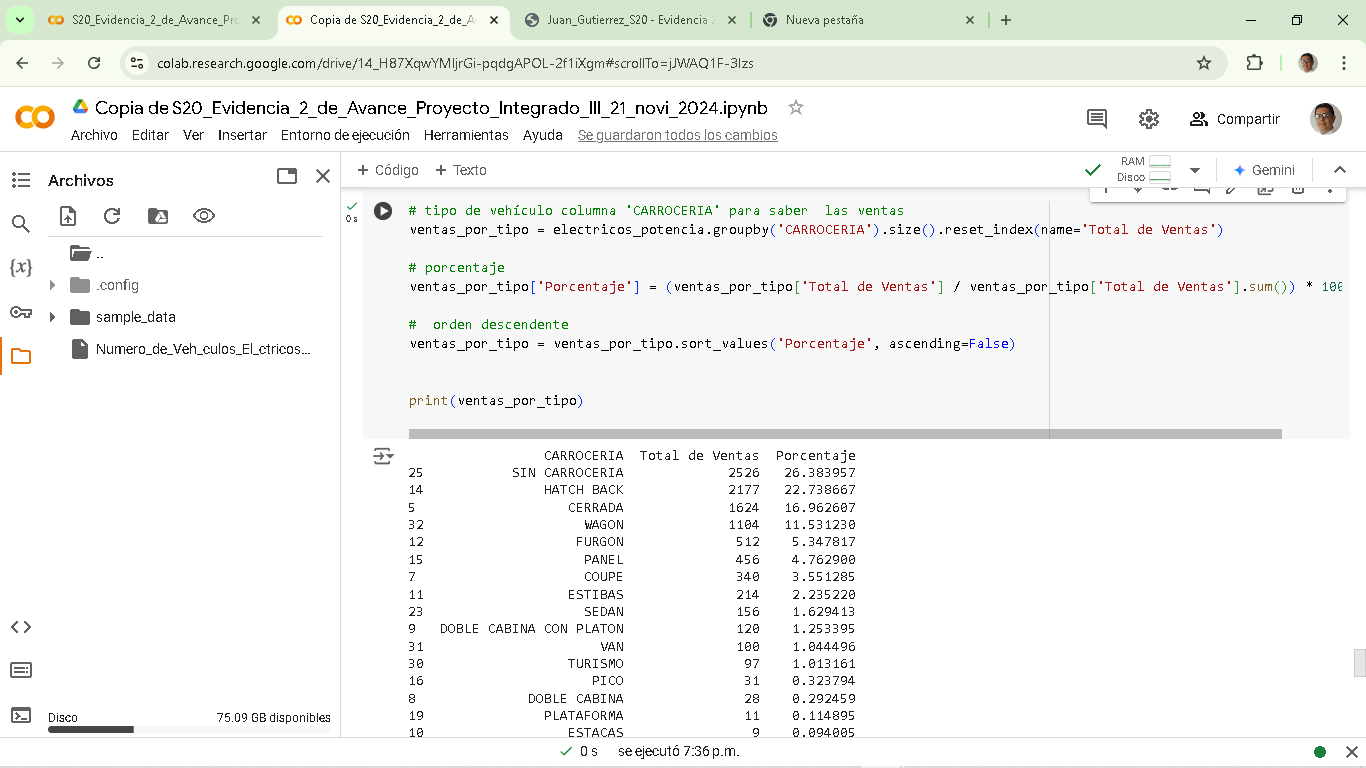
**Número de ventas:** Total, de registros de ventas por marca

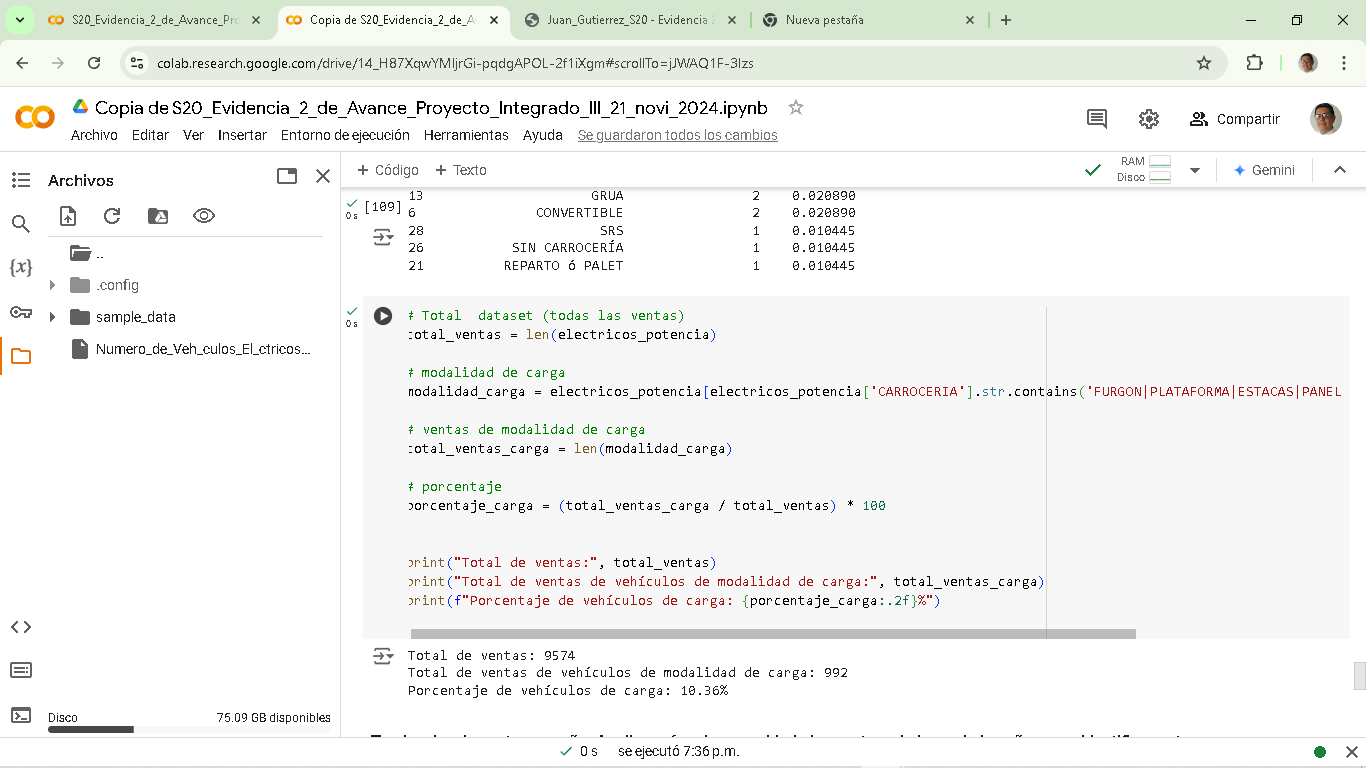
**Promedio de cilindraje y potencia:**

Comparar estos promedios de ventas entre las diferentes modelos eléctricos y los híbridos (diésel o gasolina)

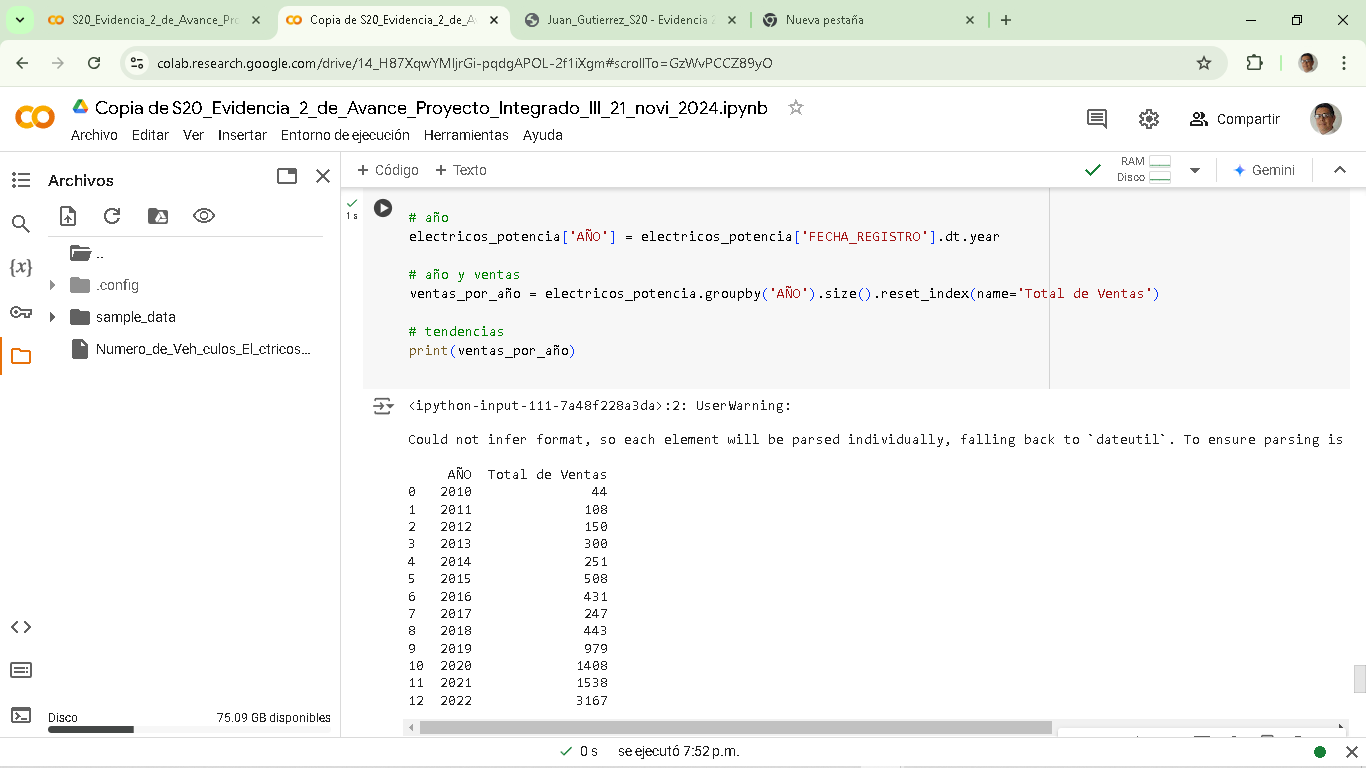
**Distribución por servicio:** Análisis del porcentaje de ventas por tipo de vehículo (sedan, carga)

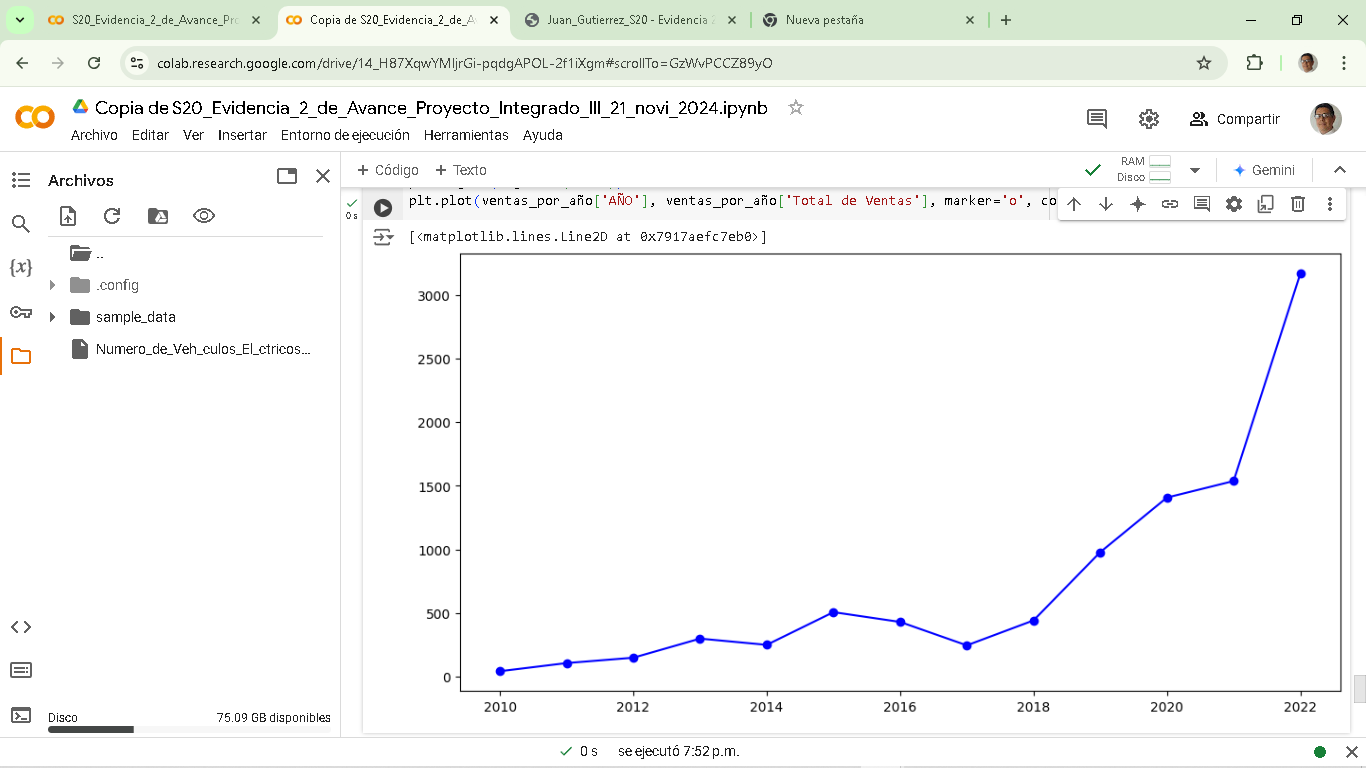
se encontró que:

SEDAN --- 156 ---- **1**.629413 % y el Total, de ventas de vehículos de modalidad de carga: 992 Porcentaje de vehículos de carga: **10**.36%

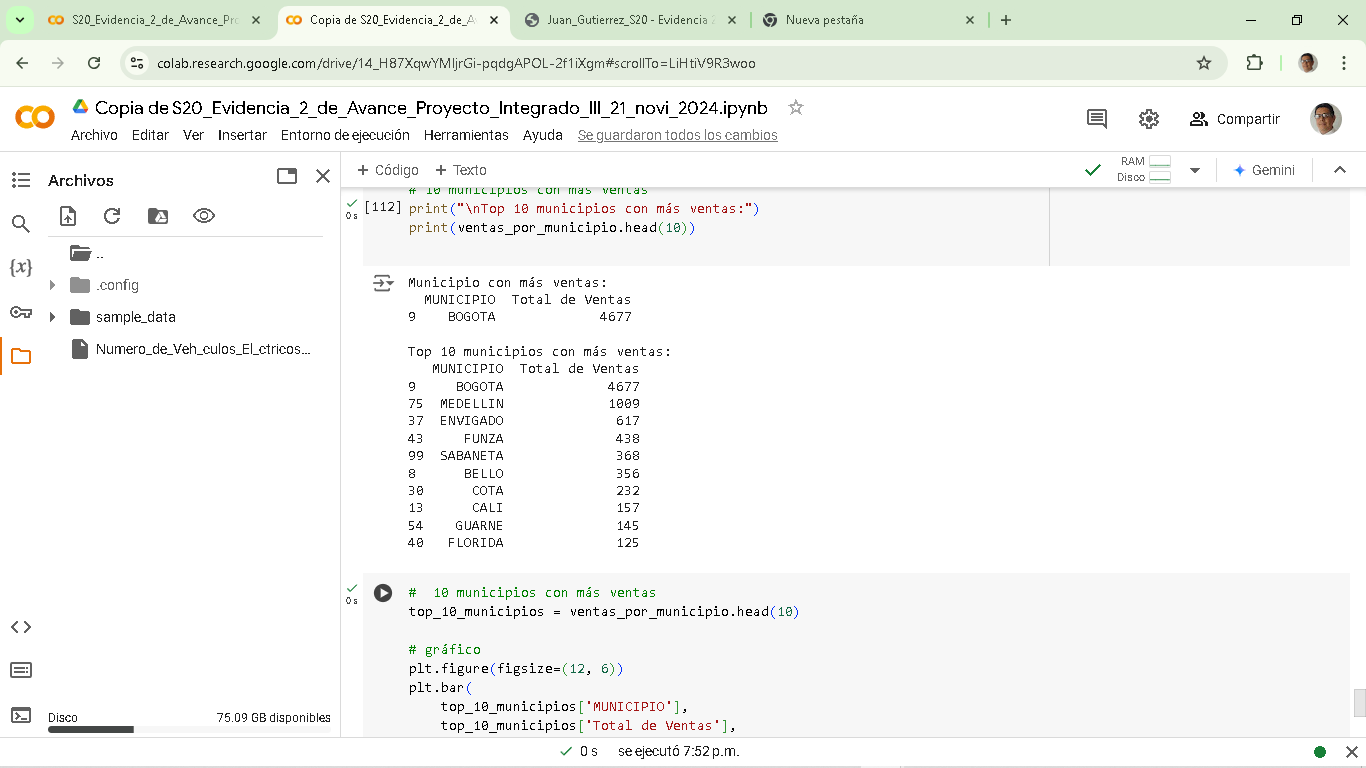


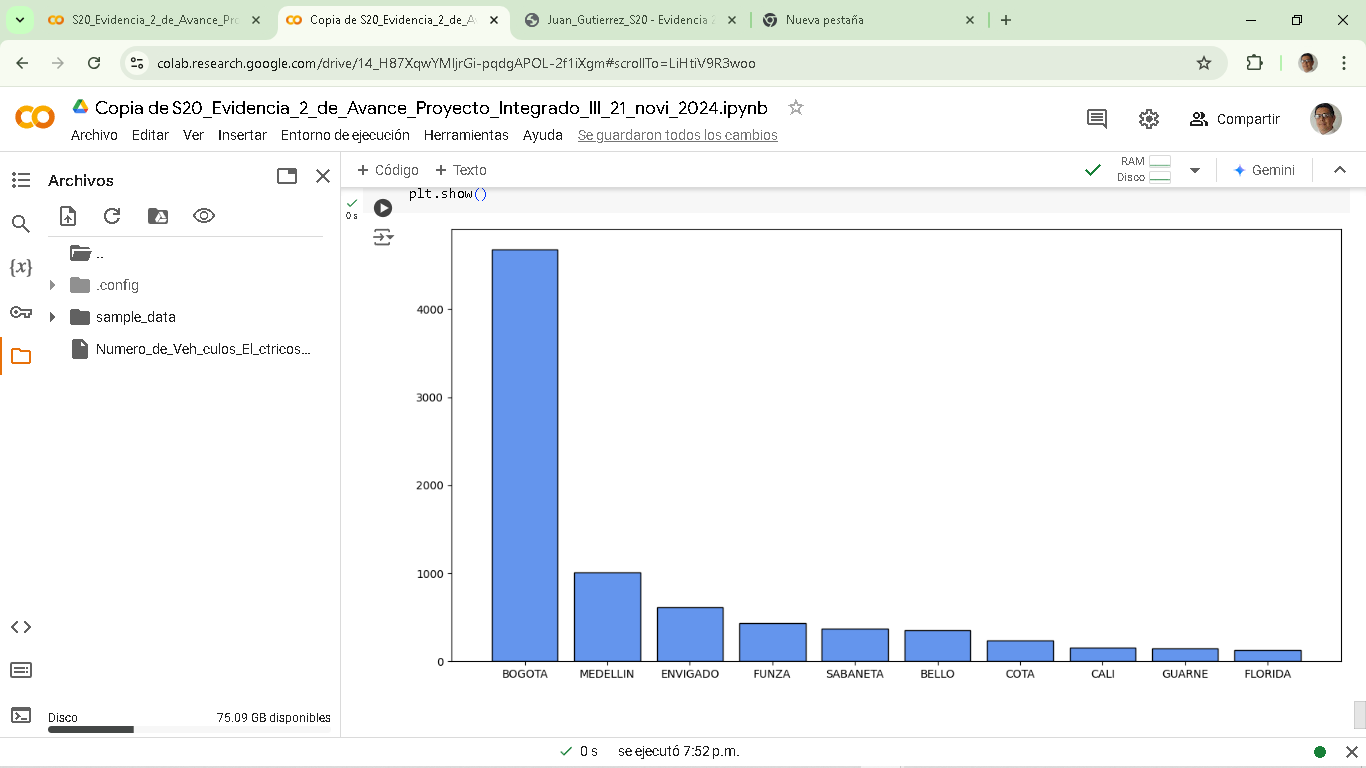
**Tendencias de ventas por año:**  Analizar cómo han cambiado las ventas a lo largo de los años para identificar patrones.





**Análisis por municipio o departamento:** Identificar en qué municipios se realizan más ventas y si hay diferencias significativas



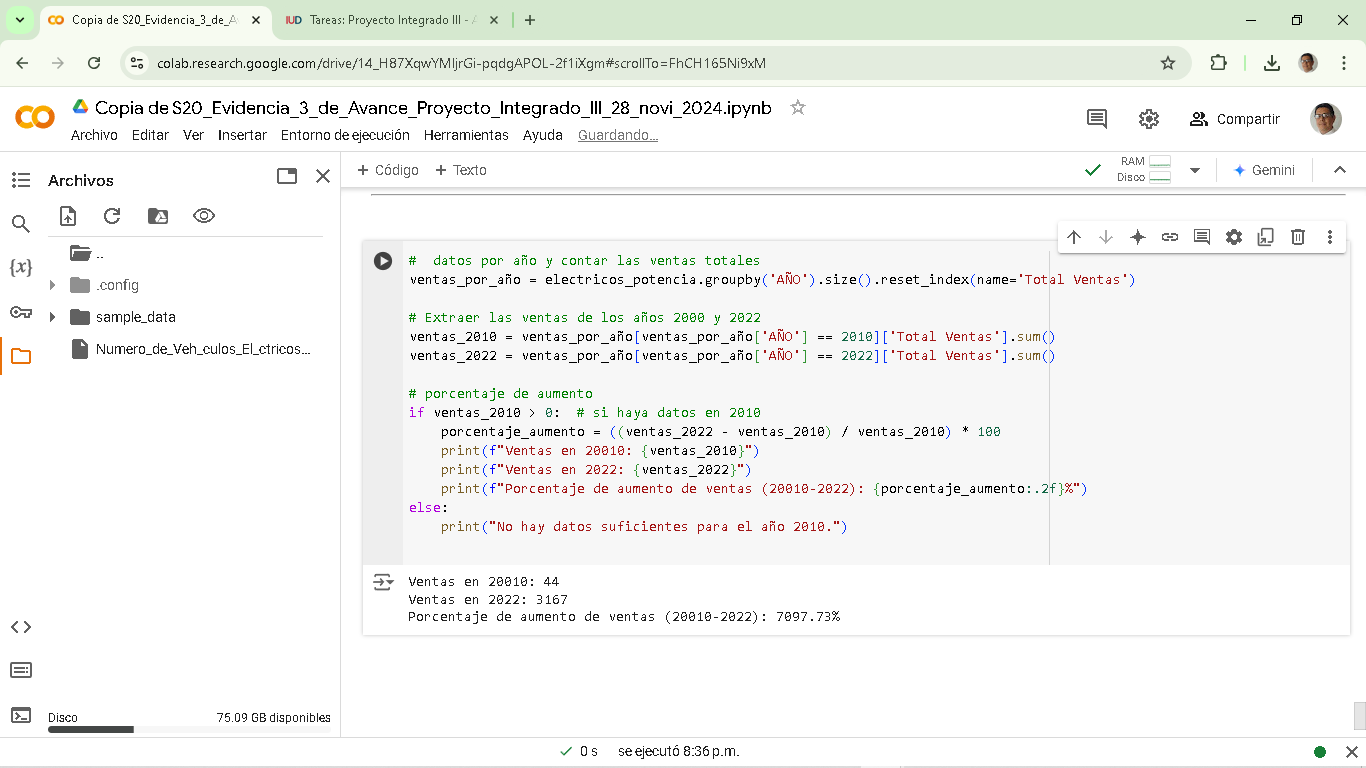


**¿Qué factores influyen en el aumento de las ventas de automóviles eléctricos en Colombia?**

Ventas en porcentaje entre los años 2010 y 2022

presenta un aumento muy alto

Ventas en 20010: 44

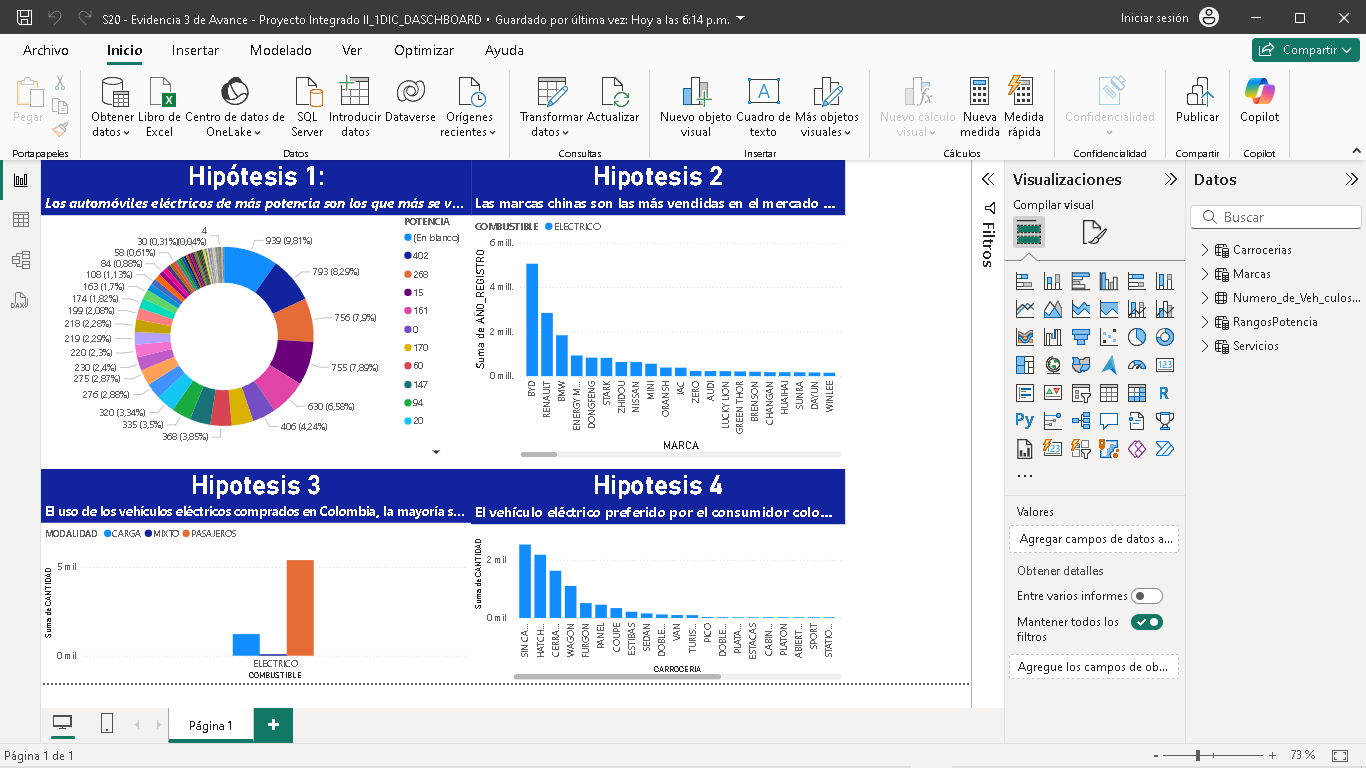


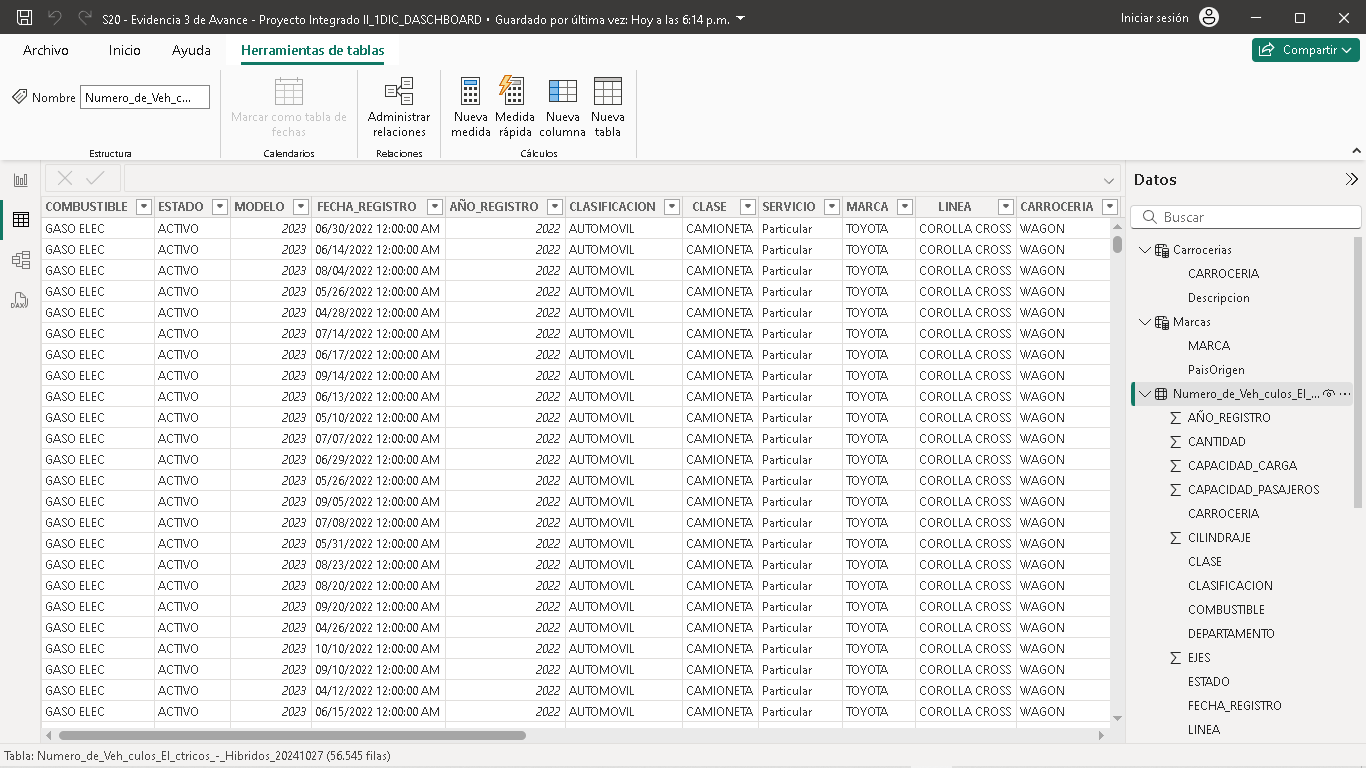
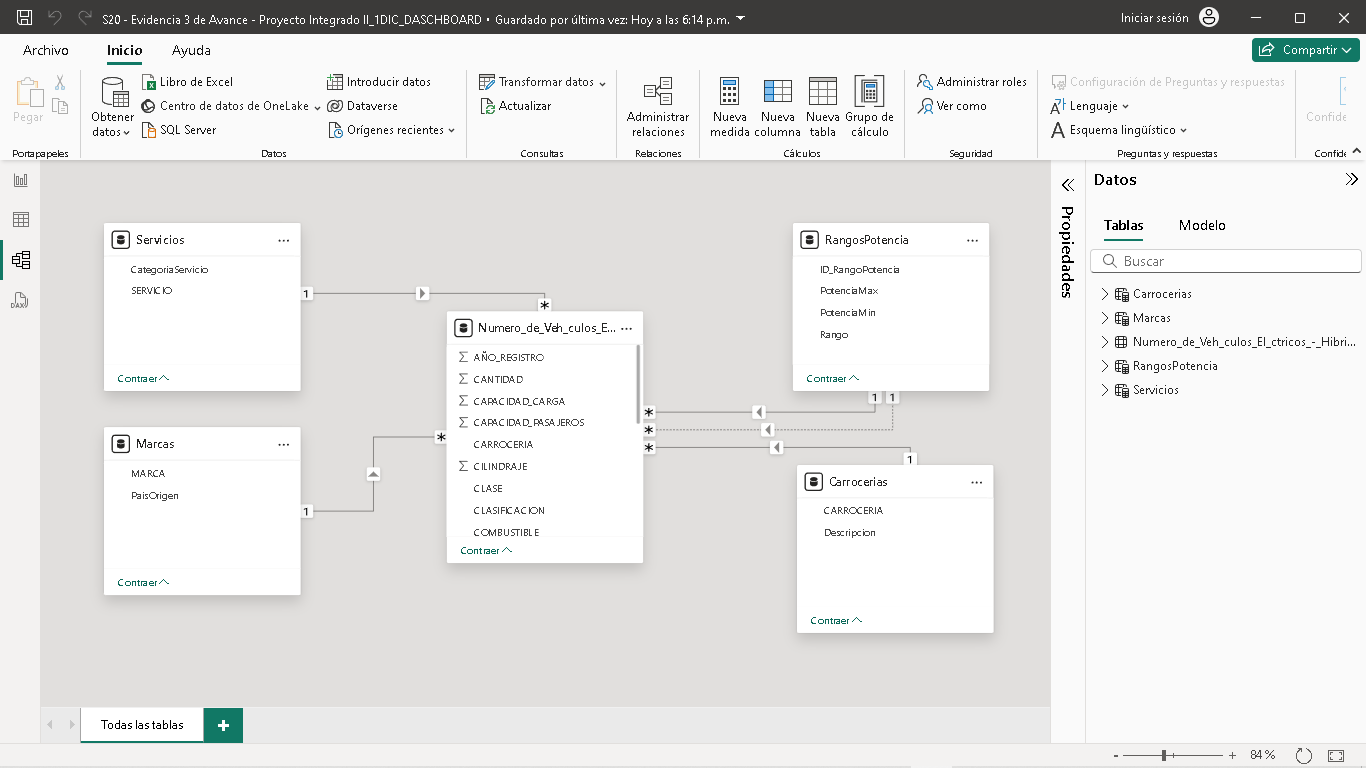
Ventas en 2022: 3167

Porcentaje de aumento de ventas (20010-2022): 7097.73%

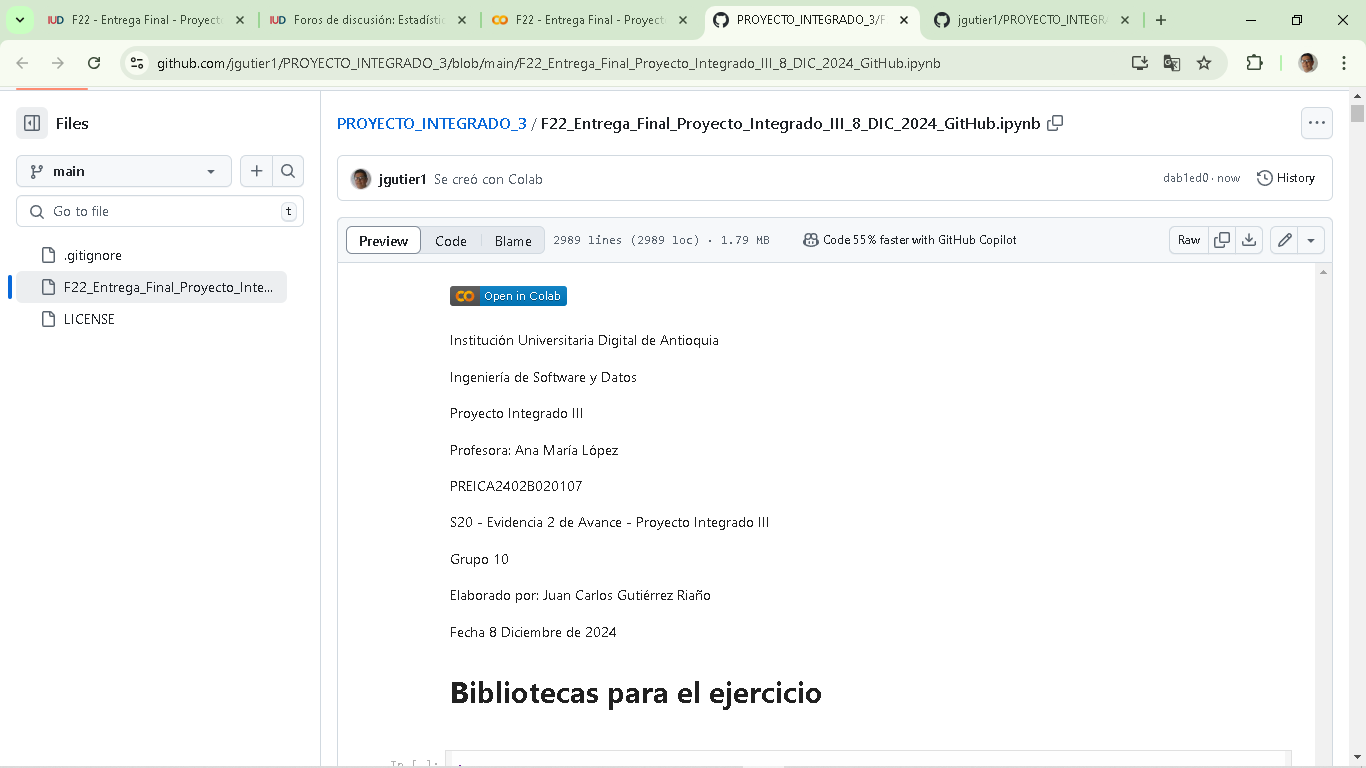
Aunque se presentan varios factores que puedan afectar las ventas de vehículos eléctricos en Colombia estos no determinan el gran auge de estos entre estos años, y a pesar de que el costo no era una variable que teníamos en la base de datos se presume un porcentaje de impedimento era el precio el cual en Colombia no ha representado un gran impedimento en ventas la proyección para los próximos años será mayor de acuerdo a este análisis

1. **Dashboard**





1. **Creación de repositorio - Inicio de nuestro portafolio**



1. **Link video explicativo final**
2. **Conclusión**

El aumento de las ventas de vehículos eléctricos en Colombia tiene un índice al alta si los factores económicos, ambientales y tecnológicos mejoran a la vez las políticas gubernamentales y de privados se optimizan se puede mantener el crecimiento, es necesario mejorar la infraestructura, costos iniciales y educación de los usuarios

En este ejercicio se identificaron y analizaron valores atípicos en varias variables (ejes, peso, potencia, cilindraje, capacidad de carga y capacidad de pasajeros)

observe que la mayoría de las distribuciones están sesgadas hacia valores bajos, lo que refleja la predominancia de vehículos con características estándar o pequeños

Se identificaron valores muy altos que no corresponden a la lógica del análisis o serán vehículos especializados, Finalmente, la limpieza me permitió garantizar que los datos sean consistentes para su posterior análisis

De acuerdo a la anterior en la investigación pueden determinar que tres de las hipótesis 1,3 y 4 se pueden contestar como favorables a la pregunta inicial, mientras que la 2 es aún para un estudio Posterior, pues se determina que la procedencia de fabricación de un vehículo eléctrico es del origen según el nombre de la marca pero indagando más a fondo se presume que empresas occidentales muy conocidas venden estos vehículos con nombres muy parecidos a los chinos , entonces los da datos del Dataset puede ser desactualizados por los dos años de diferencia a la investigación, se presume que la mayoría de carros que se venden en Colombia son chinos por el liderazgo de la empresa BYD.