

Reporte Fase 1: Descubrimiento del proyecto.

1. Título del proyecto.

Nombre de proyecto: “**Análisis de valoración de vehículos para el mercado automotriz**”.

2. Objetivo del proyecto.

Definición del objetivo del análisis desde un punto de vista de negocio.

El objetivo principal de este análisis de datos sobre venta de vehículos es aprovechar la información disponible para mejorar la estrategia comercial de cada fabricante de automóviles, con el propósito de tener una planeación en la toma de decisiones estratégicas dentro del mercado automotriz, a fin de incrementar ventas, mejorar la competitividad y fortalecer el posicionamiento de cada fabricante en un lapso de 12 meses.

Con este proyecto busco:

1. Definir estrategias que permita el posicionamiento de cada fabricante.

Analizar diferentes factores de los cuales incluye: marca, modelo, año, tipo de combustible o si es híbrido, tamaño de motor, de los cuales influye un precio de venta para el consumidor.

Tener claro una estrategia de precios que permita ser competitivo, maximice sus ventas y maximice un margen de ganancias para cada segmento.

2. Detectar tendencias de compra y preferencias de los consumidores.

Se busca evaluar que fabricantes y modelos de automóviles tienen mayor rentabilidad dentro del mercado, identificar las preferencias de los consumidores en base al tipo de combustible (gasolina o diésel) o si prefiere los autos híbridos y anticipar la evolución de la demanda de cada uno.

Reconocer patrones de compra, por ejemplo, cuales son los meses del año en donde el consumidor llega a comprar un automóvil, con el objetivo de planificar lanzamiento y campañas de marketing.

3. Impulsar decisiones estratégicas a través de datos.

Comparar el desempeño de ventas entre los distintos fabricantes y sus diferentes modelos, del mismo modo, detectar cuales las fortalezas y debilidades de cada fabricante, como puede ser el precio, motor, tipo de

combustible, y apoyar en la definición de estrategias de diferenciación para cada fabricante.

3. Descripción del problema.

Desde una perspectiva de negocios, en la actualidad los consumidores no tienen una idea clara sobre qué automóvil comprar, si comprar un vehículo económico, combustible tradicional (gasolina o diésel), incluso dudan si es buena opción comprar un vehículo híbrido o incluso eléctrico. Tengamos en mente que la industria automotriz atraviesa uno de los cambios más significantes en las últimas décadas dentro de lo que respecta, la transición energética y tecnológica hacia vehículos híbridos y eléctricos, los fabricantes enfrentan grandes retos para adaptar su oferta y demanda, mantener su rentabilidad y responder a los cambios en las preferencias del consumidor actual.

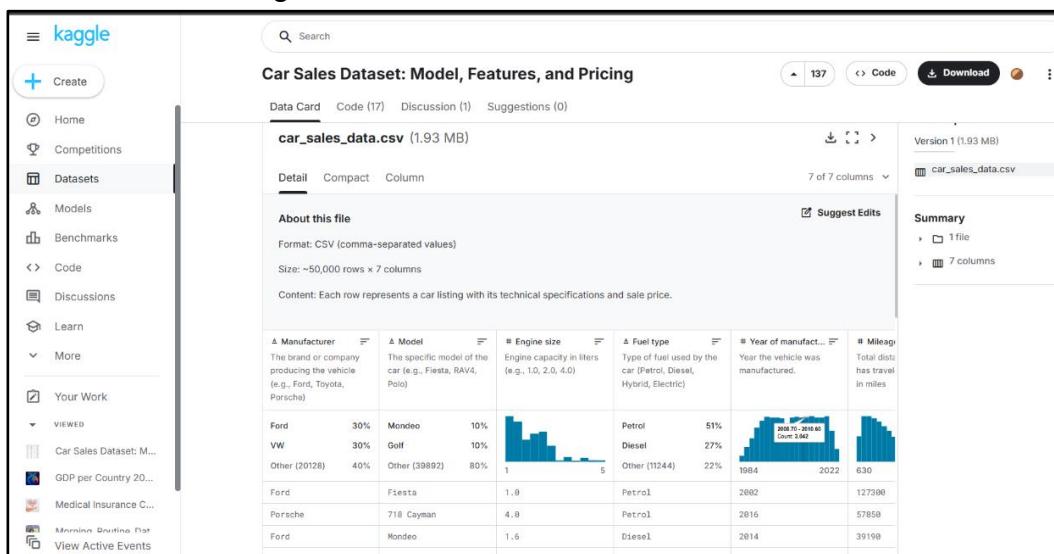
La incertidumbre recae principalmente en la compresión de qué tan rápido el consumidor dejarán de comprar vehículos de combustible tradicional (gasolina o diésel) y hacia qué segmentos recae la demanda (híbridos o eléctricos). Este cambio depende de varios factores, por ejemplo, el factor del medio ambiente persiste al momento de adquirir un vehículo, también factores de regulaciones, sin dejar de lado variables como lo son: precio de salida y el poder adquisitivo de cada consumidor.

En conclusión, el problema recae en cómo los fabricantes del sector automotriz pueden adaptarse a las nuevas preferencias del consumidor el cual puede ser abordado eficazmente con este análisis de datos. Este análisis no solo permite reducir la incertidumbre y anticipar tendencias a largo plazo, también brinda la competitividad en el mercado automotriz.

4. Recursos disponibles.

- **Tecnologías y herramientas:**

Para realizar el análisis de datos, primero se obtuvo un dataset o conjunto de datos de la plataforma Kaggle con el nombre: Car Sales Dataset, Model, Features, and Pricing.

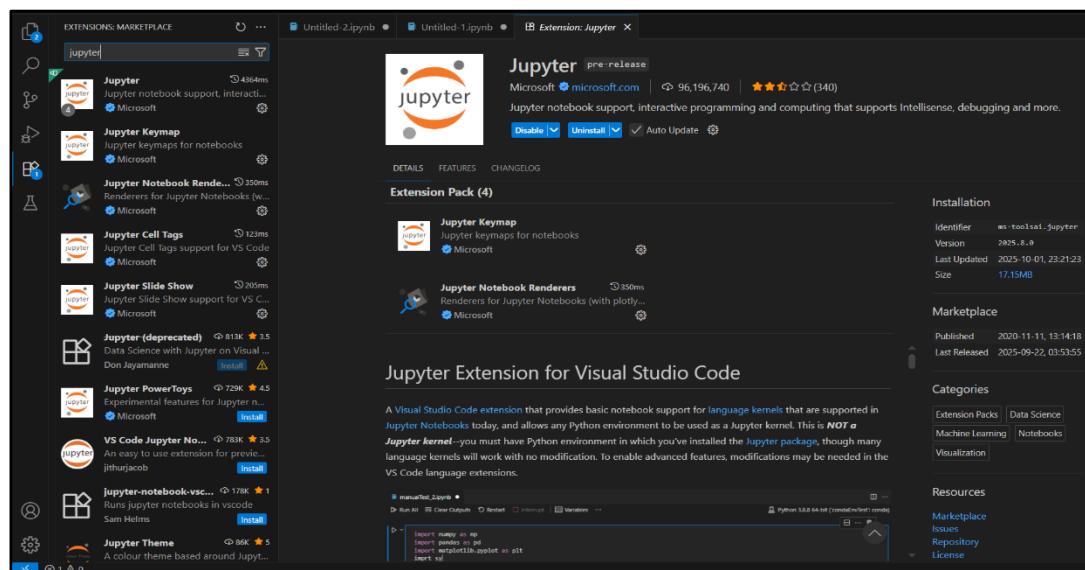


Ahora bien, para analizar el conjunto de datos se utiliza como herramienta principal Visual Studio Core, con las respectivas extensiones de Python y Jupyter Notebook.

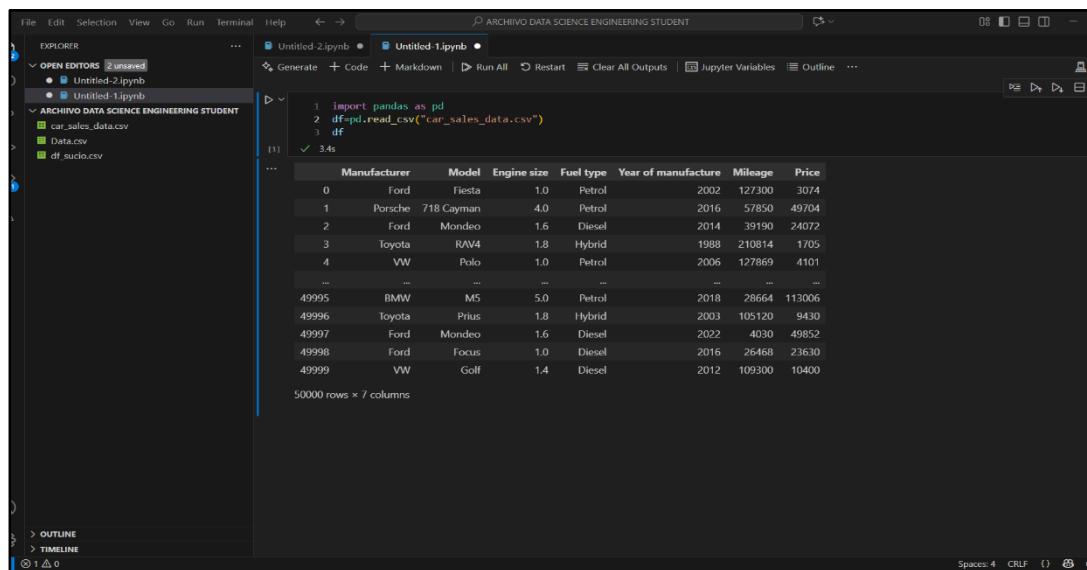
La extensión de Python dentro Visual Studio Code me sirve para facilitar y mejorar la escritura, ejecución y depuración de código Python, porque permite realizar tareas como autocompletado de código, resaltado de sintaxis, detección de errores, formateo de código, depuración paso a paso, haciendo que el desarrollo sea más rápido y eficiente dentro de este análisis de datos.



Ahora bien, la extensión Jupyter Notebook me permite crear, ejecutar y compartir documentos interactivos que integran código, texto explicativo, visualizaciones y ecuaciones, facilitando el análisis de datos.



Partiendo de las extensiones, para realizar un correcto análisis de los datos, se utiliza la extensión Pandas, el cual es una biblioteca de Python para la manipulación y el análisis de datos, ofreciendo estructuras de datos como los DataFrame (df), permiten trabajar con datos de manera eficiente, por ejemplo, me sirve para importar, limpiar, transformar, agregar y analizar conjuntos de datos desde diversas fuentes, como archivos CSV o bases de datos, facilitando el desarrollo del presente proyecto.



The screenshot shows a Jupyter Notebook environment. In the top bar, there are tabs for 'File', 'Edit', 'Selection', 'View', 'Go', 'Run', 'Terminal', 'Help', and 'Untitled-1.ipynb'. Below the tabs, there's a toolbar with icons for 'Generate', 'Code', 'Markdown', 'Run All', 'Restart', 'Clear All Outputs', 'Jupyter Variables', and 'Outline'. The left sidebar has sections for 'EXPLORER' (with 'OPEN EDITORS 2 unsaved' and 'Untitled-2.ipynb'), 'ARCHIVO DATA SCIENCE ENGINEERING STUDENT' (containing 'car_sales_data.csv', 'Data.csv', and 'df.socio.csv'), and 'OUTLINE' and 'TIMELINE' buttons. The main area contains a code cell with the following Python code:

```

1 import pandas as pd
2 df=pd.read_csv("car_sales_data.csv")
3 df

```

The code cell has a status of '3.4s' and a green checkmark. Below the code cell is a data preview table:

	Manufacturer	Model	Engine size	Fuel type	Year of manufacture	Mileage	Price
0	Ford	Fiesta	1.0	Petrol	2002	127300	3074
1	Porsche	718 Cayman	4.0	Petrol	2016	57850	49704
2	Ford	Mondeo	1.6	Diesel	2014	39190	24072
3	Toyota	RAV4	1.8	Hybrid	1988	210814	1705
4	VW	Polo	1.0	Petrol	2006	127869	4101
...
49995	BMW	M5	5.0	Petrol	2018	28664	113006
49996	Toyota	Prius	1.8	Hybrid	2003	105120	9430
49997	Ford	Mondeo	1.6	Diesel	2022	4030	49852
49998	Ford	Focus	1.0	Diesel	2016	26468	23630
49999	VW	Golf	1.4	Diesel	2012	109300	10400

Below the table, it says '50000 rows x 7 columns'.

- **Datos:**

Dentro de los datos disponibles por cada columna, tenemos 7 columnas con su respectivo dato que ayudaran durante este proceso de análisis de datos, de los cuales son:

- **1ra columna:** Tenemos los datos acerca de “**Fabricante**”, contiene el nombre de la marca o empresa responsable de la producción del vehículo, por ejemplo, BMW, Ford, Toyota, Volkswagen, Porsche, entre otros. Esta información es fundamental debido que sirve para analizar la participación de cada marca dentro del mercado automotriz, identificar qué fabricantes concentran un mayor número de ventas y analizar las preferencias de los consumidores determinada por cada marca. Además, me permite estudiar la relación entre la reputación del fabricante y variables como el precio promedio, durabilidad o el tipo de combustible más utilizado en sus vehículos.
- **2da columna:** Se tiene datos acerca de “**Modelo**”, se refiere a la denominación específica del vehículo dentro de la gama de cada fabricante, por ejemplo, RAV4 de Toyota, Z4 de BMW, Polo y Golf de Volkswagen, entre otros. Este conjunto de datos es importante para evaluar el comportamiento

de ventas a nivel de producto individual de cada fabricante, además, se pueden identificar los modelos más populares o vendidos de cada fabricante, presentan características que los hacen más atractivos para el consumidor, también es útil para estudiar la diversidad de productos de cada fabricante y su posicionamiento en el mercado automotriz.

- **3ra columna:** Dentro de esta columna de datos, tenemos “**Tamaño del motor**” el cual indica la cilindrada del motor expresada en litros, permite analizar el rendimiento, potencia y eficiencia de cada vehículo. Esta variable es relevante para relacionar las características técnicas del vehículo con factores ambientales, como el consumo de combustible, emisiones de CO₂ o la preferencia del comprador por el tipo de motor. Esta columna tiene un peso importante al estimar el nivel de desempeño del vehículo.
- **4ta columna:** Por su parte, esta columna lleva por nombre “**Tipo de combustible**” específica el tipo de energía que utiliza el vehículo, por ejemplo, gasolina o diésel, incluso si es híbrido. Estos datos resultan de gran valor para examinar las tendencias del mercado actual, incluso para entender el ritmo con que los consumidores adoptan tecnologías eléctricas o si siguen con los vehículos convencionales. Además, permite comparar precios y costos de mantenimiento entre vehículo con diferentes tipos de combustible.
- **5ta columna:** “**Año de fabricación**” muestra el año de fabricación del vehículo, esta variable posibilita calcular la antigüedad de los vehículos, el análisis de esta columna permite observar la evolución de las ventas a lo largo del tiempo, identificar los años en que se fabricaron los modelos más buscados o estudiar la relación entre el año de fabricación y kilometraje promedio. Ayuda a determinar cómo cambia la demanda de vehículos nuevos frente a los usados en distintos períodos.
- **6ta columna:** “**Kilometraje**” representa la distancia total recorrida por cada vehículo, expresada en millas, es unas variables más influyentes ya que un mayor kilometraje suele estar asociado con un mayor desgaste. El análisis de esta columna permite determinar identificar patrones de utilización según el tipo de combustible o comparar la durabilidad entre marcas y modelos.
- **7ta columna:** “**Precio**” la variable dependiente más importante dentro de un análisis de tipo económico o predictivo, la información sobre precios permite construir modelos de regresión o aprendizaje automático que estimen el valor de un automóvil a partir de sus características (marca, modelo, motor, año, combustible y kilometraje). Incluso posibilita estudiar cómo factores externos, como tendencias del mercado o las condiciones económicas de cada consumidor, influyen en el precio de venta.

5. Hipótesis iniciales.

- **1ra Hipótesis: Relación entre el tamaño del motor y el precio de venta.**

Planteamiento: Vehículos que poseen un mayor tamaño de motor (mayor cilindrada) tienden a tener un precio de venta más elevado que aquellos con motores de menor capacidad.

Fundamento: El tamaño del motor es una de las características técnicas más influyentes en el valor comercial de un automóvil, en general, los motores de mayor cilindrada ofrecen una potencia superior, una aceleración más rápida es uno de los factores que suelen asociarse con vehículos de gama media o alta. Estos vehículos también incorporan con frecuencia materiales, sistemas y tecnología más avanzada, lo que se refleja directamente en un costo de producción superior, por ende, en un precio más elevado al consumidor.

Además, desde la perspectiva del consumidor, la preferencia por motores grandes o pequeños varía según el propósito del vehículo, por ejemplo, los automóviles compactos o urbanos suelen tener motores más pequeños para priorizar la eficiencia de combustible, mientras que otros vehículos incorporan motores más grandes para ofrecer mayor rendimiento.

Esta hipótesis busca determinar si existe una correlación positiva y significativa entre la cilindrada del motor y el precio de venta, de ser así, podría llegar a la conclusión de que el tamaño del motor sigue siendo un factor determinante en la estrategia de fijación de precios dentro del mercado automotriz.

- **2da Hipótesis: Influencia del tipo de combustible en el precio y participación en el mercado.**

Planteamiento: Vehículos con tecnología híbrida presentan un precio de venta más elevado y una menor proporción de ventas totales en comparación con los vehículos de gasolina o diésel.

Fundamento: En la actualidad la industria automotriz, la transición hacia tecnologías más sostenibles ha impulsado la producción de vehículos híbrido e incluso eléctricos, sin embargo, el costo de estos modelos suele ser más elevado debido a las características que conlleva. Este incremento en los costos de producción repercute directamente en el precio de venta al consumidor. A pesar de los beneficios ambientales, el mercado global aún se

encuentra en un proceso gradual de adopción de estas tecnologías, factores como falta de infraestructura, precios altos y la limitada oferta de modelos pueden reducir la participación de los vehículos eléctricos e híbridos en las ventas totales.

Por otra parte, los vehículos de gasolina continúan siendo los más rentables debido a un costo más accesible, facilidad de mantenimiento y la amplia disponibilidad de modelos.

El objetivo de la hipótesis es evaluar cómo el tipo de combustible influye tanto en el precio de venta como en la distribución del mercado automotriz, los resultados permitirían conocer si los consumidores están dispuestos a pagar un precio elevado por vehículos ecológicos o se sigue en la compra de vehículos a gasolina, este análisis es especialmente relevante para fabricantes que buscan adaptar su oferta a las nuevas tendencias de movilidad sostenible.

- **3ra Hipótesis: Influencia del fabricante en la fijación del precio.**

Planteamiento: Existe diferencias significativas en el precio promedio de cada vehículo según el fabricante, lo cual refleja distintos niveles de posicionamiento de marca y estrategias de mercado.

Fundamento: En la industria automotriz, cada fabricante adopta una estrategia de posicionamiento diferente en función de su público objetivo, reputación y calidad del producto. Marcas de lujo como BMW, suelen mantener precios considerablemente más altos debido al prestigio, diseño, materiales y tecnología avanzada, por el contrario, fabricantes como Toyota se enfoca más en ofrecer vehículos de mayor accesibilidad económica, priorizando la relación costo – beneficio, durabilidad y el bajo mantenimiento.

Estas diferencias reflejan no solo estrategias comerciales distintas, sino también percepciones del consumidor sobre el valor y la calidad del vehículo, analizar el comportamiento del precio por fabricante permite identificar qué marca se encuentra mejor posicionadas, cuáles logran justificar precios y cómo se distribuye la demanda entre las diferentes categorías del mercado.

El propósito de esta hipótesis es determinar si el fabricante del vehículo ejerce una influencia estadísticamente significativa sobre el precio de venta, en pocas palabras buscar comprobar si las diferencias de precios observadas

entre cada marca responden principalmente al valor del producto y sus estrategias de posicionamiento.

6. Definición de Stakeholders clave.

1. **Fabricantes de Automóviles:** Son los principales actores en la fabricación del producto y la información, tienen interés en conocer cómo se comportan sus ventas frente a otras de la competencia, qué modelos tienen mejor desempeño comercial y qué factores influyen en el precio y demanda.

Interés en el proyecto: Analizar las tendencias de ventas por marca, modelo y tipo de combustible, del mismo modo, identificar oportunidades para ajustar precios o lanzar nuevos modelos, por último, predecir la demanda futura según las características de los vehículos.

2. **Distribuidores:** Son los intermediarios directos entre el fabricante y el consumidor, su rol principal es la venta y distribución de los vehículos.

Interés en el proyecto: Conocer qué modelos o tipo de vehículos se vender con mucha mayor rapidez, anticipar la demanda para mejorar la gestión del inventario, del mismo modo, establecer estrategias de precios o incluso promociones basadas en patrones de ventas reales e identificar las preferencias de los clientes según la región o el perfil de cada comprador.

3. **Clientes:** Representa la demanda directa del mercado, no participan directamente en el análisis, sin embargo, sus decisiones de compra son la base de toda la información del conjunto de datos.

Interés en el proyecto: Obtener mejores opciones de comprar según sus preferencias en gustos, igual se benefician de la competencia entre fabricantes mediante precios más competitivos, porque son el centro de estrategias personalizadas en cuestión de venta.

4. **Departamento de Marketing y ventas:** Encargados de interpretar los resultados del análisis para diseñar campañas, promociones y estrategias comerciales.

Interés en el proyecto: Identificar los segmentos de clientes con mayor potencial de compra, analizar qué factores influyen en la decisión de compra (marca, tipo de combustible, cilindrada, precio, etc.), del mismo modo, para determinar qué modelos requieren mayor impulso publicitario.

5. Directivos: Su papel recae en tomar decisiones estratégicas a partir de los resultados del análisis, su interés se centra en la rentabilidad, competitividad y la planificación a largo plazo.

Interés del proyecto: Evaluar el desempeño general del mercado, así mismo, identificar oportunidades de expansión o diversificación y definir inversiones en innovación o en nuevos segmentos.

7. Preguntas clave.

Preguntas que guían la investigación y descubrimiento.

1. ¿Qué periodo de tiempo cubre tu conjunto de datos y cómo las condiciones económicas y tecnológicas influyen en los resultados?
2. ¿Qué características del mercado automotriz actual justifican la necesidad de este análisis (transición a los vehículos eléctricos, inflación)?
3. ¿El conjunto de datos incluye todas las variables relevantes o faltan aspectos importantes?
4. ¿Cómo fueron recolectados los datos y cuál es su nivel de confiabilidad o actualización?
5. ¿Es necesario completar esta base con fuentes externas (por ejemplo, datos de consumo de combustible, emisiones de CO₂, satisfacción del cliente)?
6. ¿Qué herramientas estadísticas de visualización utilizarás (Python, Power BI, Excel)?
7. ¿Buscas construir un modelo predictivo del precio de los vehículos o solo comprender patrones de comportamiento del mercado?
8. ¿Qué tipo de visualización facilitan la interpretación de las tendencias (gráficos de dispersión, gráficos circulares o de barra, histograma)?
9. ¿Qué descubrimiento podrían aportar valor estratégico a fabricantes o distribuidores?
10. ¿Cómo se traducen los hallazgos del análisis en decisiones prácticas (por ejemplo, ajustar precios o lanzar nuevas versiones de los vehículos)?
11. ¿Qué implicaciones tendría el análisis para la sostenibilidad ambiental o la transición hacia vehículos eléctricos?
12. ¿Qué pasos a futuro podrían derivarse del análisis: nuevos análisis, campañas de mercado, recomendaciones en ventas o desarrollo de vehículos nuevos?

8. Fuente de datos identificadas.

- **Lista de fuentes de datos que se consideran importantes para probar las hipótesis iniciales.**

1. Datos internos del conjunto de datos (DataSet), fuente principal del análisis.

Descripción: Este conjunto de datos (DataSet) contiene información detallada sobre las ventas de vehículos nuevos incluyendo variables como fabricante, modelo, tipo de combustible, tamaño del motor, precio, año de lanzamiento.

Importancia:

Es la fuente principal para probar las tres hipótesis.

Constituye la base de todos los análisis descriptivos y predictivos.

2. Registro de fabricantes y asociaciones automotrices.

Ejemplo de fuente:

Asociación Mexicana de la Industria Automotriz (AMIA).

Datos de ventas de fabricantes (Toyota, Volkswagen, etc).

Importancia:

Permite validar el número de unidades vendida, precio oficial y tipo de modelos más comercializados.

Aporta contexto sobre participación de mercado de fabricantes y estrategias de marca.

3. Fuentes ambientales y normativas.

Ejemplo de fuente:

Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT).

Agencia de Protección Ambiental (EPA).

Importancia:

Proveen información sobre restricciones ambientales, incentivos para autos eléctricos o híbridos, políticas de transición energética.

Ayuda a explicar la disminución o aumento de ventas según el tipo de motor y combustible.

4. Fuentes económicas y financieras.

Ejemplo de fuente:

Instituto Nacional de Estadística y Geografía (INEGI).

Banco de México.

Banco Mundial.

Fondo Monetario Internacional.

Importancia:

Permite contextualizar los datos dentro del entorno económico, considerando factores como inflación, ingreso promedio y tipo de cambio. Son muy importantes para entender si los cambios de precio responden al mercado automotriz o condiciones económicas generales.

9. Justificación del proyecto.

El mercado automotriz se encuentra en una etapa de profunda transformación impulsada por diferentes factores, tecnológicos, económicos y ambientales, en este sentido, se busca comprender los patrones de venta se convierte en una necesidad estratégica tanto para fabricantes, distribuidores y analistas dentro del sector. Este proyecto de análisis de datos busca estudiar de manera detallada las relaciones entre variables clave como el tamaño del motor, tipo de combustible, precio de venta, entre otros, con el propósito de identificar tendencias de comportamiento del consumidor y factores que determinan el valor comercial de cada vehículo.

La importancia de realizar este análisis radica en que decisiones de compra y venta de vehículos están influenciadas por varios factores (económico, tecnológico, ambiental, percepción de cada marca) así que requieren una interpretación basada en datos reales, a través del estudio se pretende obtener una visión del mercado automotriz, determinando de manera significativa en el precio final y en la demanda.

Desde una perspectiva empresarial, este proyecto resulta relevante porque permite convertir la información en conocimiento para una toma de decisiones, a través del análisis de correlaciones, tendencias y comparaciones entre fabricantes, los resultados pueden servir de base para predecir comportamientos y comparaciones entre fabricantes, los resultados pueden servir de base para predecir comportamientos futuros del mercado, identificar segmentos con mayor crecimiento, ajustar la estrategia comercial ante cambios en los costos energéticos o en las preferencias del consumidor. En suma, este proyecto no solo busca describir la realidad del mercado automotriz, sino explicar las causas y consecuencias de los patrones observados, aportando el valor tanto a la industria como al ámbito académico, la información obtenida permitirá comprender mejor la dinámica de las ventas, anticipar tendencias, contribuir a un desarrollo más sostenible, competitivo y orientado.

10. ¿Cuántos datos y qué tipo son?

Son un total de 50,000 datos divididos en 7 columnas.

Son de tipo:

Manufactura: "Texto o cadena de caracteres"

Modelo: "Texto o categoría"

Tamaño de motor: "Número decimal"

Tipo de combustible: "Texto"

Año de fabricación: "Número entero"

Kilometraje: "Número entero"

Precio: "Número entero"

The screenshot shows a Jupyter Notebook interface with three open files: Untitled-2.ipynb, Untitled-1.ipynb, and Untitled-3.ipynb. The current cell (Untitled-3.ipynb) contains the following code and output:

```
1 print(df.head())
...
   Manufacturer      Model  Engine size Fuel type Year of manufacture \
0      Ford        Fiesta       1.0    Petrol            2002
1    Porsche     718 Cayman       4.0    Petrol            2016
2      Ford       Mondeo       1.6   Diesel            2014
3    Toyota      RAV4        1.8  Hybrid            1988
4      VW         Polo        1.0    Petrol            2006

   Mileage  Price
0  127300   3074
1   57850  49704
2   39190  24072
3  210814   1705
4  127869   4101
```

Below this, another cell (Untitled-3.ipynb) shows:

```
1 print(df.dtypes)
...
   Manufacturer      object
   Model             object
   Engine size    float64
   Fuel type       object
   Year of manufacture  int64
   Mileage          int64
   Price            int64
dtype: object
```

The interface includes a sidebar with file navigation, a status bar at the bottom, and a toolbar with various icons.