



# Análisis de datos en ventas de vehículos y su impacto en el precio

Nombre del estudiante: **Gonzalo Leonel Gramajo**

Comisión: **65180**

Número de entrega: **2**

Inicio de la entrega: **Semana 4**

Año 2024

## ÍNDICE

Descripción de la temática de los datos.....	2
Descripción de la hipótesis.....	2
Dataset.....	3
Objetivo del proyecto.....	3
Alcance.....	3
Usuario final y nivel de aplicación del análisis.....	4
Diagrama entidad-relación.....	4
Listado de tablas.....	4
Listado de columnas de cada tabla.....	6

## Descripción de la temática de los datos

Siguiendo la metodología SMART

**Specific (Específico):** El objetivo es evaluar cómo el estado del vehículo (condición) y el kilometraje (odometer) afectan el precio de venta (sellingprice) en diferentes marcas (make) y modelos (model). Se busca identificar patrones que permitan predecir el impacto de estas variables en el precio final y cómo estas relaciones varían según la categoría del vehículo, como sedanes, SUVs, o camiones.

**Measurable (Medible):** Identificar una correlación estadística entre el precio de venta, el estado del vehículo y el kilometraje para los distintos modelos de automóviles.

**Alcanzable (Achievable):** Utilizar herramientas de análisis de datos como Python (librerías: Pandas, Matplotlib) y conocimientos adquiridos en el curso para realizar las visualizaciones y análisis estadísticos. Los recursos necesarios están disponibles (dataset y software). [LINK AL DATASET](#)

**Relevante (Relevant):** Este análisis es crucial para entender los factores que determinan el precio de venta en el mercado de vehículos usados. Al comprador le permite identificar qué vehículos ofrecen mejor valor según su estado y kilometraje. Al vendedor, le ayuda a establecer precios competitivos basados en características objetivas del vehículo.

**Temporal (Time-bound):** El proyecto se llevará a cabo en tres etapas, cada una con fechas definidas de entrega. En la primera etapa se presentará la descripción del tema, la formulación de la hipótesis y la selección del dataset. Las etapas posteriores se enfocarán en realizar un análisis detallado de los datos y en desarrollar visualizaciones, concluyendo con la creación de un tablero en Power BI y la entrega de la documentación final del proyecto.

## Descripción de la hipótesis

**Hipótesis:** "Los vehículos en mejor estado y con menor kilometraje tienden a venderse a precios significativamente más altos, independientemente de la marca o modelo."

**Justificación:** Esta hipótesis se basa en principios ampliamente aceptados en el mercado de vehículos usados, más precisamente y acorde al conocimiento de quien redacta este informe

- Los vehículos en **mejor estado físico y mecánico** presentan menos desgaste, lo que genera mayor confianza en los compradores. Esto reduce la necesidad de reparaciones inmediatas y, por ende, aumenta el valor percibido y el precio dispuesto a pagar.
- El **kilometraje** es uno de los **indicadores clave** de desgaste de un vehículo. Los vehículos con menor kilometraje suelen asociarse con menos uso, mayor vida útil remanente y menor probabilidad de problemas mecánicos, lo que se traduce en precios más altos.
- **Los compradores tienden a priorizar la fiabilidad y la longevidad del vehículo.** Estas cualidades están directamente relacionadas con el estado y el kilometraje, lo que hace que estos atributos sean más importantes que otros factores, como características opcionales o detalles estéticos.

## Dataset

El dataset cuenta con 16 columnas:

- **year:** Año de fabricación del vehículo.
- **make:** la marca o fabricante del vehículo.
- **model:** modelo específico del vehículo.
- **trim:** designación adicional del modelo del vehículo.
- **body:** el tipo de carrocería del vehículo (por ejemplo, SUV, sedán, etc.).
- **transmission:** el tipo de transmisión del vehículo (por ejemplo, automática).
- **vin:** número de identificación del vehículo, un código único para cada vehículo.
- **state:** el estado donde está registrado el vehículo.
- **condition:** estado del vehículo, posiblemente calificado en una escala.
- **odometer:** el kilometraje o distancia recorrida por el vehículo.
- **color:** color exterior del vehículo.
- **interior:** color interior del vehículo.
- **seller:** la entidad que vendió el vehículo.
- **mmr:** Manheim market report o informe de mercado de Manheim, posiblemente indicando el valor de mercado estimado del vehículo.
- **sellingprice:** el precio al que se vendió el vehículo.
- **saledate:** la fecha y hora en que se vendió el vehículo.

## Objetivo del proyecto

El objetivo de este proyecto es desarrollar un tablero interactivo en Power BI que permite analizar las oportunidades de compra o venta de vehículos acorde a su kilometraje, época u otras condiciones. Este tablero podrá dar una noción al vendedor o al comprador si el precio de una posible compra o venta es justa, rentable o no conviene realizar el trato.

## Alcance

El proyecto abarca:

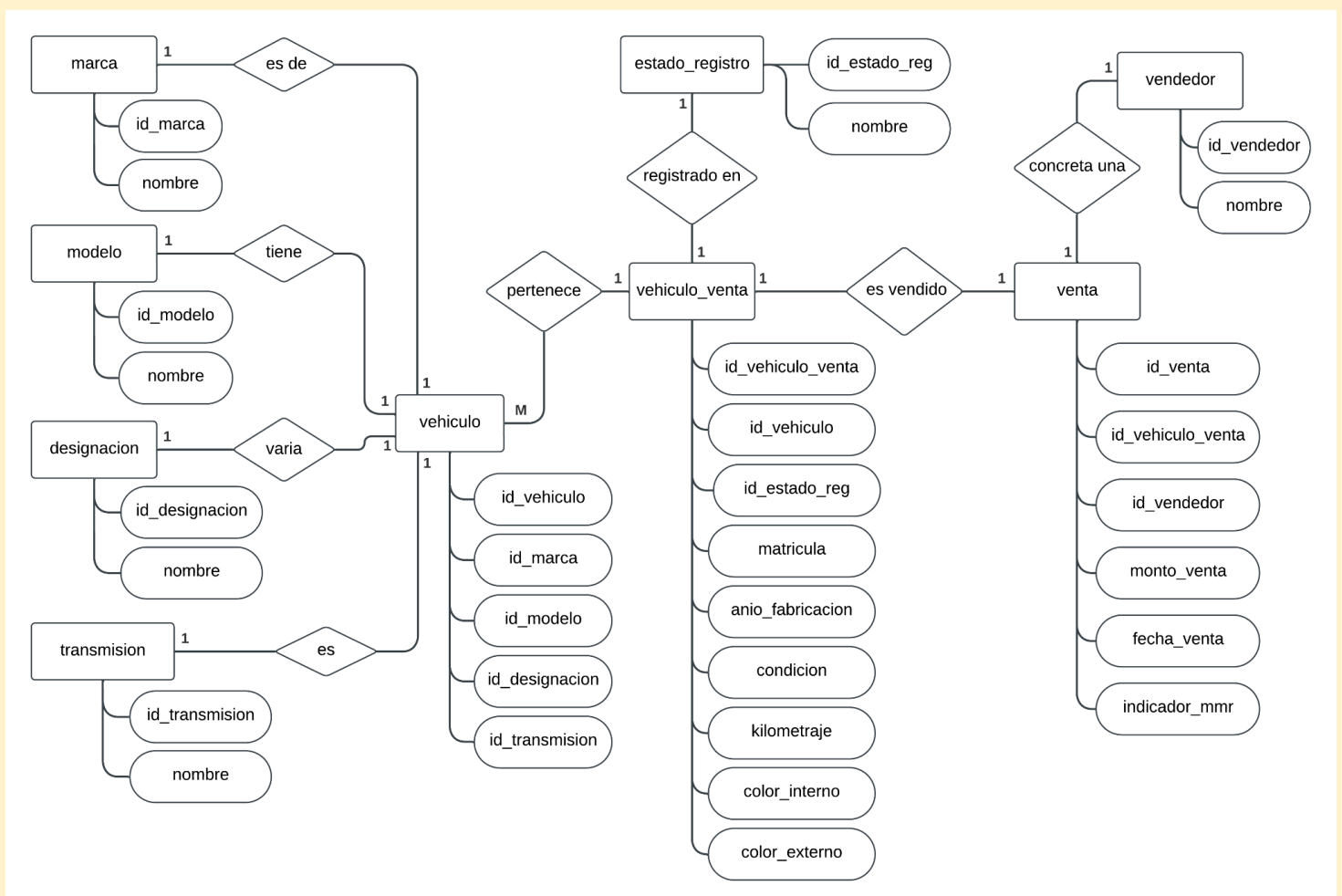
1. Recopilación, limpieza y organización de datos provenientes de las tablas realizadas en base al dataset elegido "Vehicle Sales Data".

2. Creación de un modelo de datos que refleje las relaciones entre las diferentes entidades reconocidas en el dataset “Vehicle Sales Data”.
3. Desarrollo de visualizaciones interactivas en Power BI que permitan a los usuarios explorar y analizar los datos de manera intuitiva.
4. Este documento, es decir, la documentación detallada del proceso de análisis de datos, incluyendo la descripción de las tablas y sus relaciones, junto con los hallazgos del análisis.
5. Alcance temporal delimitado por los datos que contiene el dataset, es decir desde el 3 de abril de 2015 al 27 de mayo del 2015.

## Usuario final y nivel de aplicación del análisis

El tablero está diseñado para ser utilizado por los posibles compradores o vendedores de automóviles que deseen respaldar su decisión de compra o venta acorde o en correlación a un gran cantidad de datos de ventas pasadas, logrando así definir un precio justo.

## Diagrama entidad-relación



## Listado de tablas

**Tabla “marca”:** esta tabla representa a las marcas de vehículos que aparecen en el dataset. Corresponde a los nombres que aparecen en la columna “make” del dataset.

Nombre del campo	PK o FK	Relación
id_marca	PK	con la tabla vehículo por el campo id_marca
nombre		

**Tabla “modelo”:** esta tabla corresponde a los modelos de vehículos que tienen las diversas marcas, por ejemplo, Chevrolet tiene modelos como ser Camaro, Cruze, Corsa, entre otras. Todas estas últimas están en esta tabla.

Nombre del campo	PK o FK	Relación
id_modelo	PK	con la tabla vehículo por el campo id_modelo
nombre		

**Tabla “designacion”:** esta tabla corresponde a las designaciones que las diversas marcas le dan a un modelo de vehículo. Estas pueden ser las mismas o pueden cambiar de modelo a modelo de vehículo de la misma marca. Por ejemplo, Chevrolet Cruze puede tener designación “LS”, “LT” o “LTZ”. Estas últimas son las designaciones.

Nombre del campo	PK o FK	Relación
id_designacion	PK	con la tabla vehículo por el campo id_designacion
nombre		

**Tabla “transmision”:** esta tabla corresponde a la transmisión del vehículo. Este dato está en función de la designación o del modelo, dependiendo la marca, el año. Como esto es muy variable entre marcas, se lo atribuye directamente al vehículo, aunque la cantidad de estas posibles transmisiones no es muy amplia, por ejemplo, tracción trasera, eléctrico, tracción delantera, entre otros.

Nombre del campo	PK o FK	Relación
id_transmision	PK	con la tabla vehículo por el campo id_transmision
nombre		

**Tabla “vehiculo”:** esta tabla corresponde a la descripción de un vehículo determinado. No se añaden otros campos como el de colores porque estos son muy variables e incluso cuando uno adquiere un auto, el color se lo puede cambiar a gusto, por lo que en esta tabla queda definido el auto completamente, menos por el color. Esto siempre y cuando hablemos de los datos que nos dispone el dataset, porque por ejemplo hay otros datos como la potencia o torque que no se detallan.

Nombre del campo	PK o FK	Relación
id_vehiculo	PK	con la tabla vehiculo por el campo id_marca
id_marca	FK	con la tabla marca por el campo id_marca
id_modelo	FK	con la tabla modelo por el campo id_modelo

id_designacion	FK	con la tabla designación por el campo id_designacion
id_transmision	FK	con la tabla transmision por el campo id_transmision

**Tabla “estado\_registro”:** este campo corresponde a todos los estados de los Estados Unidos de Norteamérica. Esto es porque los datos corresponden a este país y porque en ese país, los vehículos son matriculados o registrados con patentes correspondientes a ese Estado, por lo tanto el estado es una característica perteneciente a cada vehículo.

Nombre del campo	PK o FK	Relación
id_estado_reg	PK	con la tabla vehículo por el campo id_estado_reg
nombre		

**Tabla “vehiculo\_venta”:** esta tabla corresponde a las ventas que se tienen en el dataset. Cada línea sería una venta. Esta sería la tabla principal del dataset y del objetivo del proyecto.

Nombre del campo	PK o FK	Relación
id_vehiculo_venta	PK	con la tabla venta por el campo id_venta
id_vehiculo	FK	con la tabla vehiculo por el campo id_vehiculo
id_estado_reg	FK	con el tabla estado_registro por el campo id_estado_reg
matricula		
anio_fabricacion		
condicion		
kilometraje		
color_interno		
color_externo		

**Tabla “venta”:** esta tabla corresponde a los datos referidos a la venta en sí, es decir, la venta es la entidad y los datos que contiene la describen. Esta sería la tabla principal de la base de datos y del objetivo del proyecto.

Nombre del campo	PK o FK	Relación
id_venta	PK	con la tabla vehículo por el campo id_marca
id_vehiculo_venta	FK	con la tabla vehiculo_venta por el campo id_vehiculo_venta
id_vendedor	FK	con la tabla vendedor por el campo id_vendedor
monto_venta		
fecha_venta		

indicador_mmr		
---------------	--	--

**Tabla “vendedor”:** esta tabla corresponde a todas las entidades que se dedican a vender los automóviles o vehículos. Con respecto a esta entidad solo se tiene como dato el nombre.

Nombre del campo	PK o FK	Relación
id_vendedor	PK	con la tabla venta por el campo id_vendedor
nombre		

## Listado de columnas de cada tabla

**Tabla “marca”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_marca	Integer	Si		No
nombre	Varchar			No

**Tabla “modelo”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_modelo	Integer	Si		No
nombre	Varchar			No

**Tabla “designacion”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_marca	Integer	Si		No
nombre	Varchar			No

**Tabla “transmision”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_transmision	Integer	Si		No
nombre	Varchar			No

**Tabla “vehiculo”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_vehiculo	Integer	Si		No

id_marca	Integer		Si	No
id_modelo	Integer		Si	No
id_designacion	Integer		Si	No
id_transmision	Integer		Si	No

**Tabla “estado\_registro”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_estado_reg	Integer	Si		No
nombre	Varchar			No

**Tabla “vehiculo\_venta”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_vehiculo_venta	Integer	Si		No
id_vehiculo	Integer		Si	No
id_estado_reg	Integer		Si	No
matricula	Varchar			No
anio_fabricacion	Year			No
condicion	Integer			No
kilometraje	Integer			No
color_interno	Varchar			Si
color_externo	Varchar			Si

**Tabla “venta”:**

Nombre del campo	Tipo de dato	PK	FK	Null
id_venta	Integer	Si		No
id_vehiculo_venta	Integer		Si	No
id_vendedor	Integer		Si	No
monto_venta	Float			No
fecha_venta	Date			No
indicador_mmr	Integer			Si

**Tabla “vendedor”:**



Nombre del campo	Tipo de dato	PK	FK	Null
id_vendedor	Integer	Si		No
nombre	Varchar			No