

Reporte Fase 1: Descubrimiento del Proyecto

Plantilla del primer reporte de proyecto

1. Título del Proyecto

- *Nombre del proyecto.* "Análisis del consumo de gasolina de automóviles en México".

2. Objetivo del Proyecto

- Definir las principales marcas de automóviles, así como sus versiones de autos que contaminan más dependiendo de donde sean conducidos, ya sea calle o carretera. Además de los principales que contaminen el ambiente.
- "La marca de automóviles Nissan en su versión versa son los que más contaminan a nivel carretera".
- "La marca de automóviles Kia en su versión Soul son los que menos contaminan a nivel calle".
- "Los 10 autos que menos contaminan el aire de México".

3. Descripción del Problema

- Como ya sabemos, la contaminación que existe en México cada vez es más preocupante, existen diferentes razones del porque se genera esta contaminación, ya que son diferentes tipos que hace que no solo sea un factor que genere tal contaminación. Enfocados principalmente en la contaminación del aire una de las principales razones son el uso de automóviles en México. Y es que al ya no ser un lujo sino una necesidad ya que facilita el transporte además de una mayor rapidez y eficacia para trasladarse de un lugar a otro, es por qué se utiliza demasiado este

tipo de transporte, pues la mayoría de las personas ya cuenta con uno para realizar sus actividades cotidianas.

No obstante, no todos los autos son iguales, es decir, si nos ayudan para cosas muy similares y principalmente el trasladarnos, pero de acuerdo con la necesidad del cliente, este puede optar por diferentes versiones de este transporte, desde camionetas para personas que requieran cargar bastante peso, hasta autos pequeños de 4 puertas para uso personal únicamente, hay quienes optan por una camioneta cerrada para poder ir más cómodo cuando son varios integrantes en la familia. Sea como sea todos los automóviles son diferentes y, por ende, son diferente su fabricación, así como las piezas que ocupan haciendo que unos contaminen más y otros menos, también influye mucho la empresa por la que fueron creados. Es por eso por lo que en esta base de datos podemos apreciar la marca, submarca y versión de distintos autos que podemos encontrar en México, además de algunas características más, pero lo que en realidad nos importa es los distintos números que generan de acuerdo con su contaminación CO₂, NO_x, además de acuerdo con en donde estén transitando generen más o menos contaminación, calle o carretera.

4. Recursos Disponibles

- *Tecnología y Herramientas*: Identificar las herramientas tecnológicas, sistemas y software con los que se cuenta para el análisis.

- o Se utilizarán herramientas como Python y Pandas para la manipulación de datos, además de Matplotlib para visualización inicial.

- *Datos*: Breve descripción de las fuentes de datos disponibles, por cada columna que significa.

- o Marca: Empresa que fabrica el auto

- o Submarca: “Grupo” al que pertenece

- o Versión: Versión del automóvil
- o Trans.: Transmisión del vehículo (Automático, estándar, etc.)
- o Comb.: Tipo de combustible
- o Potencia (HP): Caballos de fuerza del auto
- o Tamaño (L): Tamaño motor
- o Categoría: Categoría a la que pertenece
- o R. Ciudad (km/l): Rendimiento en Ciudad
- o R. Carr. (km/l): Rendimiento en Carretera
- o R. Comb. (km/l): Rendimiento en combustible
- o R. Ajust. (km/l): Rendimiento ajustado
- o CO₂(g/km): Generación de CO₂ por kilometro
- o NO_x (g/1000km): Generación de NO_x por kilometro
- o Calificación Gas Ef. Inv.: Calificación para el auto contra el gas de efecto invernadero
- o Calificación Contam. Aire: Calificación para el auto contra la contaminación del aire

5. Hipótesis Iniciales

- Formular 3 hipótesis basadas en la intuición inicial y en el conocimiento del dominio.
- Ejemplo:
 - o *Hipótesis 1*: "Los autos de transmisión automática generan mucha menos contaminación que los que ocupan Diesel".

- o *Hipótesis 2*: "Los automóviles más pequeños son los que mejor consumo de gasolina tendrán sea cual sea el rendimiento que probar".
- o *Hipótesis 3*: "La marca Ford al ser muy antigua, genera autos de mejor calidad a comparación de marcas con apenas unos cuantos años de creación, además de que los autos con más potencia serán los más contaminantes".

6. Definición de Stakeholders Clave

- Identificar quiénes son las partes interesadas en el proyecto y su rol.
- *Clientes Actuales*: Gente interesada en el tema, así como gente que requiera comprar un automóvil, pero sea ahorrativo para evitar el mayor gasto en combustible.
- *Departamento de Marketing*: Usará los hallazgos para diseñar campañas específicas contra la contaminación del aire.
- *Empresas de automóviles*: Al generar una base de datos se puede hacer conciencia de la gran contaminación que generan ciertos vehículos haciendo que se busque la falla y se mejore en las versiones del futuro de dicho automóvil.

7. Preguntas Clave

- Incluir las preguntas que guiarán la investigación y descubrimiento (mín. 10).
- Preguntas clave
 - o ¿Qué factores alteran la contaminación que genere un auto?
 - o ¿Cuáles son los autos que contaminan menos?
 - o ¿Qué generan los autos para contaminar el aire?

- o ¿El rendimiento de ciudad es diferente a carretera?
- o ¿Qué es el Nox y que tan diferente es al CO2?
- o ¿Hay una marca que sus autos tengan similares números?
- o ¿Entre más nueva sea la marca mayor sus autos contaminan?
- o ¿La transmisión influye en la contaminación?
- o ¿El tipo de combustible afecta la contaminación que el auto genera?
- o ¿La versión del auto influye en la contaminación?
- o ¿La categoría entre más “liviana” sea ocupa menos gasolina y por ende menos contaminación a diferencia de autos que su principal función sea la carga?
- o ¿Por qué existen diferentes tipos de rendimientos, pero con números muy similares?

8. Fuentes de Datos Identificadas

- Listar las fuentes de datos que se consideran importantes para probar las hipótesis iniciales.
 - Información sobre cada tipo de rendimiento
 - Información sobre las marcas de los principales autos
 - Características del Nox y CO2
 - Datos sobre cada transmisión y cómo funcionan
 - Características de cada tipo de combustible
 - Las diferentes categorías y en que consiste cada una
 - Como se obtiene la calificación para contaminación de gas de efecto invernadero y para calificación contaminación del aire

- Como definen la “Potencia (HP) en los autos
- Datos de los cilindros y su función
- Información sobre el “Tamaño (L) “de los autos y a que se refieren con exactitud

9. Justificación del Proyecto

- Explicar la importancia del análisis en media cuartilla.
- La contaminación que generan los autos es un tema muy importante hoy en día ya que como comentaba anteriormente, los autos cada vez se vuelven más una necesidad por lo que no es nada del otro mundo ver que las calles se saturan a ciertas horas de automóviles, ya sea por ir al trabajo, la escuela o simplemente dar la vuelta. Pero al ocupar combustible este genera una contaminación haciendo que sea casi imposible de evitar pues seamos realistas, jamás se dejaran de ocupar los autos o no por ahora y en un bastante tiempo.

Por eso creo es importante conocer más a fondo sobre los autos que contaminan menos, principalmente para las personas que deseen adquirir un auto sepan cual es el que menos contamina o mejor rendimiento tiene para que lo tomen en cuenta si lo que necesitan solo es por necesidad y ya no tanto por un lujo. Además, pensando en grande podría llegar esta información hasta un lugar de gobierno para que se haga marketing de que autos son los que rinden mejor y para los que no tengan las mejores calificaciones se les tome mayor en cuenta para que regularicen esos números y sino de lo contrario empezar a implicar sanciones o incluso prohibir la distribución de ese vehículo en México, pues como sabemos, en algunos lugares de México el aire ya esta demasiado contaminado, como CDMX, donde se aplican varias “reglas” como el hoy no circula.

Estamos aun con tiempo para empezar a cambiar nuestras vidas en un futuro, ya que este problema cada vez se vuelve algo irremediable, y que mejor iniciar un cambio con algo que no podremos evitar, pero si cambiar, el uso del automóvil.

11.- ¿Cuántos datos y que tipo son?

En total son 4692 filas con 18 columnas donde existen 6 columnas de tipo entero (int), 6 de

tipo objeto (object) y 6 de tipo decimal (float). Agrego imagen para tener más detalle.

```
<class 'pandas.core.frame.DataFrame'>
Index: 4692 entries, 0 to 5794
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Marca                                4692 non-null   object
1   Submarca                             4692 non-null   object
2   Versión                              4692 non-null   object
3   Modelo                              4692 non-null   int64
4   Trans.                              4692 non-null   object
5   Comb.                                4692 non-null   object
6   Cilindros                            4692 non-null   int64
7   Potencia (HP)                       4692 non-null   int64
8   Tamaño (L)                          4692 non-null   float64
9   Categoría                            4692 non-null   object
10  R. Ciudad (km/l)                    4692 non-null   float64
11  R. Carr. (km/l)                     4692 non-null   float64
12  R. Comb. (km/l)                     4692 non-null   float64
13  R. Ajust. (km/l)                    4692 non-null   float64
14  CO2(g/km)                           4692 non-null   int64
15  NOx (g/100km)                       4692 non-null   int64
16  Calificación Gas Ef. Inv.            4692 non-null   int64
17  Calificación Contam. Aire            4692 non-null   float64
dtypes: float64(6), int64(6), object(6)
memory usage: 696.5+ KB
```

Autor: Carlos Alexander Castillo Gomez