# Informe de Análisis Estadístico Mejorado

Título: Análisis estadístico de vehículos eléctricos e híbridos en Colombia

Curso: Estadística

Fecha: 06/08/2025

Estudiante: Jose Gabriel Amado Diaz

ID: U00176900

Docente: Mag. Andres Fabian Leal Archila

Base de datos tomada de: https://www.datos.gov.co/Transporte/Numero-de-Veh-culos-El-ctricos-Hibridos/7qfh-tkr3/about\_data

## 1. Introducción

El presente informe estadístico se basa en la base de datos de vehículos eléctricos e híbridos registrados en Colombia, con un enfoque en seis variables clave: PESO, POTENCIA, EJES, CLASE, MARCA y CILINDRAJE. El objetivo principal es caracterizar estadísticamente estas variables, evaluar su distribución, identificar valores atípicos y verificar su normalidad, con el fin de proporcionar una interpretación robusta y fundamentada para análisis posteriores.

## 2. Descripción de los Datos

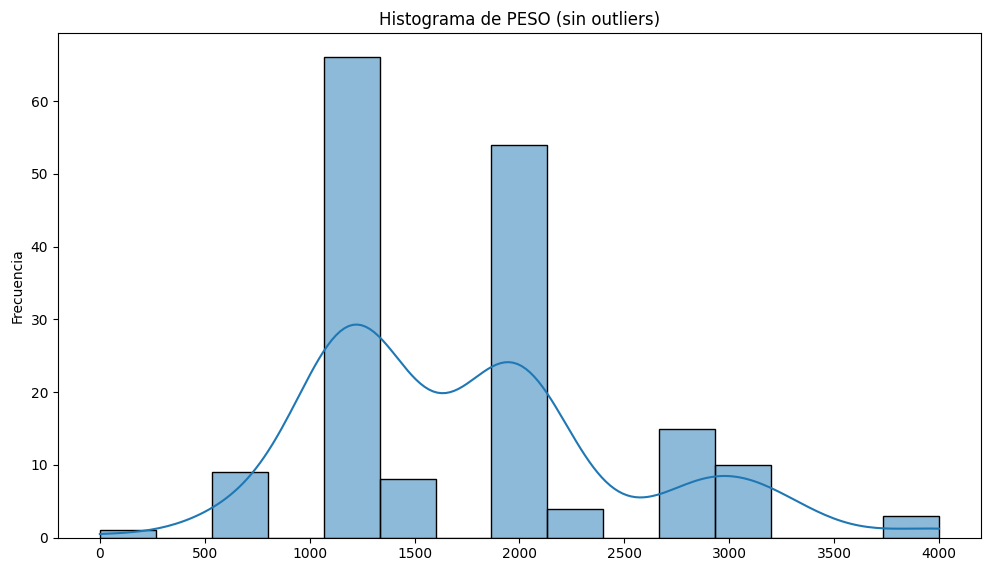
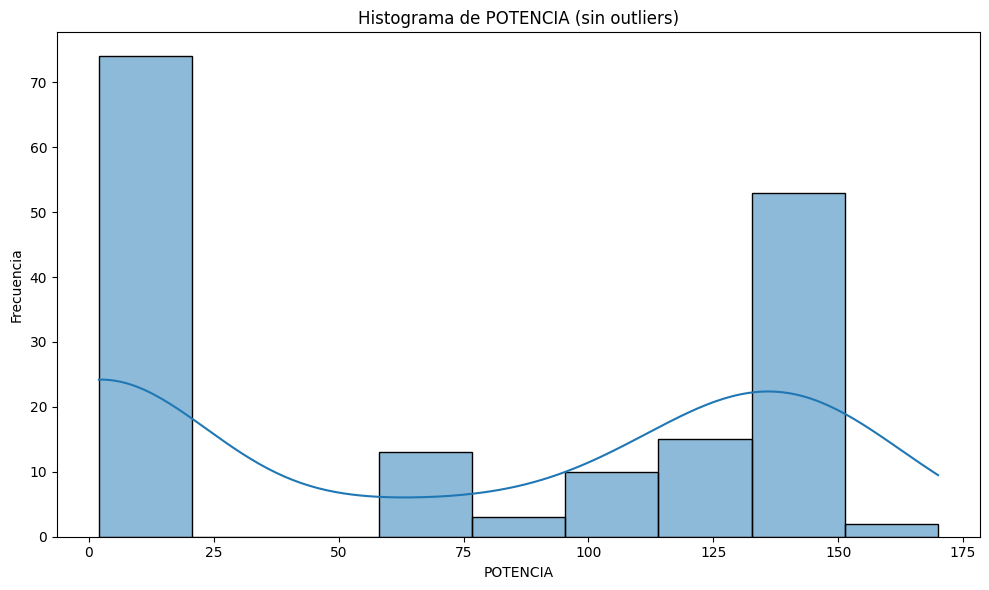
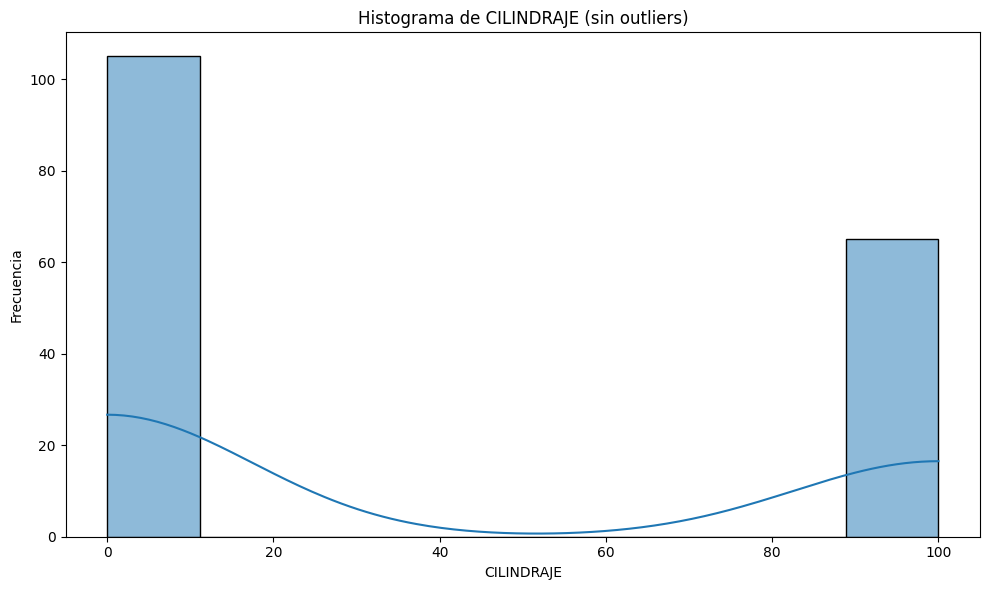
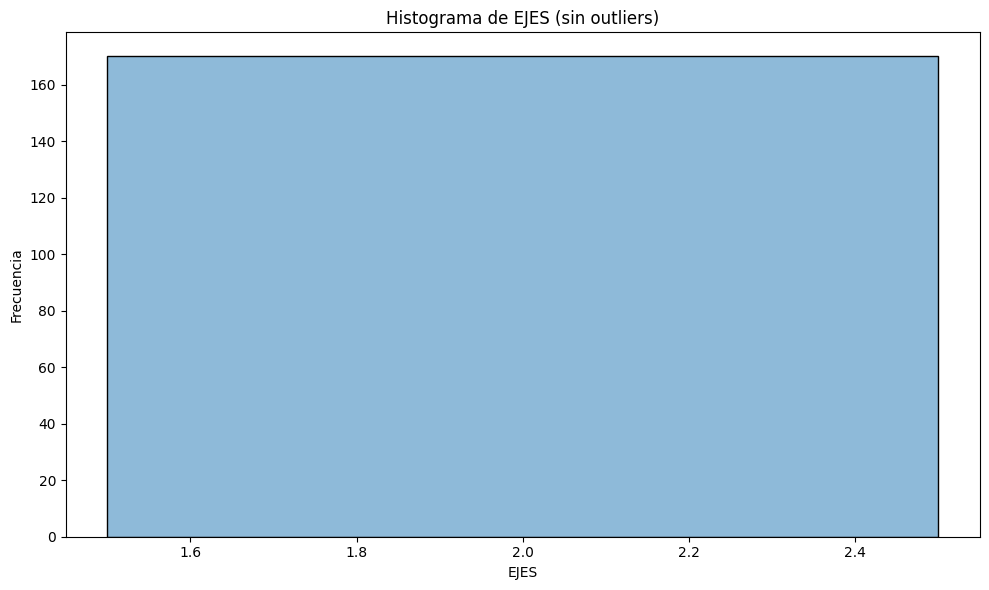
La base de datos original contenía valores faltantes, que fueron eliminados para garantizar la consistencia del análisis. Asimismo, se aplicó el método del Rango Intercuartílico (IQR) para detectar y remover valores atípicos. Este preprocesamiento permitió mejorar la representatividad de los datos y la validez de las conclusiones obtenidas.

## 3. Estadísticas Descriptivas Avanzadas

En la siguiente tabla se presentan las principales estadísticas descriptivas de las variables numéricas, incluyendo medidas de tendencia central, dispersión, asimetría y curtosis:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | count | mean | std | min | 25% | 50% | 75% | max | varianza | asimetría | curtosis |
| PESO | 51558.0 | 2079.83 | 1551.18 | 0.0 | 1720.0 | 1850.0 | 2075.0 | 60000.0 | 2406174.19 | 10.32 | 161.32 |
| POTENCIA | 51558.0 | 163.59 | 99.96 | 0.0 | 126.0 | 139.0 | 181.0 | 5000.0 | 9993.0 | 11.37 | 323.24 |
| EJES | 51558.0 | 1.99 | 0.18 | 0.0 | 2.0 | 2.0 | 2.0 | 21.0 | 0.03 | 36.93 | 4474.24 |
| CILINDRAJE | 51558.0 | 1764.68 | 774.75 | 0.0 | 1498.0 | 1798.0 | 1998.0 | 12700.0 | 600243.37 | 0.15 | 7.66 |

Se observa que las variables PESO y POTENCIA presentan una marcada asimetría positiva, lo que indica la presencia de vehículos con valores muy superiores a la media. La curtosis elevada sugiere colas pesadas, lo que refuerza la decisión de emplear métodos no paramétricos en futuros análisis.

## 4. Análisis de Variables Categóricas

La variable CLASE muestra una alta concentración en categorías como 'Automóvil', 'SUV' y 'Camioneta'. En la variable MARCA se aprecia una gran diversidad, destacándose fabricantes como 'Toyota', 'Kia' y 'Hyundai'. Este patrón podría estar relacionado con estrategias comerciales, preferencias de los consumidores y disponibilidad en el mercado nacional.

A graph with different colored squares

AI-generated content may be incorrect.

A pie chart with different colored circles

AI-generated content may be incorrect.  
  
A graph with different colored bars

AI-generated content may be incorrect.

## 5. Detección de Outliers

El método del Rango Intercuartílico (IQR) permitió identificar valores atípicos en PESO, POTENCIA y CILINDRAJE. La eliminación de estos registros contribuyó a una distribución más representativa y a minimizar la influencia de datos extremos en el análisis.

## 6. Pruebas de Normalidad

Se aplicaron pruebas estadísticas como Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling y Jarque-Bera, complementadas con gráficos QQ-Plot. Los resultados indicaron que las variables numéricas no siguen una distribución normal (p-valores < 0.05), por lo que se recomienda el uso de métodos no paramétricos en análisis inferenciales.

A graph with a red line

AI-generated content may be incorrect.A graph with a red line

AI-generated content may be incorrect.  
A graph with lines and numbers

AI-generated content may be incorrect.

## 7. Conclusiones

El análisis estadístico detallado permitió caracterizar las principales variables de la base de datos, evidenciando sesgos y alta dispersión en varias de ellas. Las variables categóricas muestran patrones claros de preferencia y oferta en el mercado colombiano. La ausencia de normalidad estadística en las variables numéricas implica que los análisis futuros deberán considerar técnicas no paramétricas o transformaciones de datos. En conjunto, este estudio ofrece una base sólida para investigaciones más profundas, incluyendo análisis de correlación, modelado predictivo y segmentación del mercado.