

Análisis de Datos de Automóviles

Este proyecto se enfoca en el análisis de datos recopilados mediante web scraping sobre las características de automóviles. El objetivo es aplicar los conocimientos adquiridos en diferentes espacios curriculares de la carrera de Ciencia de Datos e Inteligencia Artificial. El equipo de trabajo está conformado por los docentes Nahuel Pratta y Marcos Ugarte, junto con los estudiantes Erick López, Christian Nüesch y Débora Zurita.





Objetivos del Proyecto

1

Análisis de Datos

El proyecto se centra en el análisis de datos recolectados sobre características de automóviles, con el fin de aplicar los conocimientos adquiridos en diferentes espacios curriculares de la carrera.

2

Tratamiento de Datos

El dataset contiene valores atípicos, errores, datos ausentes e inconsistencias, por lo que se procederá a "curar" el conjunto de datos para poder realizar un análisis efectivo.

3

Visualización de Datos

Se generaron gráficos utilizando la librería matplotlib para tener un abordaje visual de los datos y poder entenderlos mejor.

Características del Dataset

Formato

El dataset se encuentra en formato RAW, lo que significa que contiene errores, valores atípicos, valores ausentes e inconsistencias.

Columnas

El dataset comprende varias características de automóviles ordenadas en doce columnas, obtenidas de un sitio web en Argentina a principios de enero de 2023.

Librerías Utilizadas

Para trabajar con el dataset se importaron las librerías de Python numpy, pandas y matplotlib.pyplot.

Análisis Inicial

Se leyeron los datos desde un archivo CSV y se analizaron los tipos de datos, valores nulos y uso de memoria.

Análisis Estadístico

Estadísticas Básicas

Se realizaron análisis estadísticos básicos sobre las columnas numéricas, como mínimo, máximo, media, varianza y desviación estándar.

También se calcularon los valores válidos sobre el total de datos.

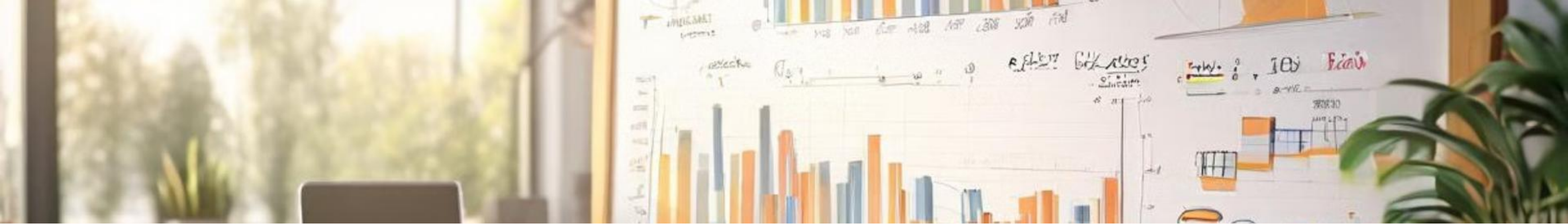
Tratamiento de Datos Faltantes

Para los datos ausentes, se rellenaron con la moda o se eliminaron filas, mientras que para los datos en formato incorrecto se aplicaron transformaciones usando expresiones regulares y la función `apply()` de pandas.

Visualización de Datos

Se generaron gráficos usando la librería `matplotlib` para tener un abordaje visual de los datos y entenderlos mejor.

El objetivo era poner en práctica los conocimientos adquiridos en los espacios curriculares "Estadística y exploración de datos", "Procesamiento de Datos" y "Ciencia de Datos".



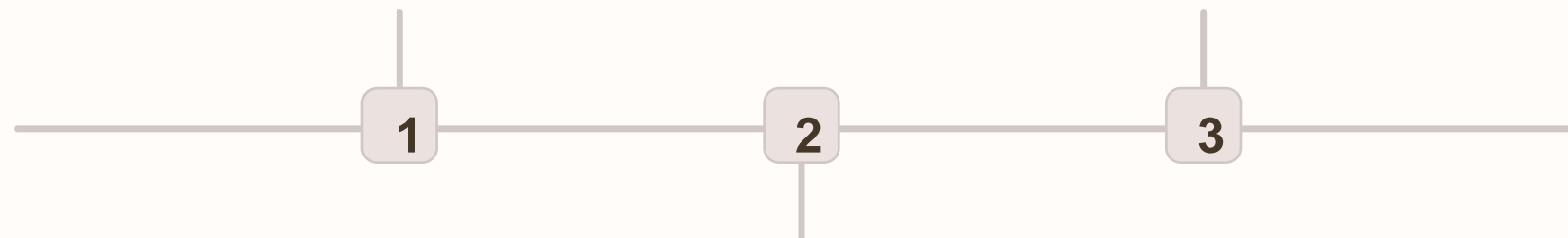
Próximos Pasos

Análisis Avanzado

Una vez que el dataset está limpio y preparado, se pueden realizar análisis más avanzados, como modelos de machine learning para predecir características de los automóviles.

Presentación de Resultados

Los hallazgos y conclusiones del proyecto se presentan en el documento final y en la notebook de Python trabajada, con el fin de compartir los conocimientos adquiridos.



Visualizaciones Interactivas

Se incluyen visualizaciones interactivas utilizando herramientas como pyplot y squarify, para permitir a los usuarios observar los datos de manera gráfica.



Conclusión

1

Análisis de Datos

El proyecto se centró en el análisis de datos recopilados sobre características de automóviles, con el objetivo de aplicar los conocimientos adquiridos a lo largo de la carrera.

2

Tratamiento de Datos

Se realizó un proceso de "curación" del dataset, tratando los datos ausentes y en formato incorrecto, para poder realizar un análisis efectivo.

3

Visualización de Datos

Se generaron gráficos utilizando la librería matplotlib, lo que permitió tener un abordaje visual de los datos y entenderlos mejor.

Agradecimientos

Queremos agradecer a los docentes Nahuel Pratta y Marcos Ugarte.

Disfrutamos de haber aplicado los conocimientos adquiridos en la carrera y esperamos continuar aprendiendo y creciendo.

