

INFORME DE ANÁLISIS

Predicción de Datos en el Mercado Automotor Proyecto Integrador



Durante proceso de limpieza de datos se aplicaron acciones significativas en el conjunto de datos '[original_dataset.csv](#)' para garantizar la calidad y la preparación óptima de los datos. Estos resultados, junto con las métricas de memoria obtenidas, establecen una base sólida para análisis y modelos predictivos futuros, destacando la eficiencia del conjunto de datos en términos de tamaño de memoria. El conjunto de datos limpiado, ahora con 28 variables, se exportó como '[cleaned_dataset.csv](#)', listo para análisis y modelado subsiguientes. Los resultados claves incluyen:

Tamaño del Conjunto de Datos:

Número de variables: 26

Número de observaciones: 205

Características Generales:

Ausencia de filas duplicadas y valores nulos en todas las columnas.

Columnas discretas: 8

Columnas continuas: 8

Catégoricas: 10

Optimización de la Memoria:

Tamaño total en memoria: 42.57 KiB

Tamaño promedio de registro en memoria: 0.21 B

Estos valores indican una eficiencia notable en el uso de memoria, crucial para análisis de datos eficientes y rendimiento.

Normalización y Transformación de Datos:

Normalización de nombres de columnas.

Mapeo de valores a las columnas 'door_number', 'cylinder_number', 'symboling'.

Agregadas las variables 'calificacion_riesgo' (Extremadamente Riesgoso, Riesgoso, Muy Riesgoso, Neutral, Poco Seguro, Seguro) y 'clasificacion_precio' (1= valor por debajo de la mediana, 0= si está por encima), permitiéndonos clasificar los vehículos como baratos o caros.

Redondeo del valor en columna 'price' y convertida a enteros.

Exploración de Outliers:

Identificación visual de outliers, los cuales se retuvieron debido al tamaño relativamente pequeño del conjunto de datos y su posible relevancia en futuros análisis.