

Proyecto integrador del módulo

Analista de Datos

Docentes:

- Marcos Ugarte
- Nahuel Pratta

Alumnos integrantes:

- Christian Nüesch

Origen de los datos:

<https://www.kaggle.com/datasets/ahmettalhabektas/argentina-car-prices>

Este dataset comprende varias características de automóviles ordenadas en diferentes columnas, obtenidas mediante peinado web (web scraping) de un sitio web en Argentina a principios de enero de 2023.

Descripción de cada columna:

(Los precios son para enero del 2023)

- Precio:
 - El precio del auto.
 - Tipo de dato: entero.
 - Valores posibles de la columna: múltiples valores numéricos enteros.
- Marca:
 - La marca del auto.

- Tipo de dato: cadena de caracteres.
- Valores posibles de la columna: "Audi", "Baic", "BMW", "Chery", "Chevrolet", "Citroën", "Dodge", "DS", "Fiat", "Ford", "Honda", "Hyundai", "Jeep", "Kia", "Mercedes-Benz", "Mini", "Mitsubishi", "Nissan", "Peugeot", "Porsche", "RAM", "Renault", "Subaru", "Suzuki", "Toyota", "Volkswagen", "Volvo".
- Modelo:
 - El modelo del auto.
 - Tipo de dato: cadena de caracteres.
 - Valores posibles de la columna: múltiples valores de cadena de caracteres.
- Año:
 - El año del auto.
 - Tipo de dato: entero.
 - Valores posibles de la columna: "1995", "1996", "1997", "2000", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021", "2022".
- Color:
 - El color del auto.
 - Tipo de dato: cadena de caracteres.
 - Valores posibles de la columna: "Azul", "Beige", "Blanco", "Celeste", "Dorado", "Gris", "Gris oscuro", "Marrón", "Naranja", "Negro", "Plateado", "Rojo", "Verde", "Violeta".
- Combustible:
 - El tipo de combustible que utiliza.
 - Tipo de dato: cadena de caracteres.
 - Valores posibles de la columna: "Diésel", "Híbrido/Nafta", "Nafta", "Nafta/GNC".
- Puertas:
 - La cantidad de puertas del auto.
 - Tipo de dato: entero.
 - Valores posibles de la columna: "2", "4", "5".
- Caja:
 - El tipo de caja de cambios.

- Tipo de dato: cadena de caracteres.
- Valores posibles de la columna: "Automática", "Manual".
- Motor:
 - El tamaño del motor (en litros).
 - Tipo de dato:
 - Valores posibles de la columna: múltiples valores numéricos con un decimal.
- Carrocería:
 - El tipo de carrocería.
 - Tipo de dato: cadena de caracteres.
 - Valores posibles de la columna: "Coupé", "Crossover", "Furgón", "Hatchback", "Minivan", "Monovolumen", "Pick-Up", "Rural", "Sedán", "SUV".
- Kilómetros:
 - El kilometraje del auto.
 - Tipo de dato: entero.
 - Valores posibles de la columna: múltiples valores numéricos enteros.
- Moneda:
 - La moneda en que se cotiza.
 - Tipo de dato: cadena de caracteres.
 - Valores posibles de la columna: "dólares", "pesos".

Estado de los datos

Los datos están en RAW (o datos crudos), lo que significa que están en su forma original, sin ningún tipo de procesamiento, limpieza o transformación. Estos datos fueron recogidos directamente de la fuente y no han sido manipulados para eliminar errores, duplicados, valores faltantes, etc.

No se le realizaron operaciones de limpieza o transformación, por lo tanto contienen errores, valores atípicos, valores ausentes y algunas inconsistencias.

Los datos están en distintos formatos (texto y diferentes tipos de números) y contienen toda la información capturada de la fuente, lo que los hace útiles para la práctica y el aprendizaje.