EDA Used Cars Dataset

Abstract

The goal of this project was to EDA to thinking about While many parameters influence a car buyer, the decision always involves *emotion*. The first thing to consider before taking any decision is to have as much good quality data as possible. We have then to perform analysis and extract information's.
In this EDA project , I study a public dataset available on KAGGLE. The dataset is scraped from **Craiglist**, a very popular classified ads in the US, and north America. It contains information's from around 450k car classified ads. It is updated every 18 months and contains a fair number of properties for each transaction (26 columns).

Design

This project was analyzed by reviewing the most used cars in United State, and the factors that's go into making that decision. There is many things consider to help to make a decision the car's mileage, condition, manufacturer ,type, price etc.

The result were presented in graphic images that easily simulate this information's.

Data

The dataset contains 26 features, 426880 unique values.

A few feature highlights include: id, url, region, region url, price, year, Manufacturer, model, condition, cylinders, fuel, odometer, title status, transmission, vin, drive, size, type, paint color, image url, description, country, state, lat, long, posting date.
number pages, rating count, and text reviews count.

Algorithms

1- Read Data
2- Check if there is missing values (for numerical calculate the mean, and for categorical the mode).
3- Transformation (Data transformation is a method to change the structure, format, and value of the data so that our data become more accessible for analysis.)
4- Visualization:

- outlier detection in numerical
- show some Categorical features
- Answer multiple questions:
  - **-The most common manufacturers in dataset based on number of cars?**
  - **-The Highest Price Based on Type Car?**
  - **-The most transmission type used based on manufacturer?**
  - **-The type of cylinders based on price of Cars.**
  - **- Price of type of cars for each type of fuel.**

Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting
- Tableau for interactive visualizations

Communications

5 Most Highest Price Based on Type of Cars

The type of cylinders based on price of Cars

Price of type of cars for each type of fuel

The type of cylinders based on price of Cars