Analysis of Automotive Industry

Dataset:

The dataset used for Analysis is taken from the Source: 1985 Model Import Car and Truck Specifications, 1985 Ward's Automotive Yearbook.

Data is having Data has 26 Attributes, it has missing values as '?', Datatypes of Column: normalized-losses, Horsepower & Price are Object they should be an integer.
This data set consists Specification Details, Risk Rating & relative average loss payment per insured vehicle year.

Objective of this Analysis to get idea on:

- Which factors affecting the price?
- Which type of vehicles are mostly produced?
- What is the risk level of various models?
- Which are the Various Manufacturers producing the Vehicles?

Tools & Libraries:

Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks.

Data Description:

The dataset contains the following Columns:

- Symboling: - Risk Level of Vehicle, varies from -3 (Low) to +3 (High)
- Normalized-losses: - Relative average loss payment per insured vehicle year
- Make: - Manufacturer Name
- Fuel-type: - Fuel type used for Vehicle
- Aspiration: - Engine Aspiration (i.e., Standard & Turbo)
- Num-of-doors: - Number of Doors available to Vehicle
- Body-style: - Type of Vehicle Body
- Drive-wheels: - Drive wheel type of Vehicle (FWD, RWD, 4WD)
- Engine-location: - Where engine is located in vehicle
- Wheel-base: - Size Wheel base
- Length: - Length of Vehicle in cm
- Width: - Width of Vehicle in cm
- Height: - Height of Vehicle in cm
- Curb-weight: - Curb Weight of Vehicle in kg
- Engine-type: - Type of engine
- Num-of-cylinders: - Cylinders used in engine
- Engine-size: - Size of Engine in cm
- Fuel-system: - Fuel System used
- Bore: - Engine parameter to calculate Displacement
- Stroke: - Engine parameter to calculate Displacement
- Compression-ratio: - Compression ratio of engine
- Horsepower: - Engine Horsepower

- Peak-rpm: - Revolution per minute at peak
- City-mpg: - Miles per gallon in city
- Highway-mpg: - Miles per gallon on Highway
- Price: - Price of Vehicle

Data Cleaning:

I made the following changes:

- Removed “?” form the Dataset
- Converted Columns normalized-losses, price, & horsepower in to Integer
- Renamed Columns from 'normalized-losses' to 'loss' & 'body-style' to 'body'.

EDA:

I looked at the different-different trends of the data and below are the few highlights of analysis.

1.Dataset has Outliers in Normalized Losses, Engine Size, Horsepower & Price.
2.Jaguar and Mercedes-benz having highest engine Size 258 & 234 Respectively.
3.Volksvagon & Nissan having more relative average loss, 256 & 231 respectively.
4.Jaguar and Mercedes-benz having highest price.
5.Toyoto has produced most Vehicles.
6.Volvo has average Low Risk=-1.27& Porsche has average highest risk=3.
7.BMW has highest relative average loss payment per insured vehicle year = 190
8.Jaguar has highest average price 32.25K.
9.Diesel Vehicles are Mostly produced in that year.
10.Diesel Vehicle has highest Price compare to gas vehicle.
11.Convertible Vehicles has highest horsepower & Price compare to hatchback, wagon, hardtop and sedan
12.Front Wheel Drive Vehicle are mostly produced compare to Rare Wheel and Four Wheel Drive Vehicles