Data Visualization
Assignment 1

Team Aspire

- Sneh Patel – C0869367
- Malav Shah – C0870066
- Dhruvrajsinh Chavda - C0867457
- Nandiniben Patel – C0869773

Abstract *The dataset can be used by car manufacturers and dealerships to identify which features have the greatest impact on car prices, as well as to provide more accurate pricing estimates to customers.:*

***Keywords*—.** Car features and MSRP (Manufacturer's Suggested Retail Price)

❖ INTRODUCTION

Exploratory Data Analysis (EDA) is a process of analyzing and summarizing data to gain insights and understanding of the data. In this report, we will perform EDA on the dataset of car features and MSRP (Manufacturer's Suggested Retail Price)

❖ OBJECTIVES

The objectives of analyzing Car Features and MSRP data may vary depending on the specific needs of the analysis. common objectives are:

- To identify the relationship between different car features and their impact on the MSRP.
- To understand the distribution of MSRP in the dataset and identify any outliers or extreme values.
- To identify any missing or inconsistent data and clean the dataset for further analysis.
- To identify the most popular car features and their impact on the MSRP.
- To visualize the data using graphs and charts to gain insights and communicate the findings effectively.
- To develop a model that can predict the MSRP based on different car features.
- To identify any trends or patterns in the data that can be used to make predictions about future car prices.

❖ ABOUT DATA

- **Make**: The make of the car.
- **Model**: The model of the car.
- **Year**: The year of the car.
- **Engine Fuel Type**: The type of fuel the car uses.
- **Engine HP**: The horsepower of the car.
- **Engine Cylinders**: The number of cylinders in the car's engine.
- **Transmission Type**: The type of transmission the car has.
- **Driven Wheels:** The type of wheels the car uses.
- **Number of Doors:** The number of doors the car has
- **Market Category**: The category the car belongs to in the market.
- **Vehicle Size**: The size of the car.
- **Vehicle Style**: The style of the car.
- **Highway MPG**: The miles per gallon the car can achieve on the highway.
- **City MPG**: The miles per gallon the car can achieve in the city.
- **Popularity:** The popularity of the car.
- **MSRP**: The Manufacturer's Suggested Retail Price of the car.

❖ Steps Involved:

- **Data Cleaning**: The first step in the EDA process is to clean the data set.

  We checked for any missing values.
  Engine Fuel Type - 3
  Engine HP – 69
  Engine Cylinders- 30
  Market Category-3742

- **Data Preprocessing**: The collected data will be preprocessed to remove any outliers, missing values, and errors. The data will also be normalized to ensure that all variables are on the same scale.

- **Data visualization**: Data visualization is the representation of data in a visual or graphical format. It is a way to communicate complex information and patterns in an intuitive and easily understood manner. By using visual elements such as charts, graphs, and maps, data visualization can help users to understand patterns and relationships quickly and effectively in data that would be difficult to discern from raw data.

- **Feature Selection:** Feature selection is a process of selecting the most relevant and important features from a data set to improve the accuracy and efficiency of a model.

- In the case of the Car Features and MSRP data set, we ca perform feature selection to identify the most importar car features that have the greatest impact on the MSRP.
- This will involve performing feature selection usin techniques such as correlation analysis and featur importance ranking.

❖ Conclusion

- In this report, we performed EDA on the dataset of car features and MSRP.
- We found that the distribution of MSRP is right skewed, with most of the cars having a lower MSRP and few cars having a higher MSRP.
- We also found that there is a positive correlation between Engine HP and MSRP, with cars having a higher Engine HP having a higher MSRP.
- We also found that the Luxury, High-Performance, and Exotic market categories have a higher median MSRP compared to other market categories.

❖ Insights

Share the insights from the data that you have achieve from the EDA exercise.

- Here are some possible insights that could be gained from an EDA exercise on car features and MSRP:
  - **Correlation between car features and price:** By analyzing the correlation between different car features and the MSRP, we can determine which features have the greatest impact on the price of a car. For example, we might find that luxury features like leather seats, advanced infotainment systems, and high-end audio systems are strongly correlated with higher prices.
  - **Distribution of car prices**: By examining the distribution of car prices in the dataset, we can get a sense of the range of prices and the most common price points. This can help us understand the market segment that the cars in the dataset are targeting.
  - **Relationship between car brand and price:** We might find that certain car brands are associated with higher prices, while others are more affordable.
  - The market positioning of different car manufacturers and their target customers.
  - **Impact of fuel efficiency on price**: We might find that cars with higher fuel efficiency tend to have higher prices. This could be because fuel-efficient cars are often associated with advanced technologies that can be expensive to produce.
  - Overall, an EDA exercise on car features and MSRP can provide valuable insights into the factors that drive the price of a car and help us understand the competitive landscape of the automotive industry.