

Assignment 1: Visualization

7PAM2000 Applied Data Science 1

JEWEL GEORGE

MSc Data Science

Student ID: 22048634

Repository Link: <https://github.com/jewelgeorge/ADS-Assignment1>

INTRODUCTION

For the purpose of visualization of data, three types of visualization methods are used here. For this the datasets are taken from an open-source database. In each of the sections, details of every data are included. The datasets are read by using 'pandas' function and all of the graphs for visualization are plotted using 'pyplot' function.

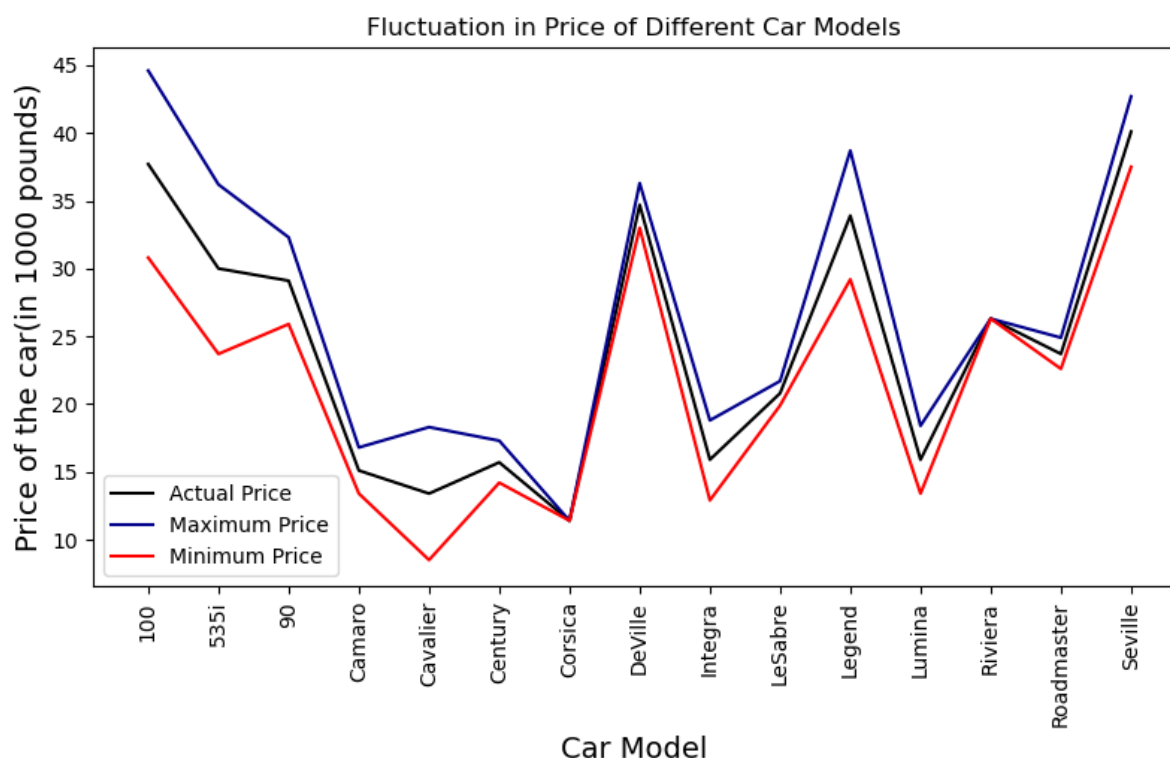
DATASET

The dataset that I have used is taken from a free open-source database named 'Kaggle'. After examining the dataset clearly, I have found that it contains information of different types of car models. Since, the dataset was huge to plot for simpler graphs, I have chosen only few relevant data for each of the plots by filtering the dataset using 'pandas'. Thus, the filtered dataset includes the vehicles (cars) minimum price, actual price, maximum price, type, mileage per gallons of fuel used by the cars both in cities and highways separately.

Dataset is taken from the below given link: -

<https://www.kaggle.com/datasets/yashpaloswal/cars93?select=Cars93.csv>

Visualization 1: LINE PLOT (FLUCTUATION IN PRICE OF DIFFERENT CAR MODELS)



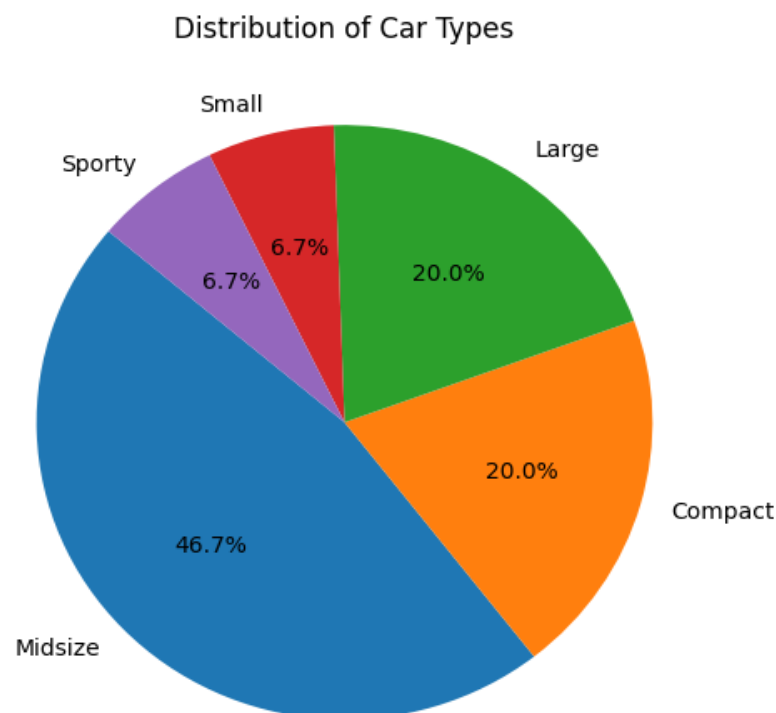
The visualization method used here is Line Plot.

Since, few data are possible to plot in the x and y axis, the first fifteen car models are selected using a data frame named car_models, to which the dataset used for visualization is imported. The first fifteen car models are plotted on the x axis and the price of the cars are plotted along the y axis of the line graph.

Three lines having different colours are used to plot the graph. The black coloured line represents the actual price, the blue coloured line represents the maximum price and the red coloured line represents the minimum price of the fifteen car models.

From the graph, it is clear that the actual, maximum and minimum price of car models named 'Corsica' and 'Rivera' always stayed the same at 12000 and 26000 pounds respectively, whereas, the price of all other cars fluctuated from their actual price. The car model '100' reached the top maximum price of around 45,000 pounds and the car model 'cavalier' hits the all-time lowest price of 8000 pounds. However, the maximum and minimum price of the car models, 'Camaro', 'LeSabre', and DeVille were always stayed similar to their actual prices. These are the conclusions that can be drawn from the above given line graph.

Visualization 2: PIE CHART (DISTRIBUTION OF CAR TYPES)



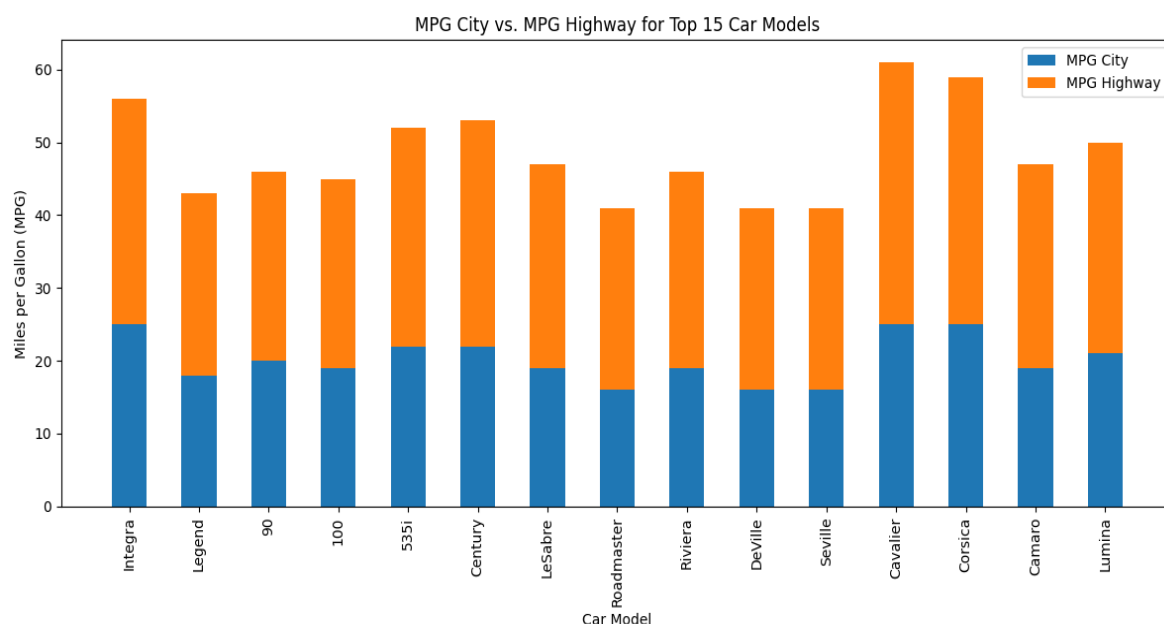
The visualization method used here is a pie chart.

The above given pie chart illustrates a diverse representation of car types in the dataset. The same data set is used, but the distribution of car types is shown using the data of the 'type' of cars among the first fifteen unique models. The type of car includes 'Sporty', 'Small', 'Large', 'Compact' and 'Midsize'.

From the pie chart it is clear that 'Midsize' cars constitute the largest portion of the car types, making up 46.7% of the distribution. This suggests that a significant number of the cars in the dataset fall into the midsize category. Both 'Large' and 'Compact' cars each contribute 20% to the distribution. This indicates a balanced representation of these two car types in the dataset. 'Sport' and 'Small' cars have smaller percentages, each comprising 6.7% of the distribution. This suggests that these types of cars are less common in the dataset compared to 'Midsize', 'Large', and 'Compact' cars.

In summary, the pie chart provides a visual representation of the distribution of car types, highlighting the prevalence of midsize cars and a relatively balanced distribution among 'Large', 'Compact', 'Sport', and 'Small' car types.

Visualization 3: STACKED BAR GRAPH (MILAGE PER GALLONS OF FUEL USED BY DIFFERENT CAR MODELS)



The third visualization method used for visualization is a stacked bar graph.

A stacked bar graph is plotted using the car model of the top fifteen cars from the same dataset along the x-axis and (MPG) "Miles Per Gallon" along the y-axis. The blue bars represent the mileage per gallon in cities and the orange graph represents the mileage per gallon while driving through highways. Thus, the graph provides an insight into the fuel efficiency of different car models in both city and highway conditions.

In Conclusion, cars with higher bars for both "MPG city" and "MPG highway" are considered more fuel-efficient, as they cover more Miles per Gallon of fuel in both driving scenarios. In this case the car model 'Cavalier' has the highest fuel efficiency with 25 MPG in city and 36 MPG while driving through highways. Similarly, cars with lower bars for both "MPG City" and "MPG Highway" are considered less fuel efficient as they cover less Miles Per Gallon of fuel in both driving scenarios. In this case the car models 'DeVille' and 'Seville' has the lowest fuel efficiency, with both of the models having 16 MPG in cities and 25 MPG on highways.

CONCLUSION

In this assignment, three visualisation methods are used, those are line plot, pie chart and stacked bar graph.

These three visualisation methods offer three different perspectives of the data set taken. That is, the line plot depicts fluctuation in price of the car models, the pie chart depicts the distribution of type of car models and the stacked bar graph represents the mile per gallon of fuel used by each of the car models in both cities and highways.