Data Preparation

Following systems have been pursued for cleaning the information as per task detail:

1. **Evacuating repetitive blank areas**:

So as to evacuate the main and trailing blank areas in the estimations of properties make fuel type, goal, number of entryways, body style, drive wheels, Engine area and fuel framework following advances are pursued.

a. As a matter of first importance, utilizing the value_counts(), the quantity of estimations of unmistakable sorts are discovered comparing to each trait

b. Also, utilizing the strip() on qualities containing the excess blank areas, all driving and trailing void areas are expelled for every one of the previously mentioned properties.

2. **Fixing case**:
So as to fix the instance of the estimations of properties make, fuel type, aspiration, number of doors, body style, drive wheels, Engine location and fuel system the accompanying advances were pursued.

a. As a matter of first importance, utilizing the value_counts(),the number of estimations of unmistakable sorts were discovered relating to each trait.

b. Also, utilizing the lower() on all qualities, the case explicit consistency was kept up.

3. **Fixing errors:**

So as to fix the errors in the estimations of characteristics make, aspiration, number of doors, drive wheels and fuel system Engine location and fuel system the accompanying advances are pursued.

a. Above all else, utilizing the value_counts(),the number of estimations of particular kinds are discovered relating to each quality.

b. Besides, the particular grammatical mistakes saw in each trait were fixed exclusively utilizing a cluster of if-elif conditions.

4**. Handling missing values and outliers:**

To fix the missing qualities saw in different numerical characteristics, they are supplanted by - 1 initially utilizing fillna(). At that point the anomalies are identified by watching the histograms plotted for highway mileage, stroke, width and length.

a. For fixing the exceptions in horsepower, every one of the qualities that are - 1 and more prominent than 225 are supplanted with the middle of the remainder of the qualities so as to perform proper information elucidation and investigation.

b. For fixing the exceptions in stroke every one of the qualities that are - 1 or more prominent than 3.7 or under 2.5 are supplanted with the middle of the remainder of the qualities so as to perform proper information translation and examination.

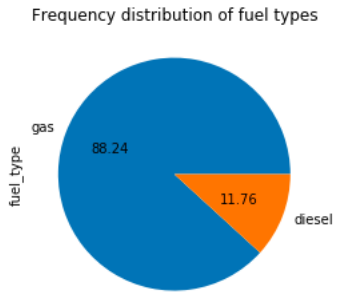
c. For fixing the exceptions in width every one of the qualities that are - 1 or under 63 are supplanted with the middle of the remainder of the qualities so as to perform proper information understanding and investigation.

d. For fixing the anomalies length every one of the qualities that are - 1 or more prominent than 155 are supplanted with the middle of the remainder of the qualities so as to perform proper information elucidation and investigation.

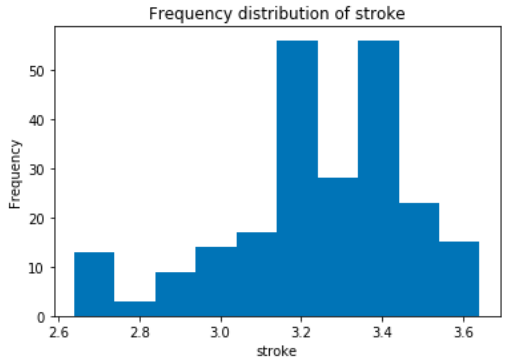
e. For fixing the anomalies in wheel base every one of the qualities that are - 1 or more prominent than 300 are supplanted with the middle of the remainder of the qualities so as to perform suitable information understanding and investigation.

1. **plotted for the below provided three attributes**:

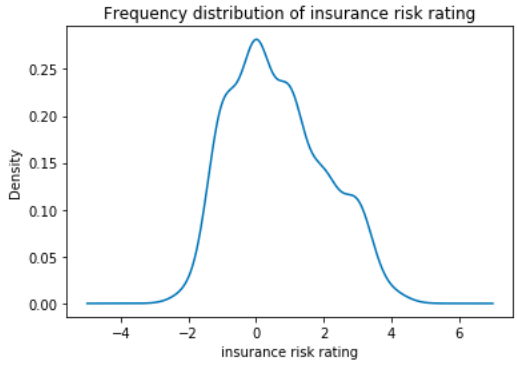
a) For nominal attribute fuel type is chosen to analyse which is more efficient amongst diesel and gas. As evident from the pie chart below gas is much more efficient and popular as a fuel compared to diesel.



b) For numerical attribute stroke is chosen to analyse the average stroke in any vehicle.

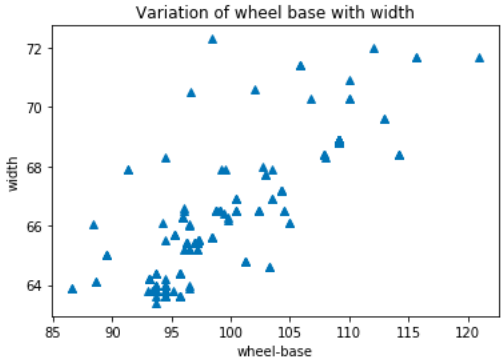


c) For ordinal attribute symboling is chosen to conclude what proportion of vehicles necessarily need insurance and what proportion of vehicles do not need insurance at all. As evident from the graph below most of the vehicles are neutral in terms of need of insurance.

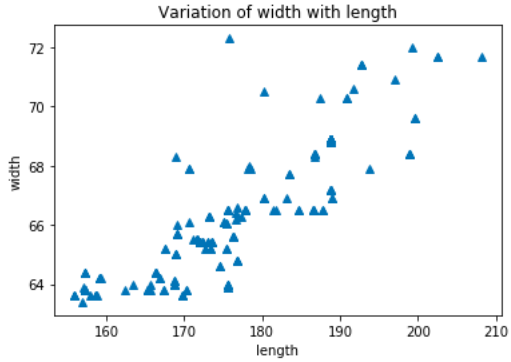


2. The graphs between these given three pairs of attributes:

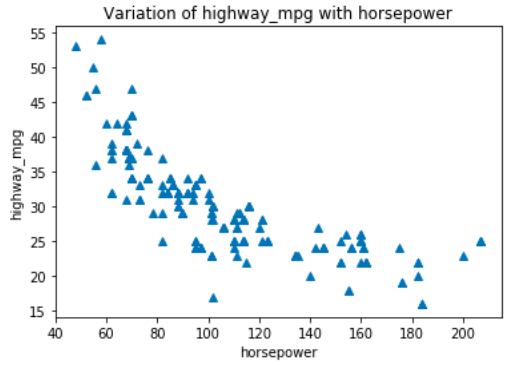
Firstly, the comparison between wheel base and width is made. As evident from the graph below there is fairly positive correlation between the two. As wheel base increases width of the vehicle also increases.



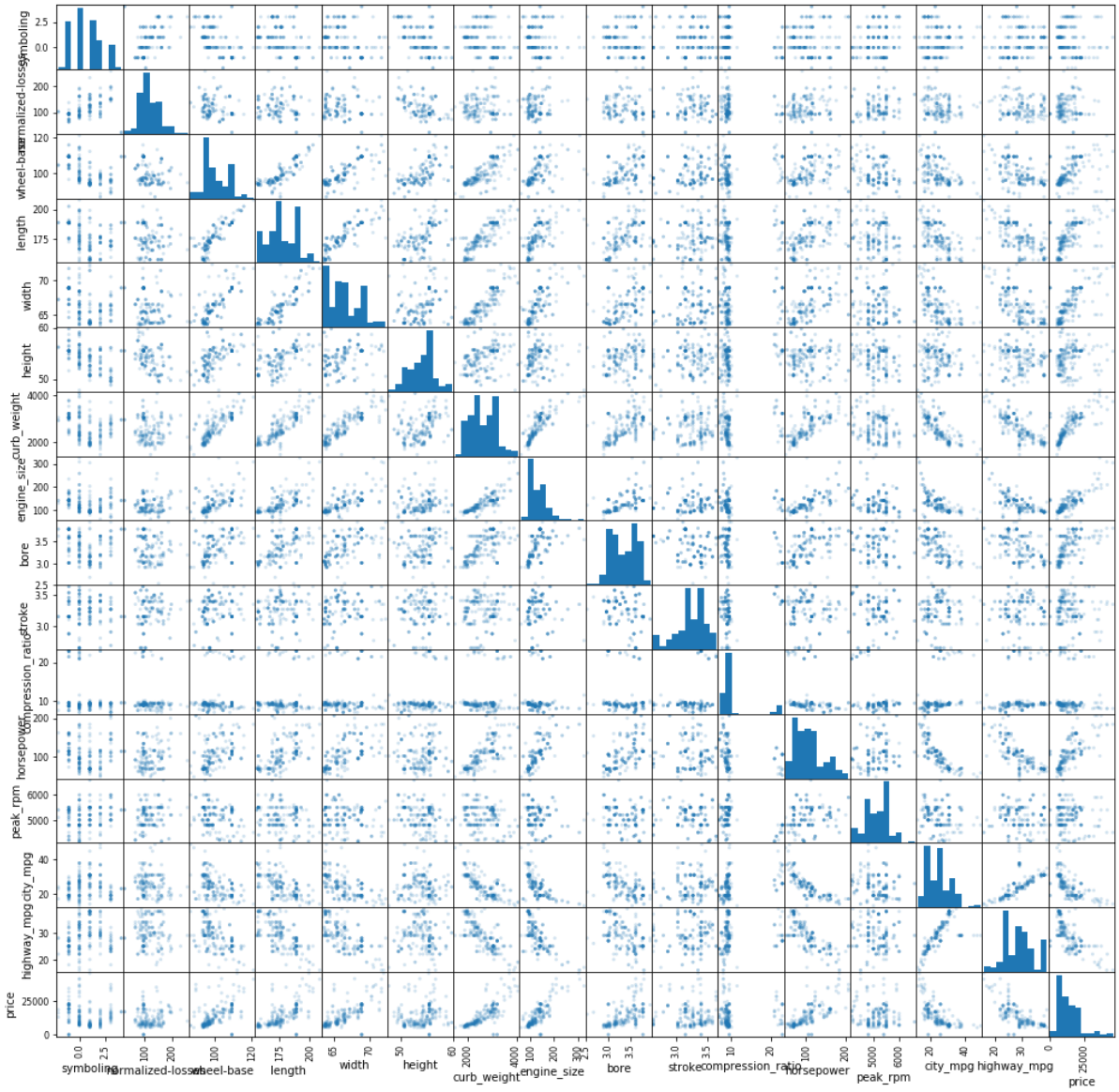
Secondly, the comparison between length and width is made. As evident from the graph below there is fairly positive correlation between the two. As the length of the vehicle increases width of the vehicle also increases.



Thirdly, the comparison between horsepower and highway mileage is made. As evident from the graph below there is fairly negative correlation between the two. As the horsepower of the vehicle increases highway mileage of the vehicle decreases.



3. **Scatter Matrix of all numerical attributes**:
The scatter matrix above shows how unique sets of attributes are identified with one another. As apparent from the scatter matrix beneath a large portion of the characteristics are feebly adversely connected to one another and others are negatively emphatically corresponded. Just few of these have zero relationship.


symboling
normalized-losses
wheel-base
length
width
height
curb_weight
engine_size
bore
stroke
compression_ratio
horsepower
peak_rpm
city_mpg
highway_mpg
price