Project Report Data Analytics 15CS30024

1)The data seems like a tabulation of distance covered by a car in a trip and average speed of the car in that trip. Arithmetic mean has a real significance to the data,since it shows what is the average distance covered by the car across all trips, and the average speed of the car across all trips.

Median might be a better measurement since it shows what is the median speed that the car achieves on a trip basis and what is the median distance the car travels on a trip basis.

2)No discrepancies seem to be present in the data.Box plots are compared in pdf file attached along with the report.

3)Categorization according to NOIR topology

#symboling : Ordinal

#normalized losses: Ratio

#make : Nominal

#fuel-type : Nominal

#aspiration : Nominal

#num-of-doors : Ordinal

#body-style : Nominal

#drive-wheels : Nominal

#engine-location : Nominal

#wheel-base : Ratio

#length : Ratio

#width : Ratio

#height : Ratio

#curb-weight : Ratio

#engine-type : Nominal

#num-of-cylinders: Ratio

#engine-size : Ordinal

#fuel-system : Nominal

#bore : Interval

#stroke : Interval

#compression-ratio: Ratio

#horsepower : Ratio

#peak-rpm : Ratio

#city-mpg : Ratio

#highway-mpg : Ratio

#price : Ratio

According to the density plots done, both peak-rpm and city-mpg follow normal distribution.

4) Done as asked

5)Done as asked

Tool used : R for plotting and analysing data in all questions