Data Cleaning and Wrangling

* Normalization: The very 1st operation aimed naming the data partitioned into different columns.
* Rename and replace operations performed helped change the column names and replace unnecessary values respectively.
* SCALING-To reduce disparities of values among columns it helps to scale the data to specific values so as to obtain desired output
* Simple Features Scaling
* Min-Max Scaling
* Z-scores Scaling

Exploratory Data Analysis

* BINNING: To sum the large data into smaller groups, here we binned the price column and grouped the price into 3 groups "Low", "Medium", "High".
* Creating Dummies: For a better understanding, we created dummies in the column “fuel-type” representing a proper partitioning in “diesel” and “gas” in “fuel-type”.
* Value\_Count: Value\_count is generated for different types of drive-wheels.

Visualization

* Box Plot: Box plot graph is plotted with price on x-axis and drive-wheels on y-axis.
  + Observation-The graph is divided in quatlite.
  + The median in each block separate higher and lower values.
  + The ”fwd” and “4wd” has highest and lowest variance respectively. “rwd” has minimum variance in price as compared to fwd.
* Scatter Plot.: “engine-size” is used as Predictor and “price” being the Target is used to plot a Scatter plot.
  + Observation-The Target “price” increases with increasing “engine-size” in linear manner with mathematical proportion.
* Group-by operation- Group-by function is applied to following columns “drive-wheels”, “bodystyle”, “price” grouping them together.
* PIVOT operation-Pivot operation is performed on drive wheels bodystyle and price columns giving a good representation of pivoted values.
* Color Bar Plot: Color map of Red and Blue color is used. Lowest to Medium range price is distributed from darkest to lightest shades of Red respectively, whereas Light blue to Dark Blue shows the distribution of Medium to Highest range of Price respectively.

-ANOVA

* Co-relation Statistics
* Regression plot
* Pearson Co-efficent
* Linear Regression