

Stats For Machine Learning

Prior to summary statistics load your data and identify and missing values or duplicates and drop them.

- **Identify the variable types in the given data set.**
- **Make a bar graph for manufacturer type**
- **Make a pie chart for showing vehicle type by percentage**
- **Calculate average Price of all cars by using correct measure of central Tendency**
- **Calculate the variance and standard deviation for numerical columns such as**
Sales_in_thousands, __year_resale_value, Price_in_thousands, Engine_size, Horsepower
- **For the same columns Calculate IQR and identify outliers in these columns**

`Q1 = df[column].quantile(0.25)`

`Q3 = df[column].quantile(0.75)`

`IQR = Q3 - Q1`

`outliers = df[(df[column] < Q1 - 1.5 * IQR) | (df[column] > Q3 + 1.5 * IQR)]`

- **Make Histograms for all Numerical columns and identify the distribution of type by calculating skewness;**

Note : If the skewness is 0, it indicates a perfectly symmetrical distribution.

If the skewness is negative, it means the distribution is skewed to the left (long left tail).

If the skewness is positive, it means the distribution is skewed to the right (long right tail).

Happy learning!