VIDYA ANALYTICS HACKATHON

**1.Importing Libraries**

Importing Necessary Libraries

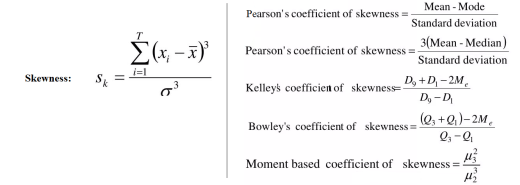
**2. Reading Data**

1.Reading in train and test data

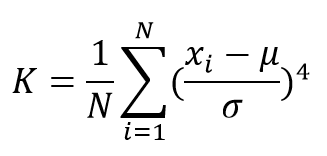
2. Inspecting Data(Checking our Data, Saving then dropping Id column, Checking Dataset's Dimensions)

**3. Skew and Kurtosis Analysis**

3.1 Skew Formula:



3.2 Kurtosis Formula:



Reasons a right-skewed predictive variable is bad:

. Mean greater than mode

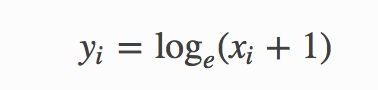
. Median greater than mode

. Mean is greater than median

This eventually can affect the performance of our modelling process, so we will decide to log-transform it in the next step.

3.3 Log-Transforming SalePrice

Log-Transforming SalePrice with np.log1p



**4 - Outliers**

4.1 Univariate Analysis

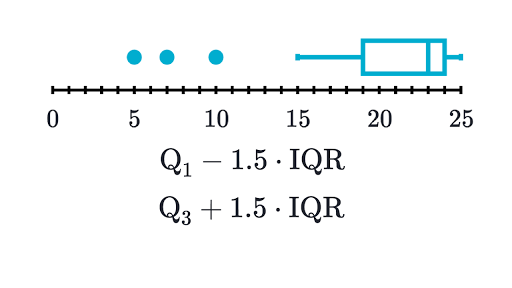
In the name 'Univariate Analysis' 'Uni' means one, and 'variate' means variable, meaning that Univariate analysis is analysis of one feature. This procedure basically tells us the distribution of each feature and information about its mean, median, and mode. This is our first step to detecting outliers.

4.2 Bivariate Analysis

Bivariate Analysis is a procedure but with two variables (hence the name 'Bi'), not one. This is our second step to detecting outliers.

4.3 Removing Outliers¶

After looking at the Visualizations of each numerical feature, we use the 1.5 IQR Rule in order to detect and remove outliers



It states that a data point is an outlier if,It is below the First Quadrant (Q1) subtracted by (1.5 x IQR),It is above the Third Quadrant (Q3) added by (1.5 x IQR)

4.4Checking for Missing Values

Combining train and test data in order to make imputing missing values easier, locating missing values

**5 - Feature Engineering**

5.1 One Hot encoding

Our dataset cannot run with categorical columns so we must One Hot Encode some columns in order to make them numerical.

**6 - High Skew Features**

Checking Skew(Create a new variable containing the dataset of only numerical features) and skew visualization.

6.1 Box-Cox Transformation(Perform Box-Cox Transformation on Skewed Features)

**7 - Modelling**

7.1 Importing Libraries for Modelling

7.2 Preparing Data for Modelling

7.3 Models

7.4 Viewing Model Performance

8.5 Stacking