**Assignment1 - Building a Good Training Set**

* Import the dataset - Diabetes.csv
* Do the missing value treatment
* Perform an exploratory data analysis with respect to the Outcome variable
* Check for outliers with appropriate plots
* Remove the outliers from the dataset
* Perform feature scaling for the features
* Find & List the most important 5 features to determine the Outcome variable
* Check with heatmap or correlation matrix to confirm the 5 selected features
* Perform PCA and generate two components for the 5 features selected
* Split the data into train and test (75:25) keeping the target variable as Outcome