**Summary**

First, we focused on **data cleansing** and **pre-processing**.

We found that many columns have null values that were more than 30%.
So we simply removed them as they won't be contributing to the analysis.

After that there were columns that contains 'Select' as one of there entry.

Now, select also stands for unknown as the ones with more than **30% of select** contribution we removed them.

After that we found that total visits and other numerical columns contain null values.
So we removed those rows.

There were only three numerical variables and rest were categorical variables.

For many categorical variables there was no variation in them so we simply removed them as they will not contribute into the analysis.

After that we did **Dummification of categorical values** and then checked out the variables with high correlation to reduce multicollinearity we dropped one set of features.

Then we broke the data into test train.

**Standardised the numerical columns**

Further we trained the model with rfe using top 15 variables and then upon selective elimination we removed   *Lead Source_Welingak Website* as it has very high p values.

We got the accuracy of 82% , sensitivity of 69% and Specificity of 88%

Then we trained the models after standardising its numerical columns  and we found that the Accuracy is 80%

Precsion is 78%

Recall is 69%