SUMMARY REPORT

The data given was cleaned up and data manipulation was done to derive insights. As a first step, missing values were identified in each column. In certain cases, columns with more than 40% missing values were dropped and in certain cases missing values were imputed with mode of the data in the column. Some interesting findings that were uncovered during exploratory data analysis(EDA) were the following.

A) Maximum number of leads are generated by Google and Direct traffic among lead sources

B) Conversion rate of reference leads and leads through welingak website is high

C) To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generating more leads from reference and welingak website.

D) API and Landing Page Submission bring higher number of leads as well as conversion from lead origin.

E) Lead Add Form has a very high conversion rate but count of leads are not very high.

F) Lead Import and Quick Add Form get very few leads.

G) In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

H) Leads spending more time on the website are more likely to be converted.

I) Website should be made more engaging to make leads spend more time.

After performing EDA, certain iterations were performed to identify variables which will effectively predict whether a conversion will happen or not. The data was spilt into training and test data. Based on the training data, predictions were made for test data and accuracy, specificity and sensitivity were measured. The final machine learning model that we built had a good level of accuracy, specificity and sensitivity (above 90%). The model can effectively predict the conversion status of the customers contacted.