Technical Summary:

- There were 37 columns in the original data set and around 9240 rows

- Removed and Imputed columns and values based on EDA

- Created dummy variables for the categorical type features

- Using RFE method selected 20 features from the original data

- Checked for p-value and VIF and dropped columns which showed high for both

- Final model was created with 16 variables for which below are the top 3 variables

    - Lead Origin - Lead Add Form

    - What is your current occupation - Working Professional

    - Total Time Spent on Website

- The Accuracy, Sensitivity, Recall, Precision was calculated, and it showed good results

Train Data

Accuracy	Sensitivity	Specificity	Precision	Recall
82.37%	82.56	82.25%	74.36%	82.56%

Test Data

Accuracy	Sensitivity	Specificity	Precision	Recall
82.90%	84.02	82.18%	74.97%	84.02%

Sensitivity for Train and Test Data indicates that the model is able to correctly identify more than 80% of positive instance. This shows the model is best suited for the given business scenario.

- Final Predictions were made, and model seems to be stable as the test and Train data shows very less differences.

Business Summary:

- 1. Marketing team should focus more on "working professionals" as they are more likely to get converted.

- 2. Marketing team should focus more on leads that spent time on website as they are more likely to get converted.

- 3. Marketing team should focus more on leads coming from Olark Chat as they are more likely to get converted.

- 4. Marketing team should skip calls to the leads who chose the option "Do not Email" as "yes" as they are not likely to get converted.

- 5. Marketing team should skip calls to the leads that have the Lead Origin as "Landing Page Submission" as they are not likely to get converted.

- 6. Marketing team should improve their social media ads. Leads from Face book & Referral sites are not getting converted