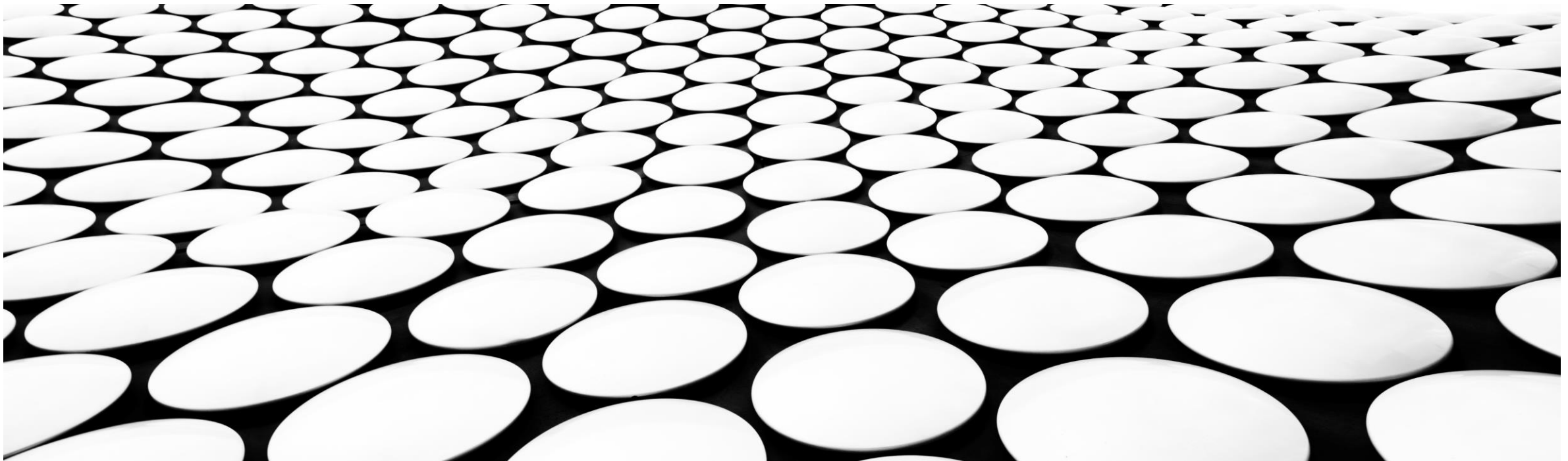# LEAD SCORING CASE STUDY

GROUP MEMBERS:
- NIKITA MUNDHRA
- PALAK GARG
- NAJEEB FAJANDAR

upGrad

PROBLEM STATEMENT:

❑ X Education sells online courses to industry professionals.

❑ The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead.

❑ Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

❑ However, the lead conversion rate is extremely poor.

❑ The typical lead conversion rate at X education is around 30%.

Business Objective:

❑ X Education would like identify the most promising leads, i.e. the leads that are most likely to convert into paying customers.

❑ The company requires a model to build wherein a lead score is assigned to each of the leads. The customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

❑ The CEO of the company has given a ballpark of the target lead conversion rate to be around 80%.

METHODOLOGY:
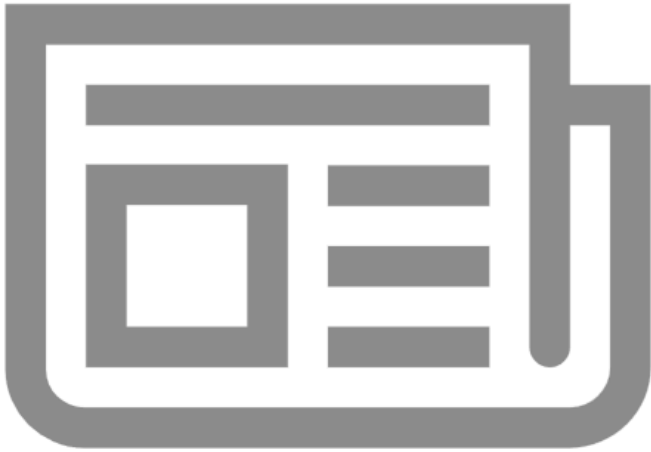
- Data cleaning and preparation

- Feature scaling and dummy variable collection

- Model evaluation

- Exploratory data analysis

- Logistic regression model building

- Conclusion and recommendation

PROBLEM SOLVING METHODOLOGY:

Data cleaning and preparation:
- Read and understand the data.
- Data cleaning. for analysis
- Exploratory data analysis
- Data preparation.

Splitting data and feature scaling:
- Splitting the data into train and test dataset.
- Feature scaling of variables.
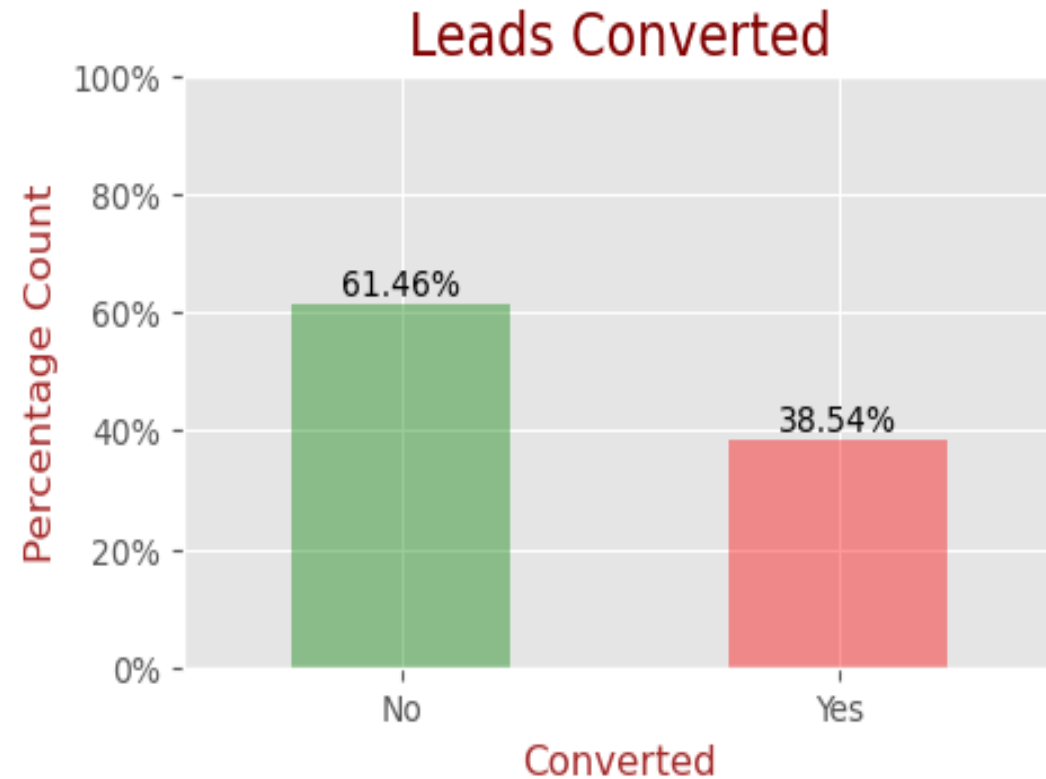
Model building:
- Recursive Feature Elimination.
- Model building.
- Plotting the ROC curve.
- Finding the optimal cutoff point.

Results:
- Making predictions on the test set.
- Determine lead score and if target final prediction is greater than 80%.
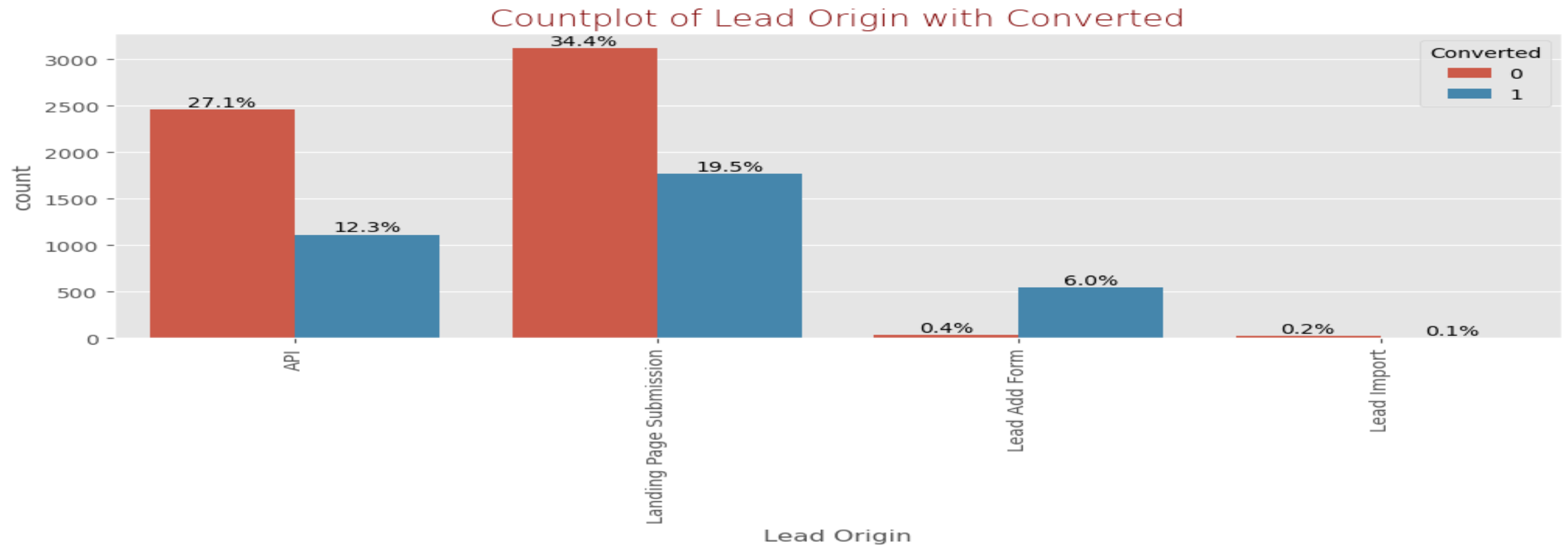- Evaluate final predication on test set.

EXPLORATORY DATA ANALSIS:

Leads Converted

Observations:
- Conversion rate is of 38.5%, meaning only 38.5% of the people have converted to leads.(Minority)
- - While 61.5% of the people didn't convert to leads. (Majority)

EXPLORATORY DATA ANALSIS:

Countplot of Lead Origin with Converted

Insights:
- Landing Page Submission identified 53.9% customers out of which 19.5% got converted, followed by API whcih identified 39.4% of the customers out of which 12.3% got converted.
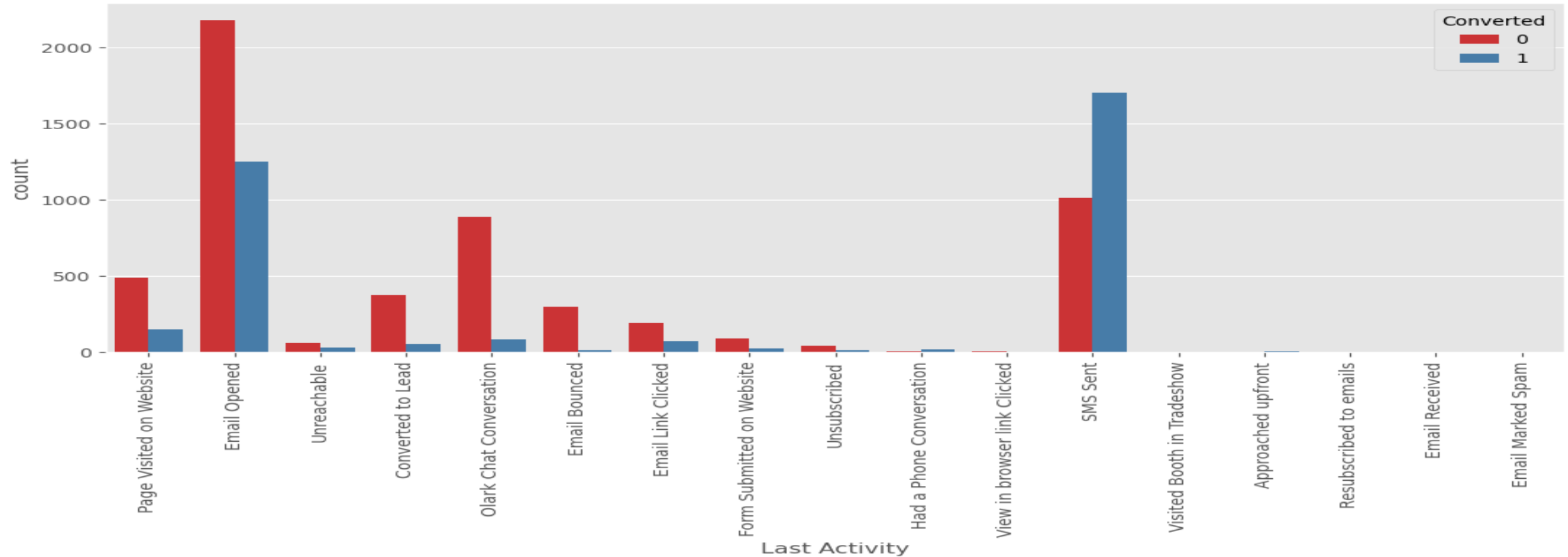- The Lead Add Form demonstrates a good conversion rate.

EXPLORATORY DATA ANALSIS:
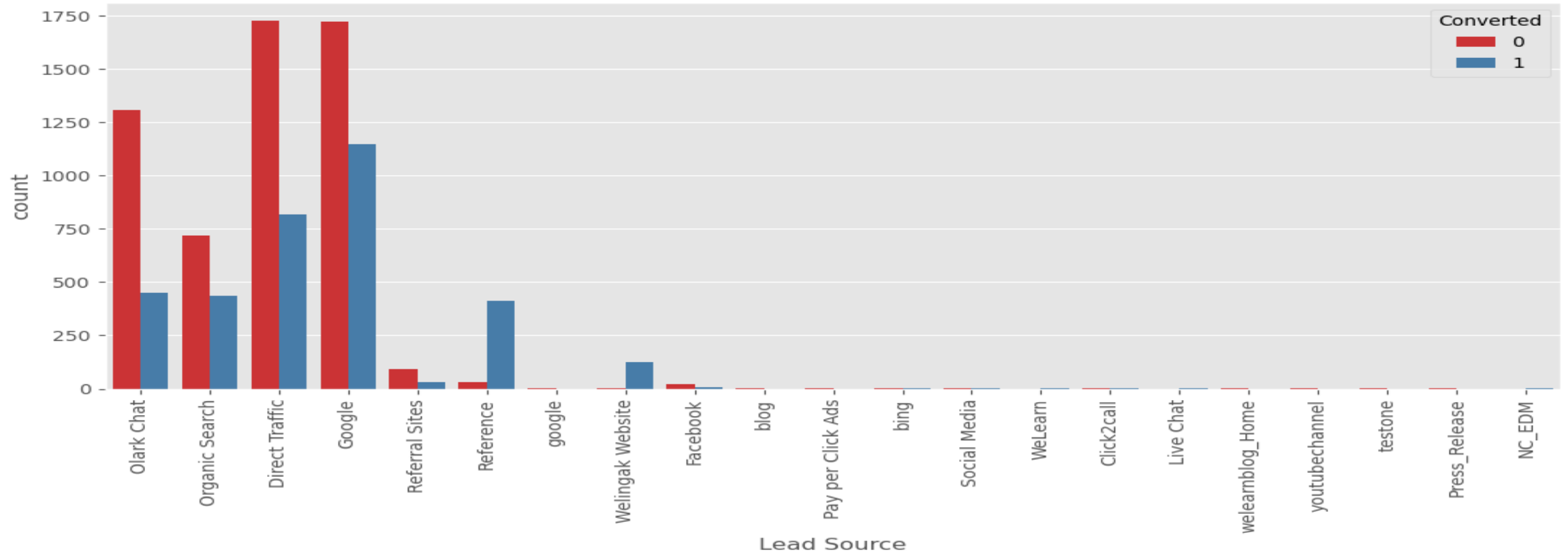
Insights:
- Majority of leads originate from Google, followed by Direct traffic and Olark chat.
- Reference and Welingak Website demonstrates a good conversion rate although the total number of leads is comparatively low.

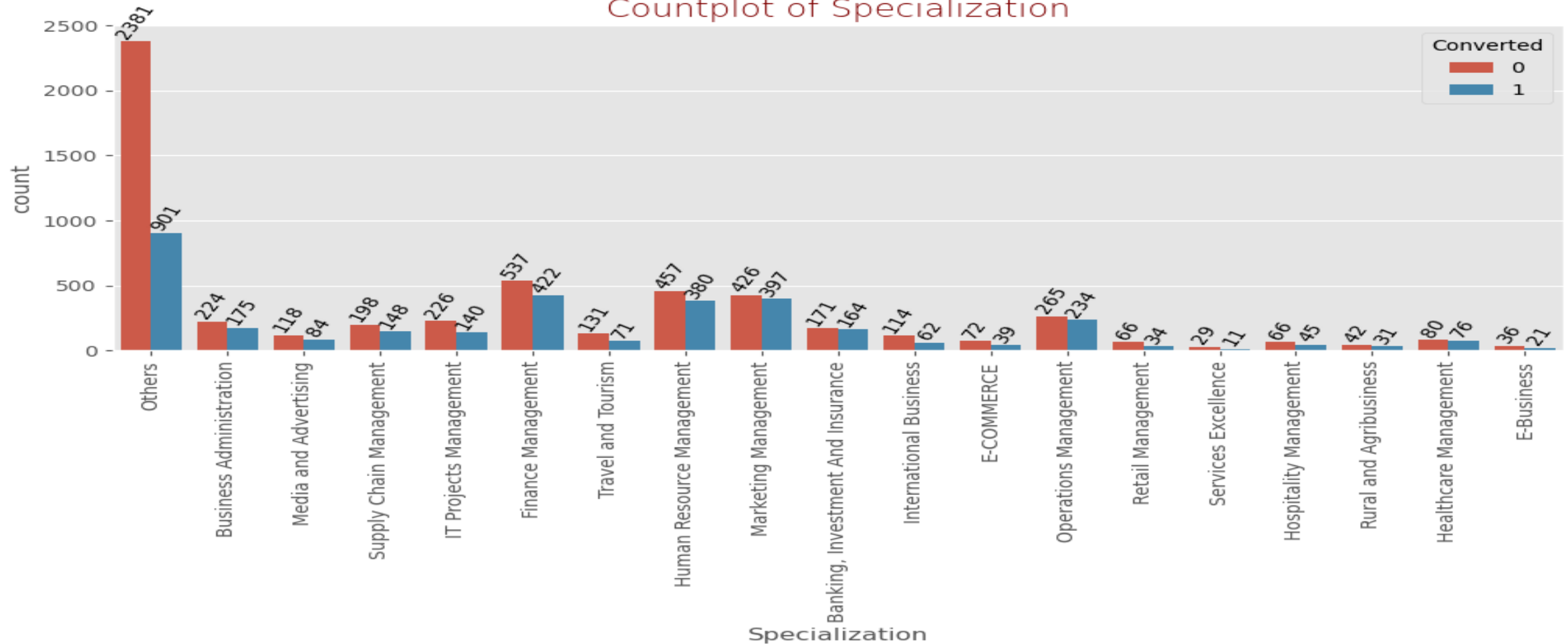# EXPLORATORY DATA ANALSIS:

Insights:
- Majority of the leads originated from 'SMS Sent'.
- The total number of leads from 'Email Opened' is highest but its conversion rate is comparatively less.
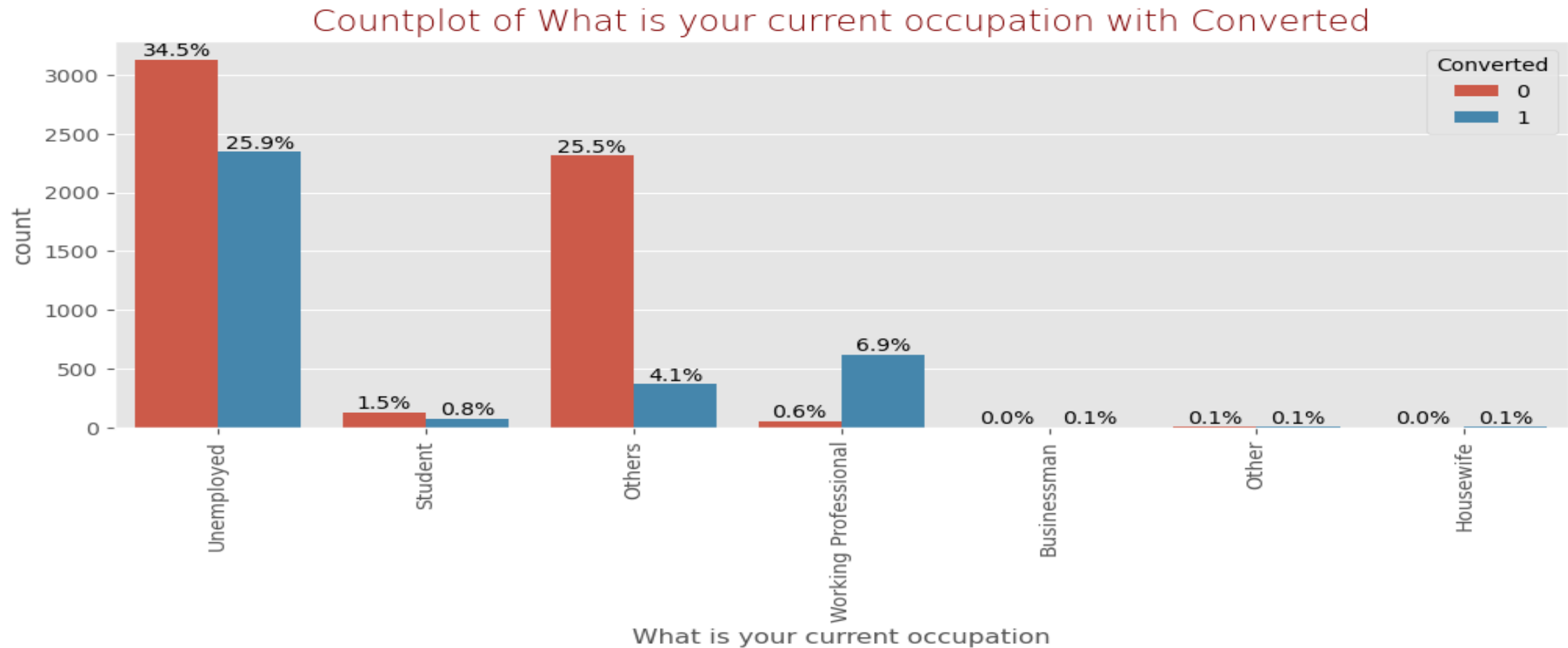
EXPLORATORY DATA ANALSIS:

Insights:
- Management have high number of leads and converted leads.

EXPLORATORY DATA ANALSIS:

Countplot of What is your current occupation with Converted

Insights:
- Unemployed customers are opting for online courses followed by Working Professional.
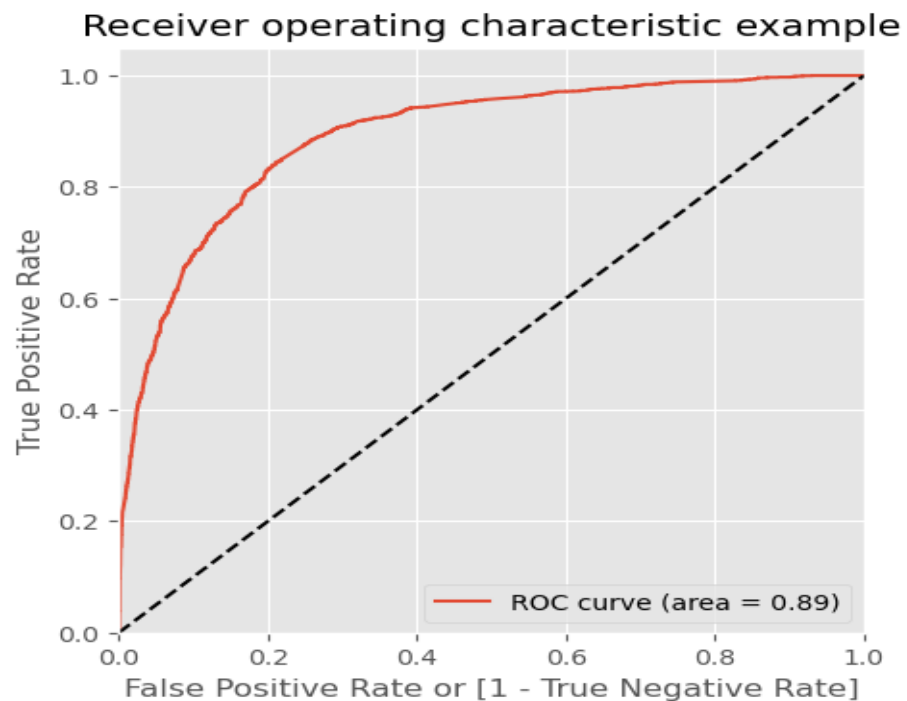
CORRELATION MATRIX

Total Time Spent on Website have a positive correlation with other variables.
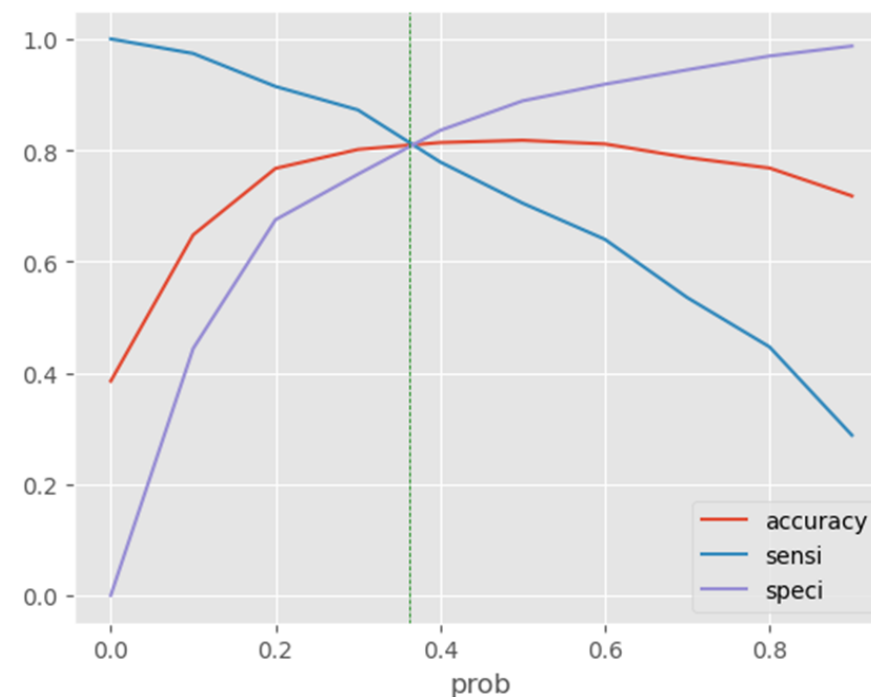
MODEL BUILDING:

ROC CURVE

By determining the area under the curve (AUC) of an ROC curve, that is, 0.89, it can be determined how good the model is. As the ROC curve is more towards the upper-left corner of the graph, it means that the model is very good.

OPTIMUM CUTOFF POINT

From the curve above, 0.34 is the optimum point to take it as a cutoff probability.

MODEL EVALUATION:

Comparing values obtained from Train and Test data:

TRAIN DATA:

- Accuracy : 80.98 %
- Sensitivity : 80.73 %
- Specificity : 81.13 %

TEST DATA:

- Accuracy : 81.67 %
- Sensitivity : 81.82 %
- Specificity : 81.57 %

Insights:

- Therefore, the goal of the target lead conversion rate to be around 80% is achieved.
- The accuracy values for both the train and test data are close to each other, with the test accuracy being slightly lower than the train accuracy, which is reasonable.
- The sensitivity and specificity values for both the train and test data are also reasonably close, which indicates that the model generalizes well and performs consistently across different datasets.

CONCLUSION AND RECOMMENDATIONS:

Focus on features with positive coefficients for targeted marketing strategies.

Develop strategies to attract high-quality leads from top-performing lead sources.

Engage working professionals with tailored messaging.

Optimize communication channels based on lead engagement impact.

More budget/spend can be done on Welingak Website in terms of advertising, etc.

Incentives/discounts for providing reference that convert to lead, encourage providing more references.

Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.

Analyze negative coefficients in specialization offerings.

Review landing page submission process for areas of improvement.