LEAD SCORING CASE STUDY

PROBLEM STATEMENT:

❖ X Education sells online courses to industry professionals.

❖ X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

❖ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

❖ If they successfully identify this set of leads, the lead conversion rate should go upas the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

BUSINESS OBJECTIVE:

❖ Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

❖ They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.
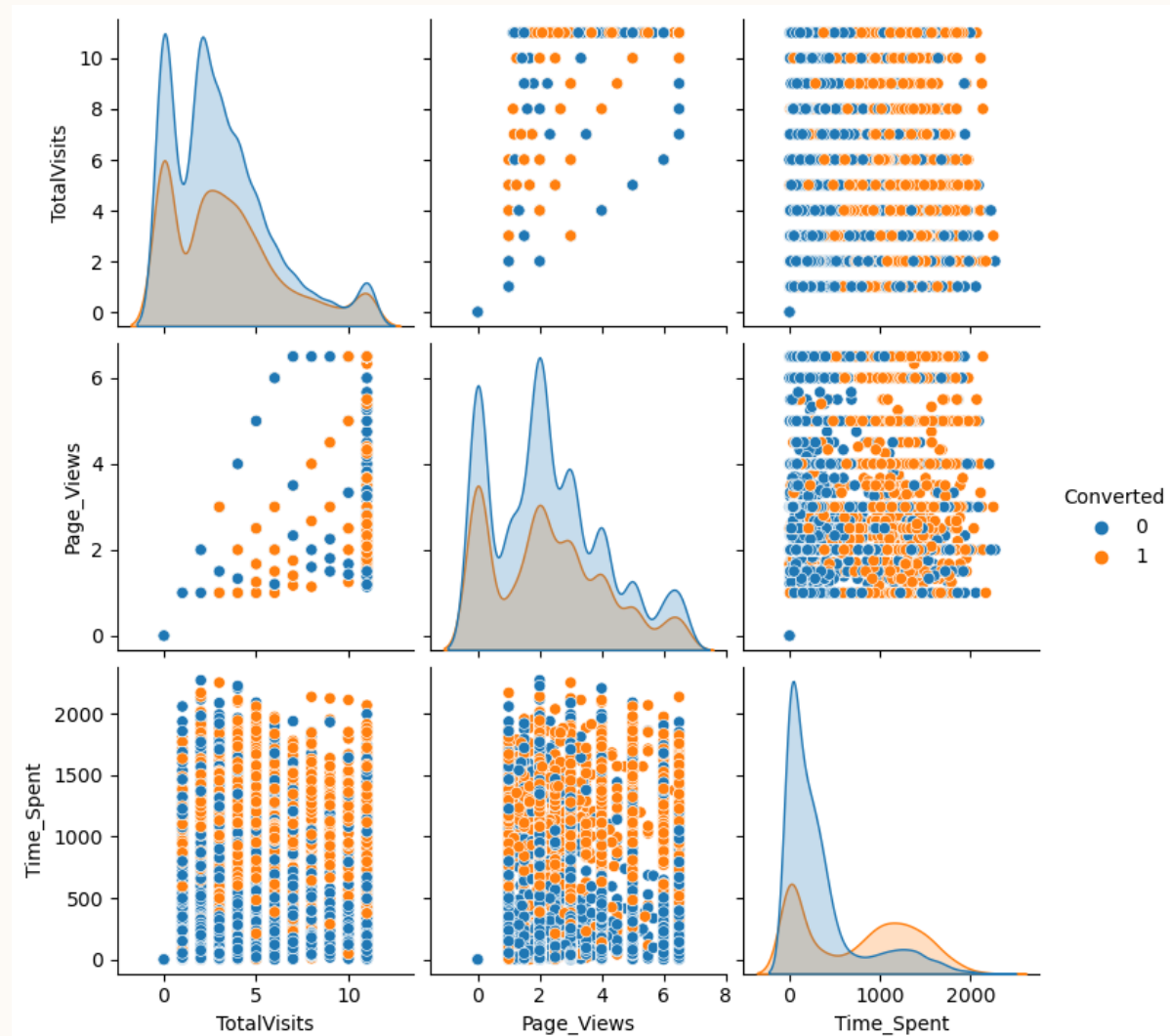
SOLUTION METHODOLOGY

- Importing the data and inspecting the data frame
- Data cleaning and data manipulation.
- EDA
- Dummy variable creation
- Test-train split
- Feature Scaling
- Correlations
- Model building (RFE VIF and p-values)
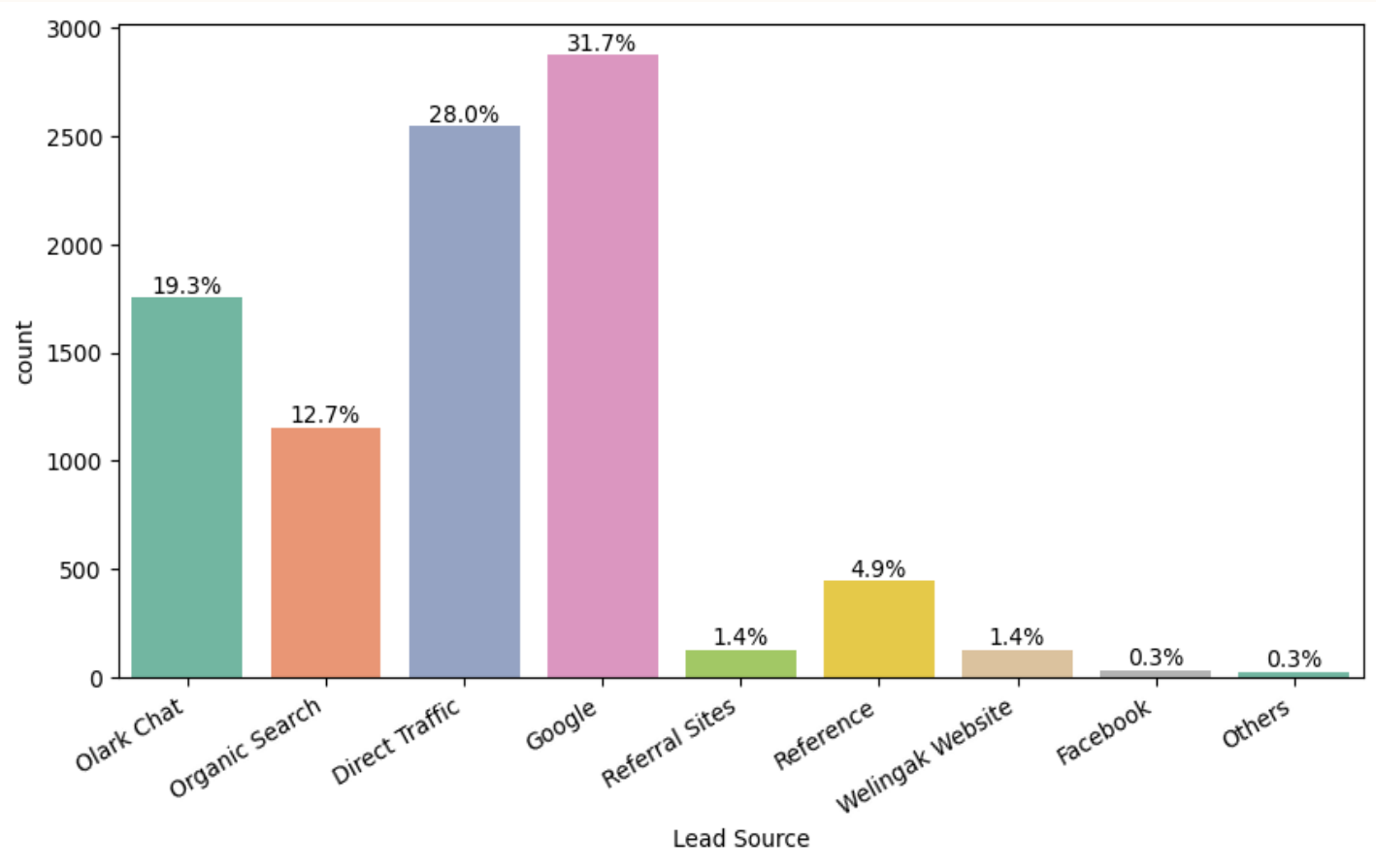- Model evaluation
- Making predictions on test Data set

EDA:

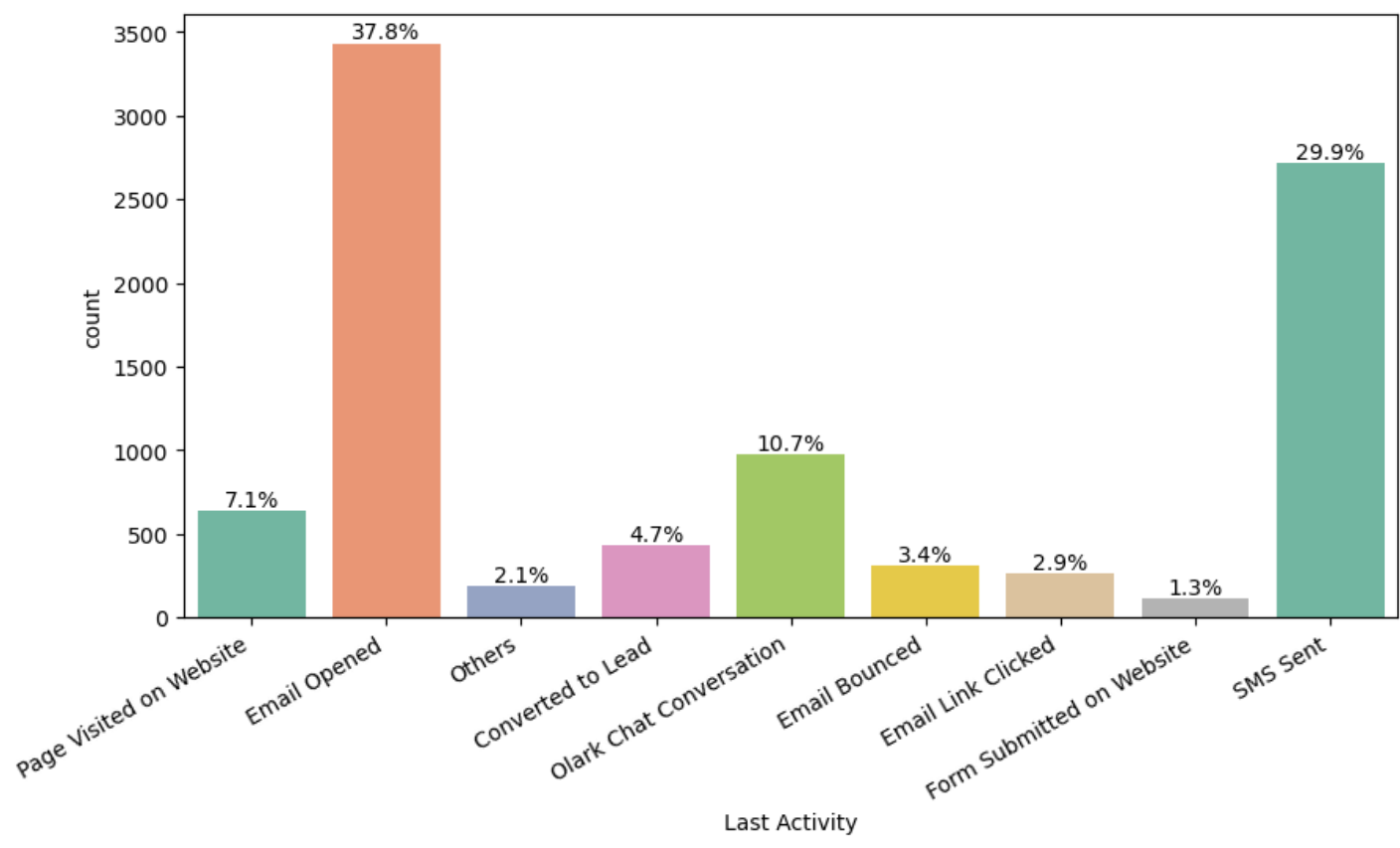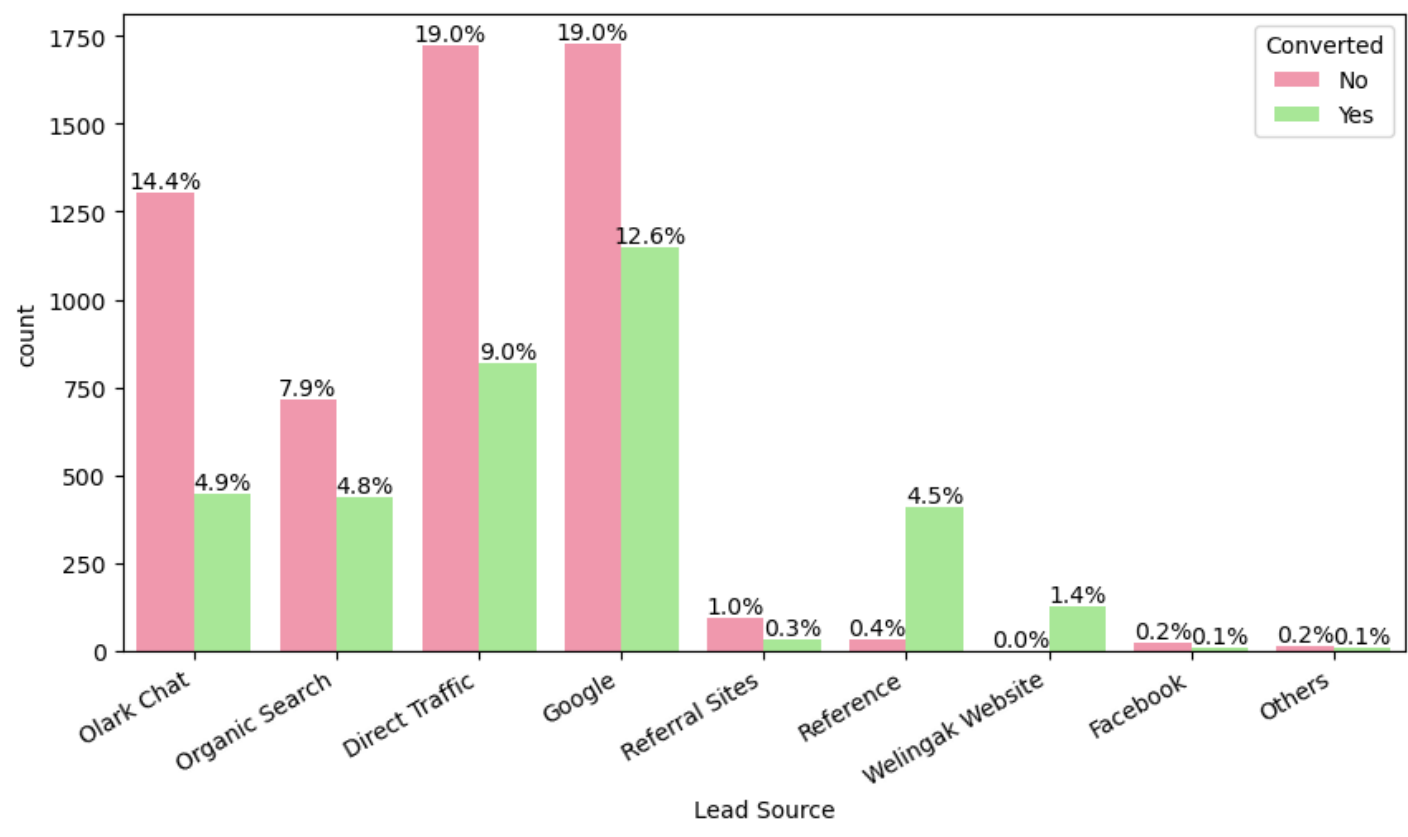## Bivariate Analysis For Numerical Variables

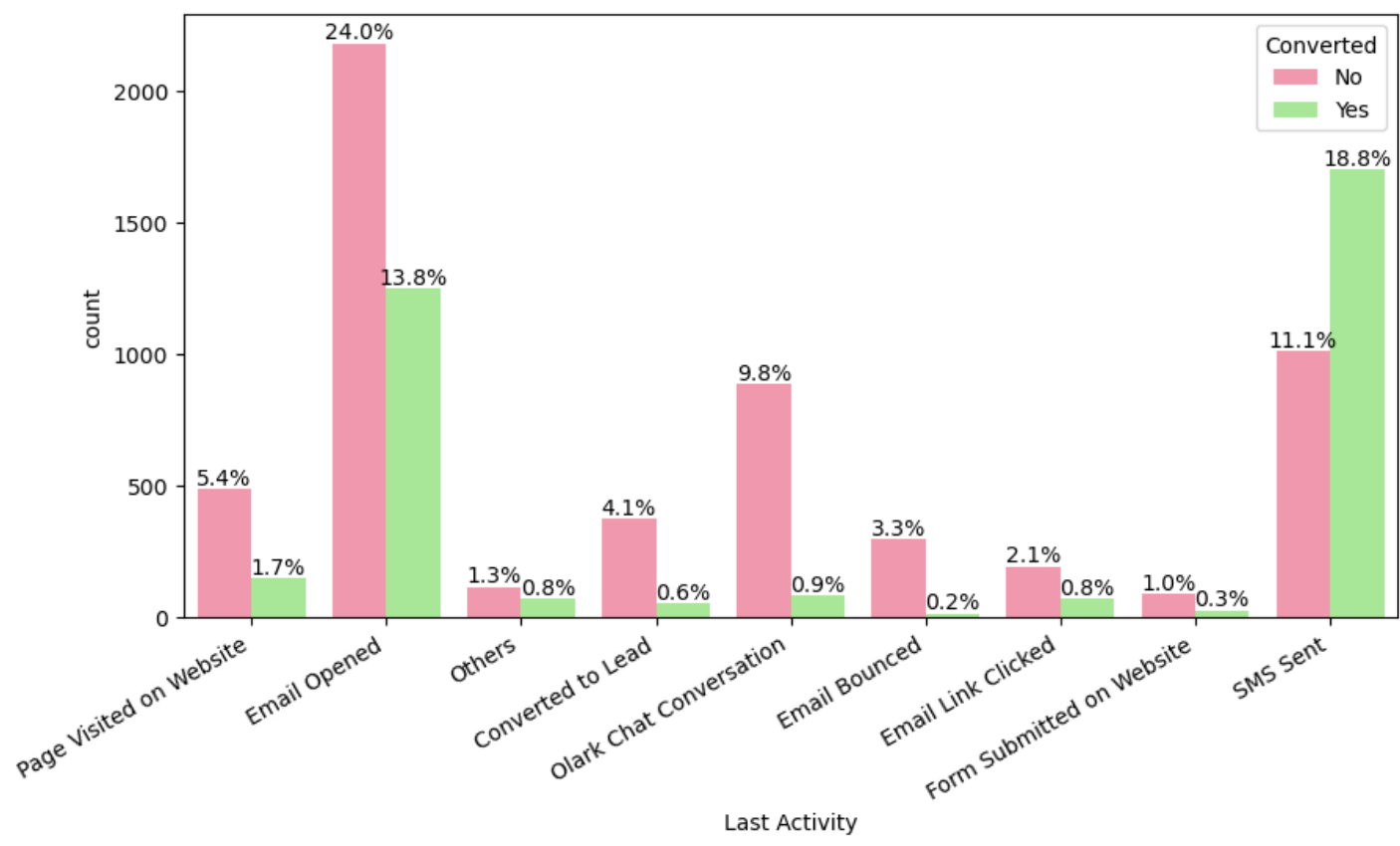# UNIVARIATE ANALYSIS OF CATEGORICAL VARIABLE
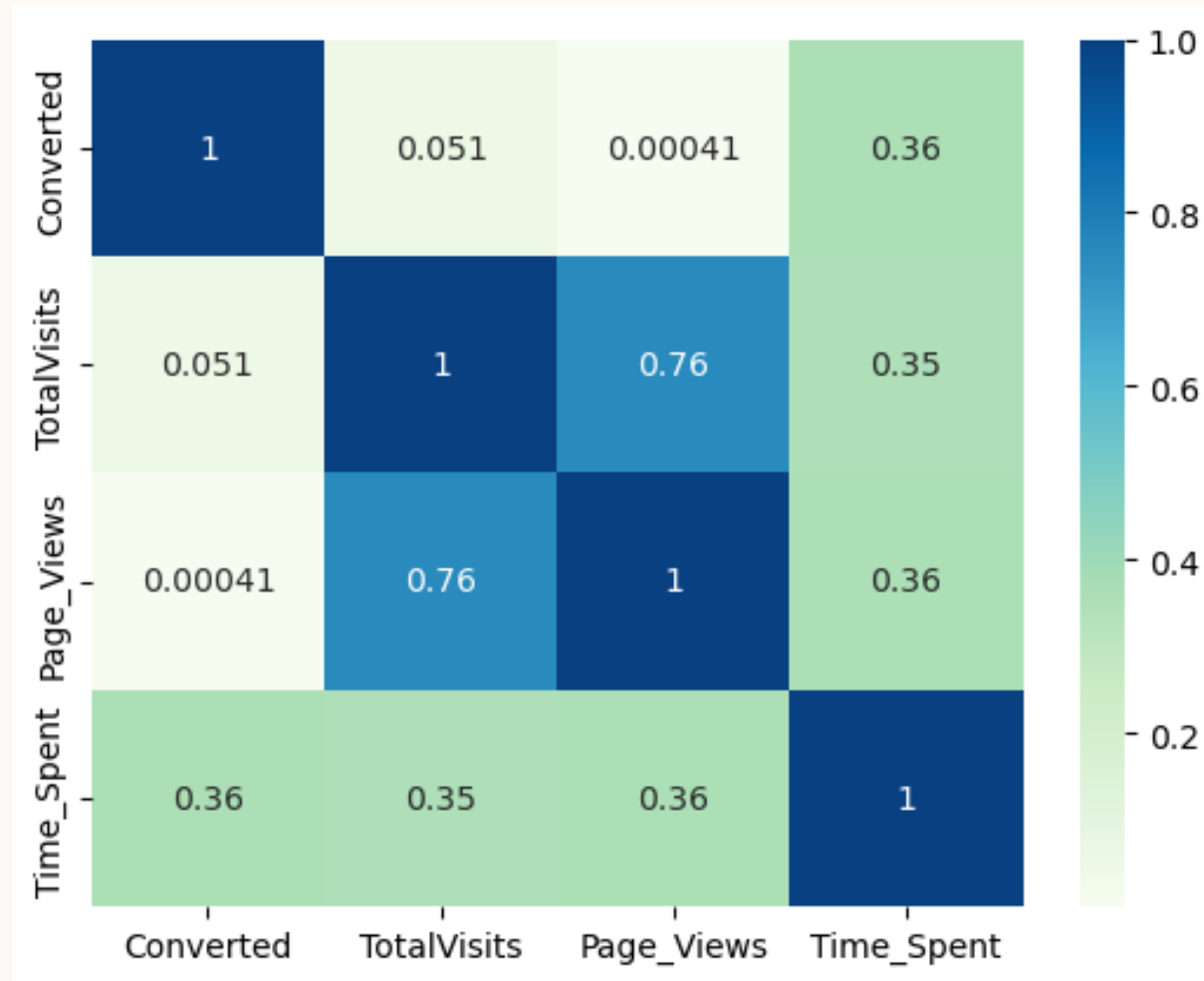
UNIVARIATE ANALYSIS OF CATEGORICAL VARIABLE

BIVARIATE ANALYSIS OF CATEGORICAL VARIABLE

BIVARIATE ANALYSIS OF CATEGORICAL VARIABLE

HEATMAP ANALYSIS

MODEL BUILDING
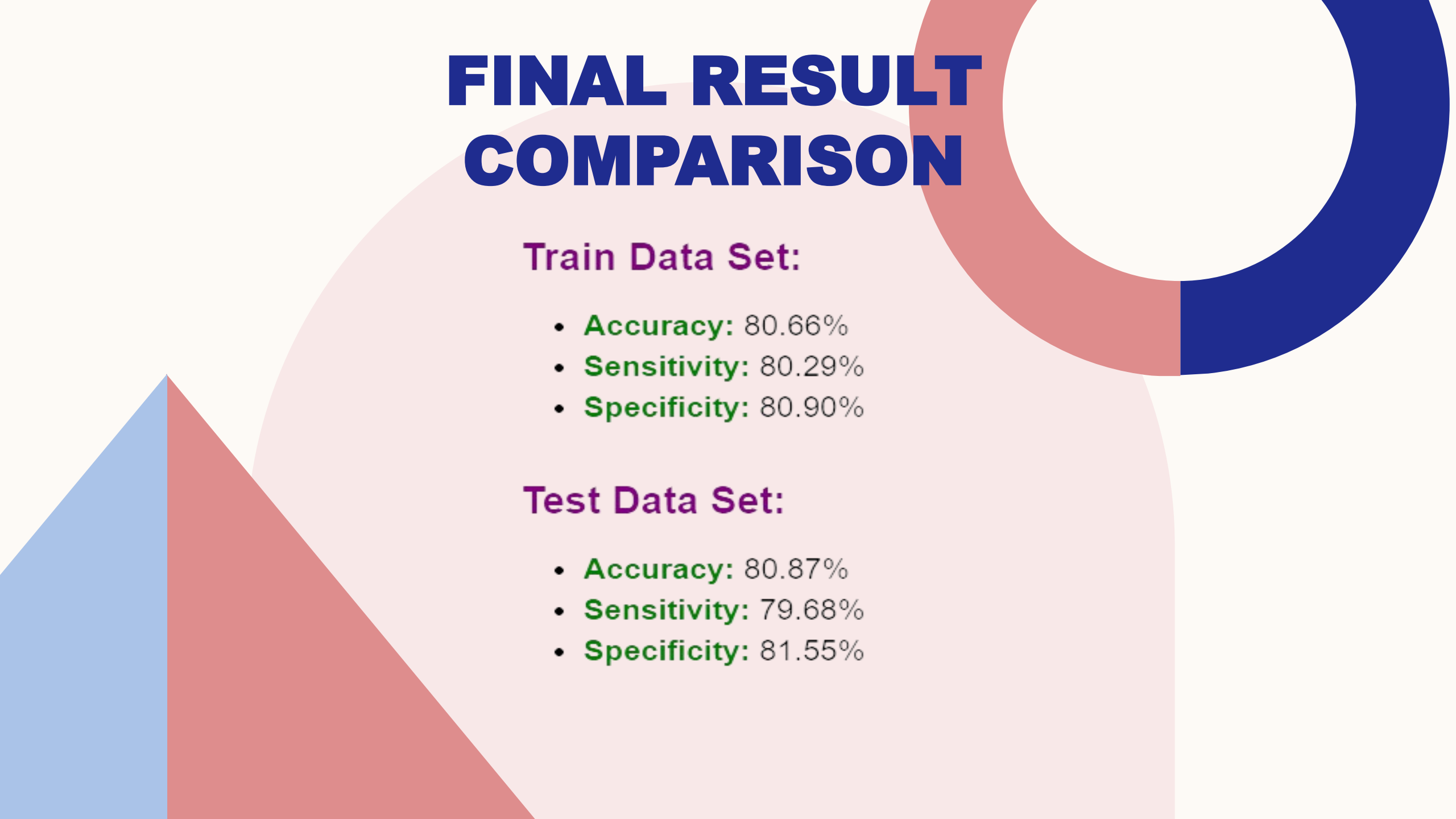
❖ Splitting the Data into Training and Testing Sets.

❖ Using RFE we will select top 15 variables.

❖ Building model using RFE selected variable until the P-value of all the Variable is > 0.05 and VIF is >5 .

❖ Model Evaluation is done using confusion matrix at cut of point of 0.346.

❖ Making Predictions on Test Data witch result in 80% accuracy.

ROC CURVE
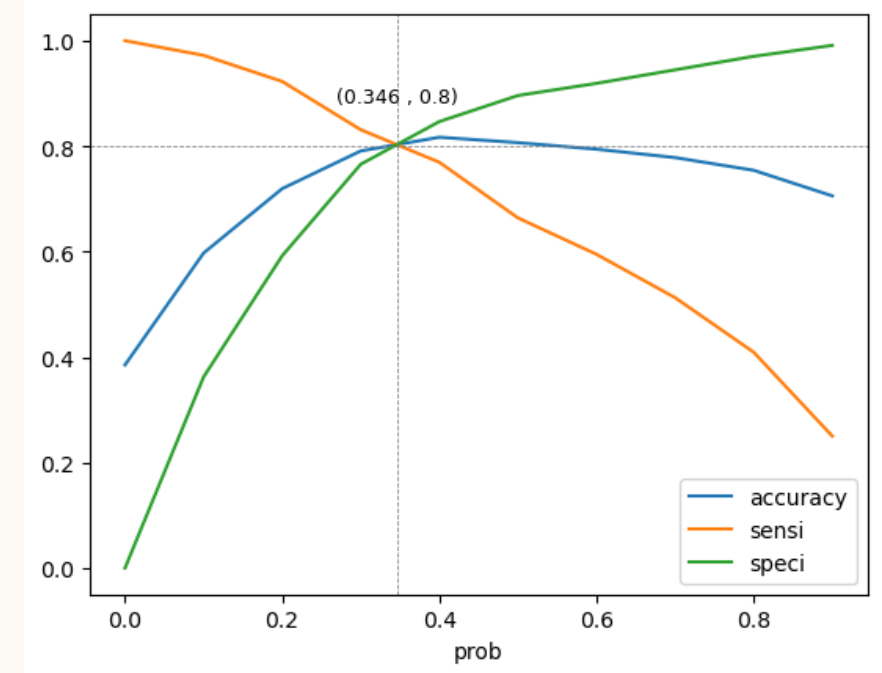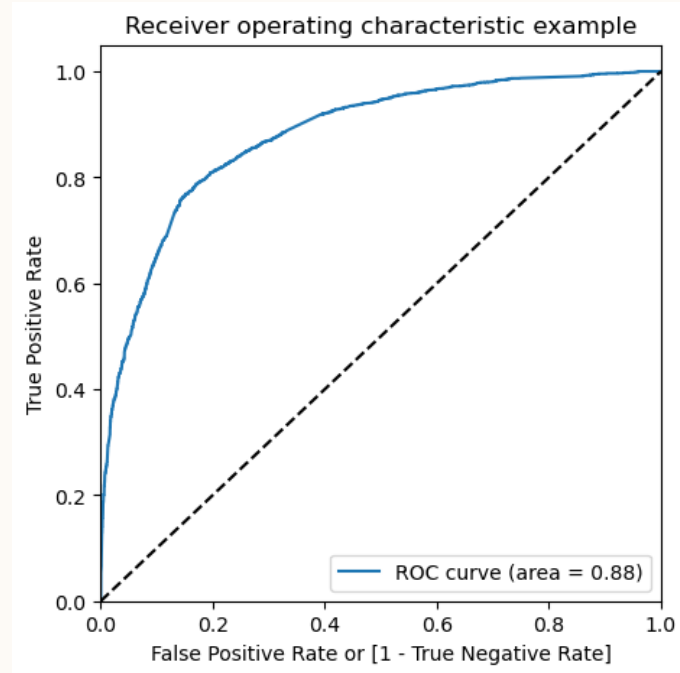
❖ Area under ROC curve is 0.88 which indicates a good predictive model.

❖ 0.346 is our Optimal cutoff point for probability threshold.

FINAL RESULT COMPARISON

Train Data Set:

- **Accuracy:** 80.66%
- **Sensitivity:** 80.29%
- **Specificity:** 80.90%

Test Data Set:

- **Accuracy:** 80.87%
- **Sensitivity:** 79.68%
- **Specificity:** 81.55%

CONCLUSION

❖ Leads who spent more time on website, more likely to convert.

❖ Leads Coming from Welingak Website & Reference have highest chances of Conversion.

❖ Most common last activity is SMS Sent & email opened which will can bring huge impact on Lead Conversion.

❖ Customer spending more time on website are very likely to take up the course.

❖ Working Professional should be targeted the most for Lead Conversion.

THANK YOU

Neelesh Parihar