

LEAD SCORE CASE STUDY

DURGESH SUNIL NIKAM

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 up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective:

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

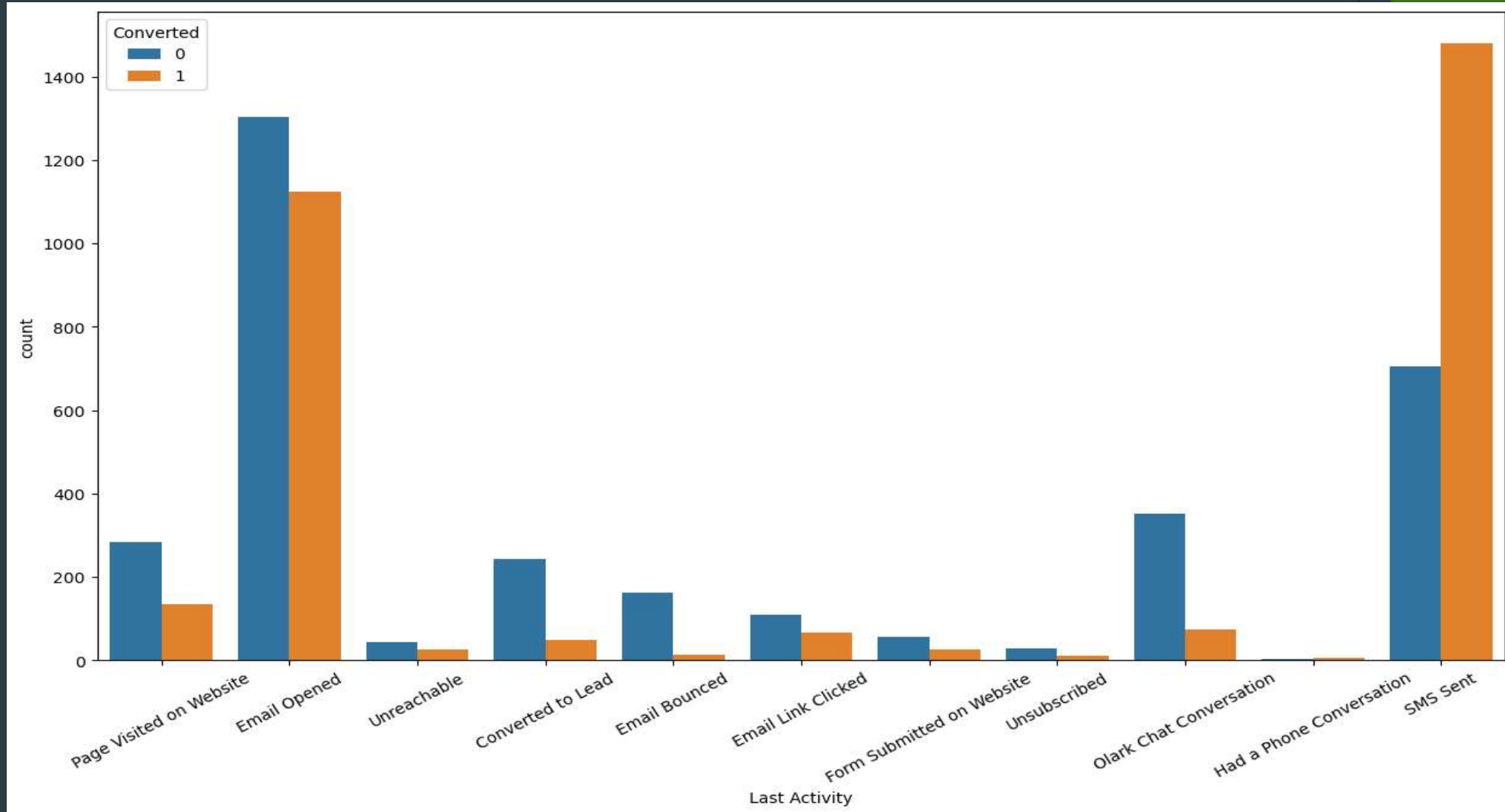
Solution Methodology

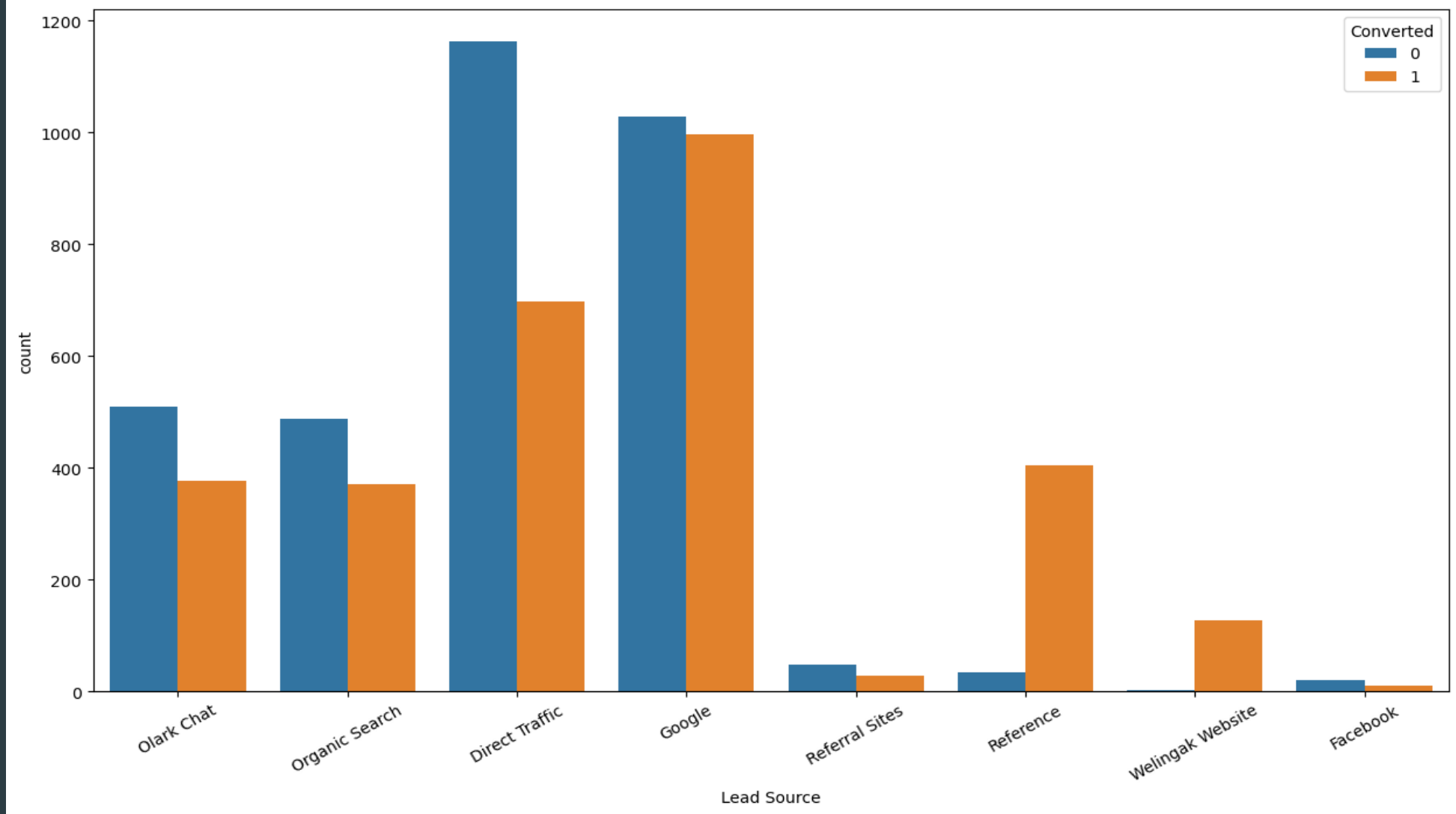
- ▶ Data cleaning and data manipulation.
 1. Check and handle duplicate data.
 2. Check and handle NA values and missing values.
 3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
 4. Imputation of the values, if necessary.
 5. Check and handle outliers in data.
- ▶ EDA
 1. Univariate data analysis: value count, distribution of variable etc.
 2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.
- ▶ Feature Scaling & Dummy Variables and encoding of the data.
- ▶ Classification technique: logistic regression used for the model making and prediction.
- ▶ Validation of the model. Model presentation.
- ▶ Conclusions and recommendations.

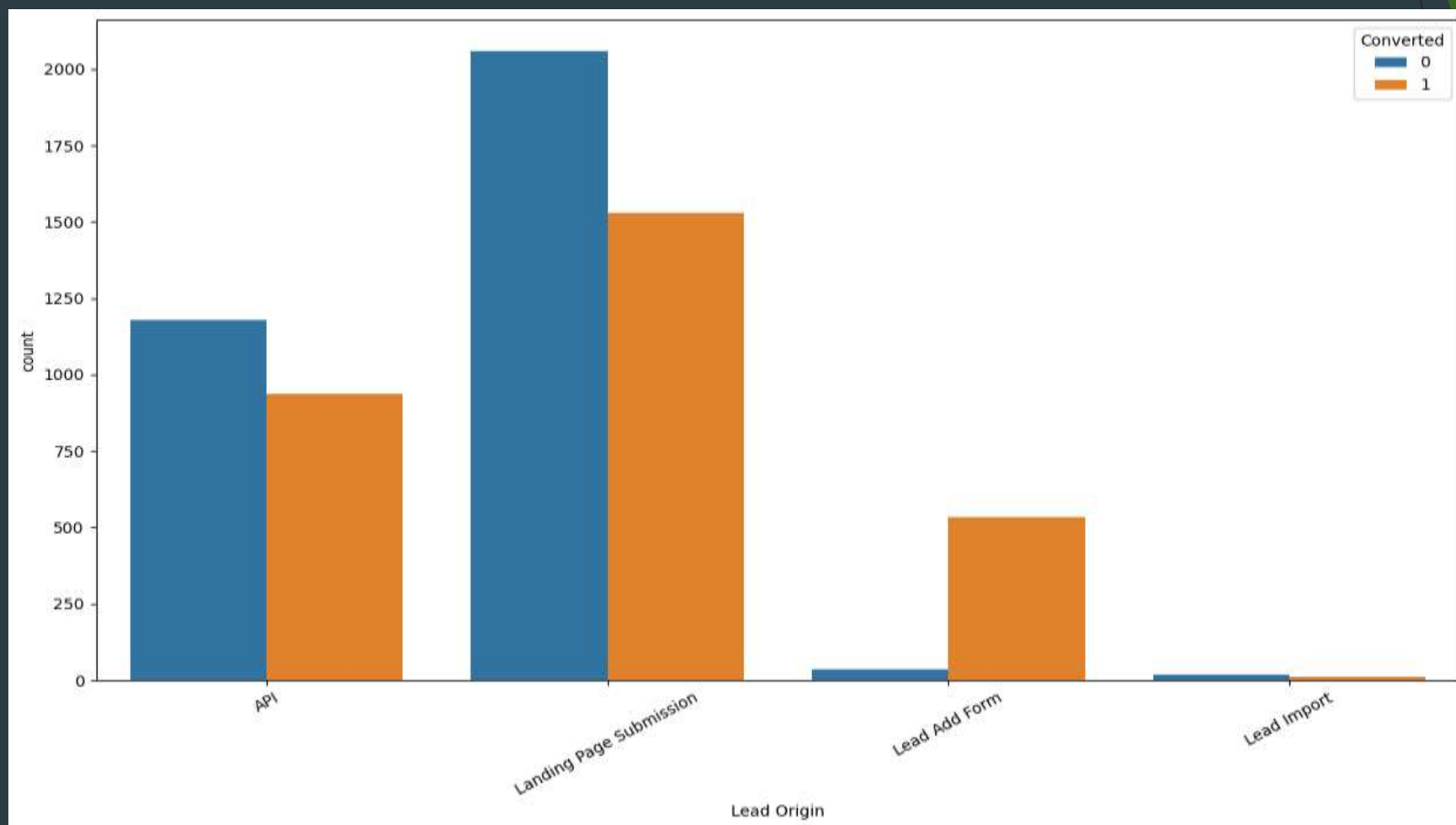
Data Manipulation

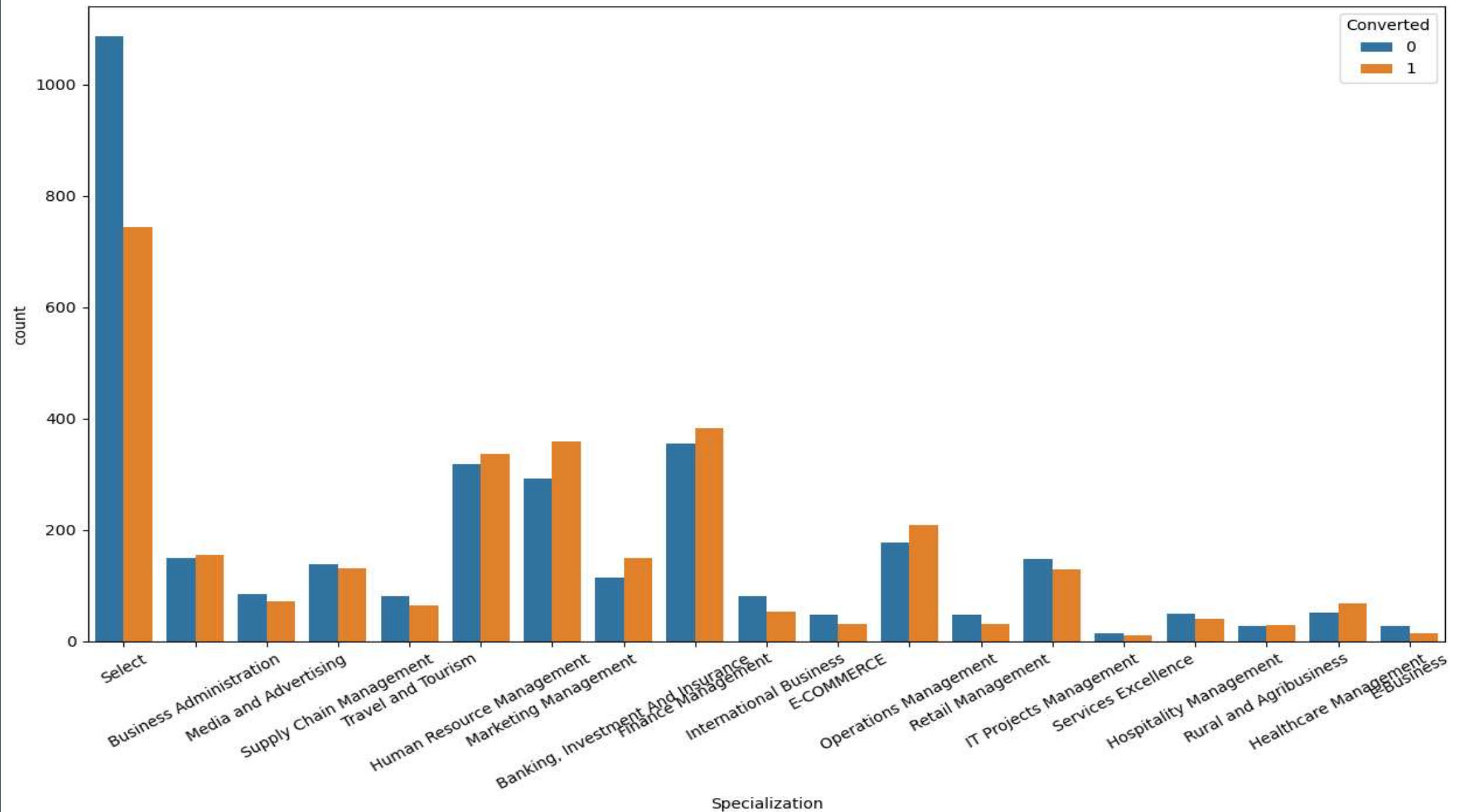
- Total number of row =9240 and columns =37
- Columns dropped with null values greater than 40% .
- Dropped single variable columns like 'Magazine','Receive More Updates About Our Courses',' I agree to pay the amount through cheque' .
- Also dropped two variable column with maximum weight age to only variable are of no use in prediction like 'Do not Email' and 'Do not call'.
- Categorical columns in which some of the variables with nunique count less than 15 are dropped as it will increase the number features which are not even that impactful
- Treated outliers using capping method to avoid the data loss.

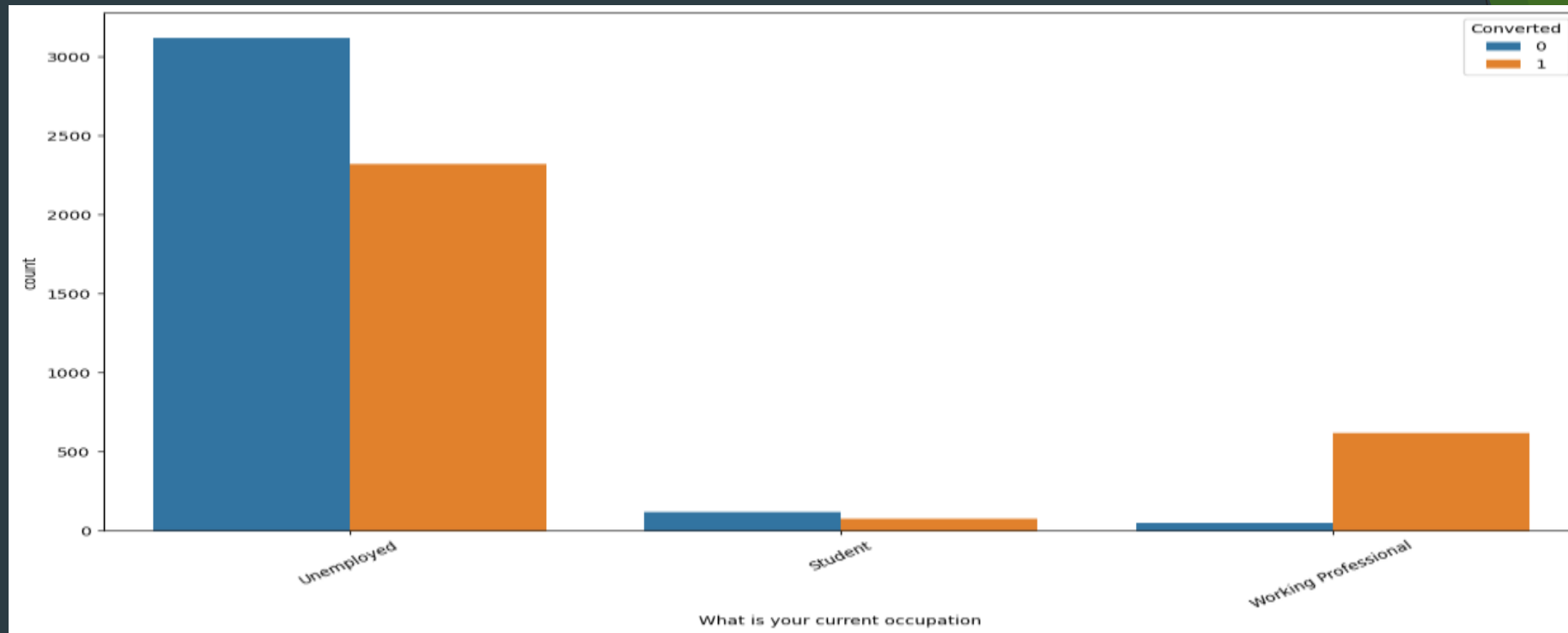
Data Visualisation





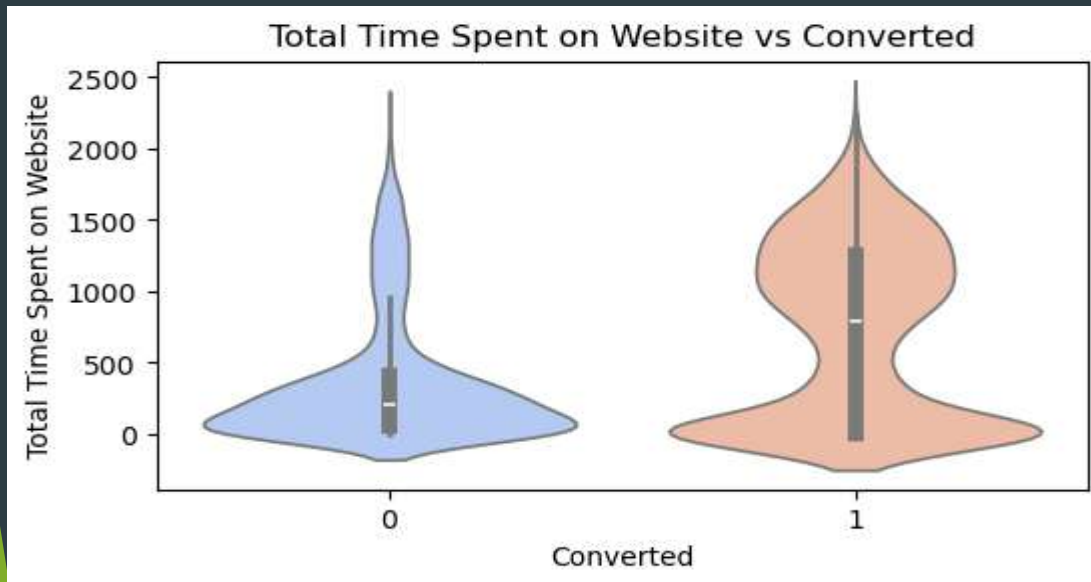
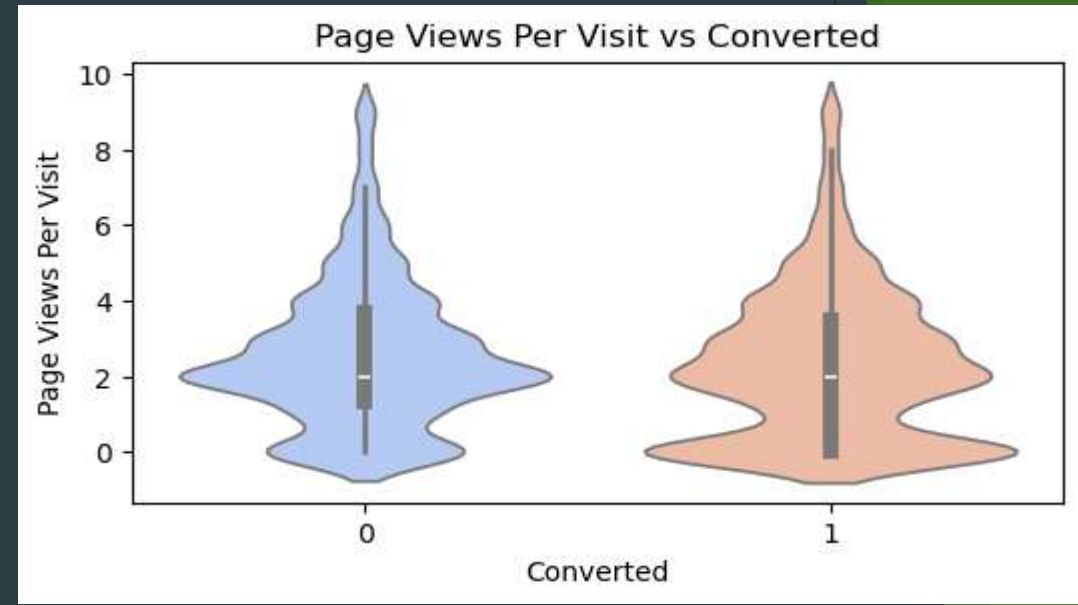
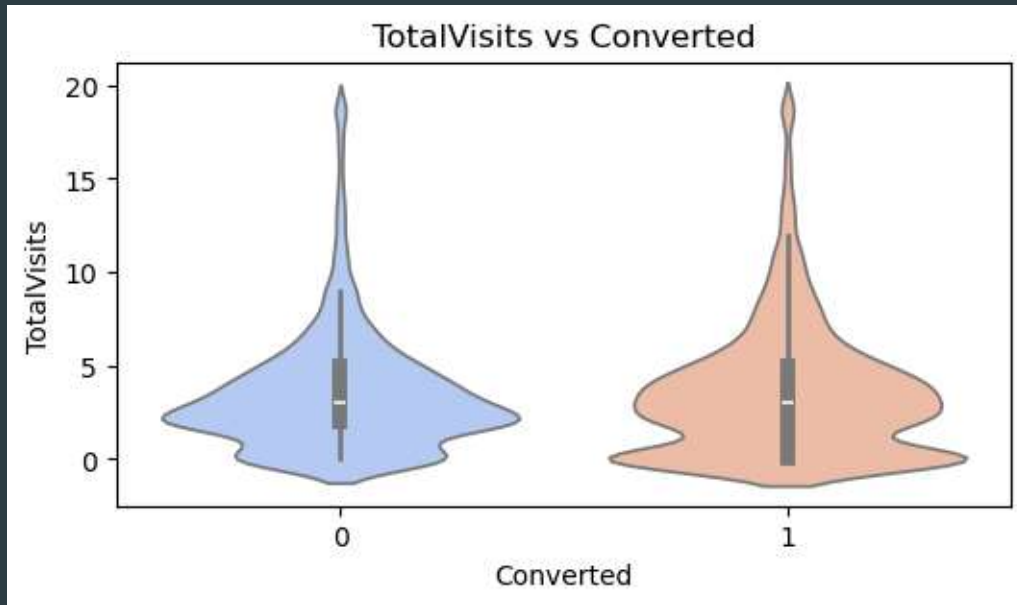






- Lead Origin from a Lead Add Form has high chances of conversion
- Lead Source from Google ,Reference and Welingak Website has shown positive response for the maximum number of times
- Last Activity SMS sent and Email Opened are likely more interested
- Lead with specialization Business Administration, Human Resource Management ,Marketing Management, Banking, Investment and Insurance, Finance and operation management are the hot leads.
- Working Professional are highly interested in getting enrolled

Numeric Variables



- As the number of Total Time spent on website increases probability of conversion also increases
- Total visits and page views per visit increase also impacting the lead decisions.

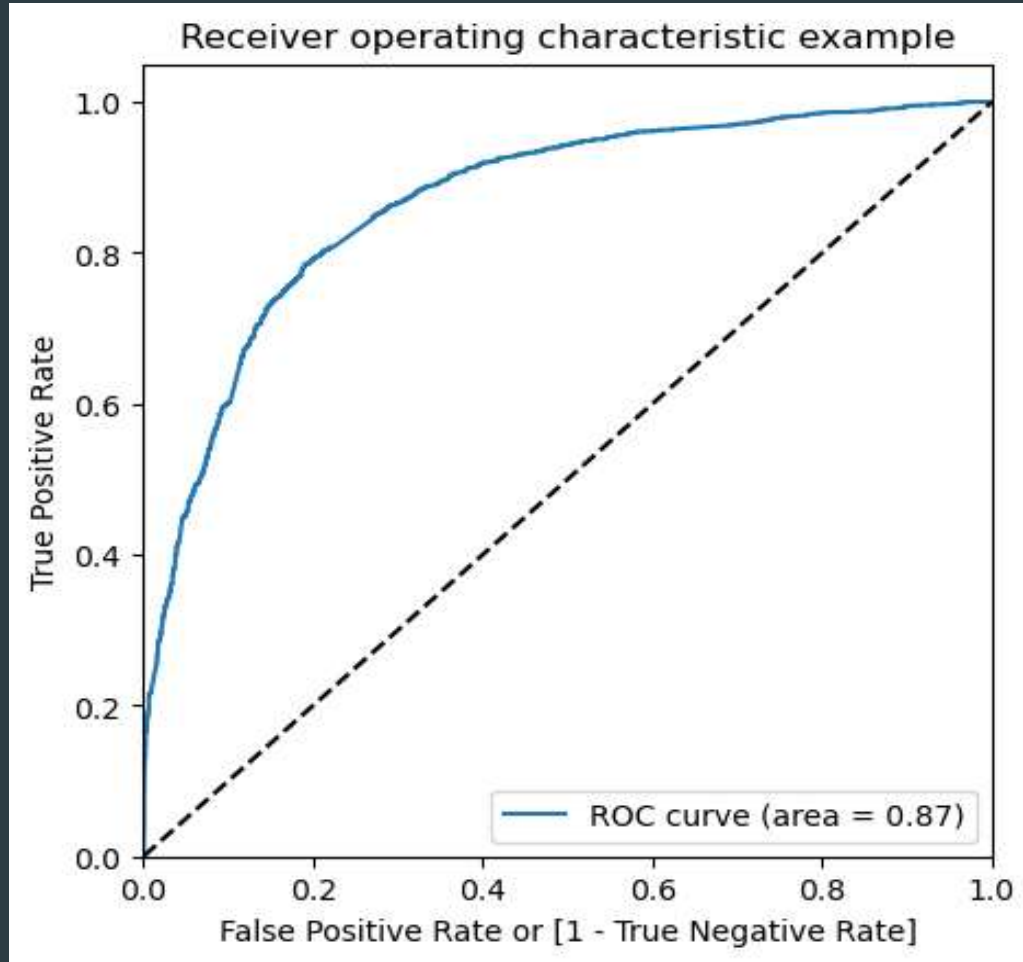
Data Conversion

- ▶ Dummy Variable are created from object type variables
- ▶ Numeric columns are normalized using minmax scaler.
- ▶ Total row for analysis 6300
- ▶ Total columns for analysis 51

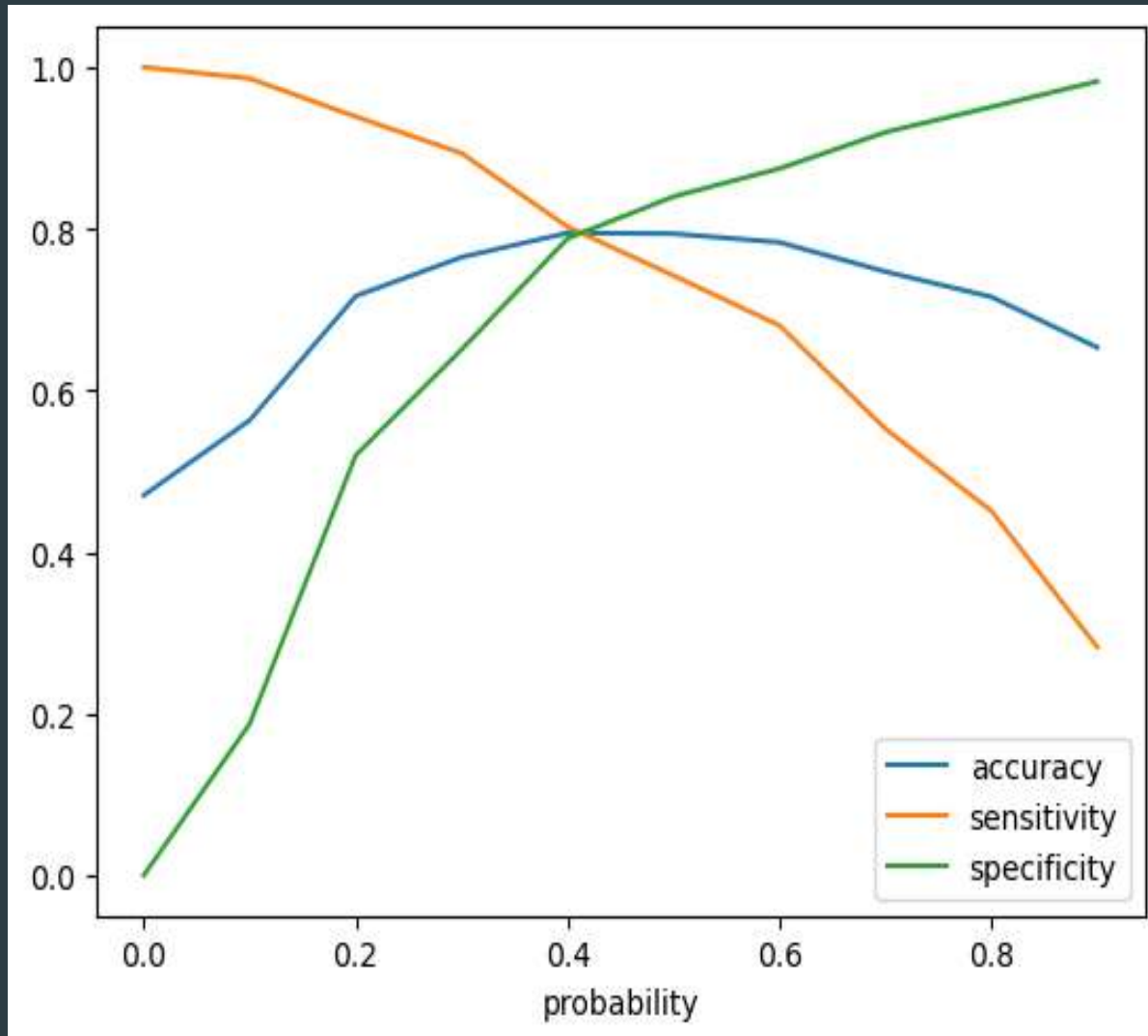
Data Modelling

- ▶ Splitting the data into Training and Test set in the ration 70:30
- ▶ Used RFE for feature selection
- ▶ 15 Features selected with the RFE
- ▶ Manually Tuning the 15 features by checking P-value less than 0.05 and VIF not greater than 5
- ▶ 12 features obtained after building 4 models
- ▶ Obtain optimal cut off 0.42 and Roc curve 0.87
- ▶ Predicted the test set Accuracy obtain 78%

ROC Curve



Area Under the ROC curve is 0.87 which is close to represents the strong discriminating power of the model.



Optimal cut-off obtained through the range of probability cutoff is 0.42

Key Insights:

- ▶ Feature Impact: The analysis revealed that leads from the Lead Add Form had a significantly higher likelihood of conversion.
- ▶ Furthermore, Google, Reference, and the Welingak Website emerged as top-performing sources, demonstrating high conversion rates.
- ▶ Activity-Based Engagement: Leads showing activities such as SMS Sent and Email Opened were more engaged and had a higher chance of conversion. This highlighted the importance of these activities as key indicators of lead interest.
- ▶ Specialization Significance: Certain specializations such as Business Administration, Human Resource Management, Marketing Management, Banking, Investment and Insurance, Finance, and Operation Management were identified as high-potential areas. Leads in these fields were found to be more likely to convert.
- ▶ Target Audience: Working Professionals were noted as a particularly interested demographic. Tailoring marketing efforts to this group and emphasizing the flexibility and benefits of the offerings could lead to higher enrollment rates.

Conclusion

- ▶ Enhance Visibility of the Lead Add Form: Improve placement and accessibility to capture more high-converting leads.
- ▶ Optimize Marketing on Key Sources: Allocate resources to enhance visibility on Google, Reference, and the Welingak Website.
- ▶ Focus on High-Interest Activities: Implement targeted follow-up strategies for leads with SMS Sent and Email Opened.
- ▶ Target Hot Specializations: Customize marketing and outreach efforts for the identified high-potential specializations.
- ▶ Address Working Professionals: Develop specialized programs and highlight benefits that align with the career needs of working professionals.