LEAD SCORING CASE STUDY-

SIDHARTH NARAIN

# INSIGHTS FROM X - EDUCATION ANALYSIS

PROBLEM STATEMENT

- X Education sells online courses to industry professionals.

- Although X Education gets a lot of leads, its lead conversion rate is very poor.

- Typical lead conversion rate at X education is around 30%.

- If 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.
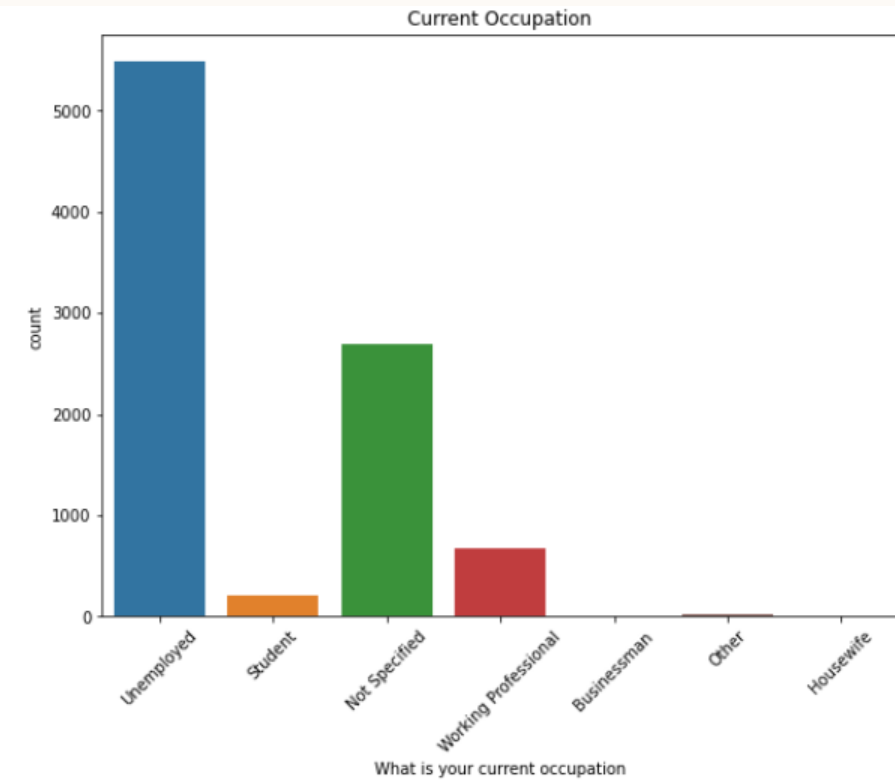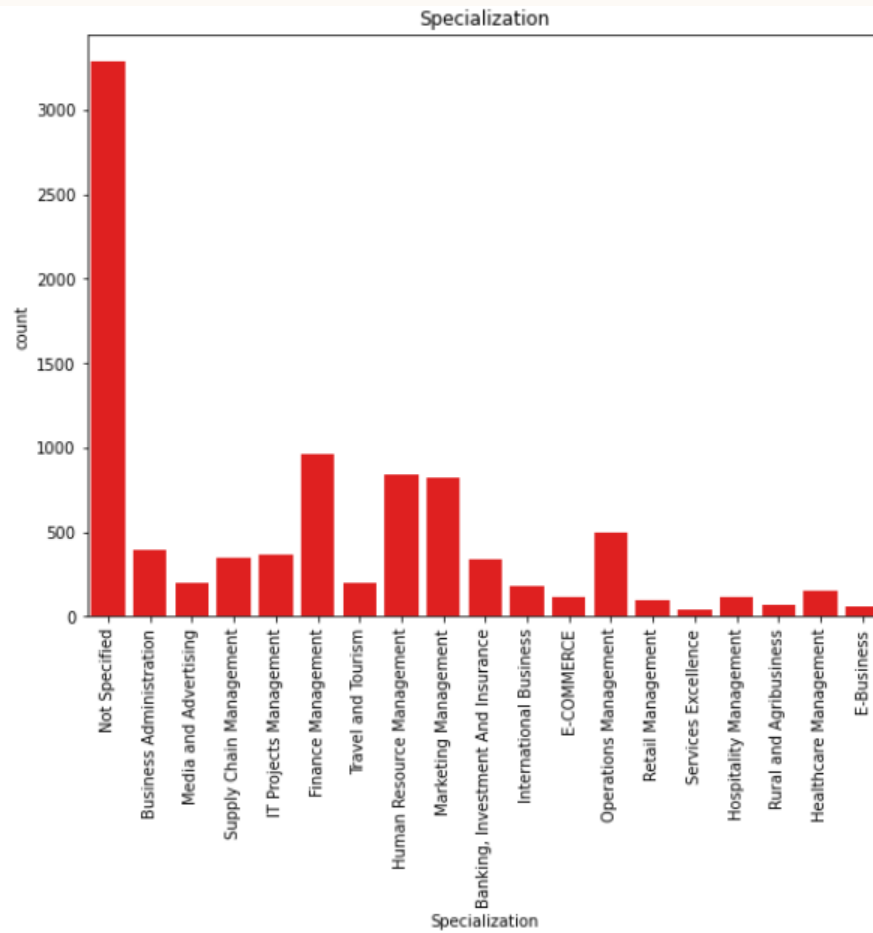
METHODOLOGY

- Data Understanding
- Cleaning the data
- EDA
  - Univariate Analysis
  - Bivariate Analysis
  - Correlation plot

- Dummy variables creation

- Splitting the data into Train and Test
  - Train data 70% and Test data 30%

- Building the Model
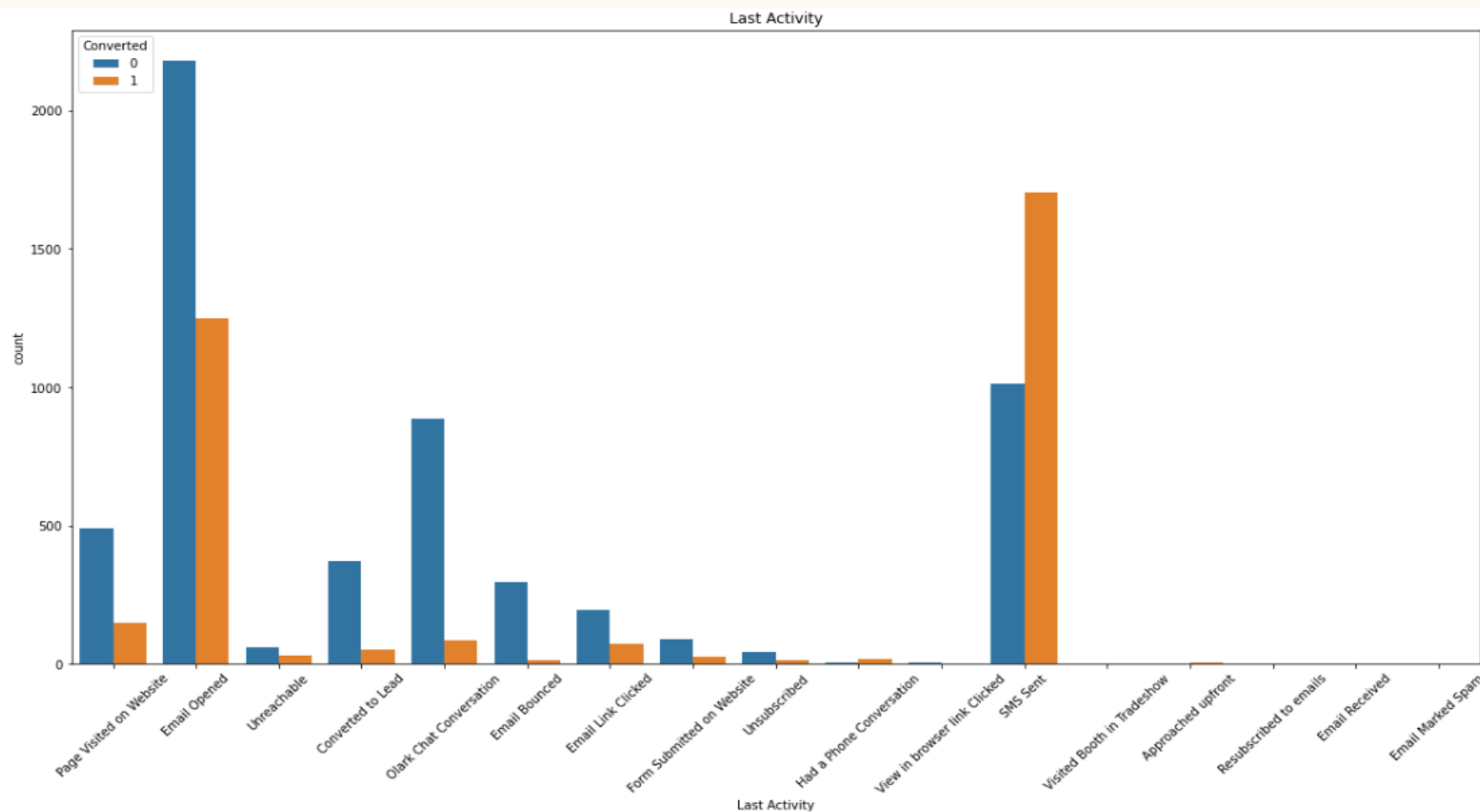 - Building a Logistic Regression Model

- Model Evaluation
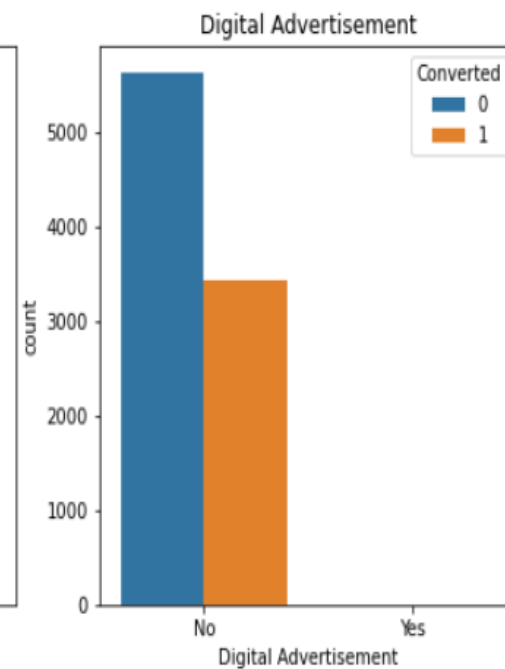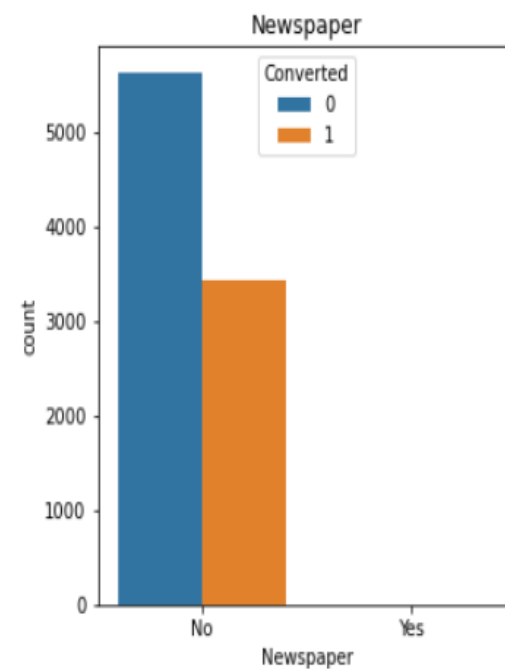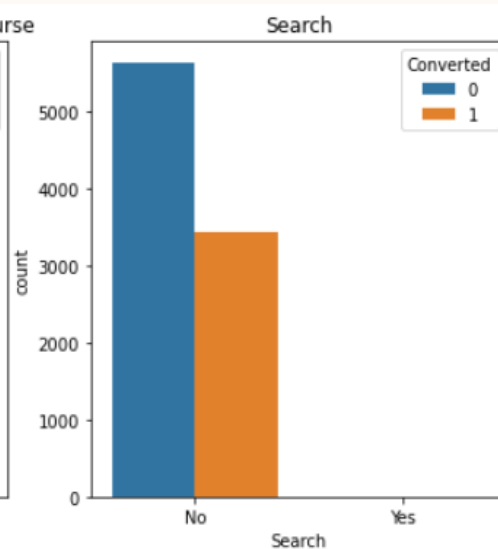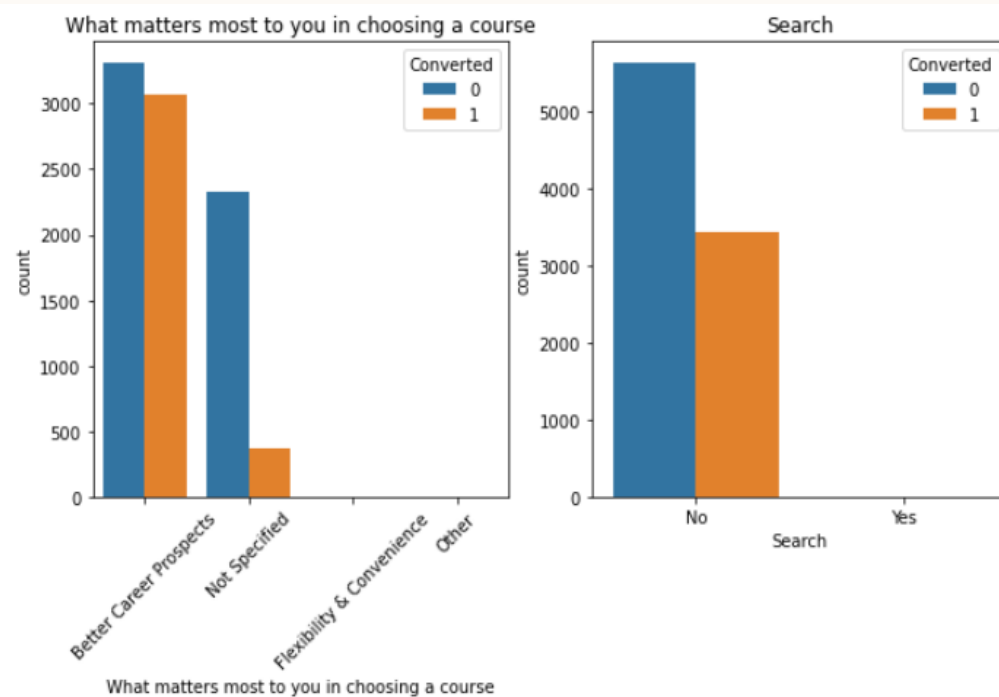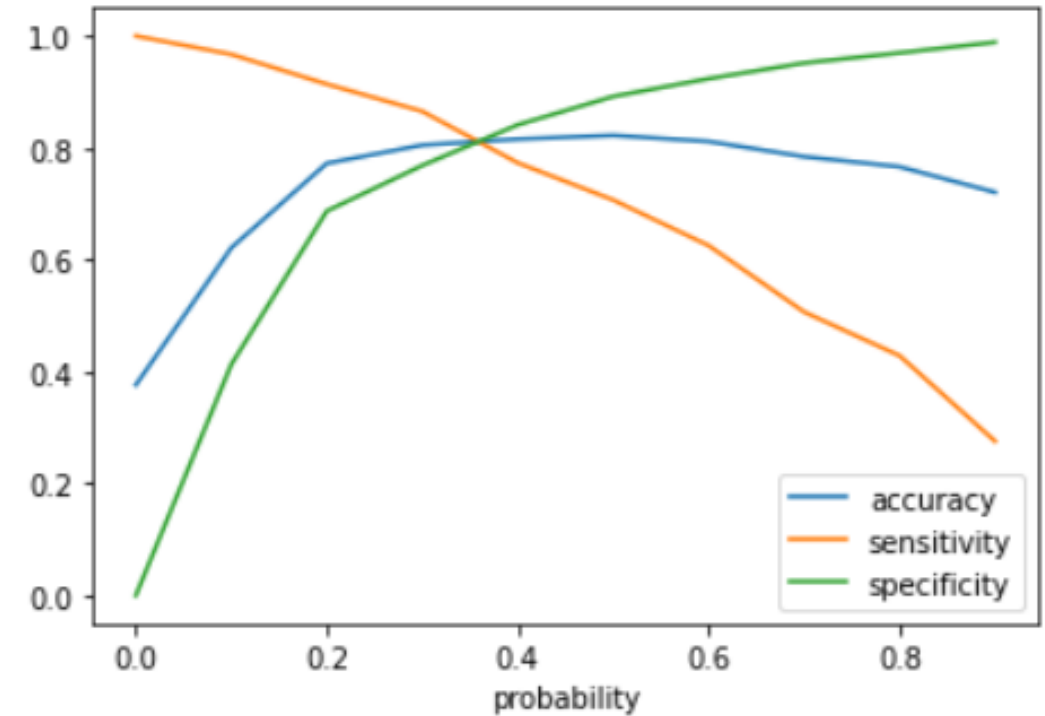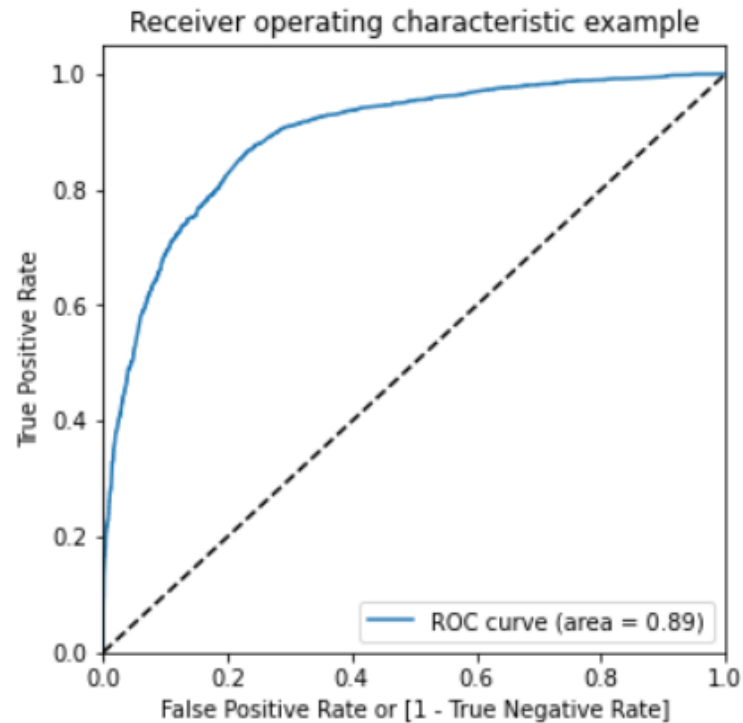
EDA
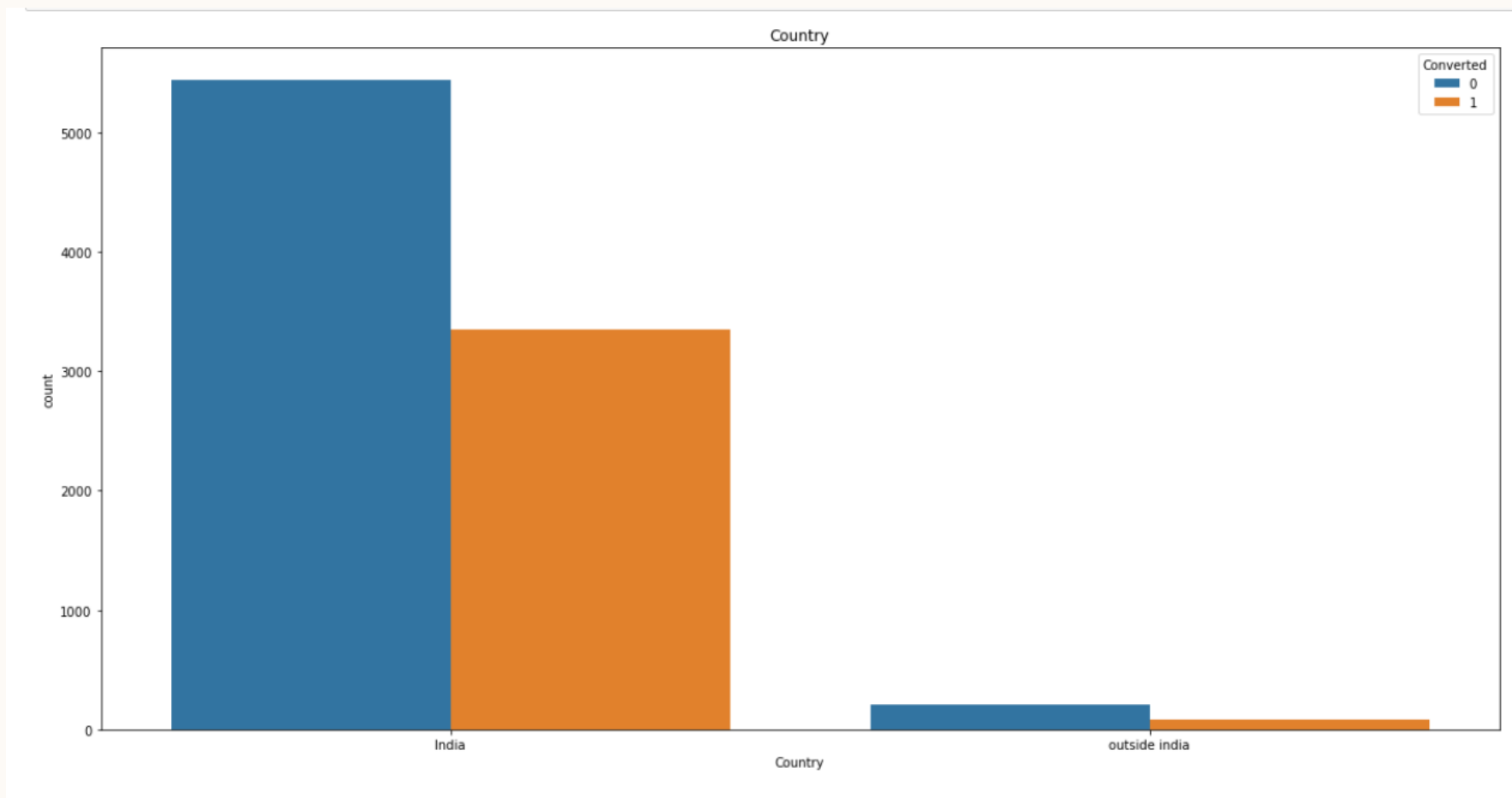
EDA

EDA

EDA

EDA

MODEL BUILDING

- The Data is Split into Training Data and Test Data.

- This split is done in the ration of 70:30 for Train and Test data respectively.

- RFE was used for feature selection to attain the top 20 variables.

- Then the rest of the variables were manually dropped based on their VIF values and the P-value (VIF < 5 and P-value < 0.05).

ROC CURVE

The Optimal Cutoff value is around 0.37

CONCLUSION

FROM THE MODEL THESE ARE THE TOP VARIABLES WHICH ARE CONTRIBUTING TO CONVERSION:

- **TOTAL TIME SPENT ON WEBSITE**
- **WHAT IS YOUR CURRENT OCCUPATION_WORKING PROFESSIONAL**
- **LEAD ORIGIN_LEAD ADD FORM**

Lead Source:

- **Google, Direct Traffic, Organic Search, Referral Sites, Welingak Website**

Lead Origin:

- **Lead Add Form**

Last Activity:

- **Olark Chat Conversation**
- **Email**
- **Page Visited on Website**

APPENDIX

- https://scikit-learn.org
- https://towardsdatascience.com

THANK YOU