



Lead Score Case Study

J VAMSI KALYAN



Steps followed

- Data Understanding
- Data Preprocessing
- Exploratory Data Analysis
- Feature Scaling
- Model Building
- Model Validation
- Conclusion and Recommendation

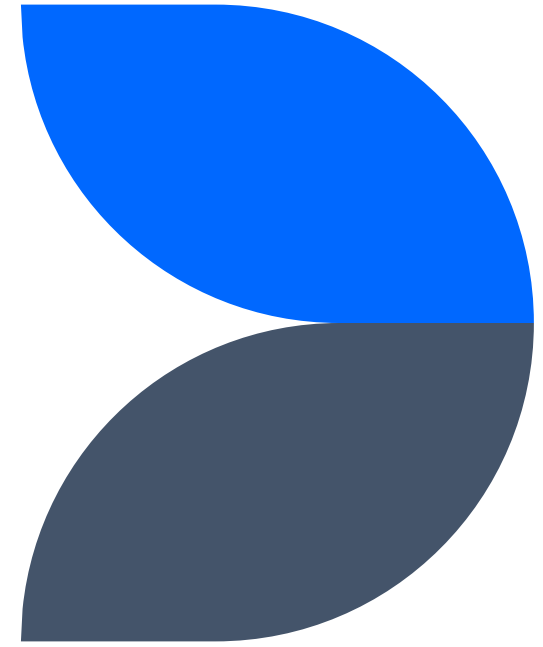
Problem Statement

- An education company named X Education sells online courses to industry professionals.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Objective

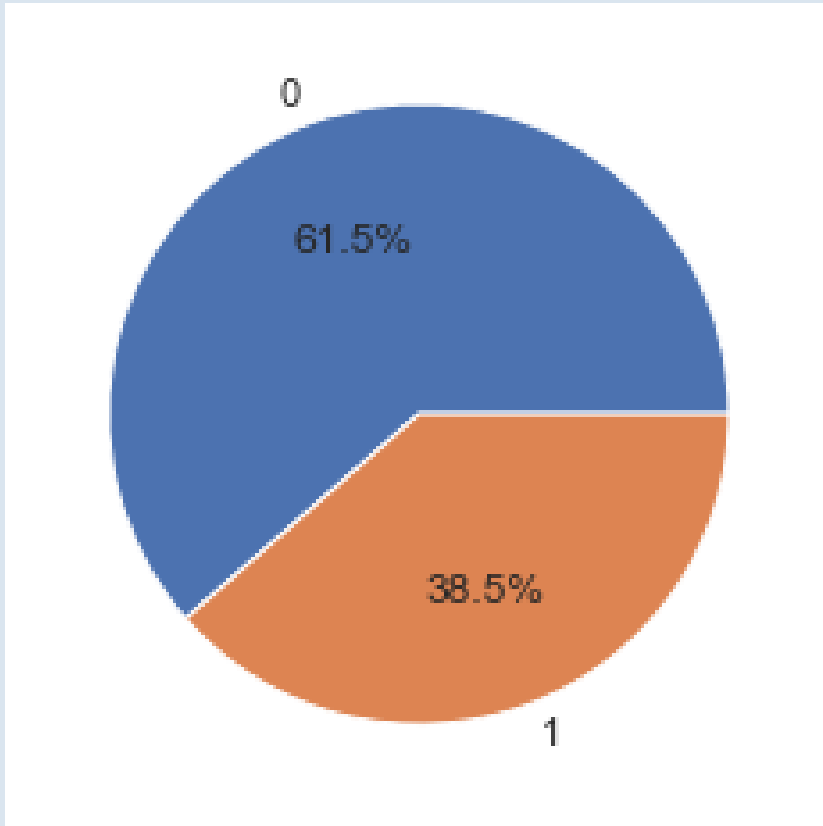
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads
- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well

Data Cleaning



- a. Few Columns have 'Select' value in the data and this as same as null
- b. Removed columns with 45% of null values
- c. Imputed null values with mode
- d. Dropped missing values in Categorical features
- e. Removed columns irrelevant for model building (Prospect_ID, Lead_Number and Last Notable_Activity)

Data Imbalance

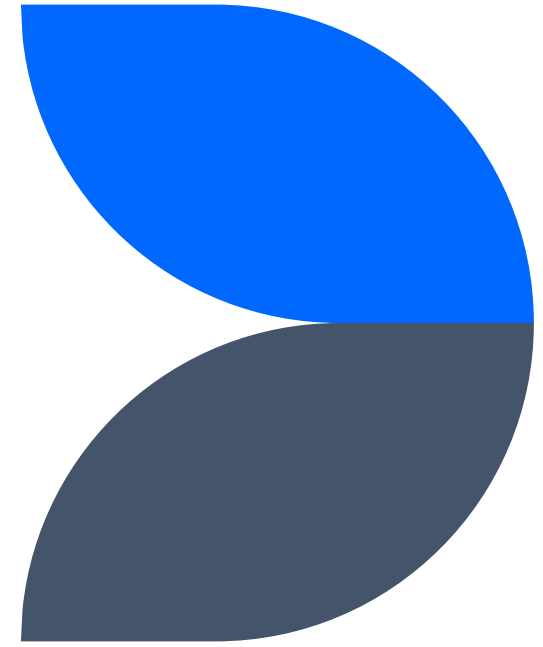


31% of leads are converted, however 61% of leads are not converted

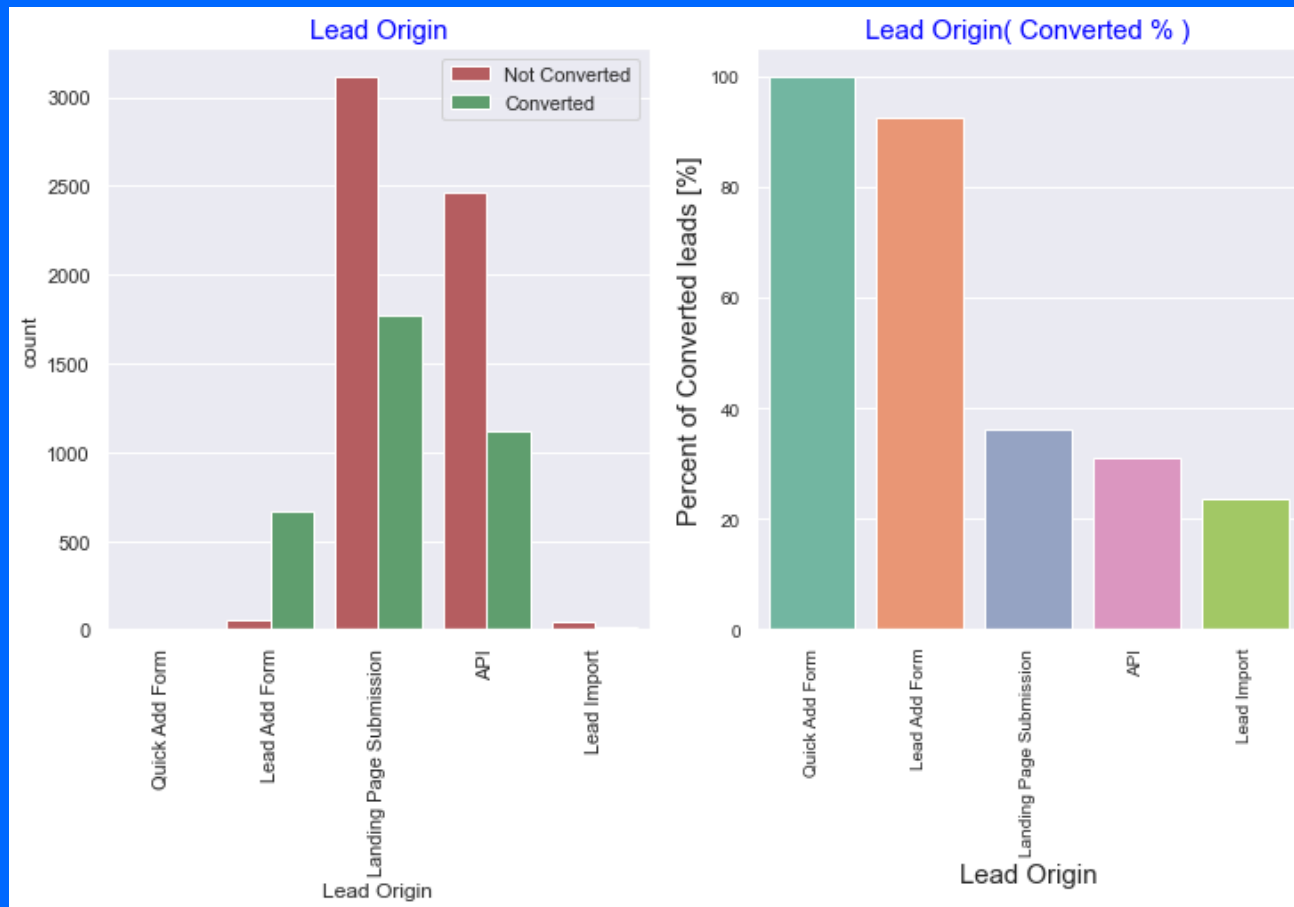
Outlier Analysis

- Outliers are very important while building a machine learning model. We have to treat them properly to increase the overall accuracy of the model.
- Outliers can be found using box plot and considered 95% of the data.

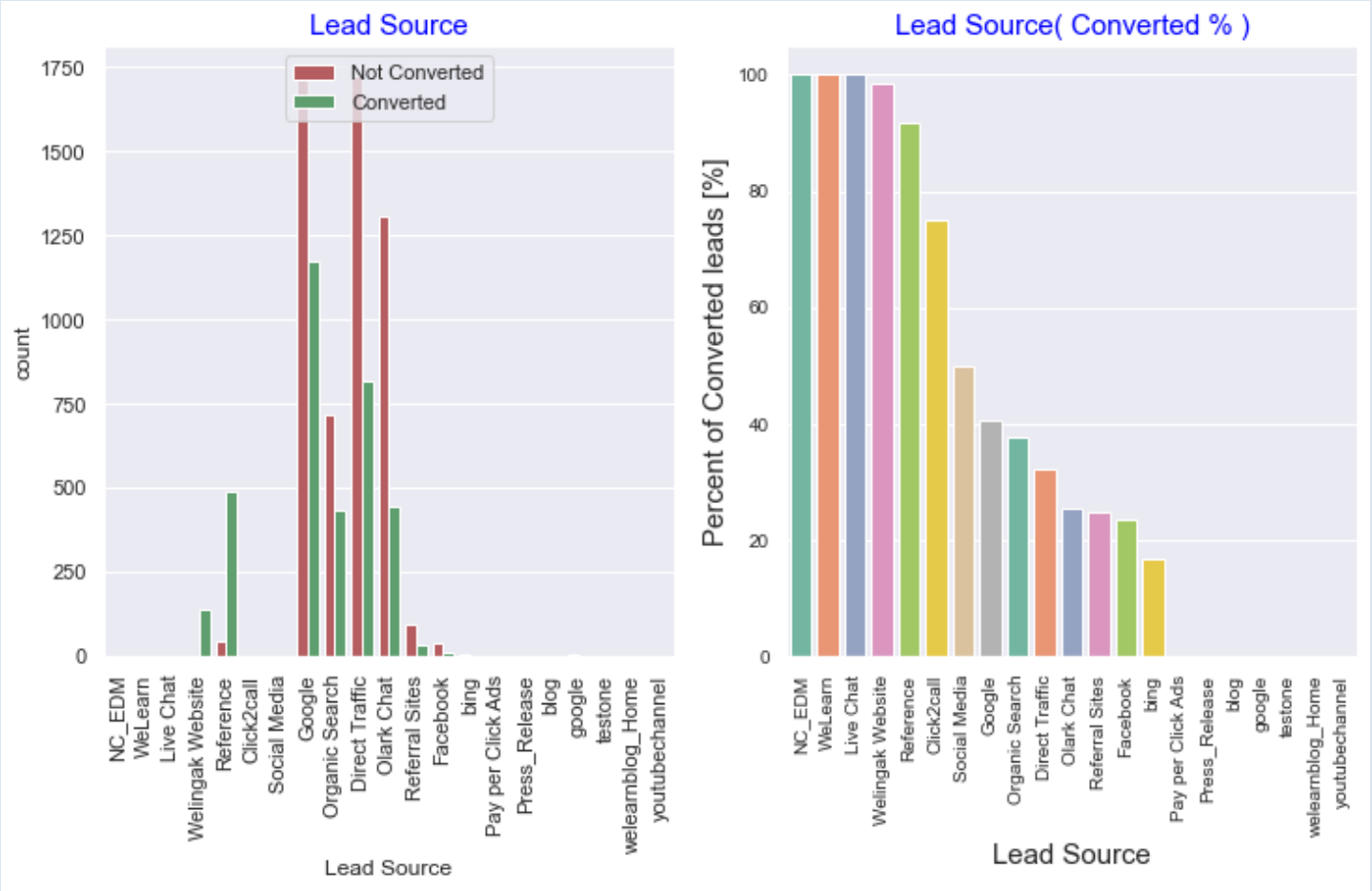
Exploratory data Analysis



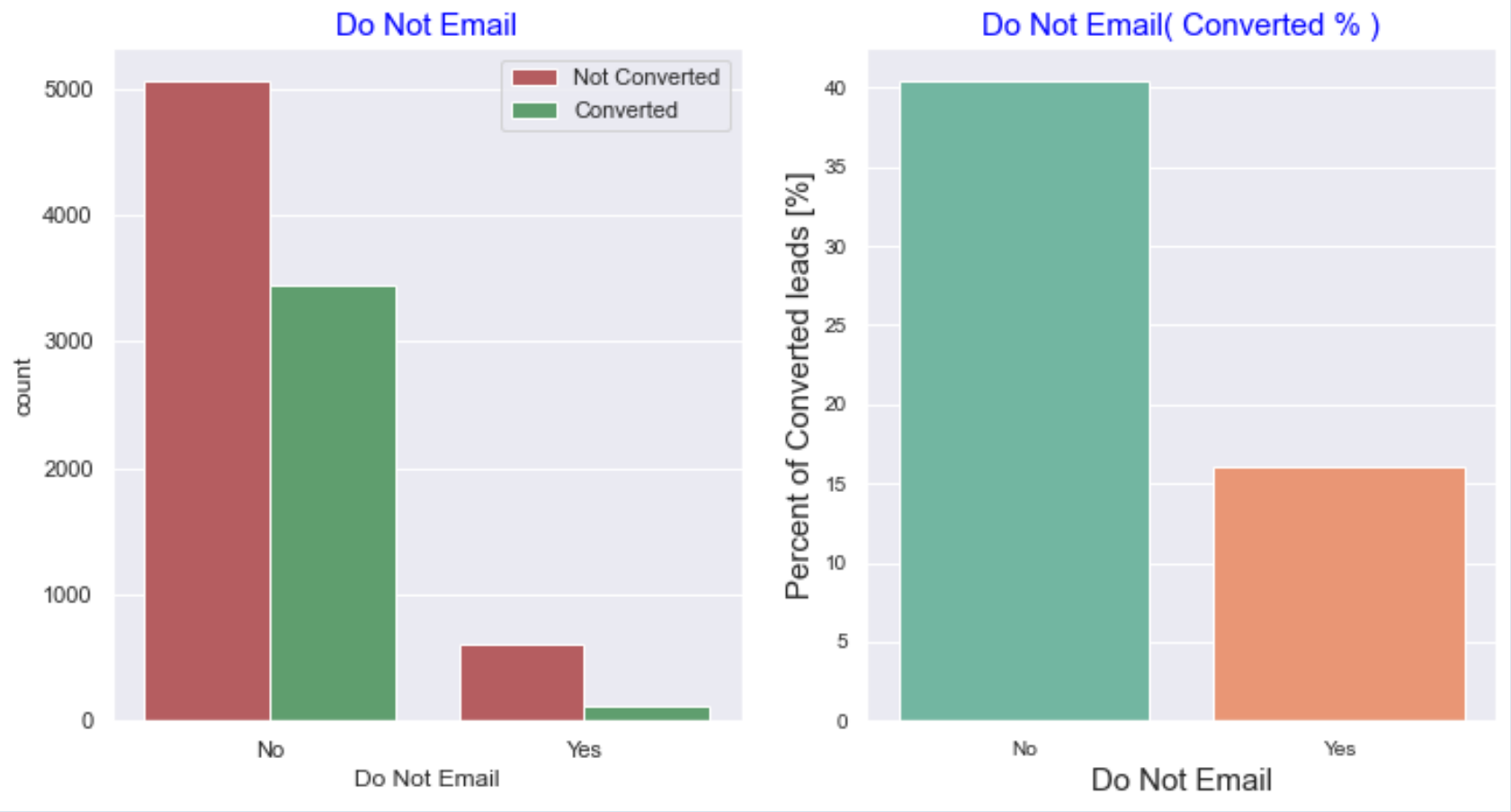
Univariate Analysis



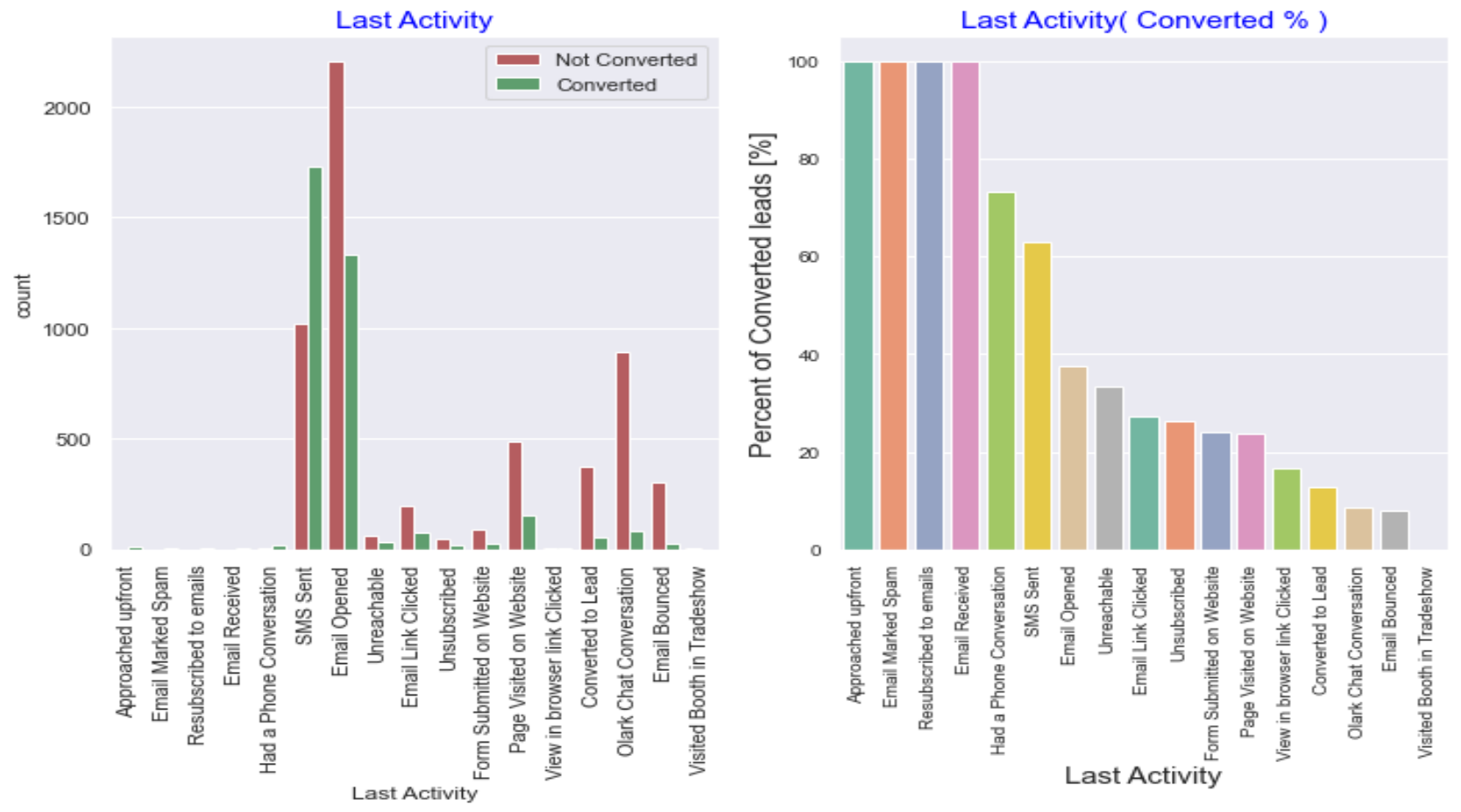
Lead Source



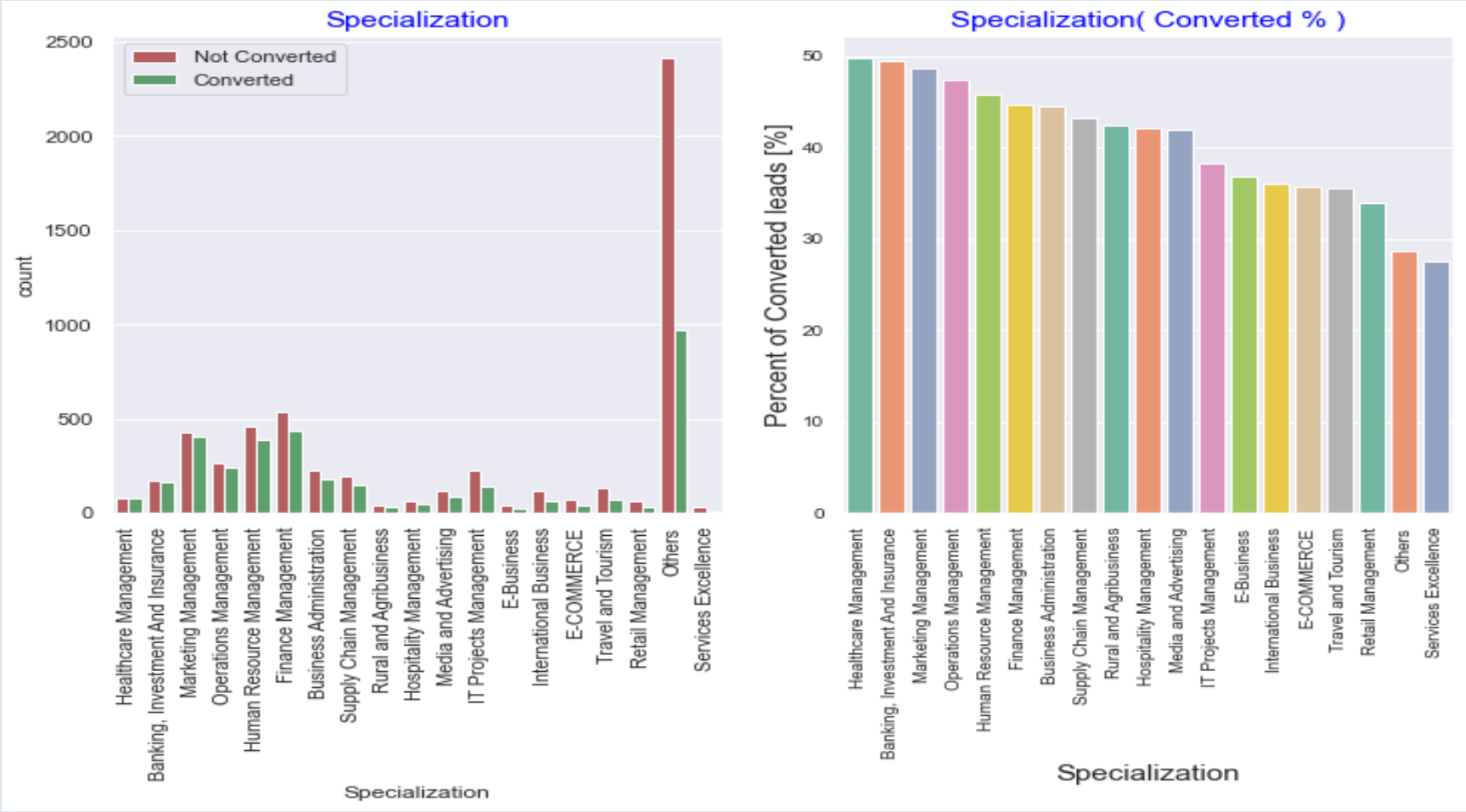
Do not Email



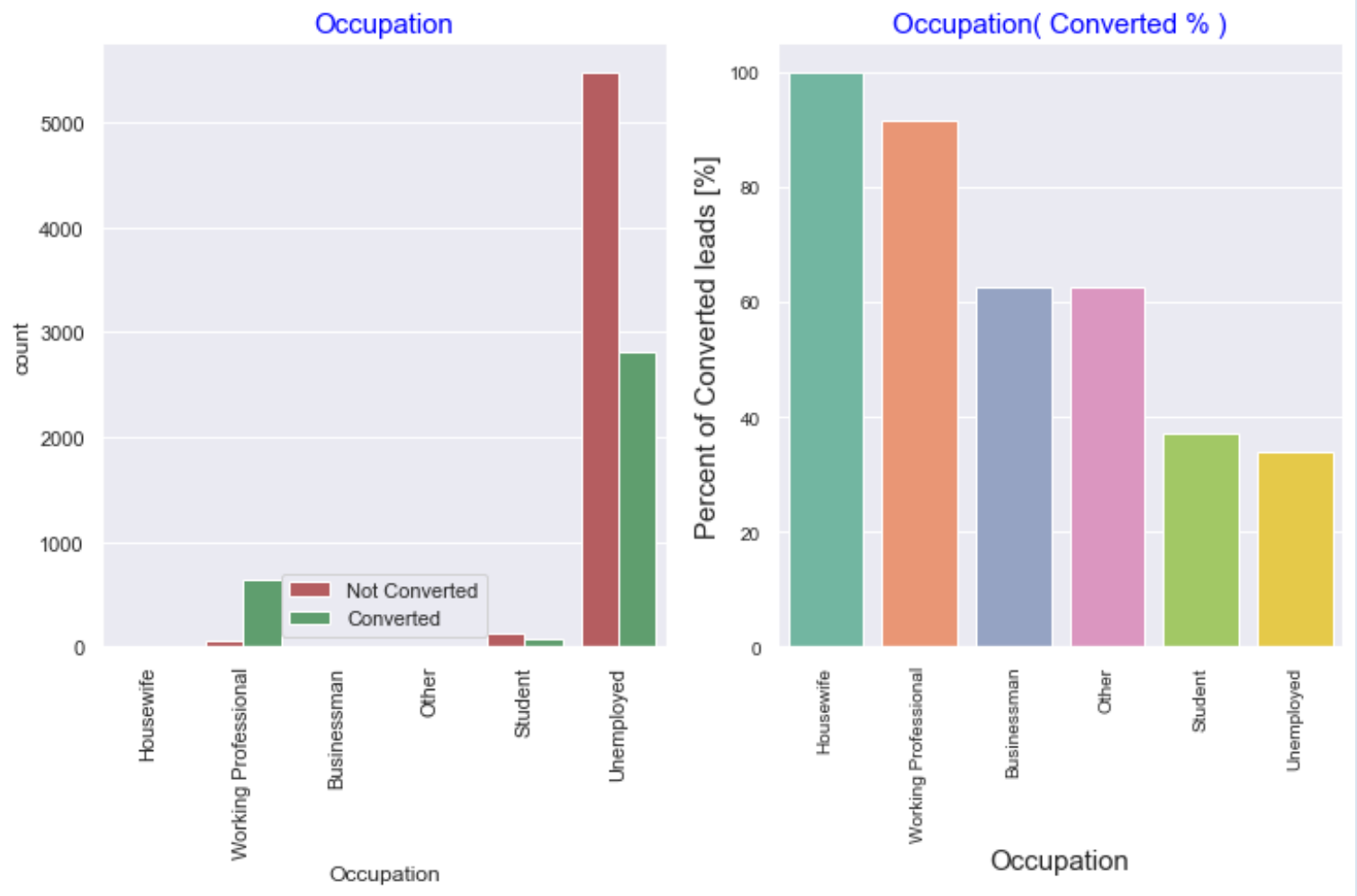
Last Activity



Specialization

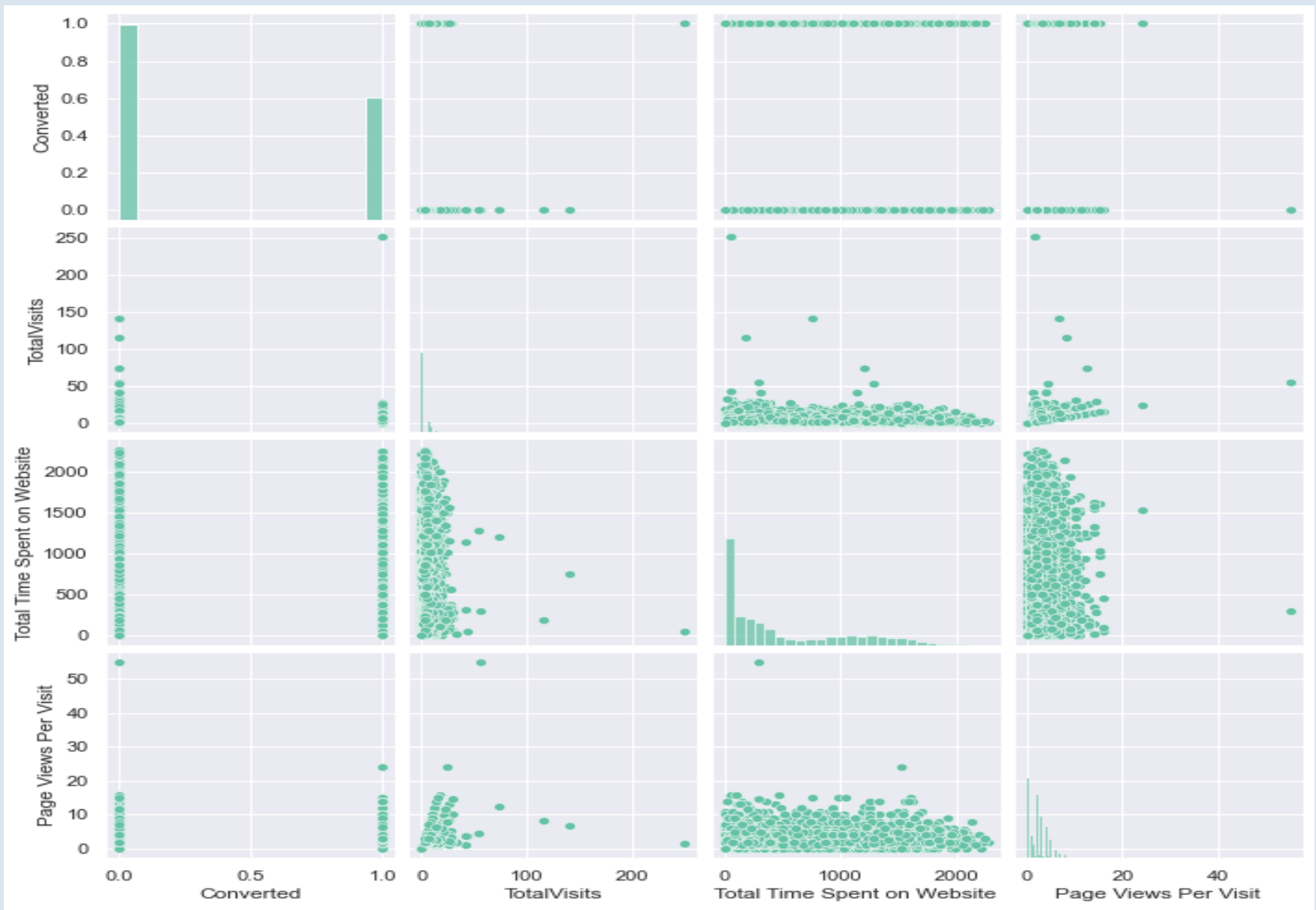


Occupation

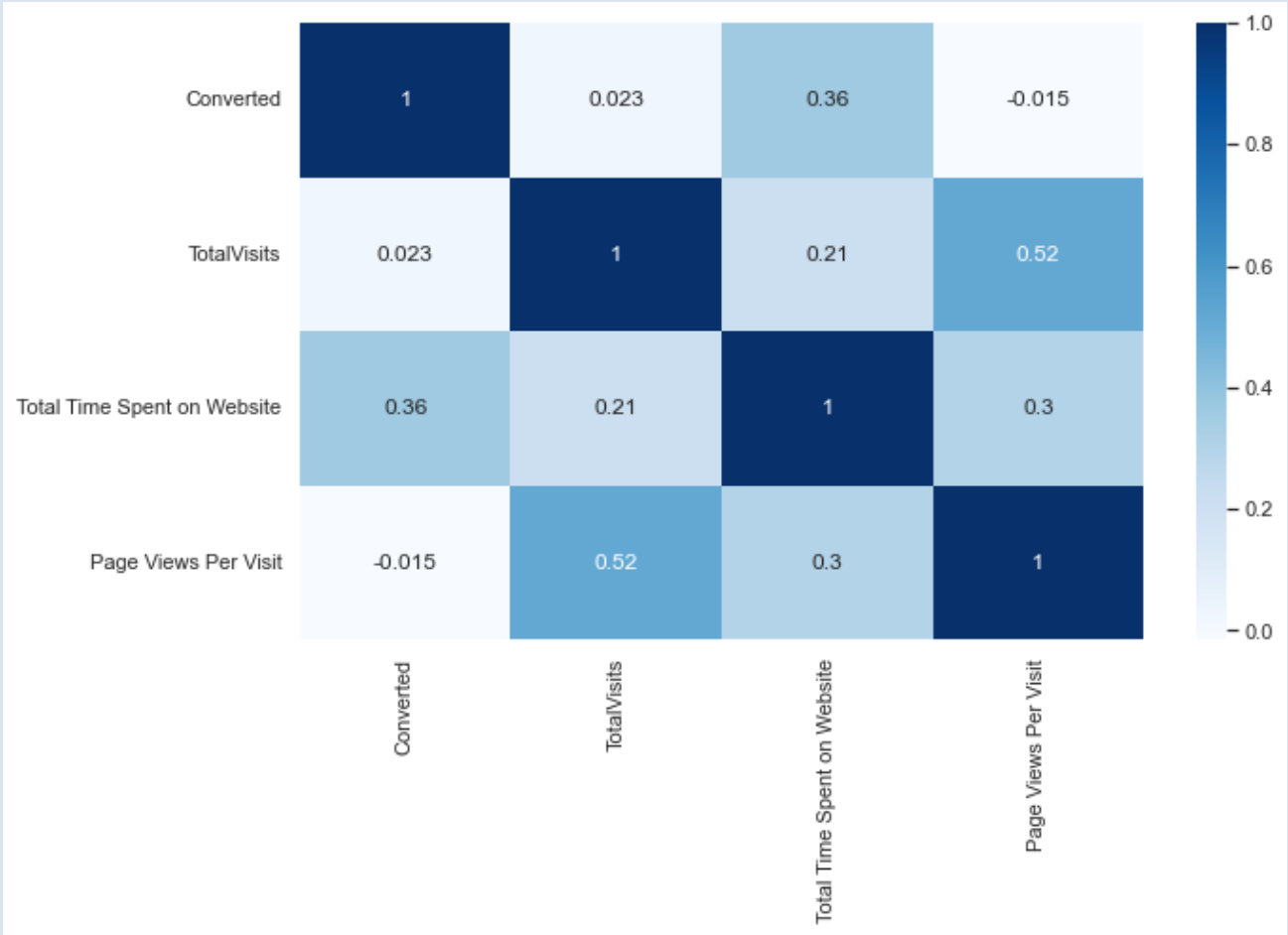


Free Copy

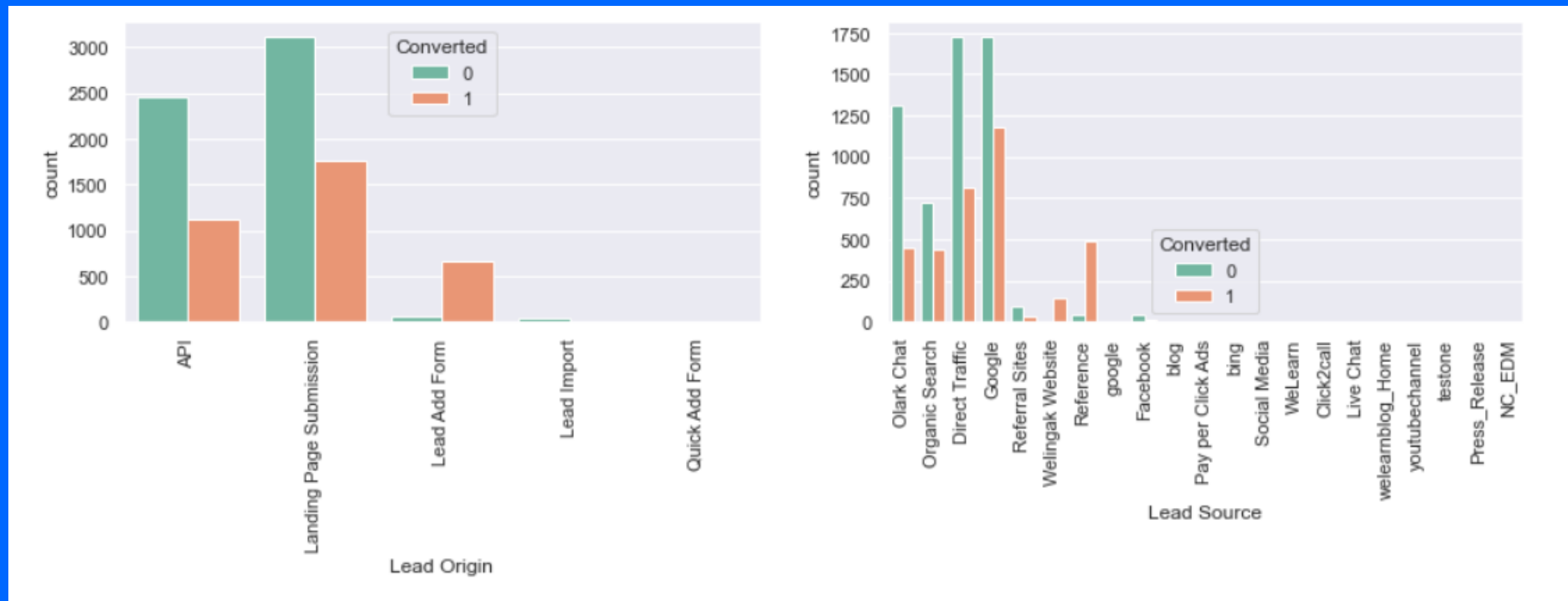


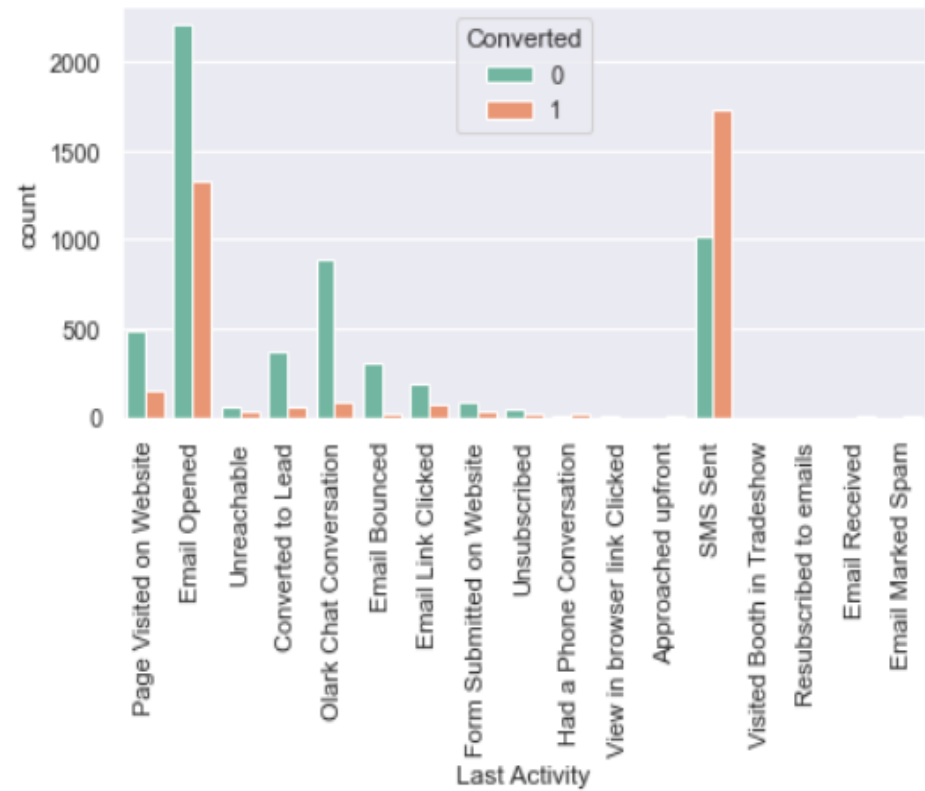
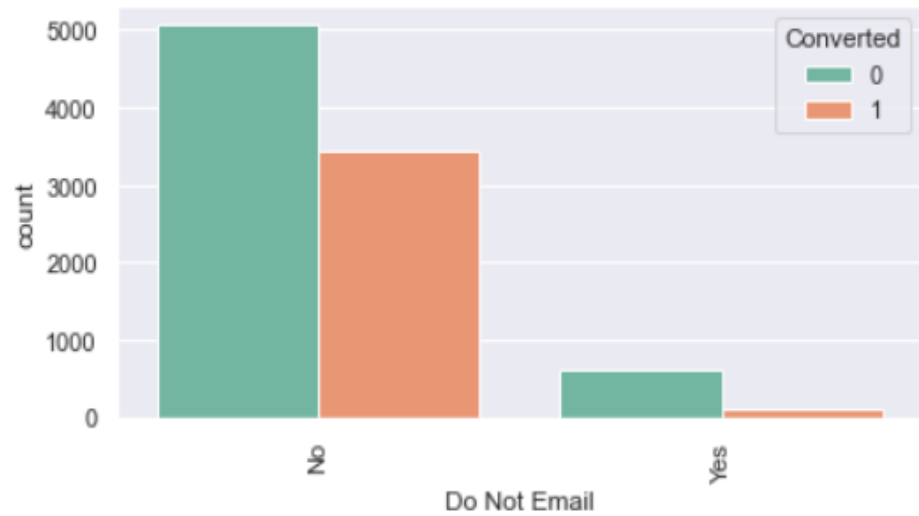


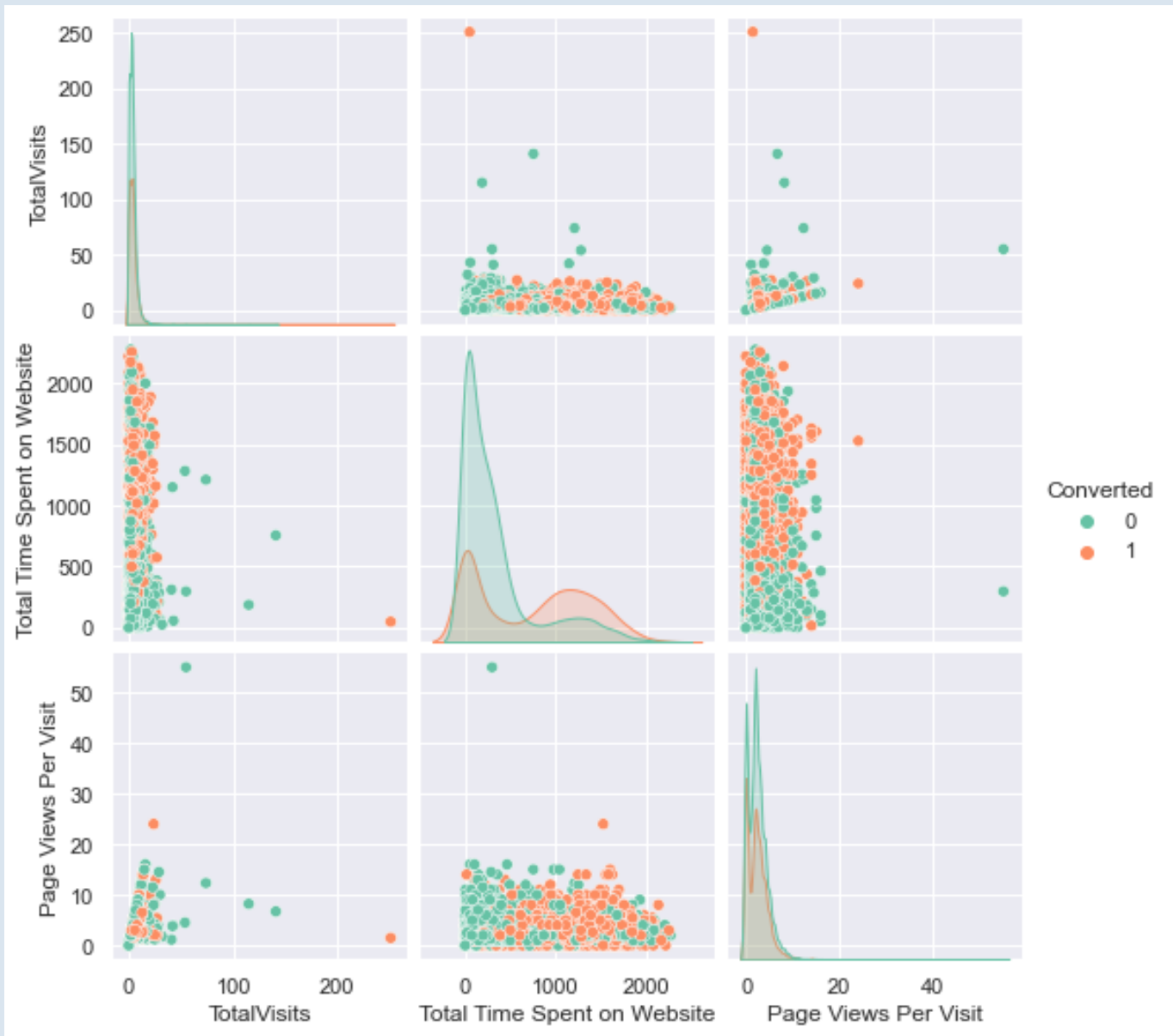
Corelation Matrix



Bivariant Analysis



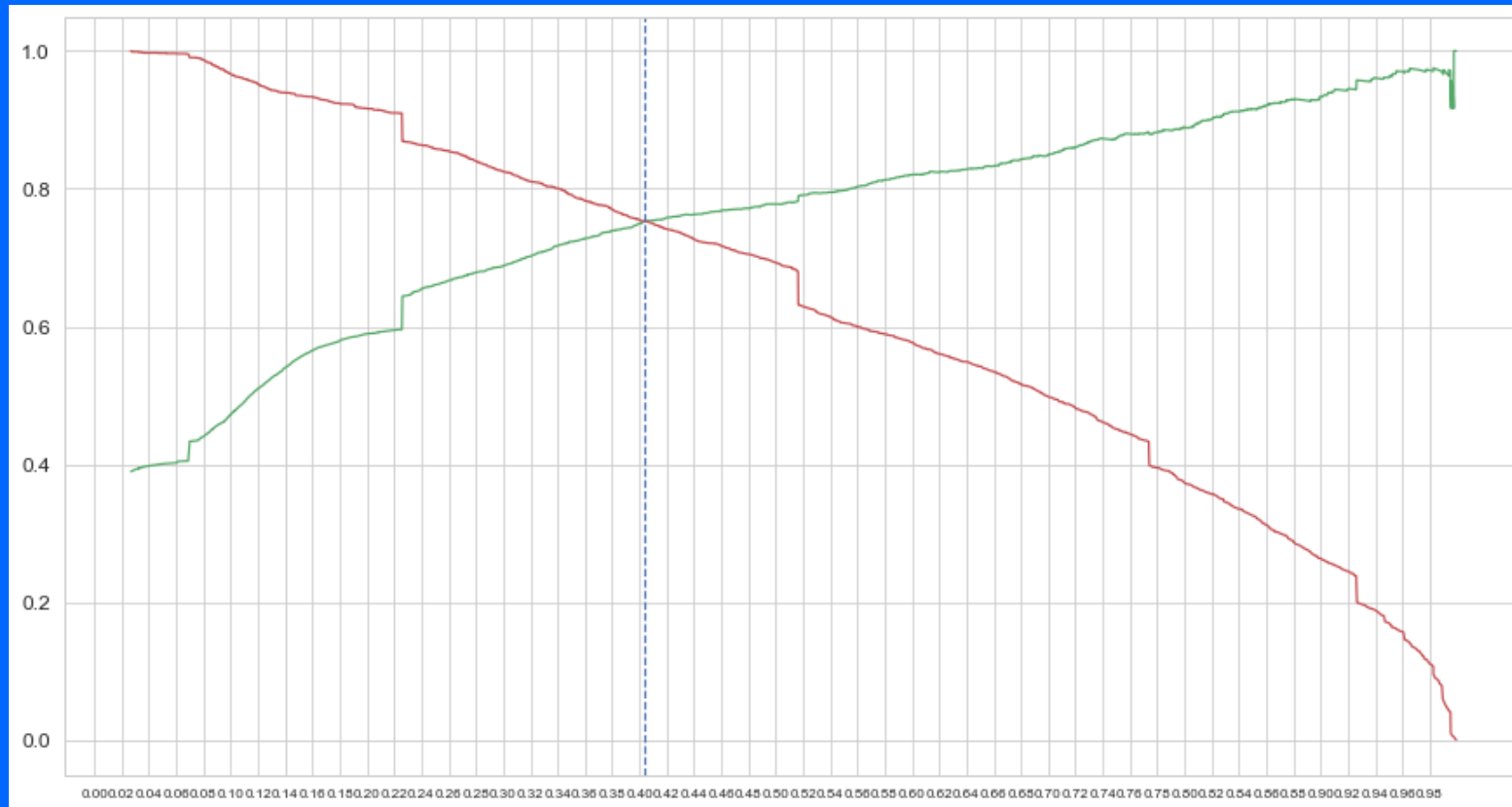




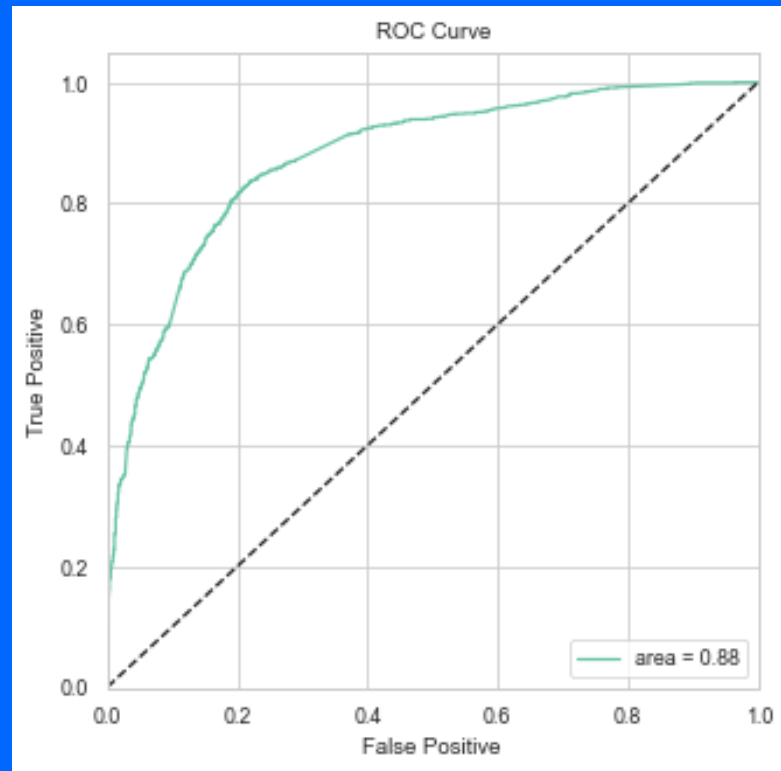
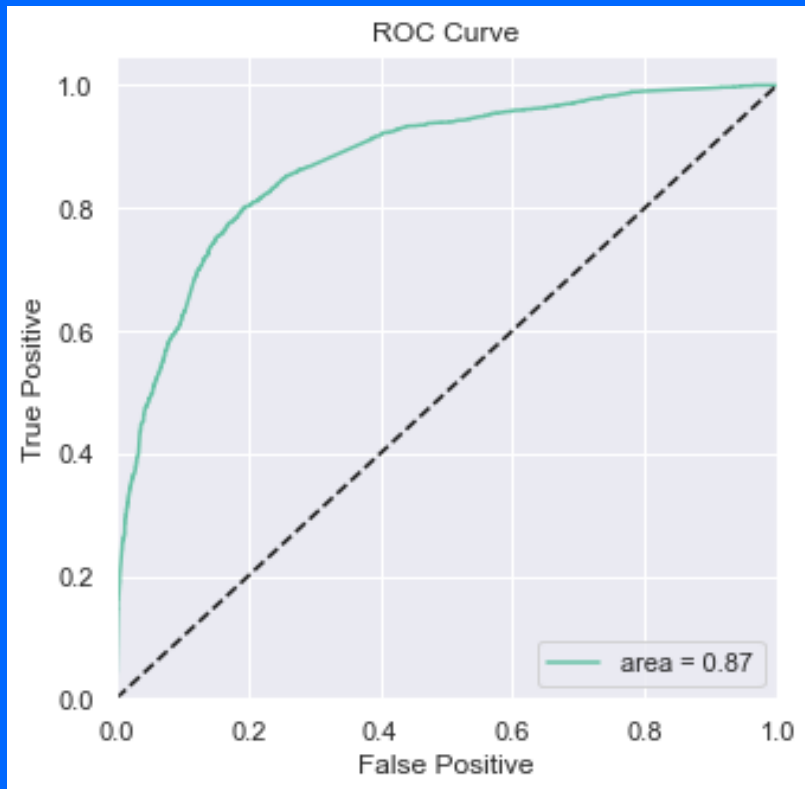
Model Building

- Created Logistic regression model with RFE for 15 features. And dropped the features having high p value and VIF. Final data have 10 features.
- Train and test data is split in 70:30 ratio
- Performed Feature Reduction to build model by removing the variable with p-value is greater than 0.05 and VIF value is greater than 5

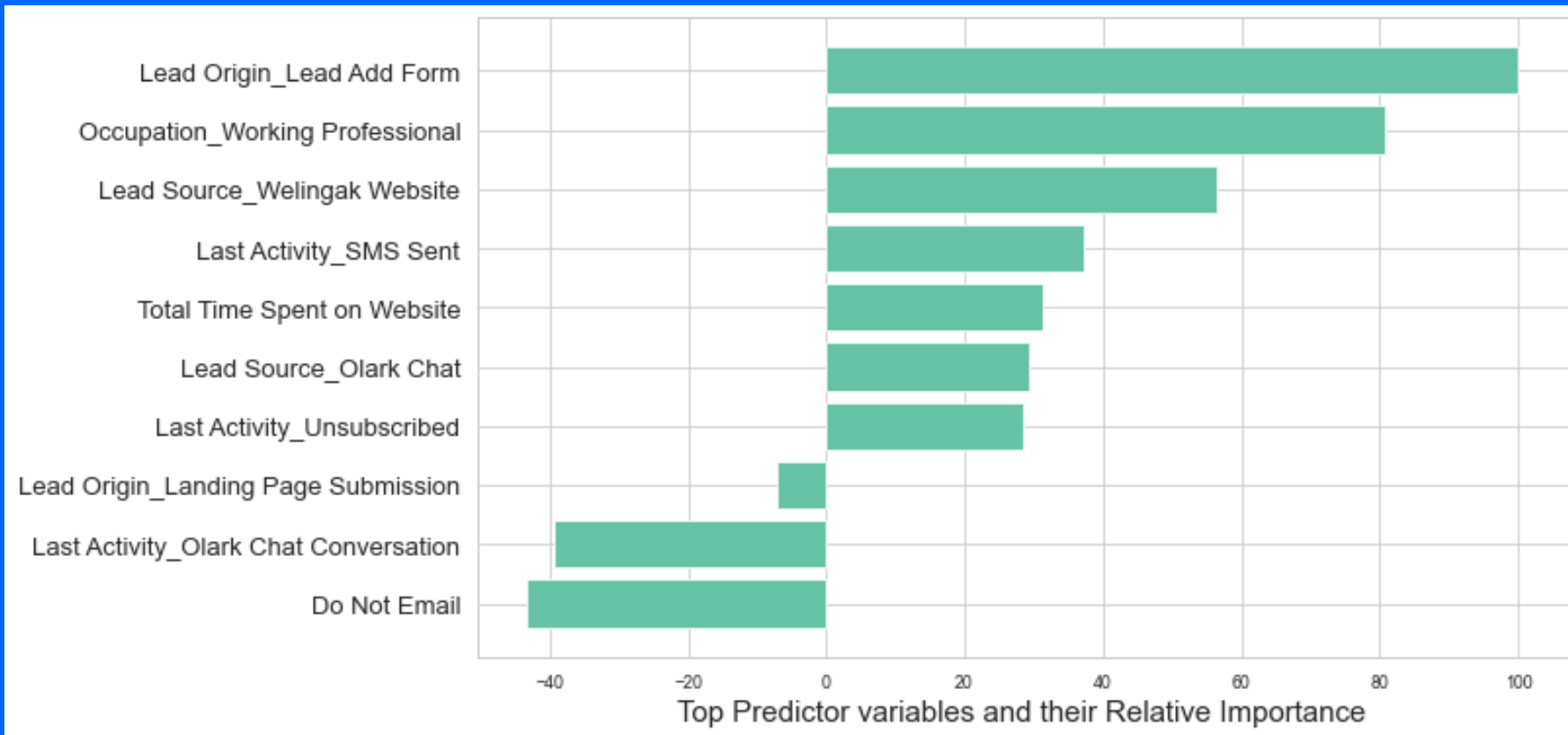
Model Evaluation



ROC Curve



Summary



Summary

- With a conversion rate of almost 90%, leads from the "Lead Add Form" rank third in terms of conversions.
- working professionals have a better conversion rate (about 90%), thus they should receive more attention while interacting.
- Welingaks website converts leads at a rate of about 98%. As it has a better likelihood of conversion, more adverts should be placed on this website to appeal to the leads from there.
- Over 90% of leads that originated from a "reference" converted. To boost the number of leads, we should entice and reward current members to recommend new people.



Thank you