Course 2: Machine Learning - I

Case Study

Leads Scoring

Group Members
Shreya
Sujash Jain
Ganesh S Diniyan

Introduction

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.

Strategy

Advertisement on websites like Google , social media platform

Referrals

Potential customer visits web / social media platform , Google

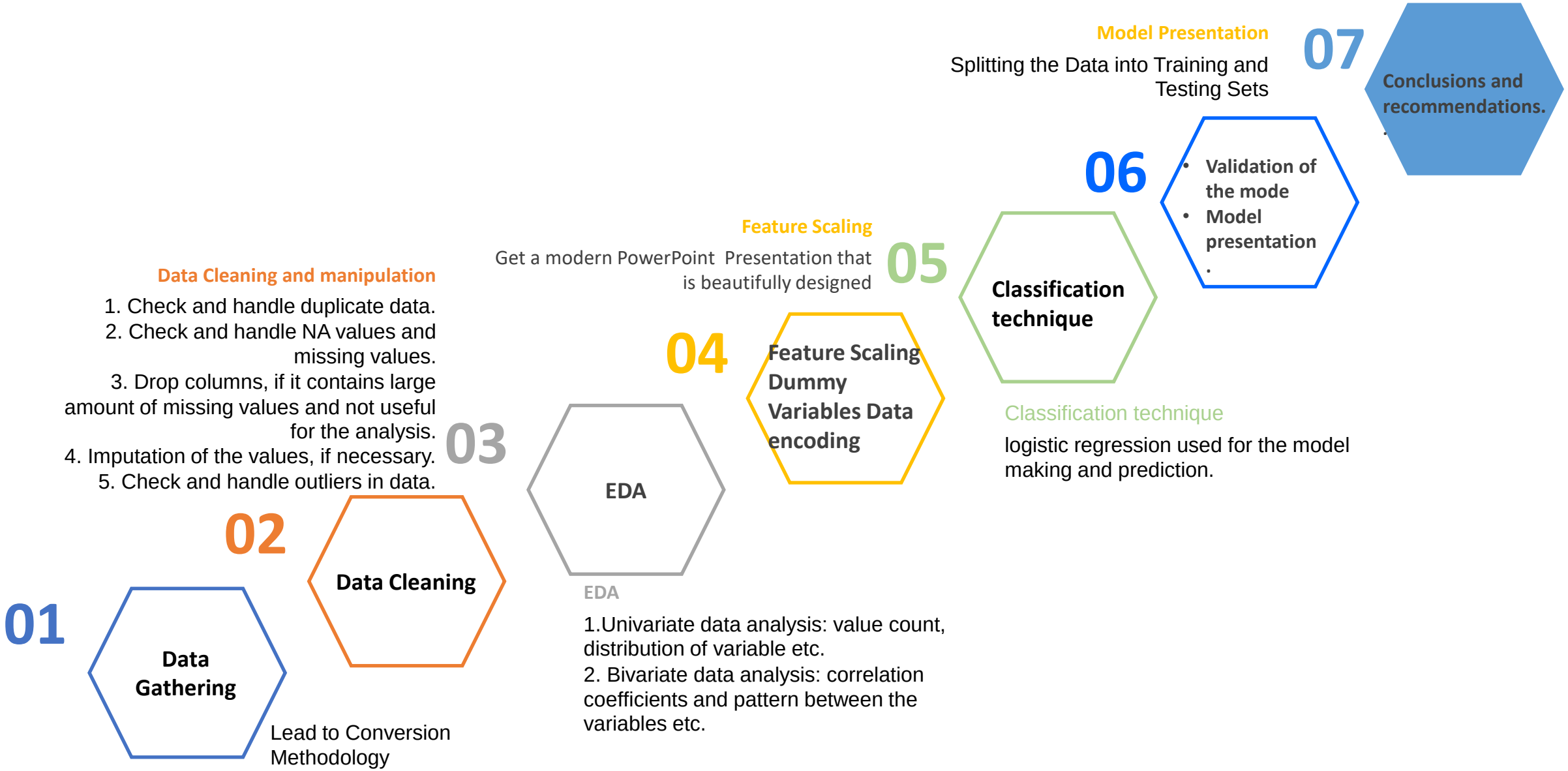
Collection of visitors Information as received from web / social media platform

Converting Leads by connecting via calls , E-Mails or Webinars

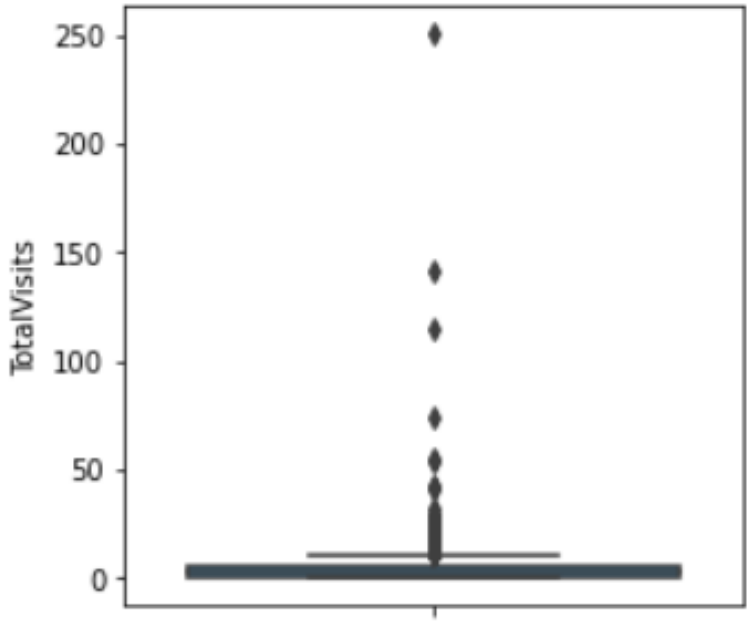
30% leads get converted

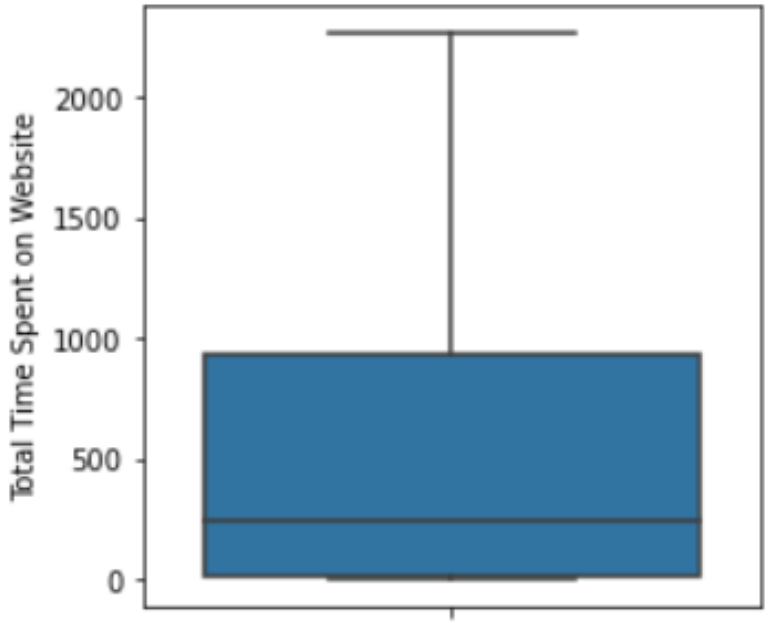
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic
- Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

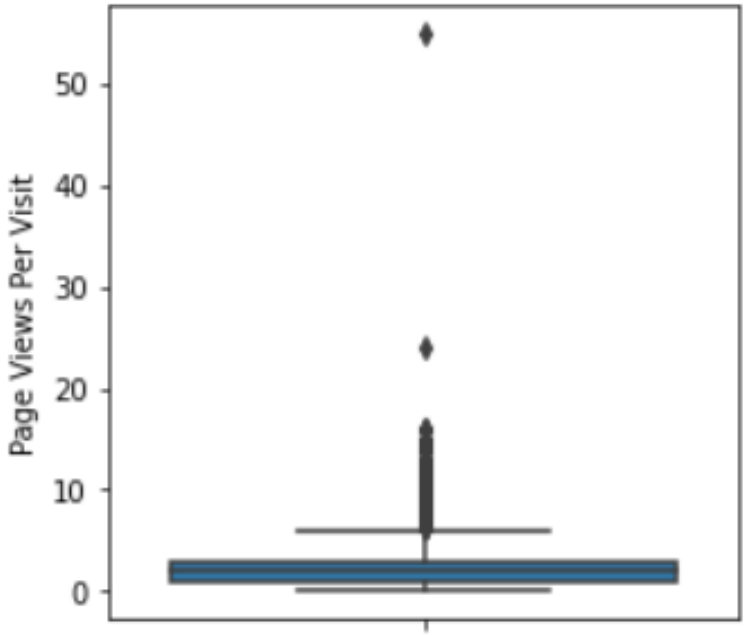
Solution Methodology



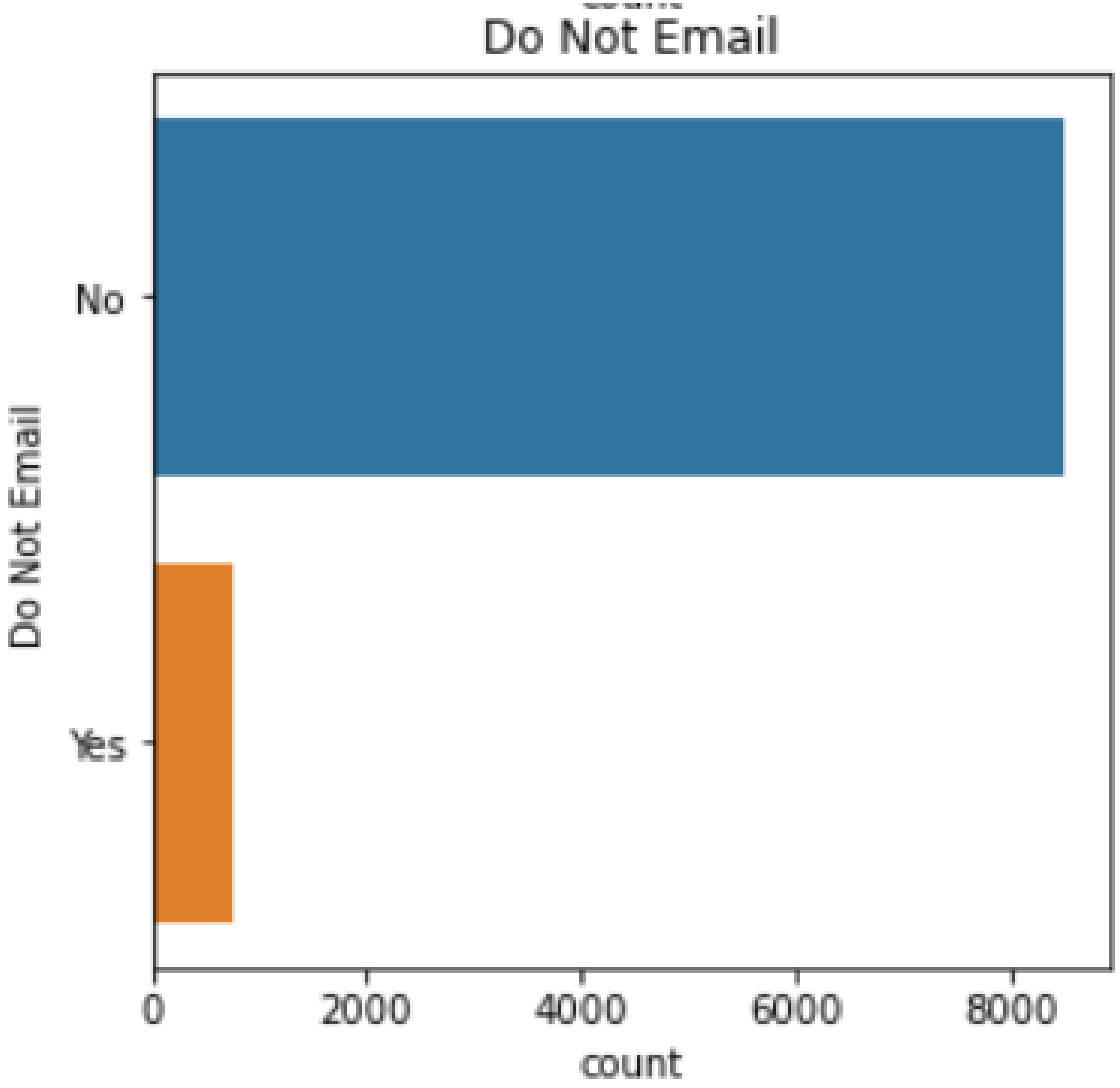
EDA – Numerical Data

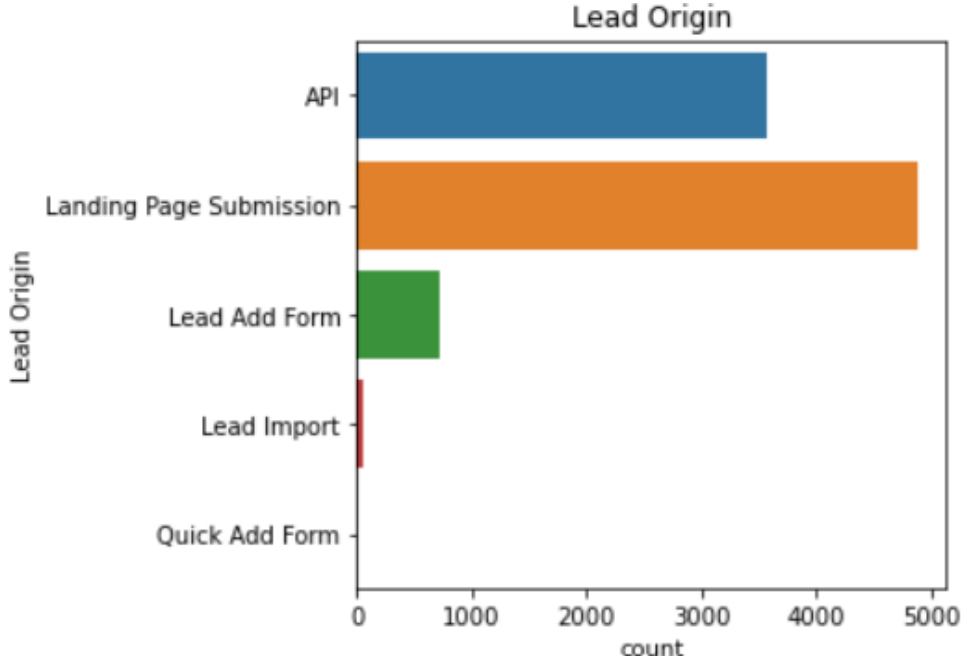


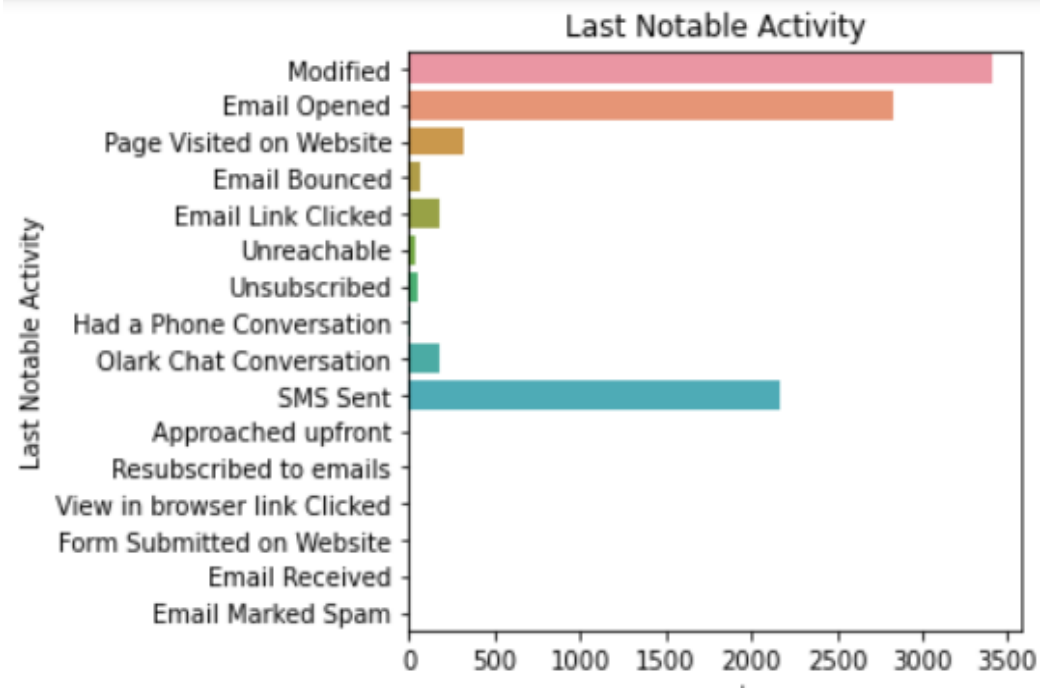




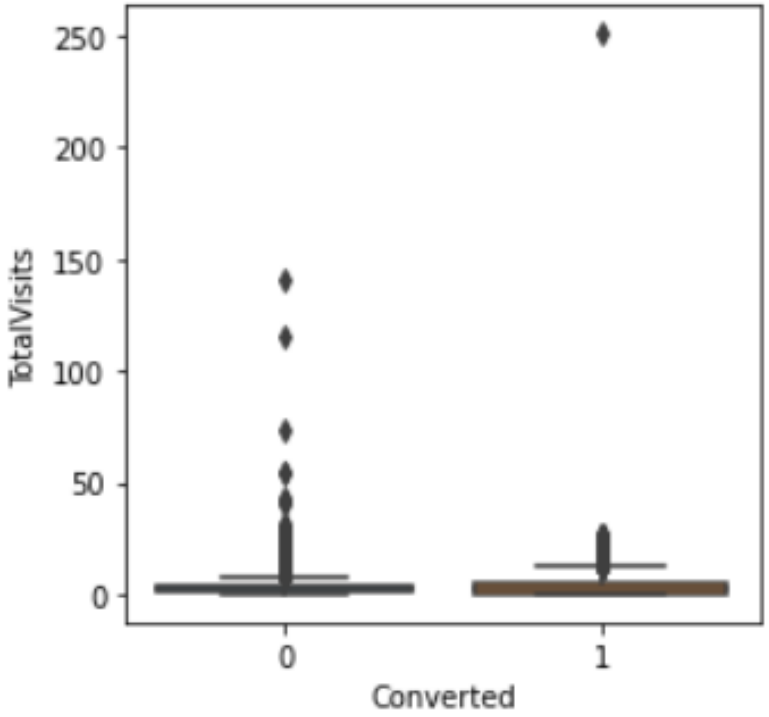
EDA – Categorical Data

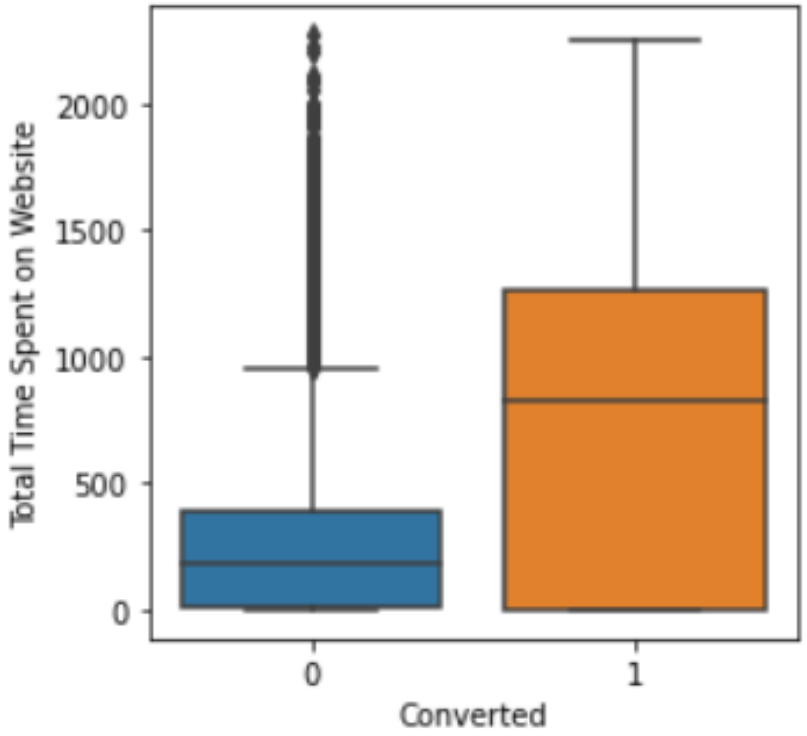


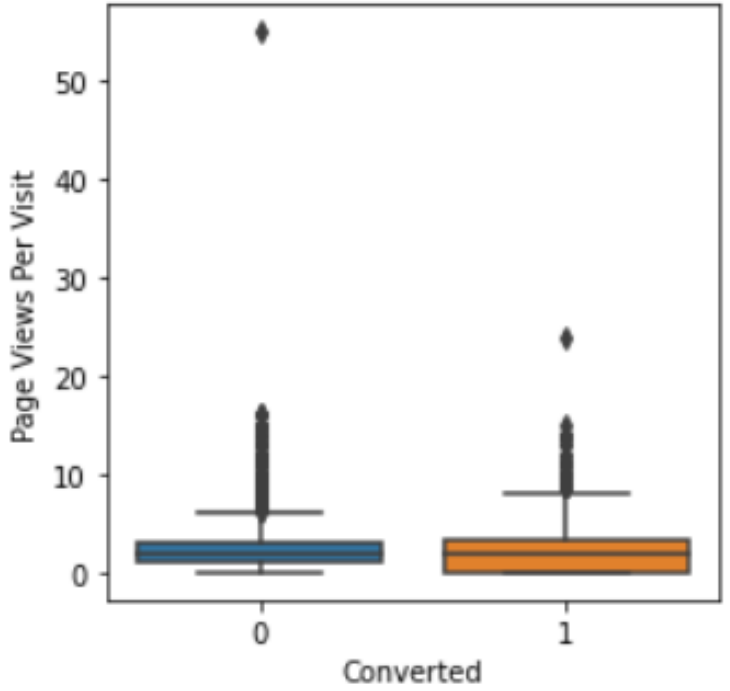




EDA – Numerical Data Conversion

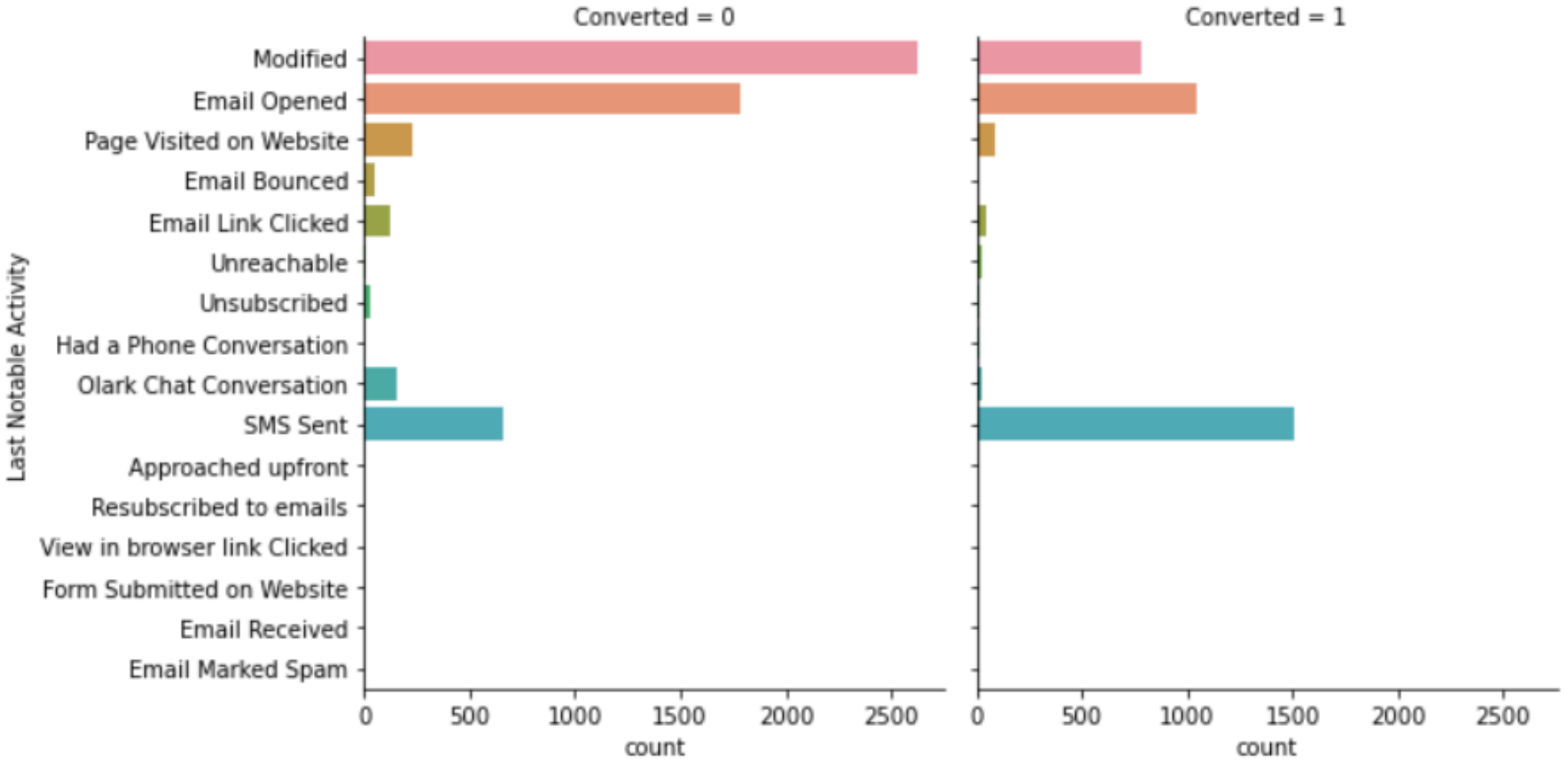




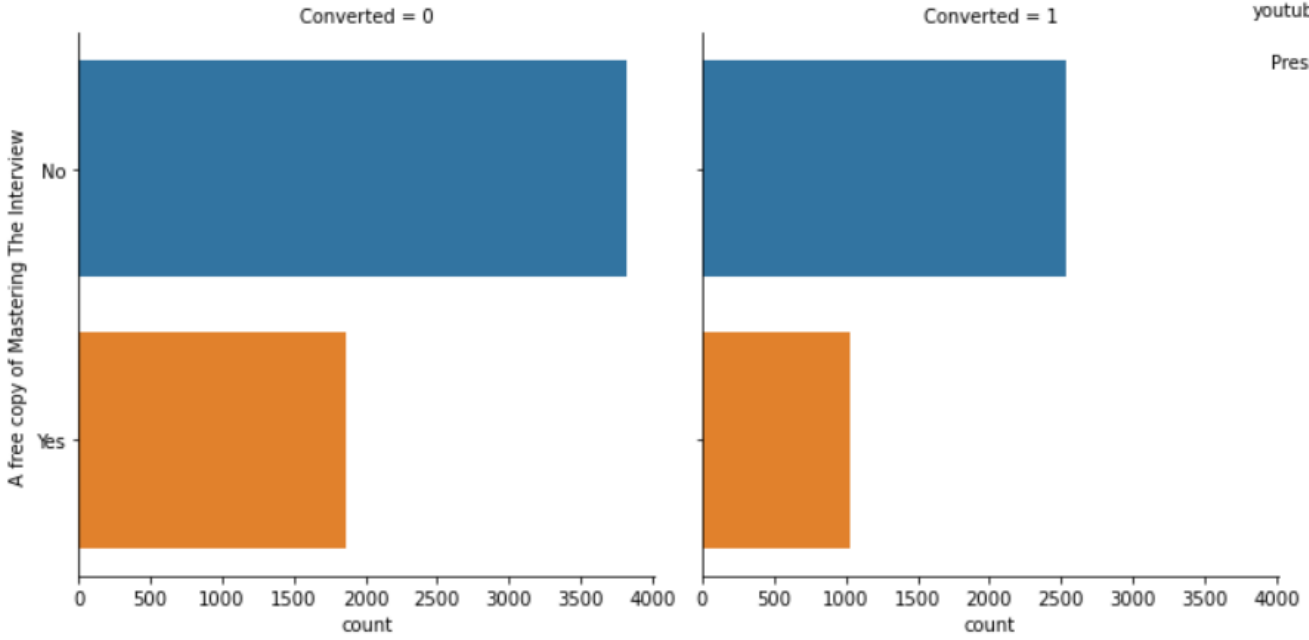


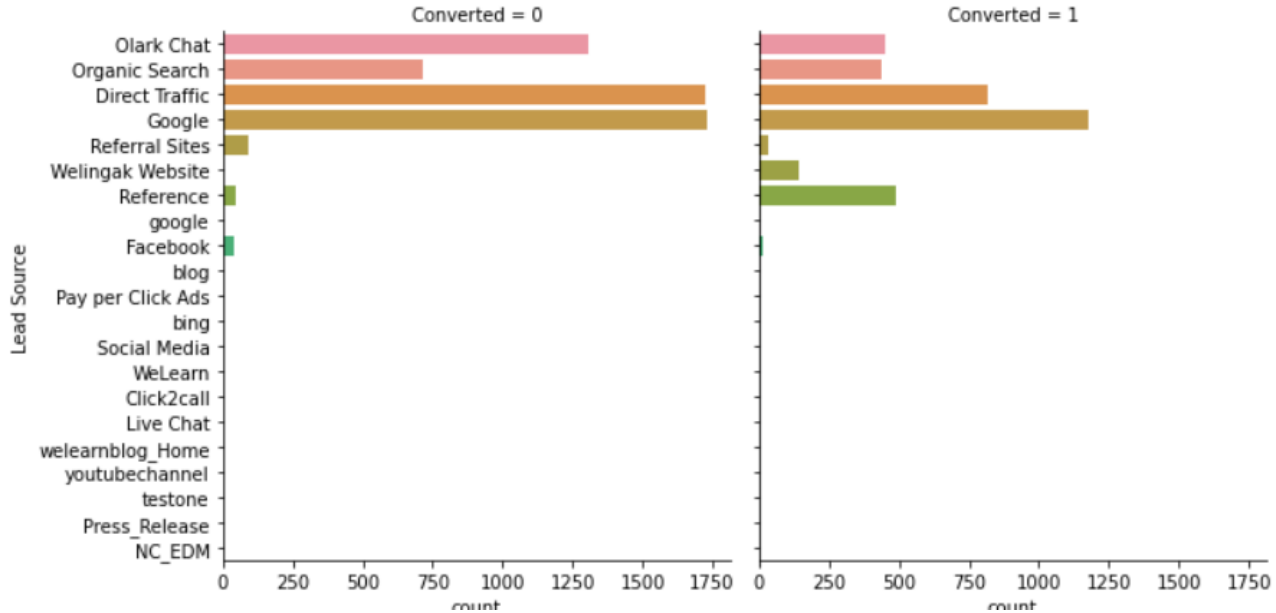
EDA plots depicting variation in numerical columns for those who Converted and those who didn't.

EDA - Categorical Data Conversion

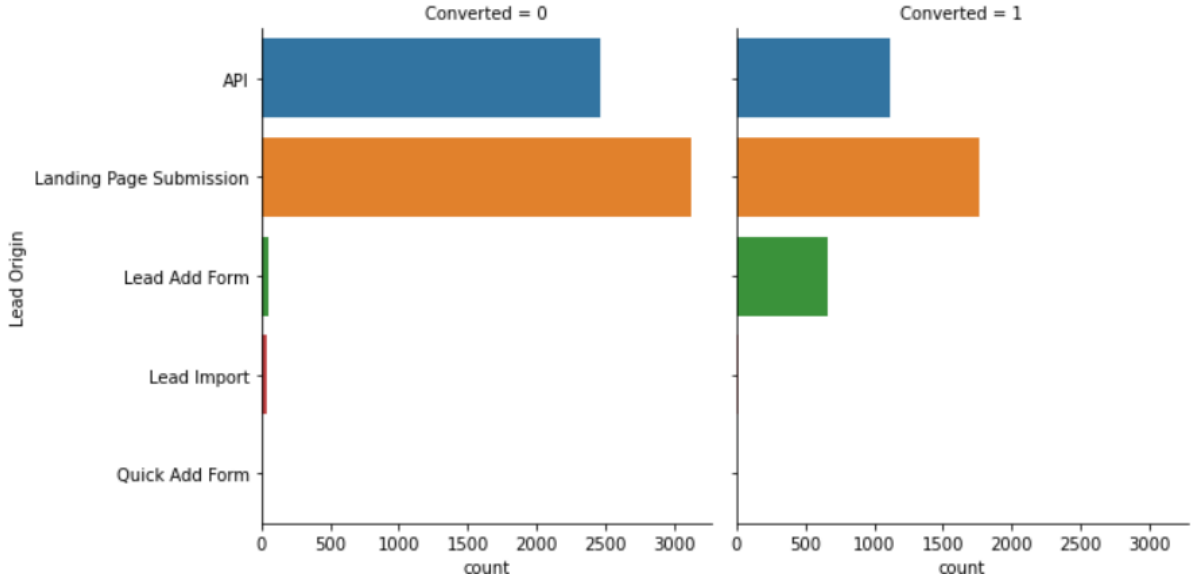


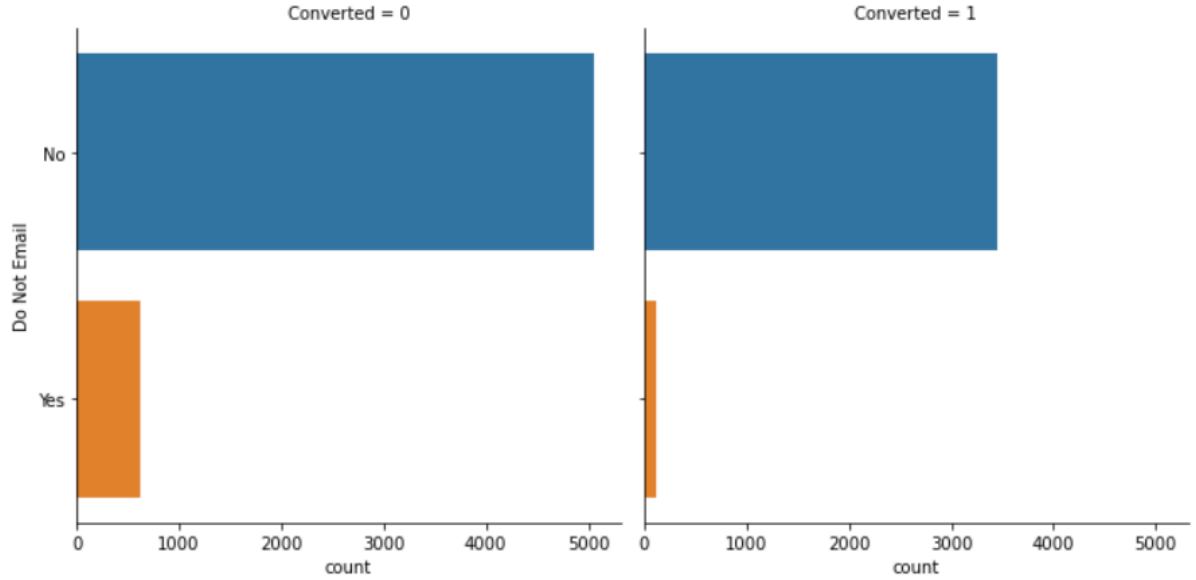
EDA - Categorical Data Conversion



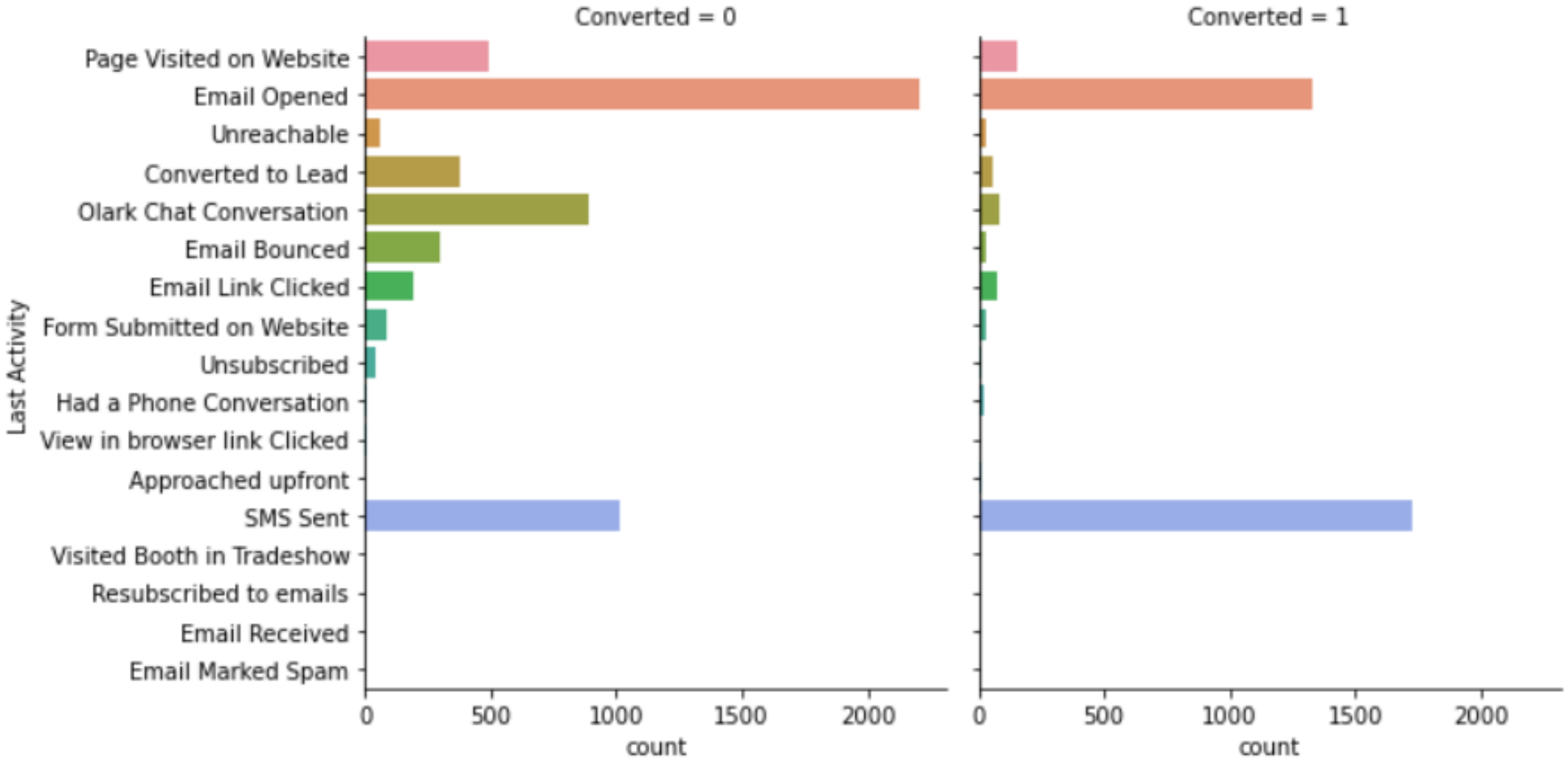


EDA - Categorical Data Conversion

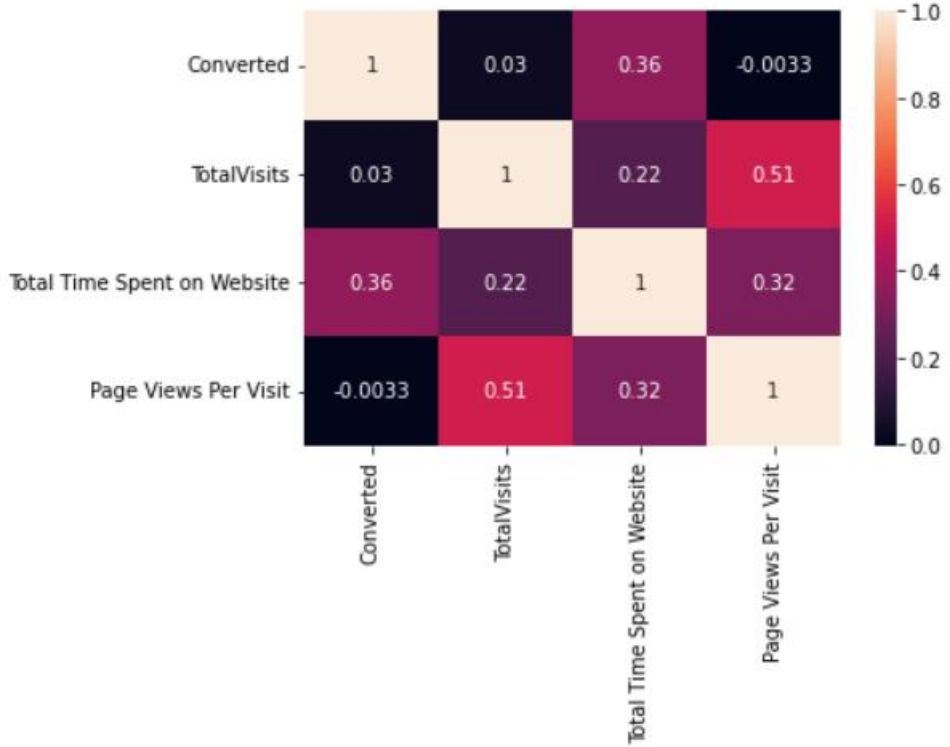


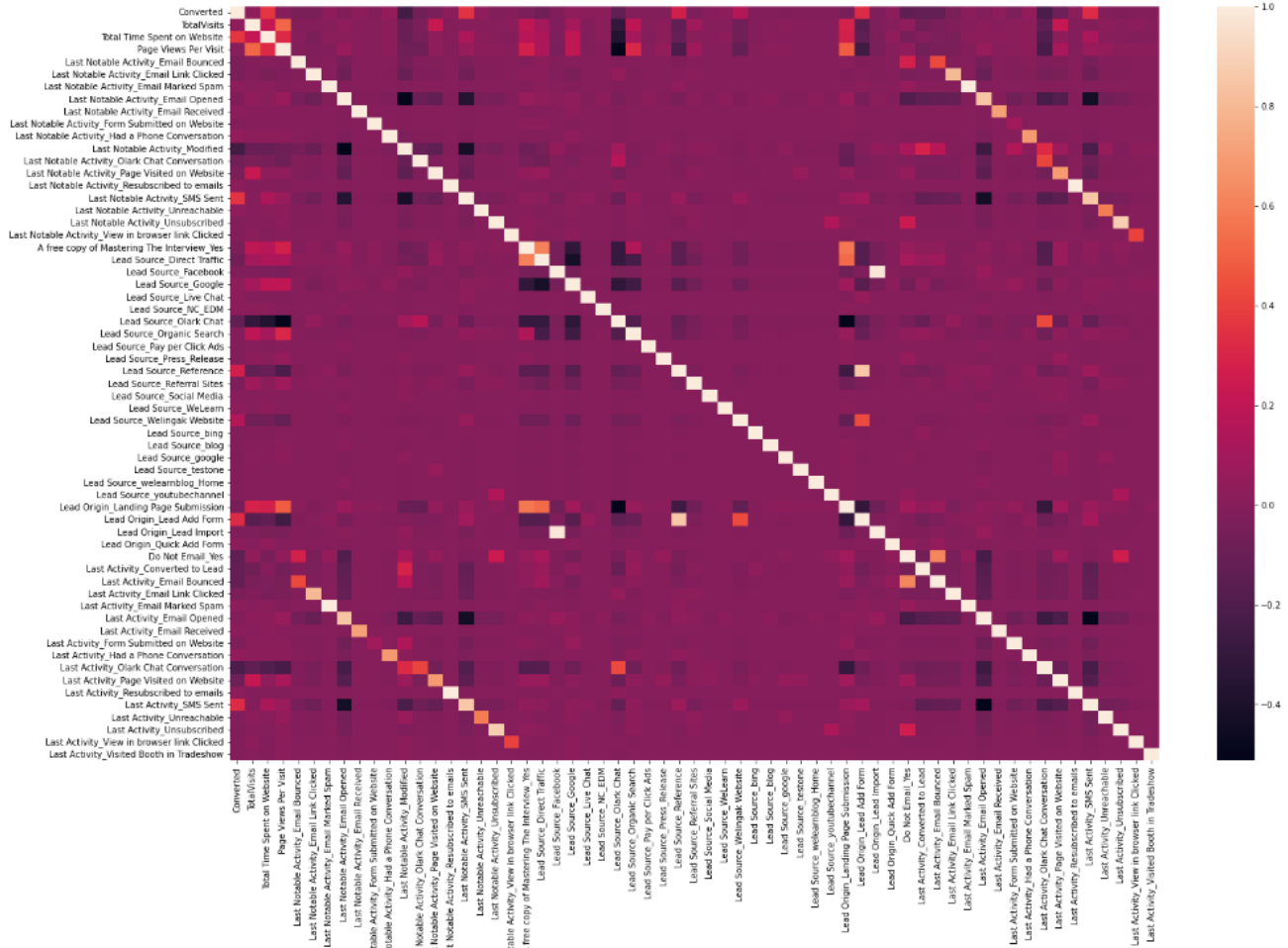


EDA - Categorical Data Conversion



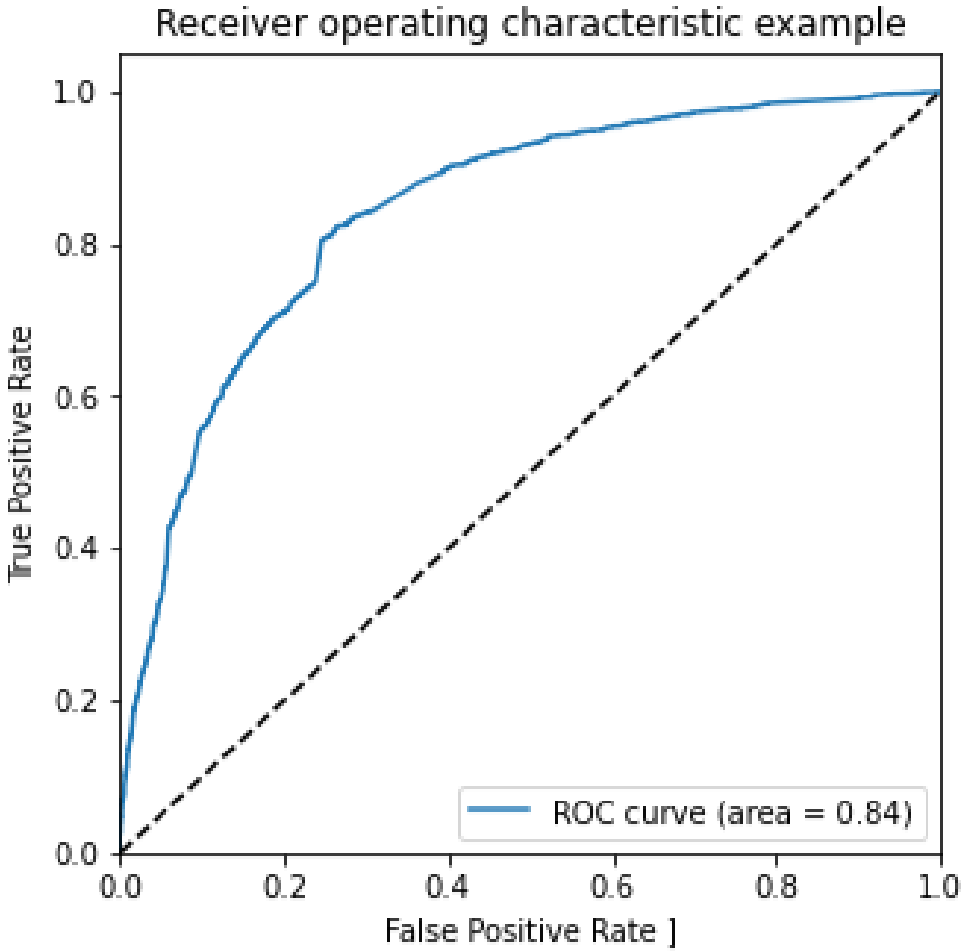
EDA –(Heat Map) Correlation of all selected columns

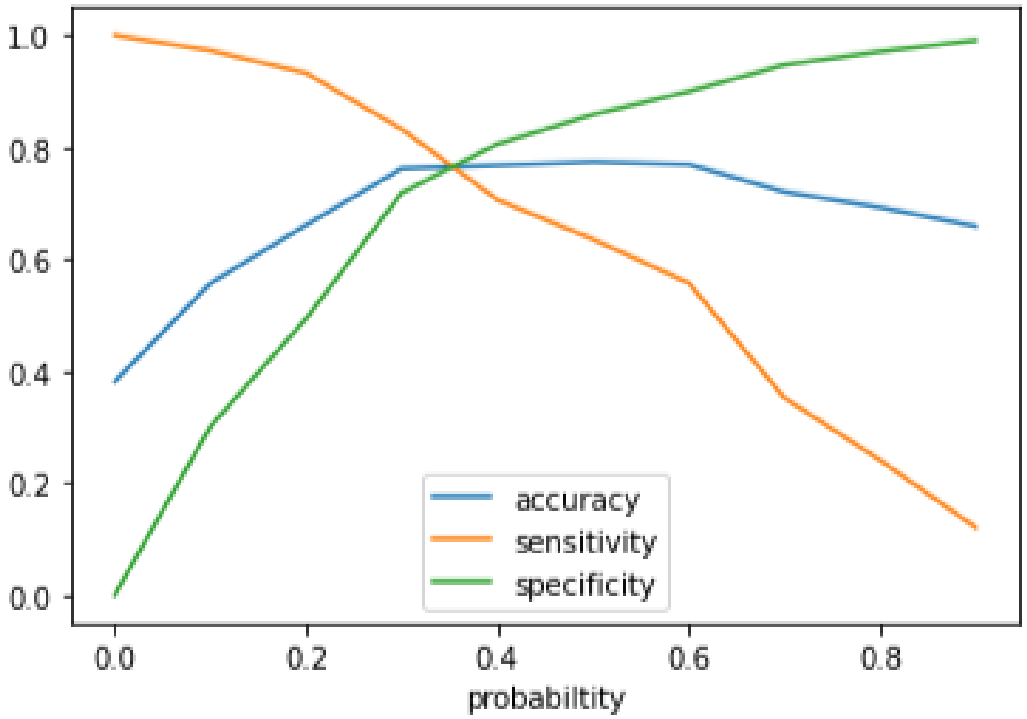




Model Building & Evaluation - Sensitivity and Specificity on Train Data Set

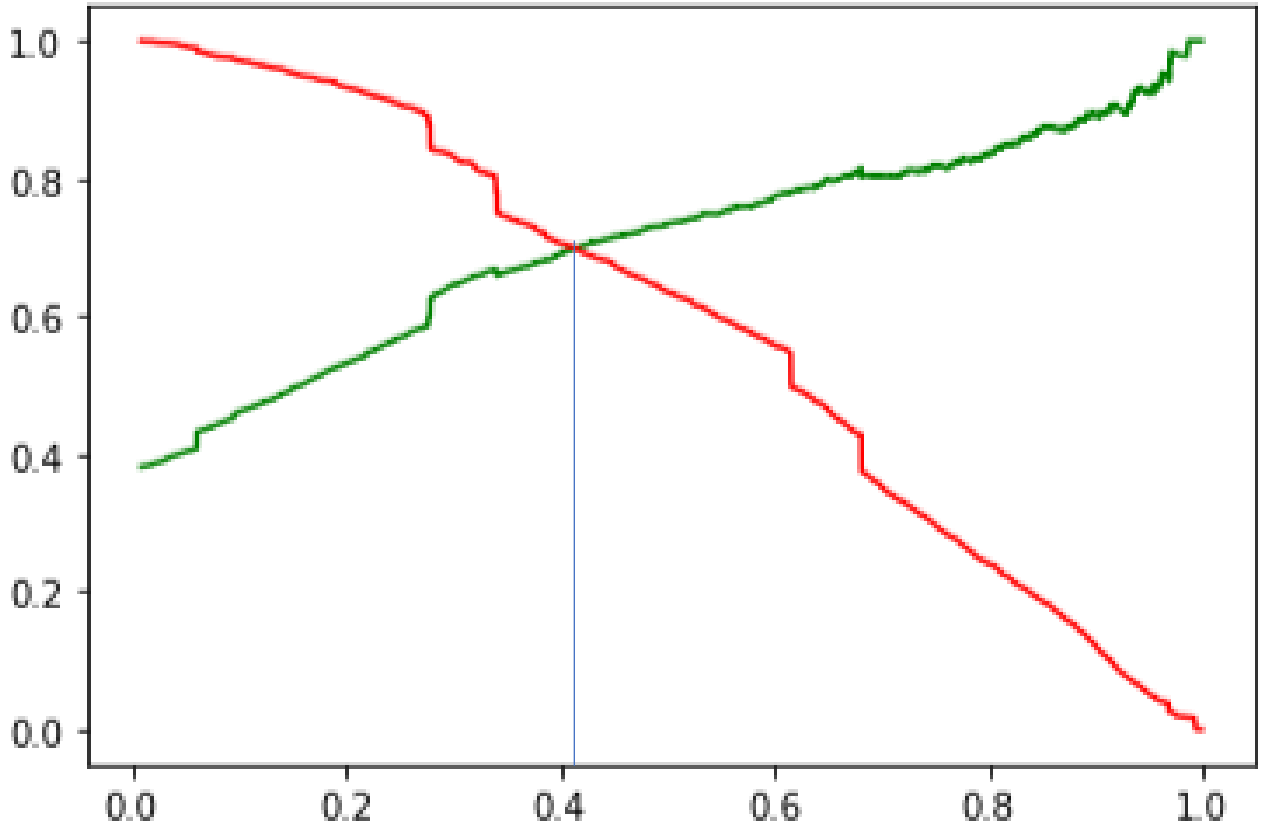
ROC Curve





The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity

Model Evaluation- Precision and Recall on Train Dataset



Precision - 79 %
Recall - 71 %

The graph depicts an optimal cut off of 0.42 based on Precision and Recall

Conclusion & Recommendation

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction. –
- Accuracy, Sensitivity and Specificity values of test set are around 81%, 79% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set.
- The top 3 variables that contribute for lead getting converted in the model are -
    1. Total time spent on website - **Increase user engagement**
    2. Lead Add Form from Lead Origin - **Improve the Olark Chat service**
    3. Had a Phone Conversation from Last Notable Activity - **Increase on sending SMS notifications**
- Hence overall this model looks good.

Thank You