X Education - Lead Scoring Case Study

Identification of Hot Leads to focus more on them and thus enhancing the conversion ratio for X Education

BY C.Arathi, Murtaza Manzoor And Tanmai Bhatnagar

X Education Company

- X Education , An education company named sells online courses to industry professionals
- Many interested professionals land on their website
- The company markets its courses on several websites like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos

X Education Company

- When these people fill up a form providing their email address or phone number, they are classified to be a lead
- 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%

Problem Statement

- X Education gets a lot of leads but its lead conversion rate is very poor
- 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

Problem Statement

- We will help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- We are required to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.

Solution

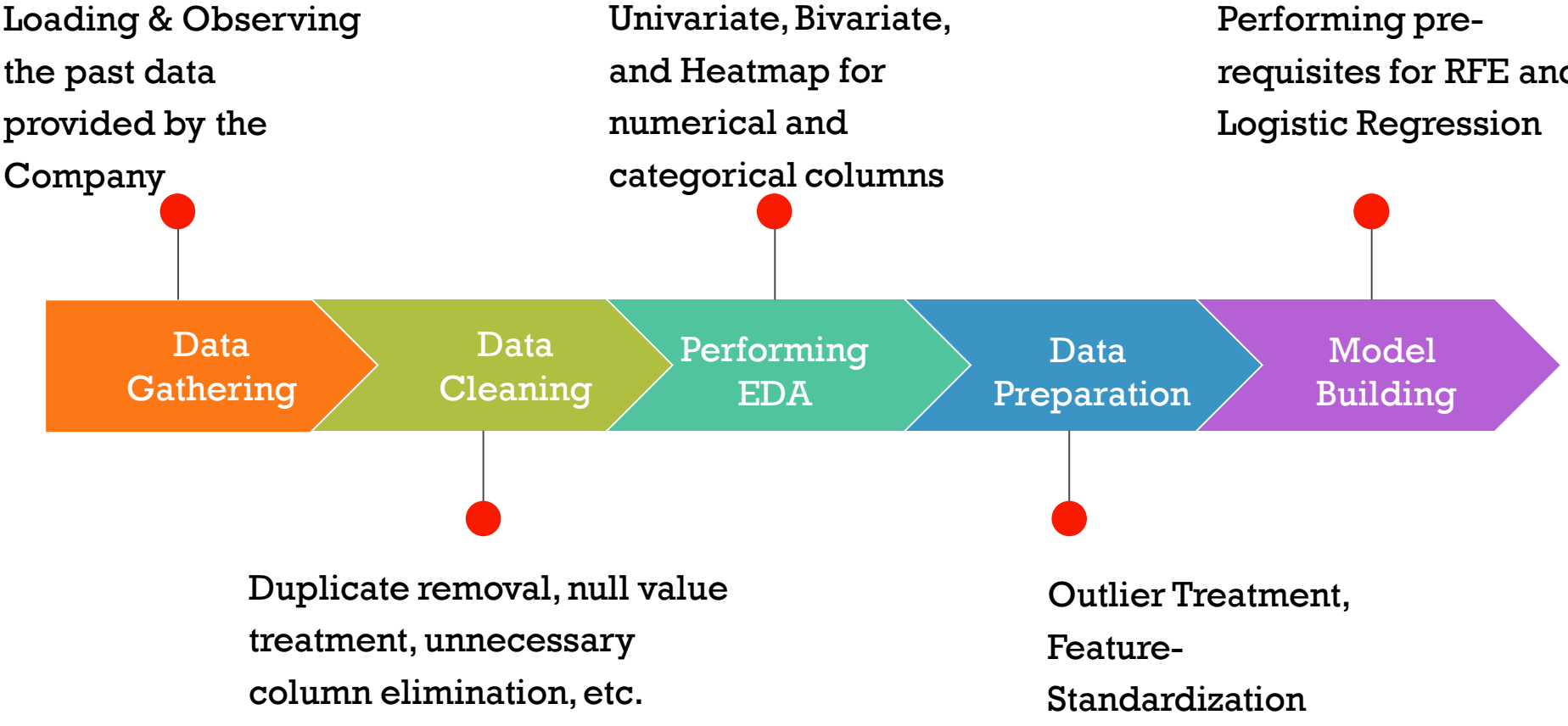
Selection of Hot Leads

For our Problem Solution, the crucial part is to accurately identify hot leads.

The more accurate we obtain the hot lead, the more chance we get of higher conversion ratio.

Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

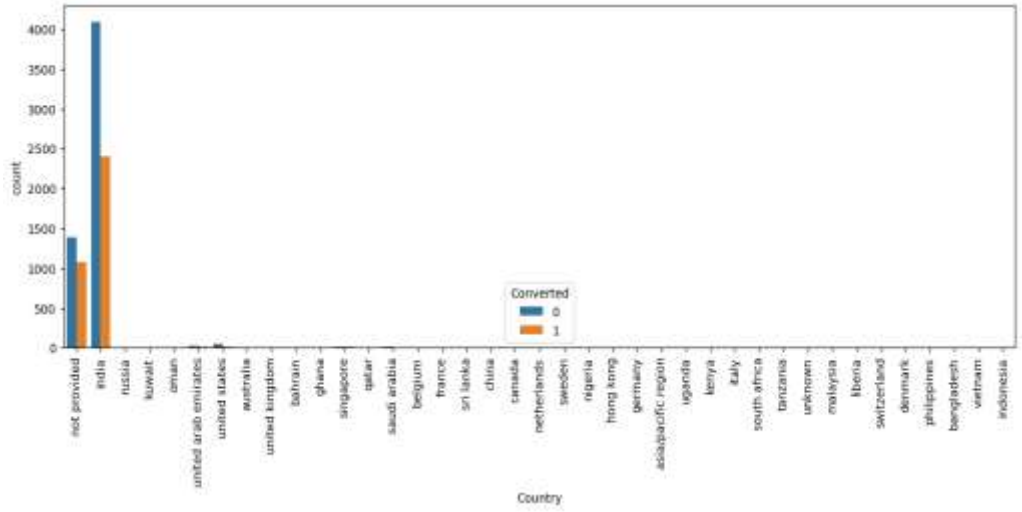
Implementation

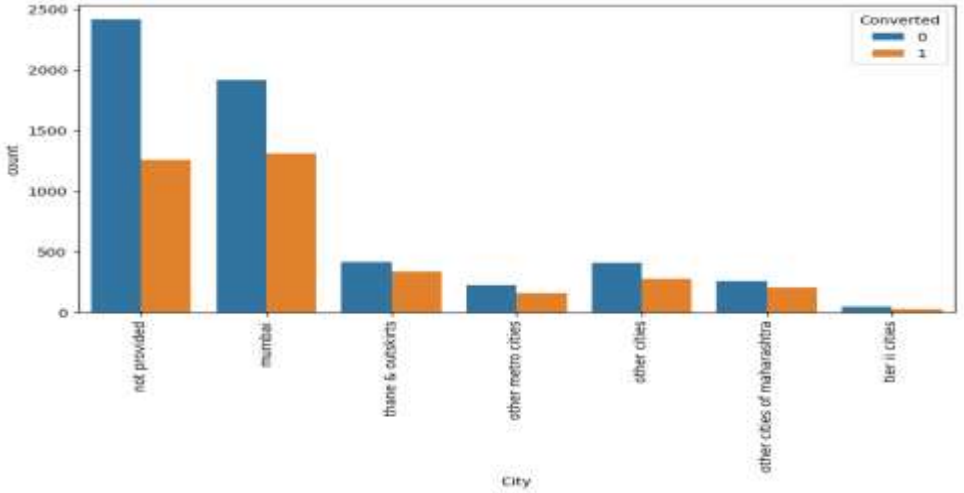

Loading & Observing the past data provided by the Company
Univariate, Bivariate, and Heatmap for numerical and categorical columns
Performing pre-requisites for RFE and Logistic Regression
Data Gathering
Data Cleaning
Performing EDA
Data Preparation
Model Building
Duplicate removal, null value treatment, unnecessary column elimination, etc.
Outlier Treatment, Feature-Standardization

Visualization

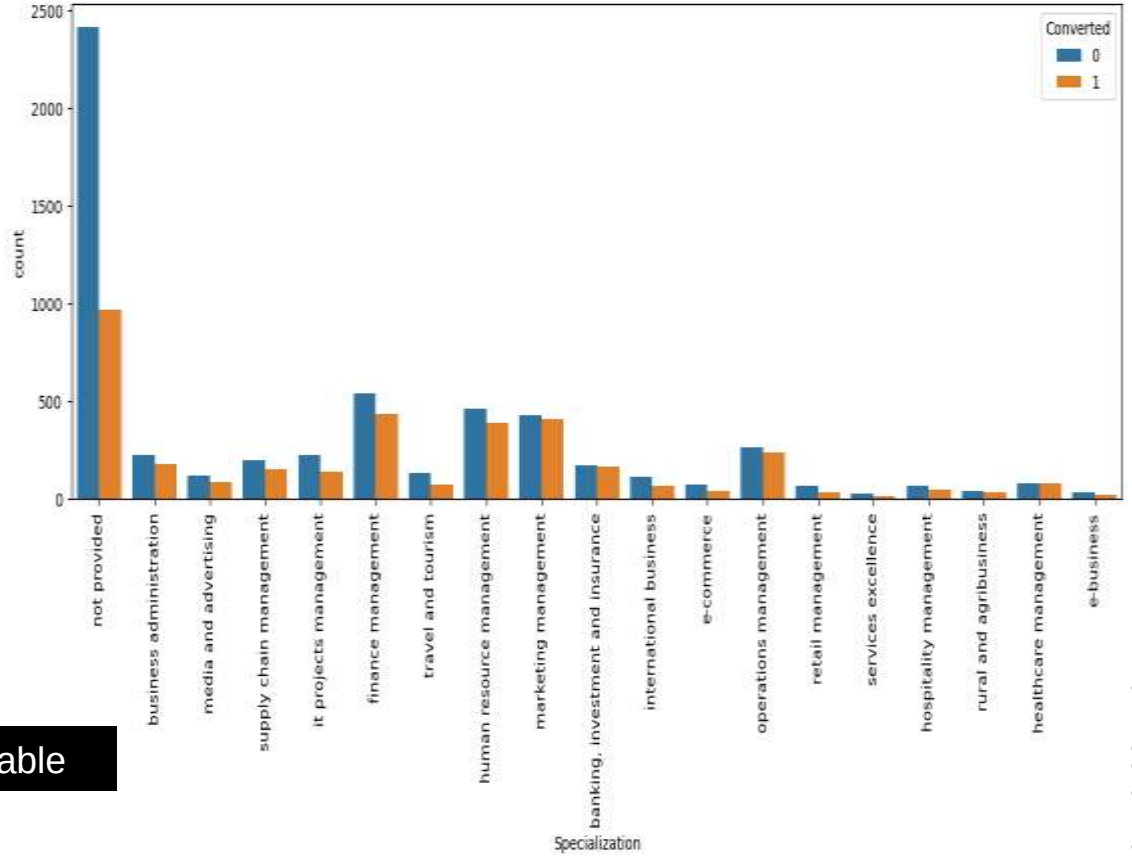
PLOTS

EDA plots depicting the highest number of values by a country and their conversion.





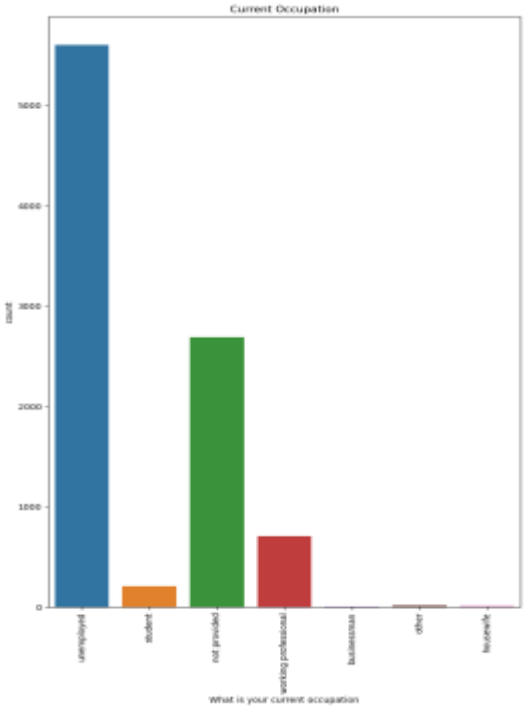
EDA plotting spread of Specialization column.

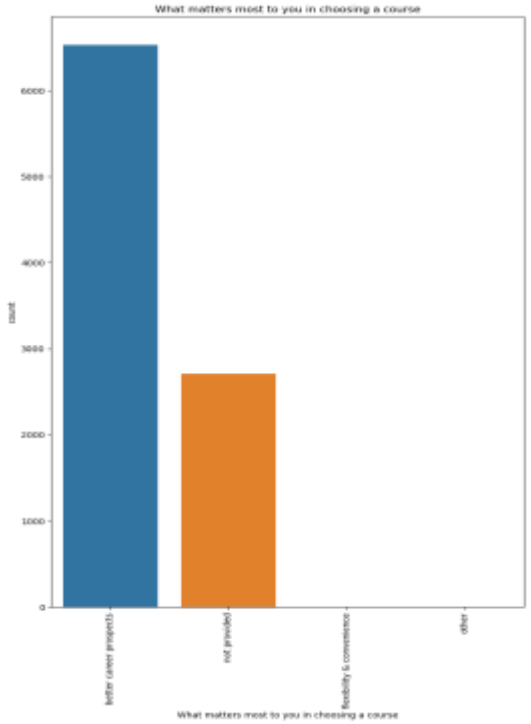


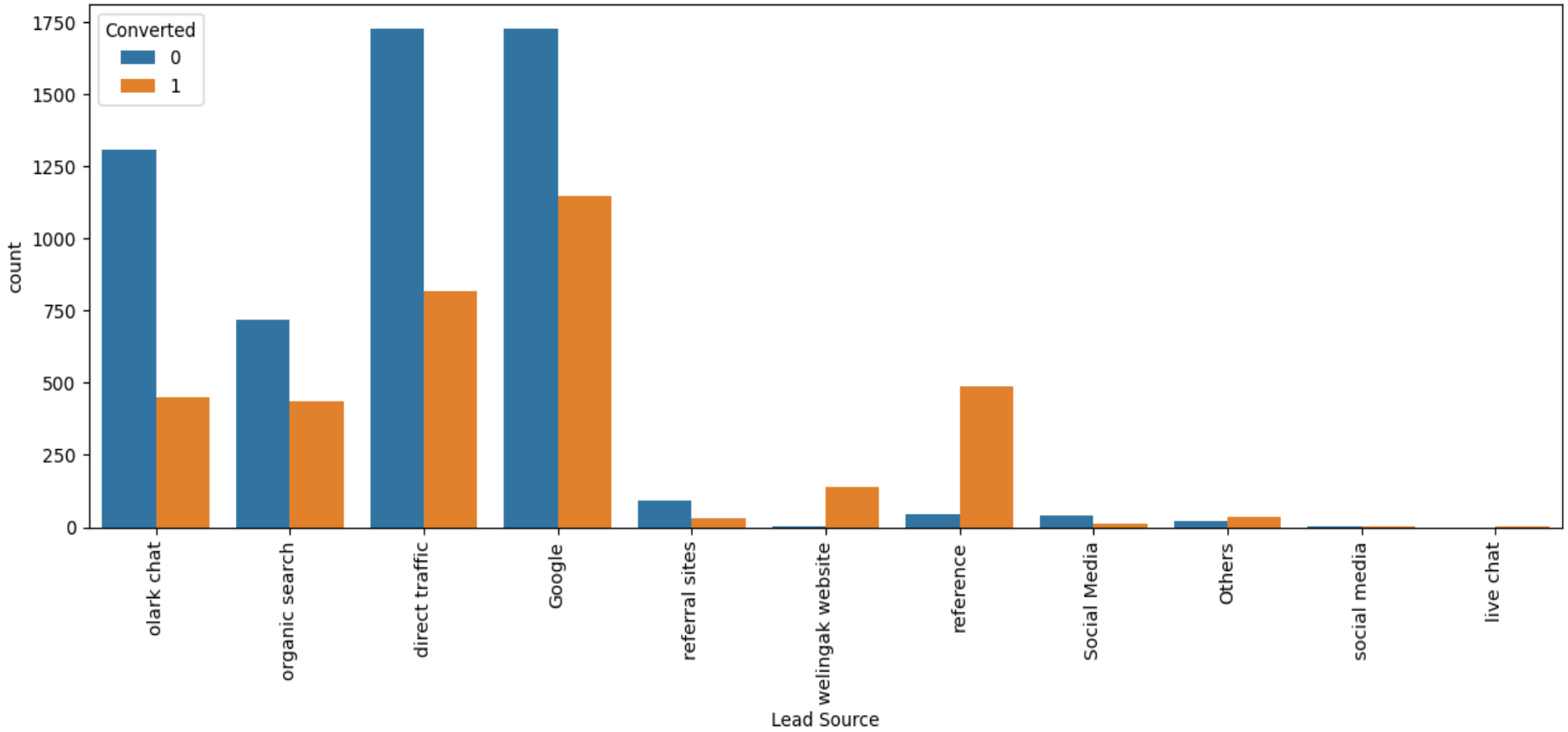
Management a significant variable

EDA plots depicting

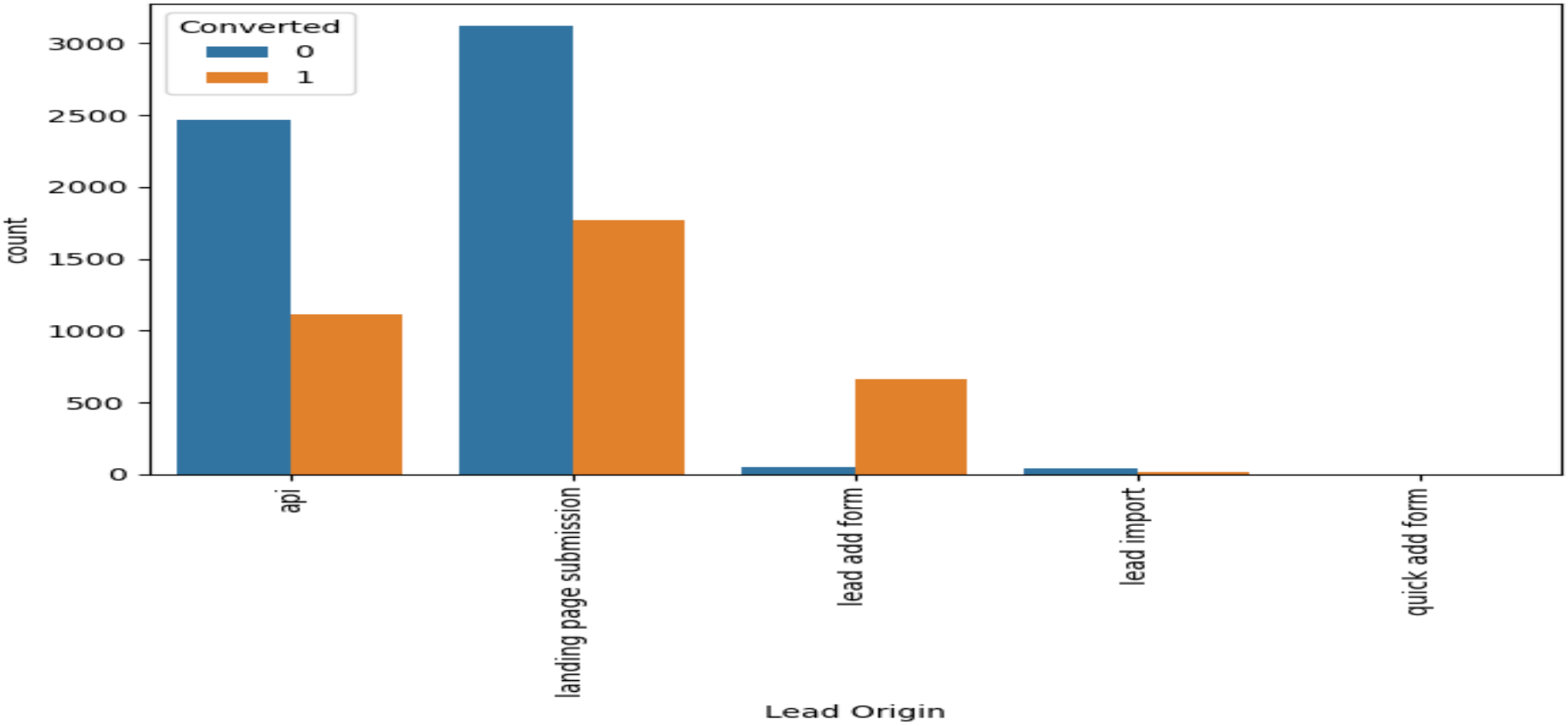
Working Professionals going for the course have high chances of joining it.
Unemployed leads are the most in terms of Absolute numbers.



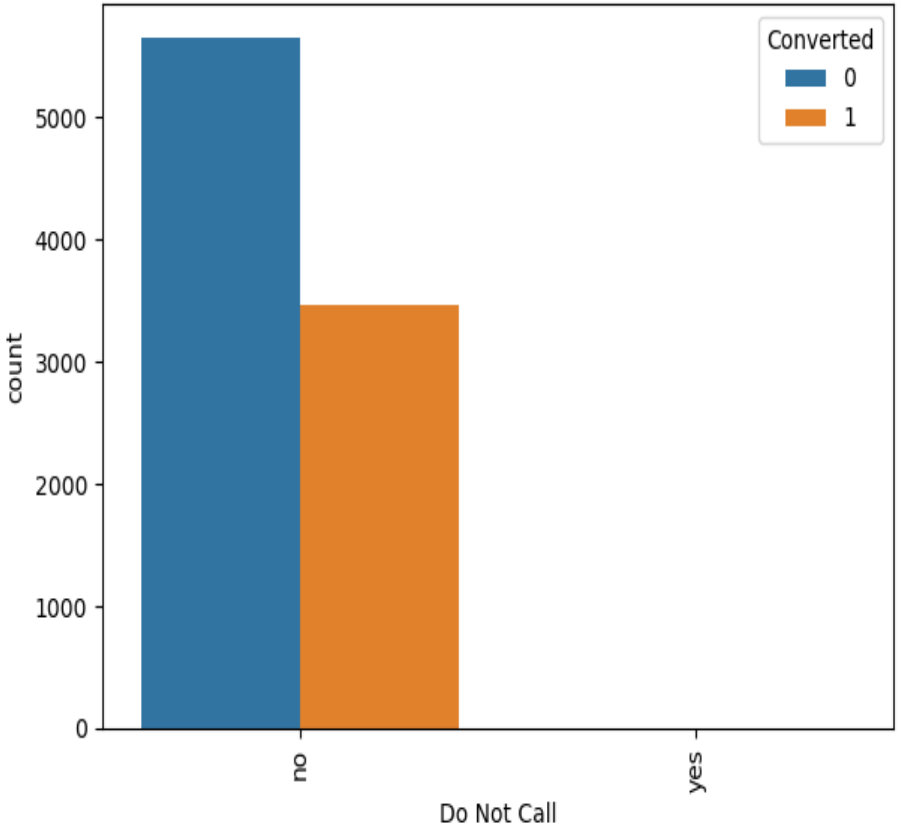


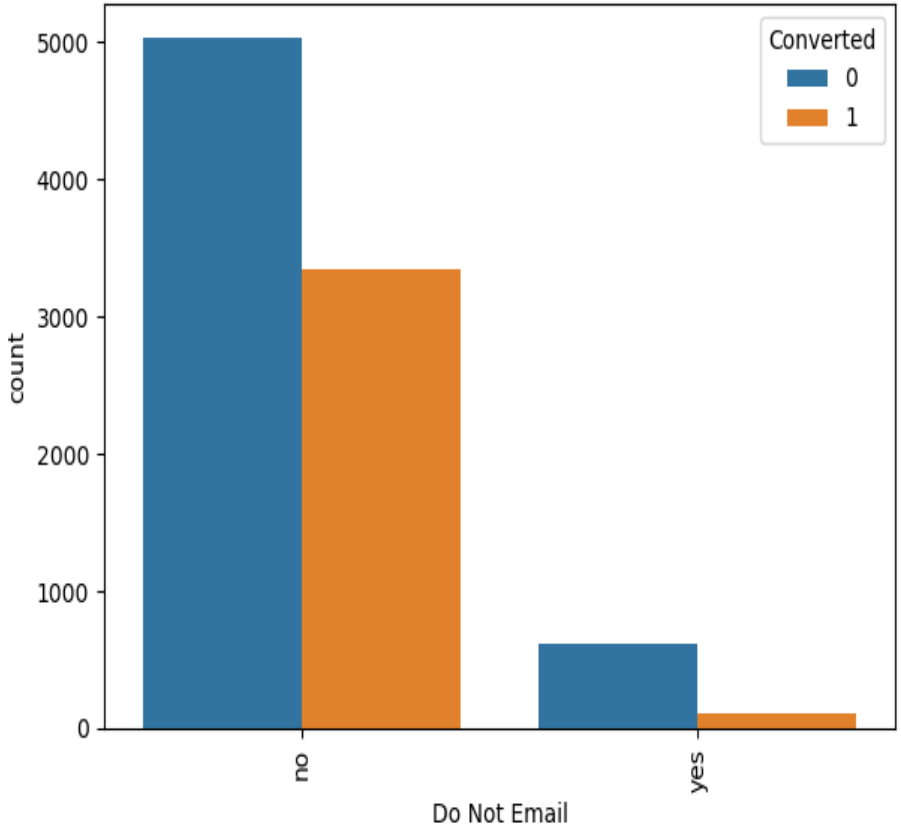


EDA plot depicting variation in categorical column (Lead Source) for those who Converted and those who didn't.

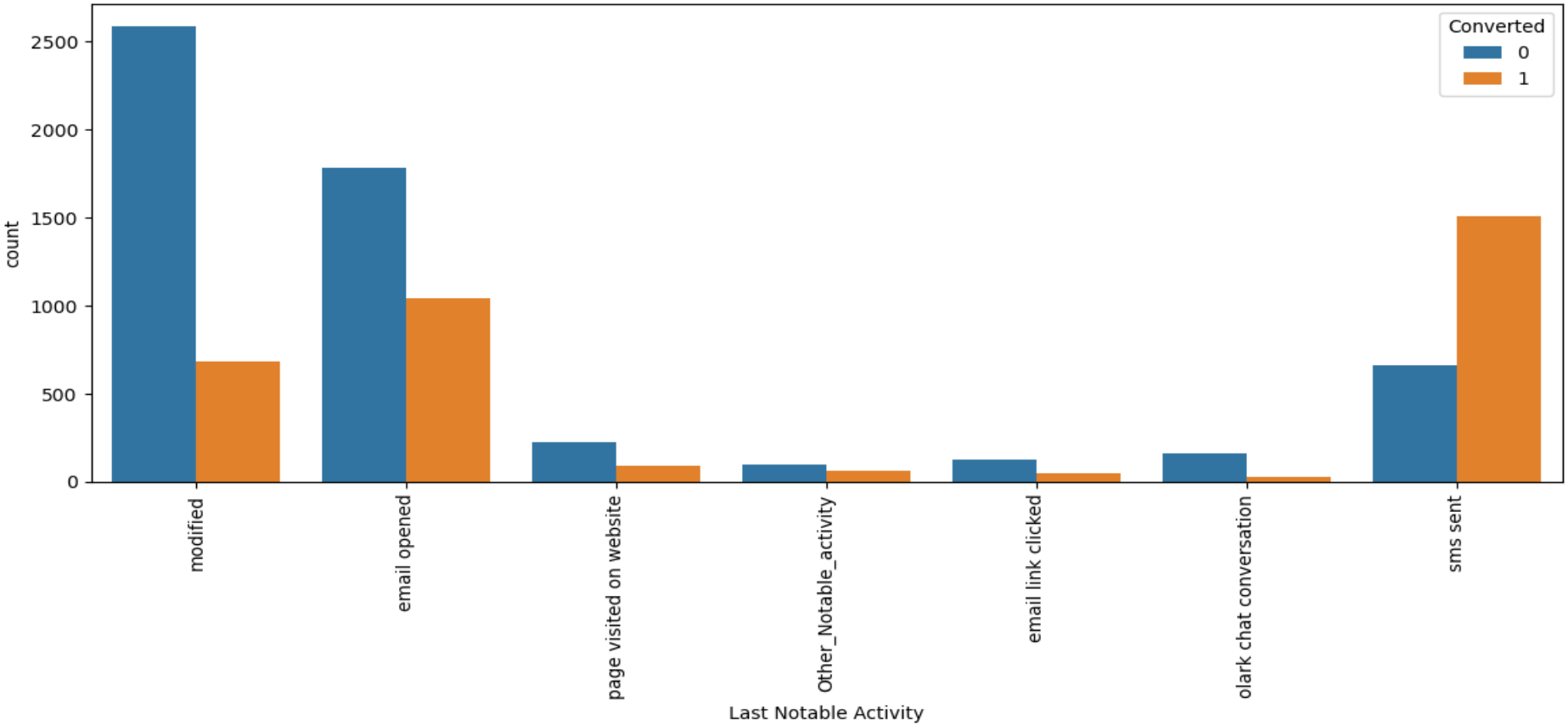


EDA plot depicting variation in categorical column (Lead Origin) for those who Converted and those who didn't.

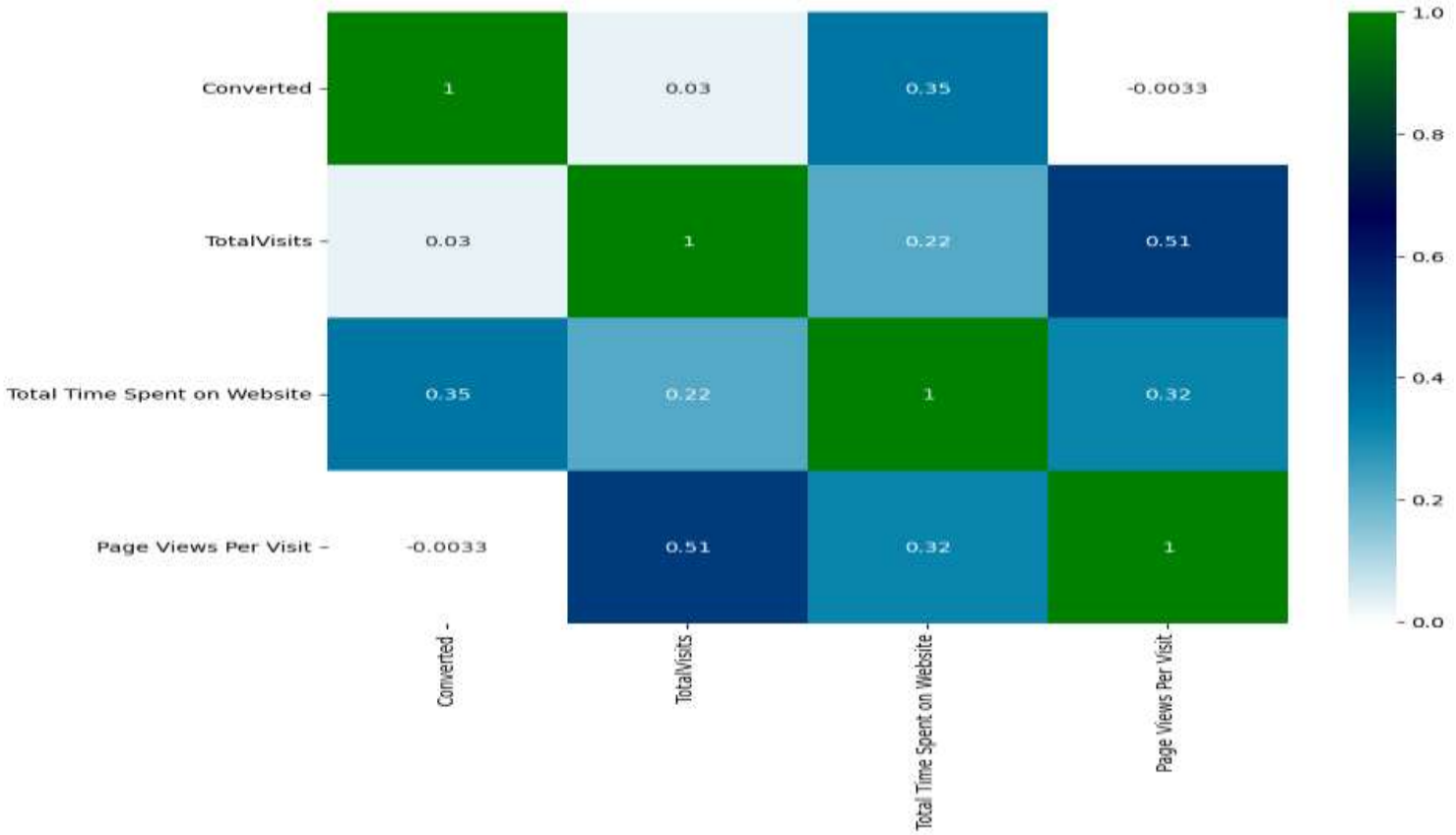




EDA plots depicting variation in categorical column (Do not Call) and (Do not Email)

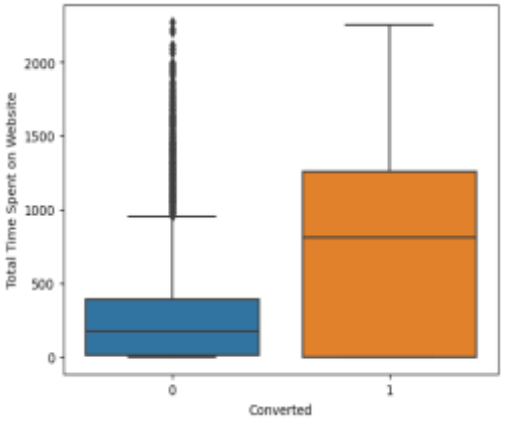


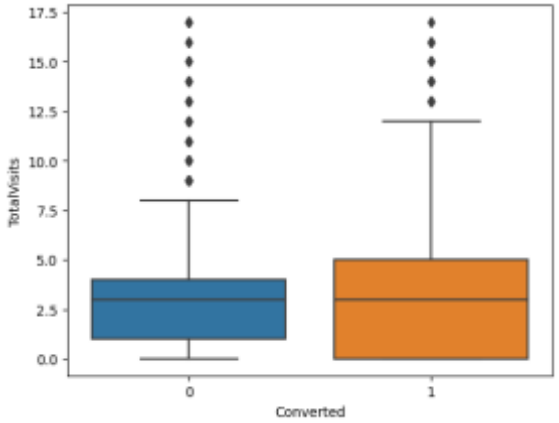
EDA plot depicting variation in categorical column (Last Notable Activity)

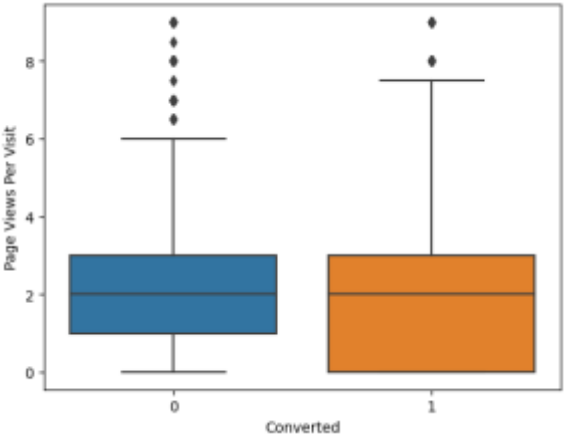


EDA plots depicting correlation (Heat Map) of all selected numerical columns.

EDA plots depicting variation in numerical columns .





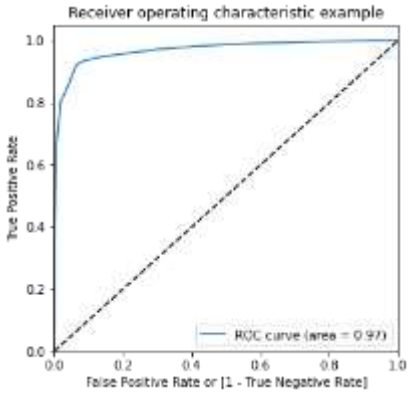


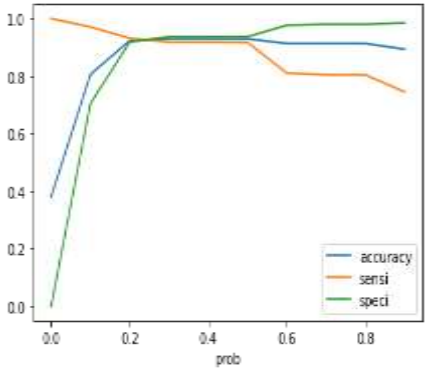
Linear Regression Final Model Parameters :

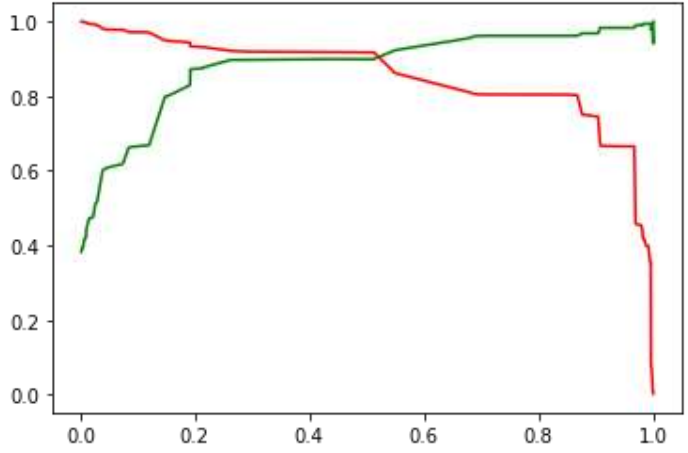
Area under ROC = 0.97

Intermediate cut-off = 0.4

Final cut-off = 0.48







Conclusion

Model Analysis

Performance of our Final Model

Overall accuracy on Test set: 0.94

Sensitivity of our logistic regression model: 0.93

Specificity of our logistic regression model: 0.94

Inferences from Model

Business Insights Derived from our Model

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model.

Top variables in model, that contribute towards lead conversion are:

- Total Time Spent on Website
- Total number of Visits
- When the Lead Source was "Google",Direct" and "Organic".

RECOMMENDATION

X Education Company needs to improve on the following variables to improve the overall conversion rate:

- Increase sending SMS notifications.
- Improve Olark Chat Service.
- Increase Total Visits by advertising etc.