CASE STUDY ON LEAD SCORING

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

Group Members:

- ✓ **Neethu Radhakrishnan**
- ✓ **Ramalingeswara Rao**

BACKGROUND

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

BACKGROUND : 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 Company's Problem

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 be now.

Focusing more on communicating with the potential leads rather than making calls to everyone

PROBLEM STATEMENT : X Education Company's Problem

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%.

PROPOSED SOLUTION :

Selection of Hot Leads → Communicating with Hot Leads → Conversion of Hot Leads

Leads Clustering

We group the leads into several groups

Depending on their propensity or like - lihood to convert, resulting in a more narrowly focused group of hot leads.

Focus Communication

We might have a smaller pool of leads to communicate.

Which would allow us to have a greater impact.

Increase conversion

We would have a greater conversion rate and be able to hit the 80% objective.

We concentrated on hot leads that were more likely to convert.

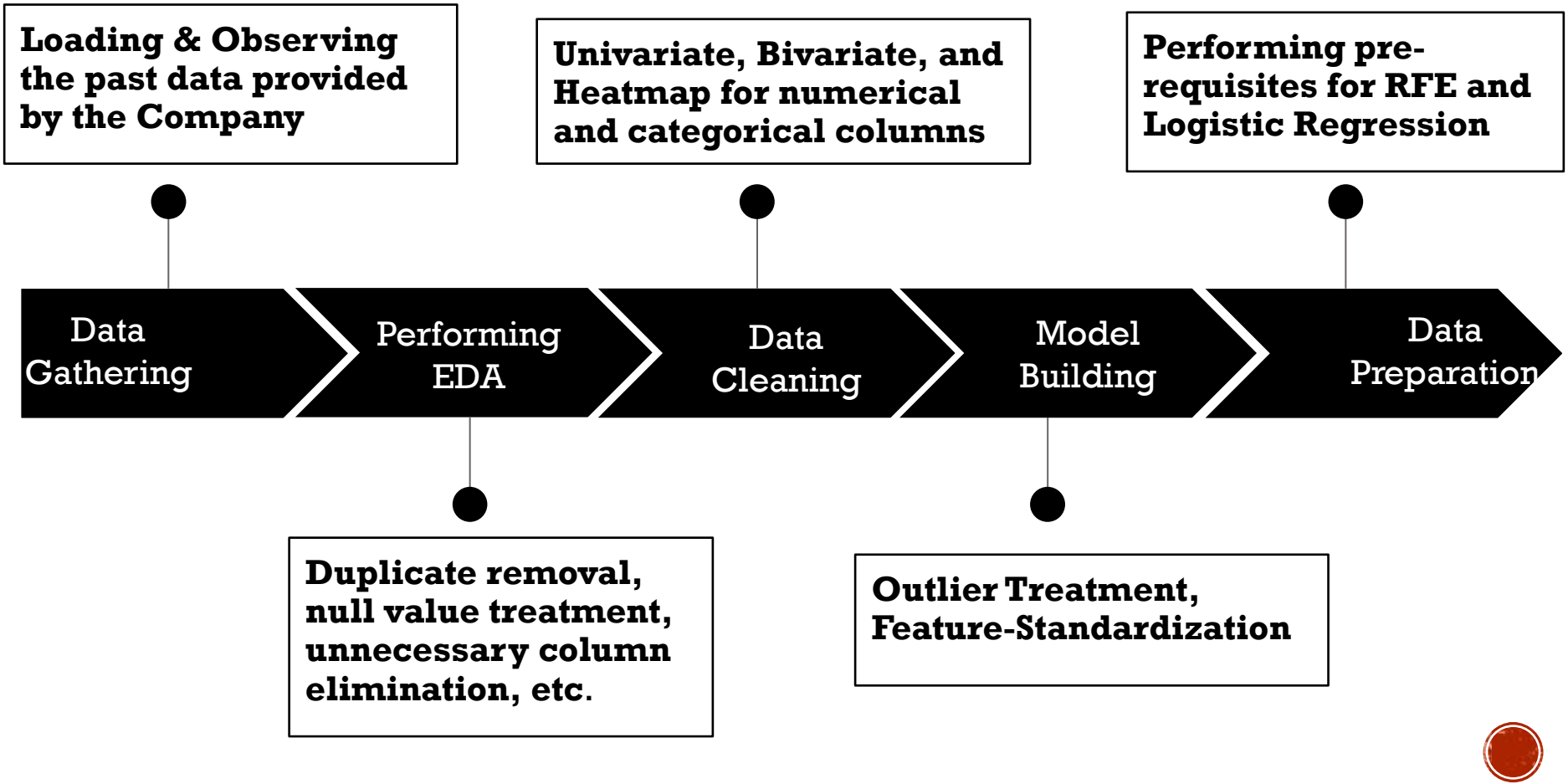
SOLUTION : Selection of Hot Leads

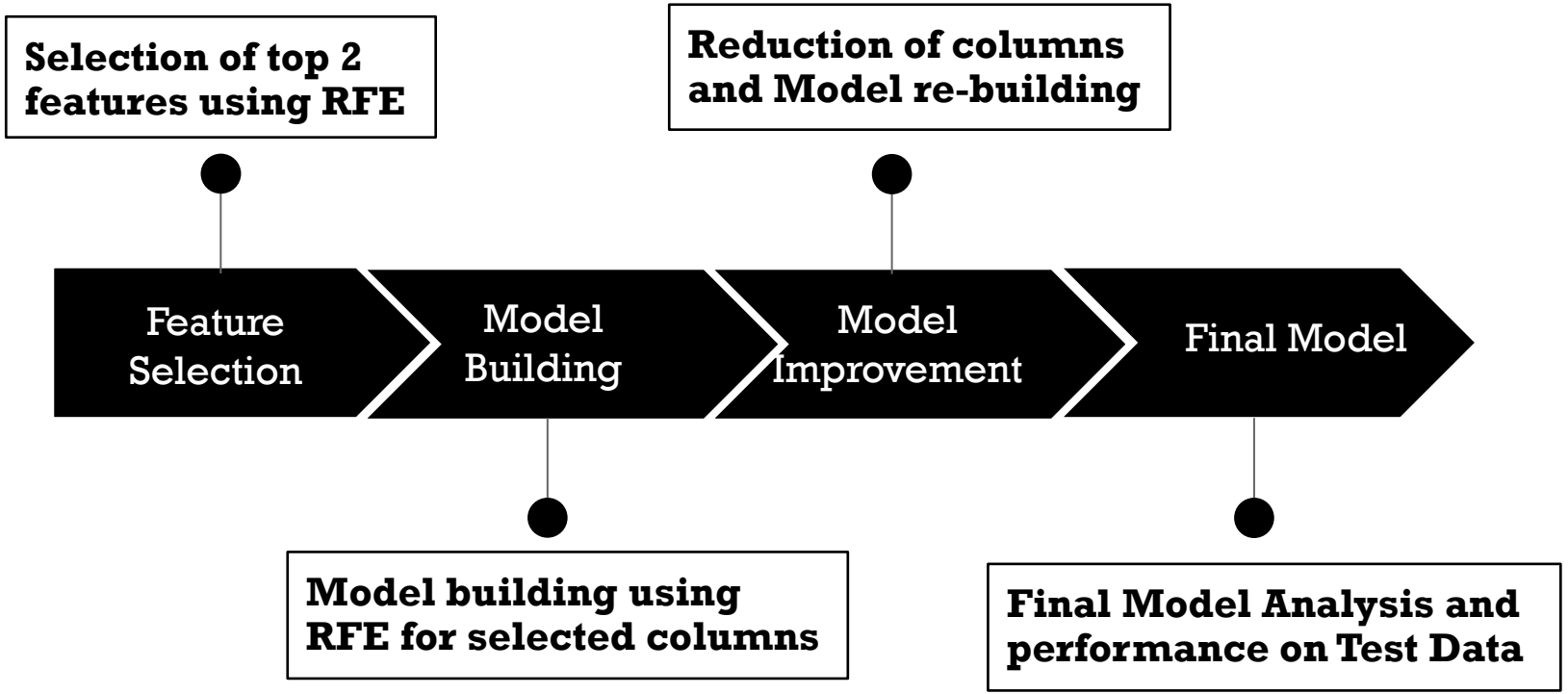
For our Problem Solution, the crucial part is to accurately identify hot leads.

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

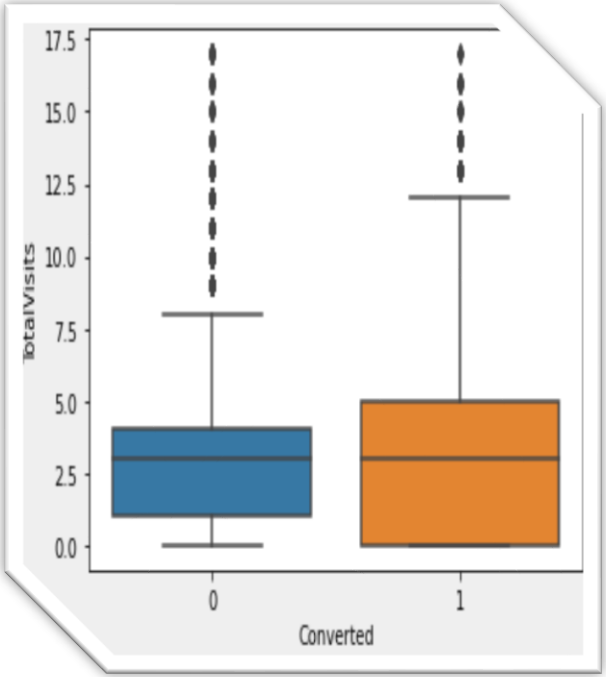
Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

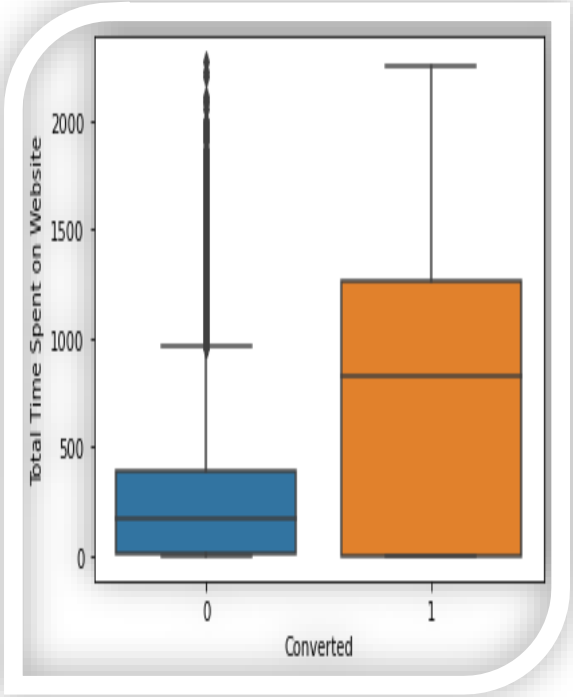
IMPLEMENTATION

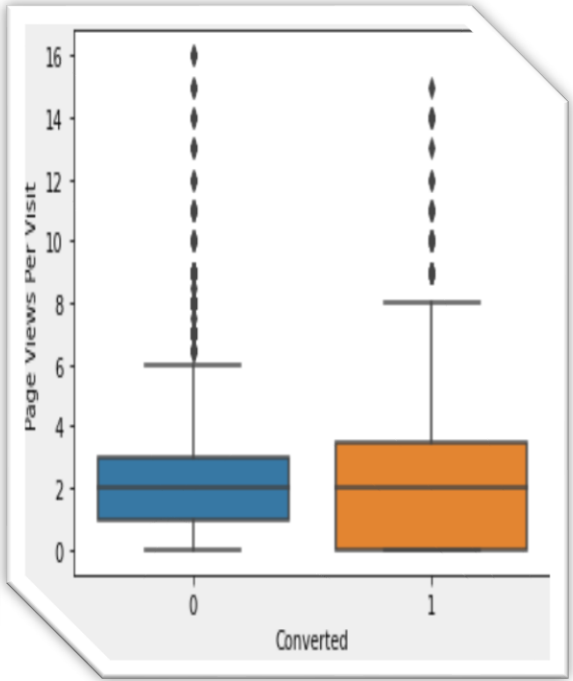

Loading & Observing the past data provided by the Company
Univariate, Bivariate, and Heatmap for numerical and categorical columns
Performing pre-requisites for RFE and Logistic Regression
Data Gathering
Performing EDA
Data Cleaning
Model Building
Data Preparation
Duplicate removal, null value treatment,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


Selection of top 2 features using RFE
Reduction of columns and Model re-building
Feature Selection
Model Building
Model Improvement
Final Model
Model building using RFE for selected columns
Final Model Analysis and performance on Test Data

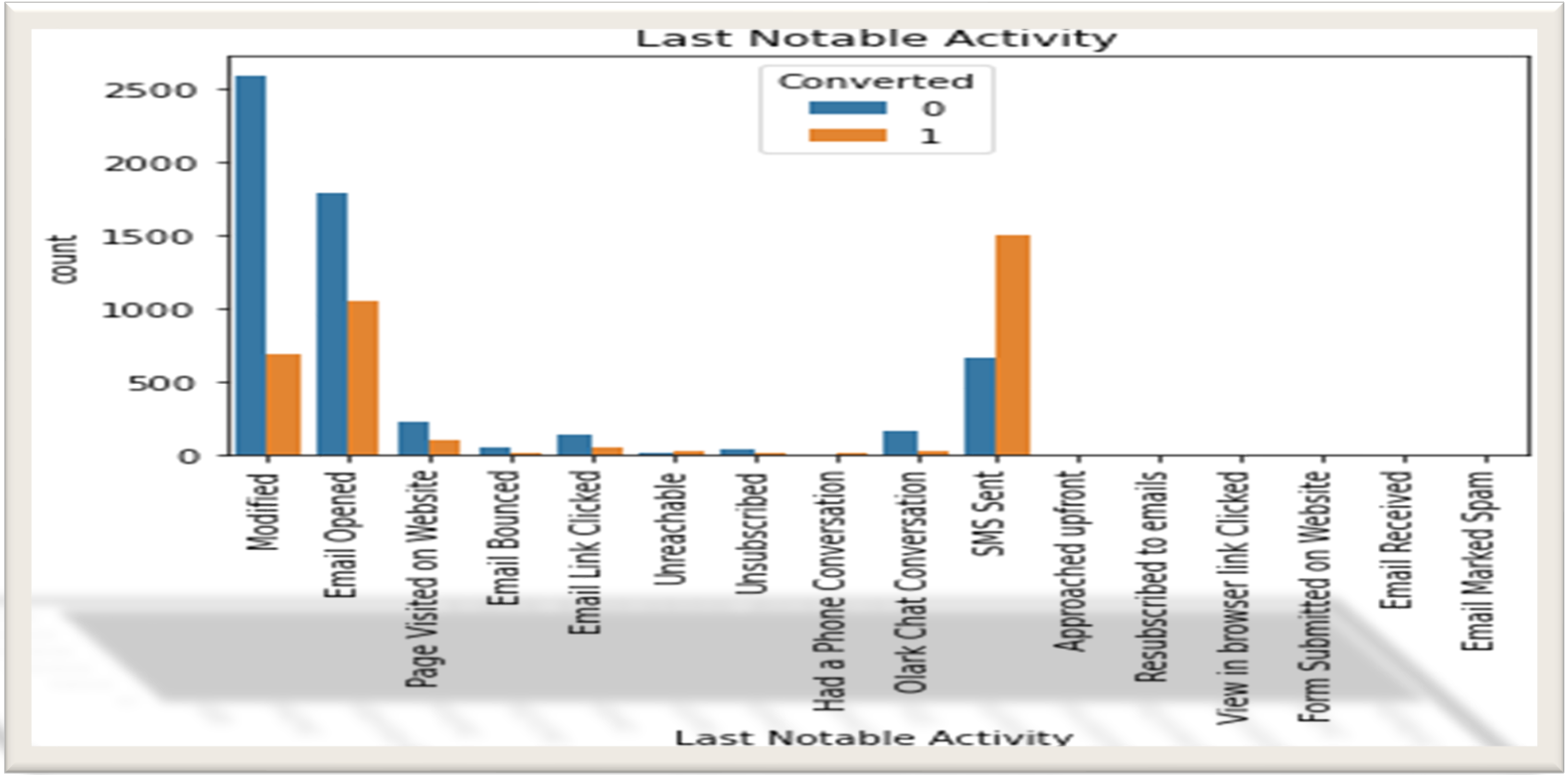
PLOTS (VISUALIZATION)



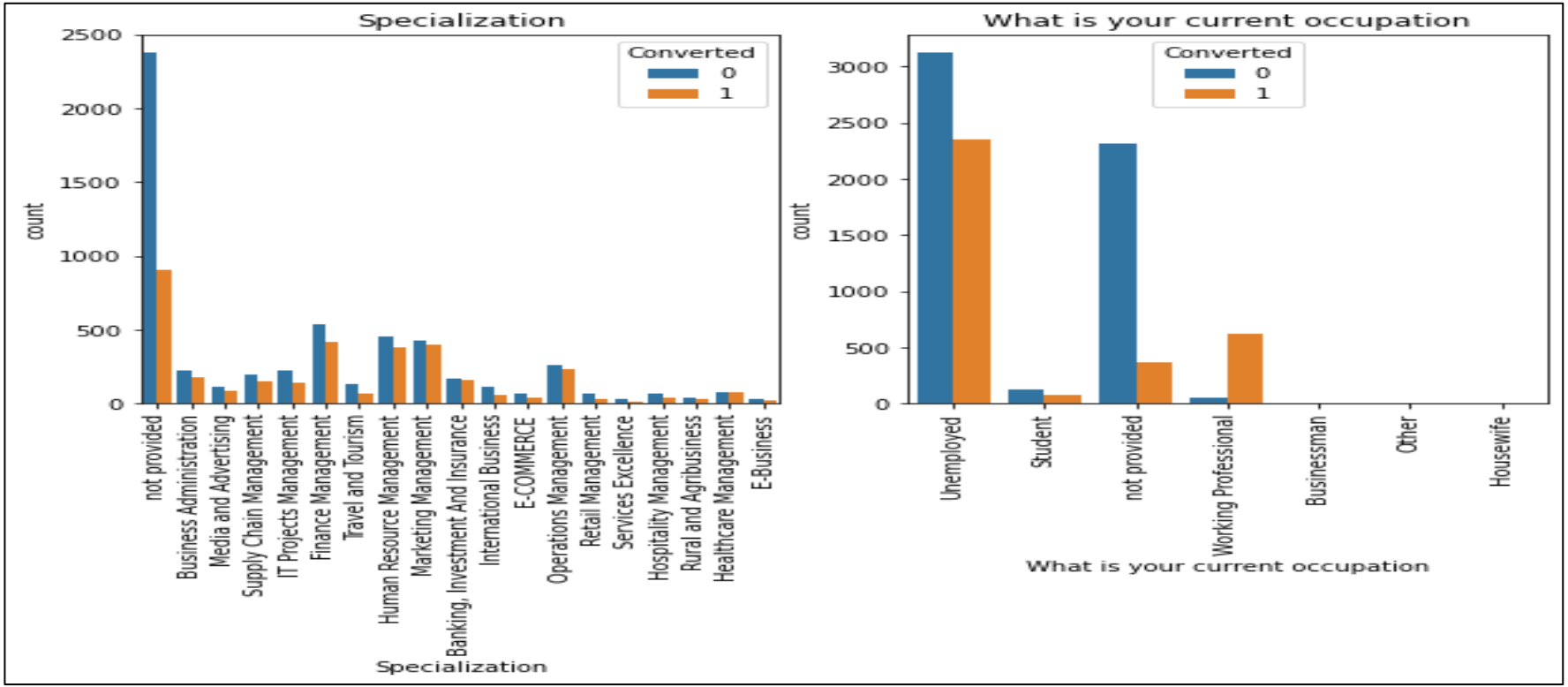




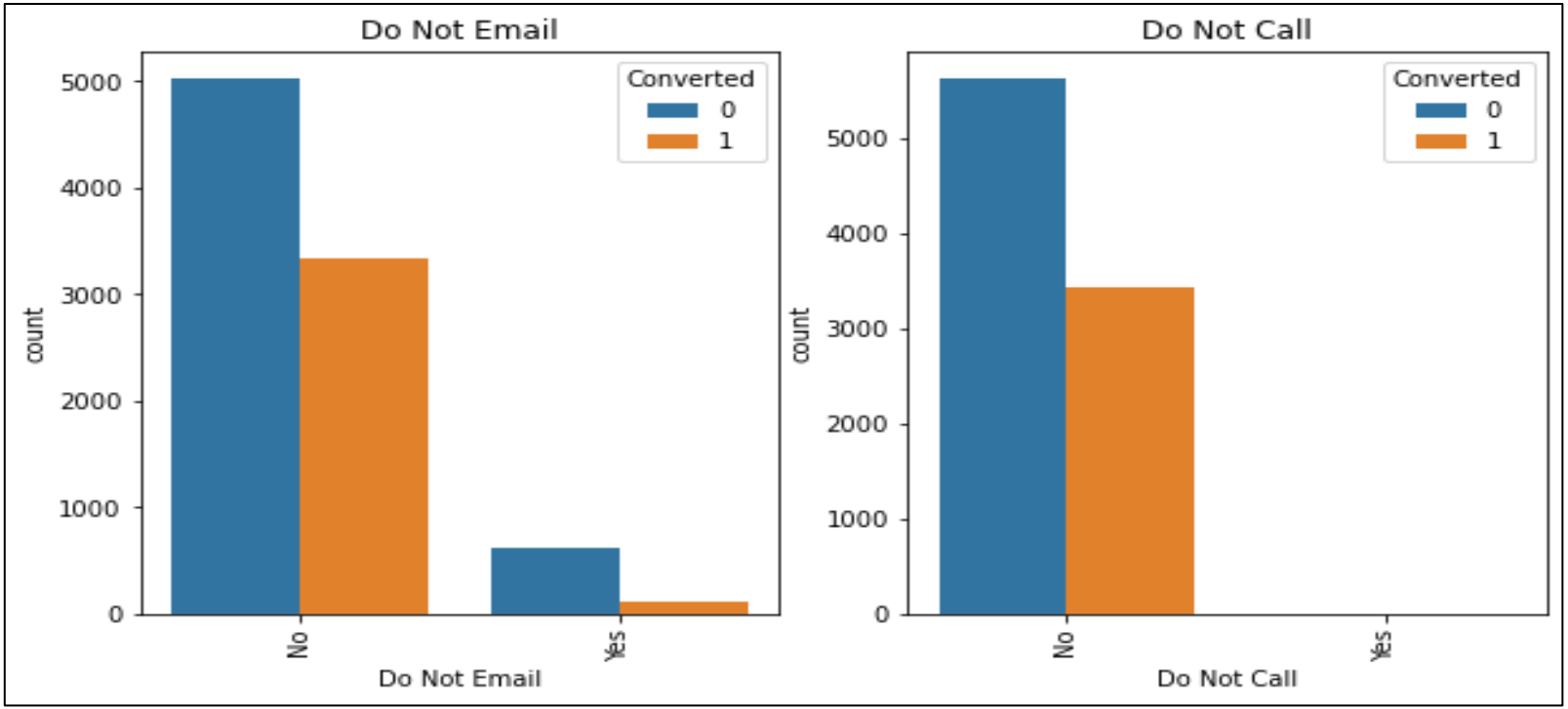
EDA plots depicting variation in numerical columns for those who Converted and those who didn't.



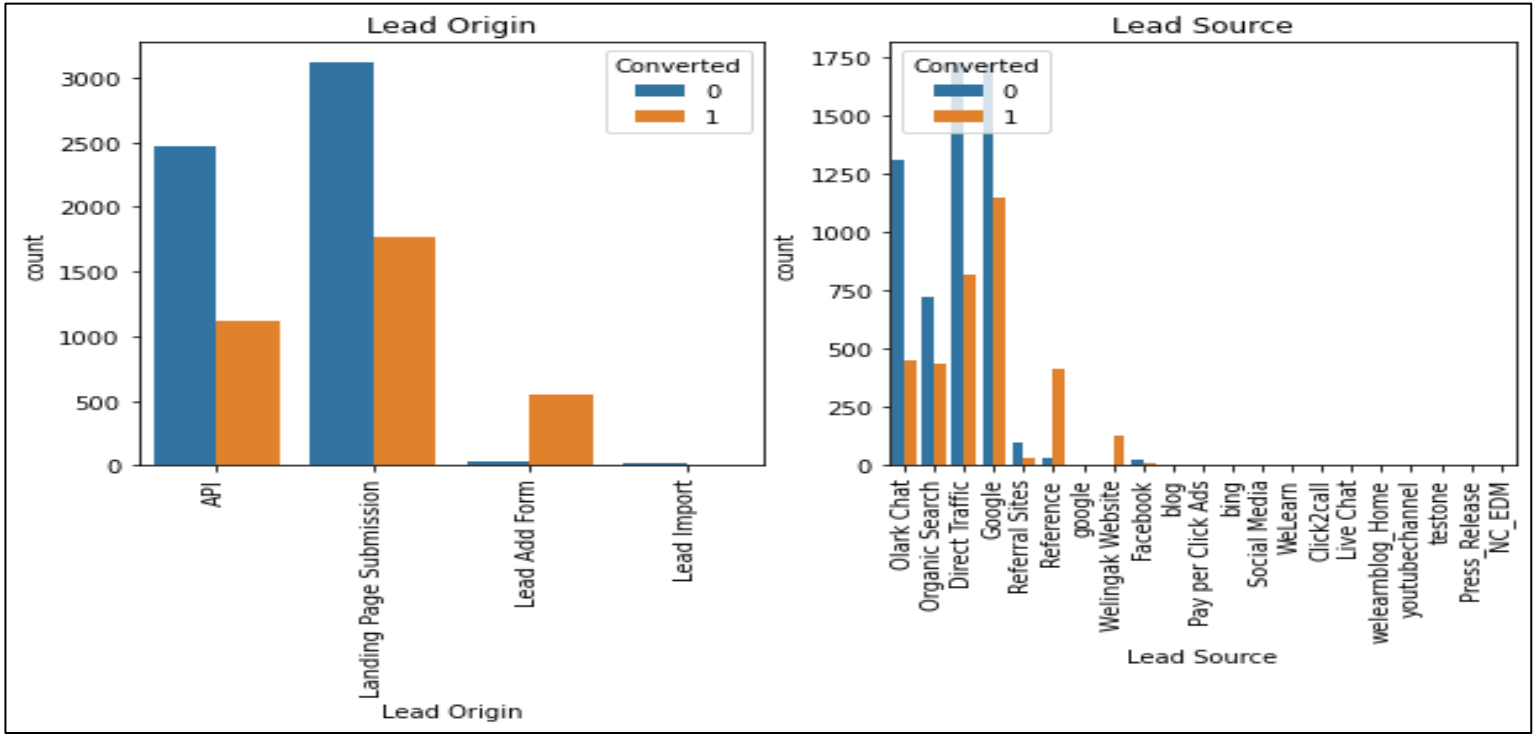
EDA plots depicting variation in categorical column (Last Notable Activity) for those who Converted and those who didn't.



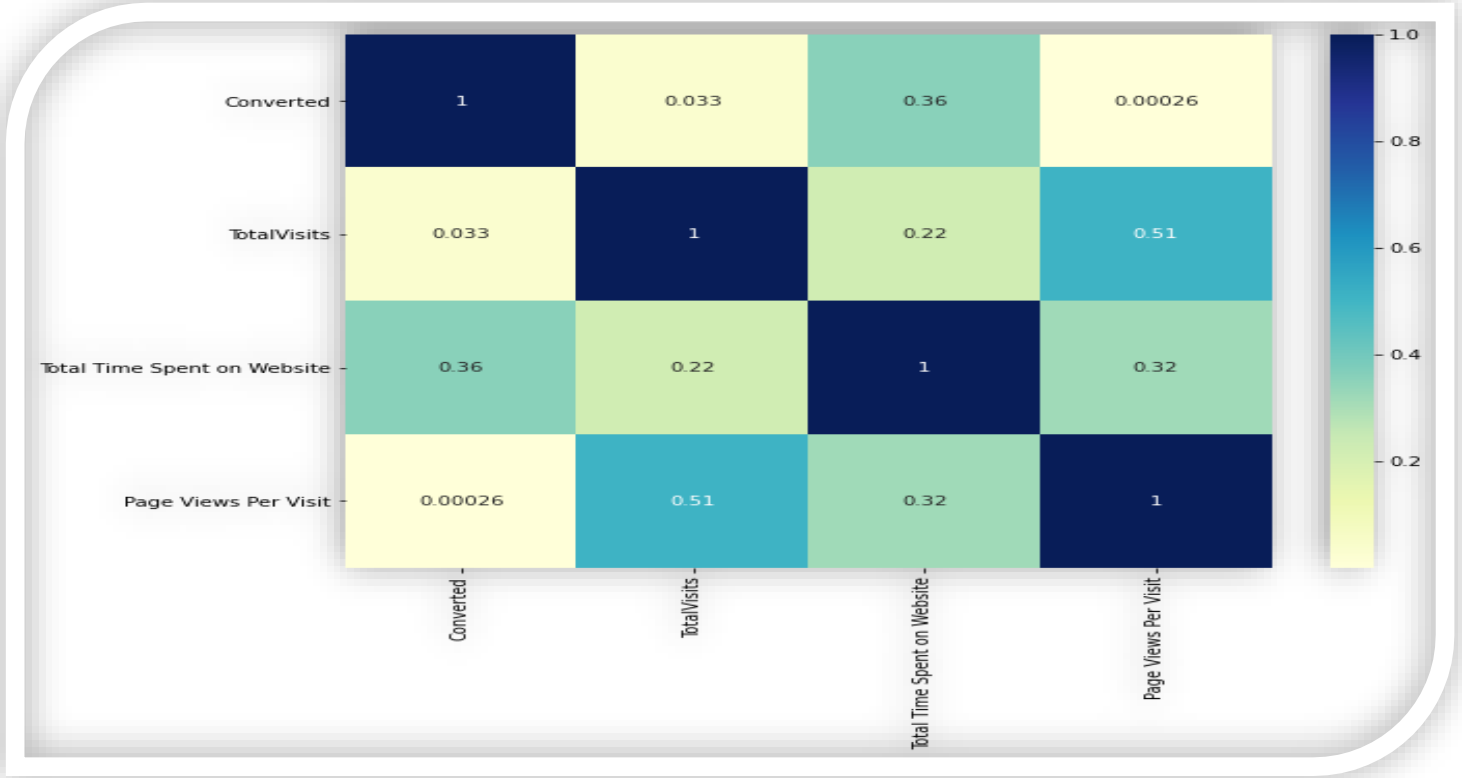
EDA plots depicting variation in categorical column (Specialization, What is your current occupation) for those who Converted and those who didn't.



EDA plots depicting variation in categorical column (Do Not Email, Do Not Call) for those who Converted and those who didn't.

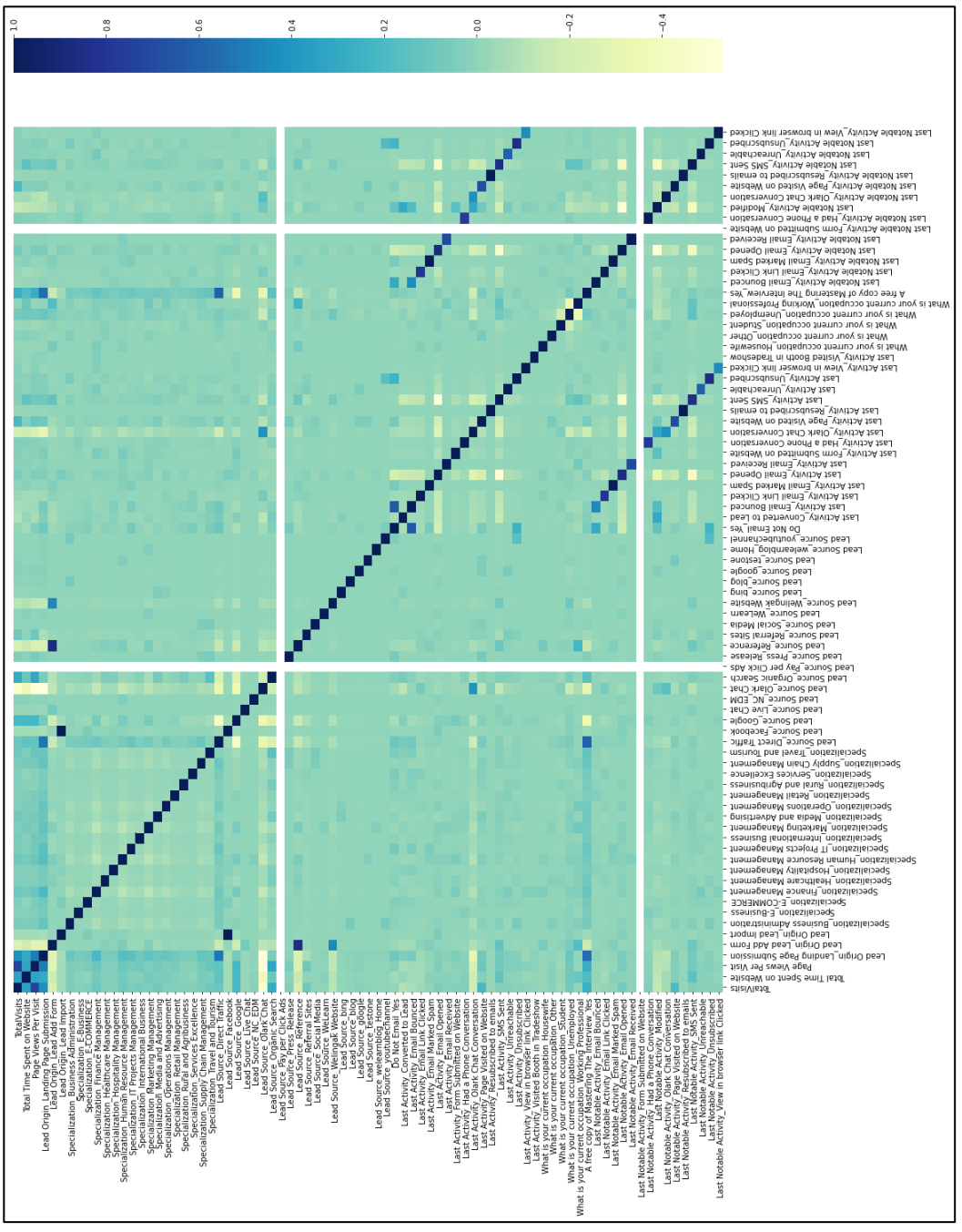


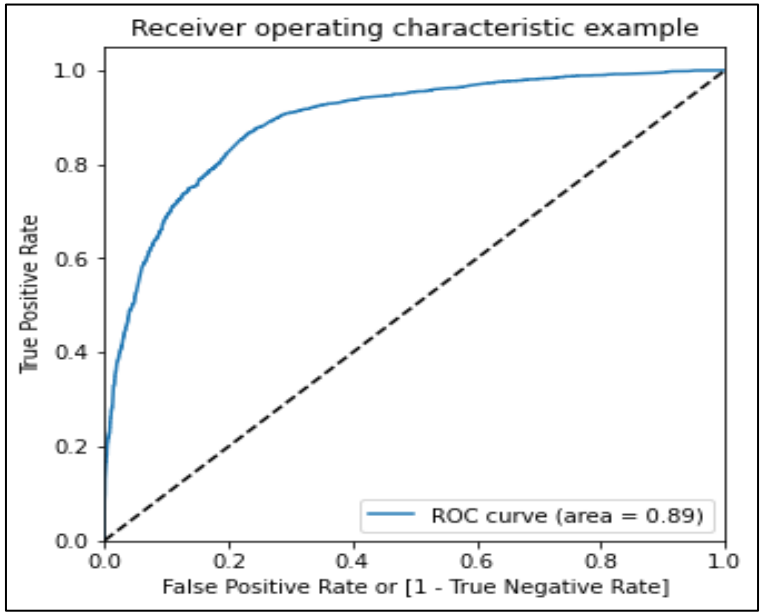
EDA plots depicting variation in categorical column (Lead Origin, Lead Source) for those who Converted and those who didn't.

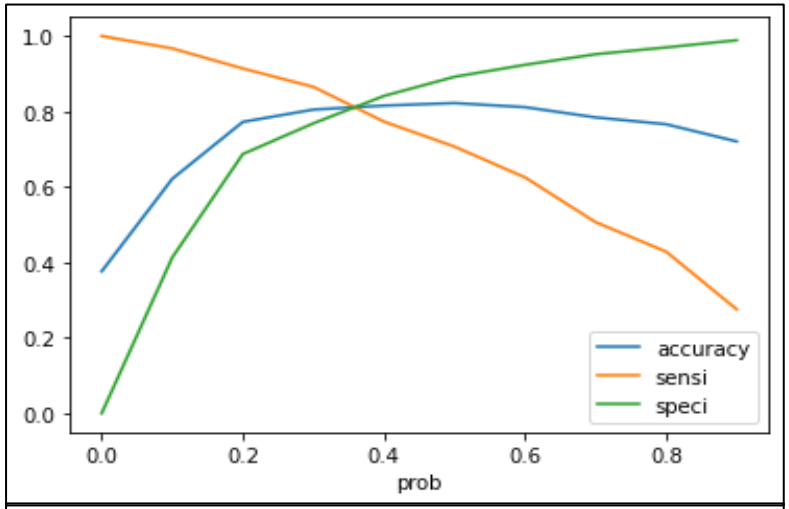


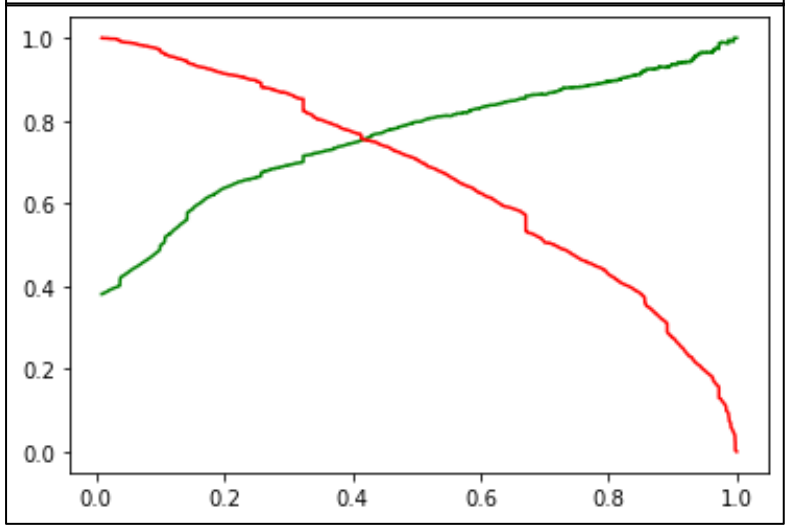
EDA plots depicting correlation (Heat Map) of all selected numerical columns.

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









Linear Regression Final Model Parameters
Area under ROC = 0.89
Intermediate cut-off = 0.35
Final cut-off = 0.41

INFERENCE/CONCLUSION

MODEL ANALYSIS : Performance of our Final Model

Overall accuracy on Test set: 80.4

Sensitivity of our logistic regression model: 80.4

Specificity of our logistic regression model: 80.6

INFERENCES FROM MODEL

Business Insights Derived from our Model

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

- TotalVisits
- Lead Source_Google
- Lead Source_Direct Traffic
- What is your current occupation_Unemployed
- Total Time Spent on Website
- Last Notable Activity_Modified
- Lead Source_Organic Search
- Last Notable Activity_Email Opened

INFERENCES FROM MODEL

Business Insights Derived from our Model

- Lead Origin_Lead Add Form
- Last Notable Activity_Olark Chat Conversation
- What is your current occupation_Working Professional
- Lead Source_Welingak Website
- Do Not Email_Yes
- Last Notable Activity_Page Visited on Website
- Lead Source_Referral Sites¶
- What is your current occupation_Student
- Last Notable Activity_Email Link Clicked

CONCLUSION-1(LR MODEL)

Our Logistic Regression Model is decent and accurate enough, with 80.4 Accuracy on Test Set, 80.4 % Sensitivity and 80.6 % Specificity.
We can vary these parameters by varying the cut-off value and thus predict Hot leads based on scenarios like availability of extra resources and vice-versa.

CONCLUSION 2 (RECOMMENDATION)

X Education Company needs to focus on following key aspects to improve the overall conversion rate:

- Increase user engagement on their website since this helps in higher conversion
- Increase on sending SMS notifications since this helps in higher conversion
- Get Total visits increased by advertising etc. since this helps in higher conversion
- Improve the Olark Chat service since this is affecting the conversion negatively