X Education - Lead Scoring Case Study

- Identifying Hot leads to increase conversion ratio for X Education by focusing more on them.

Prepared by: Kartik Mathpal

Course : DS&A (5022)
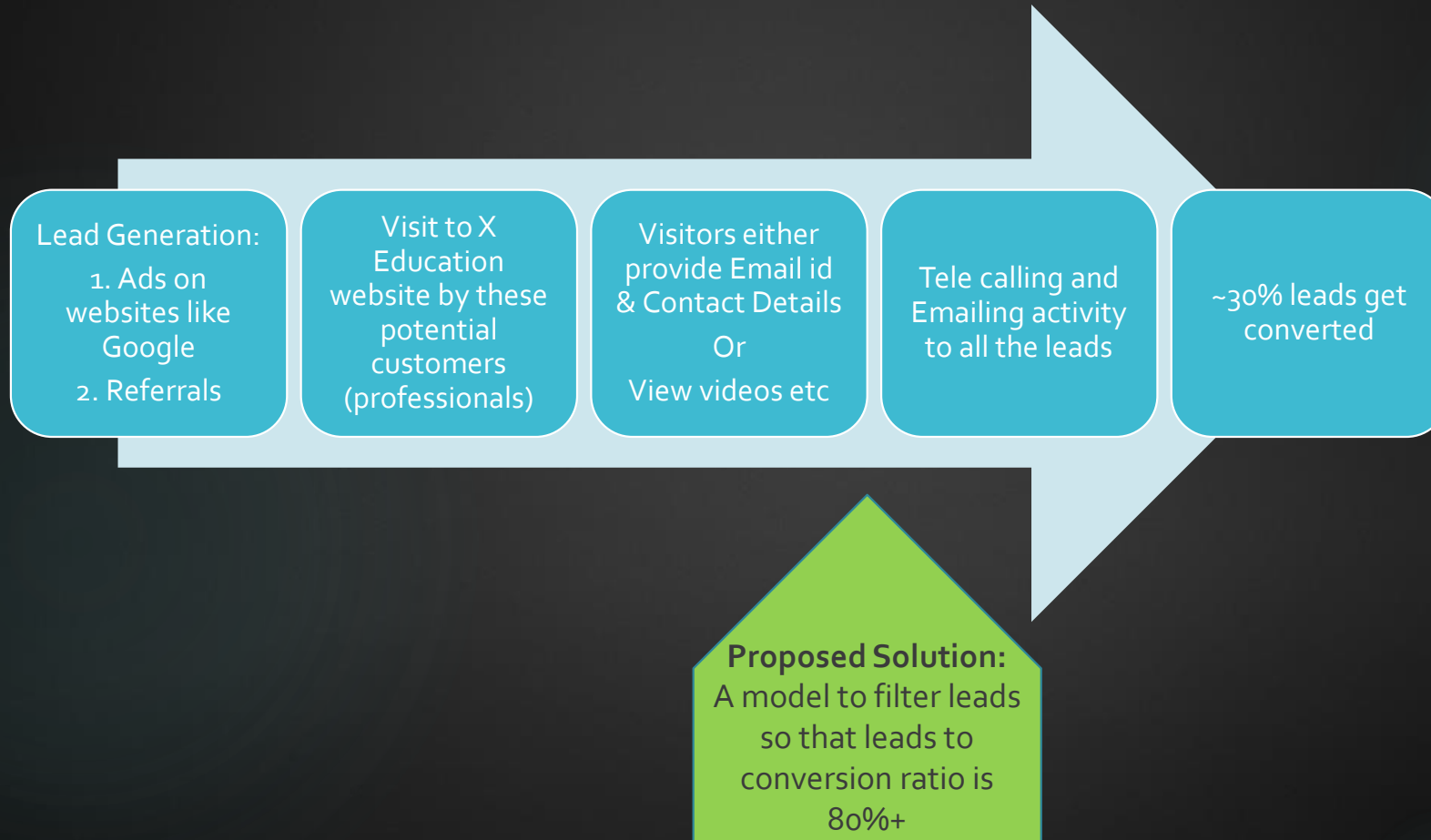
Upgrad campus

Background
- X Education Company

- X Education is an education company that sells online courses to people looking to advance in there career.

- The company markets there products on platforms like google, facebook, etc. People interested in these courses land on there website and browse through the courses, they might fill up some form or signup on their website.

- Once these people fill up the form and enter their emails and mobile numbers they are classified as leads.

- After the leads are acquired, members from the sales team start making calls and writing emails, etc. To persuade leads into purchasing the course.

- The leads that buys the course is said to be converted. The average lead conversion rate is 30%.

Problem Statement

- Although the company is getting a lot of leads, their lead conversion rate is very poor.

- The company wants to improve its lead conversion rate.

- For this they want to identify Hot leads (Potential leads with high chance of being converted), if they are able to correctly identify them then the sales team will focus more on the hot leads and conversion rate will go up.

- We have to help the company to correctly identify these leads.

- For this we have to build a model to accurately identify hot leads.

- The CEO has given us target lead conversion rate of 30%.

Lead conversion process

Lead to Conversion process

Lead Generation:
1. Ads on websites like Google
2. Referrals

Visit to X Education website by these potential customers (professionals)

Visitors either provide Email id & Contact Details
Or
View videos etc

Tele calling and Emailing activity to all the leads

~30% leads get converted

Proposed Solution:
A model to filter leads so that leads to conversion ratio is 80%+

Proposed Solution

Selection of Hot Leads

Communicating with Hot Leads

Conversion of Hot Leads

Leads Clustering

We cluster the leads into certain categories based on their tendency or probability to convert, thus, getting a smaller section of hot leads to focus more on.

Focus Communication

Since we would have a smaller set of leads to have communication with, we might make more impact with effective communication.

Increase conversion

Since we focused on hot leads, which were more probable to convert, we would have a better conversion rate, and hence we can achieve the 80% target.

Solution

- For our problem statement we need to accurately identify hot leads.

- For this we need to build a model with high accuracy in predicting hot leads.

Implementation

Loading & Observing the past data provided by the Company

Univariate, Bivariate, and Heatmap for numerical and categorical columns
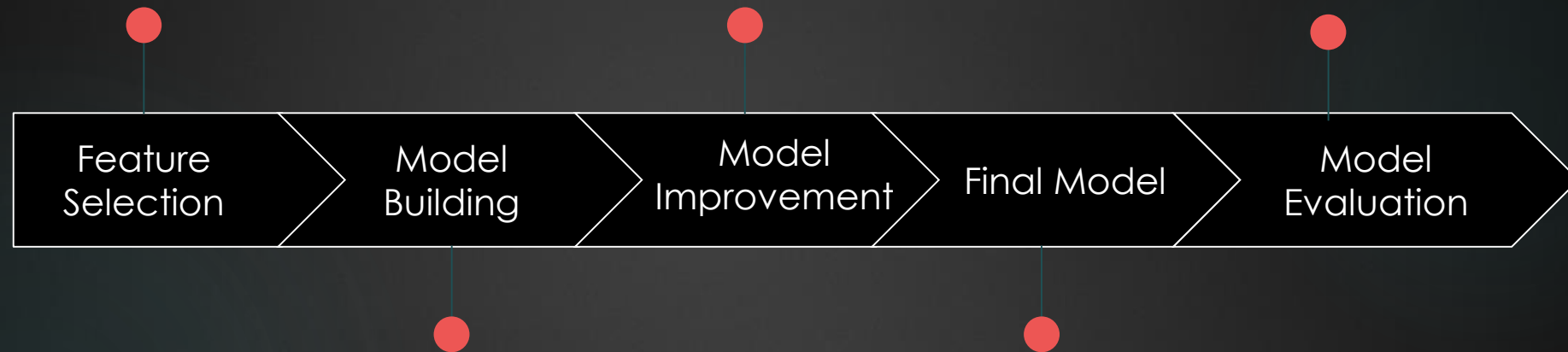
Performing pre-requisites for RFE and Logistic Regression

Data Gathering	Data Cleaning	Performing EDA	Data Preparation	Model Building
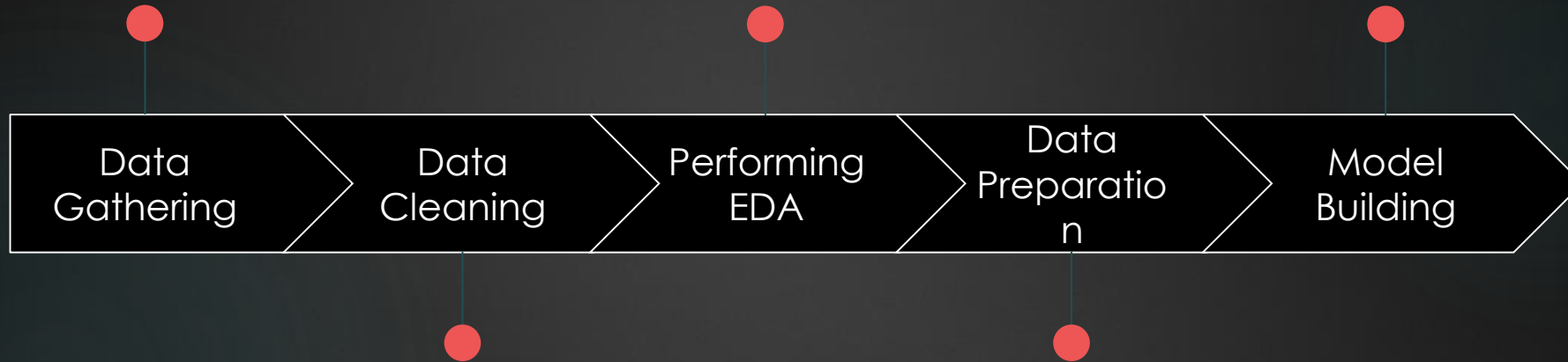
Duplicate removal, null value treatment, unnecessary column elimination, etc.

Outlier Treatment, Feature-Standardization
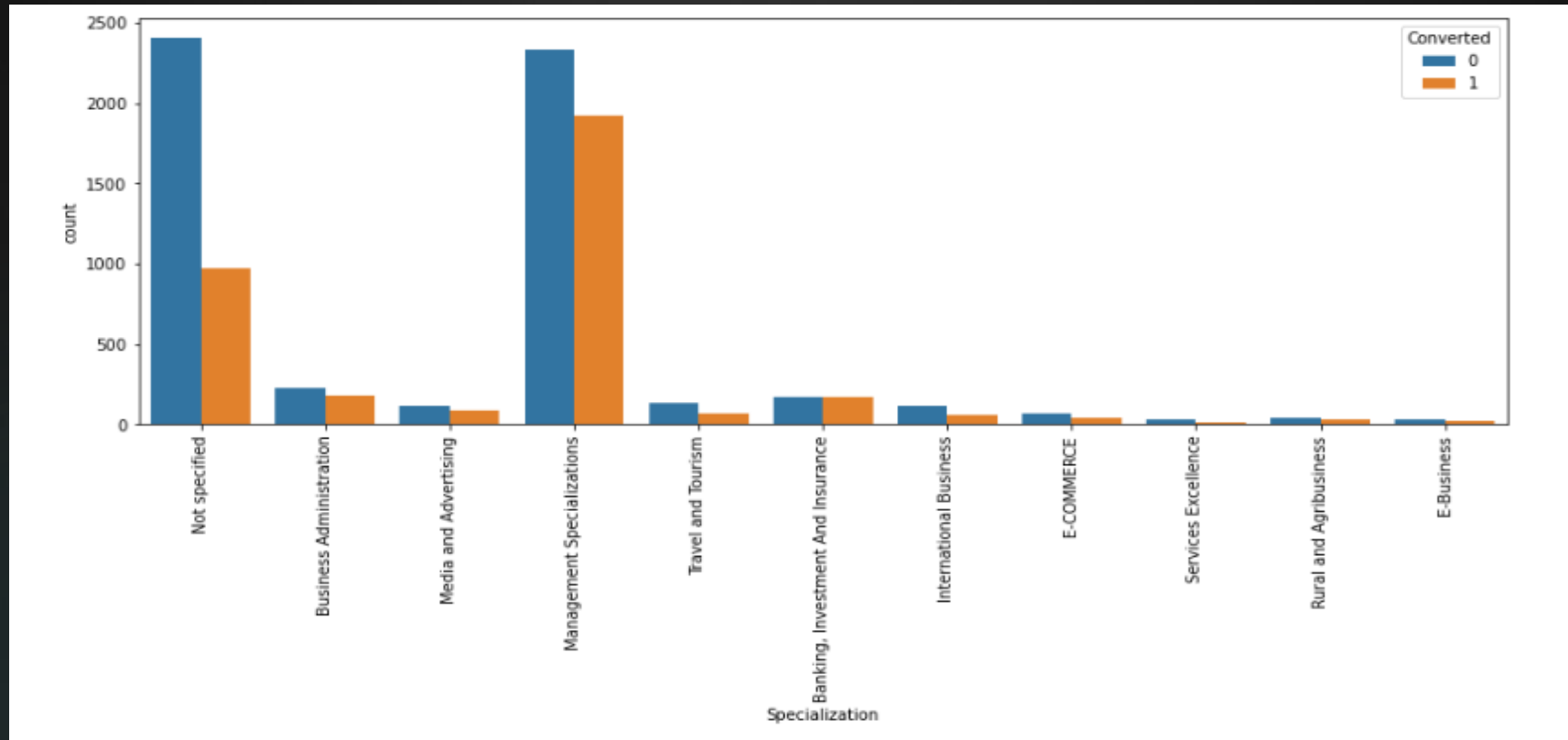
Visualization

This plot shows the distribution of converted and non-converted leads across different specializations.
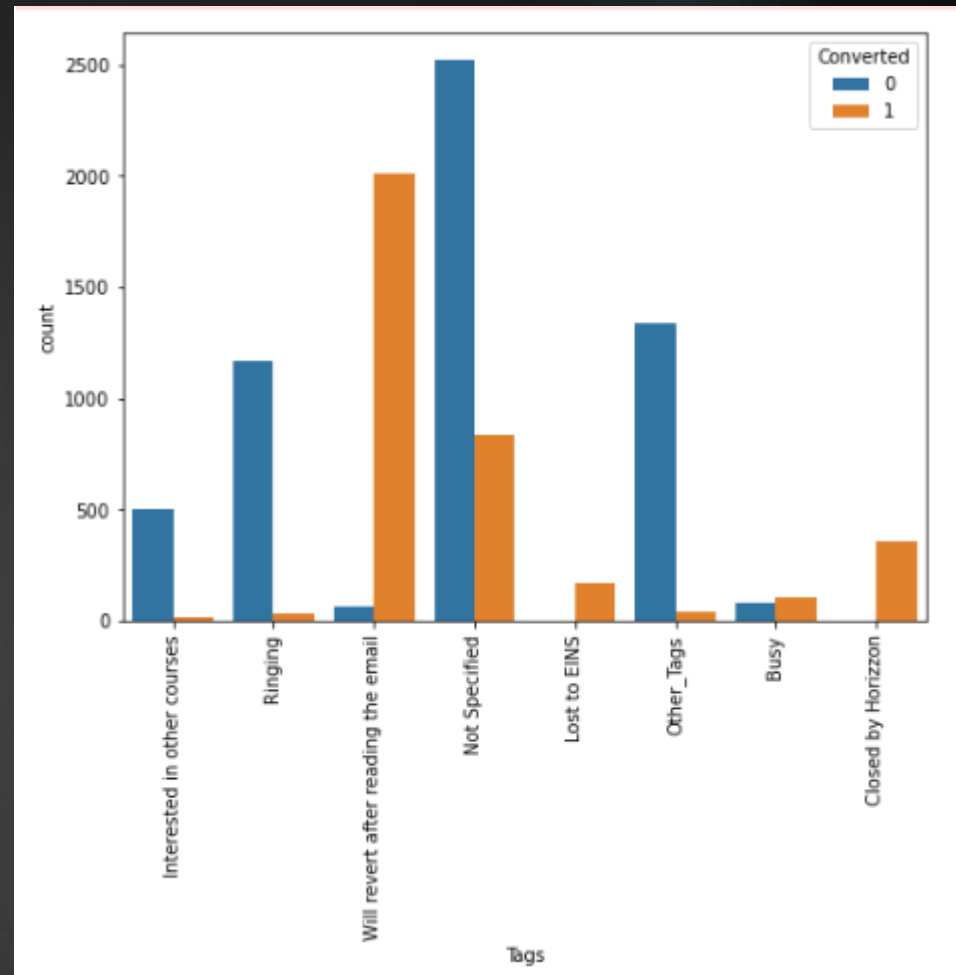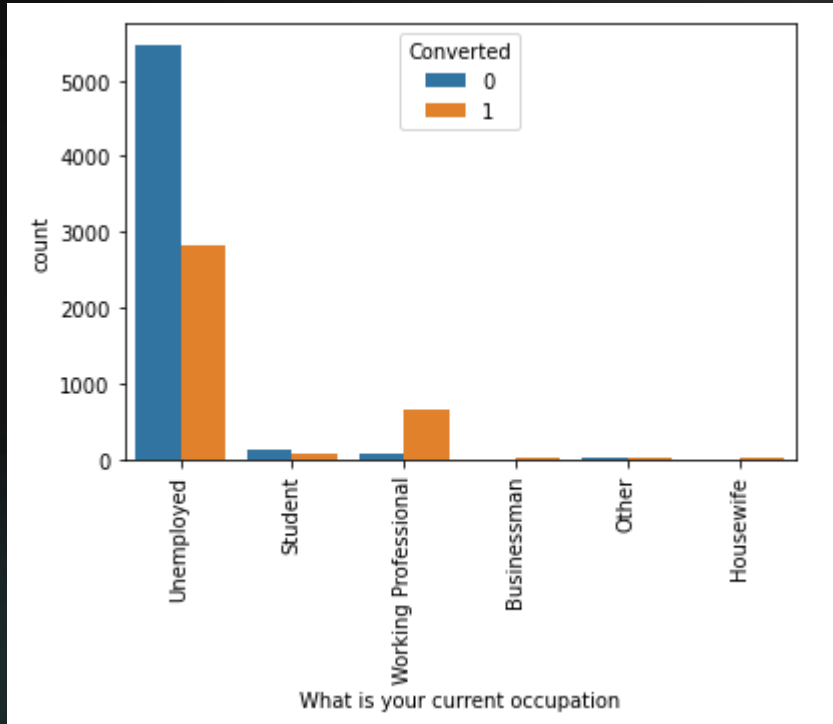
These plots shows variations in conversions and non conversions across different categorical variables ('What is your current occupation' , 'Tags')
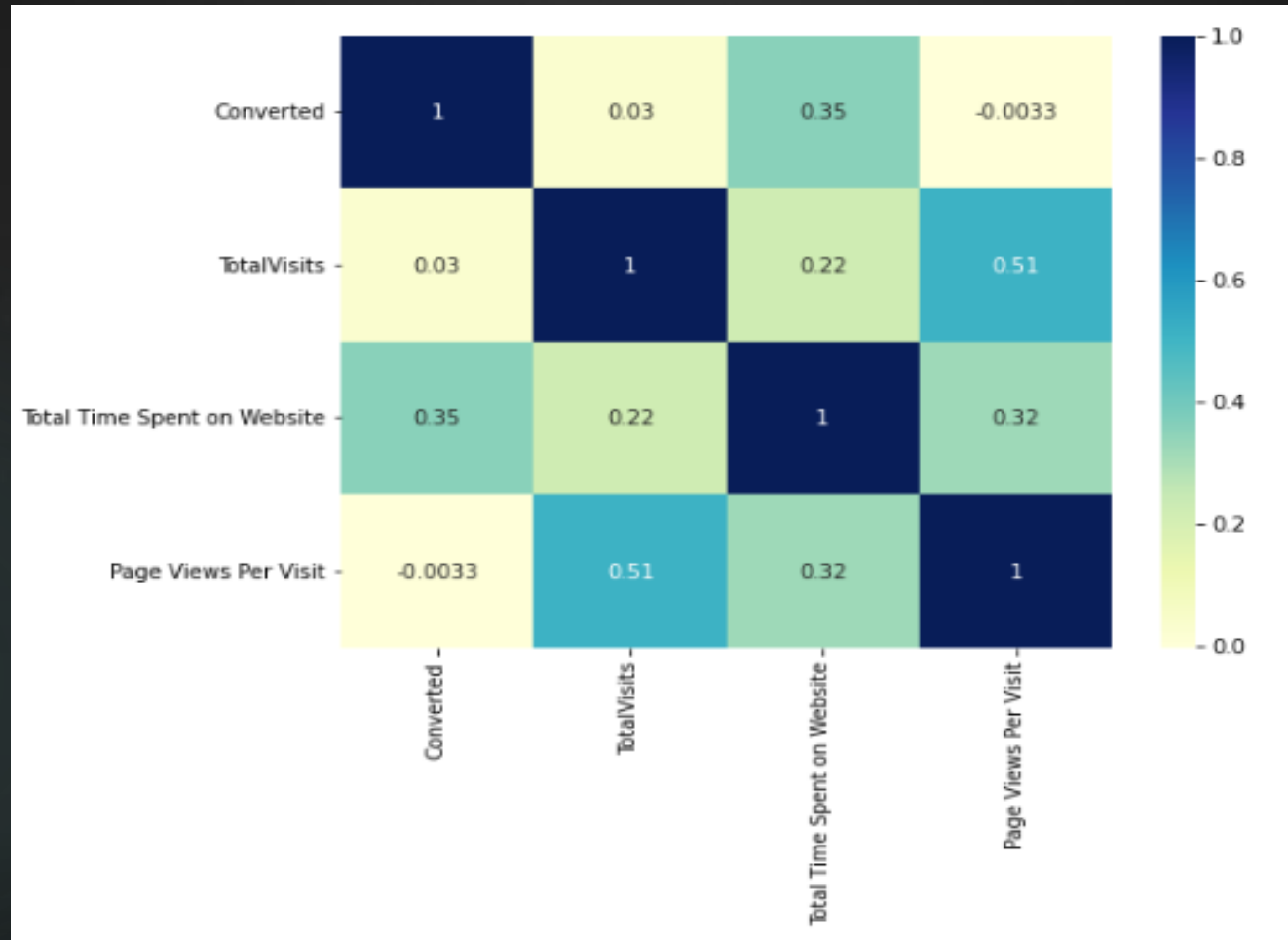
These plots shows variations in conversions and non conversions across different categorical variables ('Lead Source' , Lead Origin')
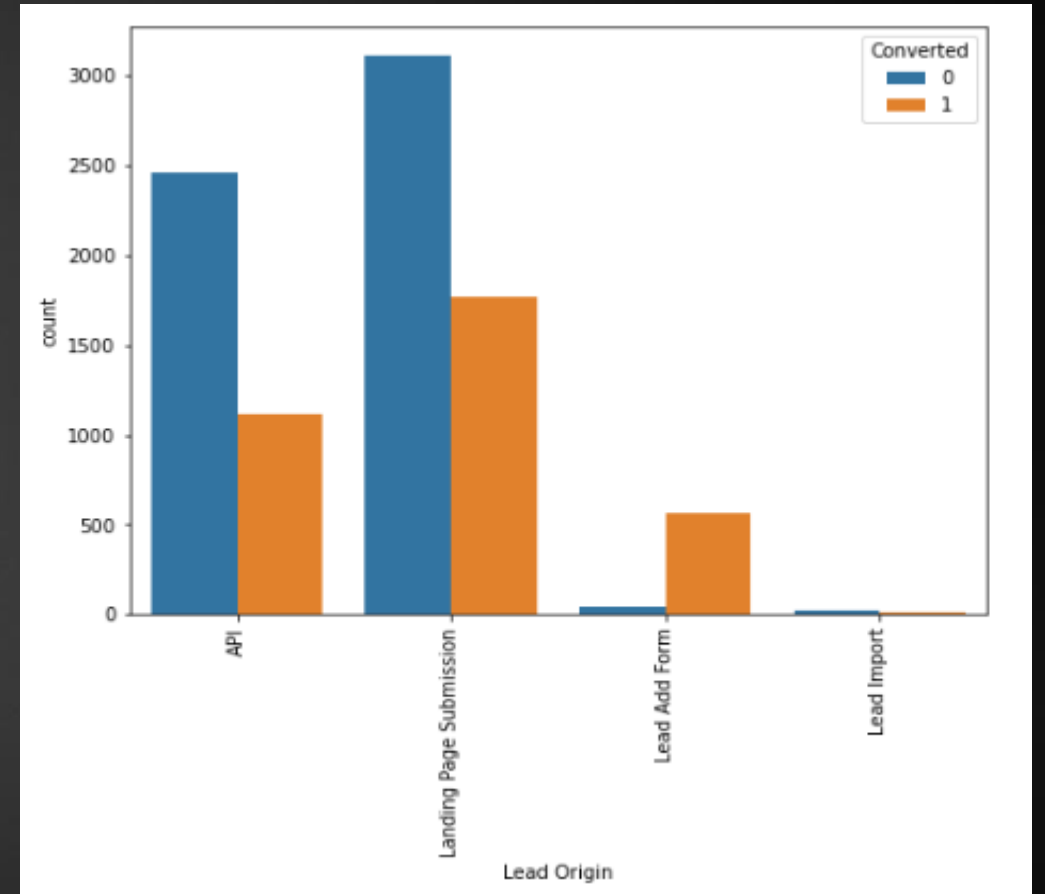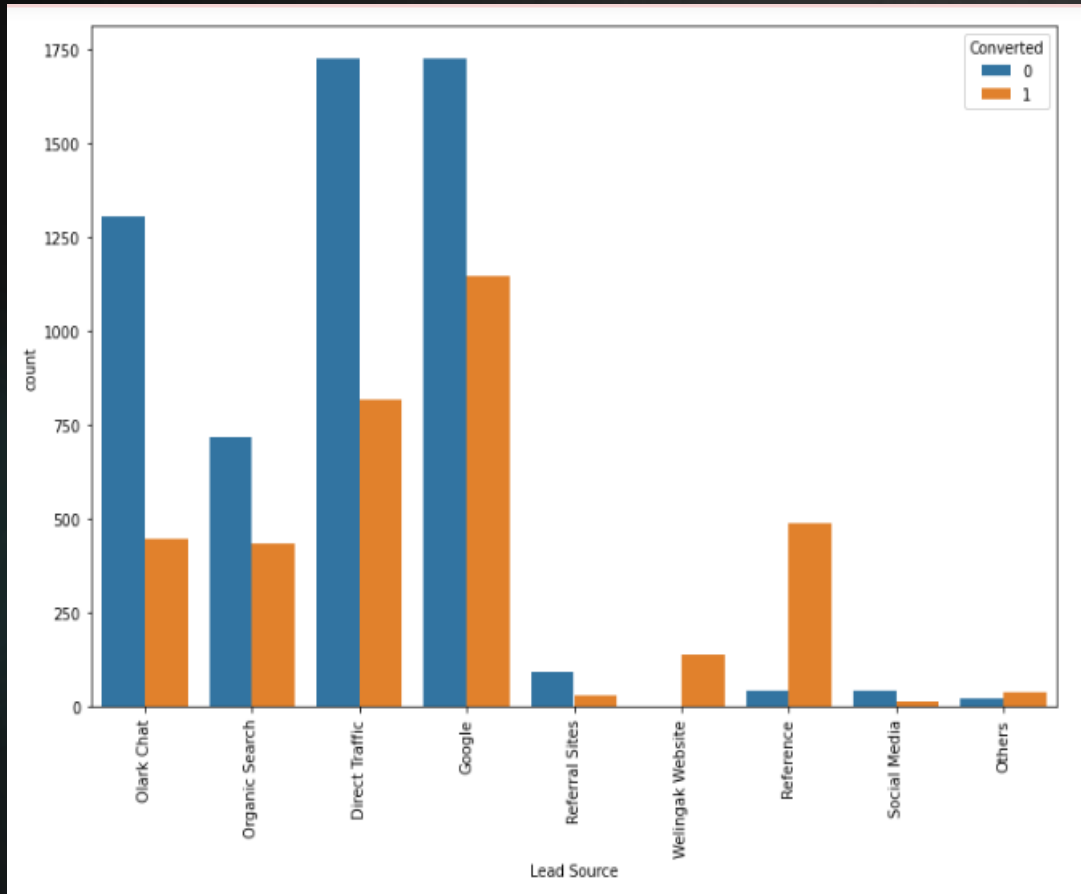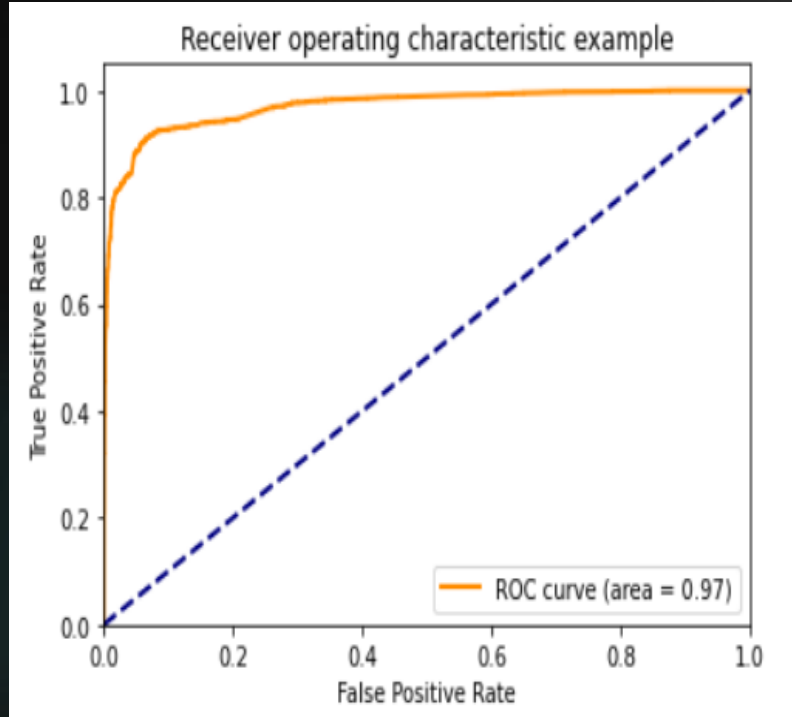
These plots shows variation in numerical columns for those who Converted and those who didn't.
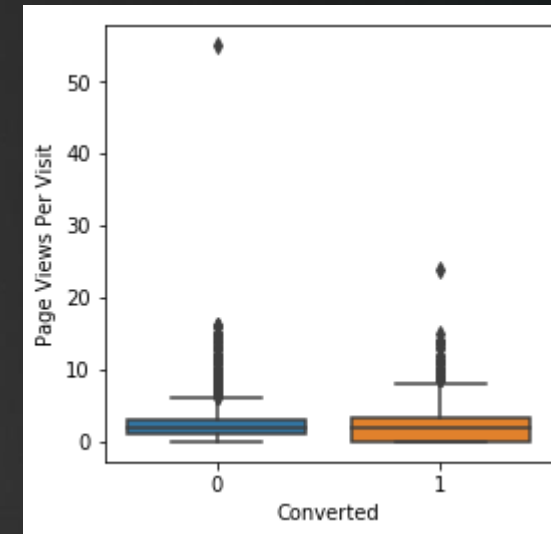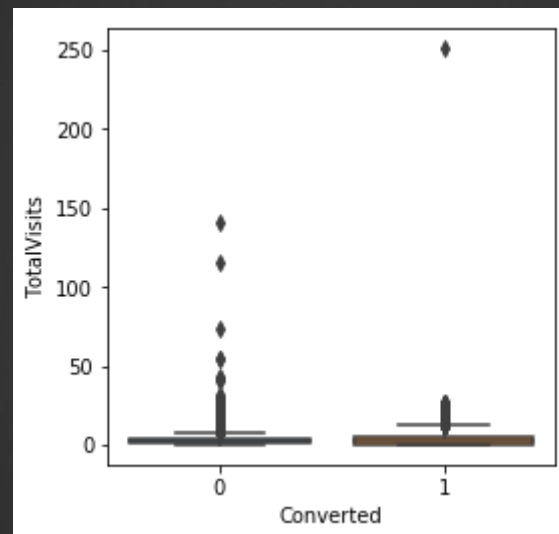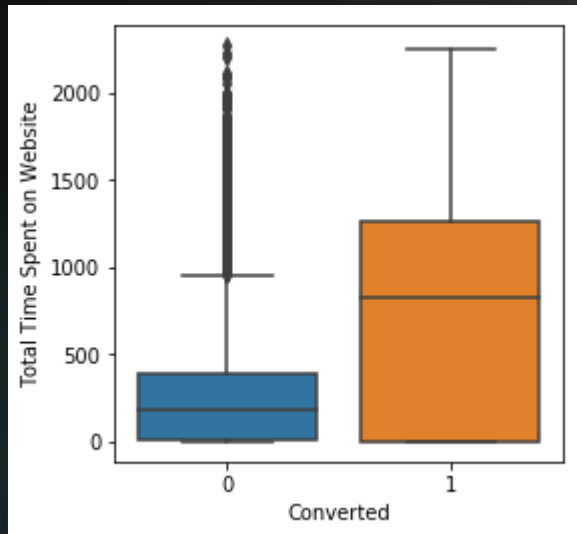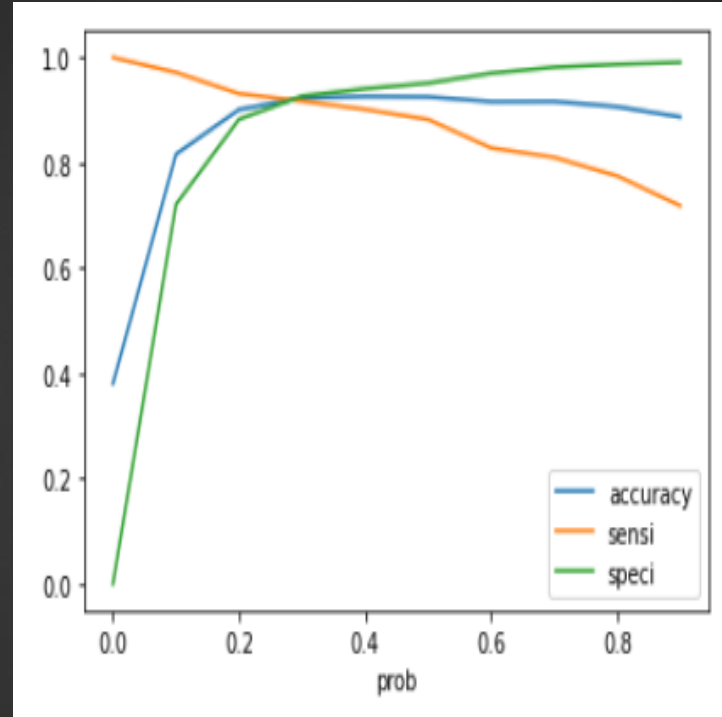
Checking correlation between various numerical variables.

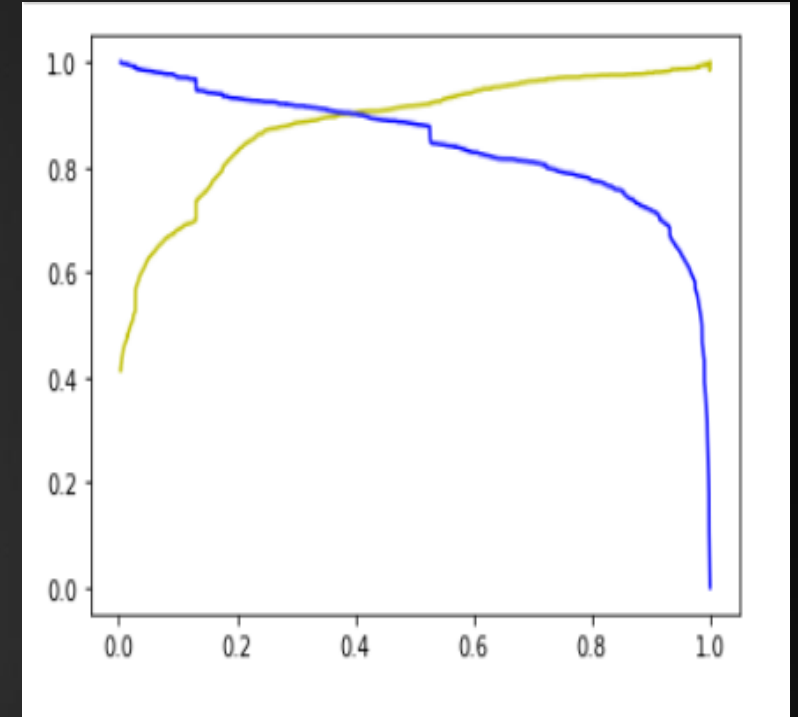**Evaluating our Final model with ROC Curve (Area under ROC = 0.97)**

0.3 is the optimum cut-off point

Precision-Recall curve for Evaluation

Model Evaluation

Evaluation:

- ▶ Overall accuracy on Test set = 92.5 %

- ▶ Sensitivity = 88%

- ▶ Specificity = 95%

Inferences

Top 3 variables that contribute towards lead conversion are:

▶ Total Time Spent on Website

▶ Last Activity_SMS Sent

▶ Lead Origin_Lead Add Form

Top 3 variables on which the company should focus:

▶ Last Activity_SMS Sent (positively impacting)

▶ Tags_Will revert after reading the email (Positively impacting)

▶ Last Activity_Olark Chat Conversation (negatively impacting)

Conclusion

Our Logistic Regression Model is decent and accurate enough, with 92.78 % Accuracy, 92 % Sensitivity and 93.26 % Specificity on the test set.

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.

Recommendations

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
- ► Lead add form has high conversion to non conversion ratio, The company can focus on this variable.
- ► Increase on sending SMS notifications since this helps in higher conversion
- ► Improve the Olark Chat service since this is affecting the conversion negatively