

# X Education - Lead Scoring Case Study

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

**Group Members: Prashant Jain  
Viplove Jain**

# 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

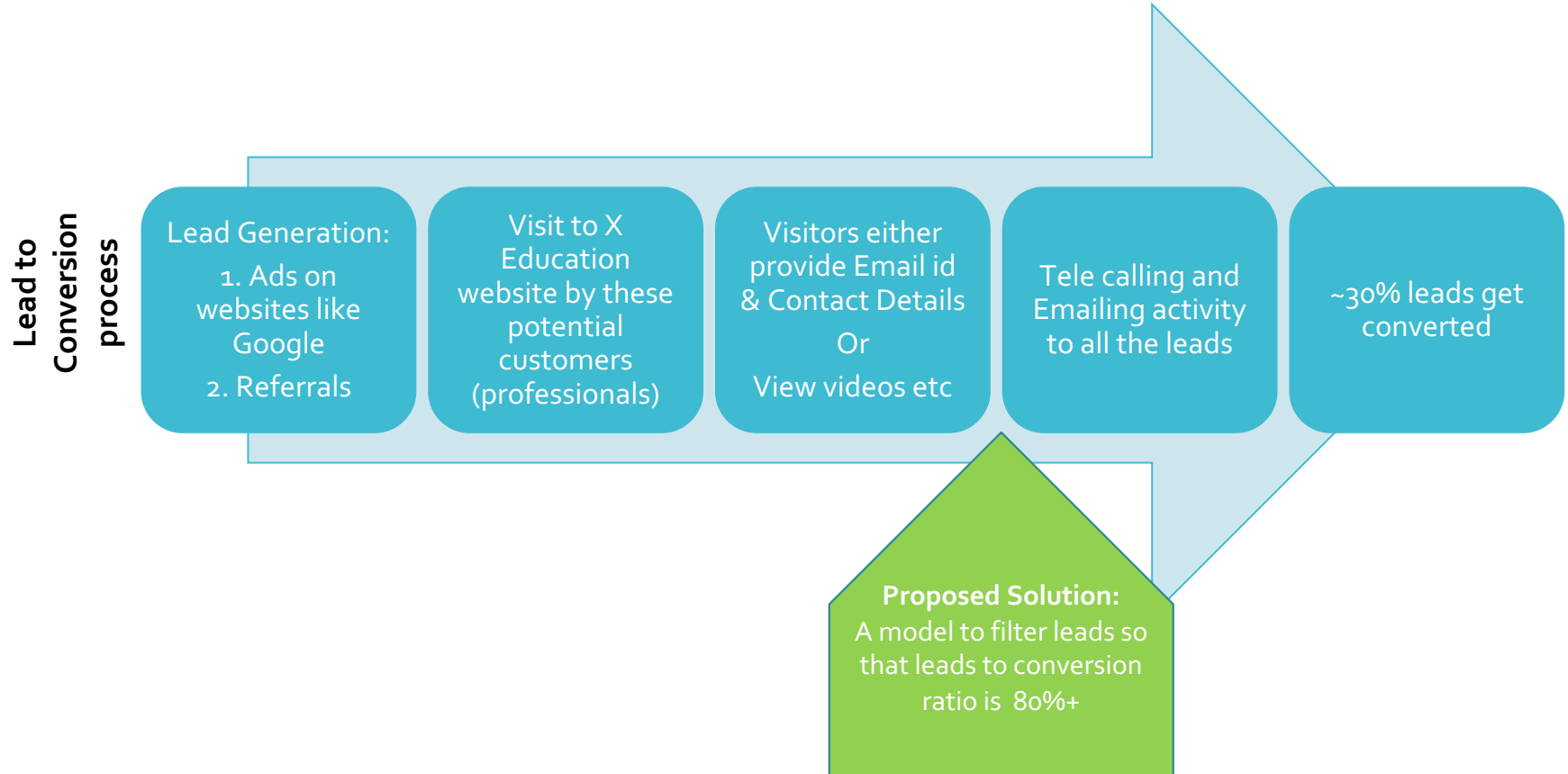
- X Education gets a lot of leads but its lead conversion rate is very poor
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be 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~~ conversion rate to be 80%.

# Lead – Conversion Process



# Proposed Solution

## Selection 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.

## Communicating with Hot Leads

### Focus Communication

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

## Conversion of Hot Leads

### Increase conversion

Since we focussed 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

## 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.

---

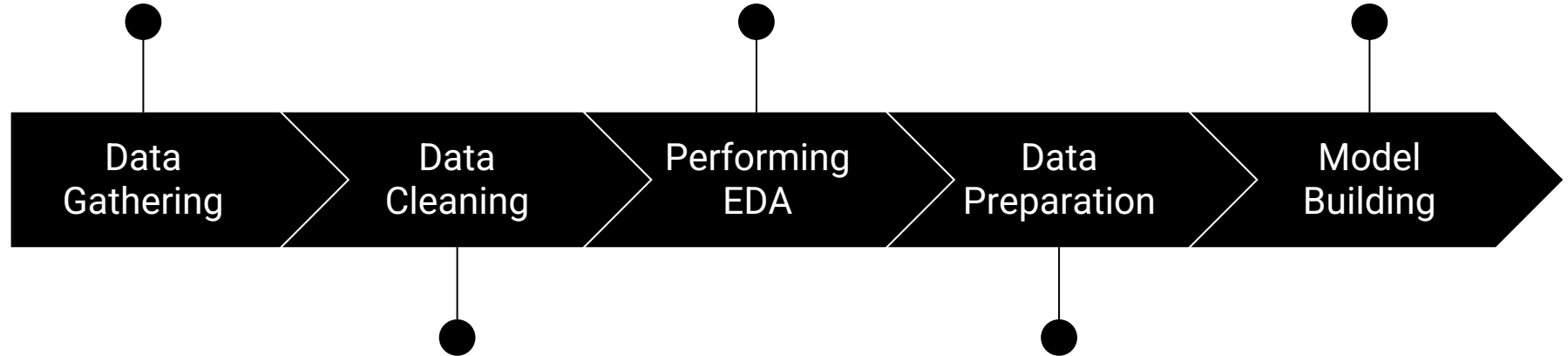


# Implementation

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



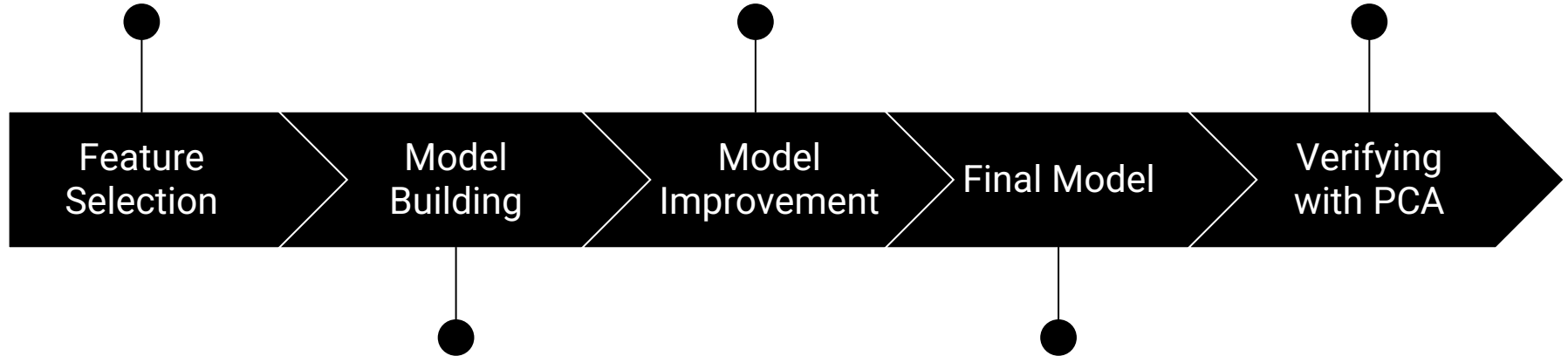
Duplicate removal, null value  
treatment, unnecessary  
column elimination, etc.

Outlier Treatment,  
Feature-  
Standardization

Selection of top 25  
features using RFE

Reduction of columns  
and Model re-building

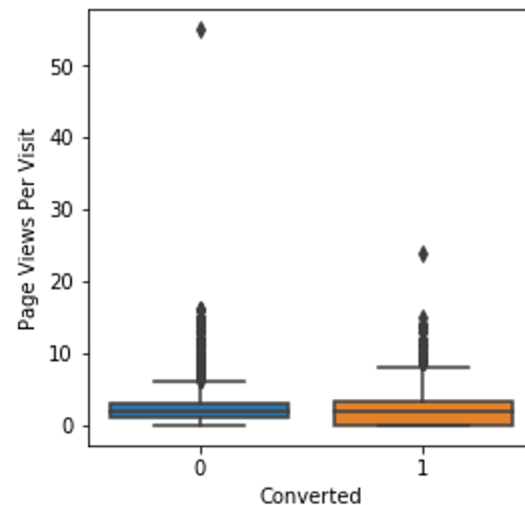
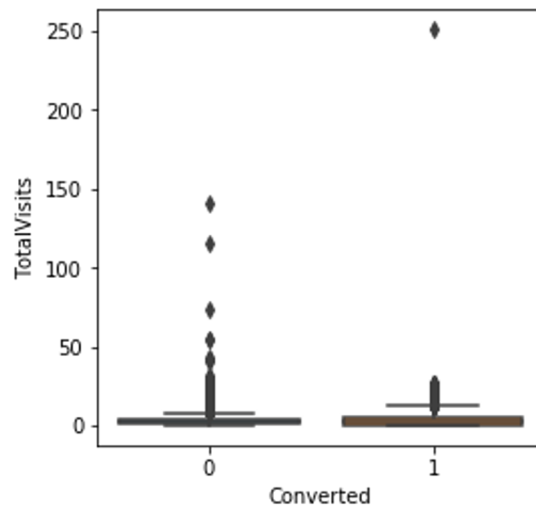
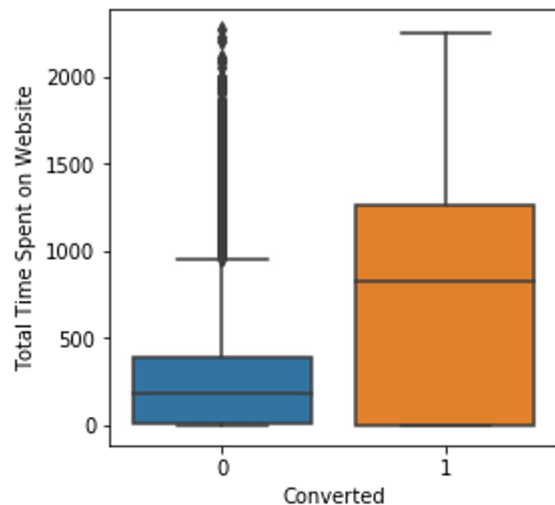
Verifying our Final Model  
Accuracy etc. with model  
built with PCA



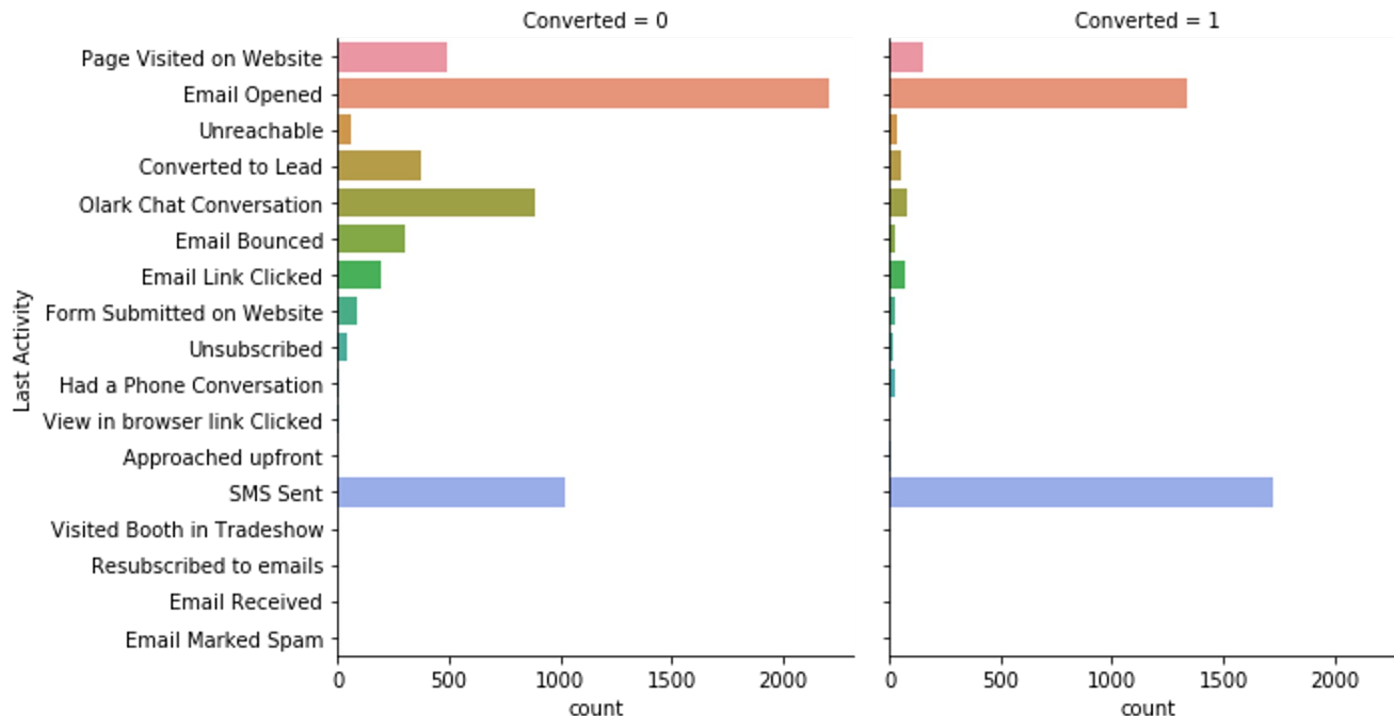
Model building using RFE for  
selected columns

Final Model Analysis and  
performance on Test  
Data

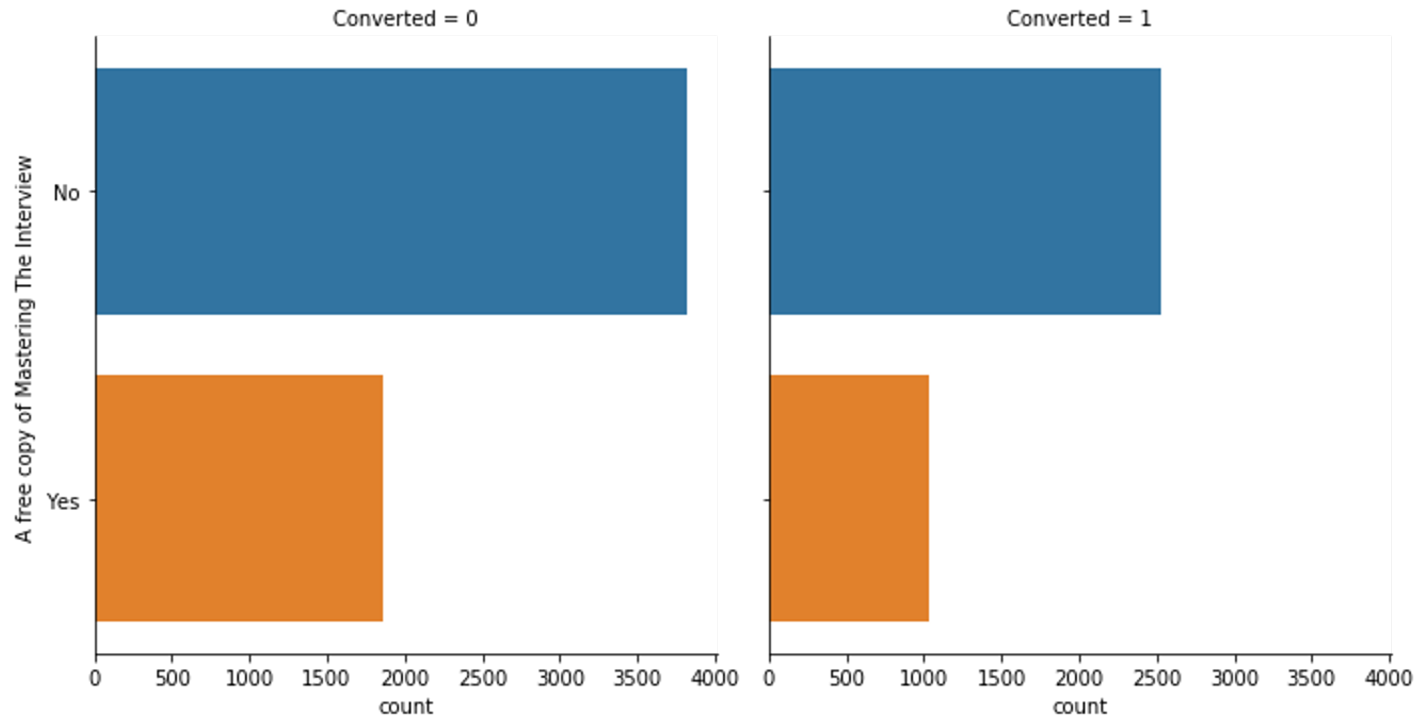
# Plots (Visualization)



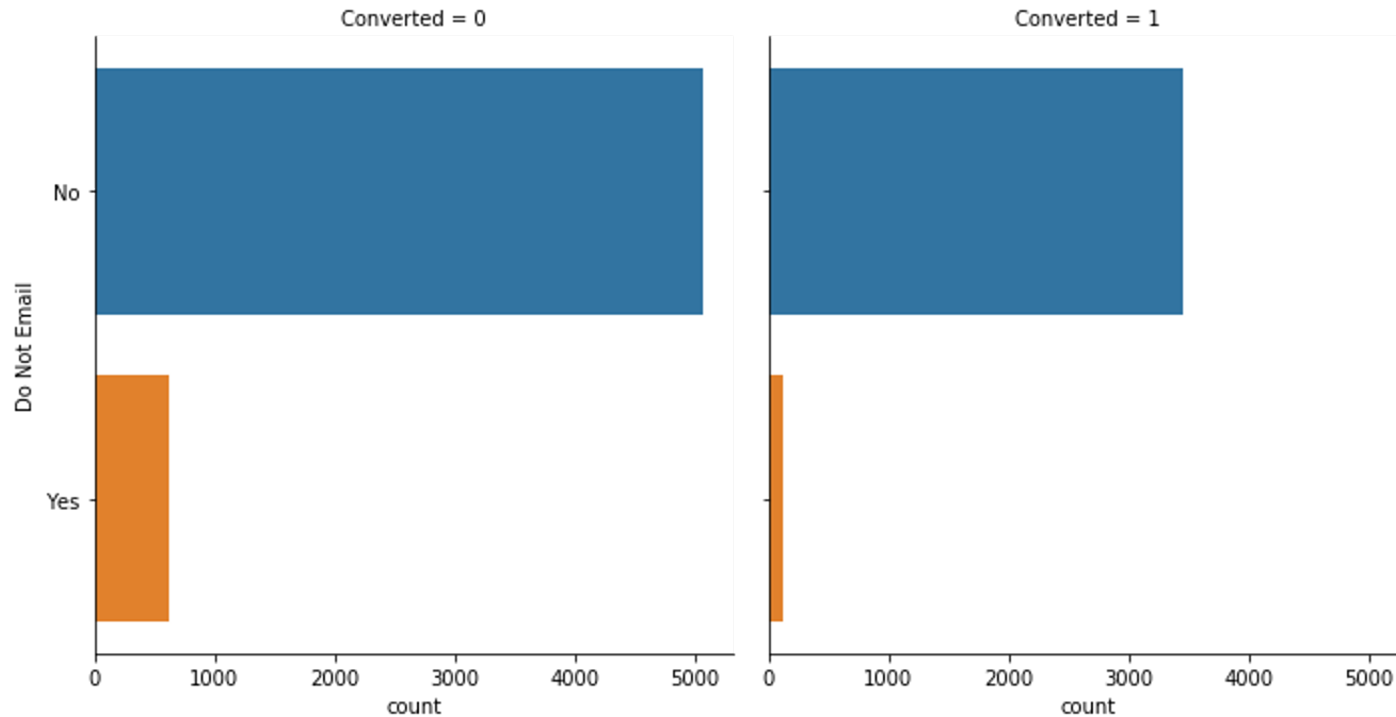
**EDA plots depicting variation in numerical columns for those who Converted and those who didn't.**



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

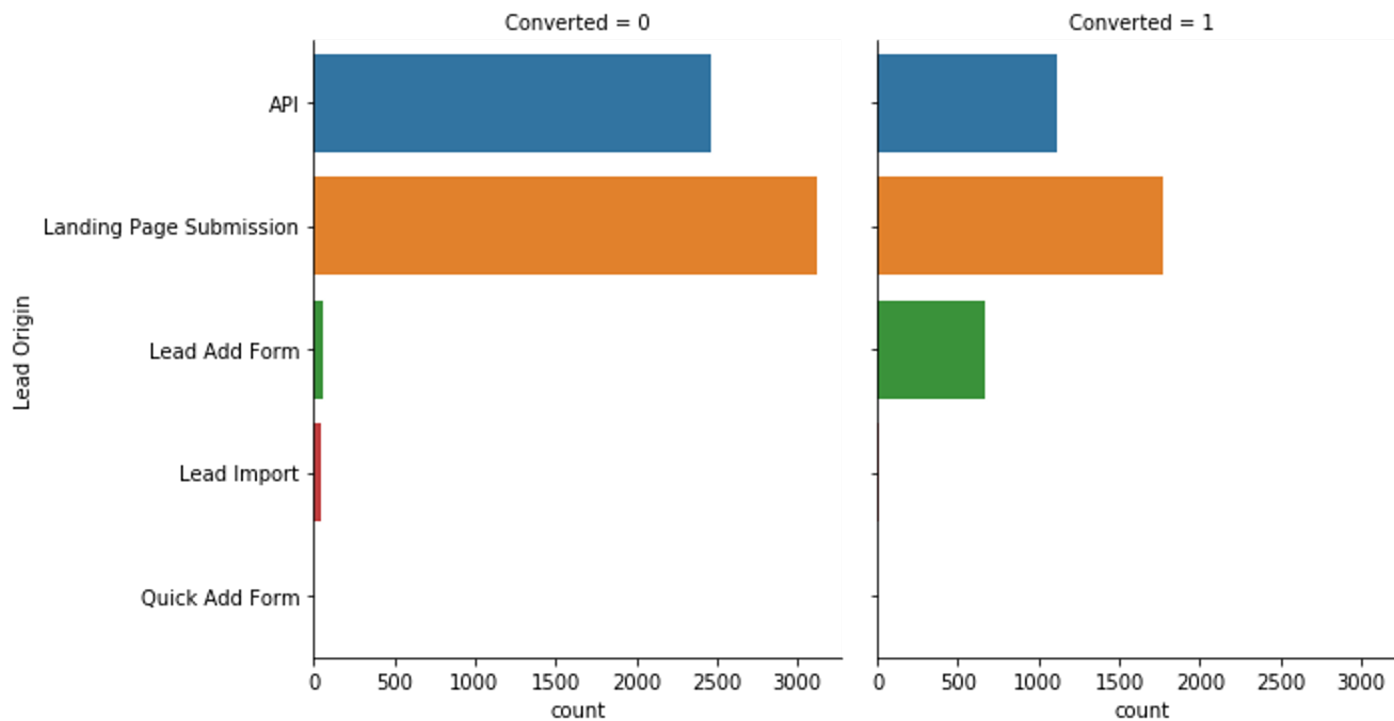


**EDA plots depicting variation in categorical column (A free copy of Mastering The Interview) for those who Converted and those who didn't.**

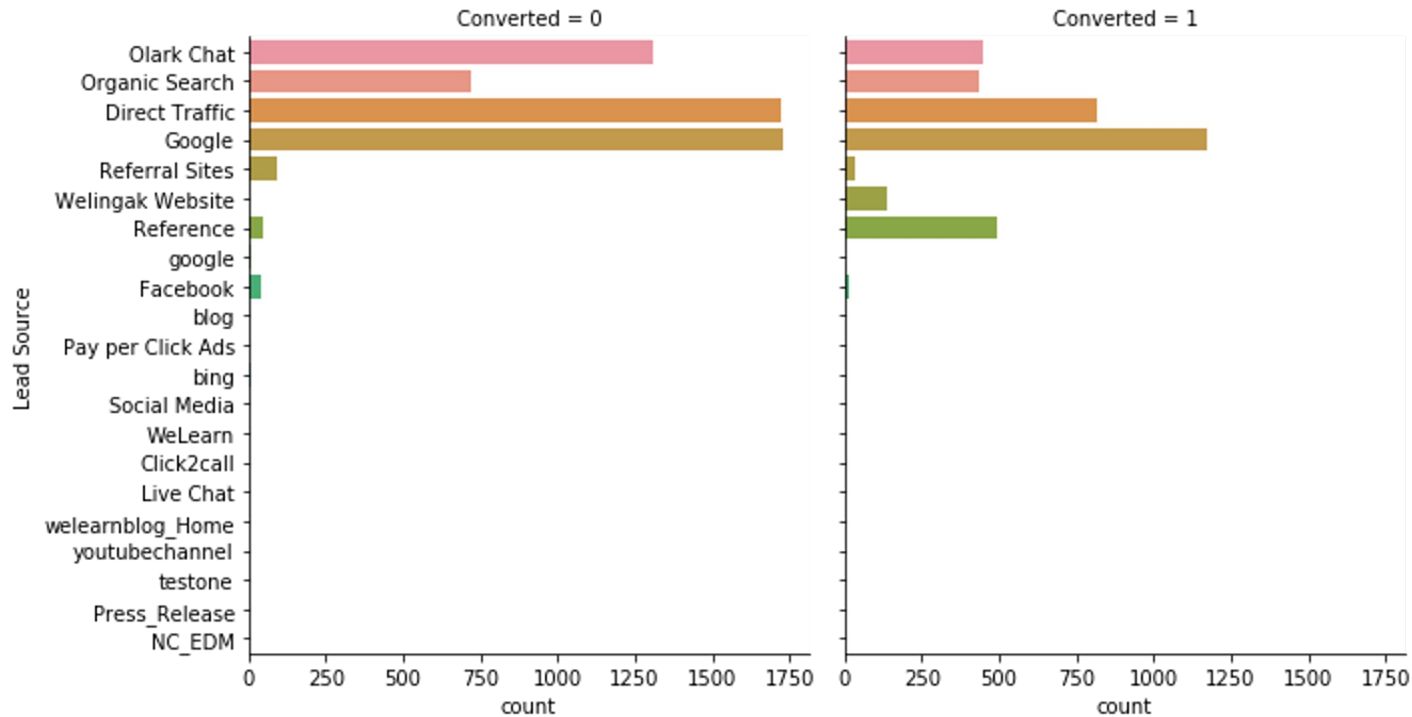


**EDA plots depicting variation in categorical column (Do Not Email) for those who Converted and those who didn't.**

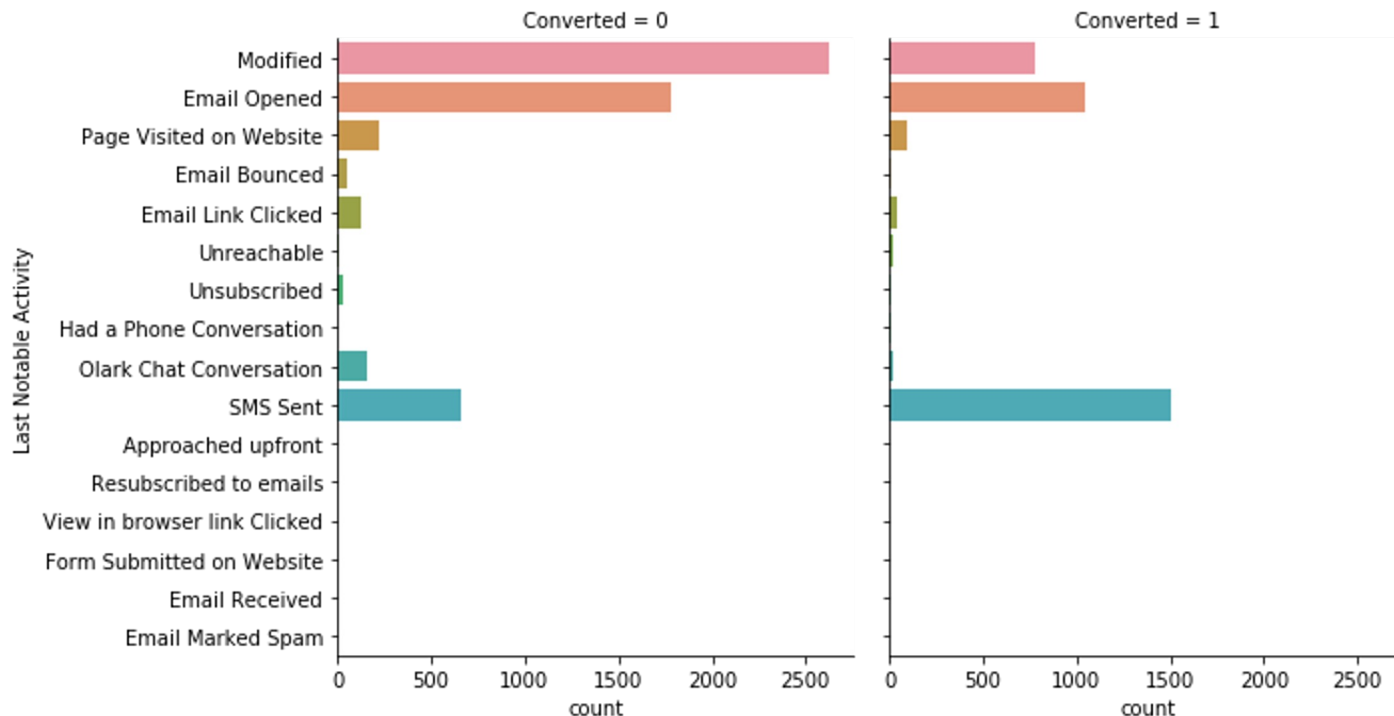




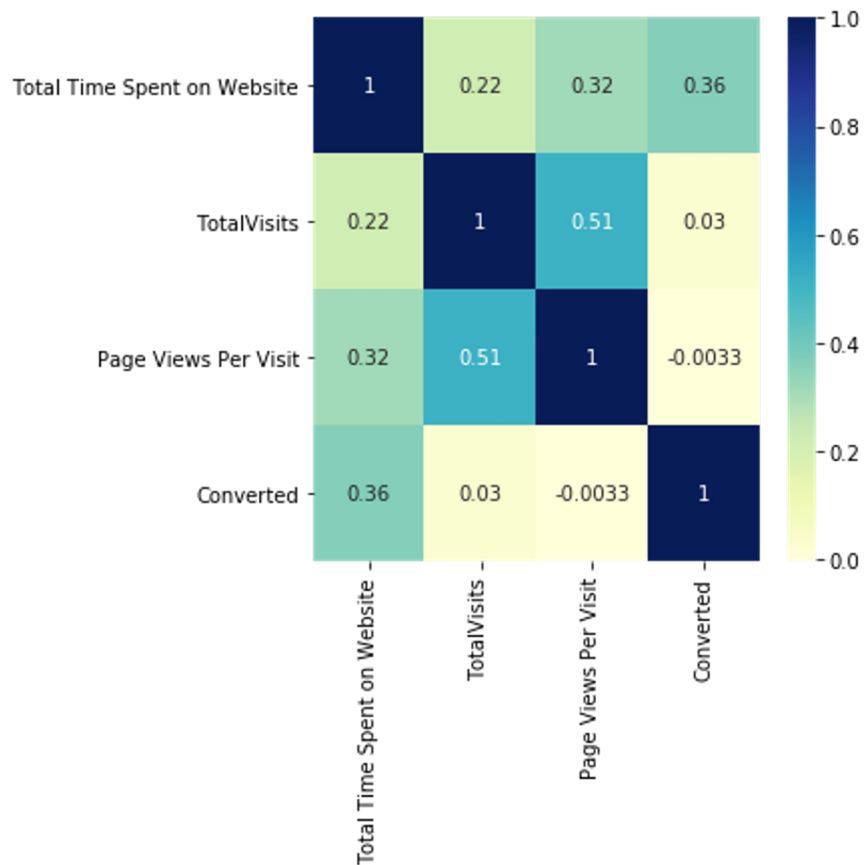
**EDA plots depicting variation in categorical column (Lead Origin) for those who Converted and those who didn't.**



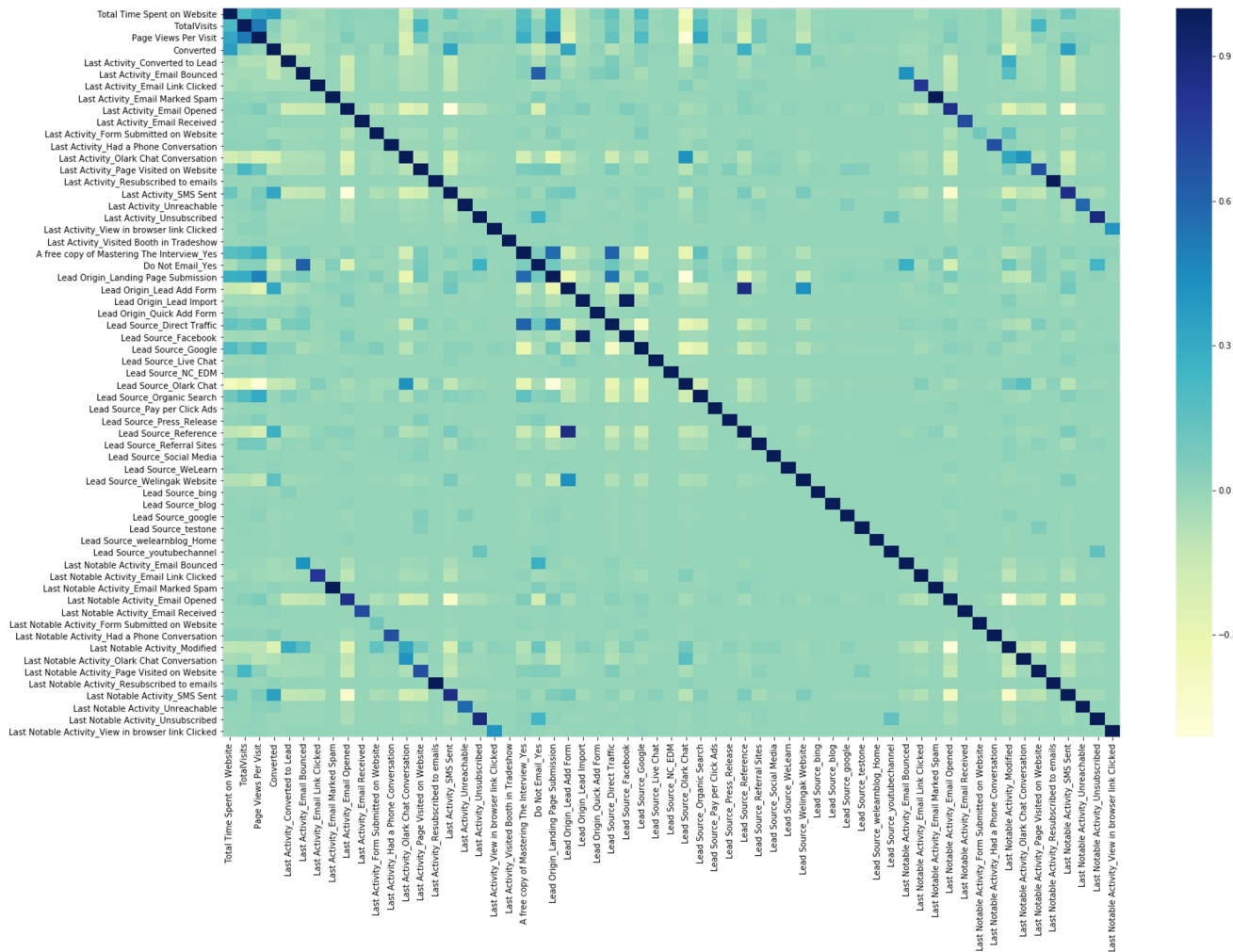
**EDA plots depicting variation in categorical column (Lead Source) for those who Converted and those who didn't.**



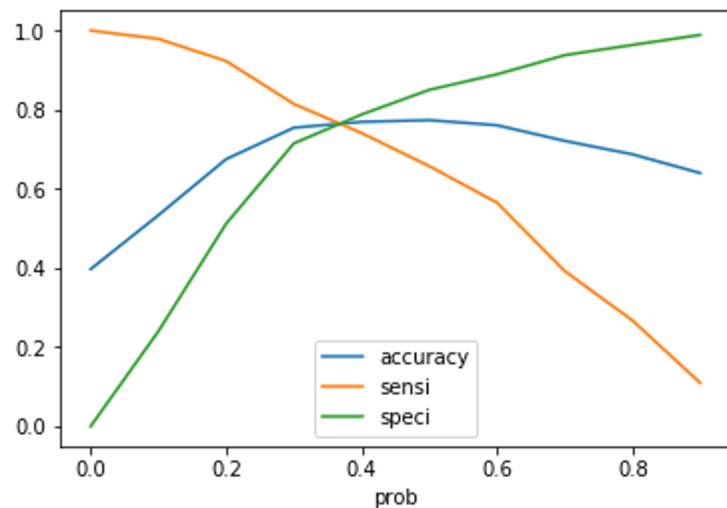
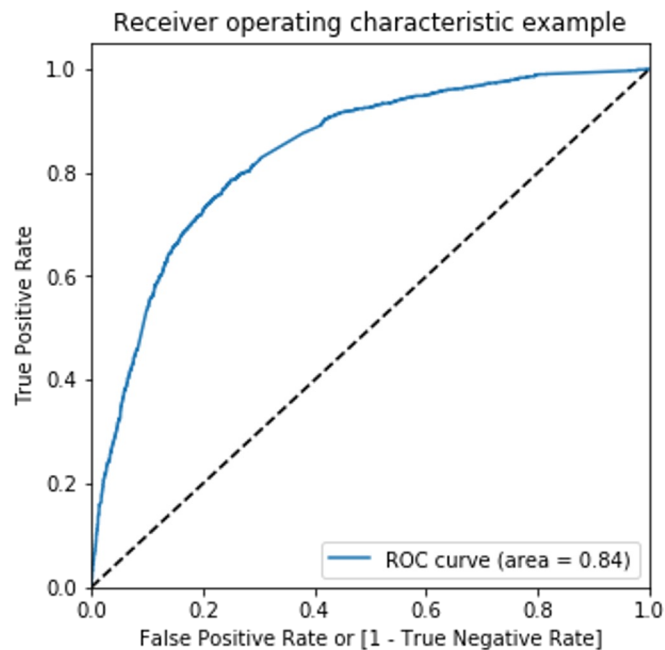
**EDA plots depicting variation in categorical column (Last Notable Activity) for those who Converted and those who didn't.**



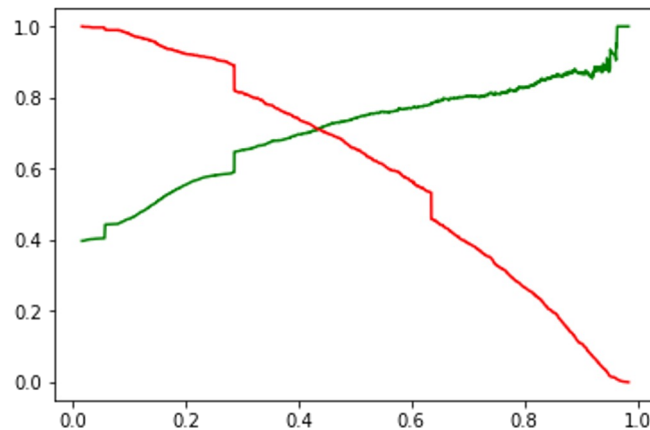
**EDA plots depicting correlation (Heat Map) of all selected numerical columns.**

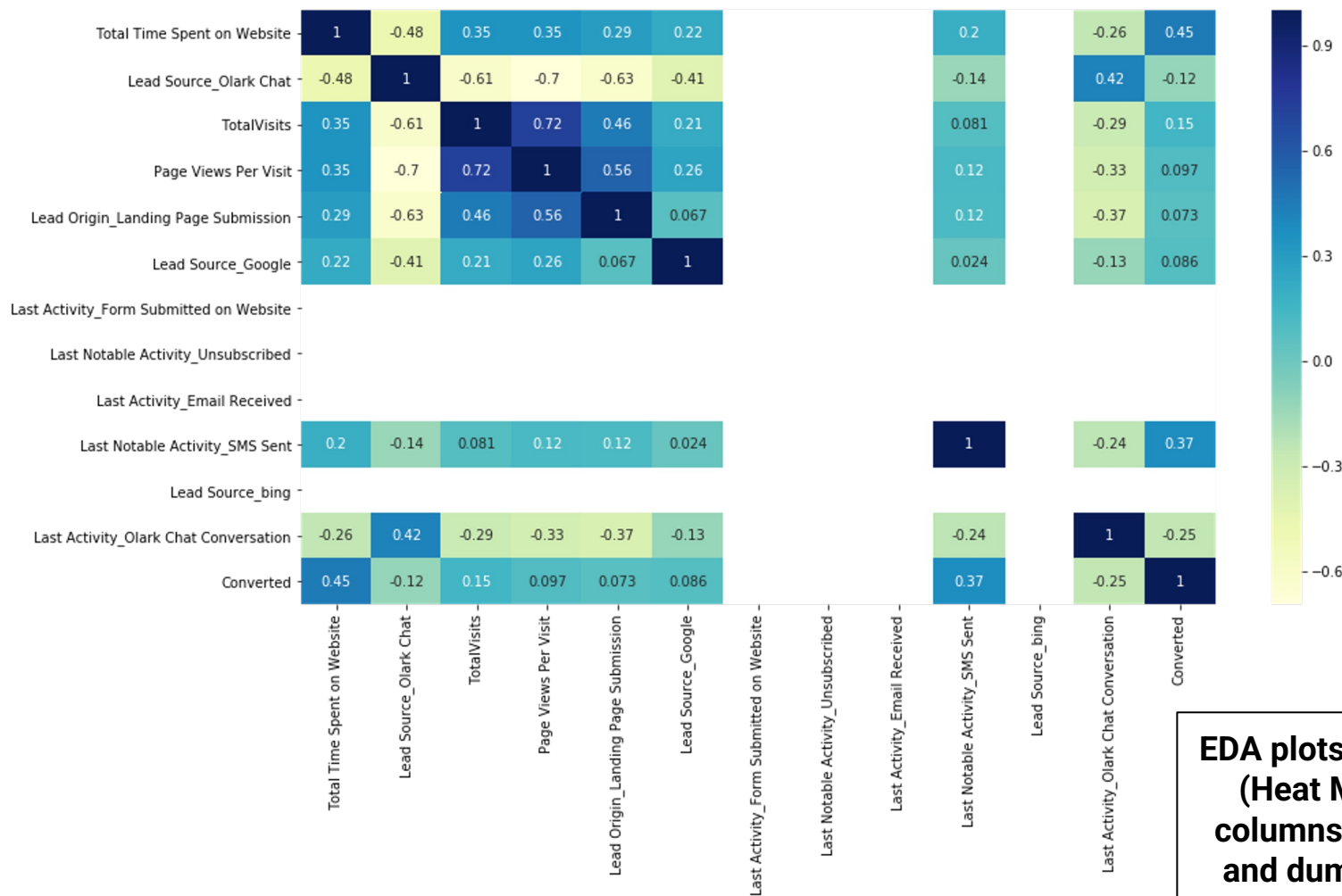


**EDA plots depicting correlation (Heat Map) of all selected columns (numerical columns and dummy columns).**



**Linear Regression Final  
Model Parameters  
Area under ROC = 0.84  
Intermediate cut-off = 0.35  
Final cut-off = 0.42**





**EDA plots depicting correlation (Heat Map) of all selected columns (numerical columns and dummy columns) in our final Model.**

Inference / Conclusion



# Model Analysis

Performance of our Final Model

Overall accuracy on Test set: 0.786

Sensitivity of our logistic regression  
model: 0.733

Specificity of our logistic regression  
model: 0.823

---

# Inferences from Model

Business Insights Derived from our Model

Top 3 variables in model, that contribute towards lead conversion are:

- Total Time Spent on Website
- Last Notable Activity\_SMS Sent
- TotalVisits

---

# Inferences from Model

Business Insights Derived from our Model

Top 3 variables in my model, that should be focused are:

- Last Activity\_SMS Sent (positively impacting)
  - Last Activity\_Olark Chat Conversation (negatively impacting)
  - Lead Source\_Olark Chat (negatively impacting)
-

# Conclusion 1 (LR Model)

Our Logistic Regression Model is decent and accurate enough when compared to the model derived using PCA, with 78.6% accuracy on the test set, 73.3% sensitivity, and 82.3% specificity. We can vary these parameters by adjusting the cut-off value and thus predict hot leads based on scenarios such as the availability of extra resources and vice versa.

# Conclusion 2 (Recommendation)

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

- Increase user engagement on their website, as this helps achieve higher conversions.
- Increase the sending of SMS notifications, as this helps achieve higher conversions.
- Increase total visits through advertising, as this helps achieve higher conversions.
- Improve the Olark Chat service, as it is currently affecting the conversion rate negatively.