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: Amritpal

BACKGROUND -

X EDUCATION COMPANY

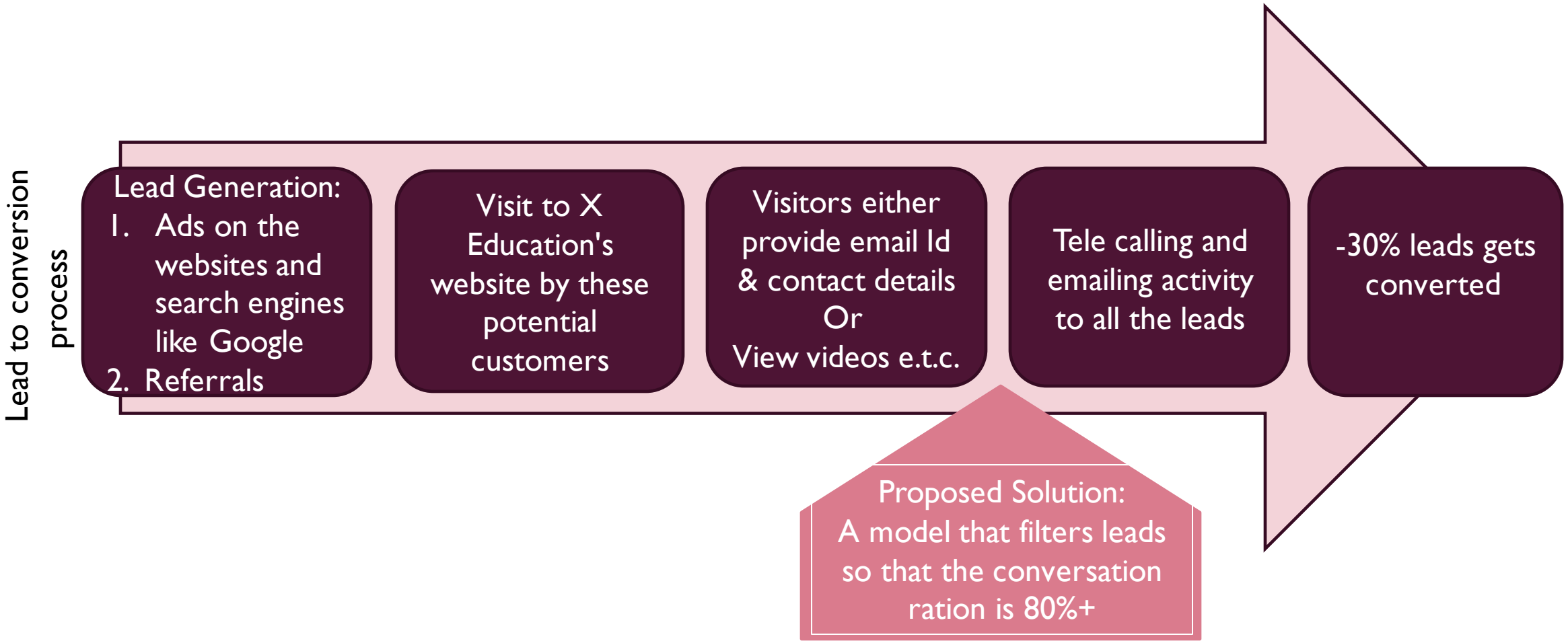
- X Education is an education/edtech company that sells online courses to industry professionals
- All the interested professionals land on their website through various sources
- The company markets its courses on several websites and search engine like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos available there.
- When the people who landed on the website fill up the form providing their email address or phone number, they are then classified as lead
- Once the leads are generated, employees from the sales team start to get in touch with them through calls, emails, etc. Through this process, some of the leads get converted into sales while most do not
- The typical lead conversion rate at X education is around 30%

PROBLEM STATEMENT -

X EDUCATION COMPANY'S PROBLEM

- Although X Education generates a lot of leads but the lead conversion rate is very poor
- In order to be more efficient, the company wishes to identify the most potential leads, which are also known as 'Hot Leads'
- If they successfully are able to identify this set of leads, the lead conversion rate should boost the sales team as the sales team will now be able to focus more on communicating with the potential leads rather than making calls to other non-potential users

LEAD – CONVERSION PROCESS



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

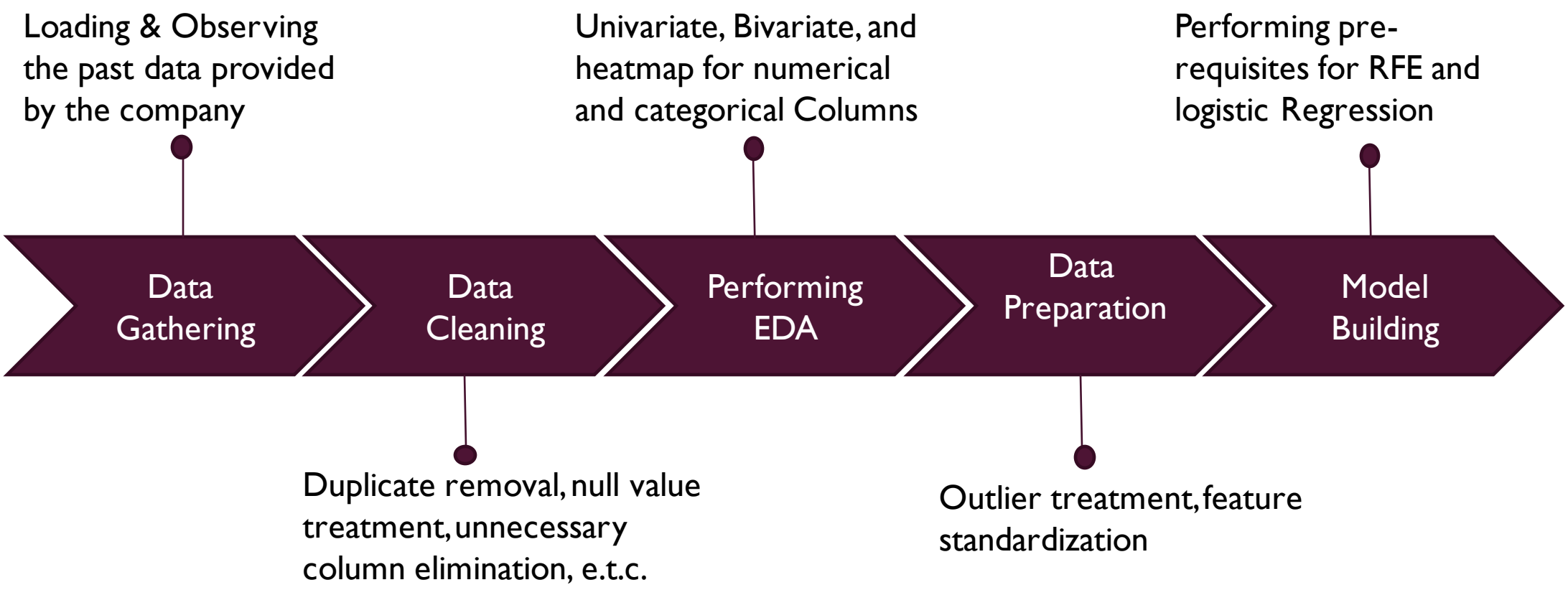
Since we focussed on hot leads, which were more probable to convert, we would have a better conversion rate, and hence we can achieve the 80% target.

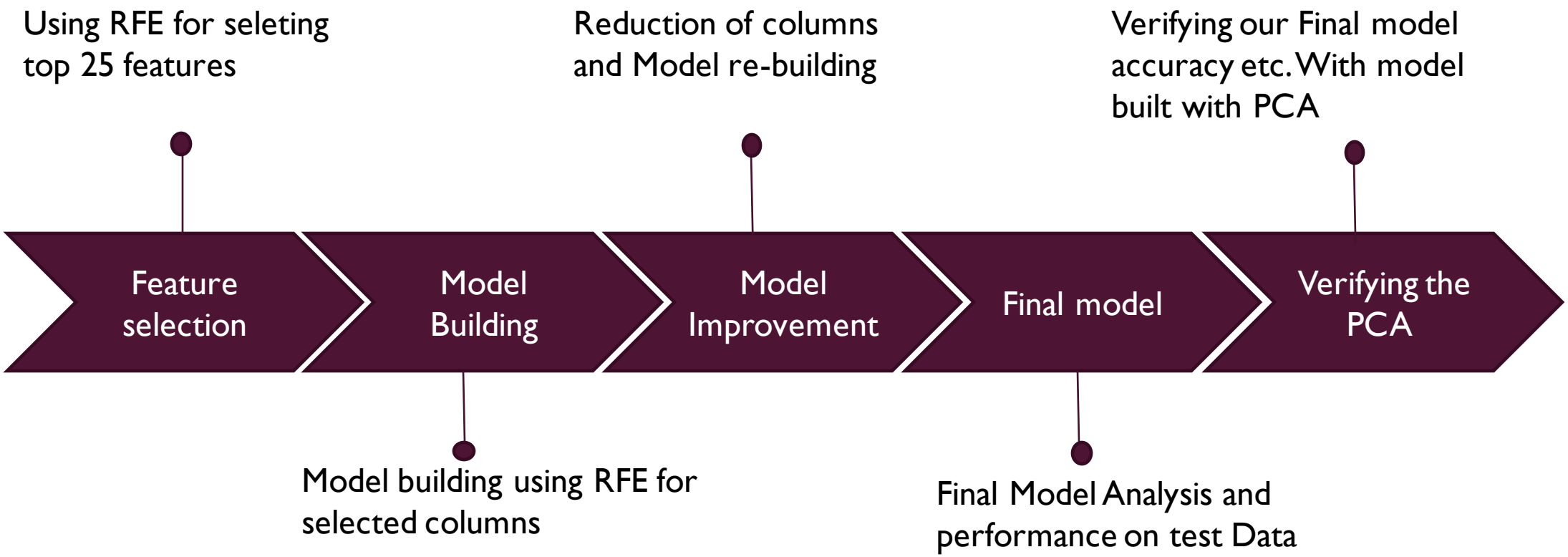
SOLUTION:

SELECTION OF HOT LEADS

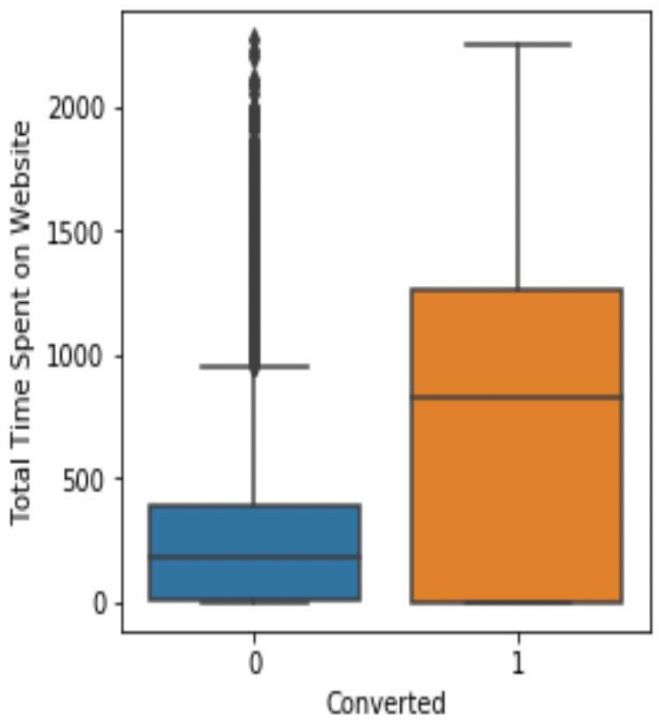
- Identifying the hot lead is the most curtail part of our problem's solution.
- The more accurate the hot leads are, the more chance we have of getting a higher conversion ratio.
- Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining the hot leads.

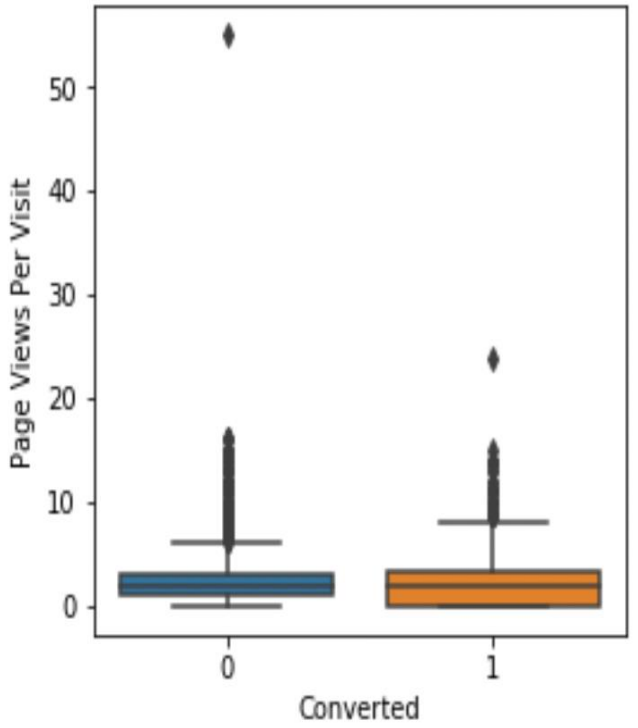
IMPLEMENTATION

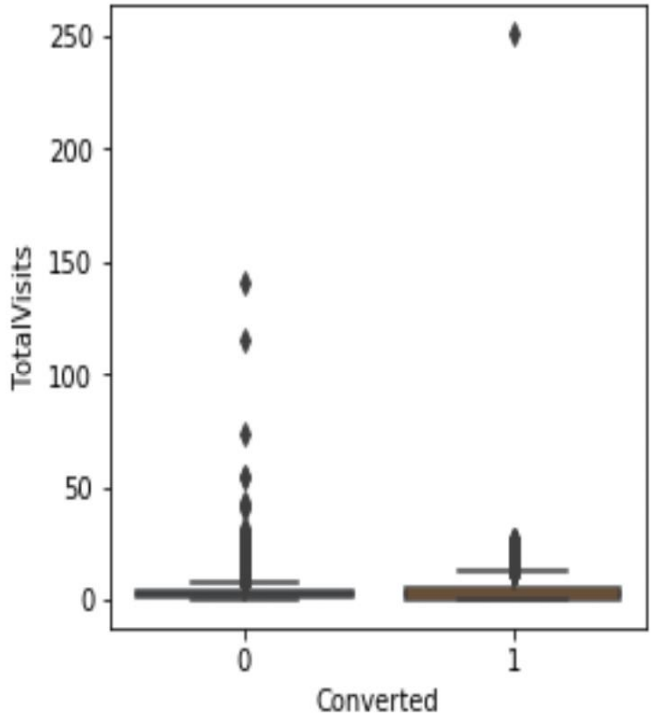

Loading & Observing the past data provided by the company
Univariate, Bivariate, and heatmap for numerical and categorical Columns
Performing pre-requisites for RFE and logistic Regression
Data Gathering
Data Cleaning
Performing EDA
Data Preparation
Model Building
Duplicate removal, null value treatment, unnecessary column elimination, e.t.c.
Outlier treatment, feature standardization


Using RFE for seleting top 25 features
Reduction of columns and Model re-building
Verifying our Final model accuracy etc. With model built with PCA
Feature selection
Model Building
Model Improvement
Final model
Verifying the PCA
Model building using RFE for selected columns
Final Model Analysis and performance on test Data

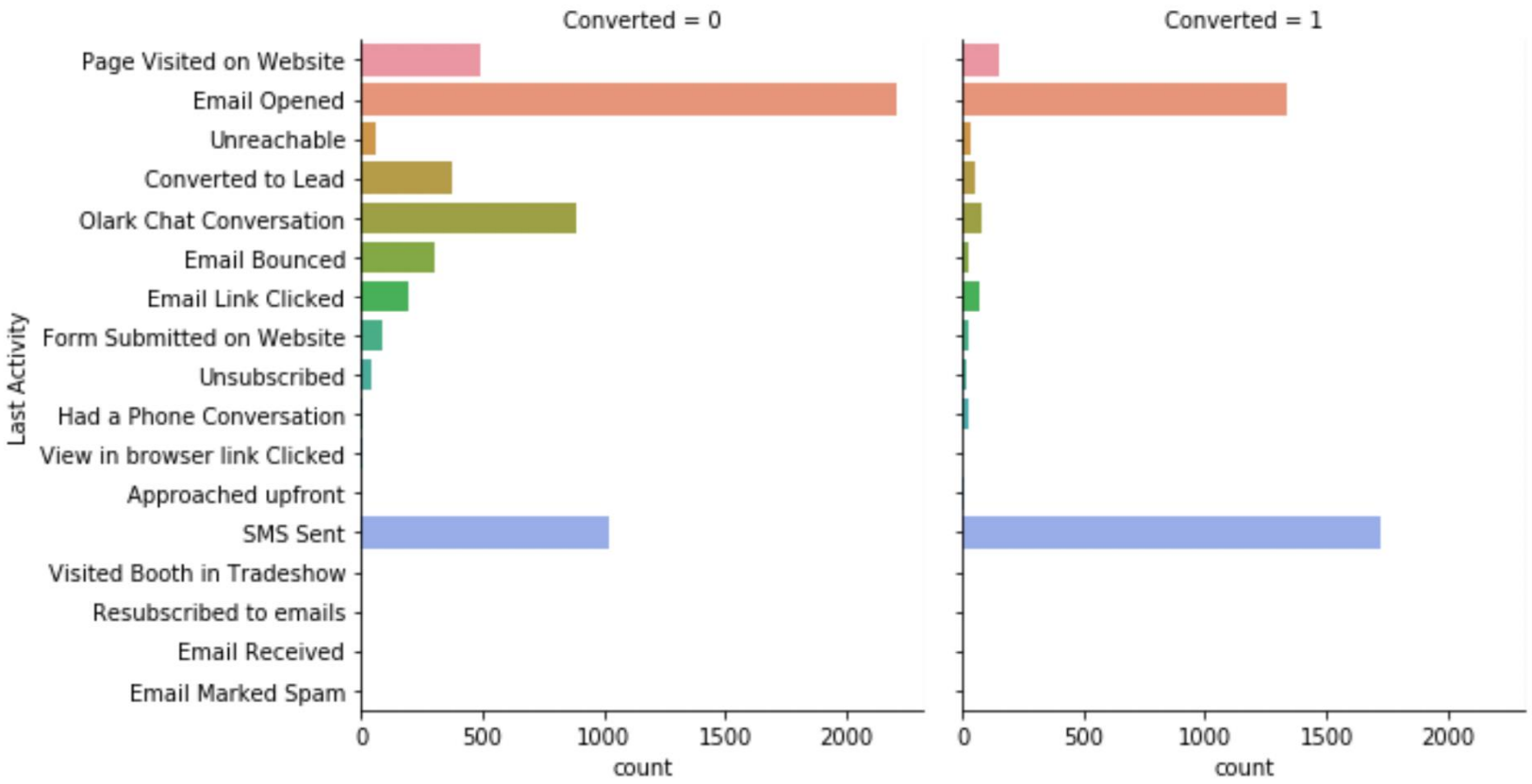
PLOTS (VISUALIZATION)



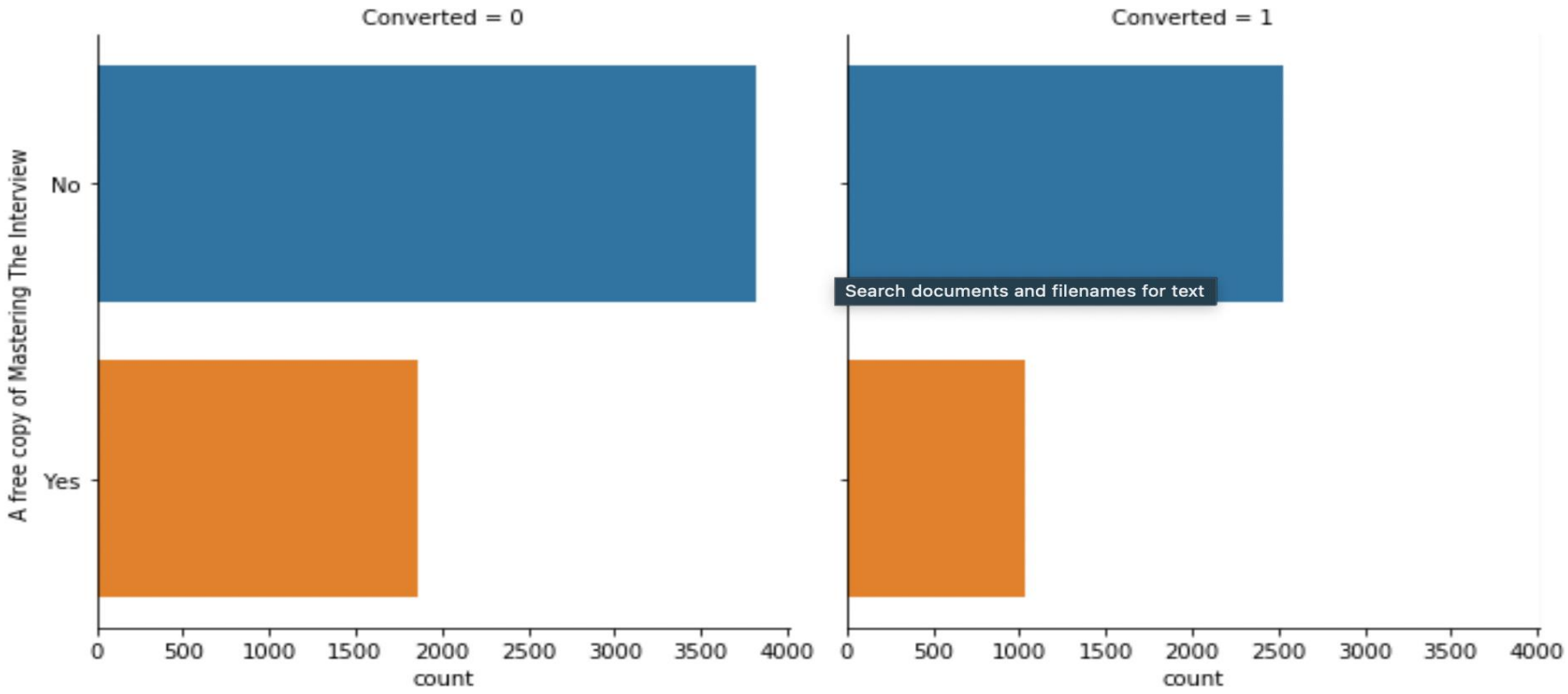




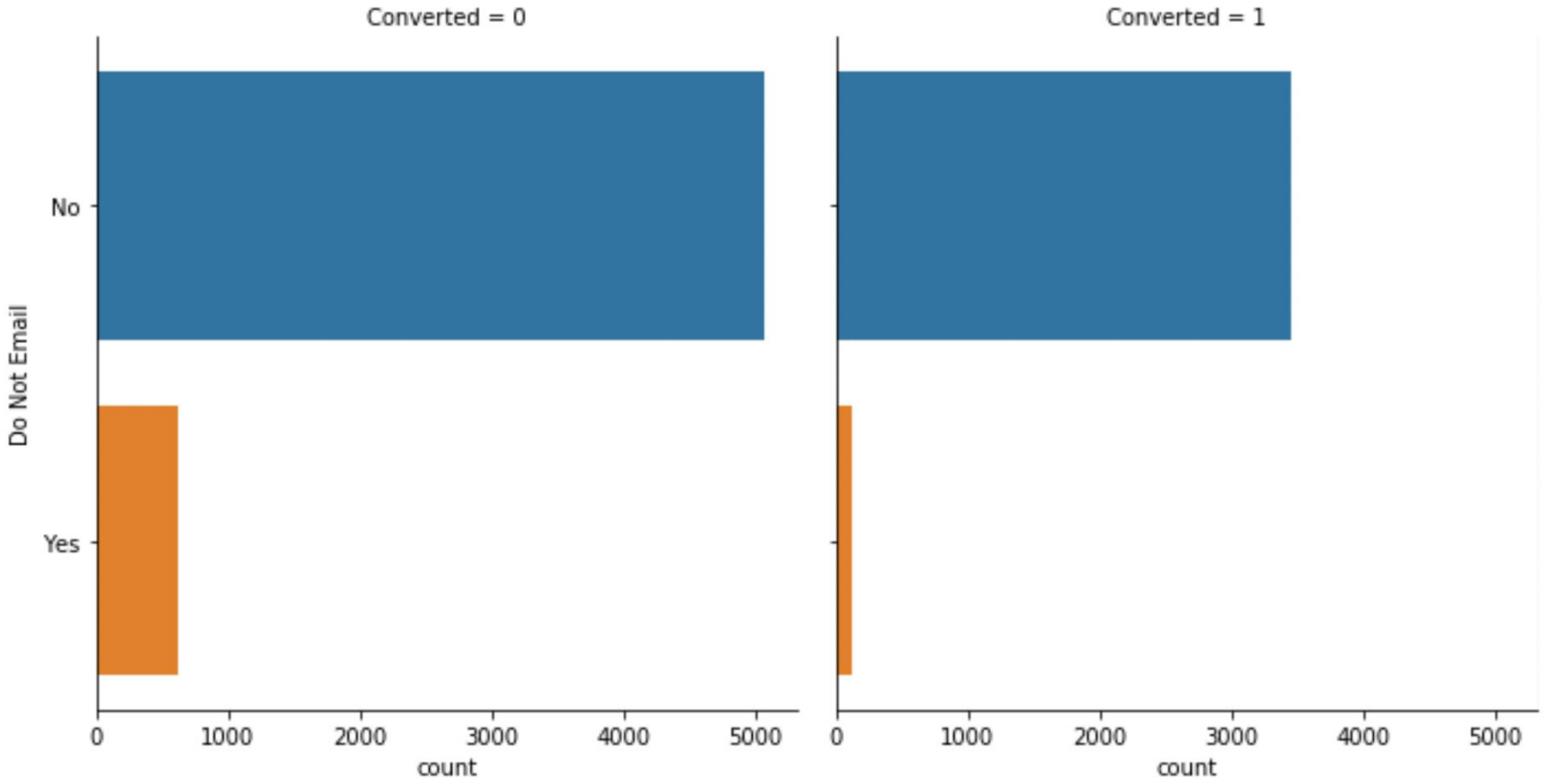
EDA plots depicting variation in numerical columns for those who converted and those who didn't



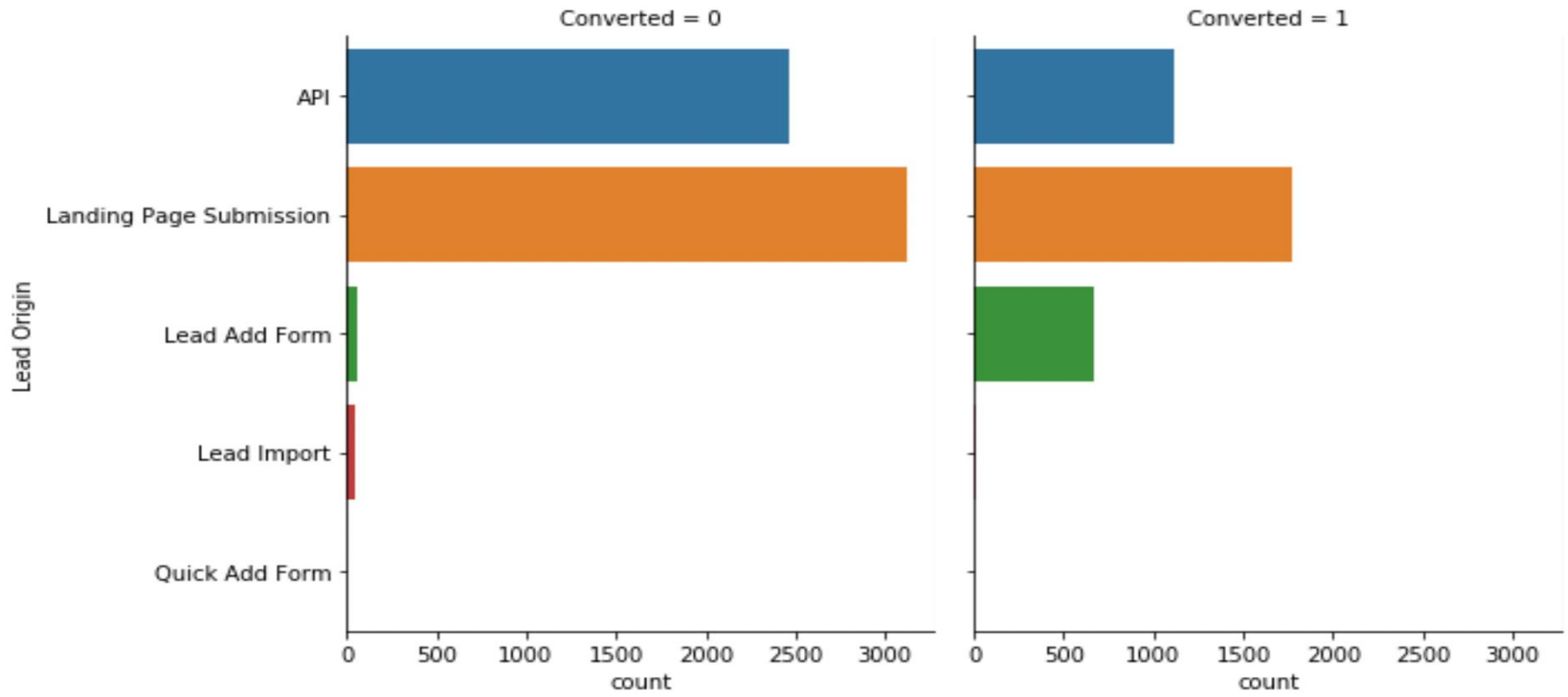
EDA plots depicting variation in categorical columns (Last Activity) for those who converted and those who didn't



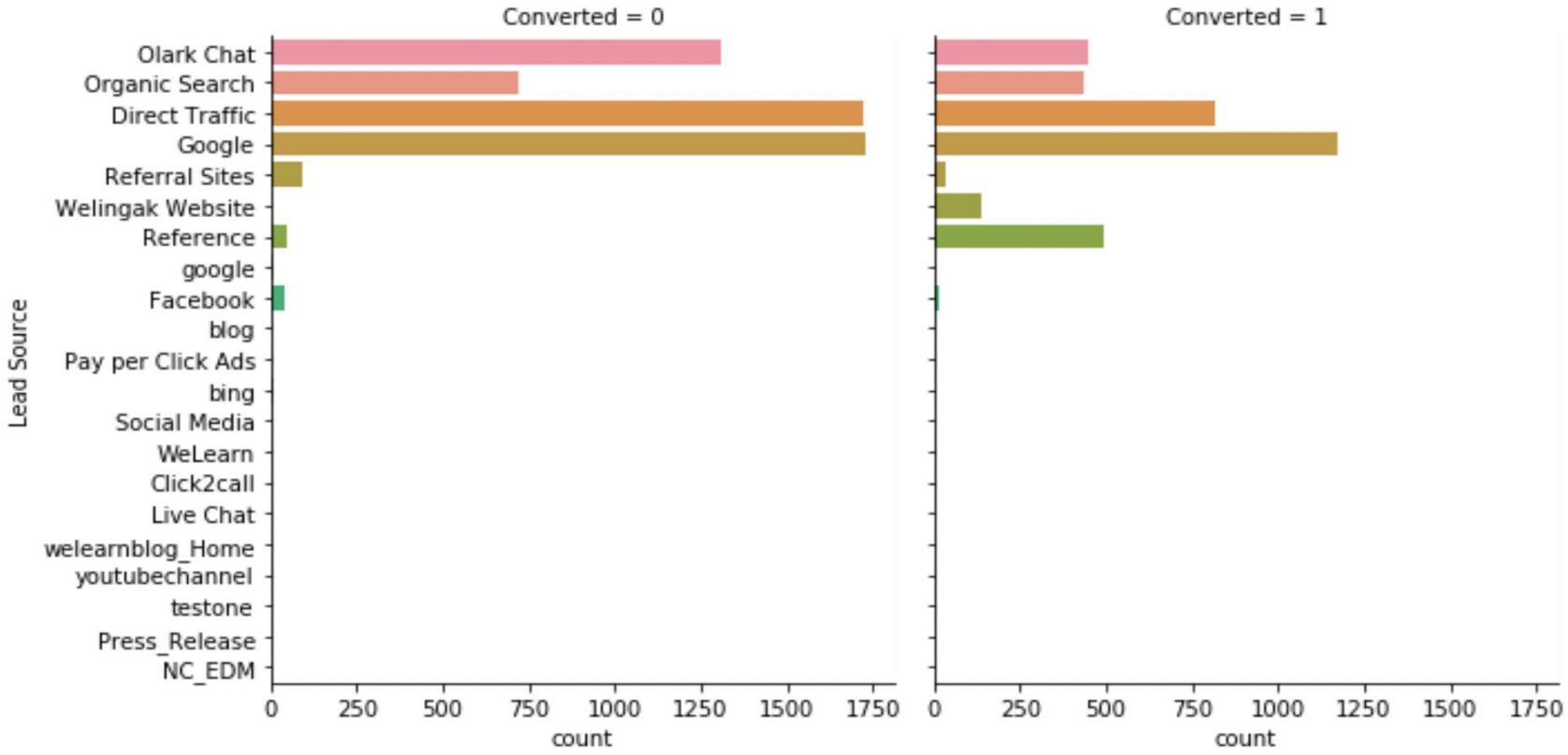
EDA plots depicting variation in categorical columns (A free copy of mastering the interview) for those who converted and those who didn't



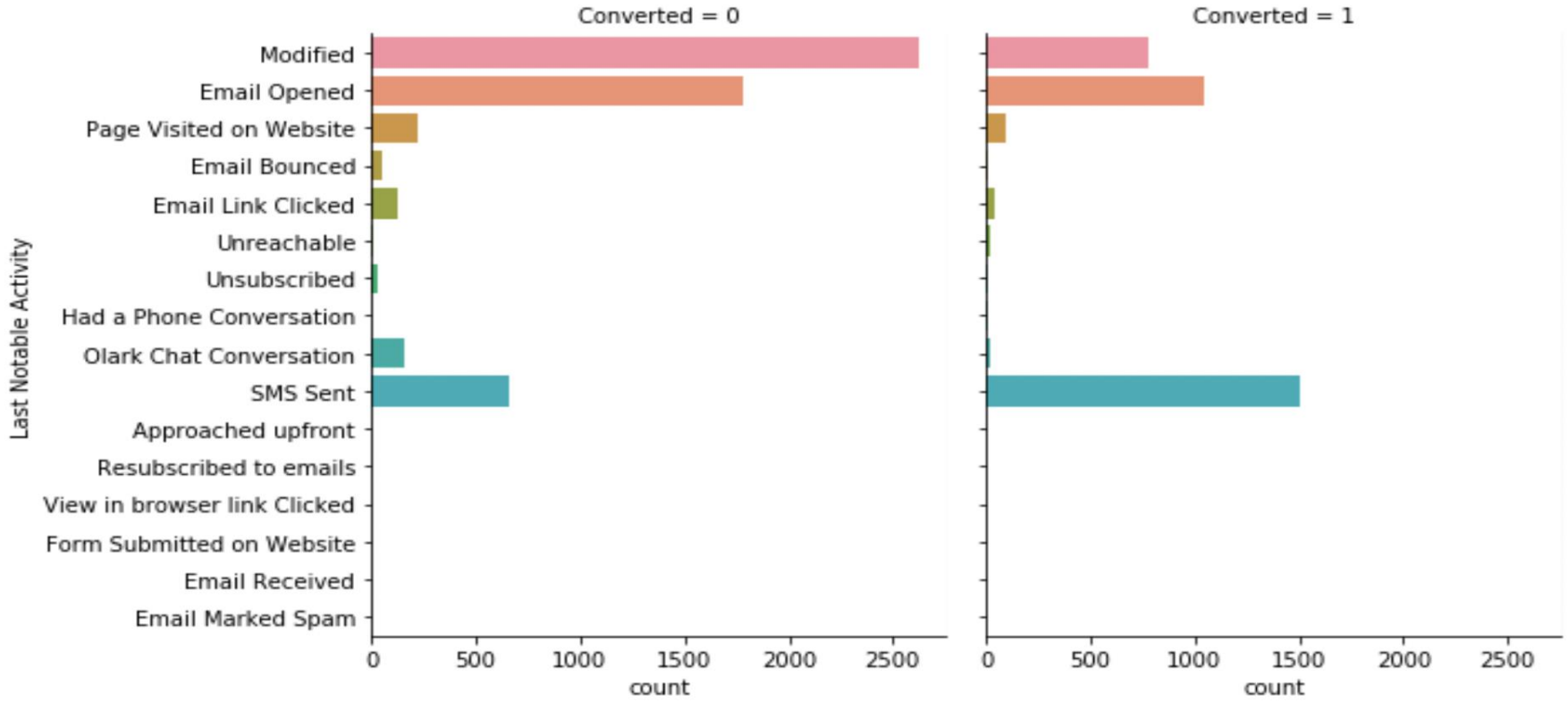
EDA plots depicting variation in categorical columns (Do not Email) for those who converted and those who didn't



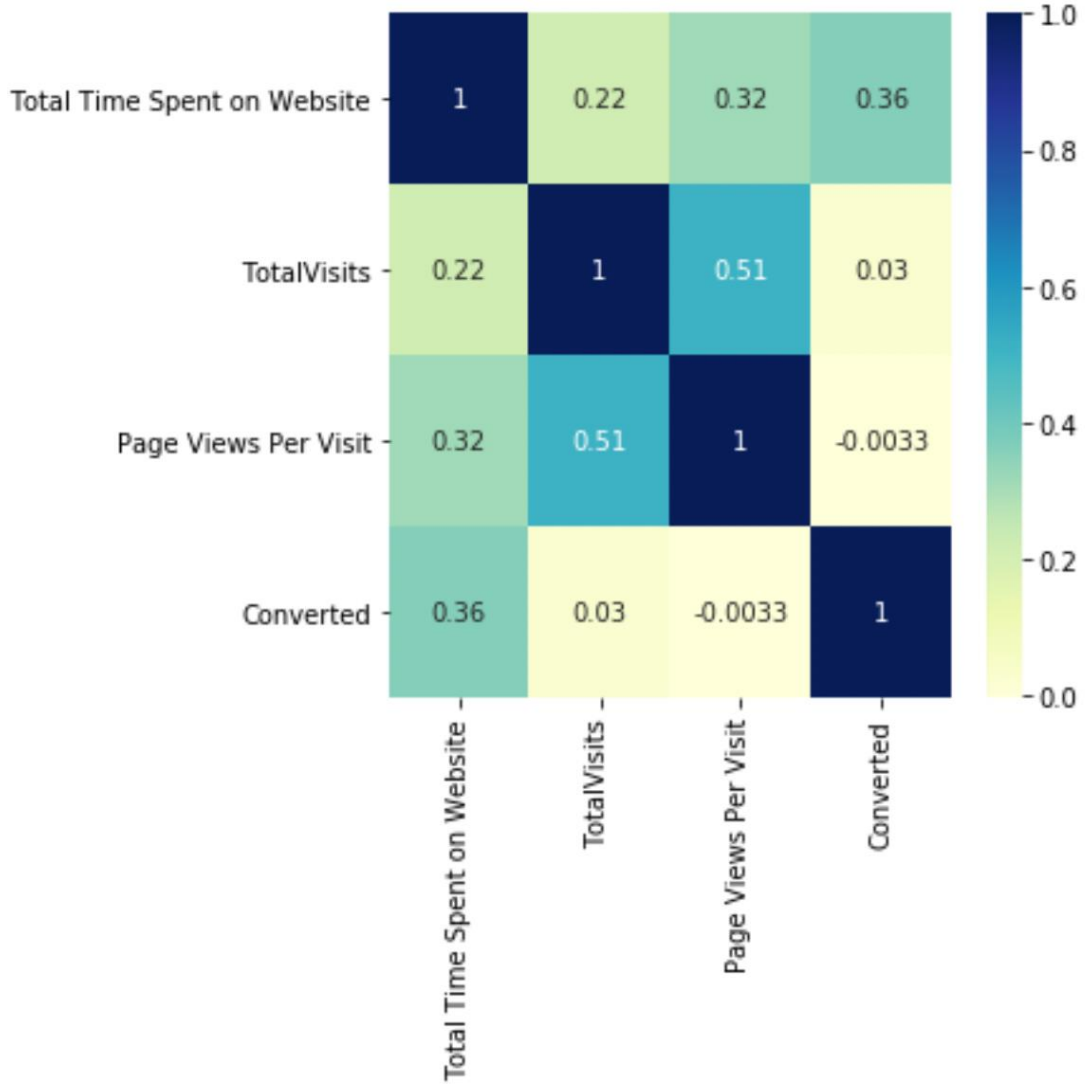
EDA plots depicting variation in categorical columns (Lead origin) for those who converted and those who didn't



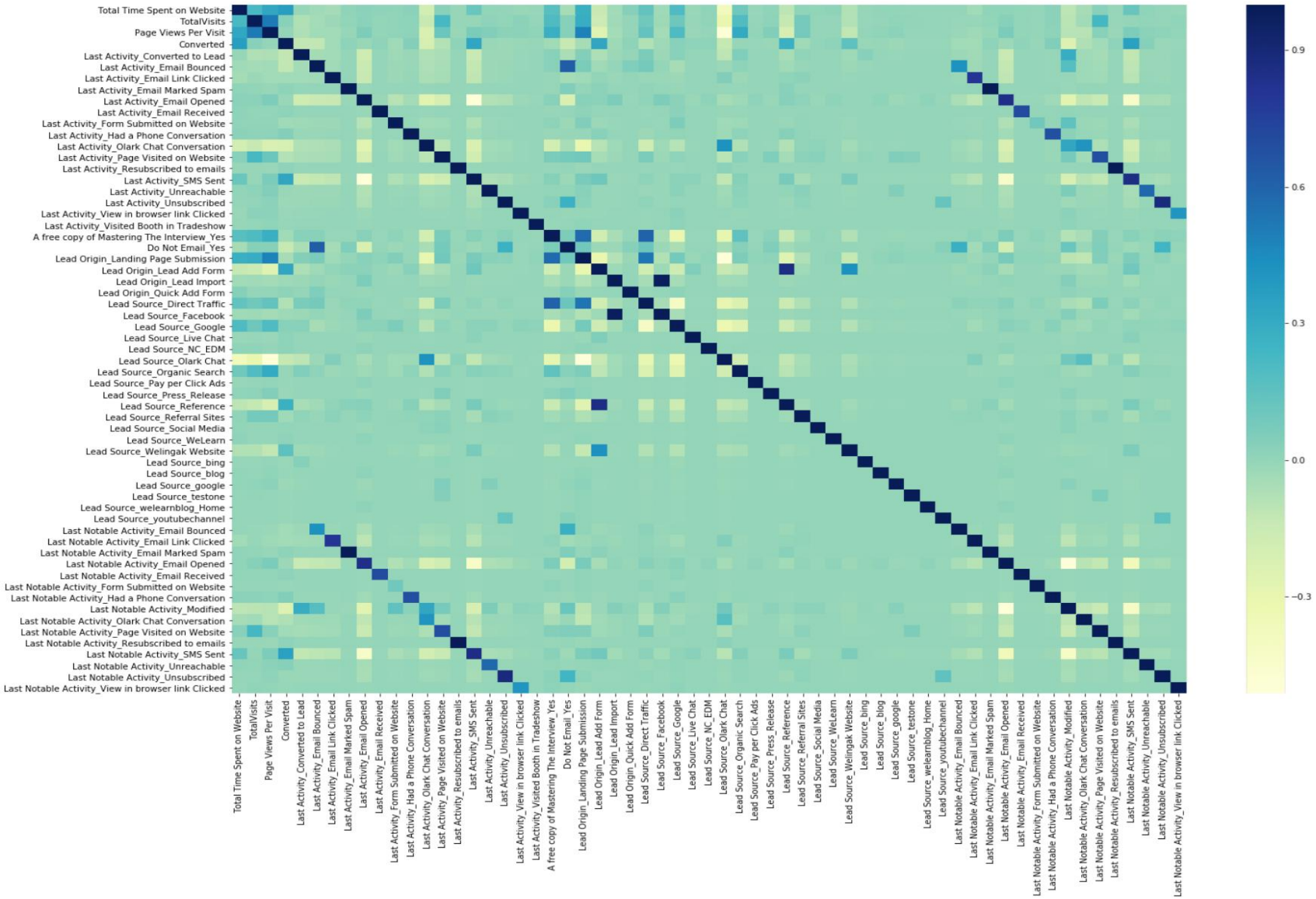
EDA plots depicting variation in categorical columns (Lead Source) for those who converted and those who didn't



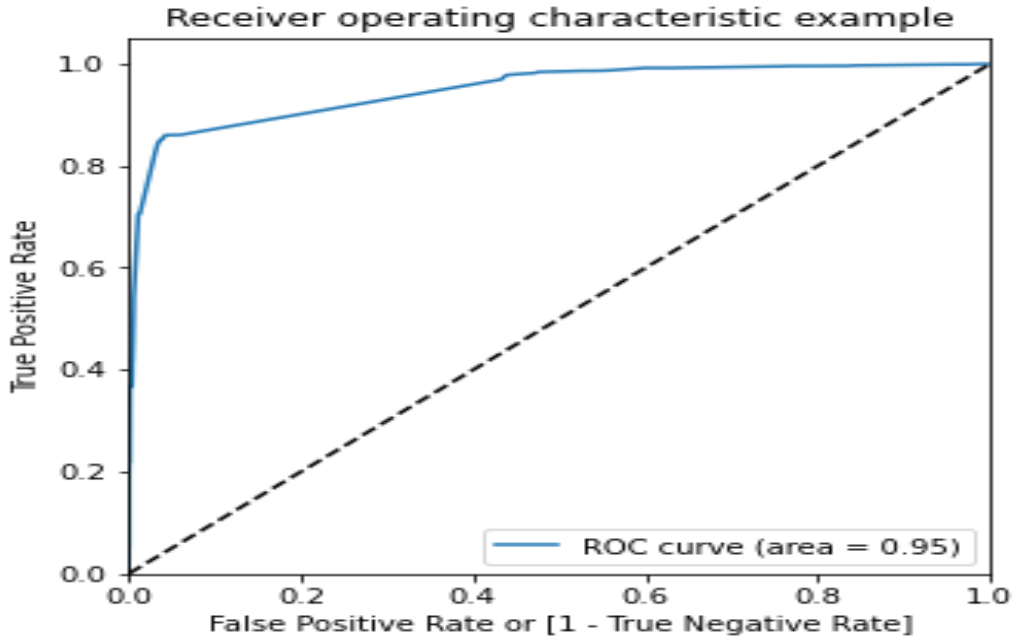
EDA plots depicting variation in categorical columns (Lead Notable Activity) for those who converted and those who didn't

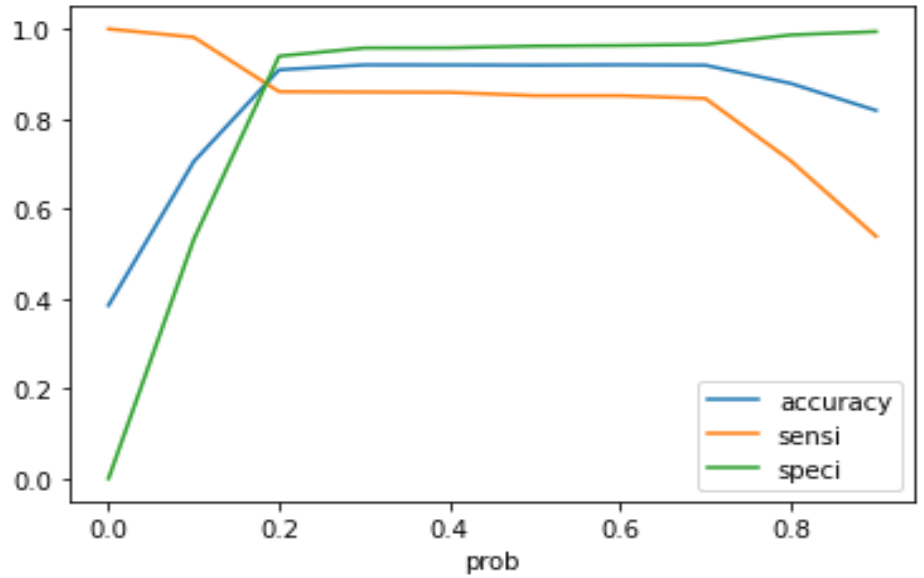


EDA plots depicting correlation (Heat Map) for all selected numerical columns

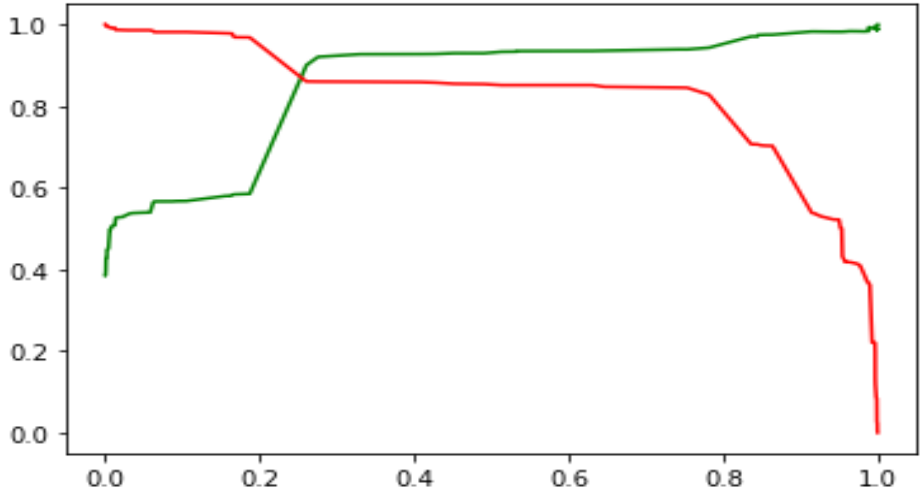


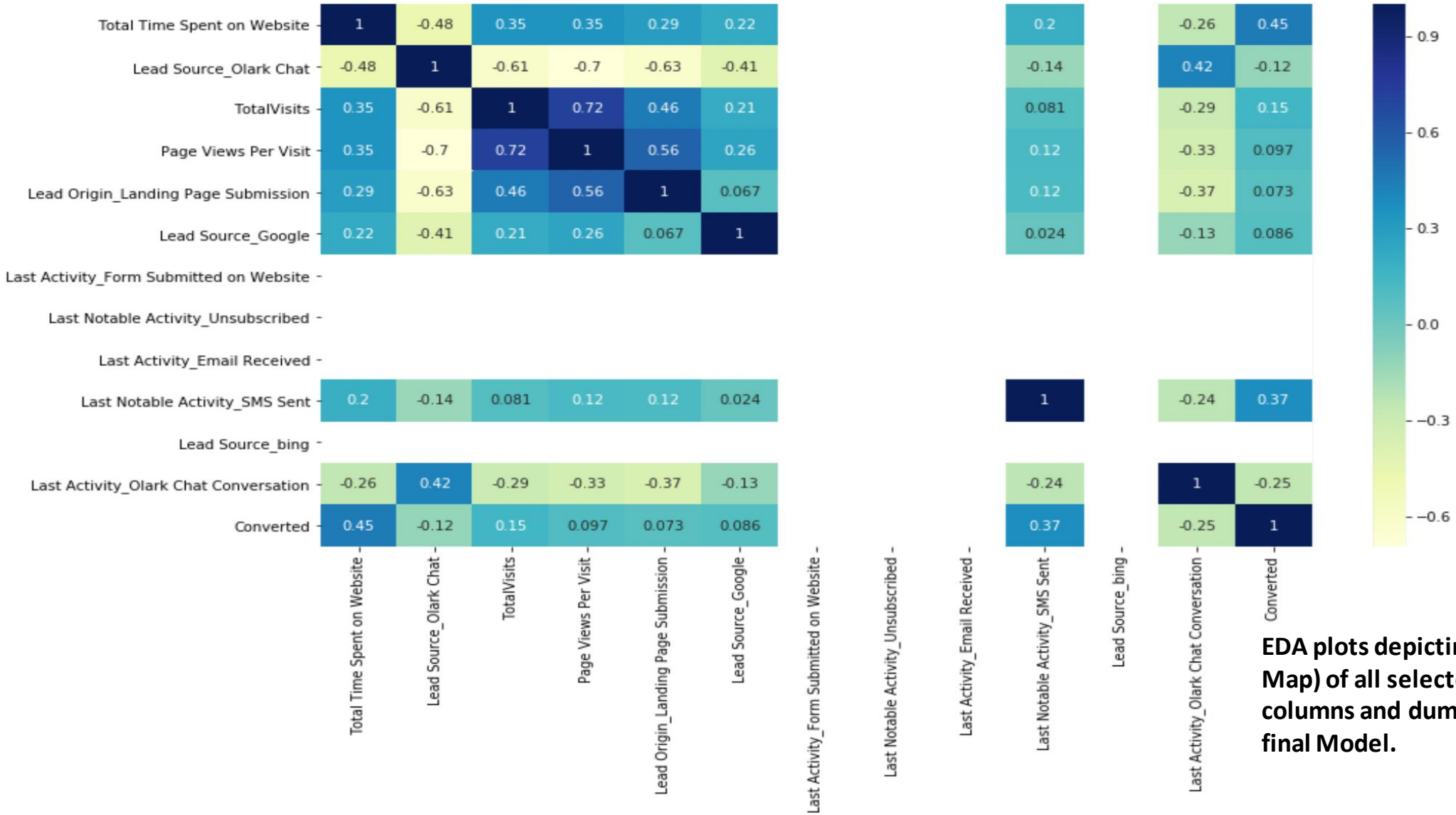
EDA plots depicting correlation (Heat Map) for all selected numerical and dummy columns





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
- Total Visits

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 Test Set, 73.3 % Sensitivity and 82.3 % 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:
- User engagement needs to be increased 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

THANK YOU