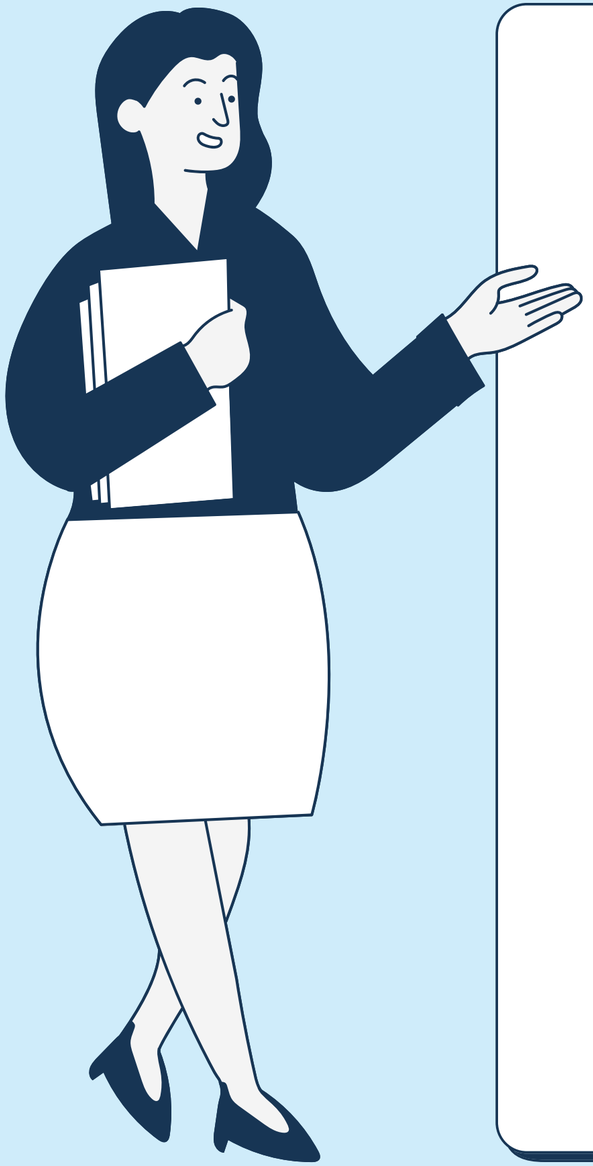
SHUJAATALI BADAMI & HARSHVARDHAN PAREEK

X Education Lead Scoring Case Study

Identification of Hot Leads in order to devote more attention to them and, as a result, increase the conversion ratio for X Education



Background

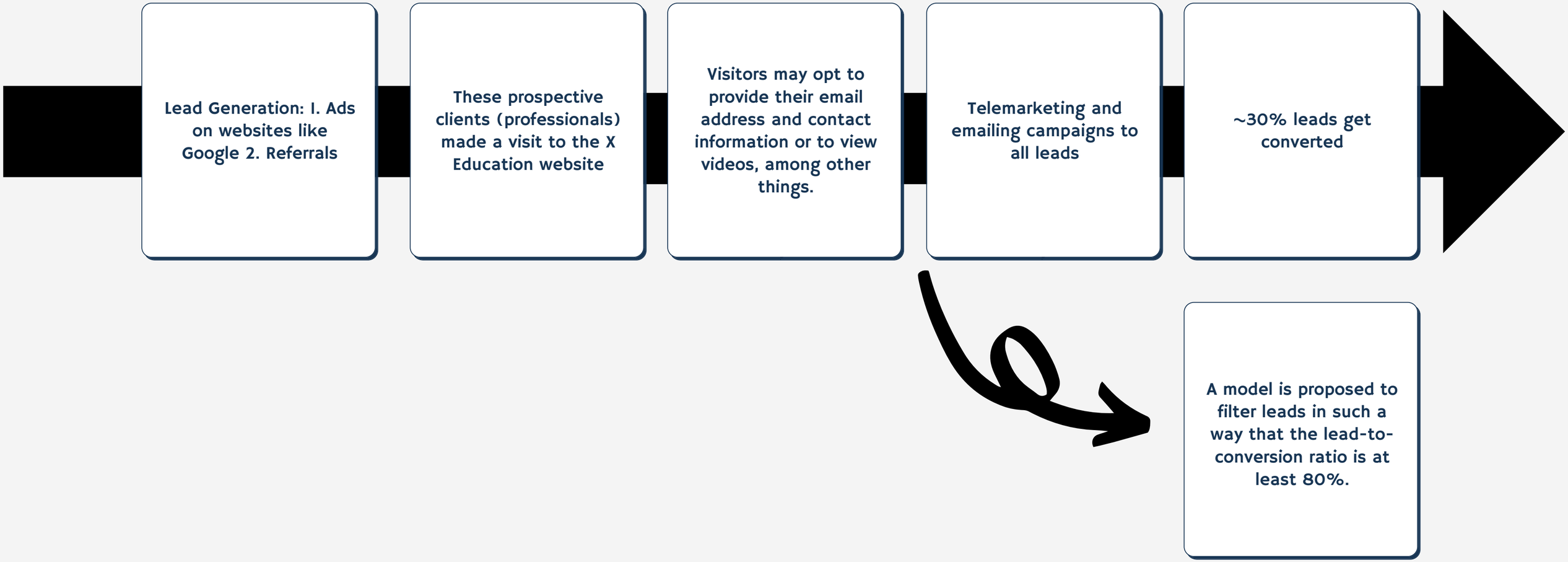
X Education is an organization that offers online courses to industry experts. Many experts that are interested in their services visit their website. The firm promotes its courses on a variety of websites, including Google. When these users arrive at the website, they may explore the courses, fill out a course registration form, or view some videos. These persons are regarded as leads when they fill out a form with their email address or phone number. Following the acquisition of these leads, members of the sales team begin making calls, sending emails, and so on. Some leads are converted as a result of this procedure, whereas the majority are not. At X schooling, the average lead conversion rate is roughly 30%.

Problem Statement

X Education receives a large number of leads, but its lead conversion rate is quite low. To improve the efficiency of this process, the organisation aims to identify the most promising leads, commonly known as 'Hot Leads.' If they are successful in identifying this group of prospects, the lead conversion rate should increase since the sales staff will now be focused more on connecting with the prospective leads rather than calling everyone. We will assist them in identifying the most promising prospects, i.e. those most likely to convert into paying clients.

We must create a model in which we give a lead score to each lead so that clients with higher lead scores have a greater conversion potential. The CEO, in particular, has said that the objective lead conversion rate should be about 80%.

Lead – Conversion Process



Proposed Solution

Selection of Hot Leads
Leads Clustering

We categorize the leads based on their proclivity or propensity to convert, resulting in a more manageable segment of hot prospects to work on.

Communicating with Hot Leads
Focus Communication

Due to the fact that we would have a smaller pool of leads to communicate with, we may be able to have a greater impact via good communication.

Conversion of Hot Leads
Increase conversion

Because we concentrated on hot leads who were more likely to convert, we would have a higher conversion rate and therefore be able to meet the objective of 80%.

Solution

The critical component of our Problem Solution is precisely identifying hot leads. The more precise the hot lead we acquire, the greater the likelihood of a higher conversion ratio. Given our goal conversion rate of 80%, we'd want to maximize our lead generation accuracy.

Implementation

Data Gathering

Loading & Observing the past data provided by the Company

Data Cleaning

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

Performing EDA

Univariate, Bivariate, and Heatmap for numerical and categorical columns

Data Preparation

Outlier Treatment, FeatureStandardization

Model Building

Performing prerequisites for RFE and Logistic Regression

Feature Selection

Selection of top 25 features using RFE

Model Building

Model building using RFE for selected columns

Model Improvement

Reduction of columns and Model re-building

Final Model

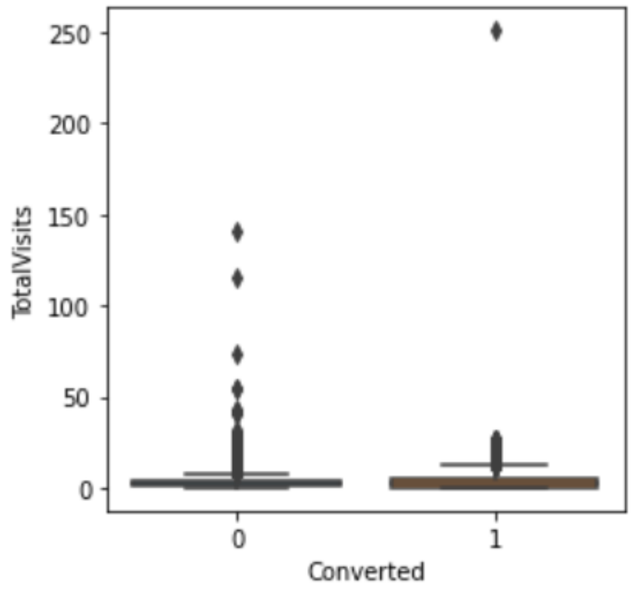
Final Model Analysis and performance on Test Data

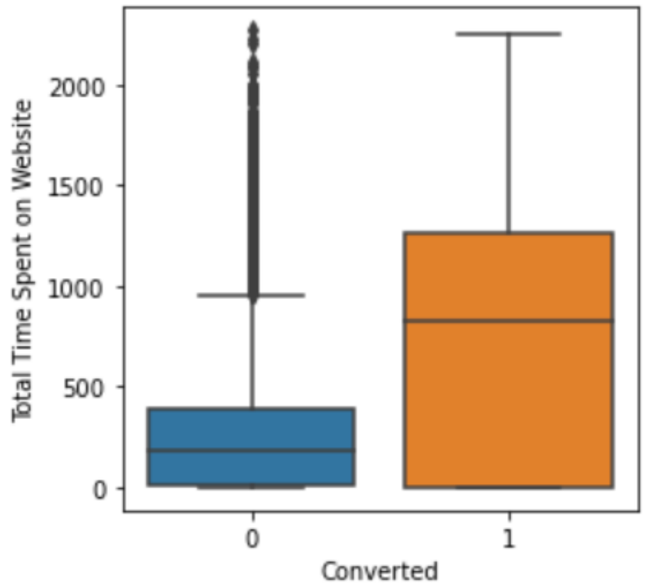
Verifying with PCA

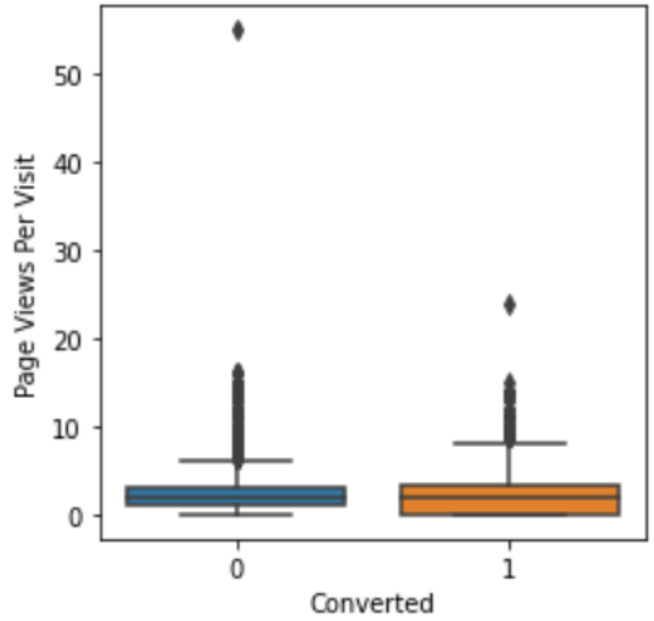
Verifying our Final Model Accuracy etc. with model built with PCA

VISUALISATION

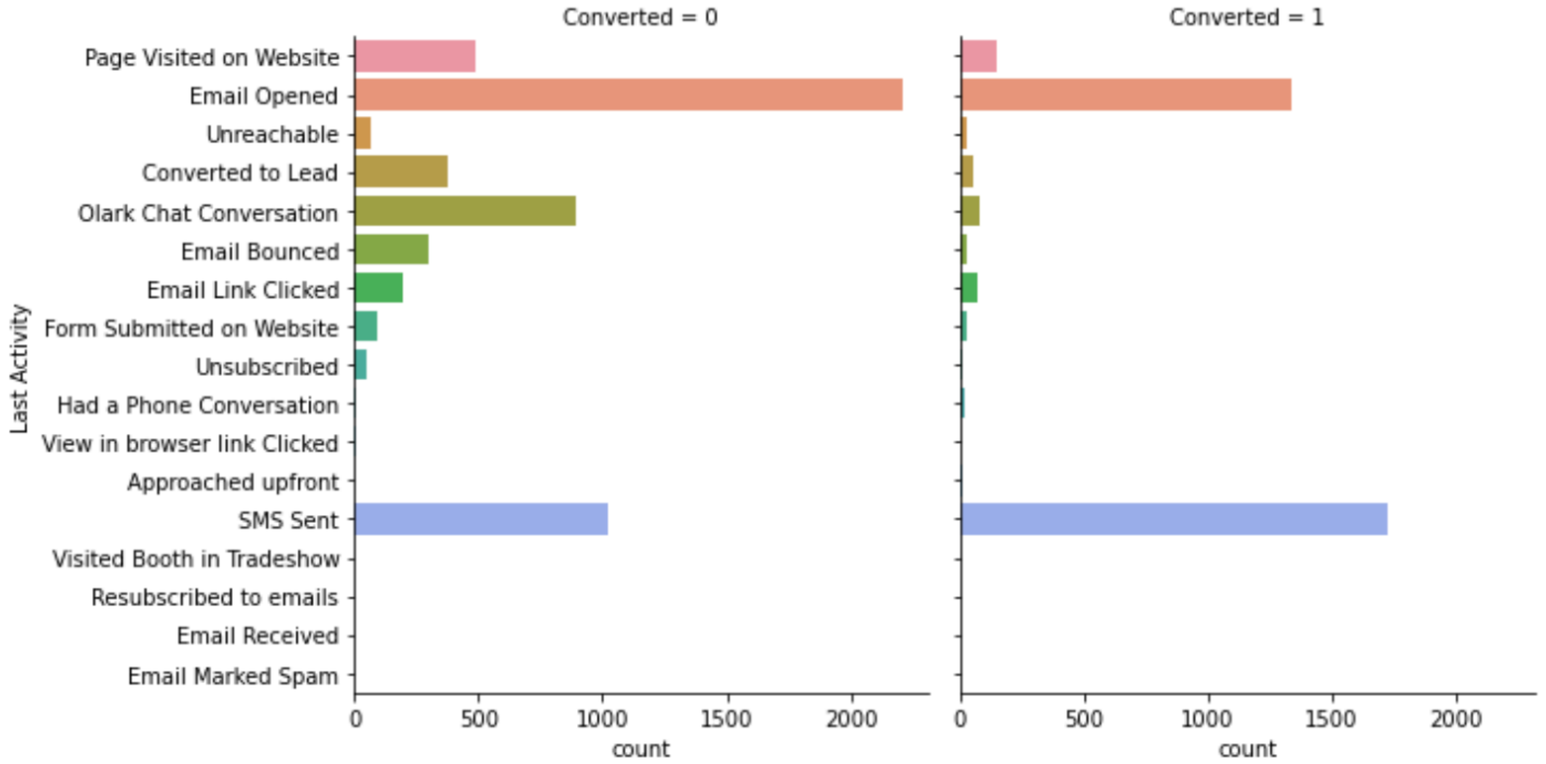
EDA charts illustrate the numerical column variance between those who converted and those who did not.



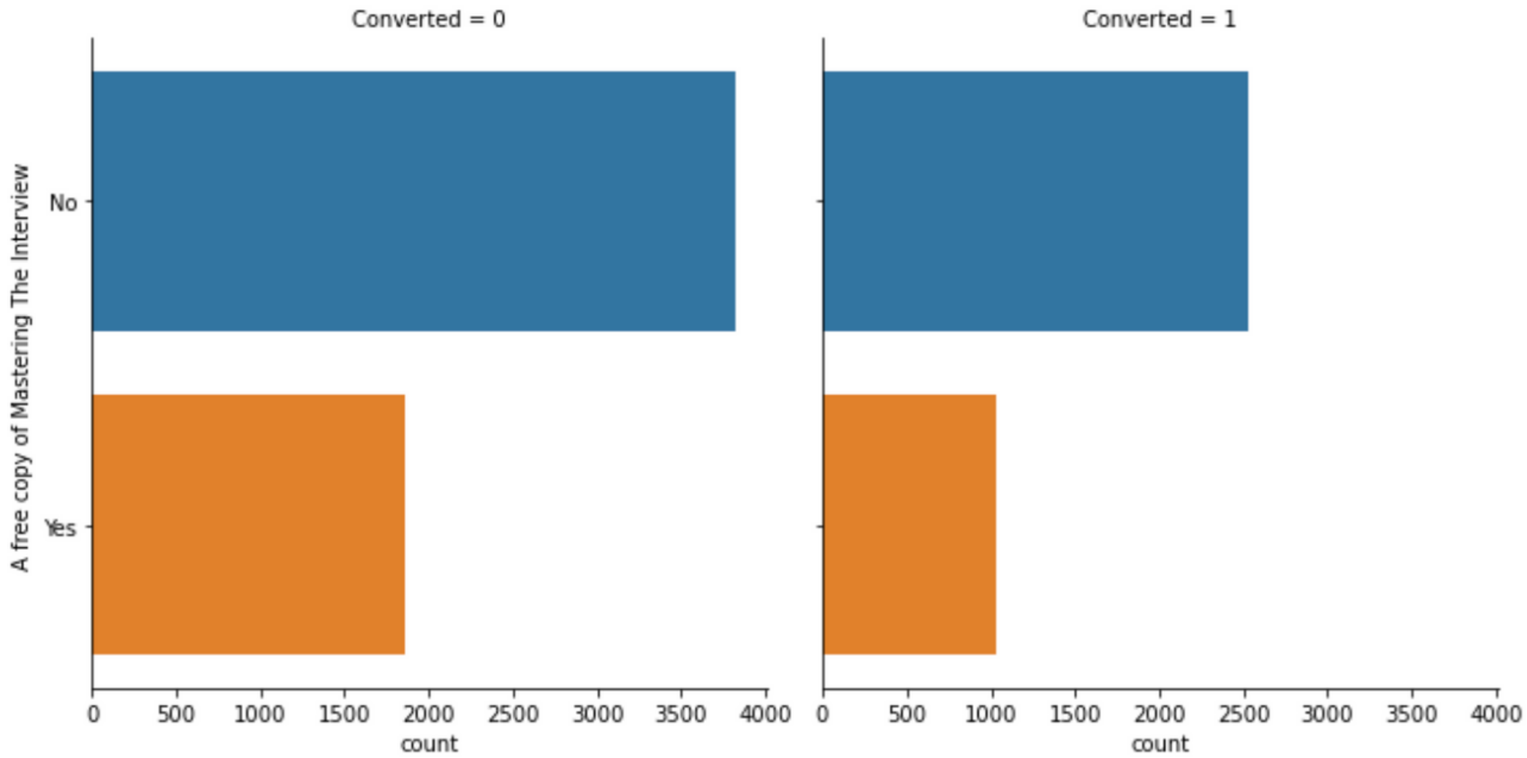




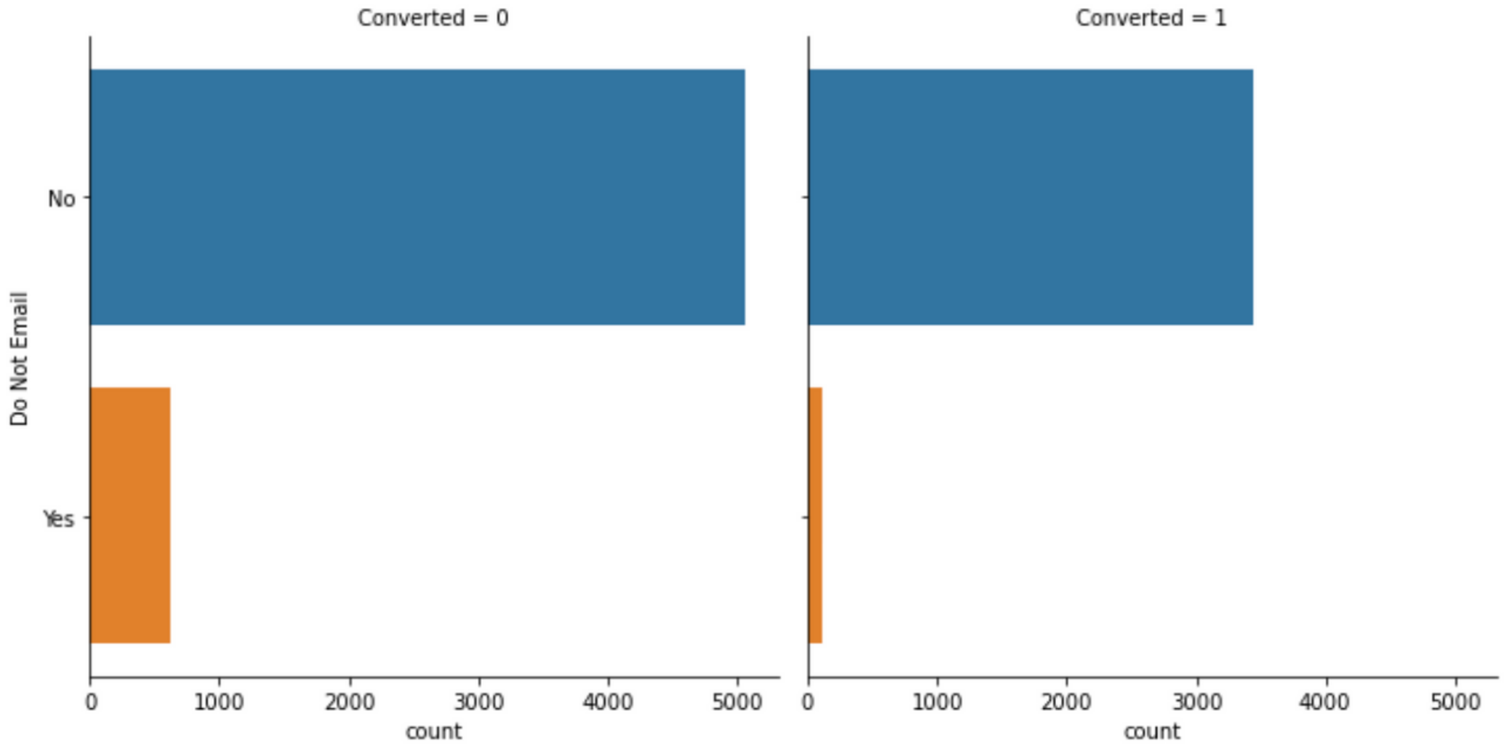
EDA graphs illustrate the difference in the categorical column (Last Activity) between those who converted and those who did not.



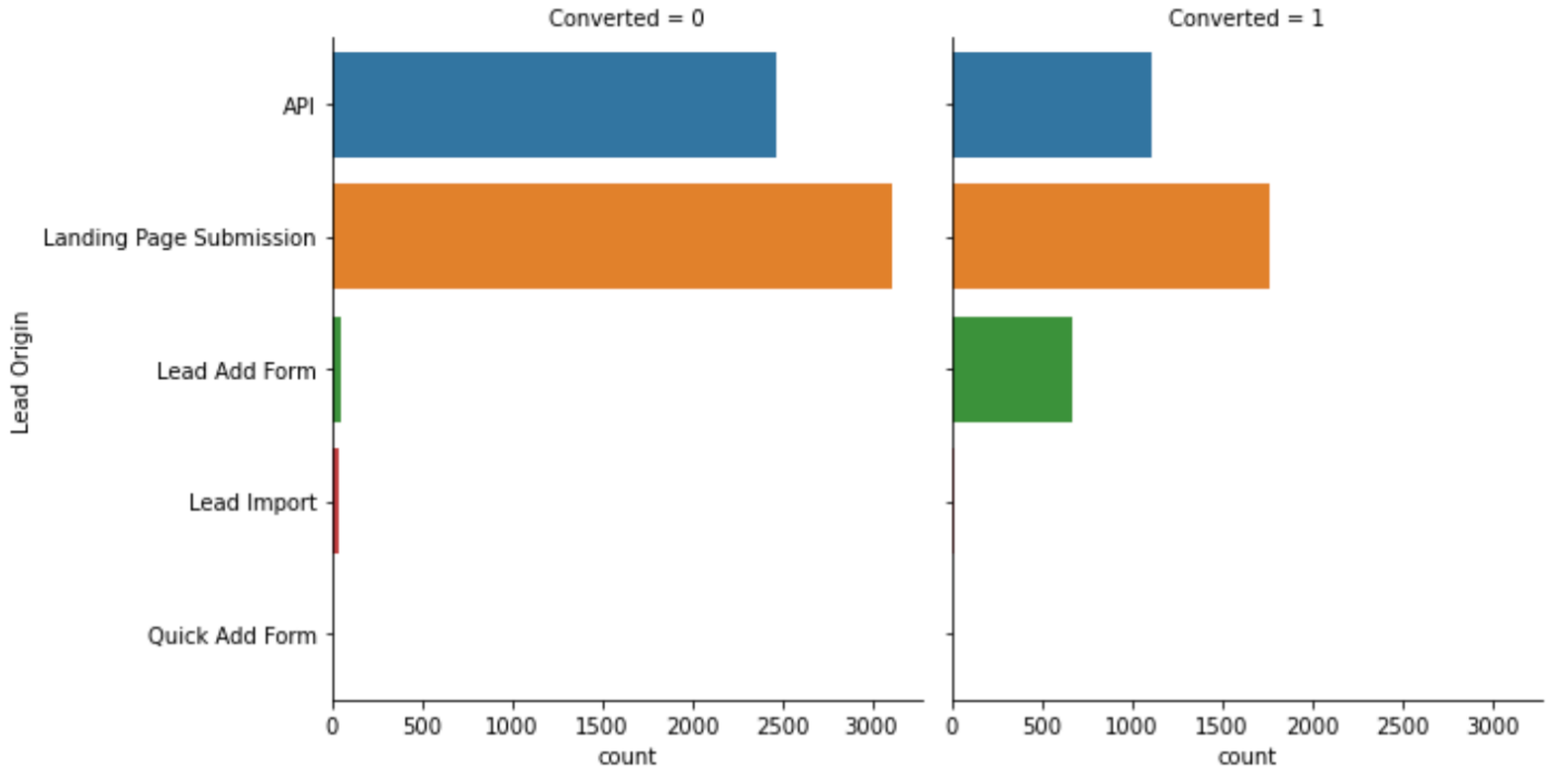
EDA plots illustrating categorical column variance (a complimentary copy of Mastering The Interview) for those who converted and those who did not.



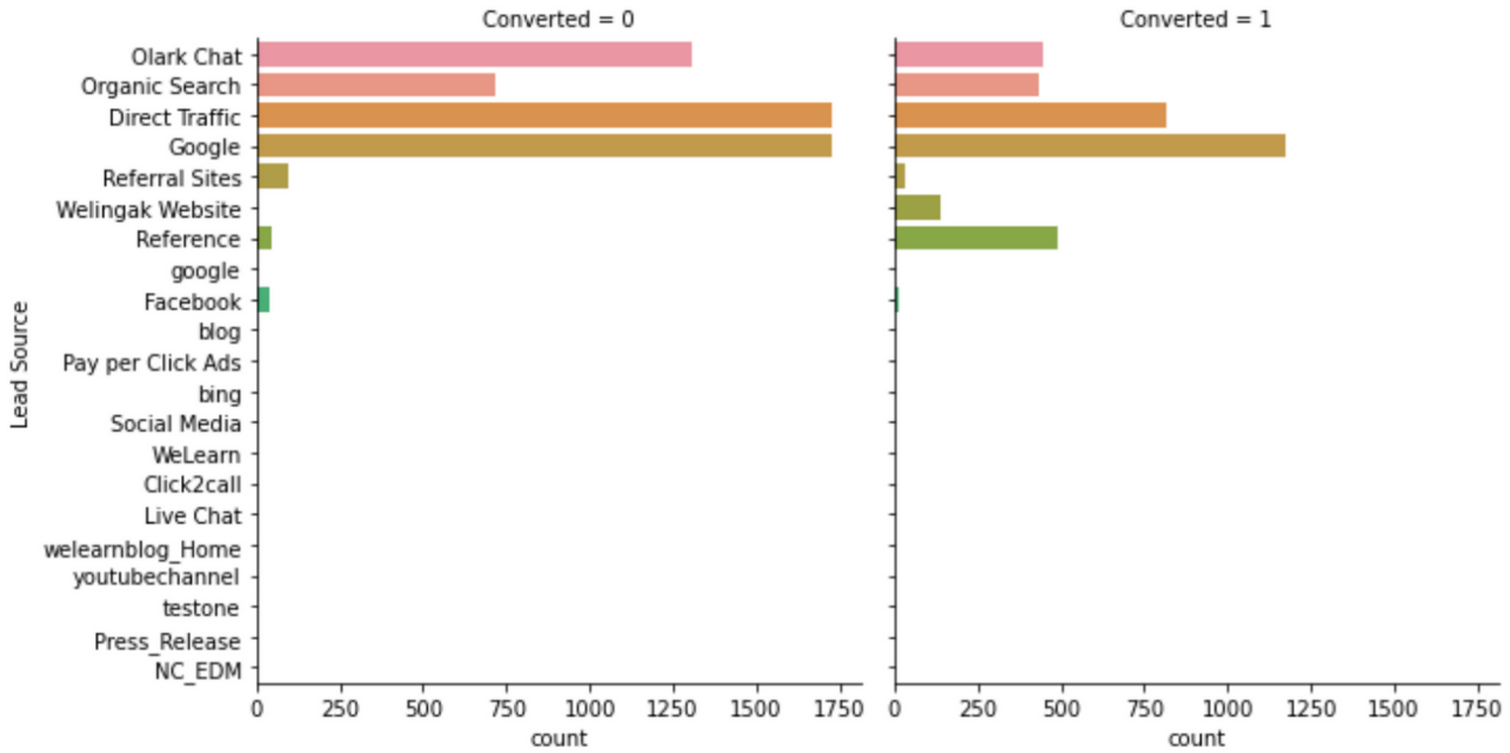
EDA graphs illustrating the difference in the categorical column (Do Not Email) between those who converted and those who did not.



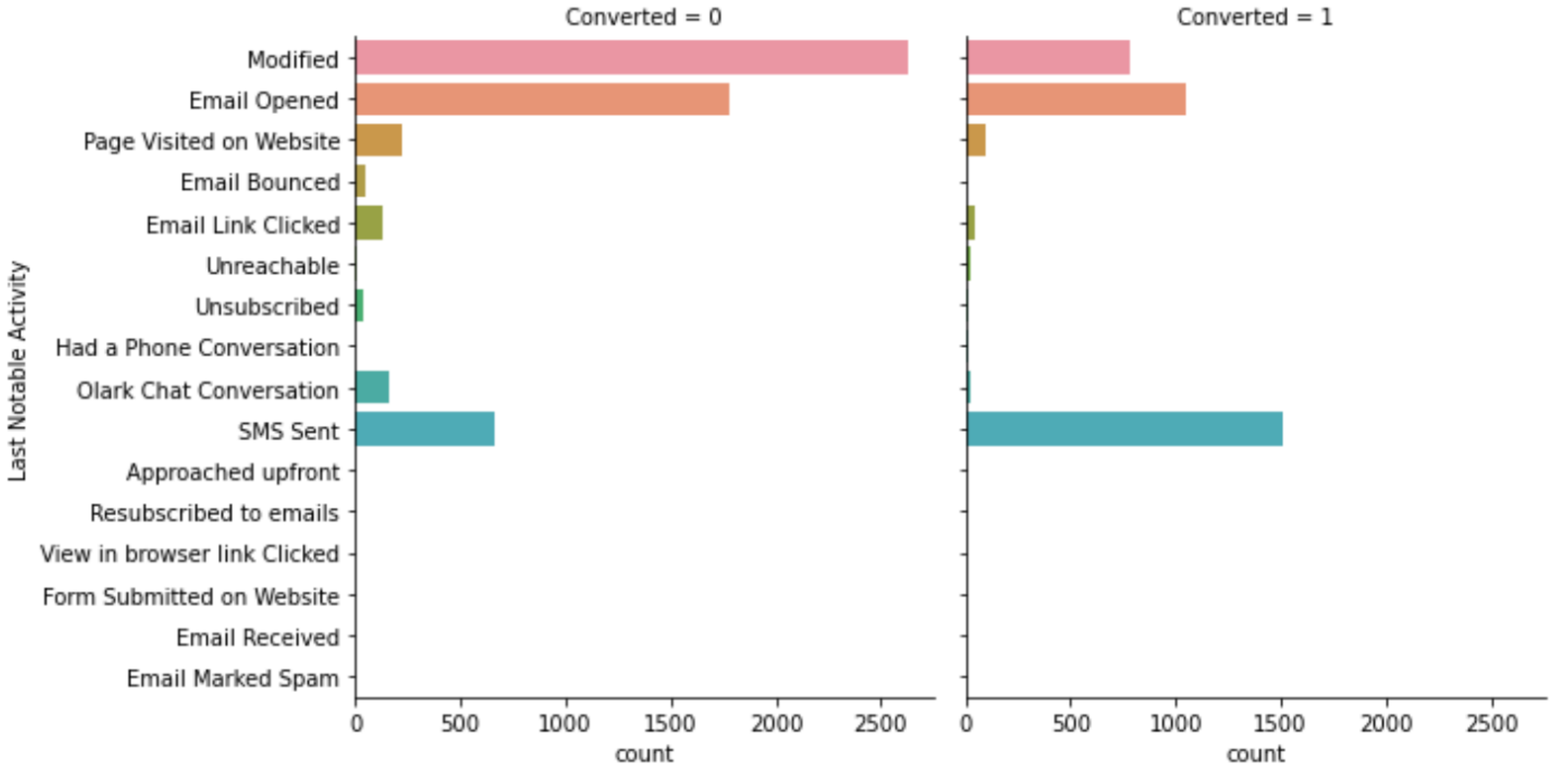
EDA graphs illustrate the difference in the categorical column (Lead Origin) between those who converted and those who did not.



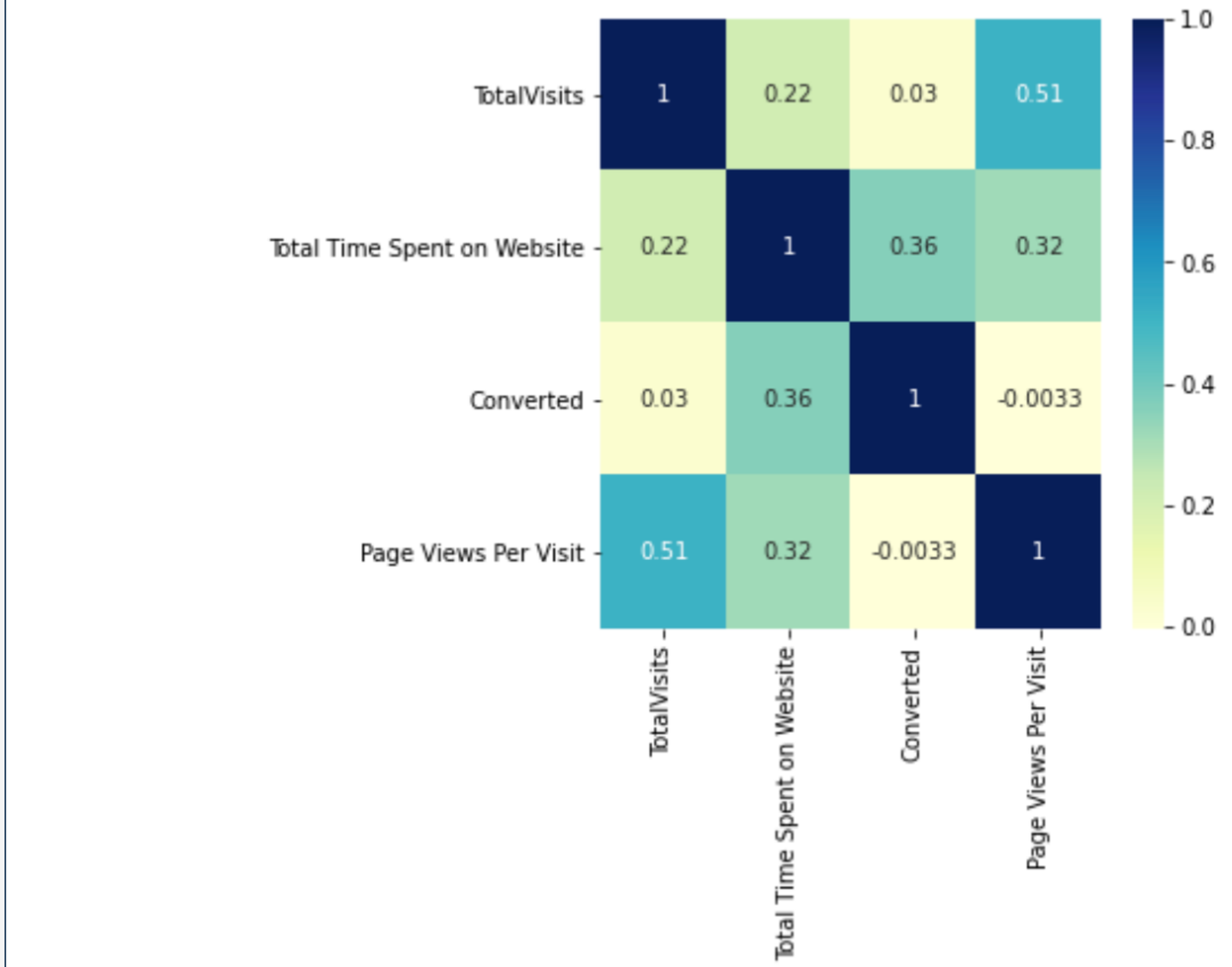
EDA graphs illustrate the difference in categorical column (Lead Source) values between those who converted and those who did not.



EDA charts illustrating the difference in the category column (Last Notable Activity) between those who converted and those who did not.

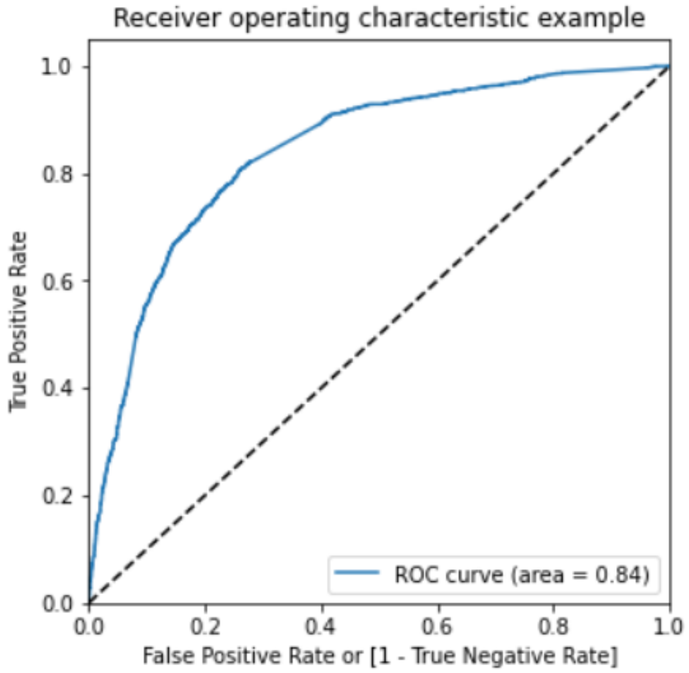


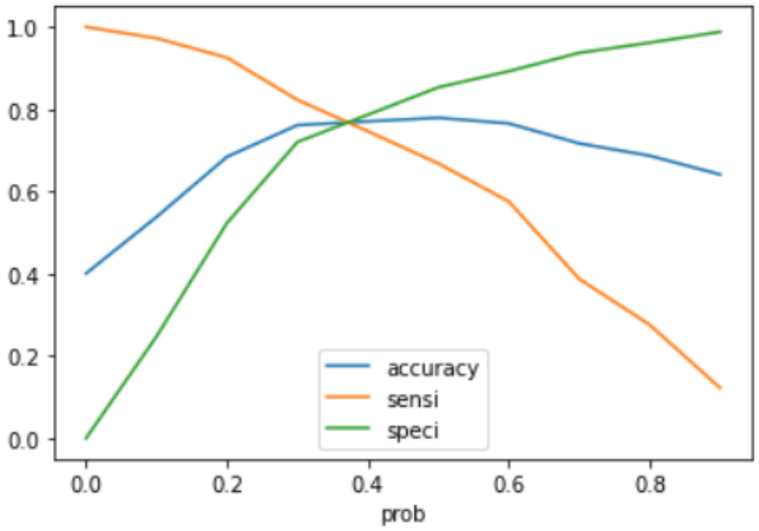
EDA graphs illustrating the correlation (Heat Map) of all numerical columns chosen.

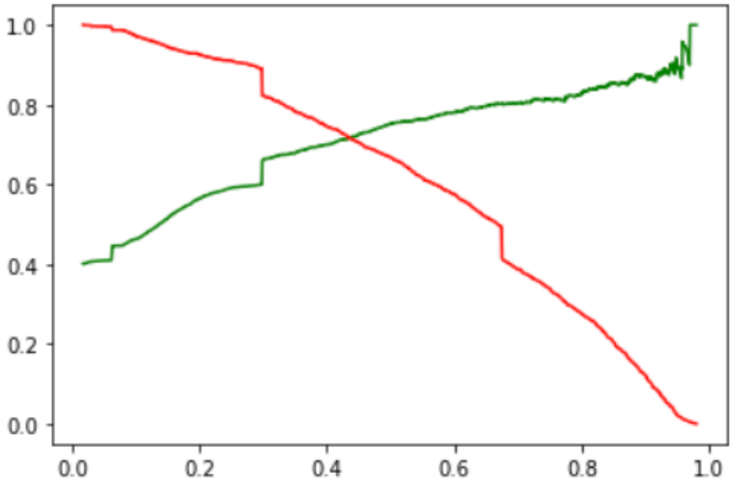


Correlation plots (Heat Maps) of all chosen columns using EDA (numerical columns and dummy columns).

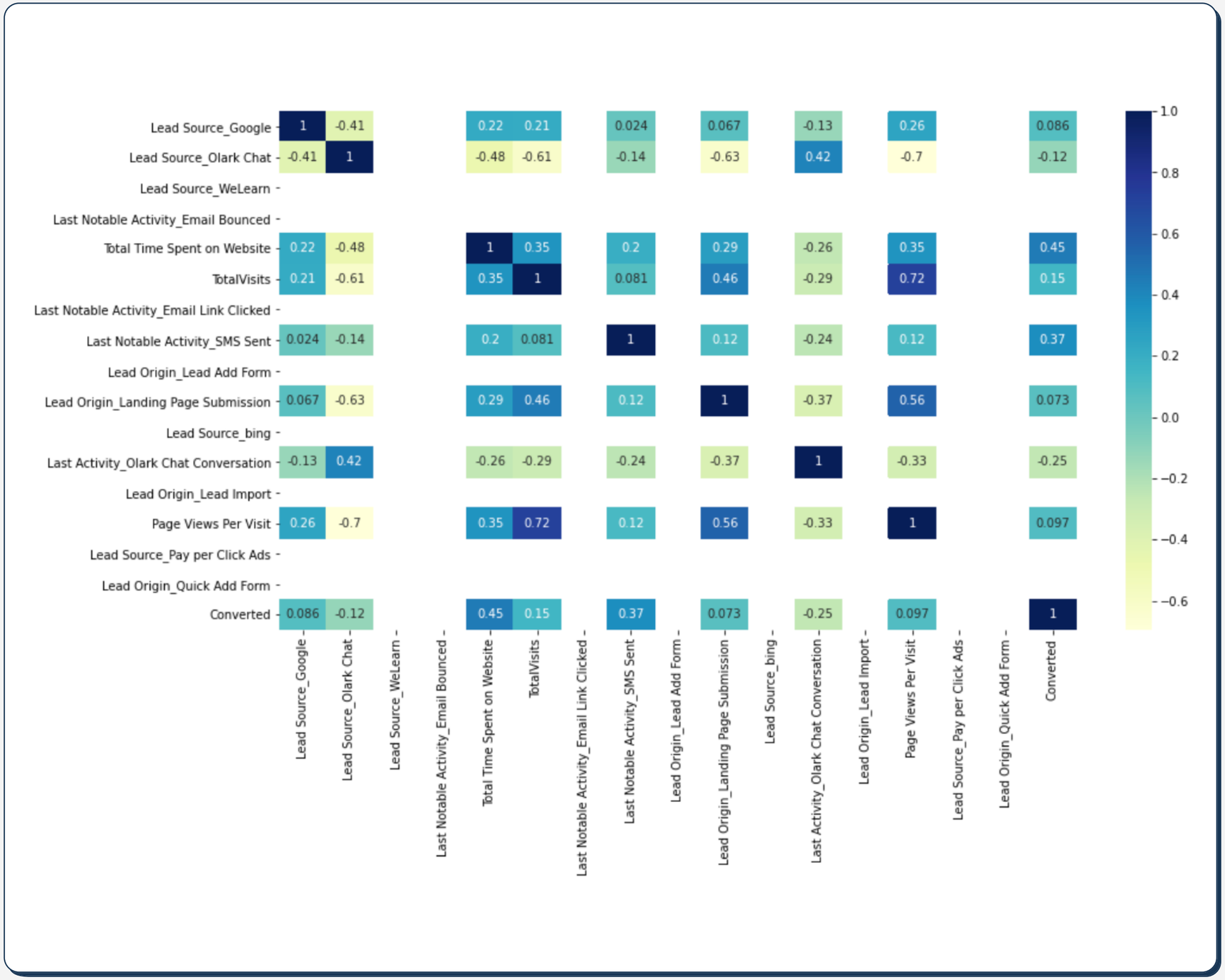
Linear Regression Final
Model Parameters Area under ROC = **0.84**
Intermediate cut-off = **0.35**
Final cut-off = **0.42**







Correlation plots (Heat Maps) of all chosen columns using EDA (numerical columns and dummy columns)



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

The top three factors in the model that contribute to lead conversion are as follows:

- Total Time Spent on Website
- Last Notable Activity_SMS Sent
- TotalVisits

The top three variables in my model on which we should concentrate our efforts are as follows:

- Last Activity_SMS Sent (positively impacting)
- Last Activity_Olark Chat Conversation (negatively impacting)
- Lead Source_Olark Chat (negatively impacting)

Conclusion (LR Model)

When compared to the model produced using PCA, our Logistic Regression Model is adequate and accurate, with 78.6 percent Accuracy on Test Set, 73.3 percent Sensitivity, and 82.3 percent Specificity. We may alter these factors by adjusting the cut-off value and so forecast Hot leads in response to circumstances such as the availability of more resources and vice versa.

Conclusion (Recommendation)

X Education Company must prioritize the following critical areas in order to increase overall conversion rates: Enhance user engagement on their website, since this results in more conversions. Increase the frequency of SMS alerts, as this contributes to greater conversion; Increase total visits via advertising, etc., as this contributes to higher conversion; and Improve the Olark Chat service, as this contributes to lower conversion.

Thank you!

Shujaatali Badami & Harshvardhan Pareek