Lead Scoring Case Study

Identification of Hot Leads so that X Education can focus more on them and thus enhancing the conversion ratio

- Kshitiz Thakur

- Kalpesh Kakulite

Background
X Education Company

- X Education , sells online courses to industry professionals
- Interested professionals land on their website to buy online courses
- The company markets its courses on several websites and search engines like Google. People on their website, they might browse the courses or fill up a form for the course or watch some videos
- When those people fill up the 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%

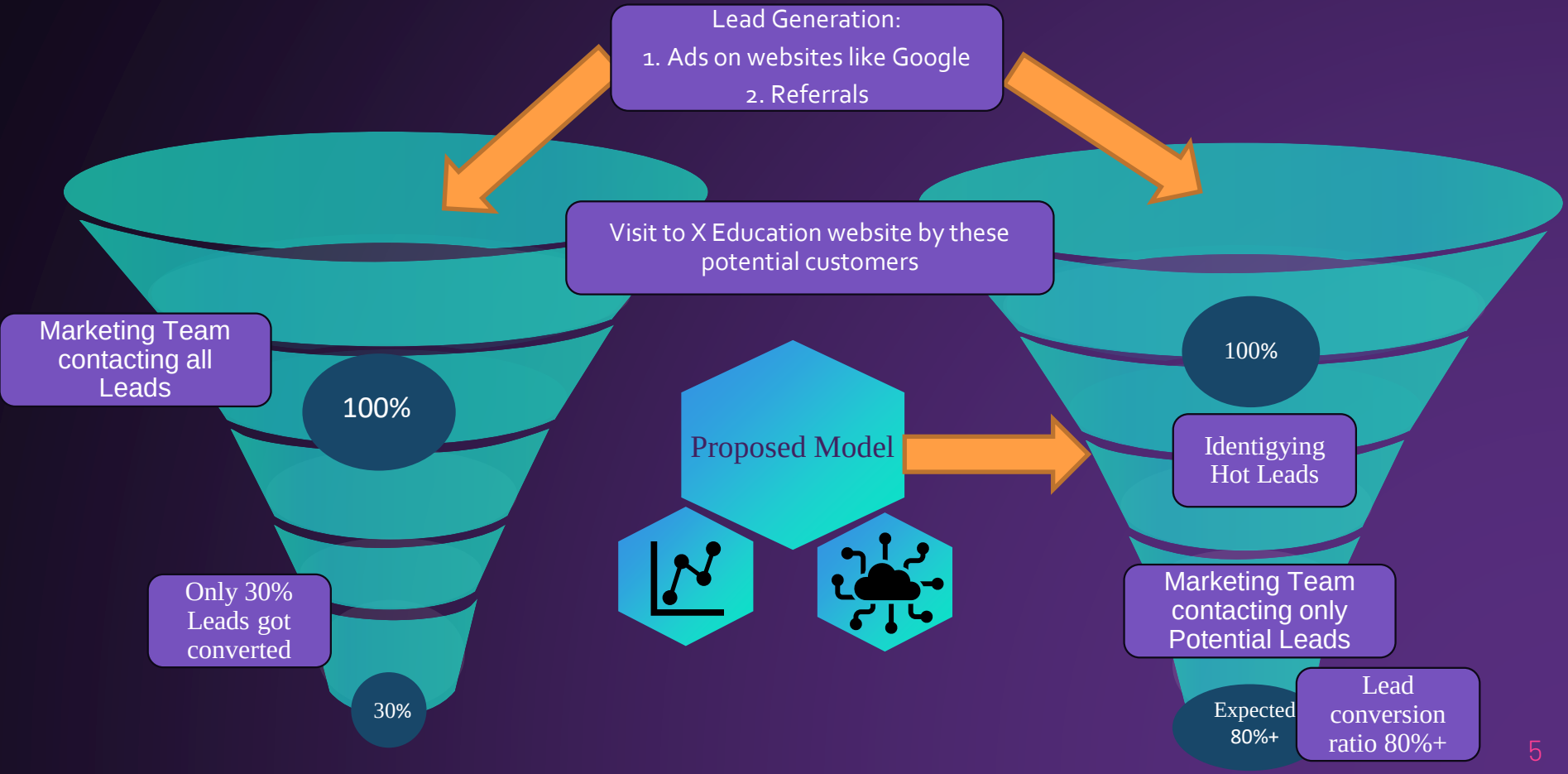
Problem Statement
X Education Company's Problem

- X Education gets a lot of leads but its lead conversion rate is very poor
- For more efficiency, 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
- So that sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

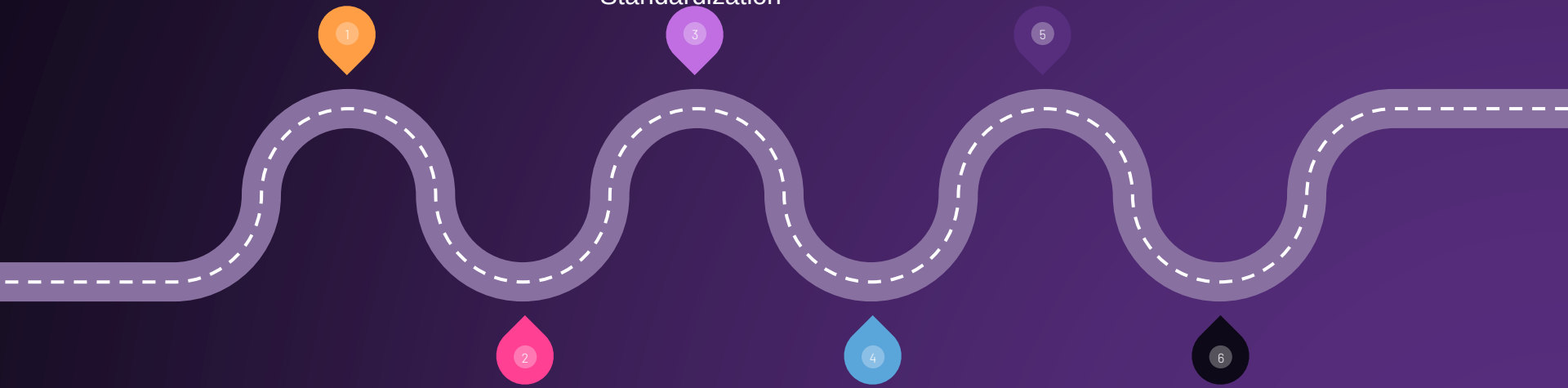
Problem Statement
X Education Company's Problem

- We will help them to select the most promising leads
- 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, has given a ballpark of the target lead conversion rate to be 80%.

Lead – Conversion Process



ROADMAP



Data Gathering

Loading & Observing the past data provided by the Company

Data Cleaning

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

Data Preparation

Outlier Treatment, Feature-Standardization

Performing EDA

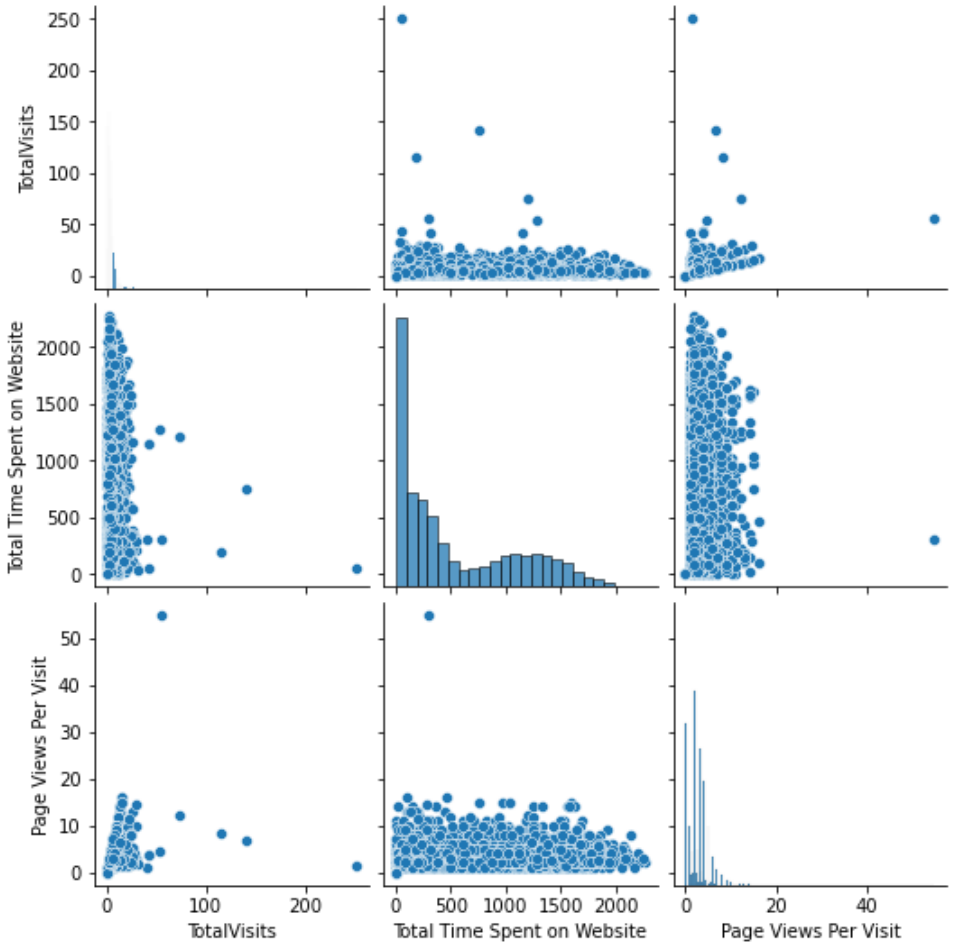
Model Building

Performing pre-requisites for RFE and Logistic Regression

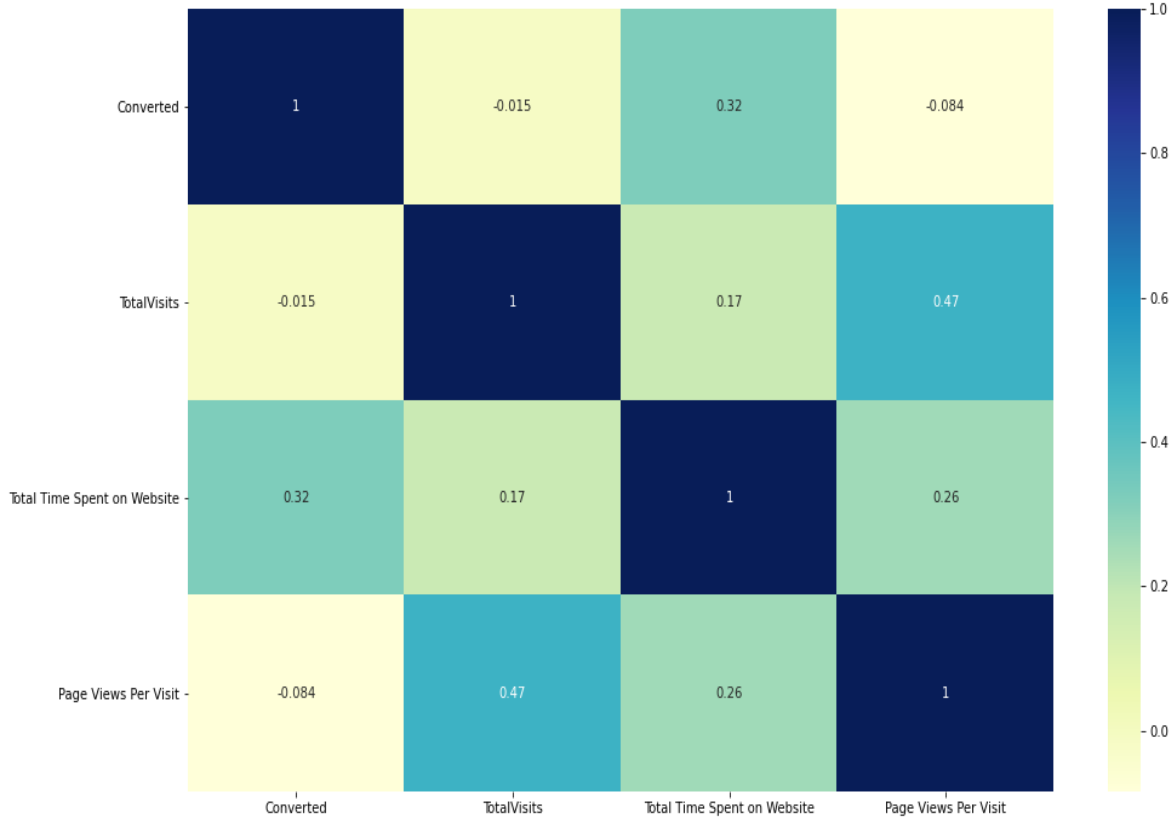
Remodeling and Finalizing Model

Selection and Reduction of important features using RFE

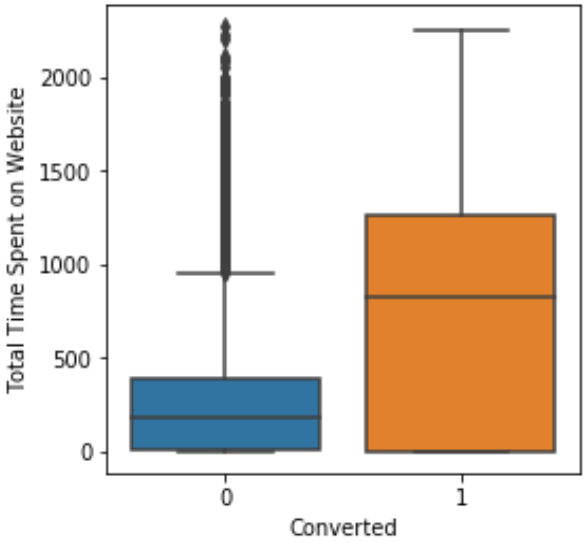
Data Visualization

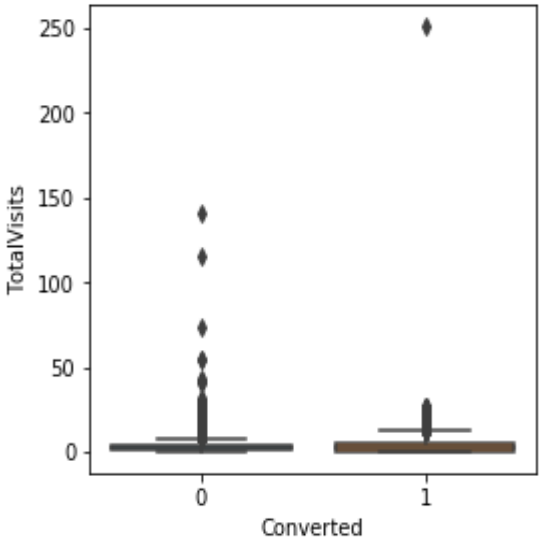


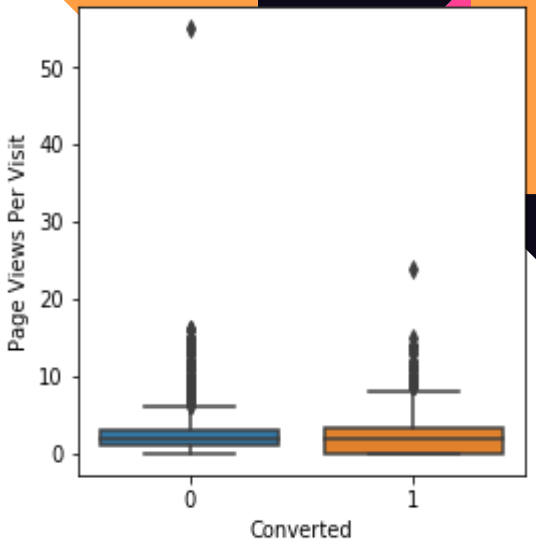
- There are a few extreme outliers present in the Variables
- Out of Total Visits, most people spent time on website



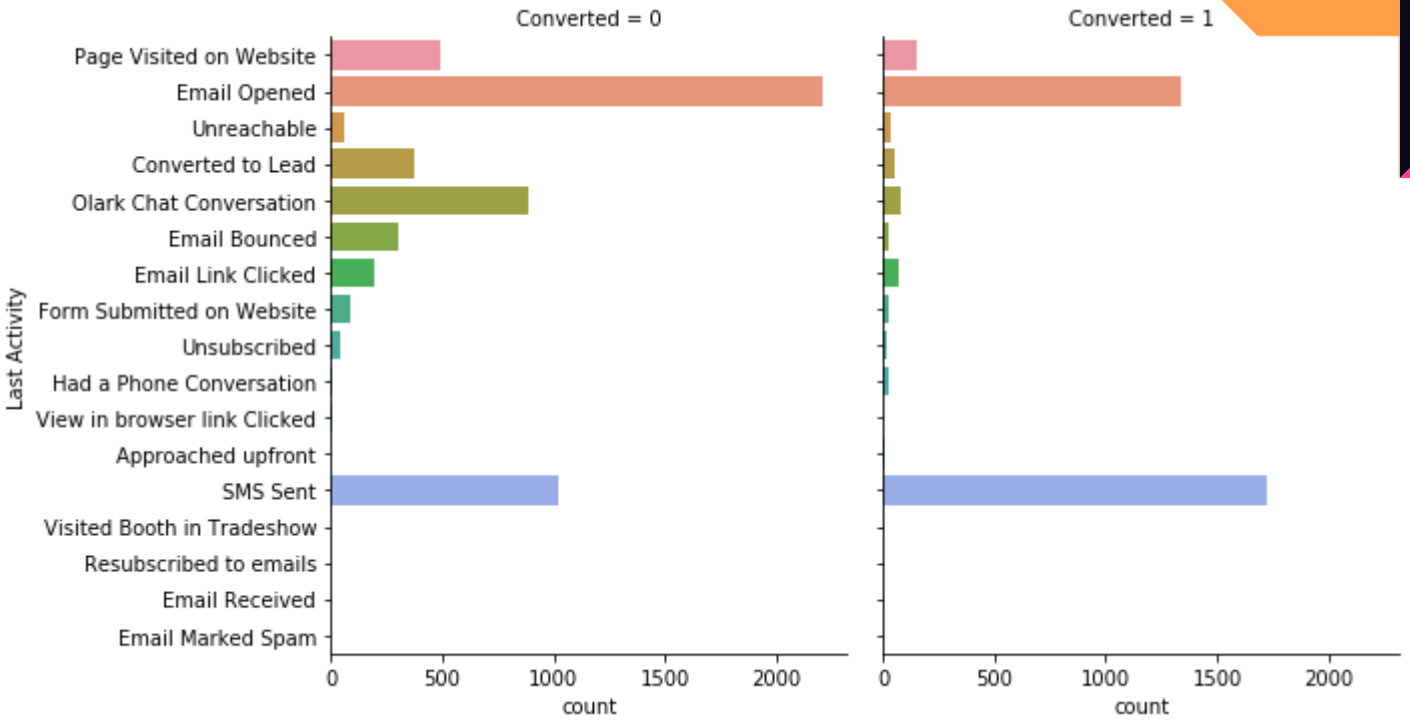
EDA plots depicting correlation (Heat Map) of all selected numerical columns.



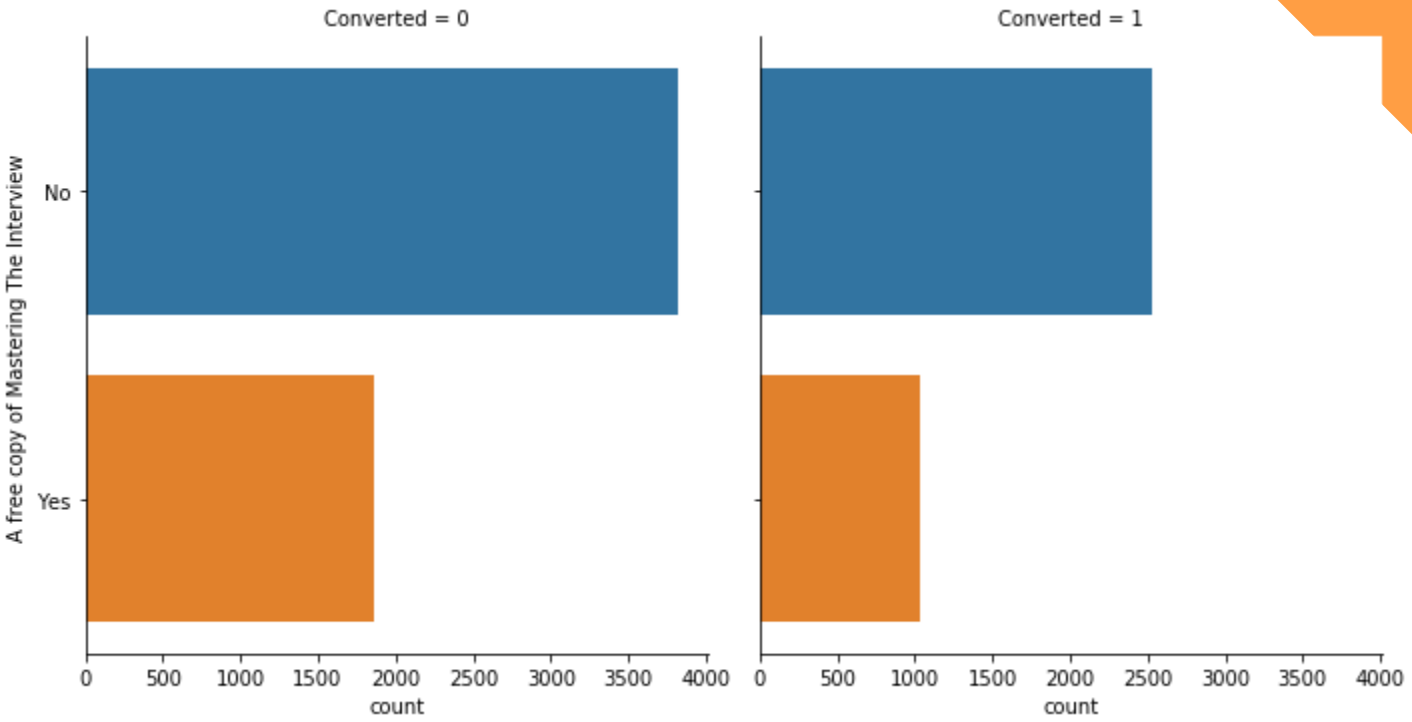




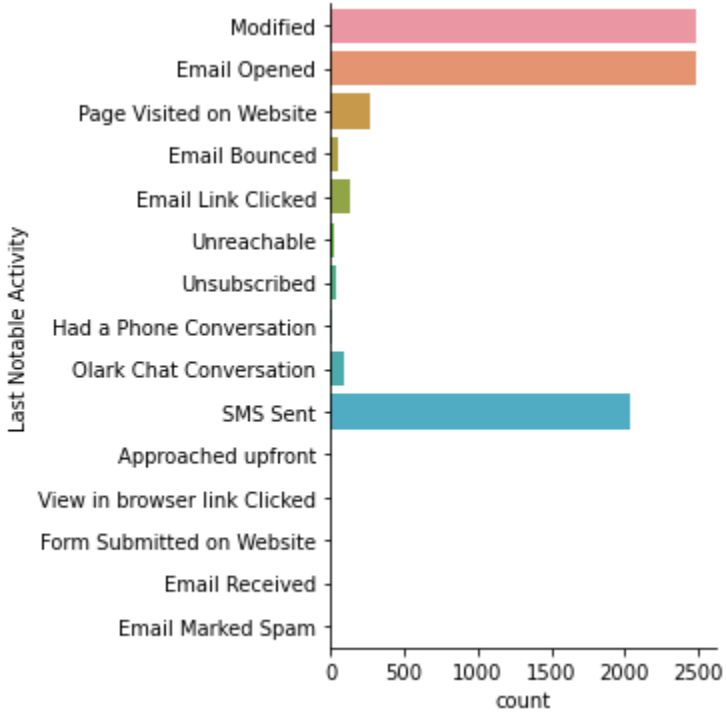
EDA plots depicting variation in numerical columns for those who Converted and those who didn't.



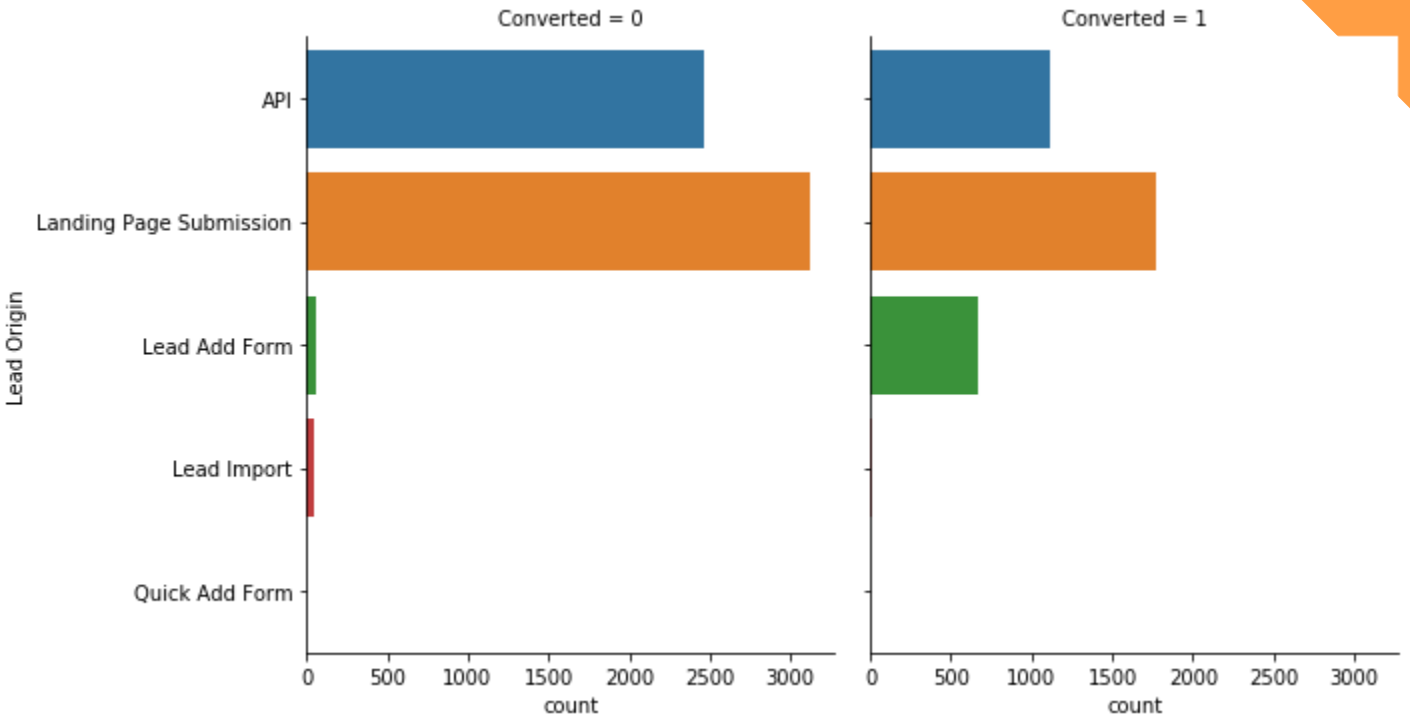
EDA plots depicting variation in categorical column (Last Activity) for those who Converted and those who didn't.



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

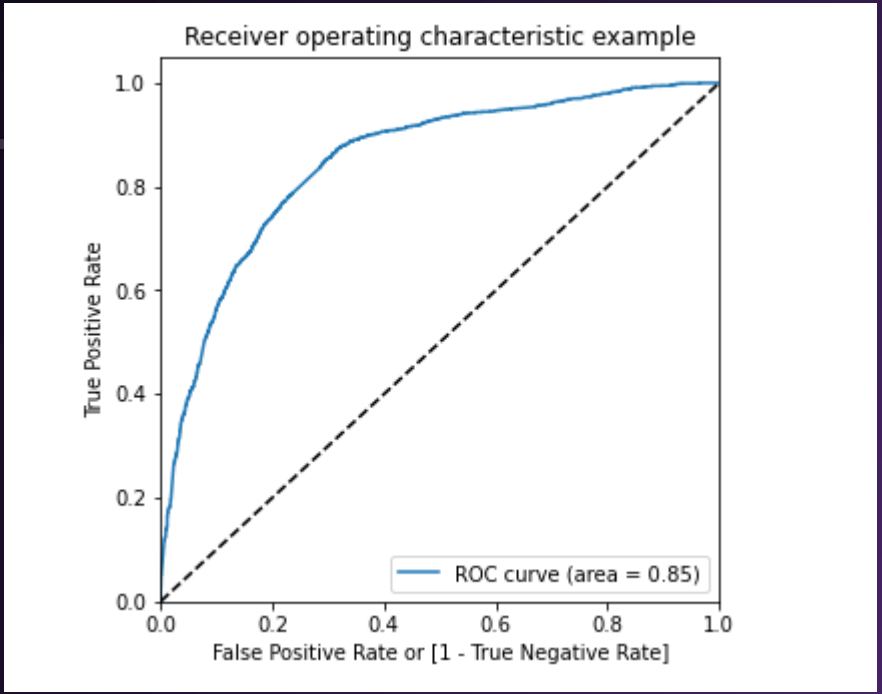


EDA plots depicting variation in categorical column (Last Notable Activity) for those who Converted and those who didn't.

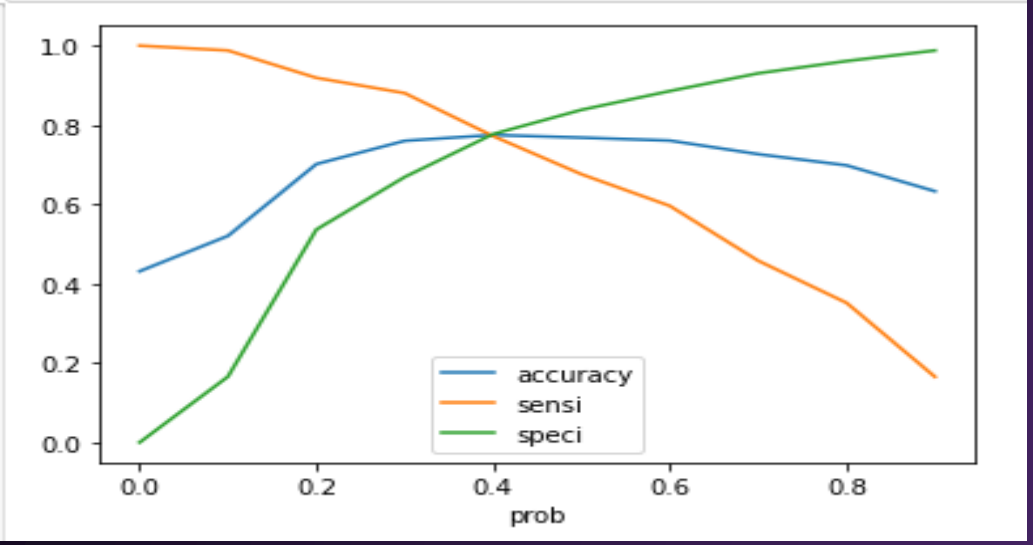


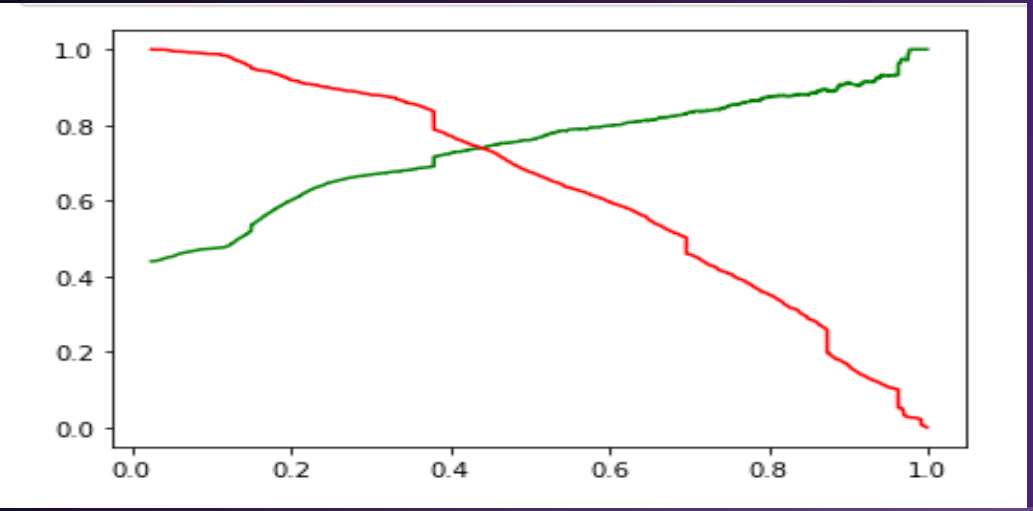
EDA plots depicting variation in categorical column (Lead Origin) for those who Converted and those who didn't.

Logistic Regression Model



The area under the curve of the ROC is 0.85 which is quite good.





Optimal values of the ‘Accuracy’, ‘Sensitivity’ and ‘Specificity’ is 0.42

Therefore 0.42 was selected as cutoff

Inference / Conclusion

Model Analysis

Performance of our Final Model

- Overall accuracy on Test set: 0.786
- Precision of our logistic regression model: 0.747
- Specificity of our logistic regression model: 0.79
- Recall of our logistic regression model: 0.754
- Sensitivity of our logistic regression model: 0.774

Inferences from Model

Business Insights Derived from our Model

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

- TotalVisits(positively impacting)
- Total Time Spent on Website(positively impacting)
- Lead Origin_Lead Add Form (positively 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, 77.4 % Sensitivity and 79 % 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
- Get Total visits increased by advertising etc. since this helps in higher conversion
- Improve the Lead Origin_Lead Add Form) since this is affecting the conversion positively.


Thank You
fn
control
option
command