

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:
Veena G
Alan S Koshy
Nishant P Mehta

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

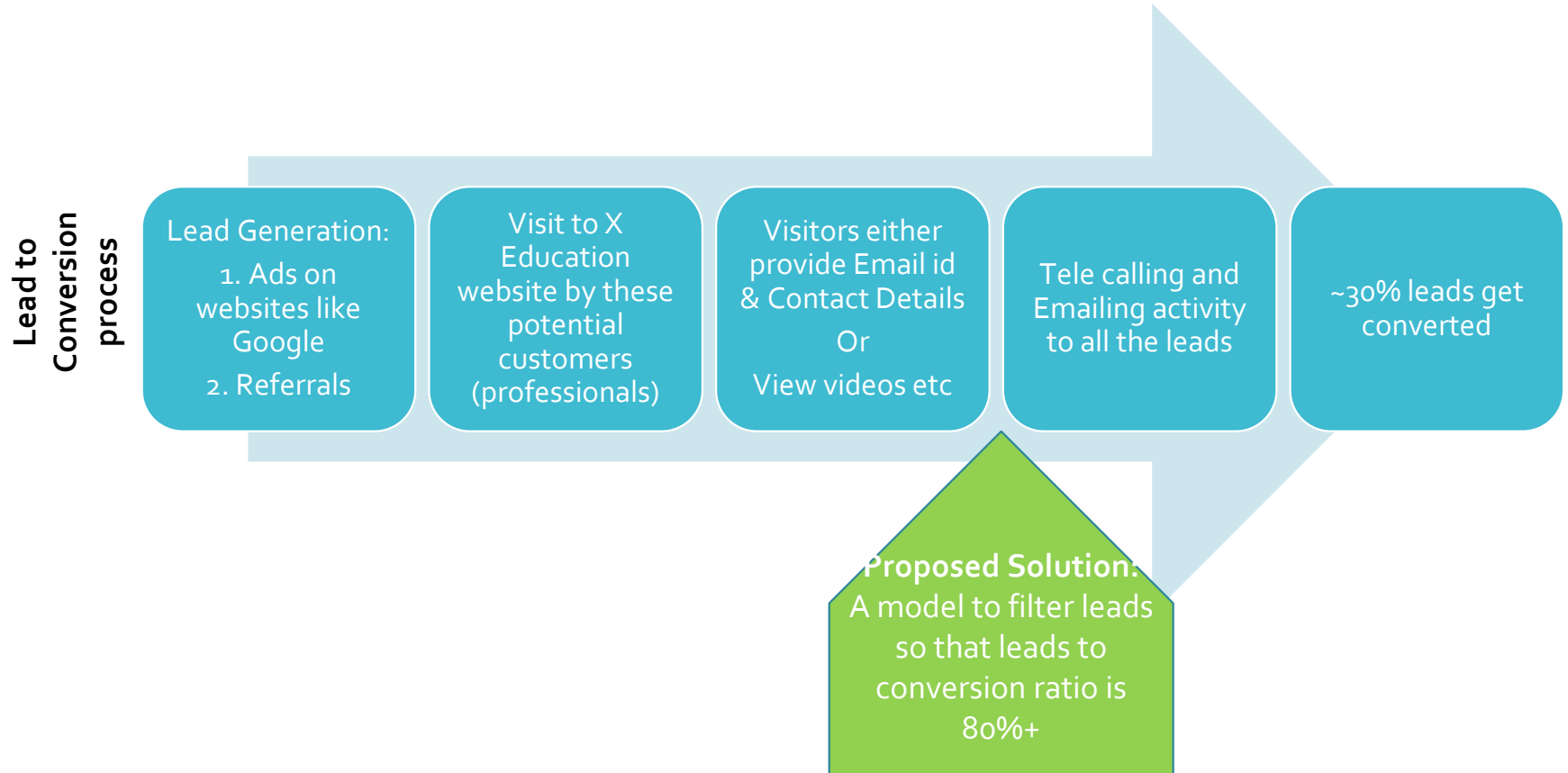
- X Education gets a lot of leads but its lead conversion rate is very poor
- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team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

Lead – Conversion Process



PROPOSED SOLUTION

Communicating with Hot Leads

Selection of 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.

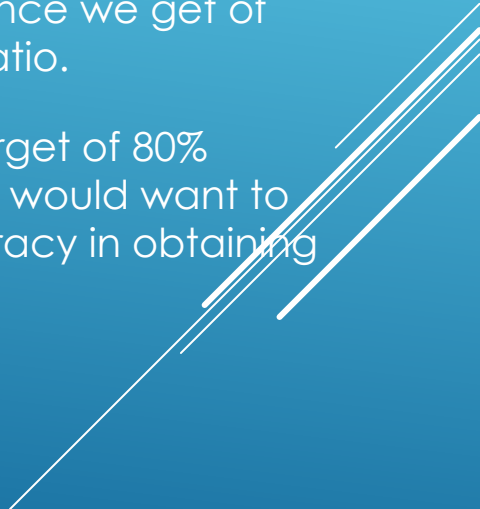
SOLUTION

Selection of Hot Leads

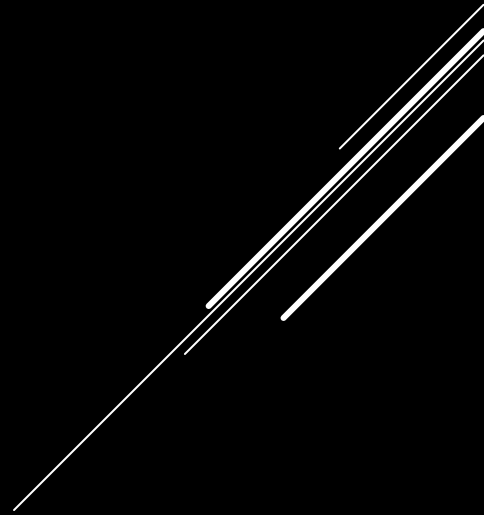
For our Problem Solution, the crucial part is to accurately identify hot leads.

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.

Several white lines of varying lengths and angles are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

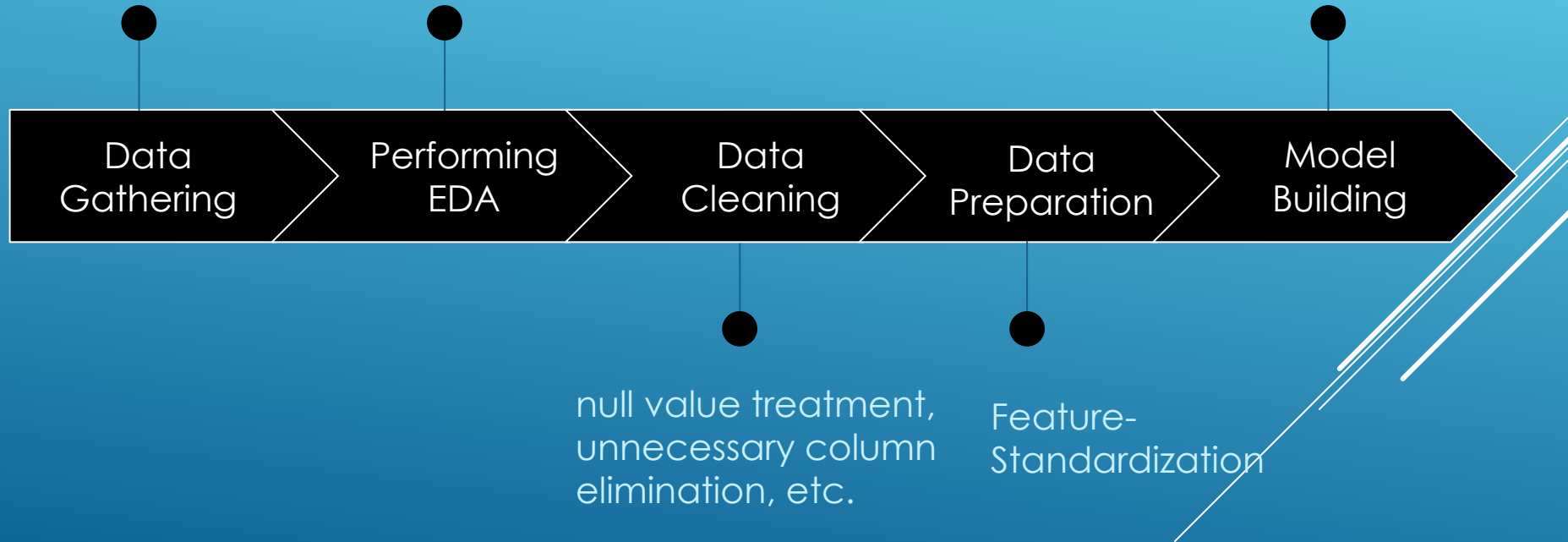
IMPLEMENTATION



Loading & Observing the past data provided by the Company

Univariate, Bivariate, and Heatmap for numerical and categorical columns

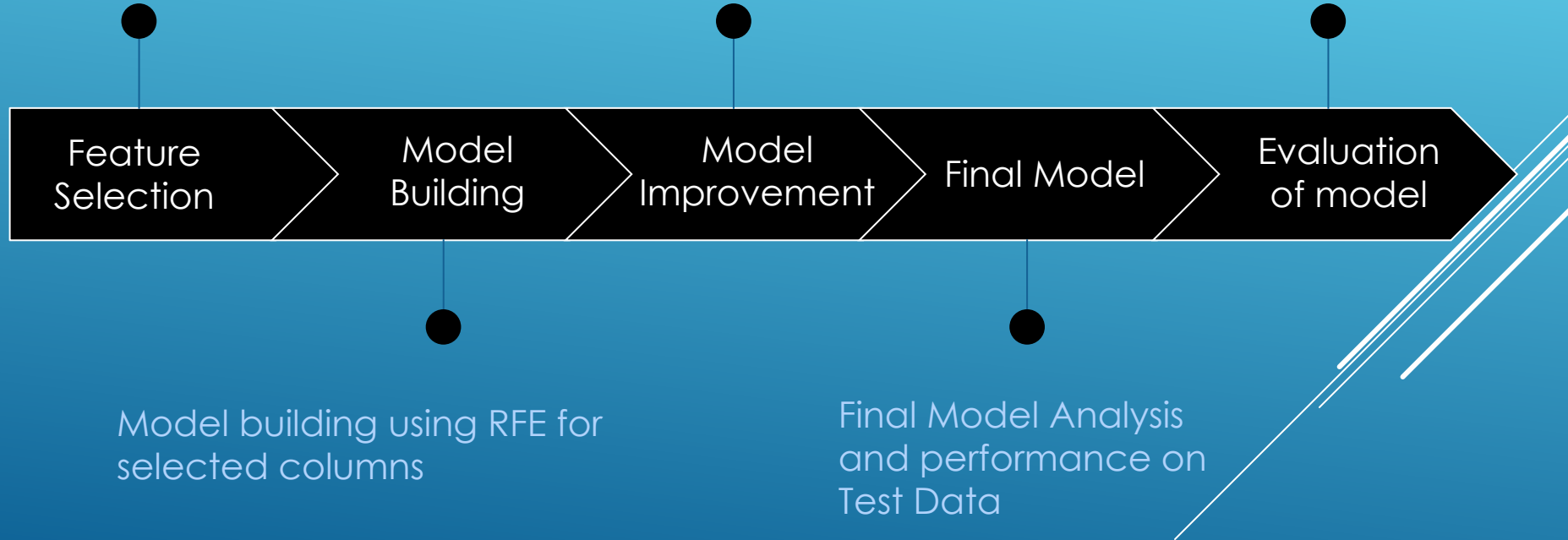
Performing pre-requisites for RFE and Logistic Regression



Selection of top 15
features using RFE

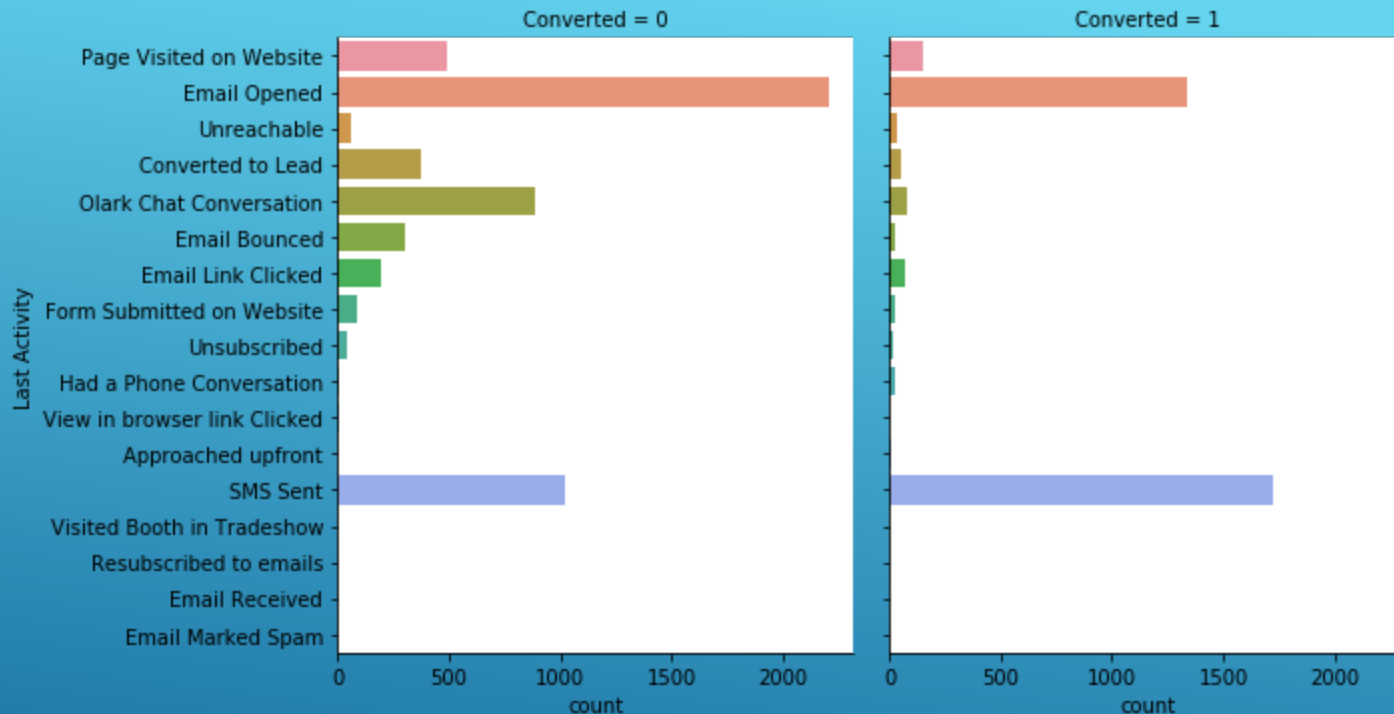
Reduction of
columns and Model
re-building

Evaluating the model
by Finding the recall
value

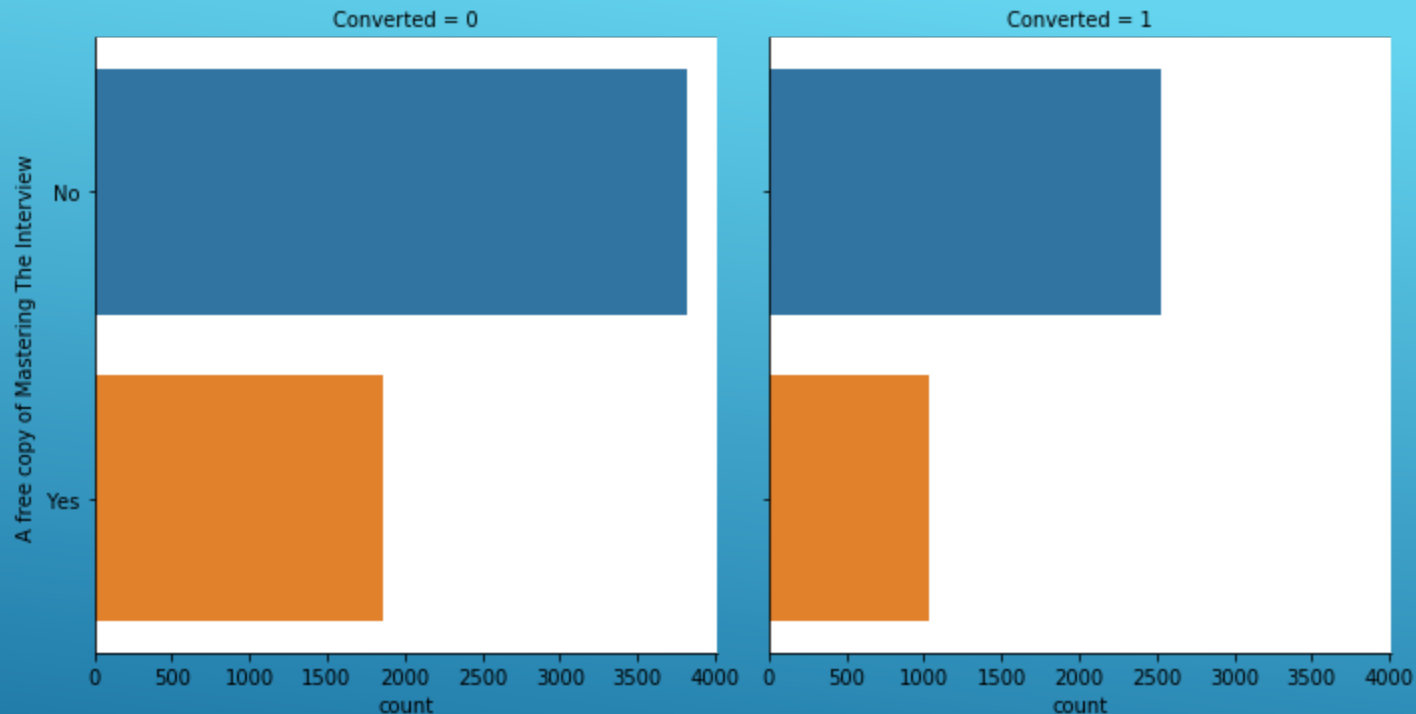


PLOTS (VISUALIZATION)

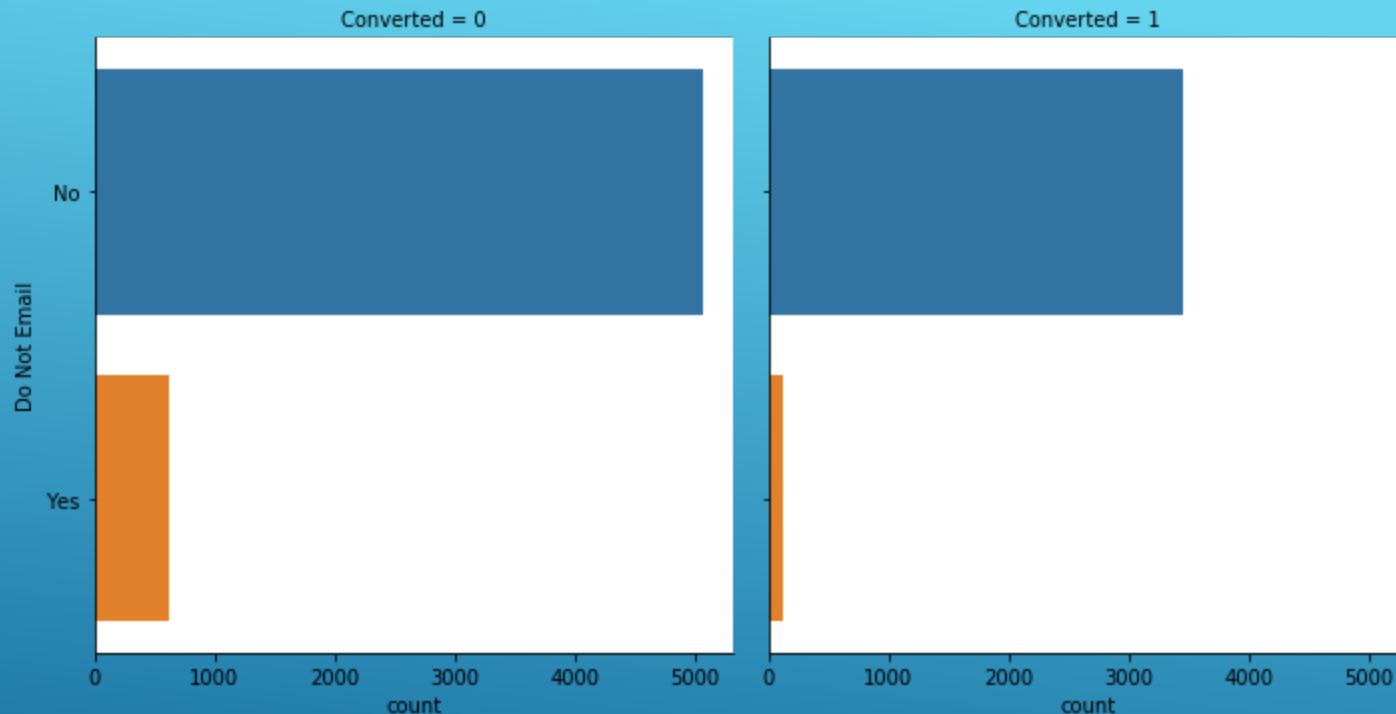




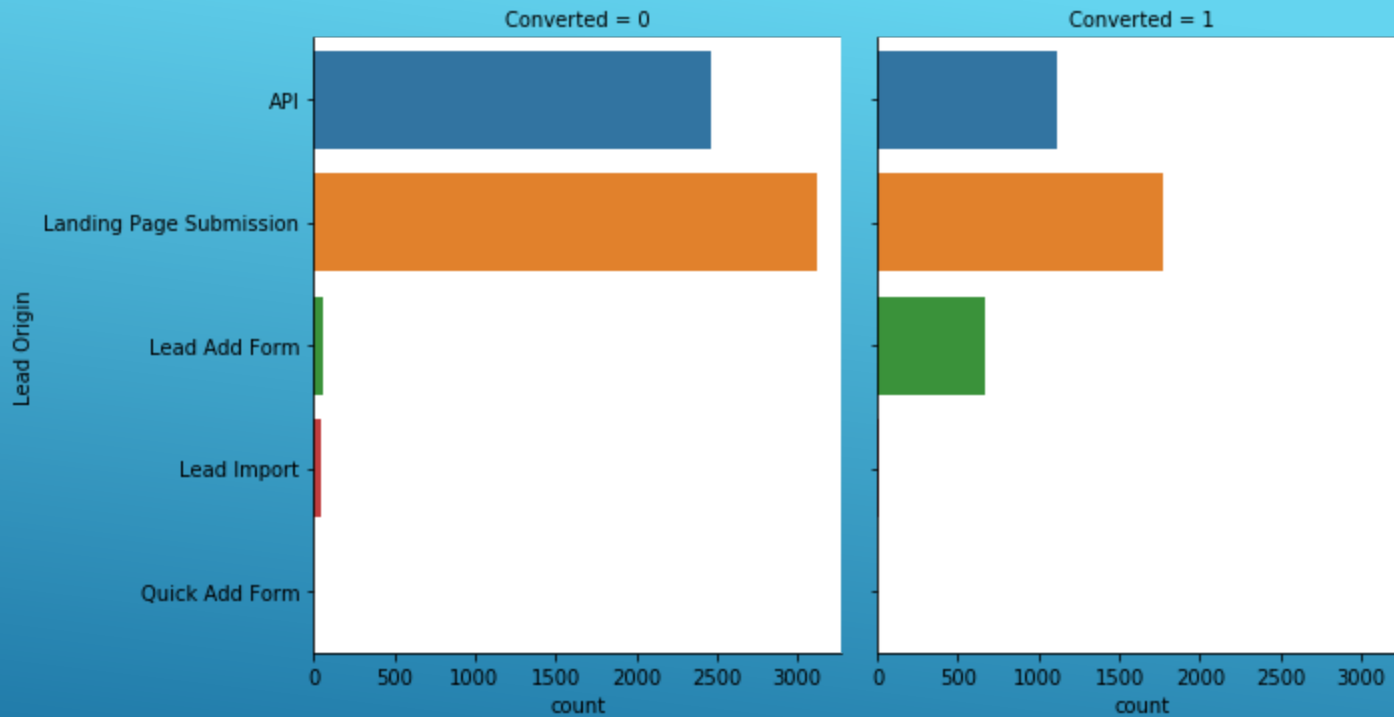
EDA plots depicting variation in categorical column (Last Activity) for those who Converted and those who didn't.



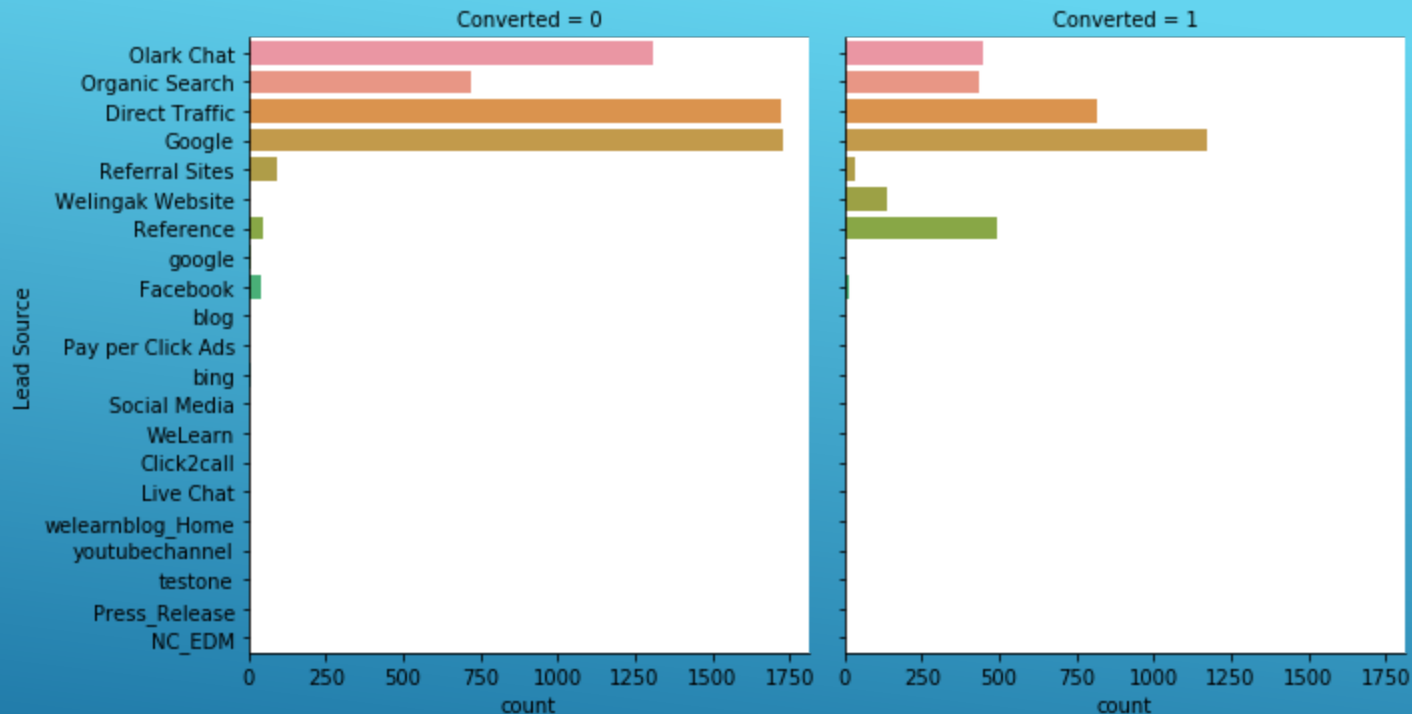
EDA plots depicting variation in categorical column (A free copy of Mastering The Interview) for those who Converted and those who didn't.



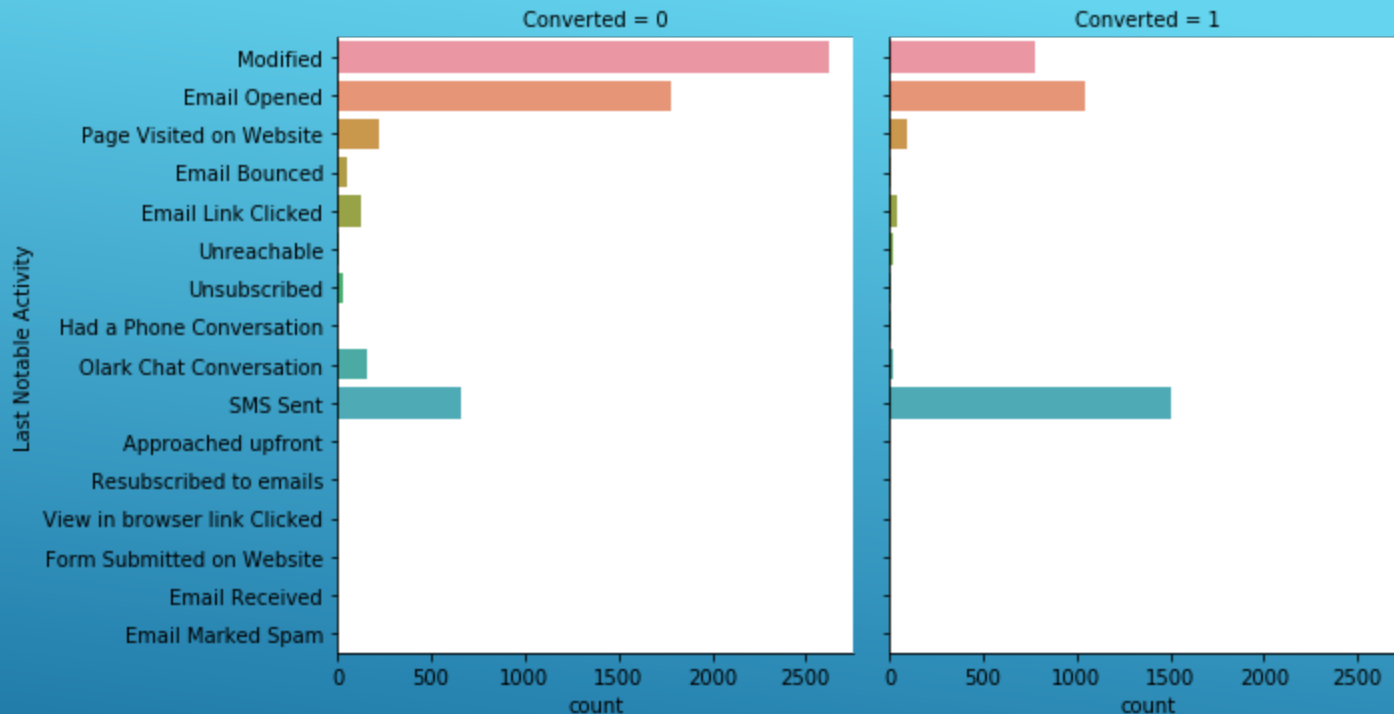
EDA plots depicting variation in categorical column (Do Not Email) for those who Converted and those who didn't.



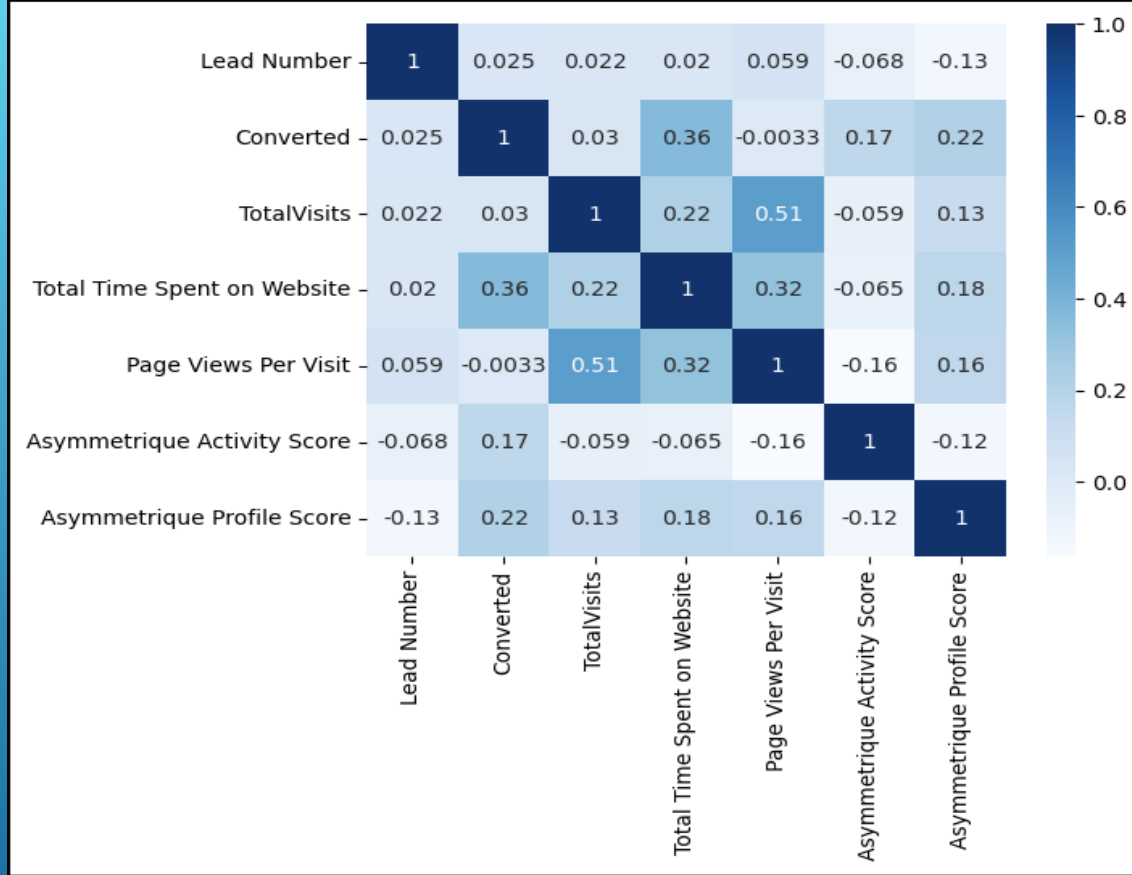
EDA plots depicting variation in categorical column (Lead Origin) for those who Converted and those who didn't.



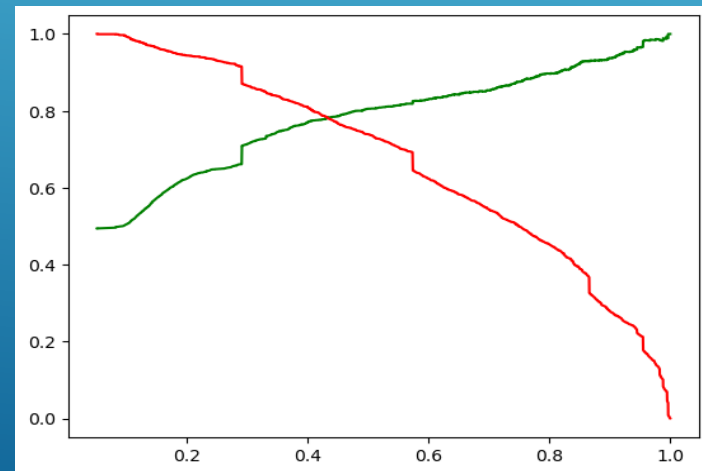
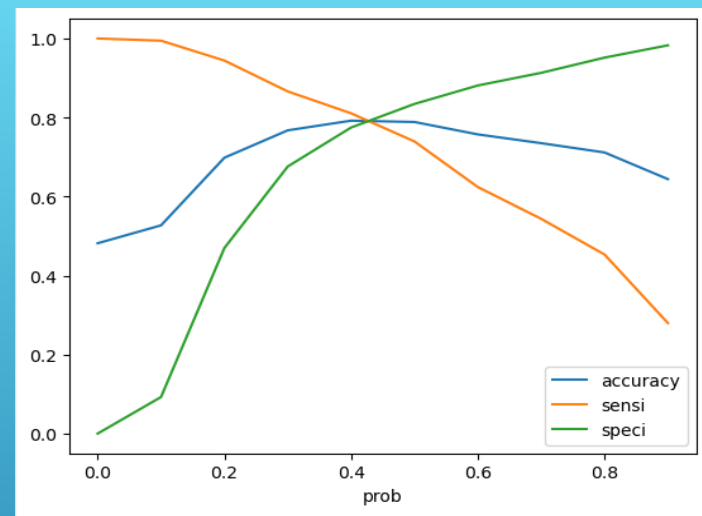
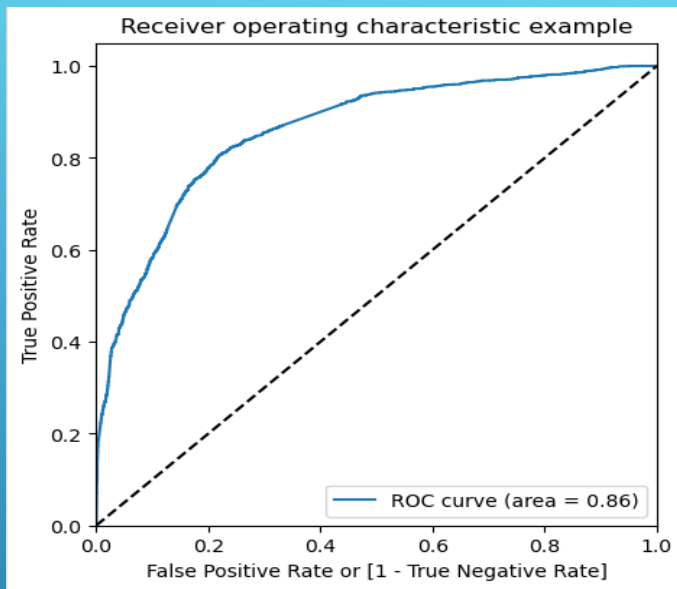
EDA plots depicting variation in categorical column (Lead Source) for those who Converted and those who didn't.



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.**



**Logistic Regression Final
Model Parameters**
Area under ROC = 0.86
Optimal cut-off = 0.42
Precision recall cut-off = 0.44

INFERENCE / CONCLUSION



MODEL ANALYSIS

Performance of our Final
Model

Overall accuracy on Test set: 0.786

Sensitivity of our logistic regression
model: 0.77

Specificity of our logistic regression
model: 0.78

Several white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

INFERENCES FROM MODEL

Business Insights Derived from
our Model

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

- Total Visits
 - Total Time Spent on Website
 - Lead Origin_Lead Add Form
- 
- Several white lines of varying lengths and angles are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

CONCLUSION 1 (LR MODEL)

Our Logistic Regression Model is decent and accurate enough. We have an accuracy value of 0.78 on Test Set, 0.77 Sensitivity and 0.78 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
- 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