X EDUCATION- LEAD SCORING CASE STUDY

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

GROUP MEMBERS-

Yechuri Mytreya

Vivekanand Mishra

Anju Bani

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.

- 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

- 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

- 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 scores save a higher conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.

LEAD CONVERSION PROCESS

Lead to the conversion process

Lead Generation:

- 1. Ads on websites like Google
 - 2. Referrals

Visit to X Education website by these potential customers (professionals)

Visitors either provide Email id & Contact Details

Or

View videos etc

Tele calling and Emailing activity to all the leads

~30% leads get converted

Proposed Solution:

A model to filter leads so that leads to conversion ratio is 80%+

PROPOSED SOLUTION

Leads Clustering

Focus Communication

Increase conversion

to convert, thus, getting a with effective communication. smaller section of hot leads to focus more on.

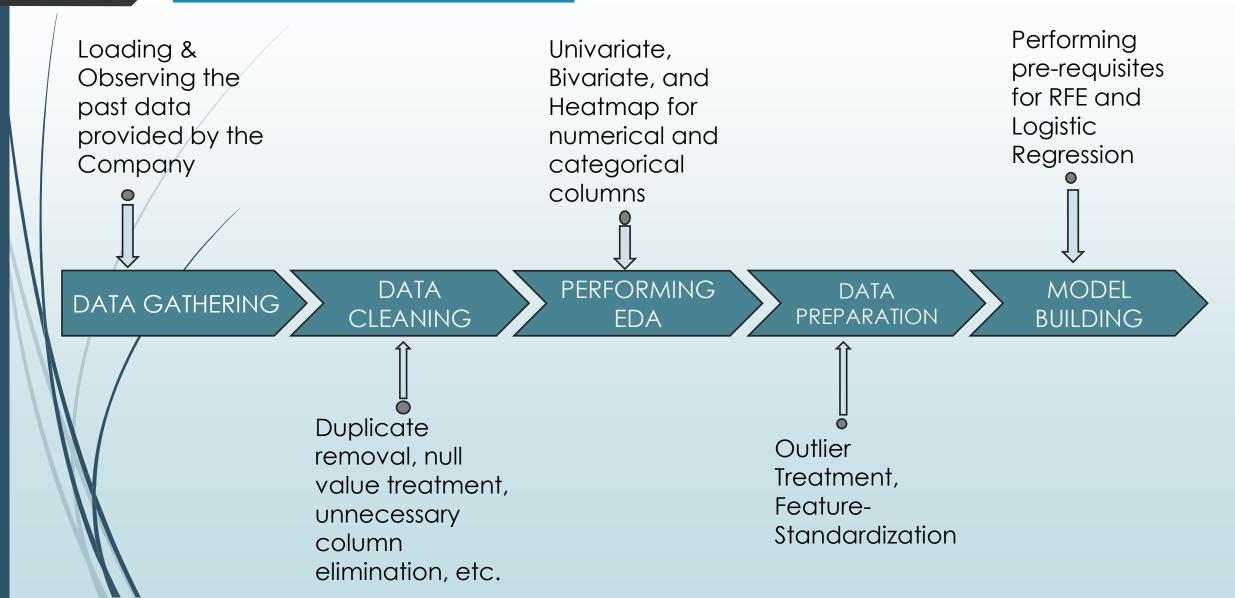
We cluster the leads into Since we would have a smaller Since we focus on hot leads. certain categories based on set of leads to communicate which were more probable to their tendency or probability with, we m make more impact convert, we would have a

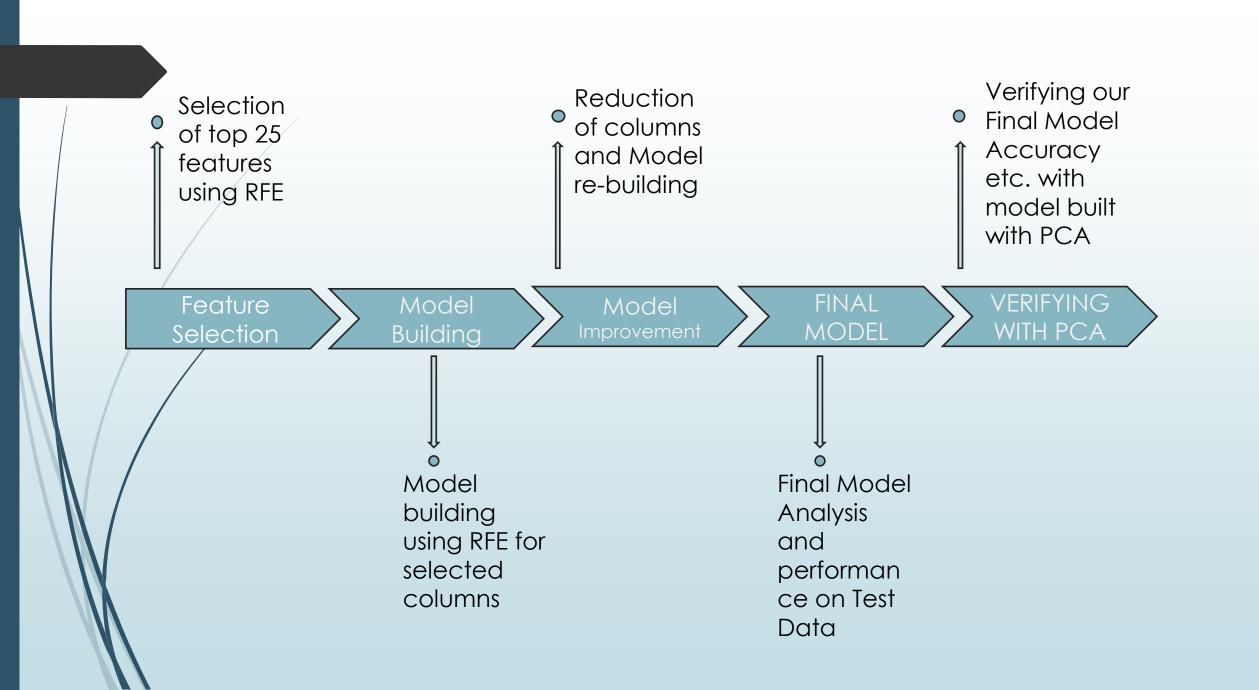
better conversion rate, and hence we can achieve the 80% target.

SOLUTION SELECTION OF HOT LEADS

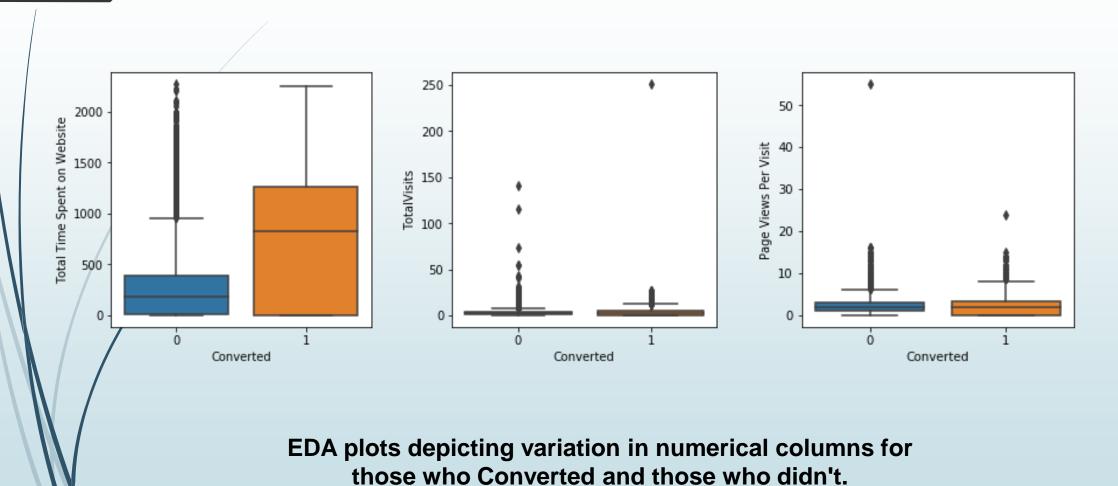
- ☐ For our Problem Solution, the crucial part is to accurately identify hot leads.
- The more accurately we obtain the hot lead, the more chance we get of a higher conversion ratio.
- ☐ Since we have a target of 80% conversion rate, we would want to obtain high accuracy in obtaining hot leads.

IMPLEMENTATION

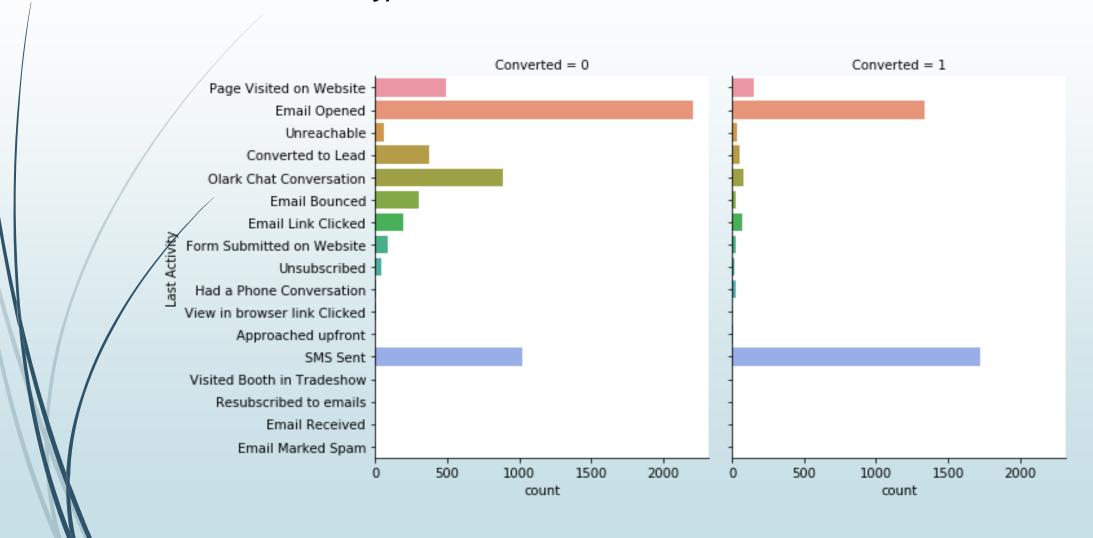




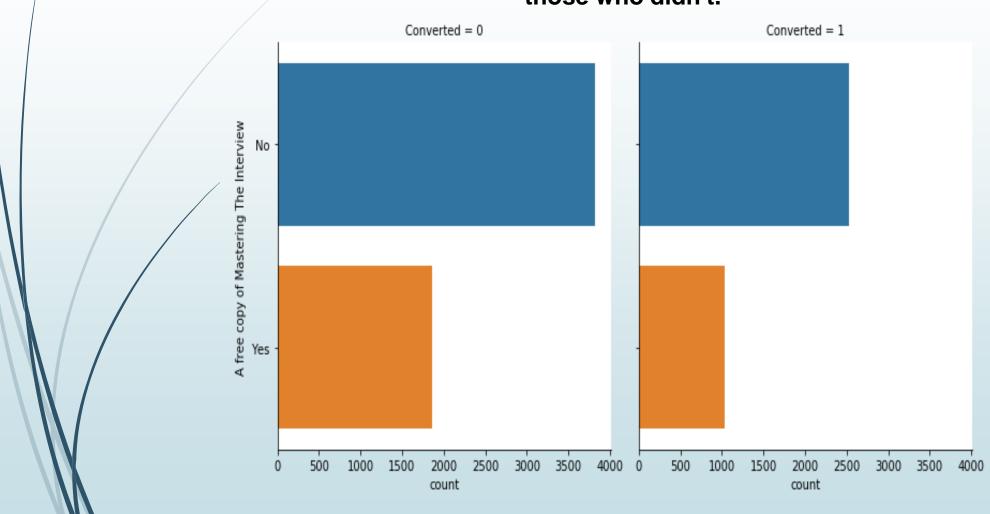
PLOTS VISUALISATION



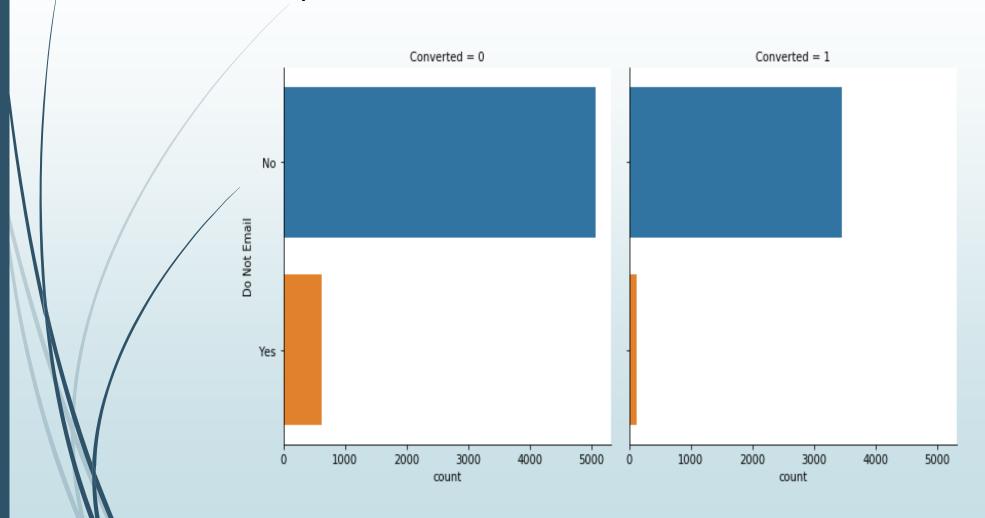
EDA plots depicting variation in categorical column (Last Activity) for those who Converted and those who didn't.



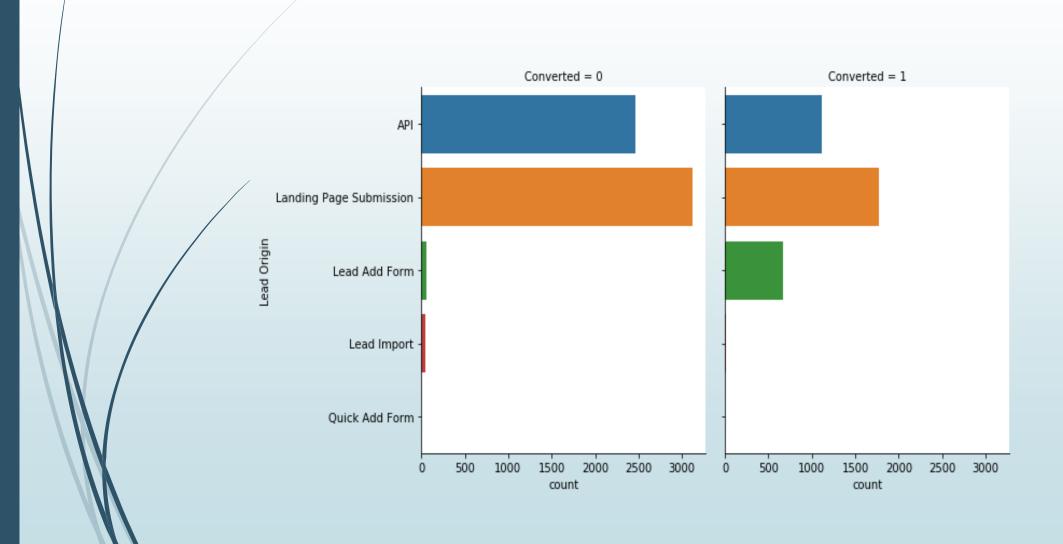
EDA plots depicting variation in categorical column (A free copy of Mastering The Interview) for those who Converted and those who didn't.



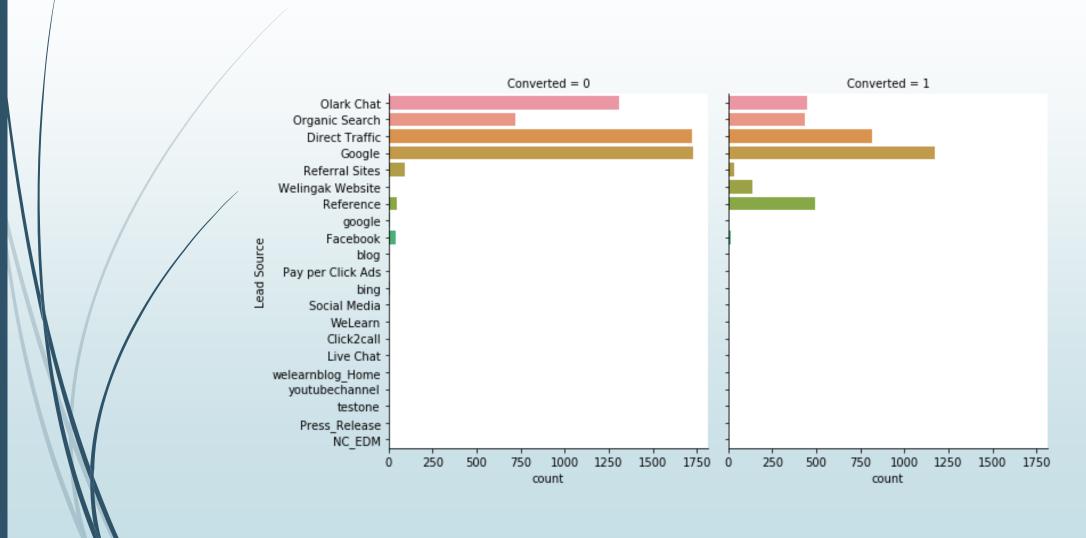
EDA plots depicting variation in categorical column (Do Not Email) for those who Converted and those who didn't



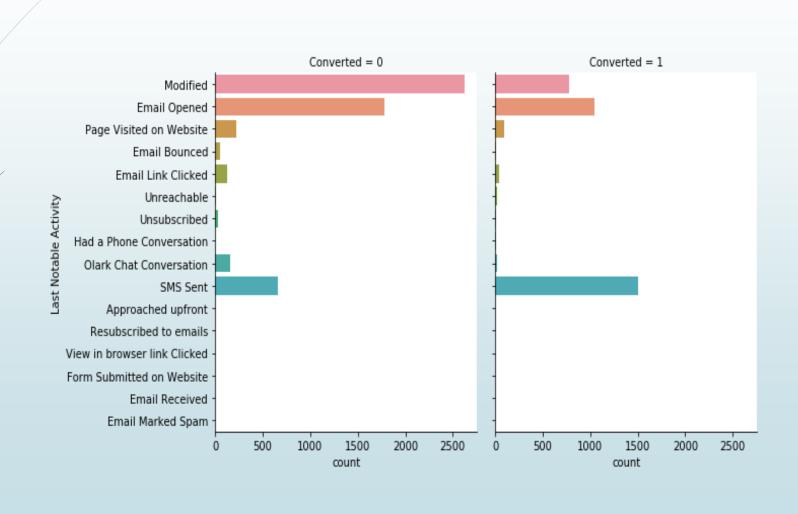
EDA plots depicting variation in categorical column (Lead Origin) for those who Converted and those who didn't.



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



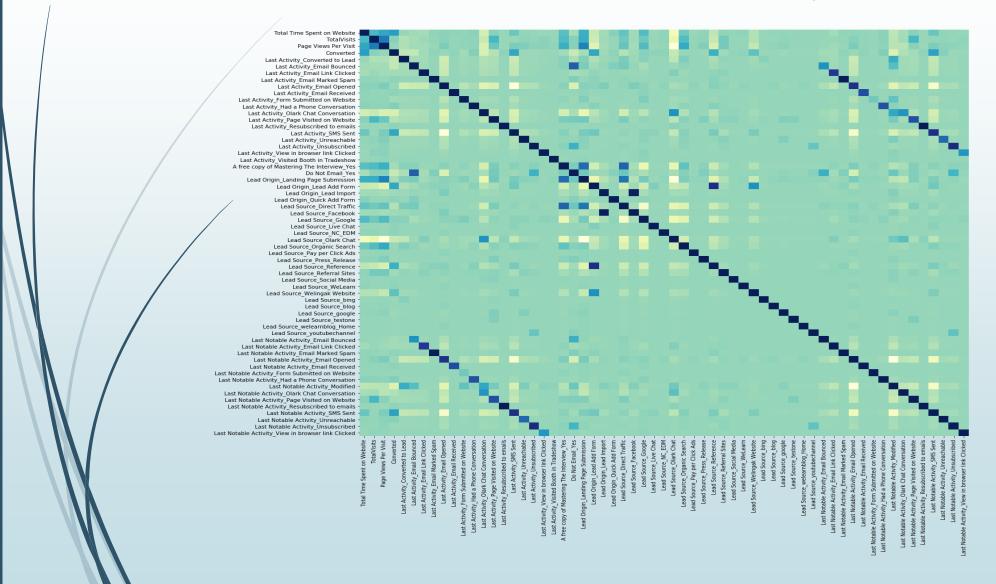
EDA plots depicting variation in categorical column (Last Notable Activity) for those who Converted and those who didn't.



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).



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

The top 3 variables in the model, that contribute towards lead conversion are:

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

The top 3 variables in my model, that should be focused on 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 the availability of extra resources and vice-versa.

Conclusion 2

- X Education Company needs to focus on the following key aspects to improve the overall conversion rate:
- Increase user engagement on their website since this helps in higher conversion
- Increase 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!!



