

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:

1. Ahinram
2. Ayushi
3. Sonali

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

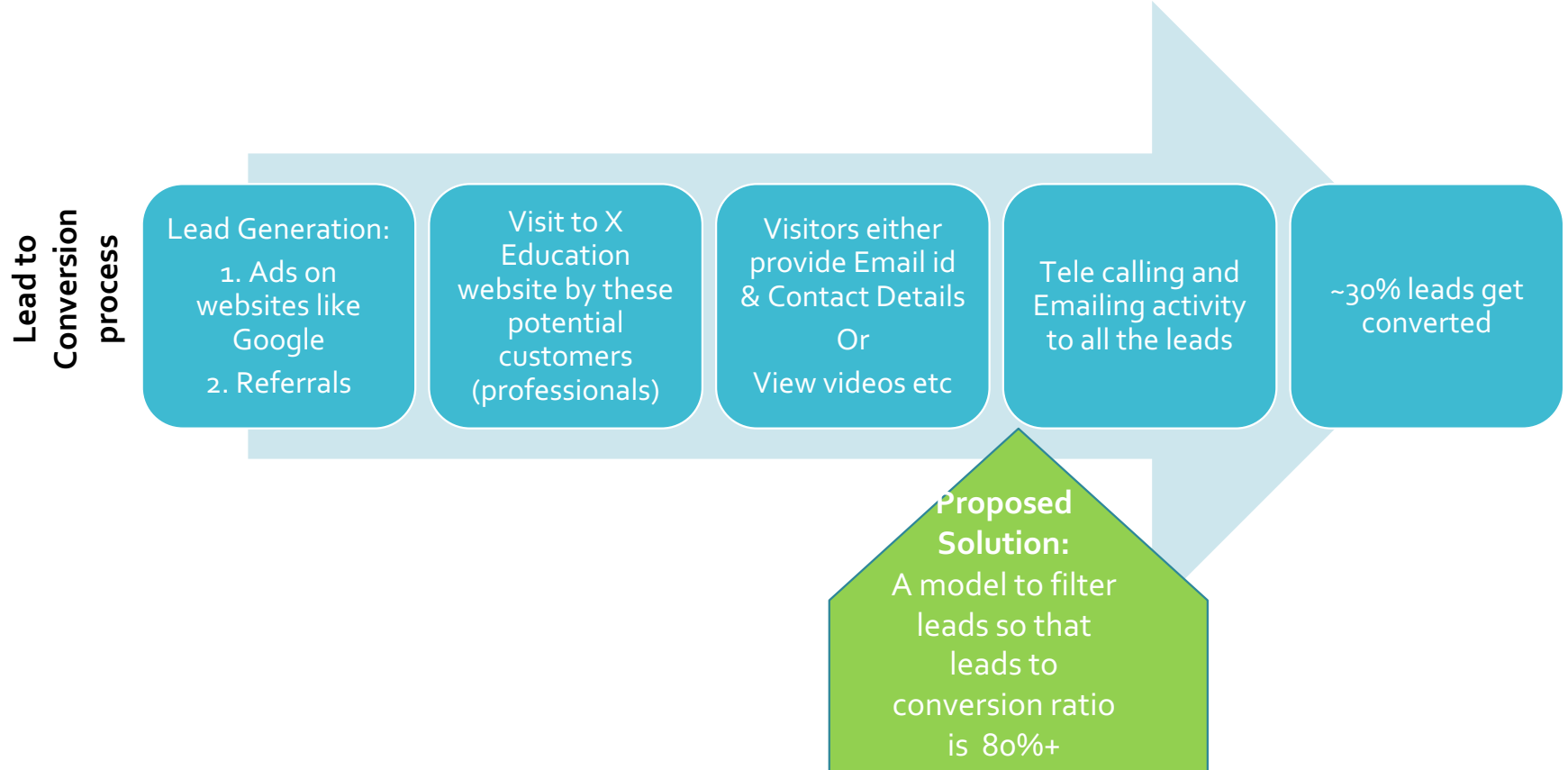
- X Education gets a lot of leads but its lead conversion rate is very poor
 - To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
 - If they successfully identify this set of leads, the lead conversion rate should go up as the sales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

Selection 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.

Communicating with Hot Leads

Focus Communication

Since we would have a smaller set of leads to have communication with, we might make more impact with effective communication.

Conversion of Hot Leads

Increase conversion

Since we focused 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.

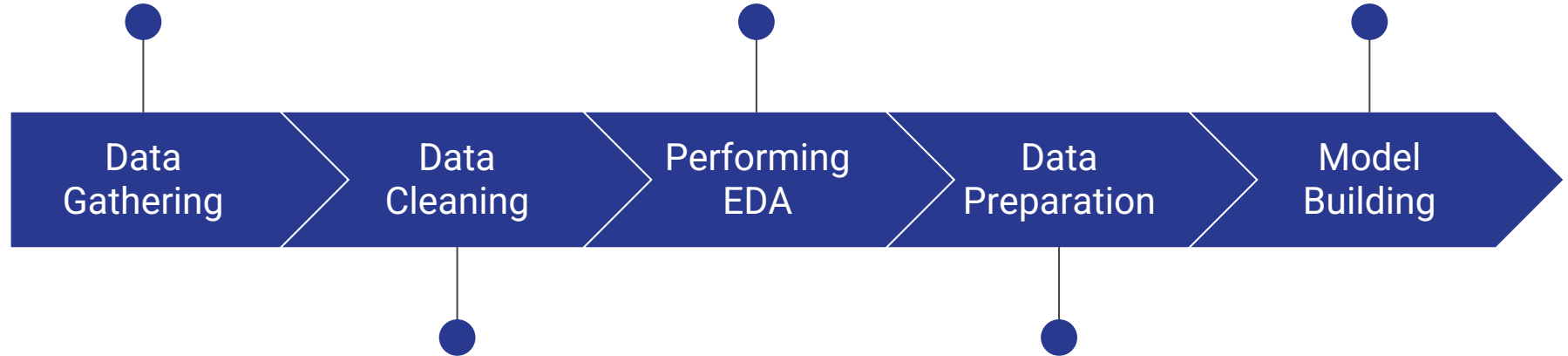
Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

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

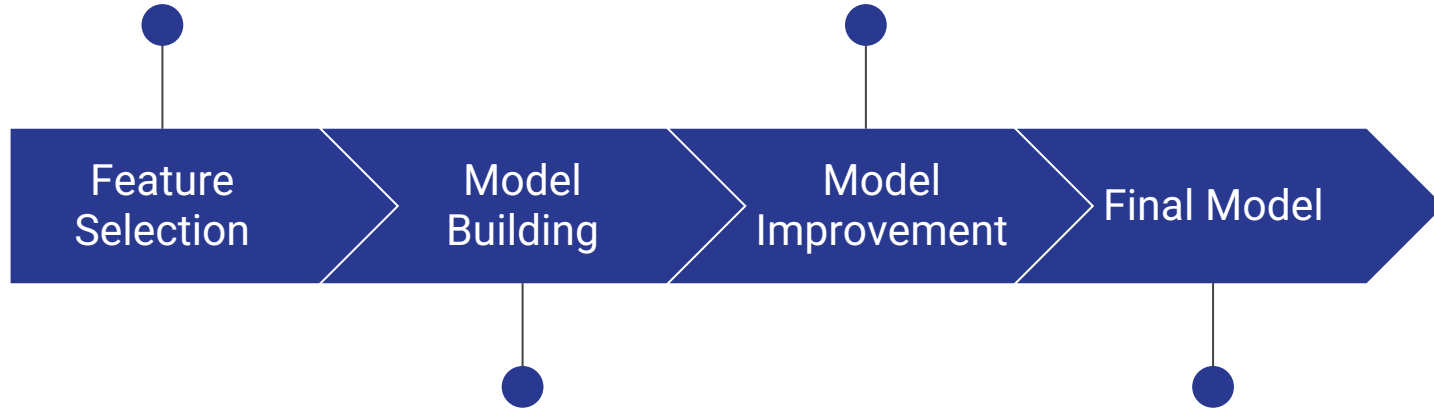


Duplicate removal, null value
treatment, unnecessary
column elimination, etc.

Outlier Treatment,
Feature-
Standardization

Selection of top 25
features using RFE

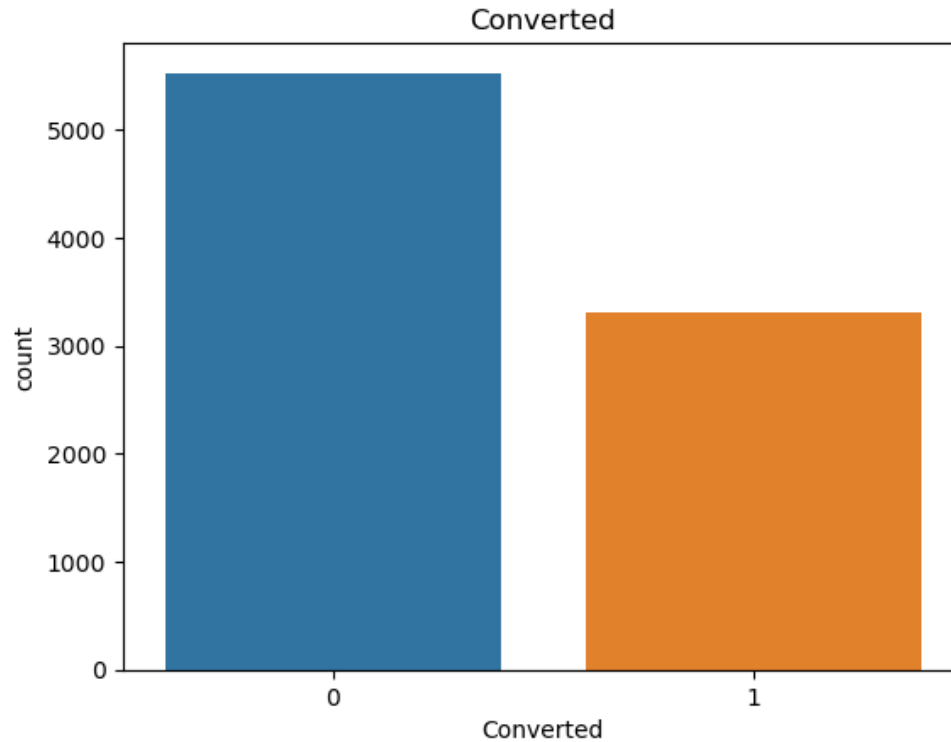
Reduction of columns
and Model re-building



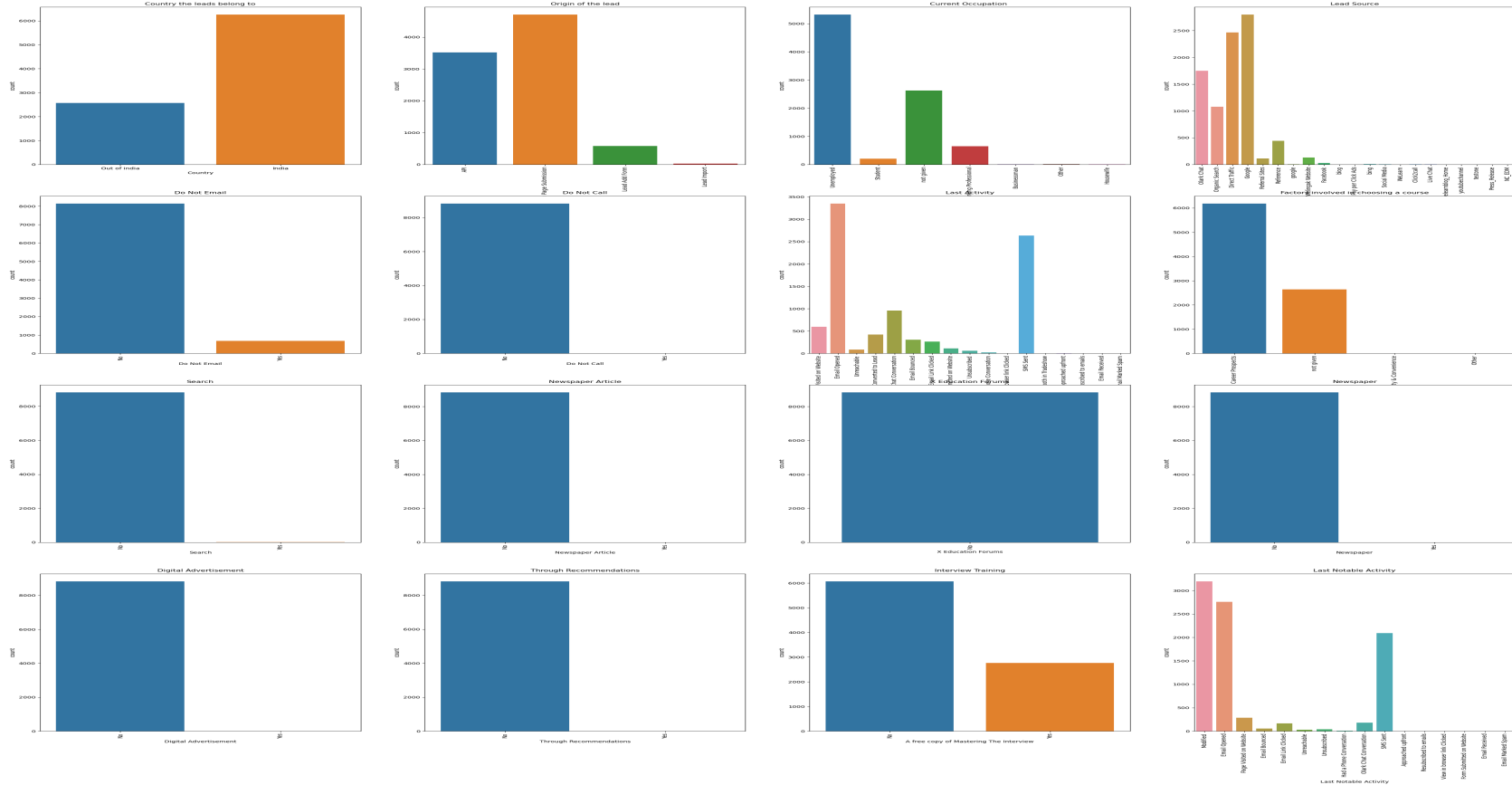
Model building using RFE for
selected columns

Final Model Analysis and
performance on Test
Data

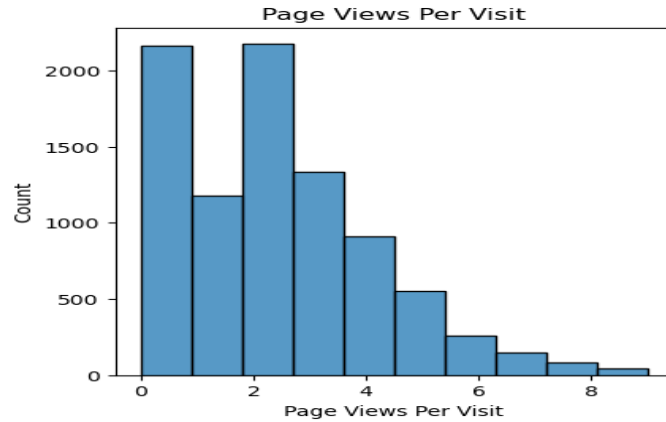
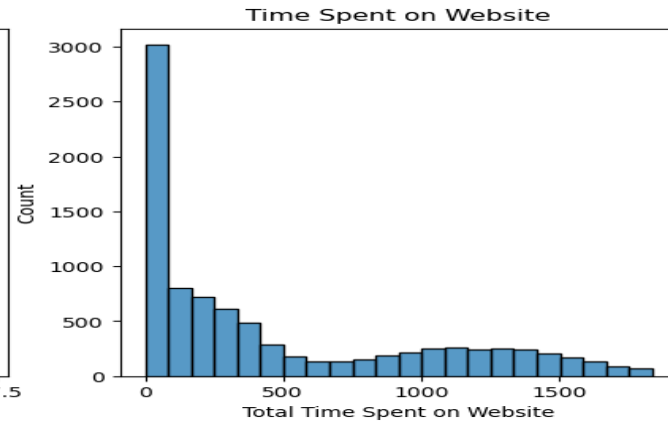
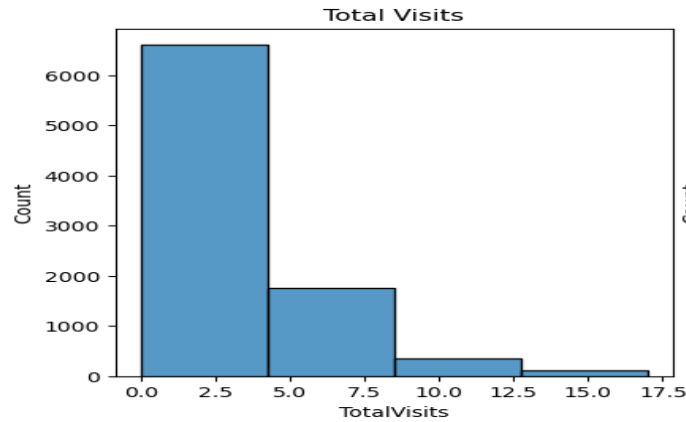
Plots (Visualization)



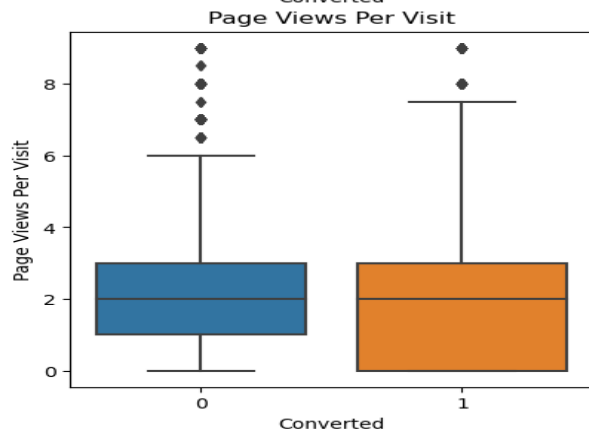
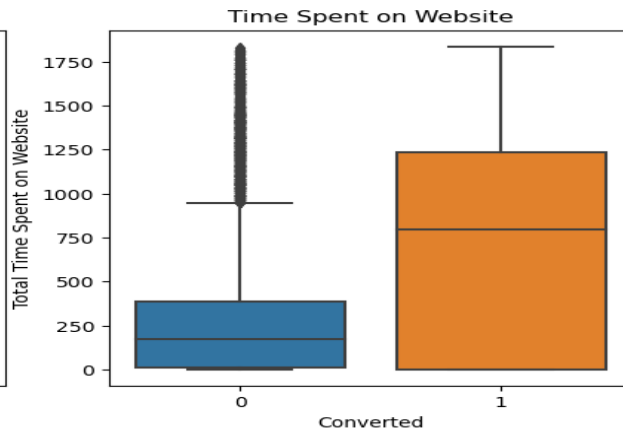
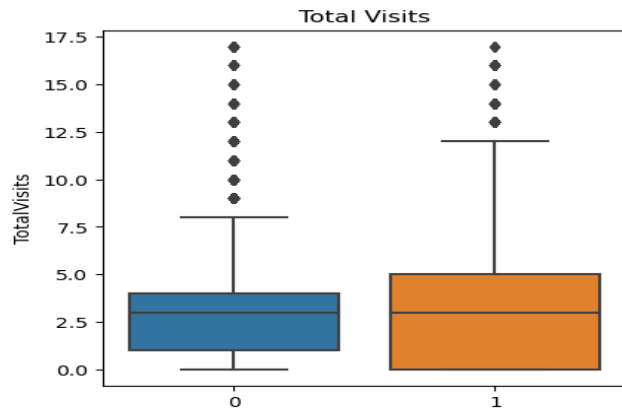
EDA plots showing Univariate Analysis for Converted and those who didn't.



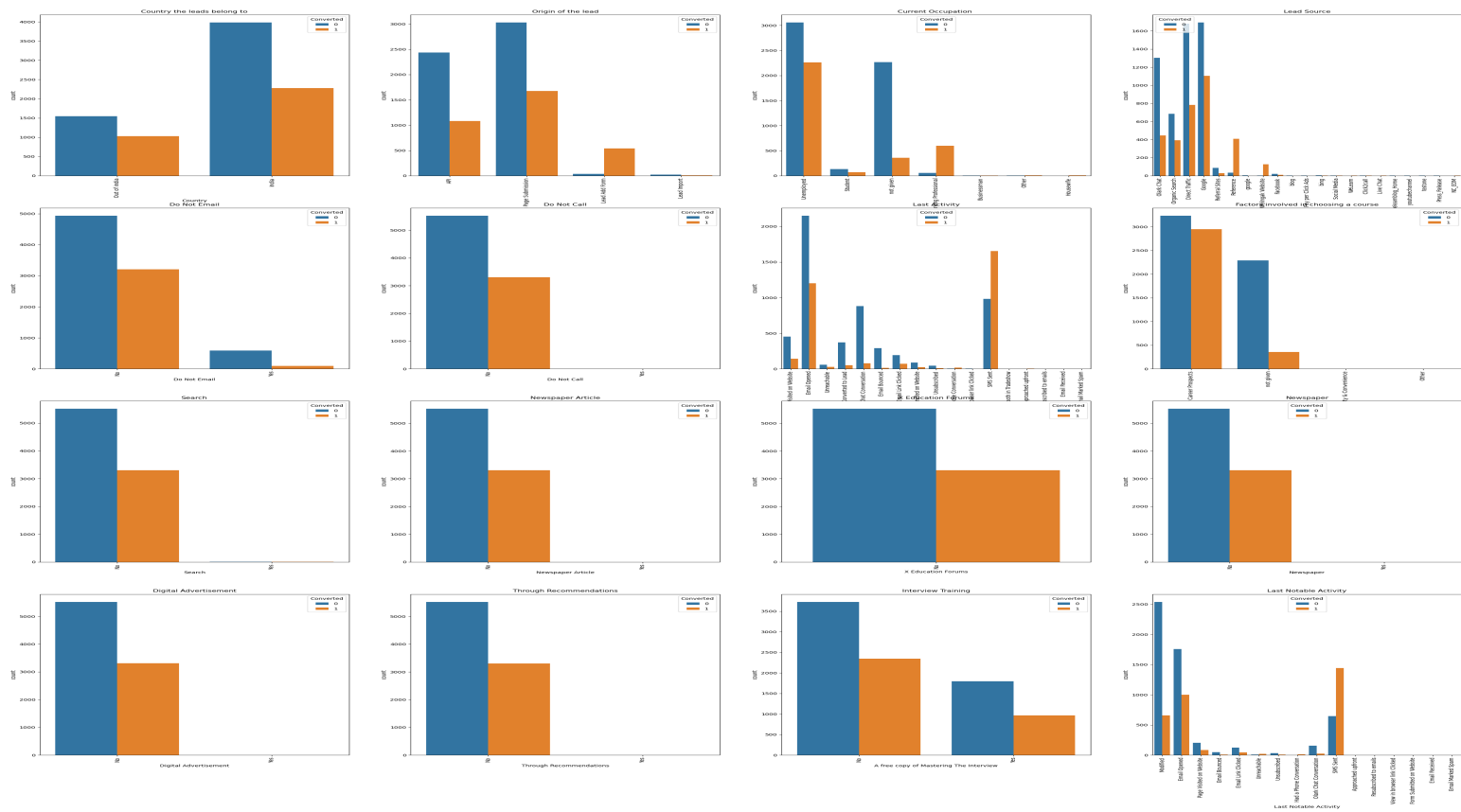
EDA plots showing categorical variables to understand the variance of the data.



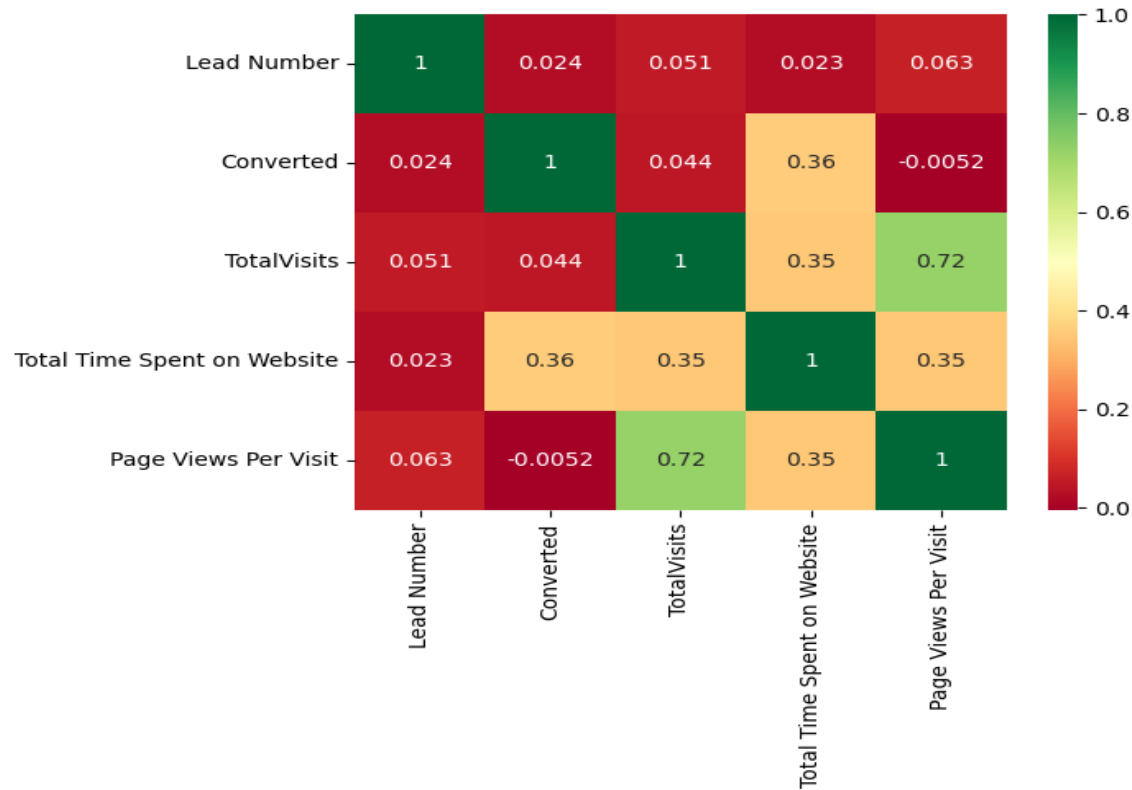
EDA plots showing Numerical variables to understand the variance of the data.



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



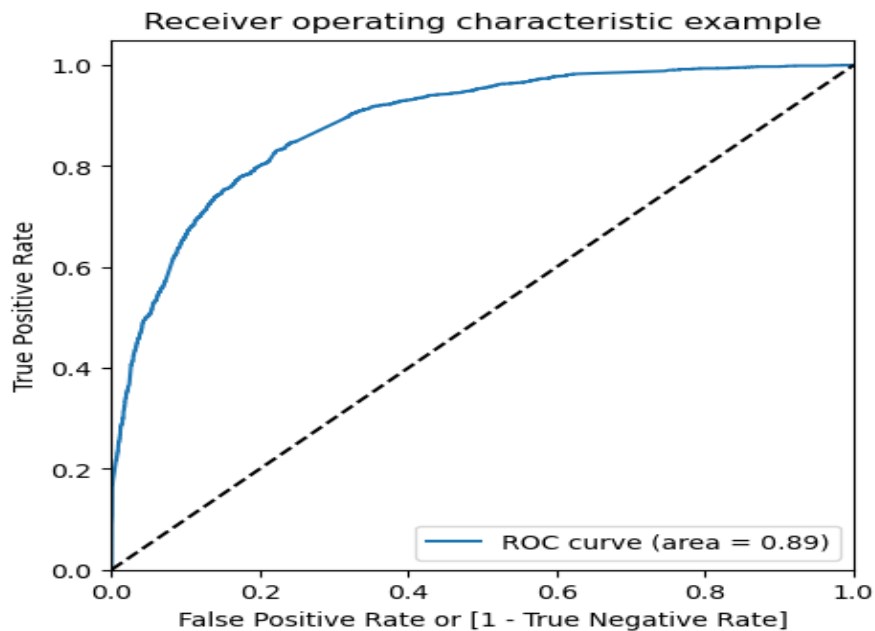
EDA plots with Bivariate analysis with all the variables Vs. Target variable.



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).

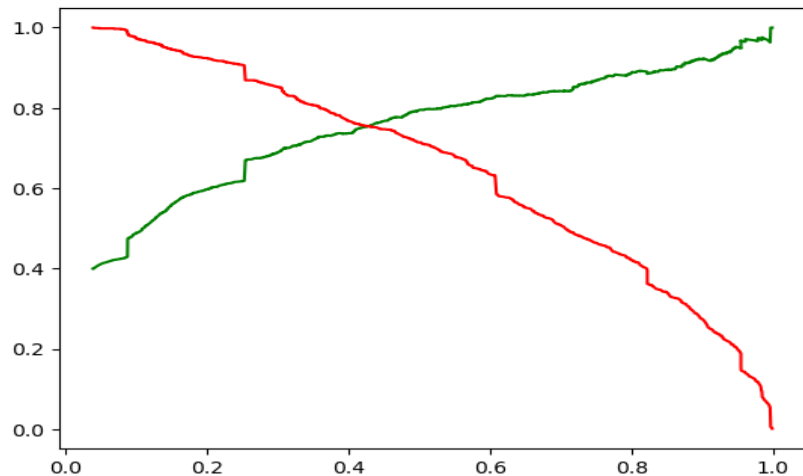
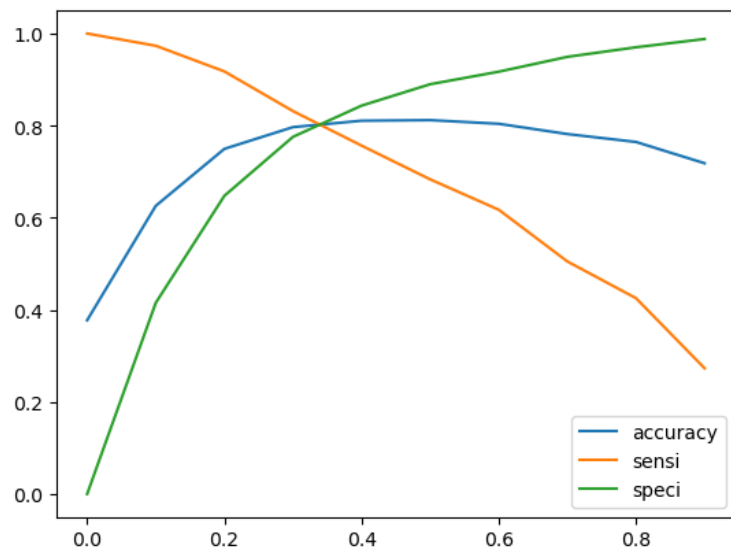


ROC Curve

Curve 1 – Area Under ROC = 0.84

Curve 2 – We can observe that 0.35 is the point where all metrics meet. 0.35 is the optimal probability (Intermediate cut-off = 0.35).

Curve 3 – It is clearly seen that the trade off is happening @ 0.4 (Final cut-off = 0.42)





Inference / Conclusion

Model Analysis

Performance of our Models

We concluded that the model 5 with 0.35 as a cut-off is delivering the Recall value of 78.7% on training dataset & 80% on the test dataset.

With the obtained cut-off we got an accuracy of 80%, Specificity of 81% & Sensitivity of 79%

For the test data we got the result of Accuracy – 80%, Specificity – 82% & Sensitivity – 80%

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
- Total number of visits
- When the lead source was: Olark chat conversation

Inferences from Model

Business Insights Derived from our Model

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

- Lead Source with elements Organic search.
 - Lead Source with elements direct traffic.
 - Lead Source with elements Olark chat conversation.
-

Conclusion 1

For this particular problem - Recall is the important metric to evaluate the model performance. From the above we can confirm that with the probability cut off of 0.35 gives the best performance.

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