X Education - Lead Scoring Case Study

Submitted By:-
Nikhil Dhiman
Raj Sharma

Background

X Education Company

- Industry professionals can purchase online courses from X Education, a company that provides education. Many experts who are interested in the courses visit their website on any given day and search for courses.
- On numerous websites and search engines like Google, the firm advertises its courses. Upon arriving at the website, these visitors may browse the courses, submit a form for the course, or watch some videos. These persons are categorised as leads when they fill out a form with their phone number or email address. Additionally, the business receives leads from earlier recommendations. Once these leads are obtained, sales team members begin calling, sending emails, etc. Some leads are converted during this procedure, but most are not. At X Education, the normal lead conversion rate is roughly 30%.

Problem Statement

X Education Company's Problem

Although X Education receives a lot of leads, it has an extremely low lead conversion rate. For instance, only approximately 30 of 100 leads they could gather in a day might actually be converted.

The goal of the business is to find the most promising leads, commonly referred to as "Hot Leads," in order to increase the efficiency of this process.

The lead conversion rate ought to increase if they are successful in finding this group of leads, since the sales team will now concentrate more on speaking with potential leads rather than calling everyone.

Problem Statement

X Education Company's Problem

- You have been asked by X Education to assist them in choosing the leads that have the best chance of becoming paying clients. The business wants you to create a model in which you give each lead a lead score so that leads with higher lead scores have a better chance of converting, while leads with lower lead scores have a lesser chance of converting. The desired lead conversion rate has been estimated by the CEO to be in the range of 80%.

DATA

X Education Company's Problem

We have access to a leads dataset from the past that has about 9000 data points.
The properties in this dataset, which include Lead Source, Total Time Spent on Website, Total Visits, Last Activity, and others, may or may not be important in determining whether a lead will be converted in the end.
The column labelled "Converted" in this instance serves as the target variable and indicates whether or not a previous lead was converted, with 1 denoting conversion and 0 denoting non-conversion.

Proposed Solution

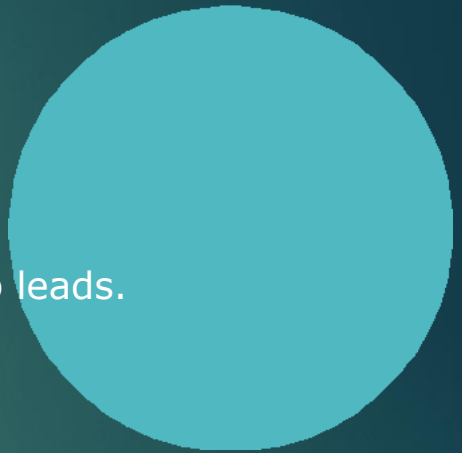
STEP – 1: Hot lead selection:
We have grouped the deals based on the percentage that can be converted, according to the ratio.

STEP – 2: Working on selected leads-
As a result of step 1, we now know which deals we can convert quickly, allowing us to focus on them more effectively.

STEP – 3: Result:-
Starting with step 2, we were able to work more productively with deals that could be quickly converted. Consequently, it raised our conversion rate.

Our Major Problem

Selection of Hot Leads



Finding the best leads is this problem's main challenge.
Our conversion rate can be raised as soon as we can identify the top leads.

Implementation of the solution

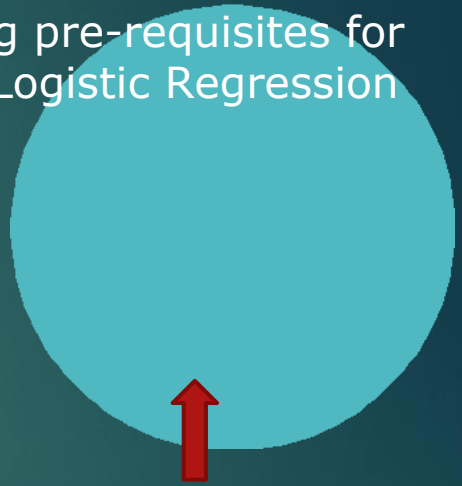
Data Gathering: Loading the datasets

Data Cleaning: Removed unnessary columns and rows

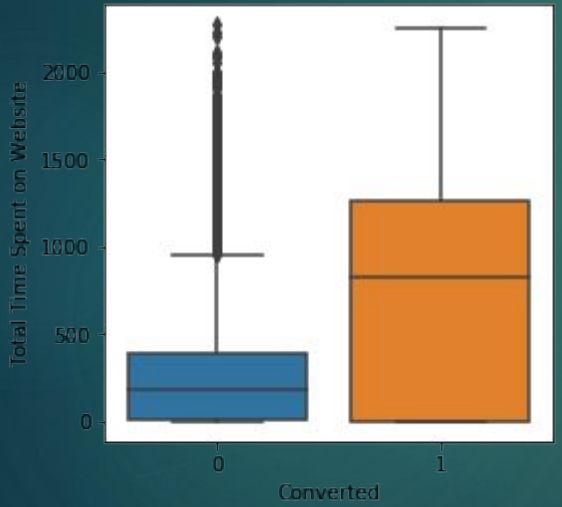
Performing EDA: Univariate and bivariate analysis

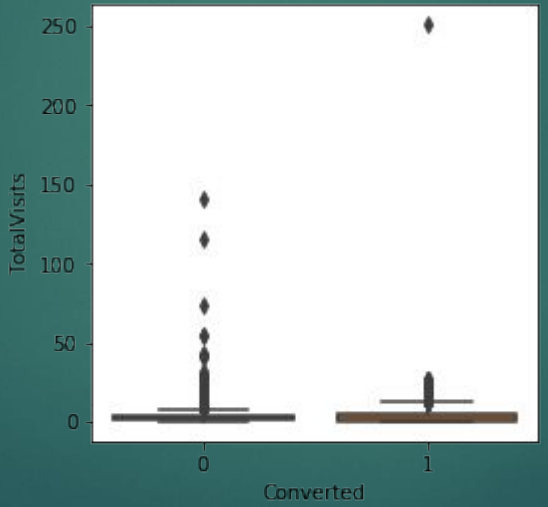
Outlier Treatment and Feature-Standardizatio n

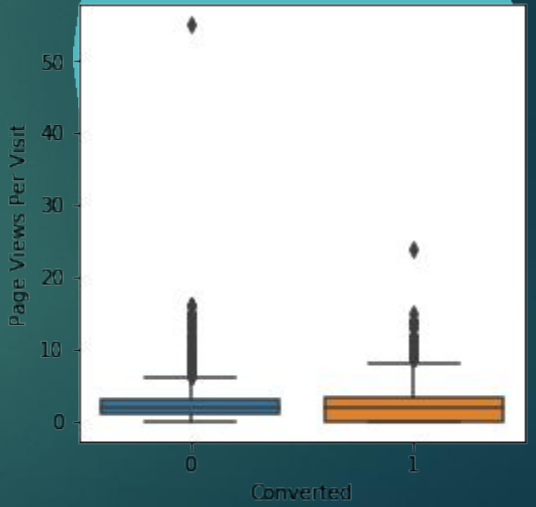
Performing pre-requisites for RFE and Logistic Regression



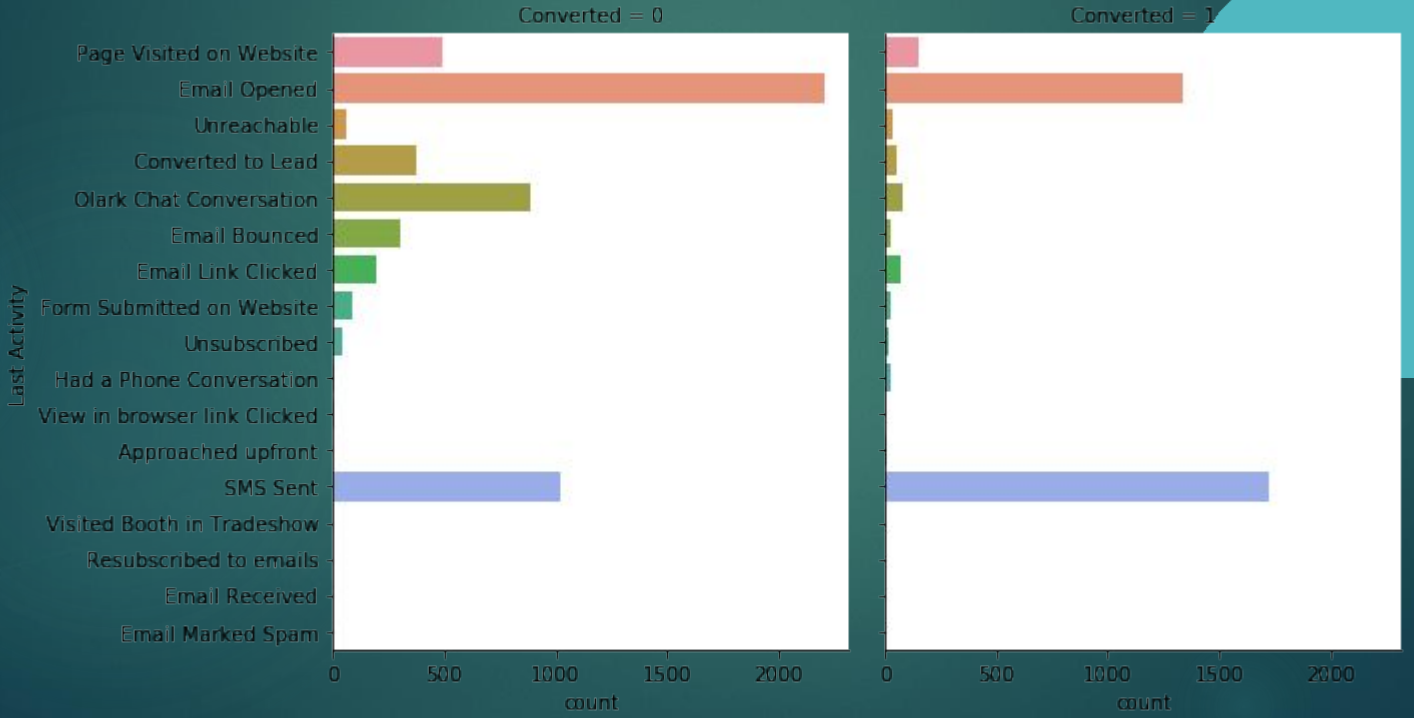
For those who converted and those who did not, numerical columns.



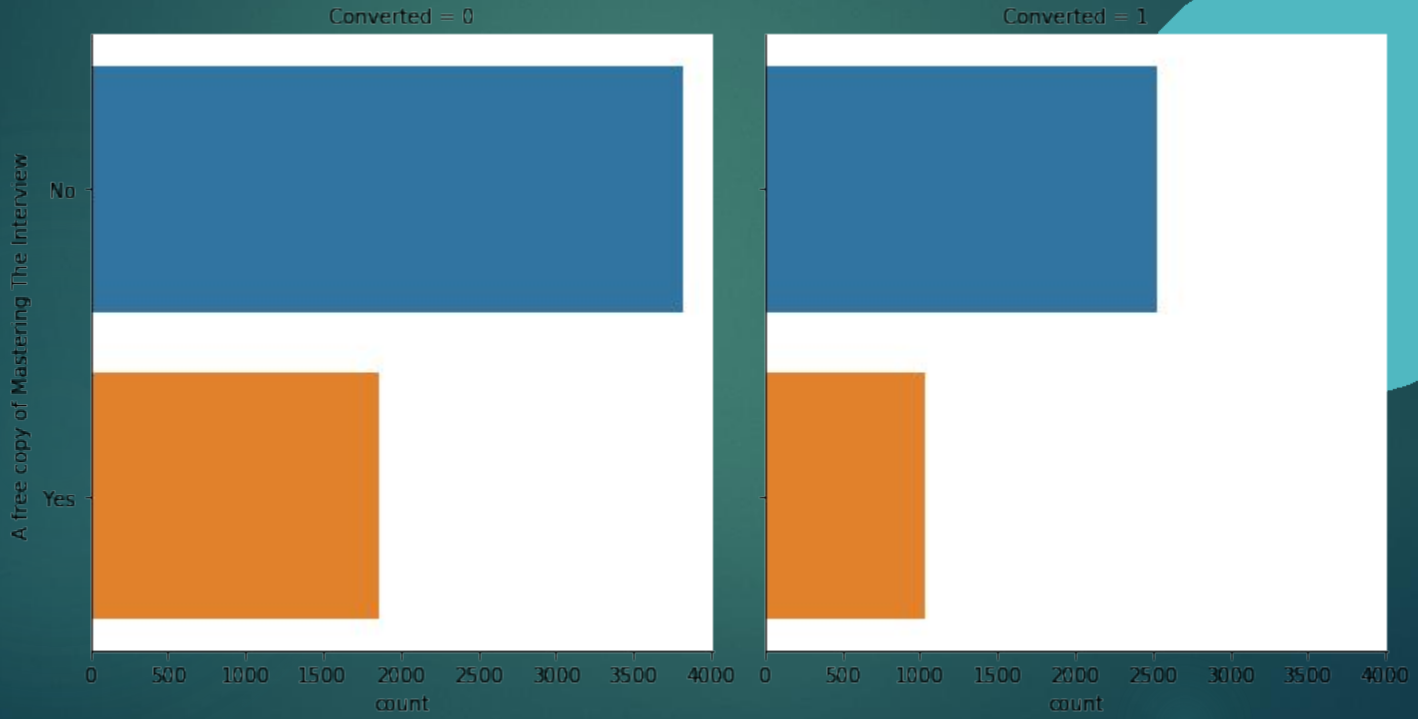




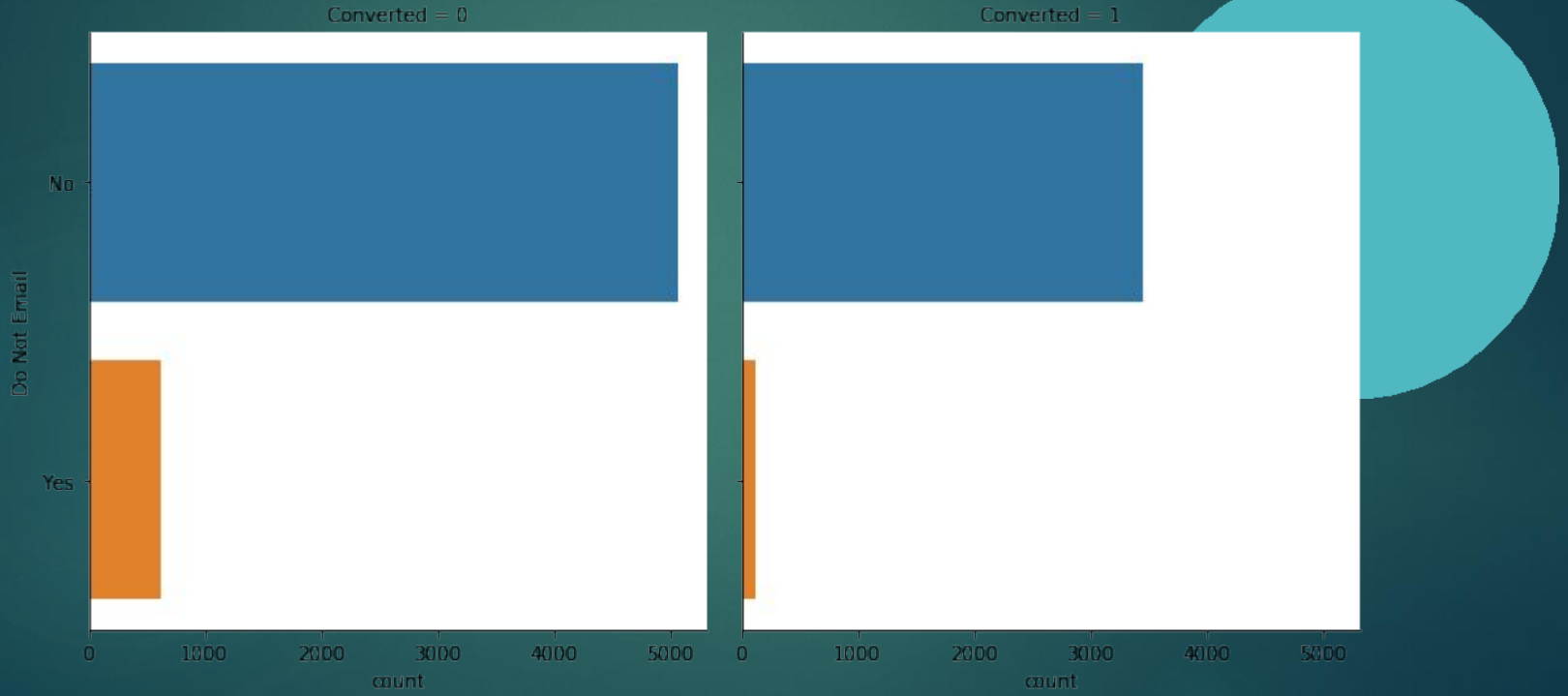
categorical column those who Converted and those who not.

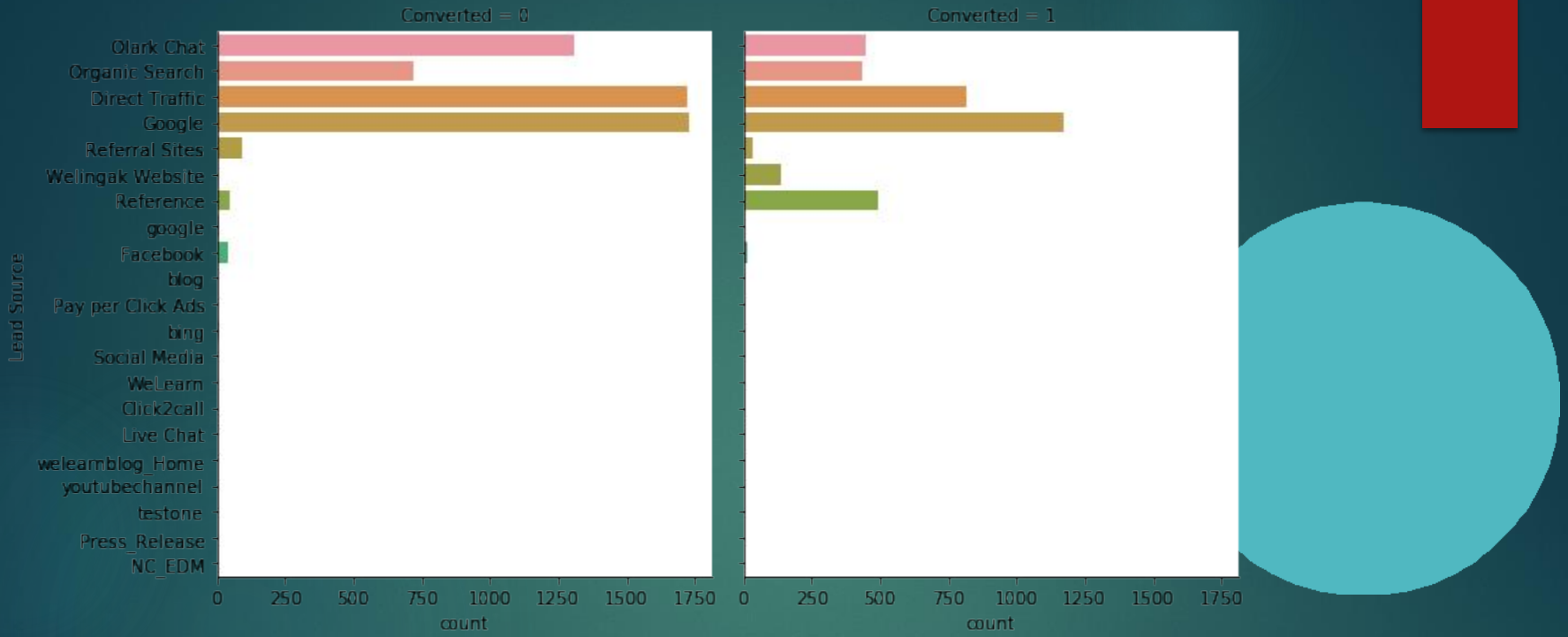


For those who converted and those who did not, there in difference in the category column.



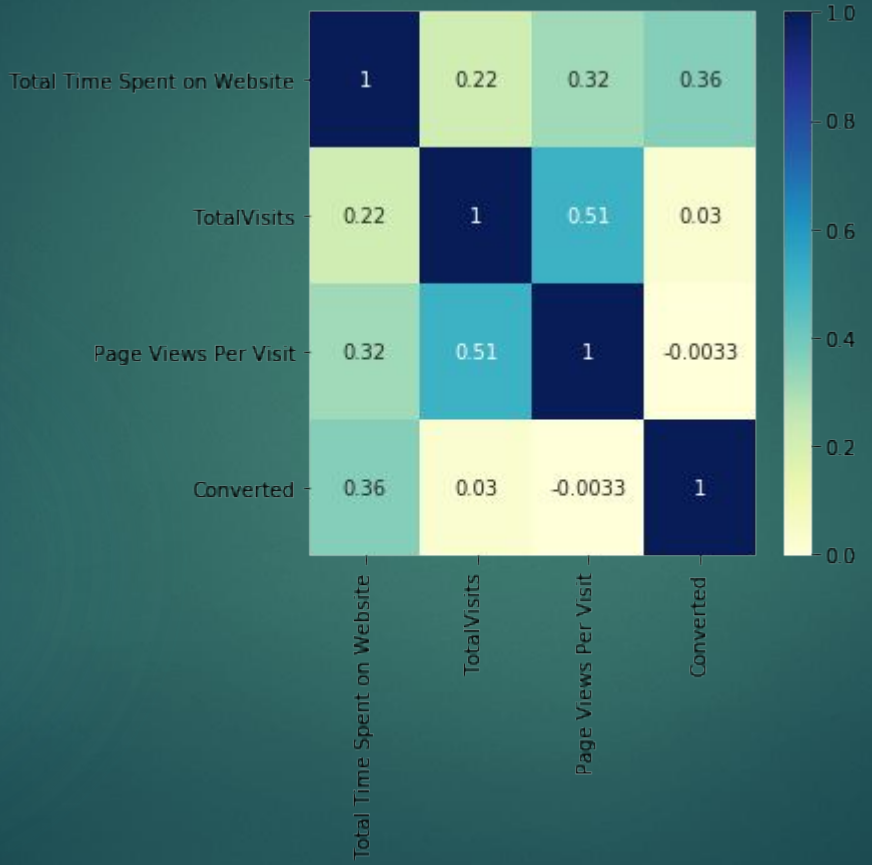
categorical column (Do Not Email) for those who Converted and those who didn't.

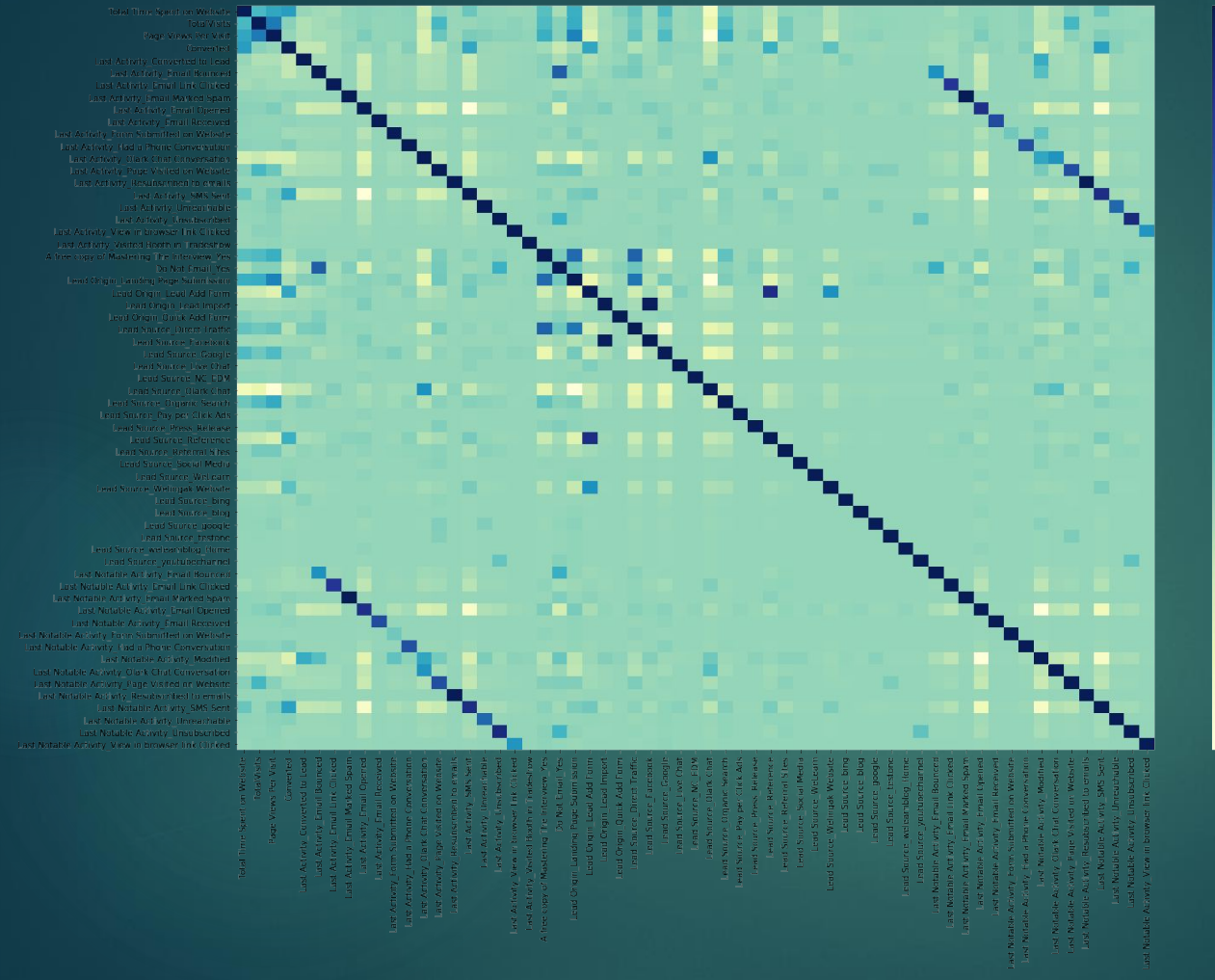




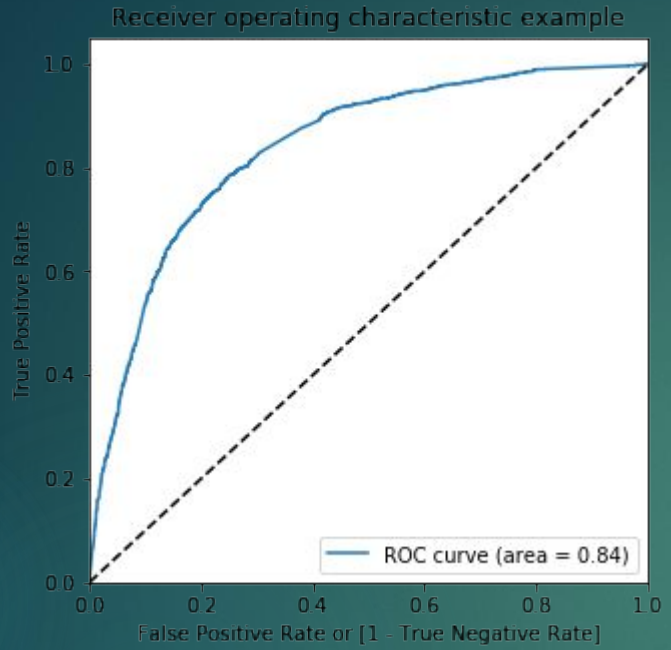
EDA plots depicting variation in categorical column (Lead Source) for those who Converted and those who didn't.

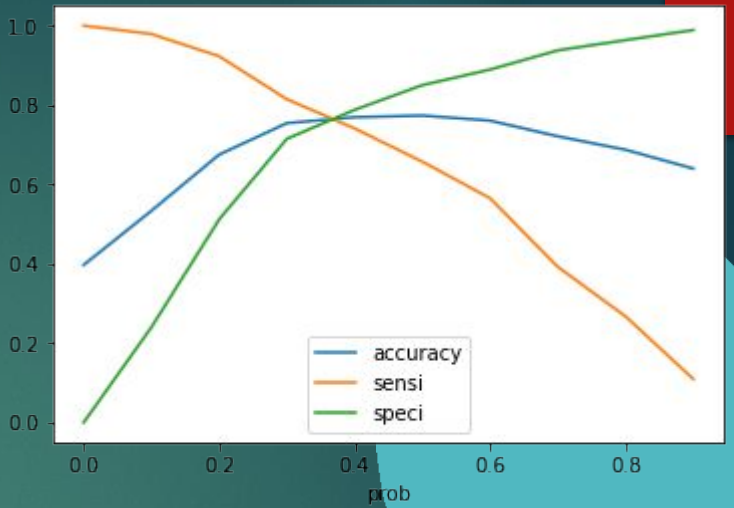
EDA plots depicting correlation (Heat Map) of all selected numerical columns.

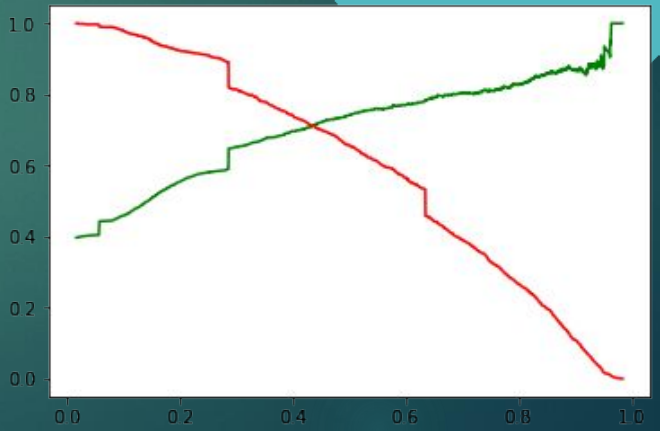




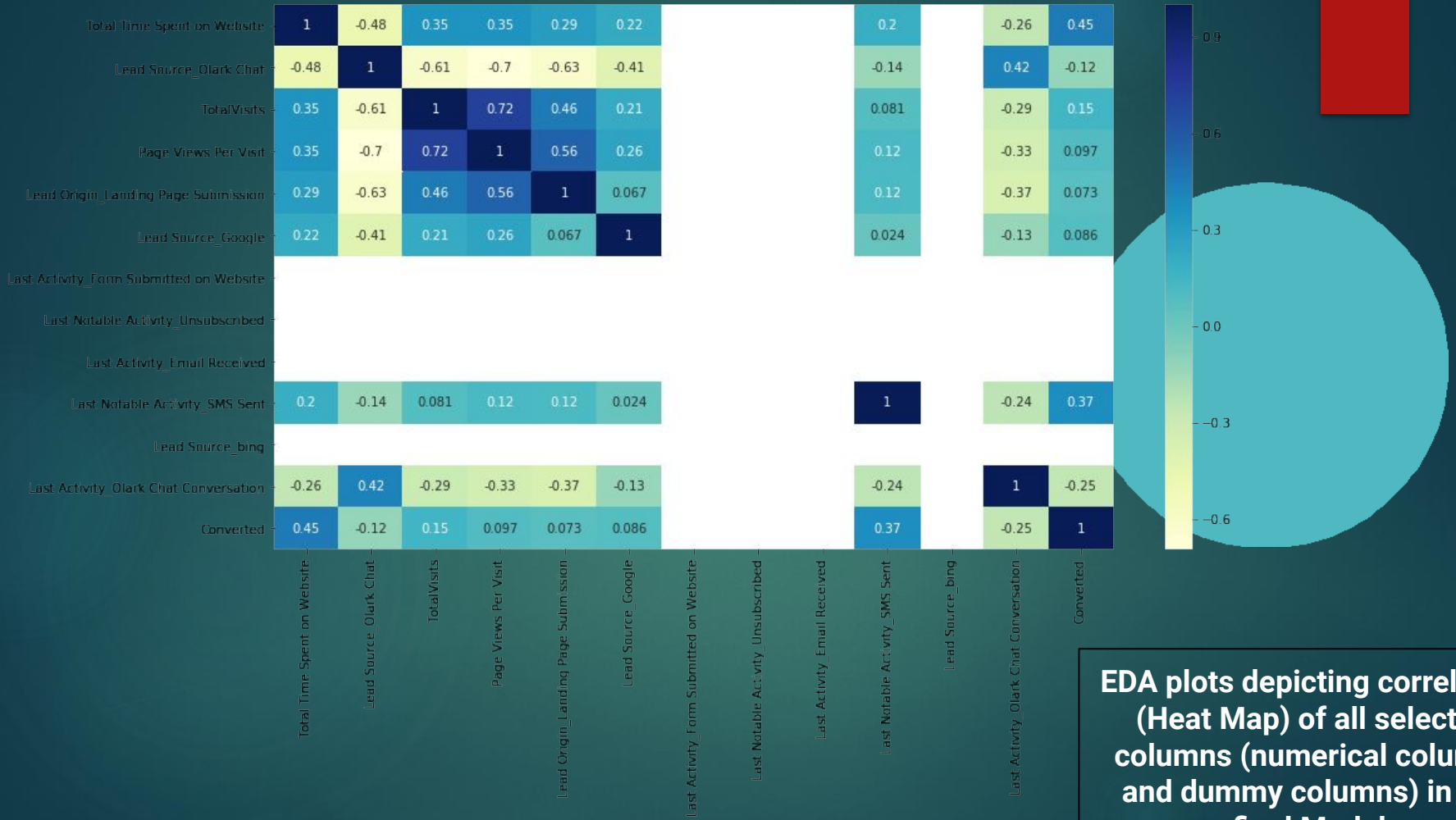
EDA plots depicting correlation (Heat Map) of all selected columns (numerical columns and dummy columns).







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



EDA plots depicting correlation (Heat Map) of all selected columns (numerical columns and dummy columns) in our final Model.

Inferences from Model

Business Insights Derived from our Model

contribute towards lead conversion are:

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

Conclusion 1 (LR Model)

OUR LOGISTIC REGRESSION MODEL IS DECENT AND ACCURATE ENOUGH, WHEN COMPARED TO THE MODEL DERIVED USING PCA
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.

Recommendations

To increase the overall conversion rate, X Education Company must concentrate on the following important factors:

Increase user engagement on the website to increase total visits through advertising, etc.
SHOULD SEND NOTIFICATIONS TO THE END USERS AS THIS AIDS IN HIGHER CONVERSION