LEAD SCORING CASE STUDY

PROBLEM STATEMENT

- X is a company selling online course to professionals
- Company markets their courses in various websites and search engines and tries to sell their courses by making calls
- In the process the company could only convert 30% of the total leads
- Now the company wants to identify the leads as hot leads which have the potential for conversion.
- With this identification they can concentrate on that particular set of leads and aim for higher conversion of leads and in-turn business.

BUSINESS OBJECTIVE

- X Company wants to identify the potential leads and build a model using which identifies the hot leads
- X Company wants to use the leads for getting higher conversion rate and wants to use the model for future identification of leads.

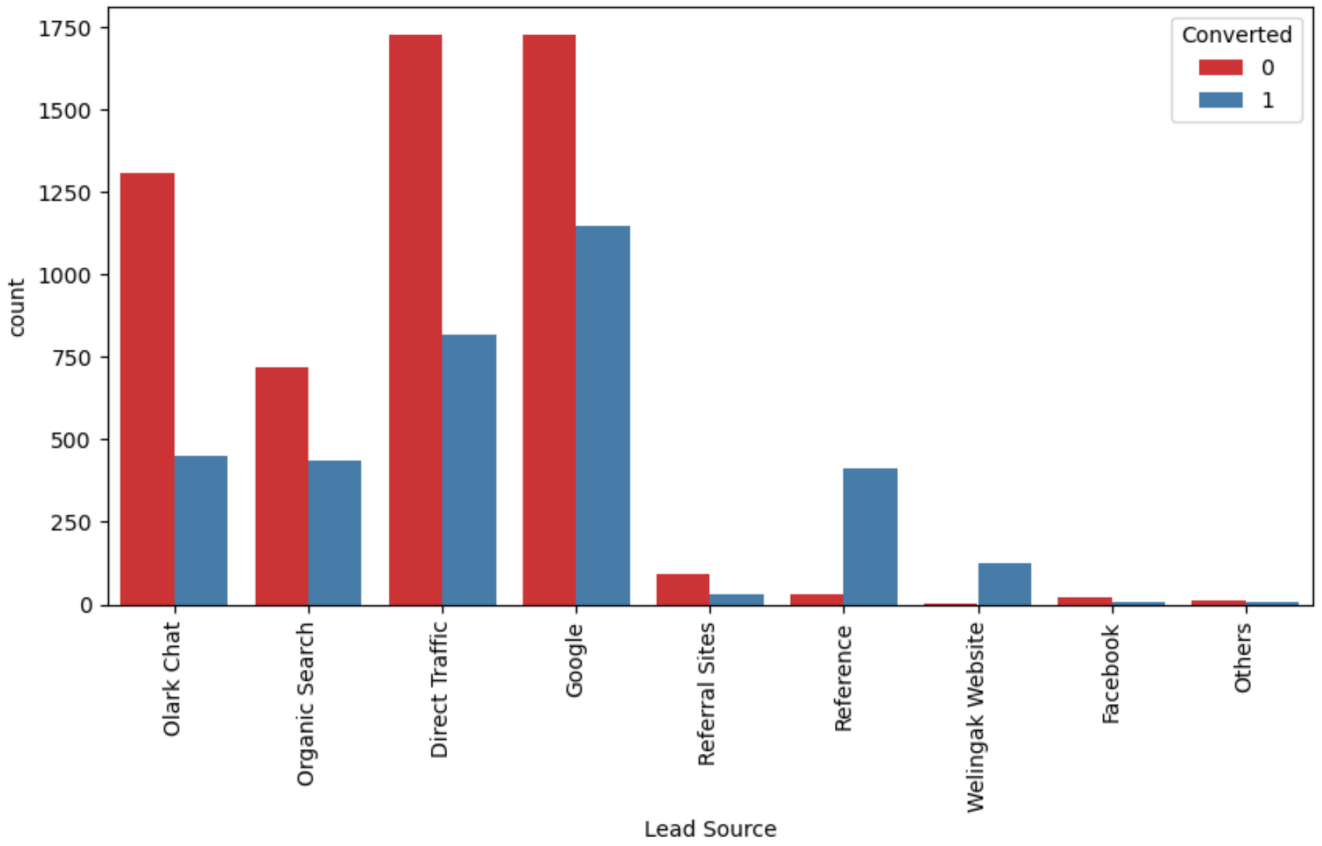
- And the GOAL is to attain 80% convertion rate of the leads using the model

IMPLEMENTATION
Data Cleaning

- The dataset is having 9240 columns and 37 rows
- Firstly, identifying the values having “Select” as input and converting the values as null or any other best possible value
- Finding out the null values in the columns and deleting the columns if the % of null values is more than 40.
- Finding out the missing values and imputing the best fit value
- “What matters most to you in choosing a course” column is deleted as the data is irrelevant for us.
- Deleted the rows which has null values after the data Cleaning.

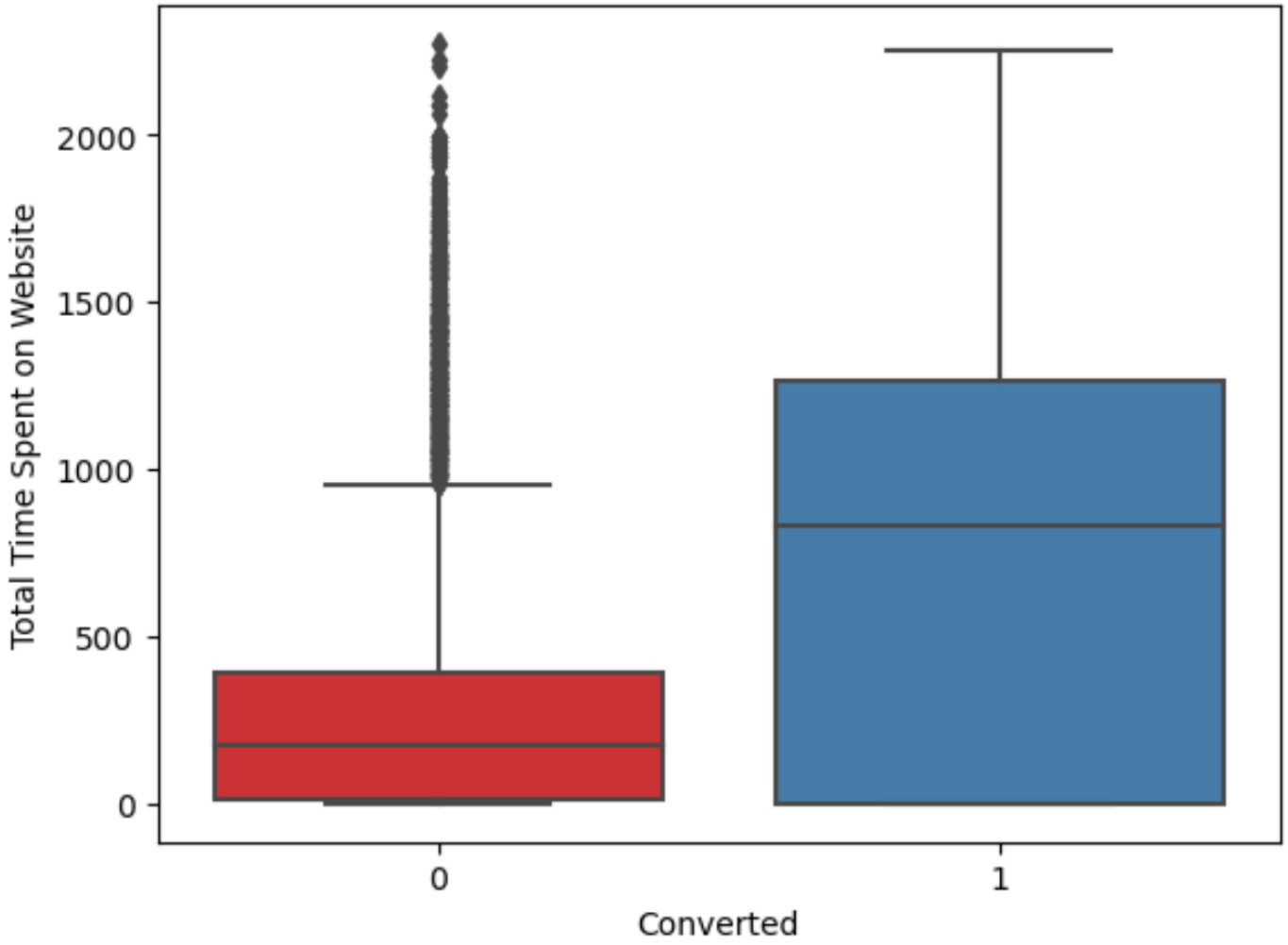
EDA Univariate Analysis – Lead Source

- Leads to be increased in Reference and Welingak Website and conversion rate has to be increased in Olark Char organic search Direct traffic and Google



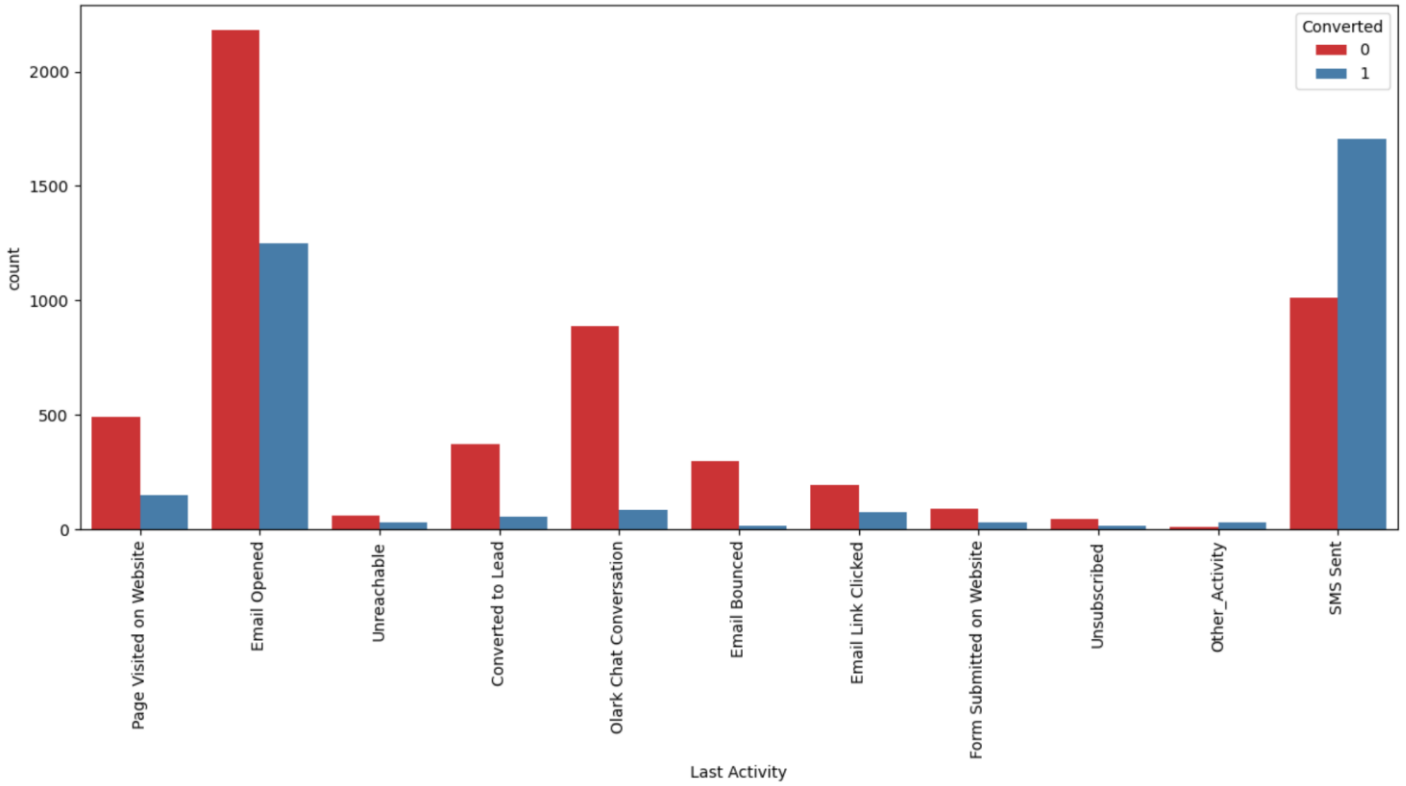
EDA Univariate Analysis – Total Time Spent on Website

- Leads conversion rate is more if the time spent is more in website



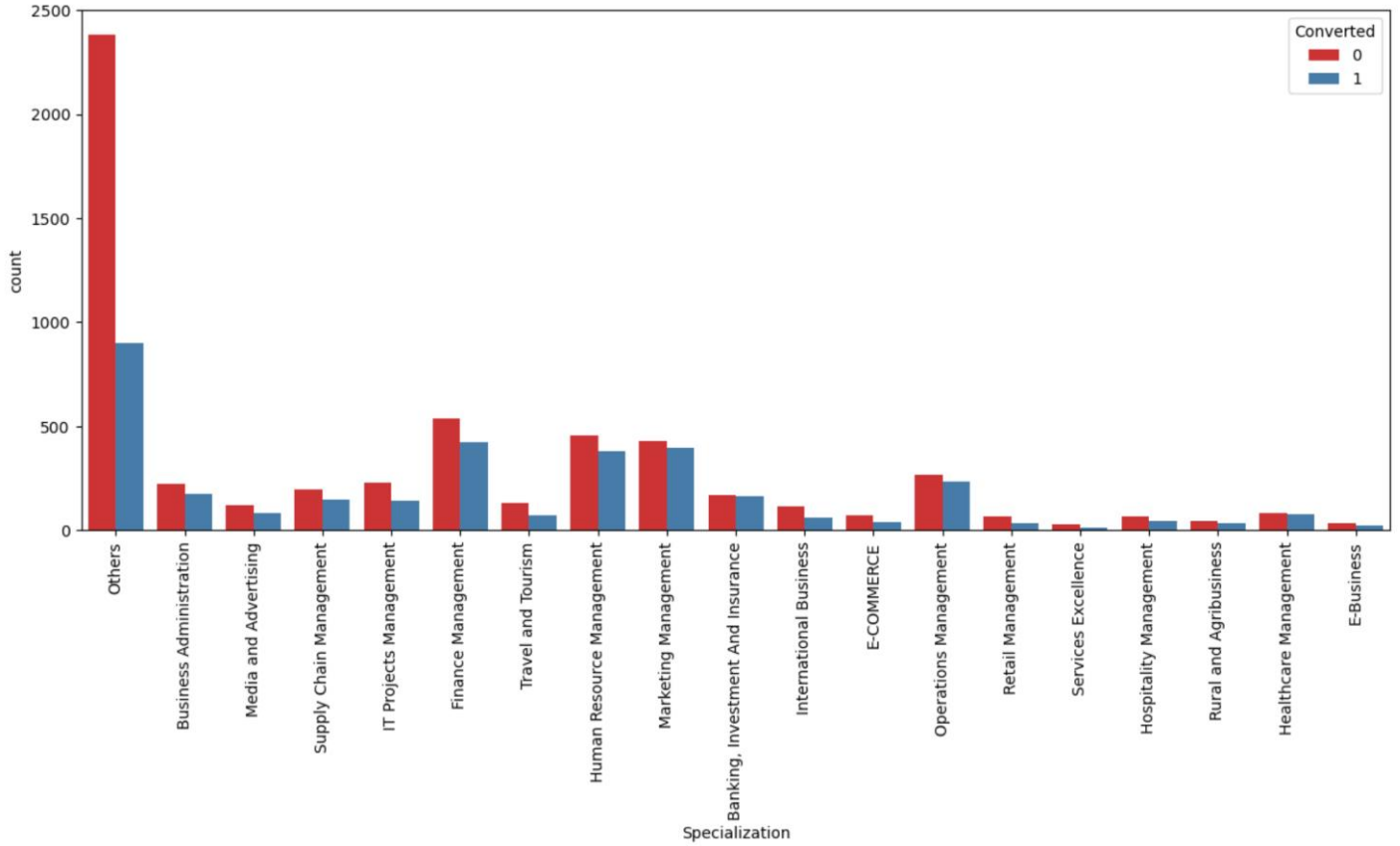
EDA Univariate Analysis – Last Activity

- Convertion rate for SMS as last activity is high and Conversions has to increased for Email Opened.



EDA Univariate Analysis – Specialization

- Conversion rate is high in HRM Mark Management ,Finance Management and Operations Mangement



EDA Analysis Conclusion

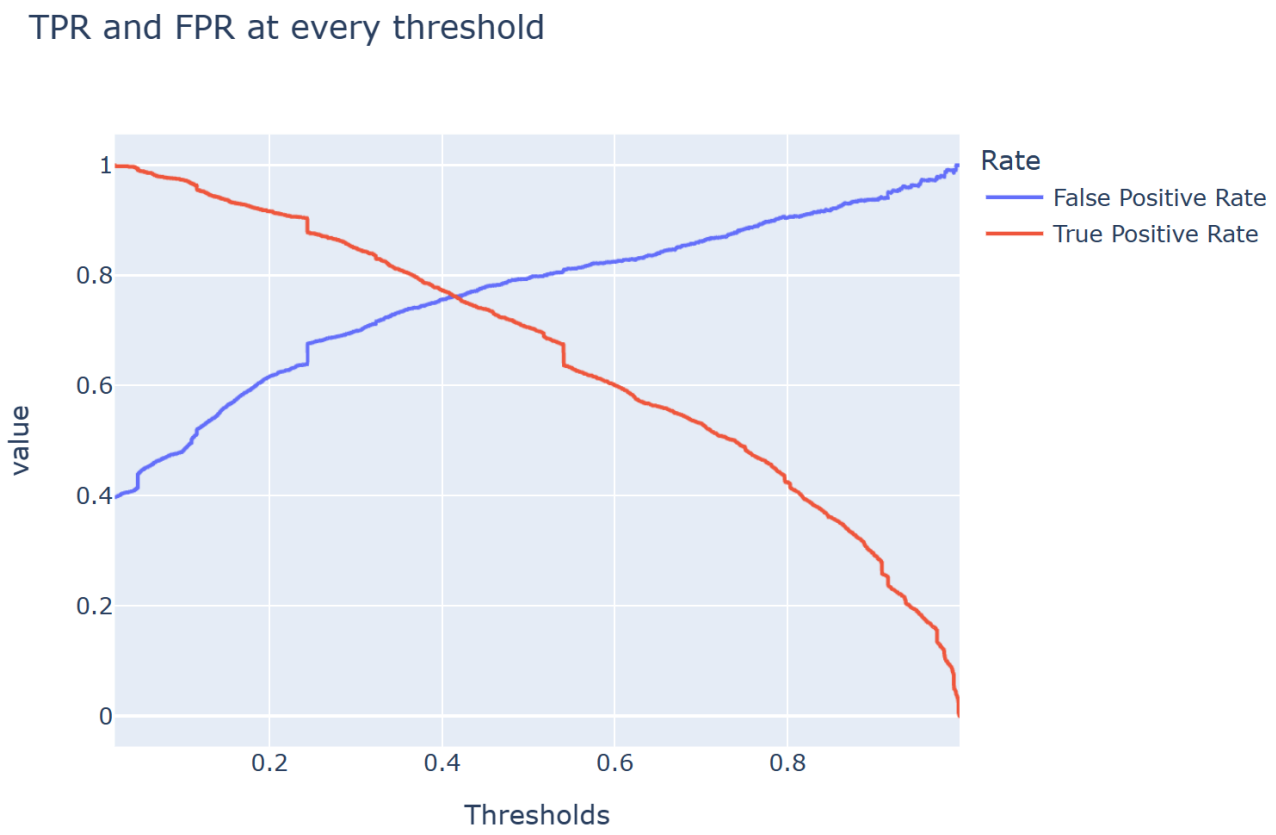
- Based on the Univariate EDA analysis we can infer that many of the variables are not adding any value for our analysis and can be deleted for further better analysis. The variables which we have deleted are
- Many of these variables yielded no inference on plotting and visualizing.
- The variables Lead Number', 'Tags', 'Country', 'Search', 'Magazine', 'Newspaper Article', 'X Education Forums', 'Newspaper', 'Digital Advertisement', 'Through Recommendations', 'Receive More Updates About Our Courses', 'Update me on Supply Chain Content', 'Get updates on DM Content', 'I agree to pay the amount through cheque', 'A free copy of Mastering The Interview' are dropped.

MODEL BUILDING AND EVALUATION

- After scaling and creating dummy variables on the data , we prepared the train and test data as 70 : 30 ratio.
- Recursive Feature Elimination technique is applied on the remaining data and removed variables whose P Value is > 0.05 and VIF is >5.
- After the model is built we calculated Accuracy , Sensitivity and Specificity and the overall accuracy comes to 81%.

MODEL EVALUATION

- Plotted Trade off curve and created confusion matrix for the data.
- Train Data
  - Accuracy – 81.3%
  - Sensitivity – 81.1%
  - Specificity – 81.5%
- Test Data
  - Accuracy – 80.8%
  - Sensitivity – 80%
  - Specificity – 81.2%



OBSERVATIONS

- Hot leads turns out to by 368 and the goal of 80% is achieved and the same can be conveyed to business team that using the model built the conversion rate of 80% can be achieved by concentrating on the Hot leads rather than call the existing leads.

RECOMMENDATIONS

- Dos
  - Calls to be made to the leads whose last actvity is as "SMS sent" , and who spent "more time on the websites"
  - Calls to be made to leads coming from "Olark Chat" and whose profession is "Working professionals"
  - Calls to be made to leads coming from "Welingak Websites" and from "refernces" as the possibilty of convertion is more

- DONTs
  - Calls to be curtailed whose last activity is "Olark Chat Conversation" and leads whose chose option as "Do not Email"
  - Calls to be curtailed whose lead origin is "Landing Page submission" and specialization is "Others as converasion possibility is less