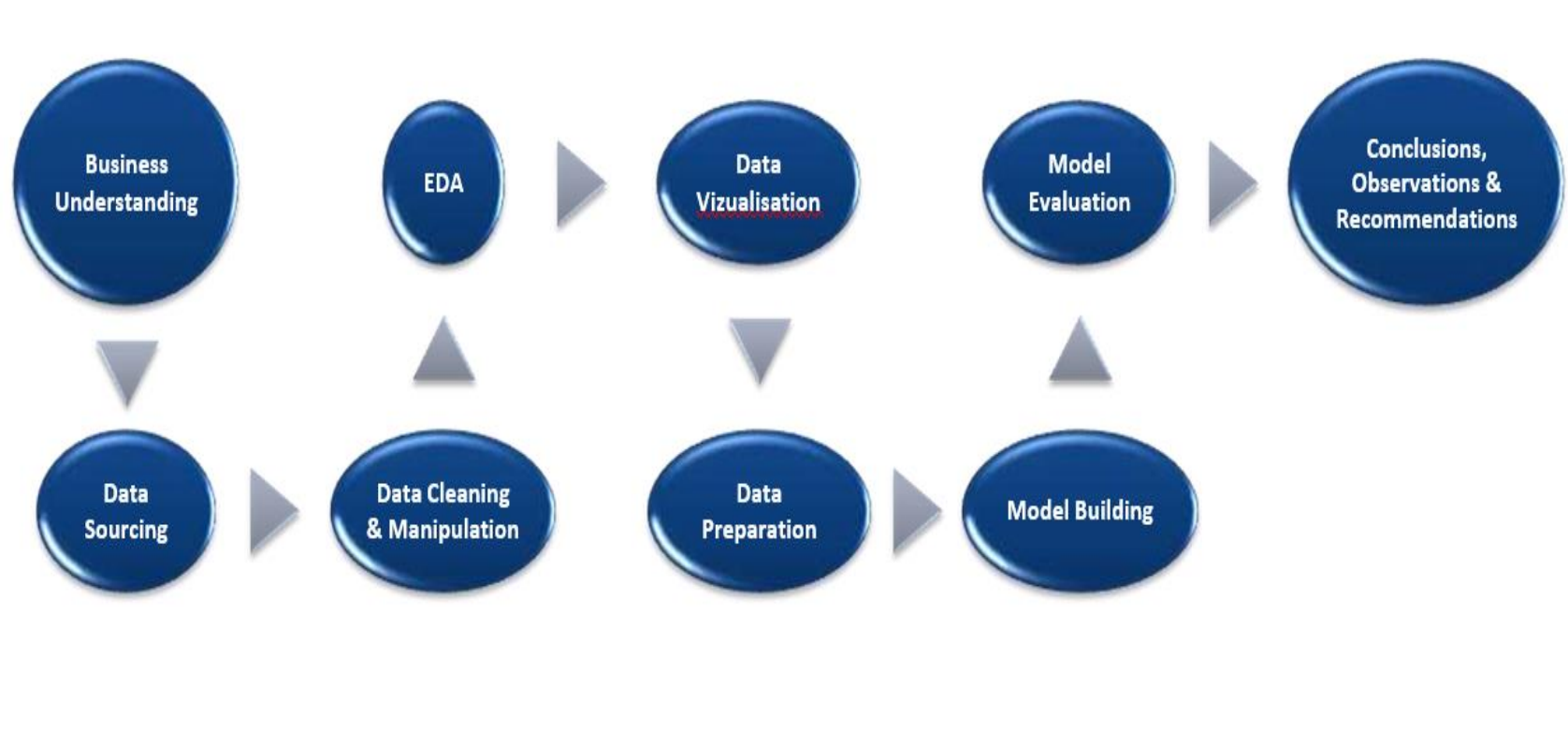
Business Problem

- **An education company named X Education sells online courses to industry Professionals**
- **Although X Education gets a lot of leads, its lead conversion rate is very poor**
- **For example, if say, they acquire 100 leads in a day, only about 30 of them are converted.**
- **The objective is to build a model to identify the hot leads and achieve lead conversion rate to 80%.**

Business Objective

The Business Objective Is To Build A Logistic Regression Model To Identify The Hot/Potential Leads And Achieve The Lead Conversion Rate To 80%."

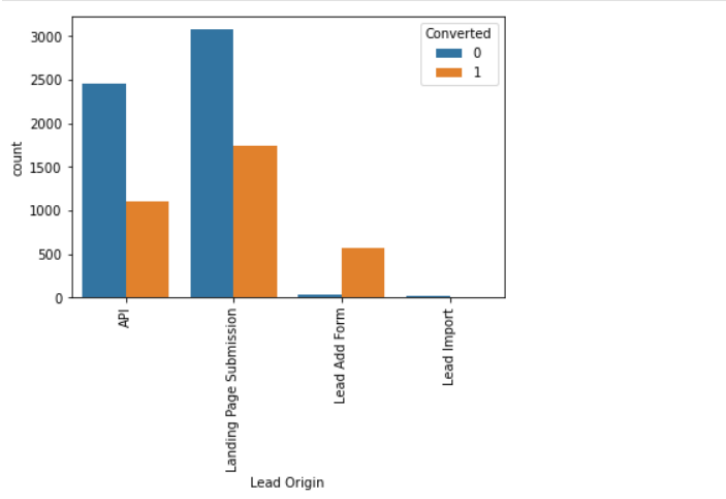
Analysis and Modelling Approach

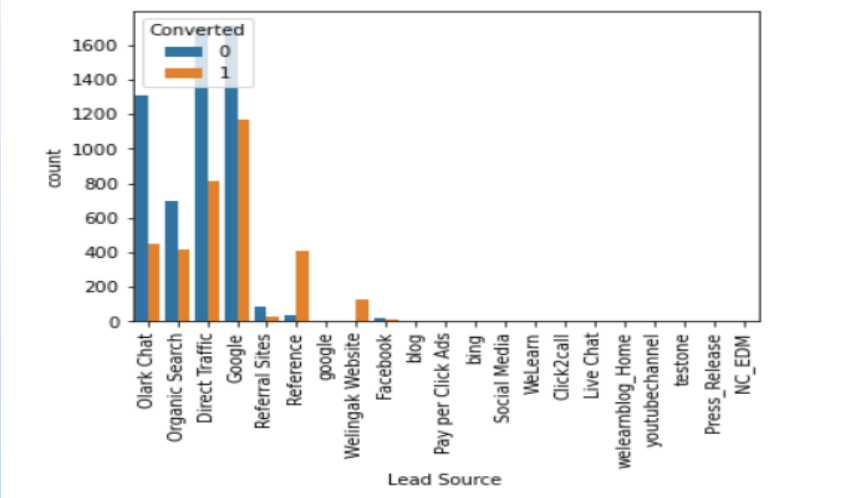


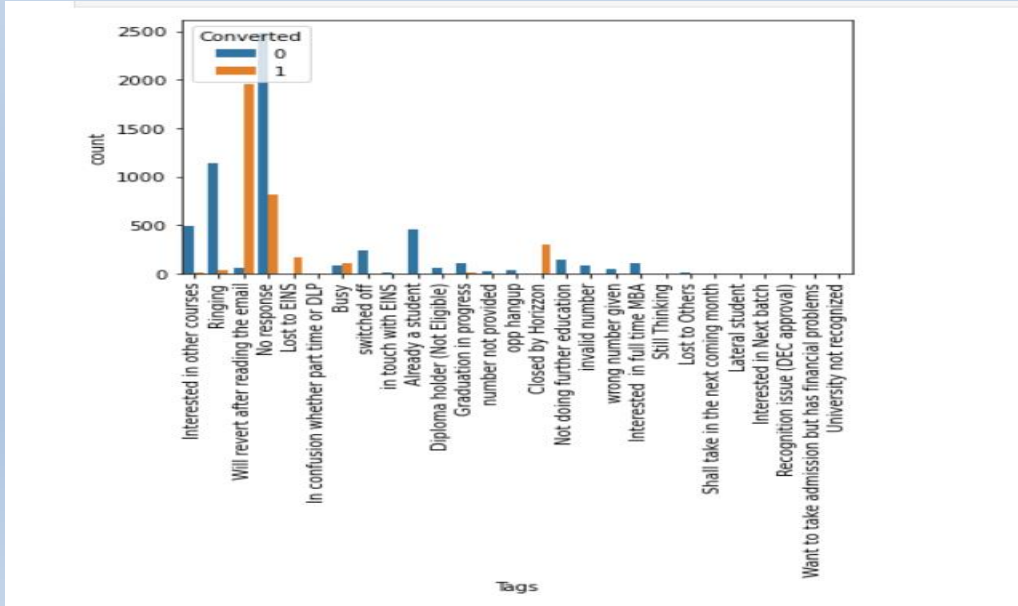
Data Understanding & Cleaning

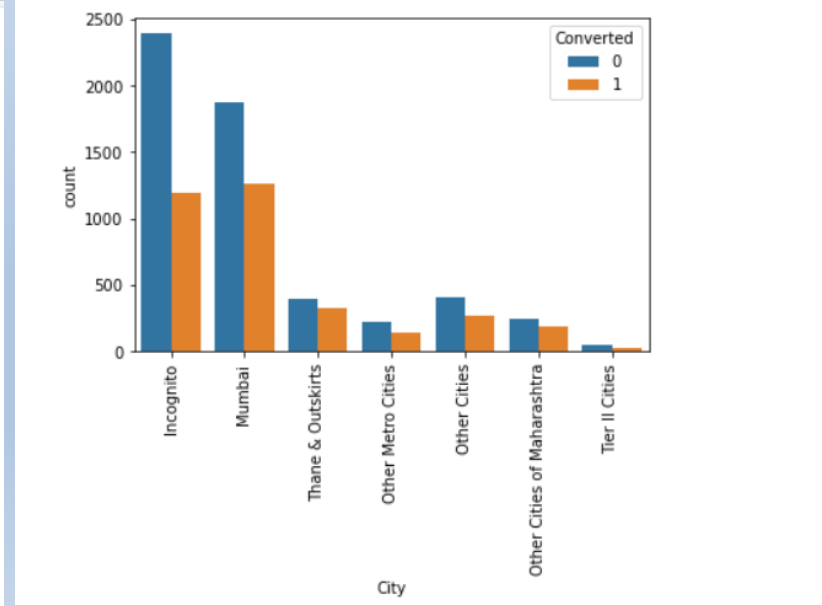
- We have 9240 Rows and 37 columns.
- Out of 37 features , 2 are ID's , 5 are continuous and 30 are categorical including binary.
- Target Variable "Converted" is binary.
- Select in multiple column replaced with 'Nan'. Later columns with more than 60% null values are dropped.
- Imputation of null values in many of the cases is done with a new category "Others" to maintain skewness of the data.
- In continuous feature the imputation is done using median.

Exploratory Data Analysis

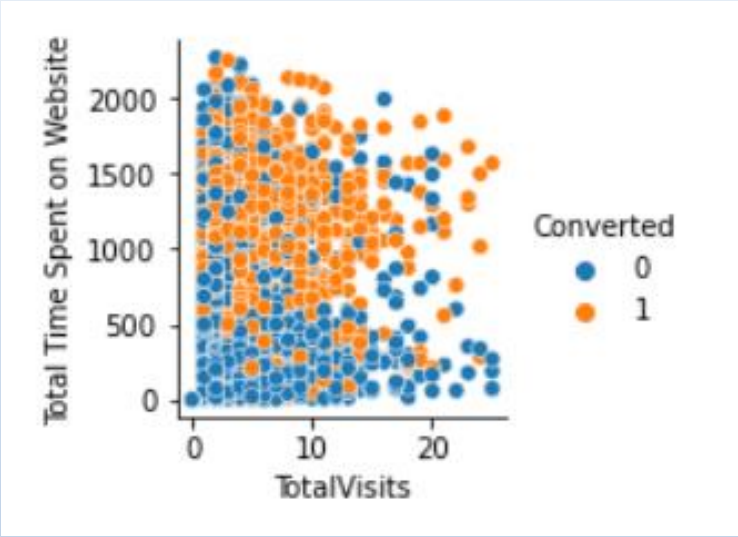


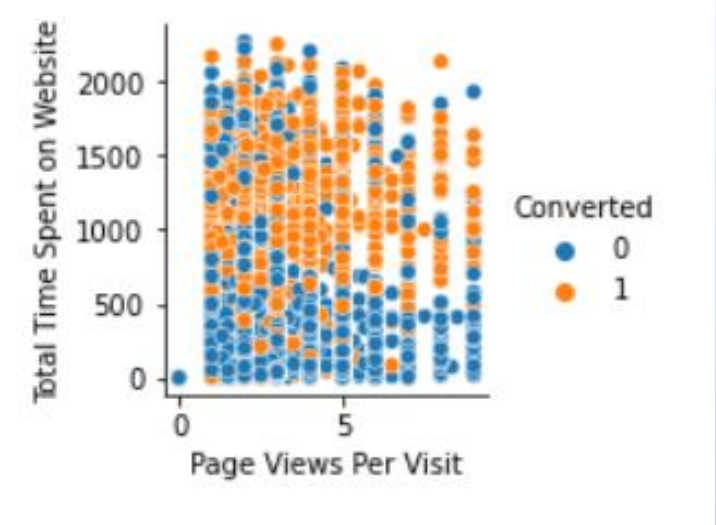


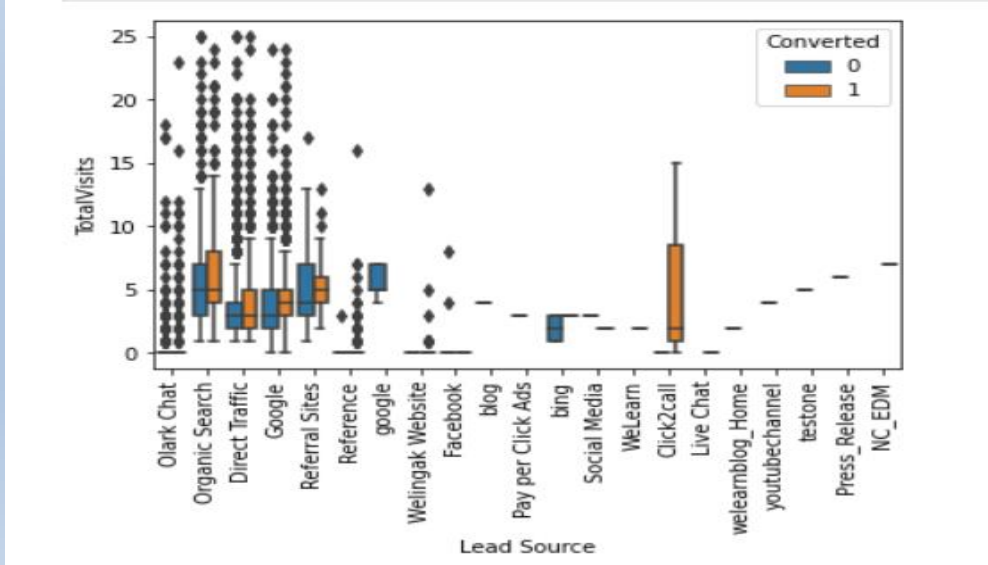




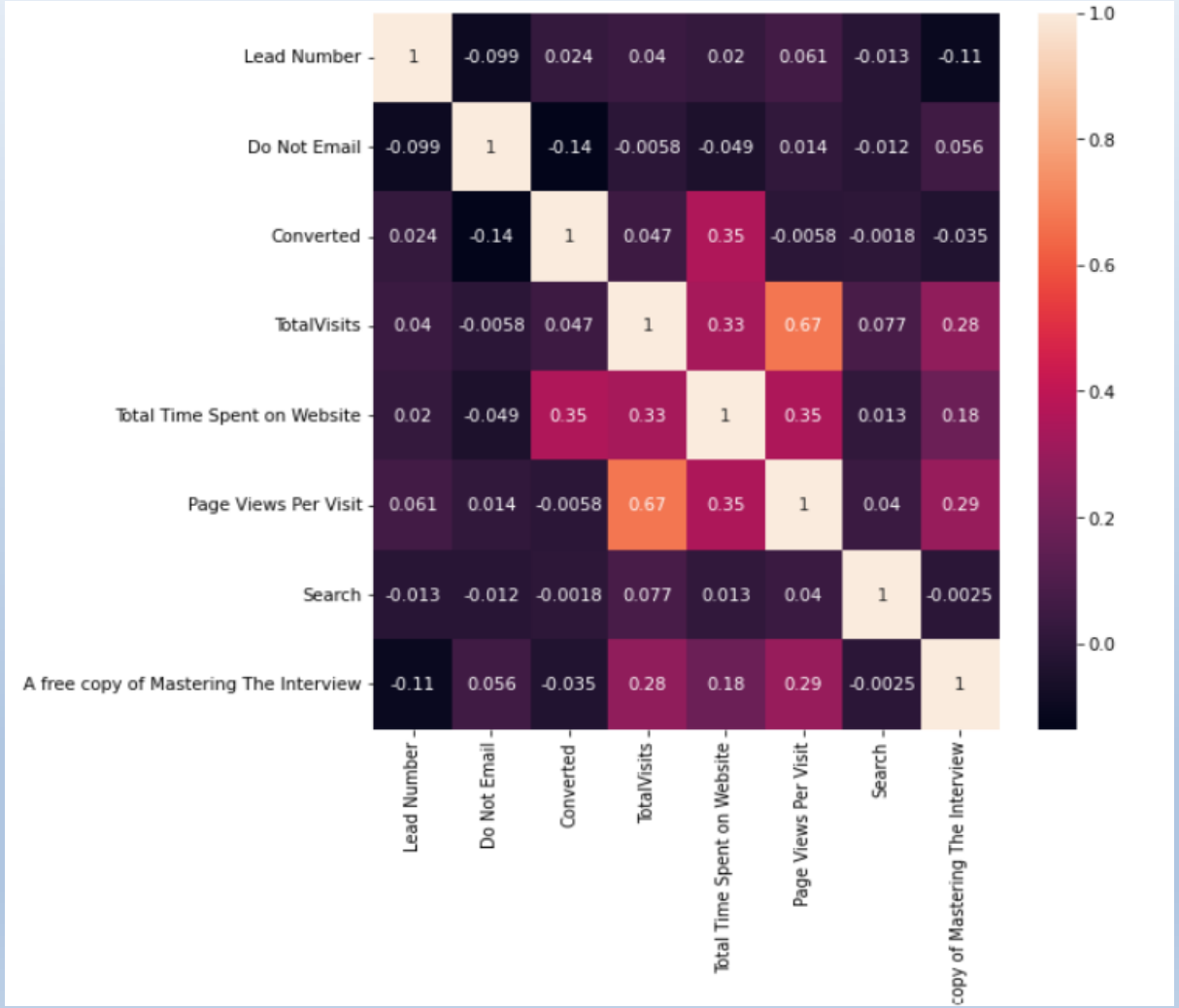
Exploratory Data Analysis



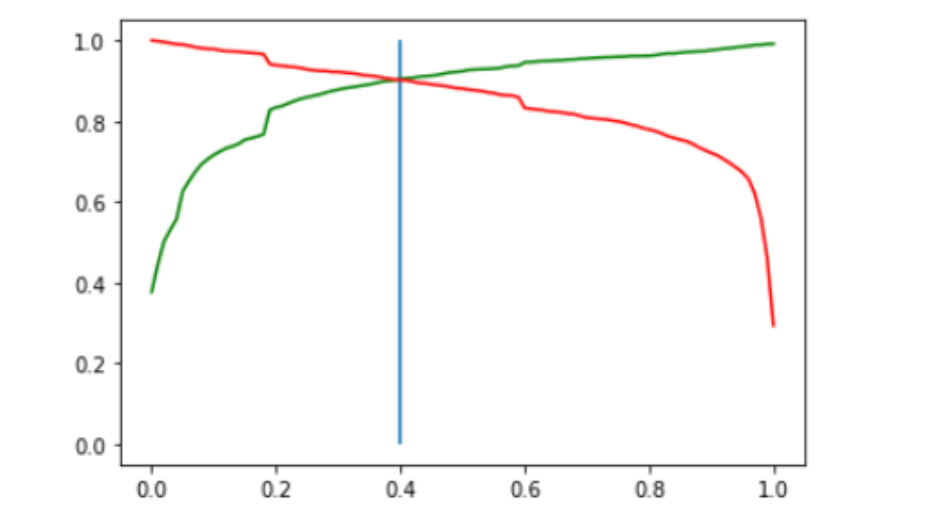




Exploratory Data Analysis



Modelling



Logistic Regression model Building with threshold 0.4

Precision Recall diagram

Train set performance

accuracy = 0.9276993816394482
Precision = 0.9063955950868277
recall = 0.9010526315789473

Test set performance

accuracy = 0.9175295857988166
Precision = 0.9000969932104753
recall = 0.8854961832061069

Conclusions

- Logistic Regression was used to score the leads between 0-100.
- Accuracy score of 0.91. Precision of 0.9 and Recall of 0.88 were achieved which means that the model is identifying the hot leads quite well.
- Optimal Probability Cut-off point was found to be 0.4, which means that any lead > than the score of 40 can be considered a hot lead.
- Our best lead source as identified by the model is the Welingak Website and if a lead spends more time on our website or responds after reading email shows the awareness for the programme and hence, they are more likely to convert.
- If a lead has higher total visits, the probability of them converting is higher.
- If the lead's activity is low and if the lead has asked us to not email them or tag is Worst Quality, this signifies that the lead is less likely to convert.