



LEAD SCORE CASE STUDY

Group Member

Neha Mathur

Priyanka Singh

Problem statement

- X Education sells online courses to industry professionals.
- 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.
- 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.

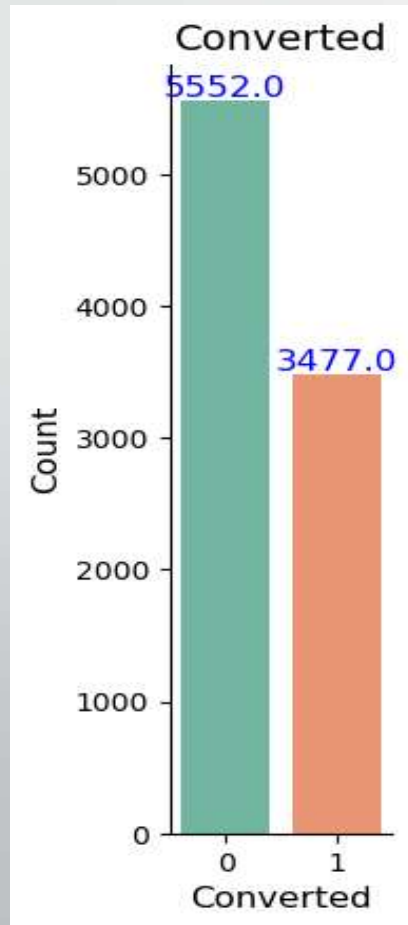
Business Objective:

- X education wants to know most promising leads.
- For that they want to build a Model which identifies the hot leads.
- Deployment of the model for the future use.

Analysis Approach

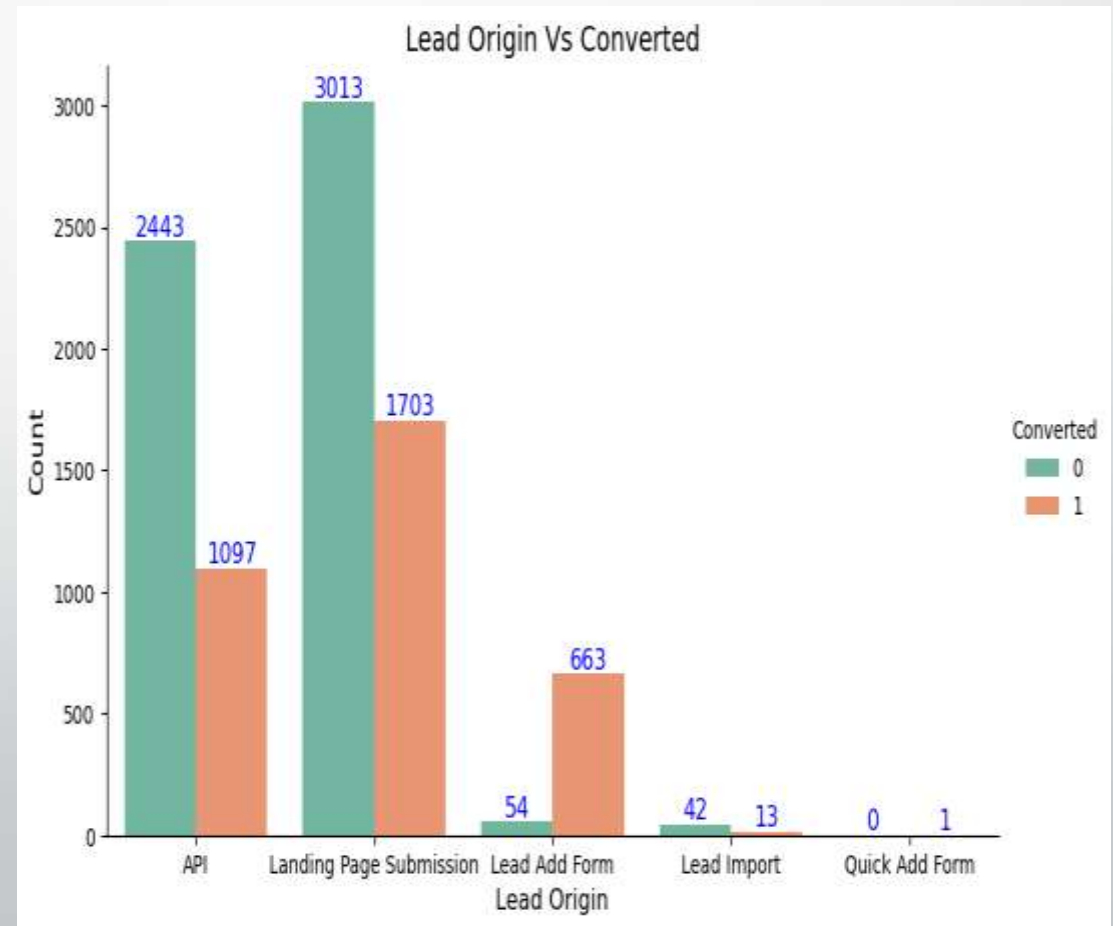
- Reading and Understanding Data.
- Data Cleaning
- Data Analysis
- Creating Dummy Variables
- Test Train Split
- Feature Rescaling
- Feature selection using RFE
- Plotting the ROC Curve
- Finding the Optimal Cutoff Point

- From the above graph, there is a overall conversion rate of around 39%



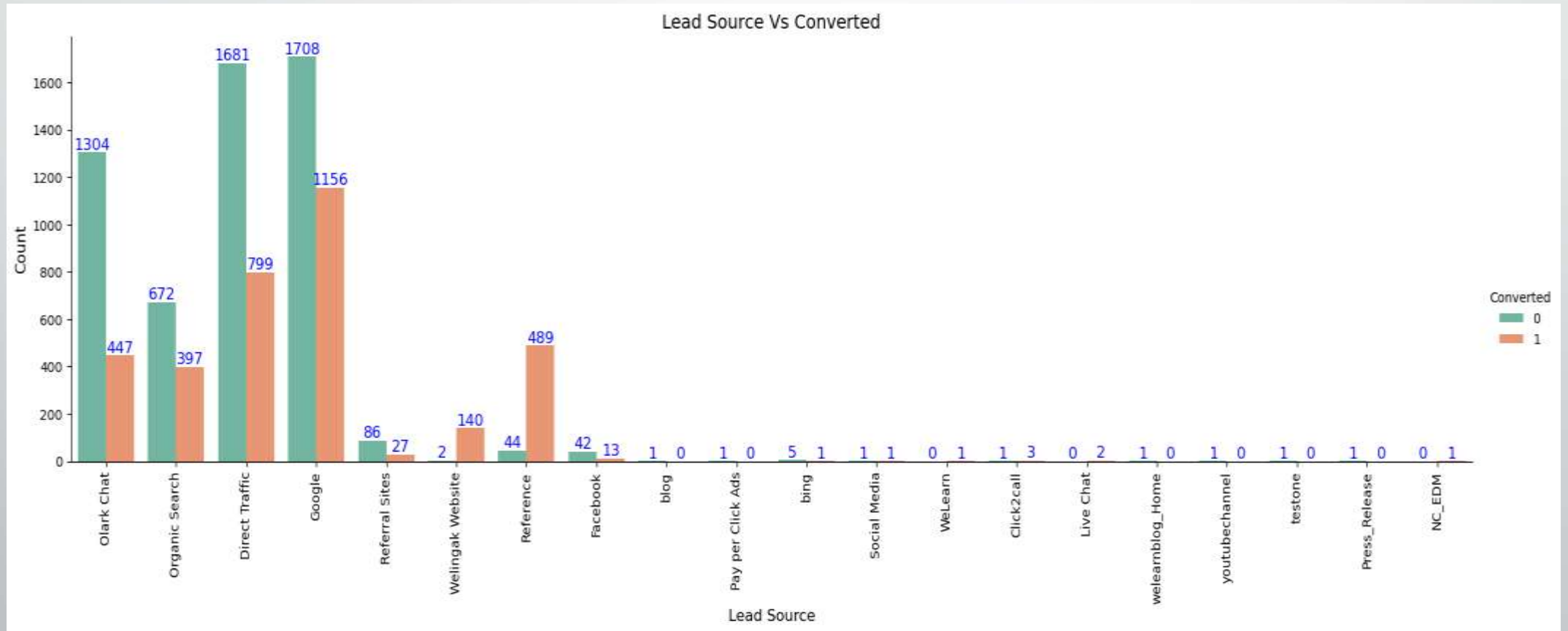
EDA

It can be seen that the maximum conversion happened from Landing Page Submission Also there was only one request from quick add form which got converted.

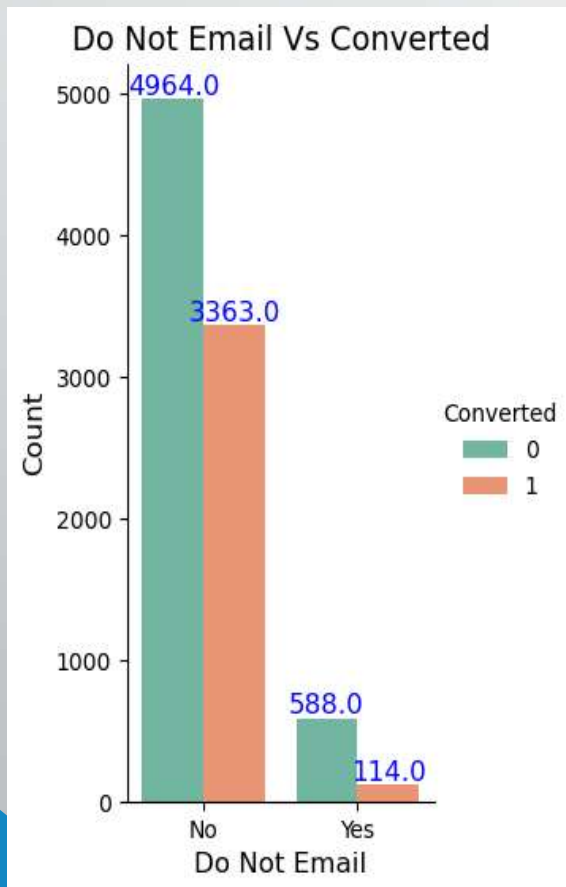


Lead sources

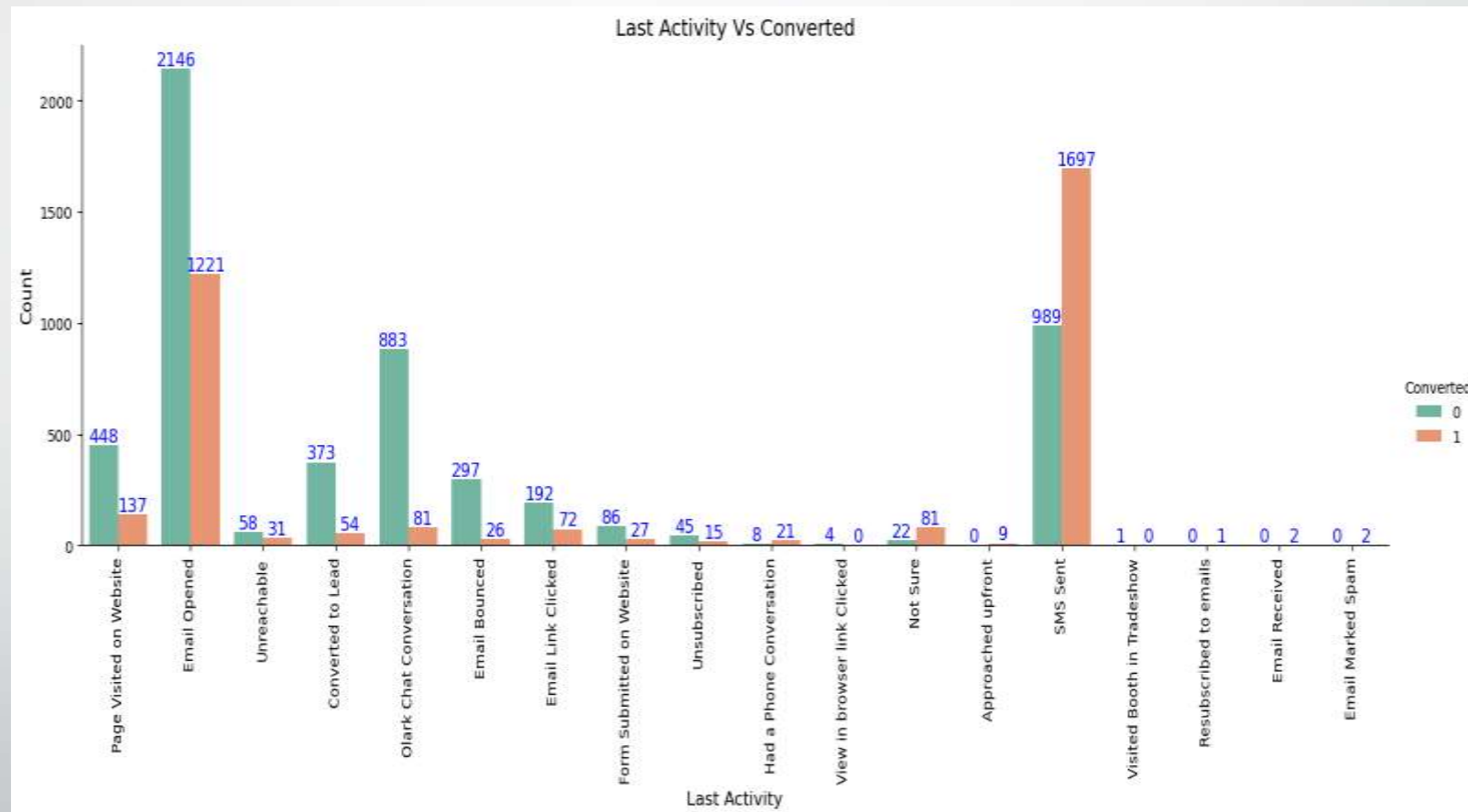
From the above graph, it can be seen that major conversion in the lead source is from google



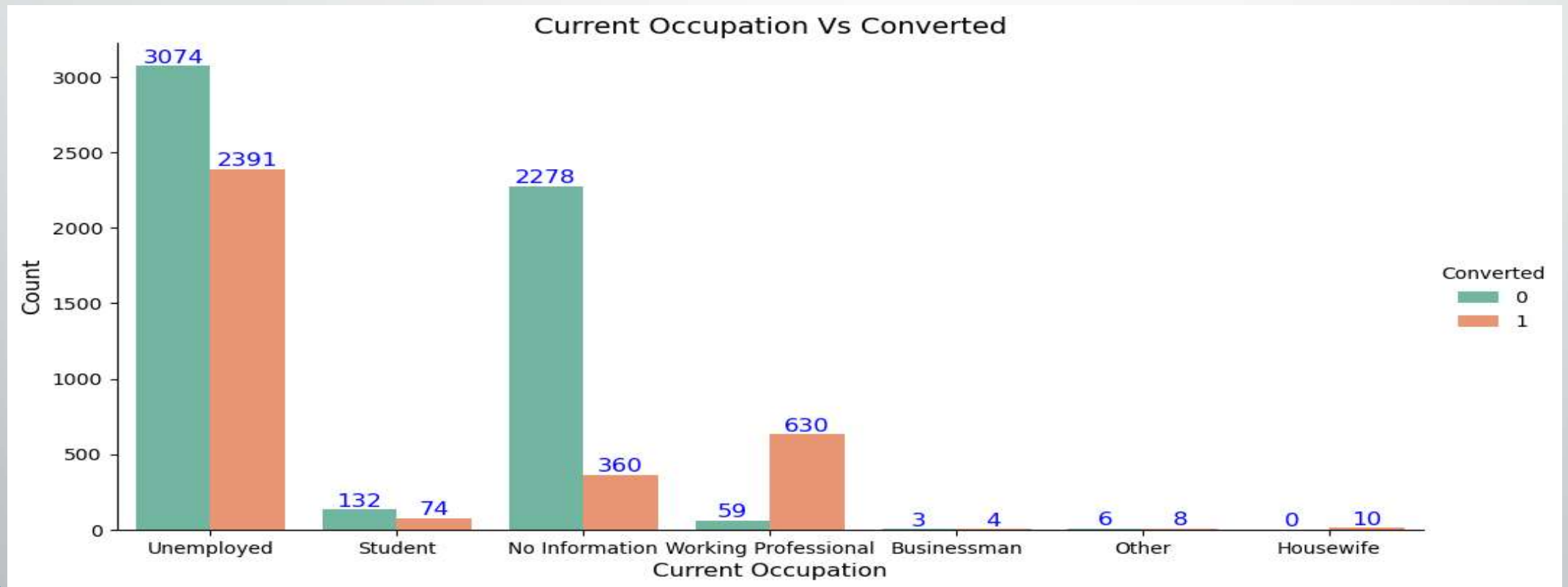
Major conversion has happened from the emails that have been sent

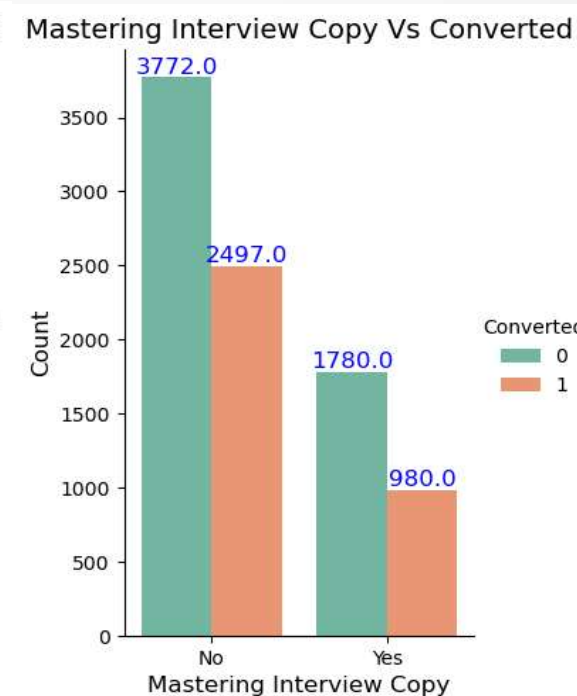
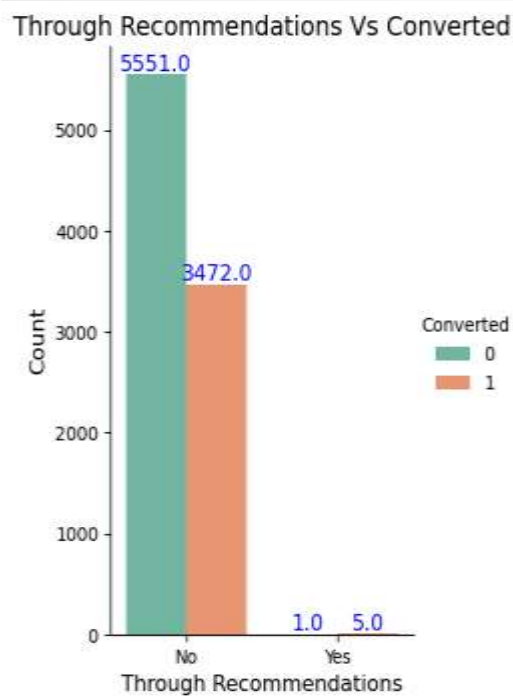
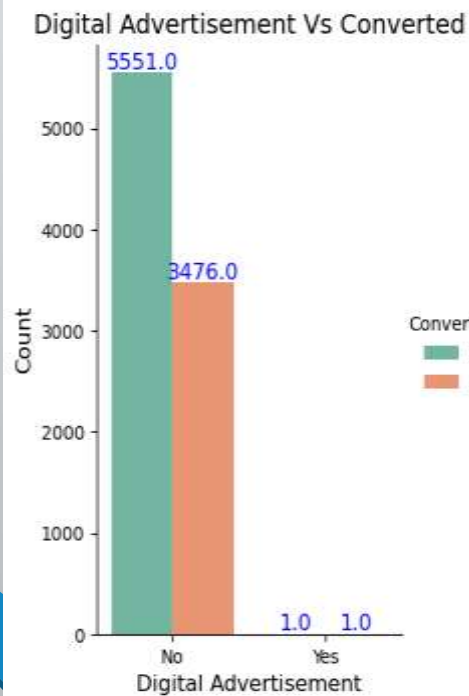
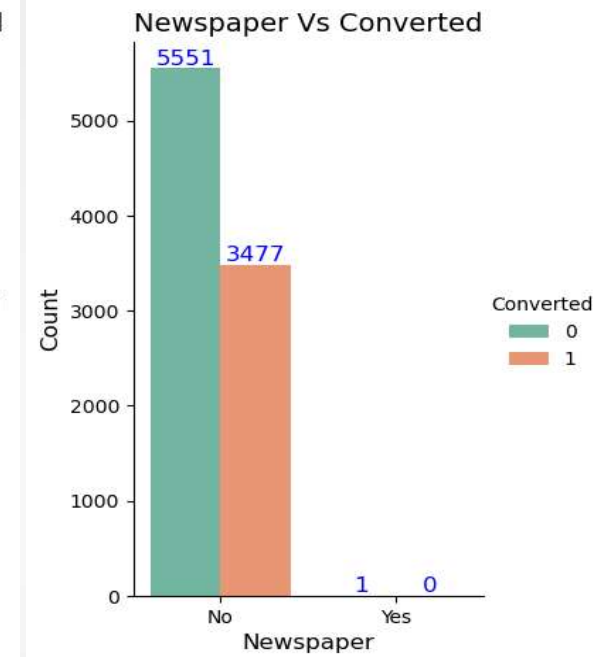
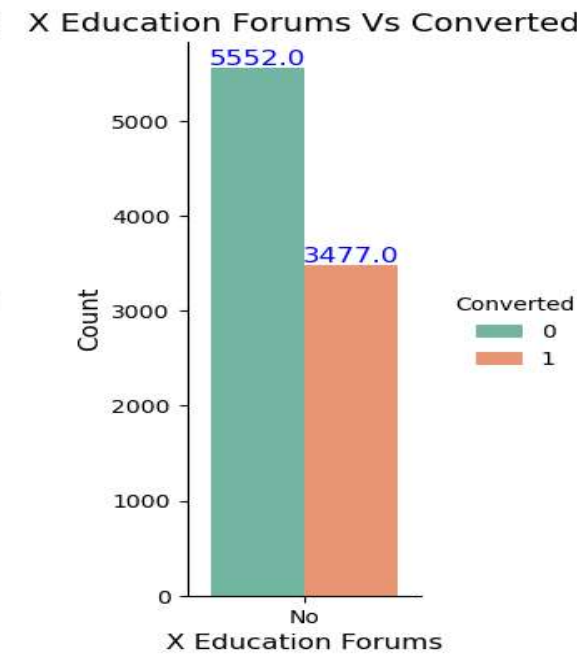
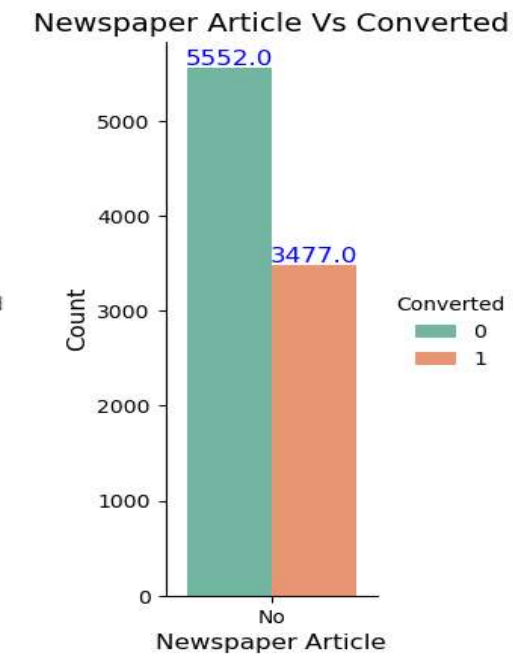
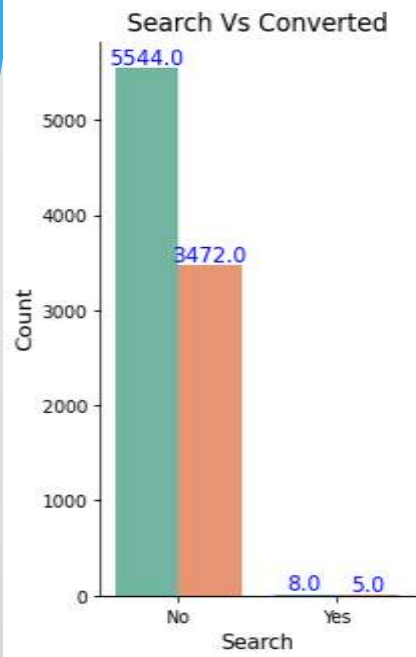


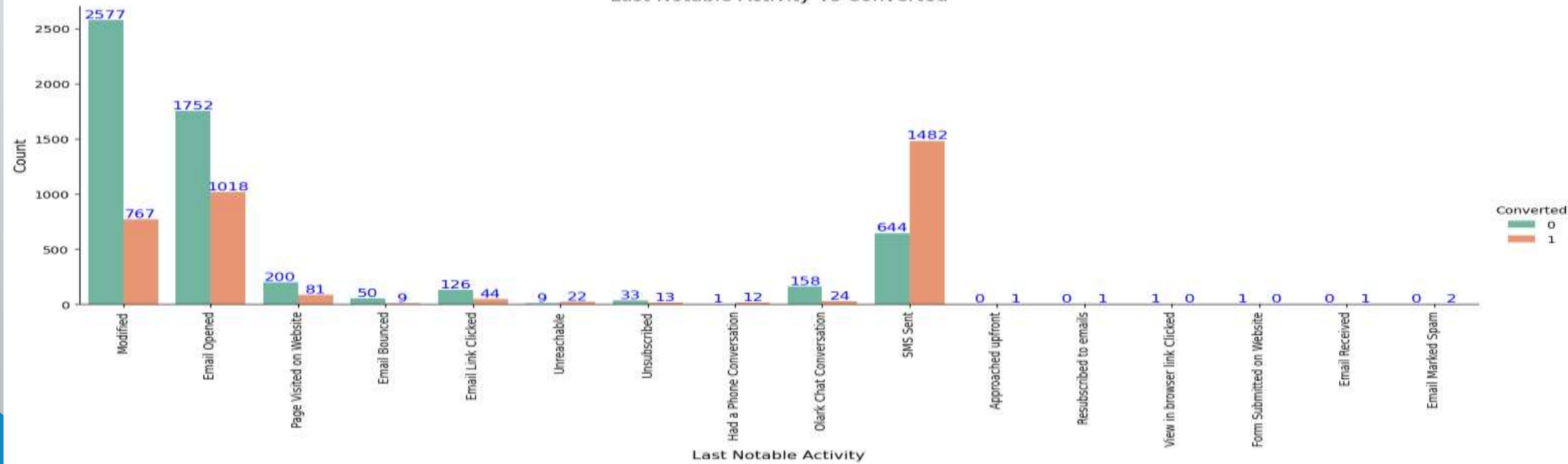
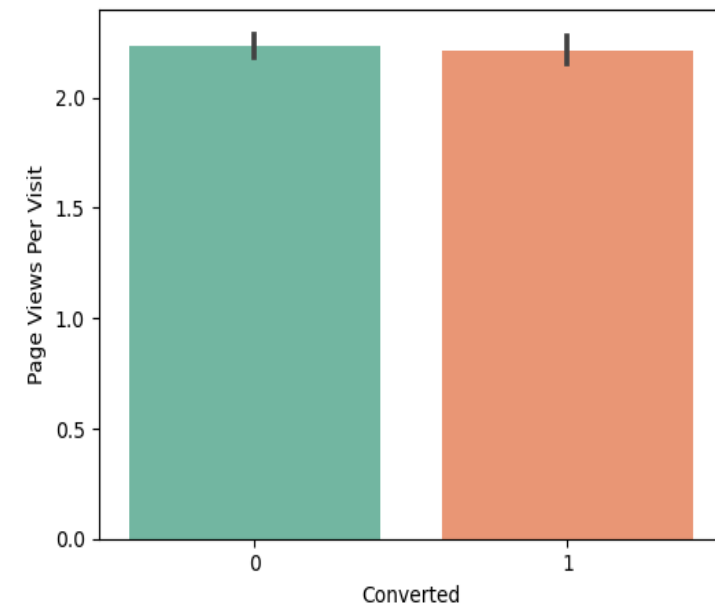
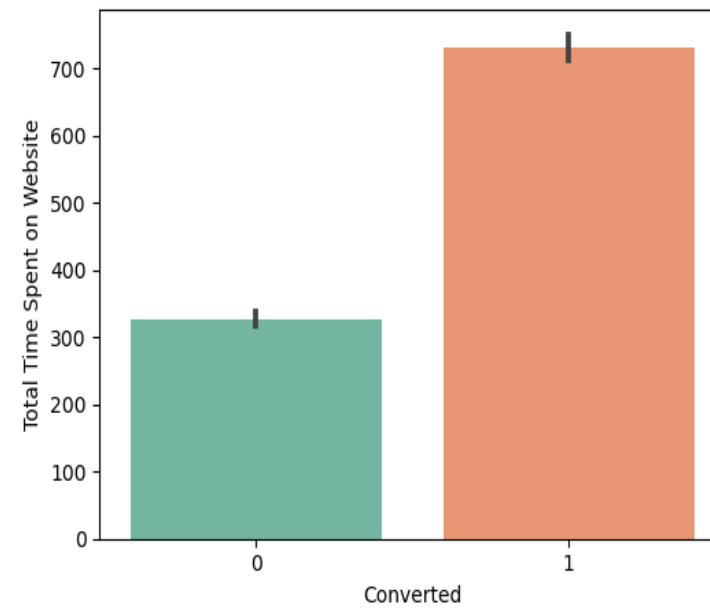
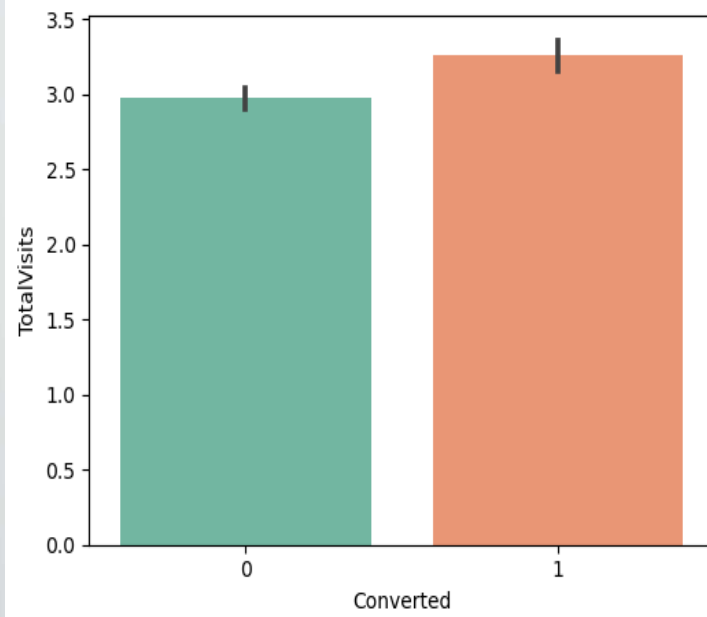
Last activity value of 'SMS Sent' had more conversion



More conversion happend with people who are unemployed. It can also be noticed from the above data that -
Out of 7 business men, 4 got converted - Out 10 housewives, all 10 leads got converted





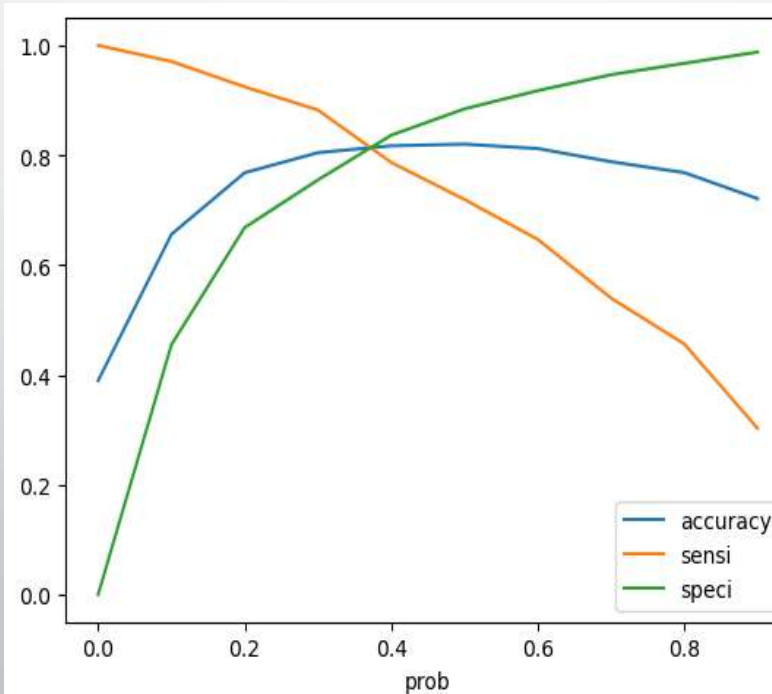
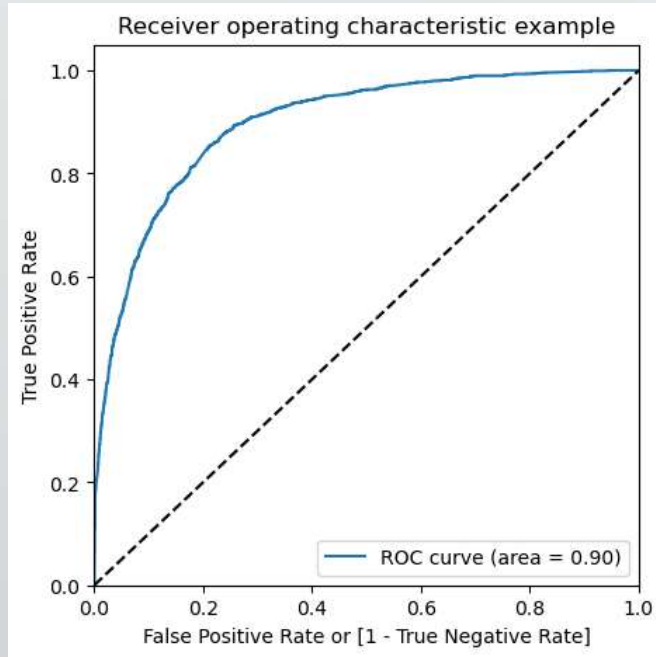


Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 20 variables as output
- Building Model by removing the variable whose p- value is greater than 0.5 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 82%

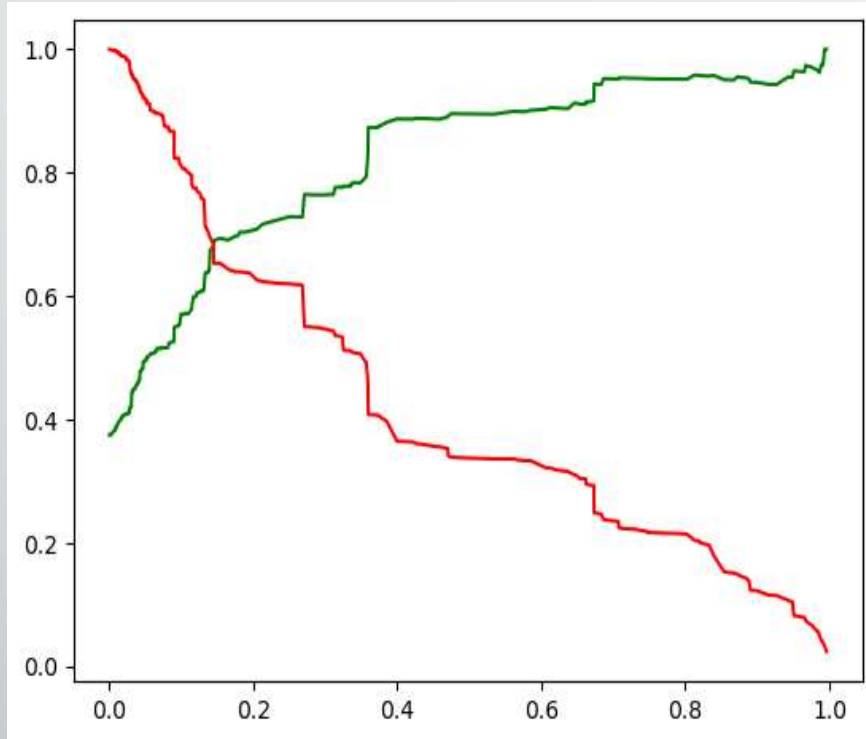
ROC Curve

Optimal cut off probability is that prob where we get balanced sensitivity and specificity



From the curve above, 0.37 is the optimum point to take it as a cutoff probability.

Conclusion



- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 82%, 81.2% and 82.1% which are approximately closer to the respective values calculated using trained set.
- Hence overall this model seems to be good.
- Thus we can conclude that
The total time spend on the Website.
- Total number of visits.

When the lead source was:

- ☐ Google
- ☐ Direct traffic
- ☐ Organic search
- ☐ Welingak website

When the last activity was:

- ☐ SMS
- ☐ Olark chat conversation

- When the lead origin is Lead add format.
 - When their current occupation is as a working professional.
- X Education can flourish by keeping these point as they have a very high chance to get almost all the potential buyers to change their mind and convert.