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:
Hitam Kumar
Sarang S
Mrutyunjay Totad

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%.

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.

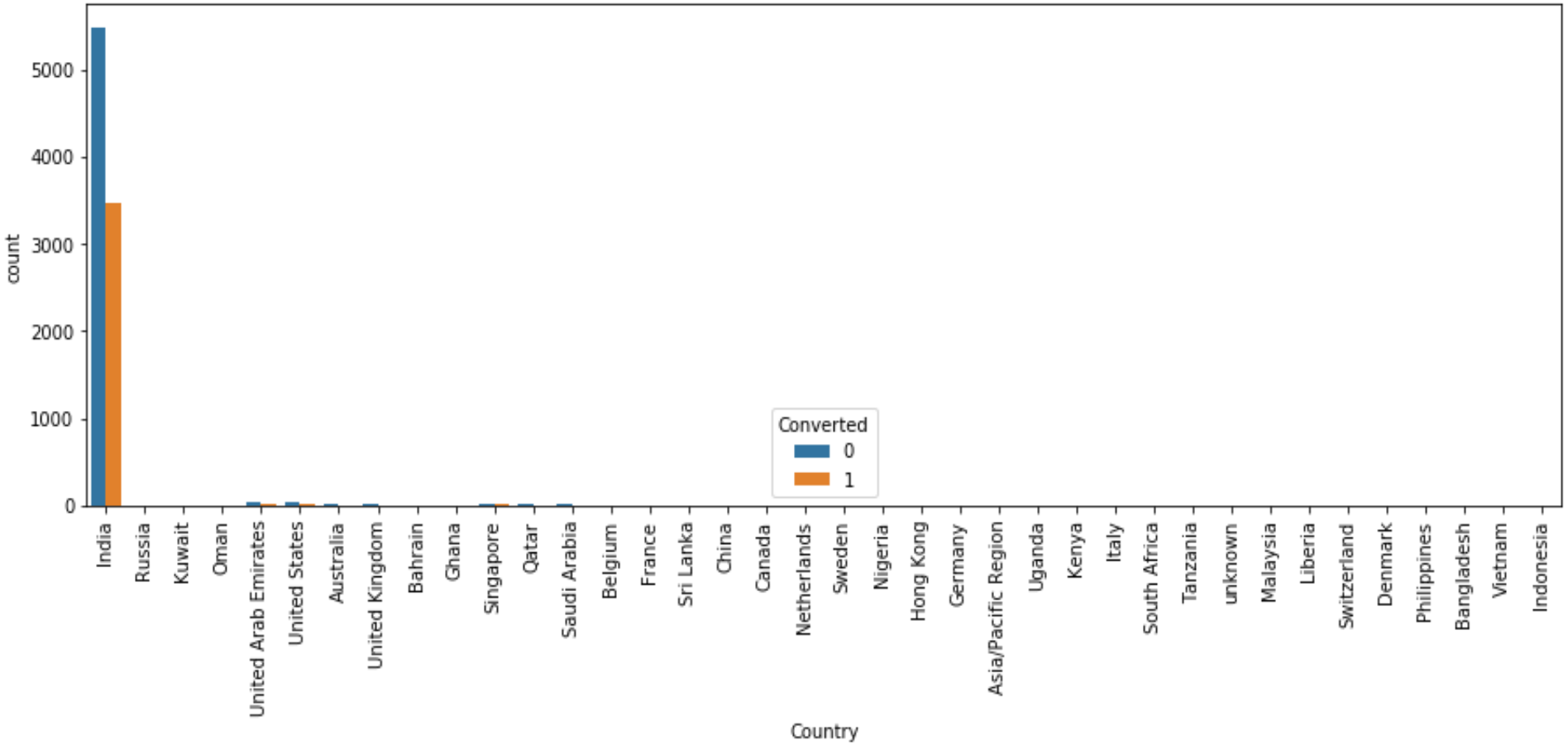
Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

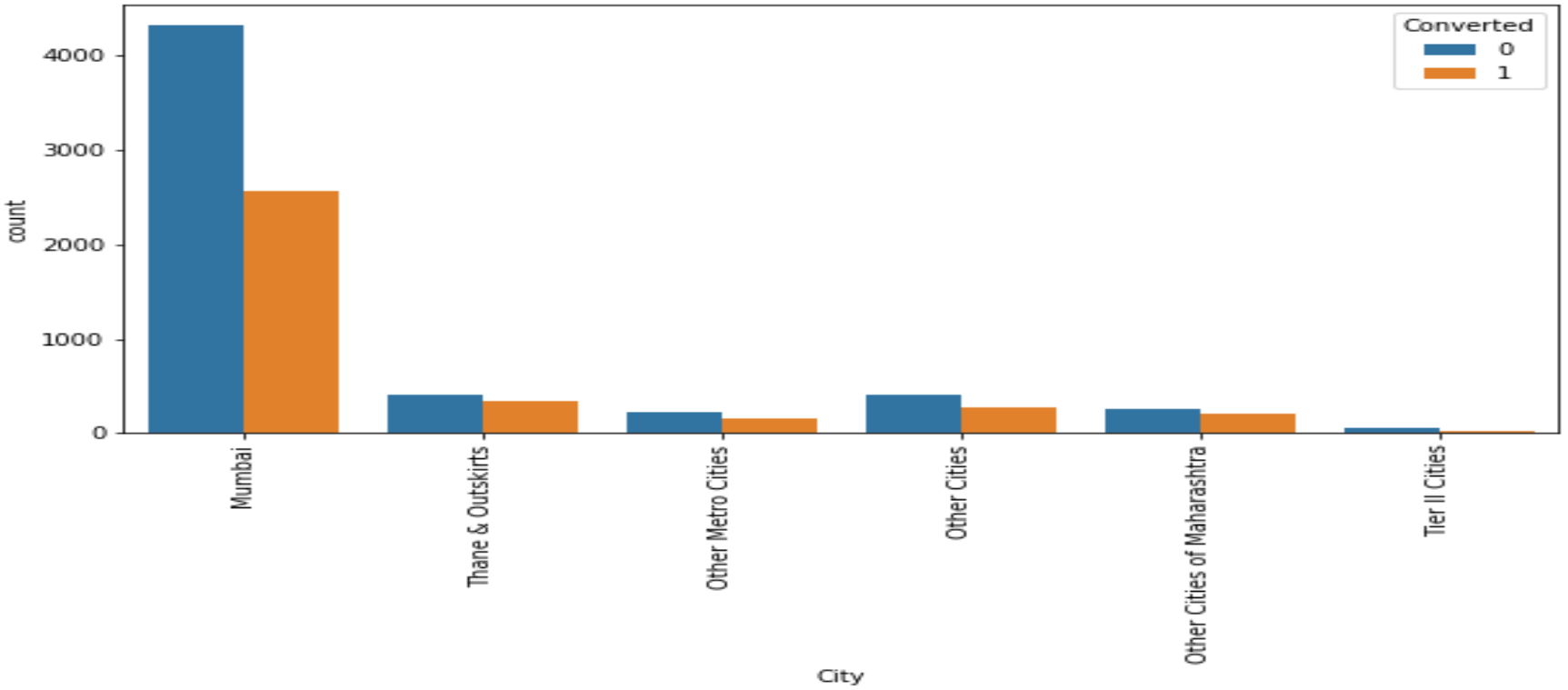
Implementation

ANALYSIS APPROACH

Plots (Visualization)

Data Analysis: Country And City VS Leads Conversion

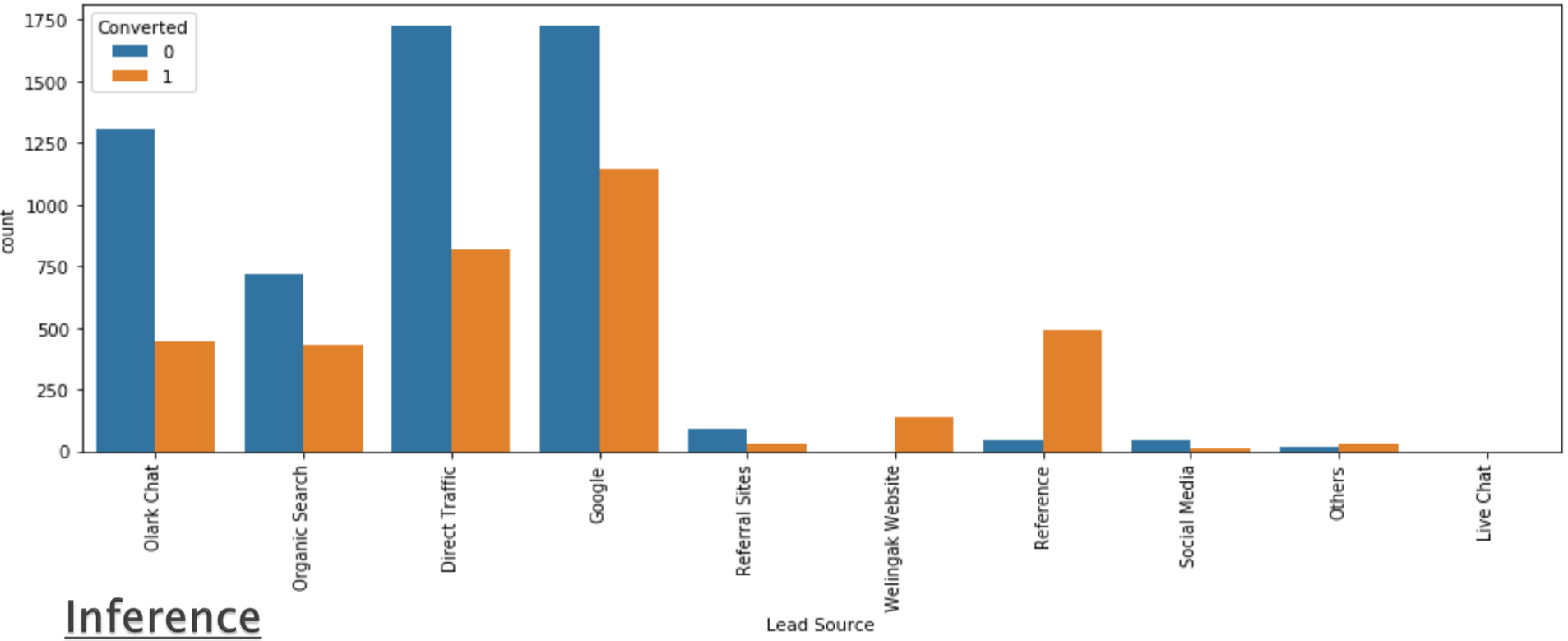




Inference:

Maximum number of leads are From India .
Number of Values for India are quite high (nearly 97% of the Data), this column can be dropped.
Majority of Leads are from Mumbai and Their Conversion rate is Low as compared to leads from that city.

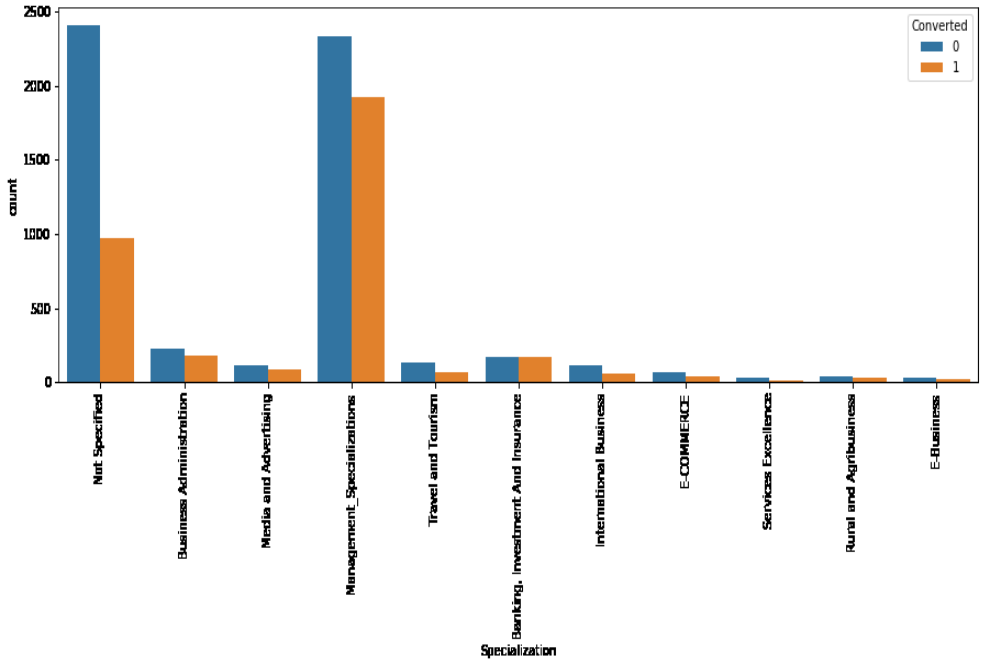
Data Analysis : Conversion Rate Vs Lead Source

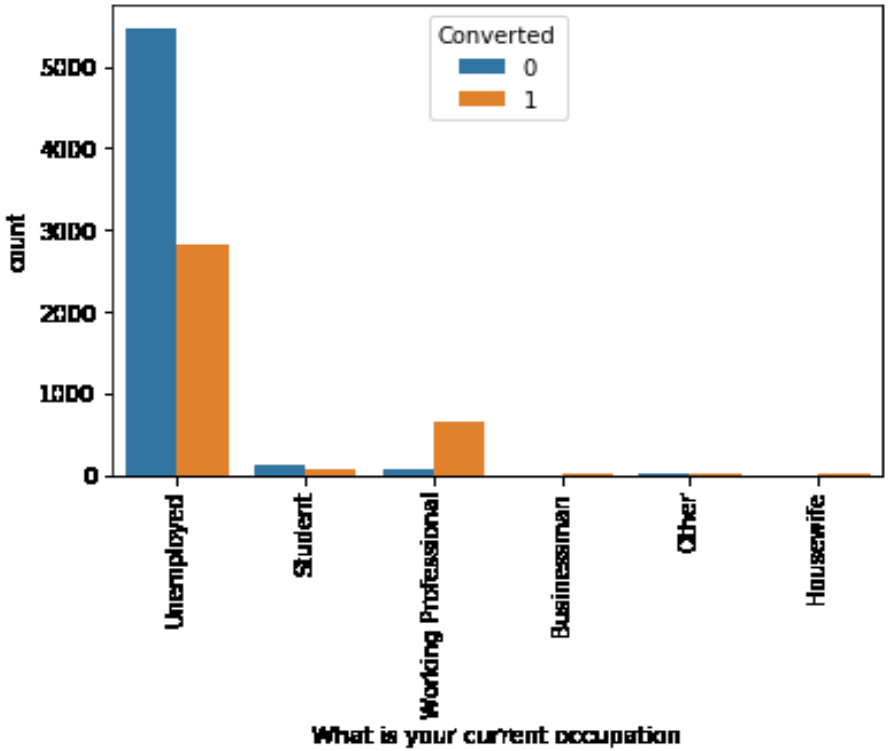


Inference

Maximum number of leads are generated by Google and Direct traffic.
Conversion Rate of reference leads and leads through welingak website is high.
To improve overall lead conversion rate, focus should be on improving lead converion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

Data Analysis:Conversion Rate Vs Lead Occupation





Inference:

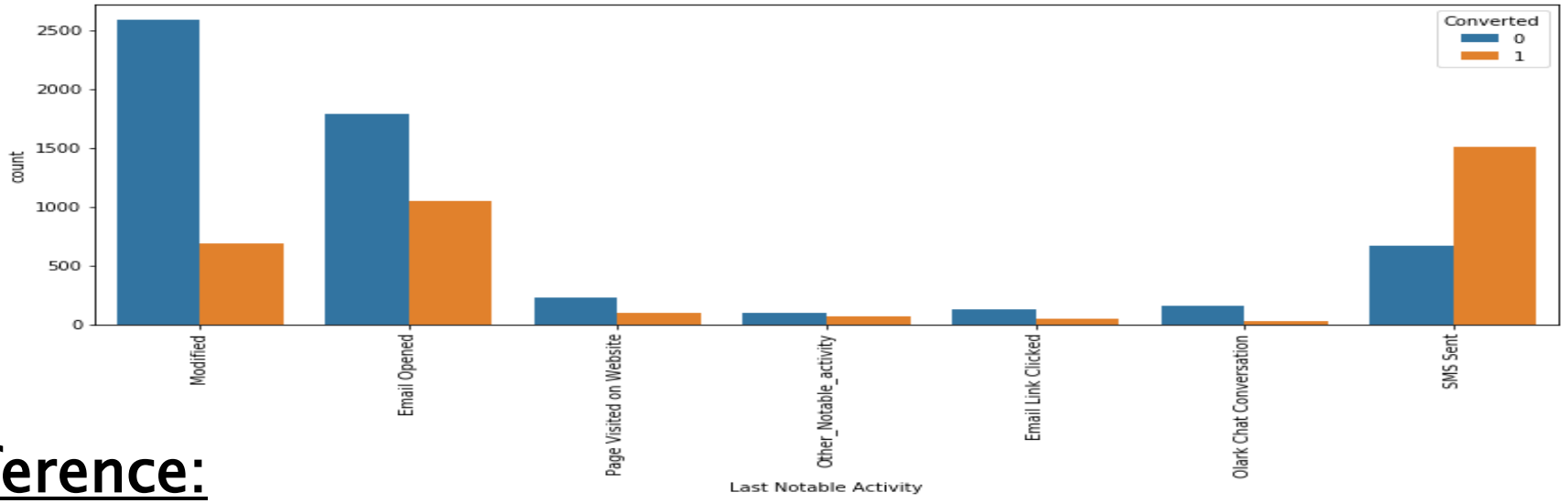
Maximum number of leads are Management_specialisation and One's who not Specified their Work history.

Conversion Rate of Management_specialisation and One's who not Specified their Work history is high.

Working Professionals going for the course have high chances of joining it.

Unemployed leads are the most in terms of Absolute numbers.

Data Analysis:Conversion Rate Vs Lead Last Notable Activity



Inference:

Maximum number of leads whose last notable activity is Modified, Email Opened and SMS sent.

Conversion Rate of SMS sent is more than 50%.

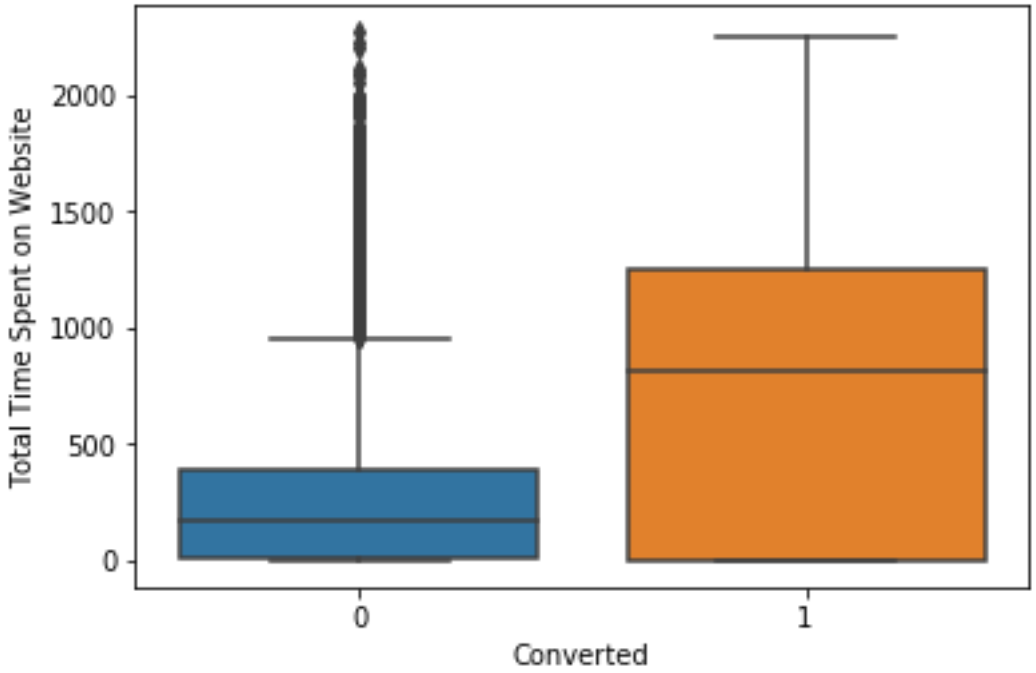
Conversion Rate of Modified and Email opened is low.

Data Analysis:Conversion Rate Vs Lead Time Spent On Website

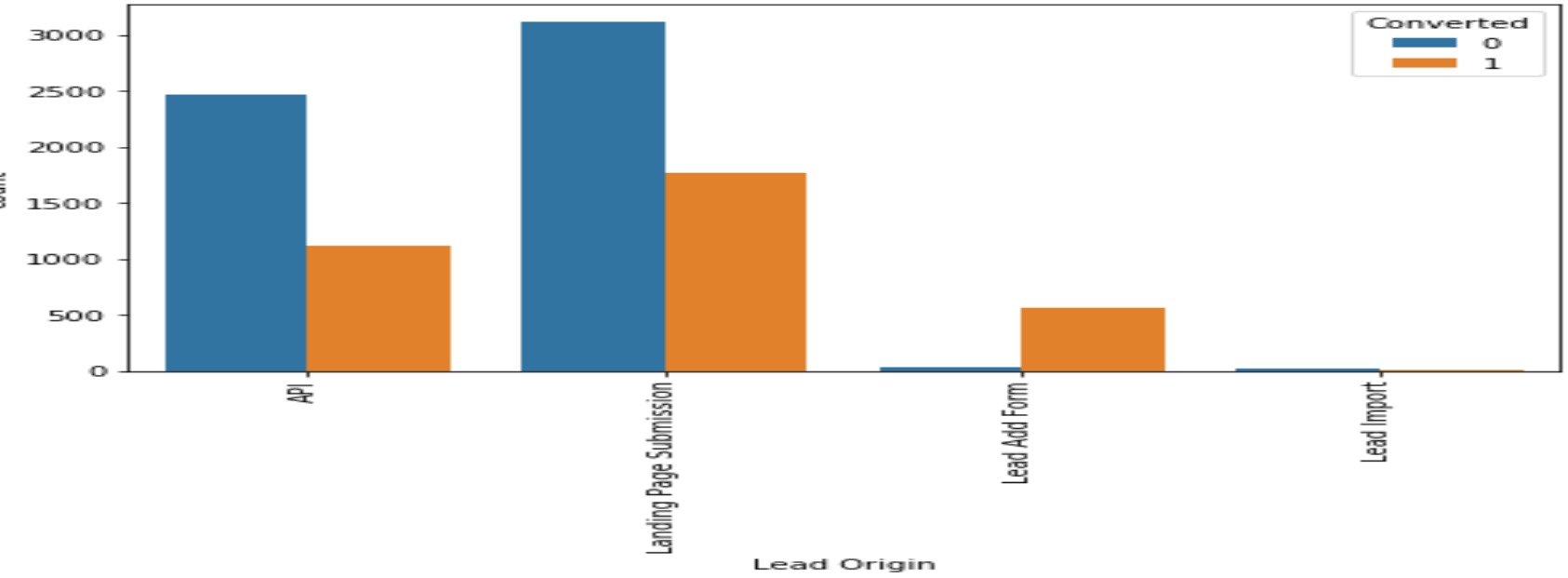
Inference:

Leads spending more time on the website are more likely to be converted.

Website should be made more engaging to make leads spend more time.



Data Analysis:Conversion Rate Vs Lead Origin



Inference:

API and Landing Page Submission bring higher number of leads as well as conversion.

Lead Add Form has a very high conversion rate but count of leads are not very high.

Lead Import and Quick Add Form get very few leads.

In order to improve overall lead conversion rate, we have to improve lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

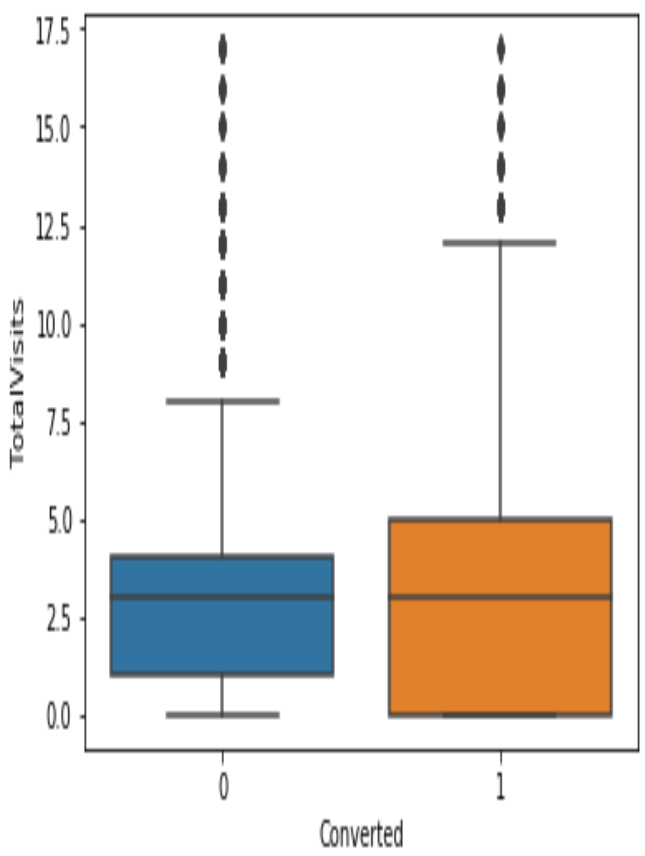
CORRELATION OF THE VARIABLES

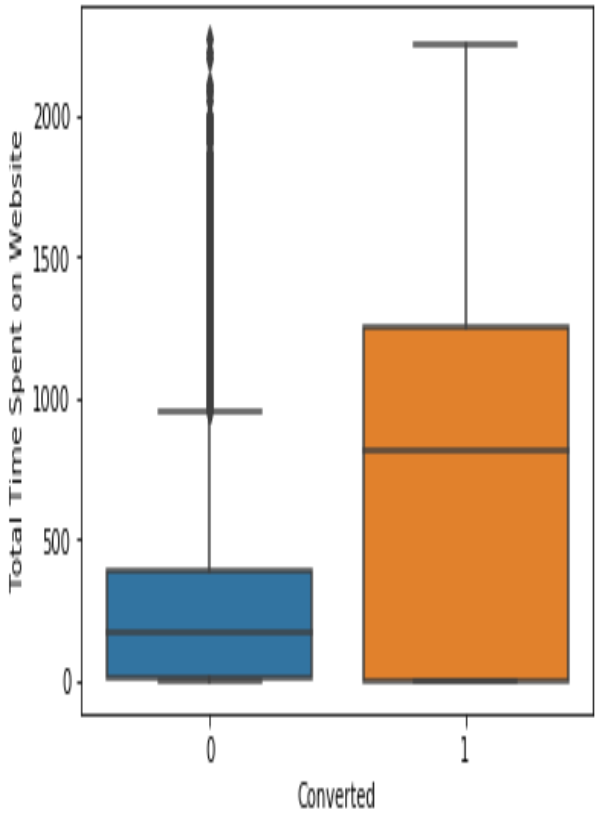
From the heat map on the right side we can conclude below points:

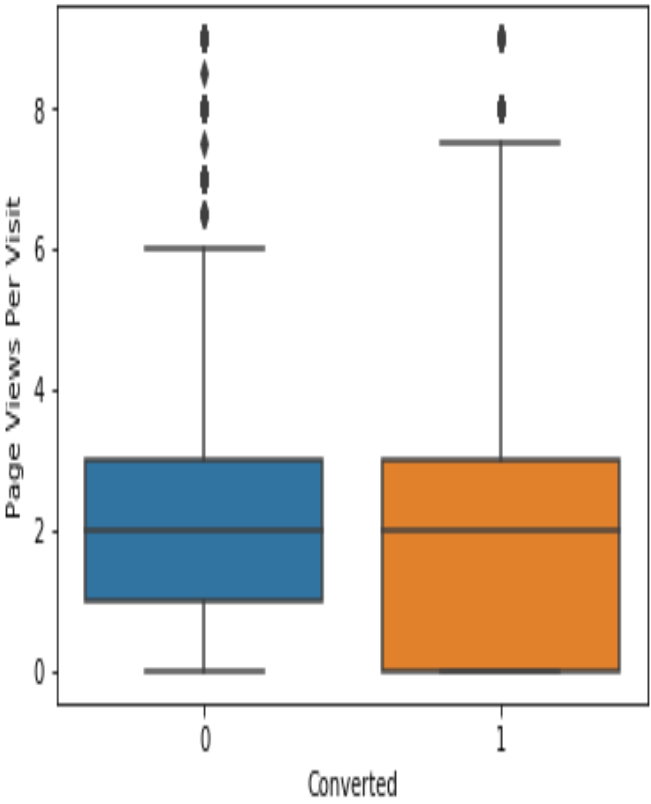
Total Visit And Page Views Per Visit Are Positively Correlated With Correlation Of -0.51

Total Time Spent on Website And Converted Are Positively Correlated With Correlation Of 0.35

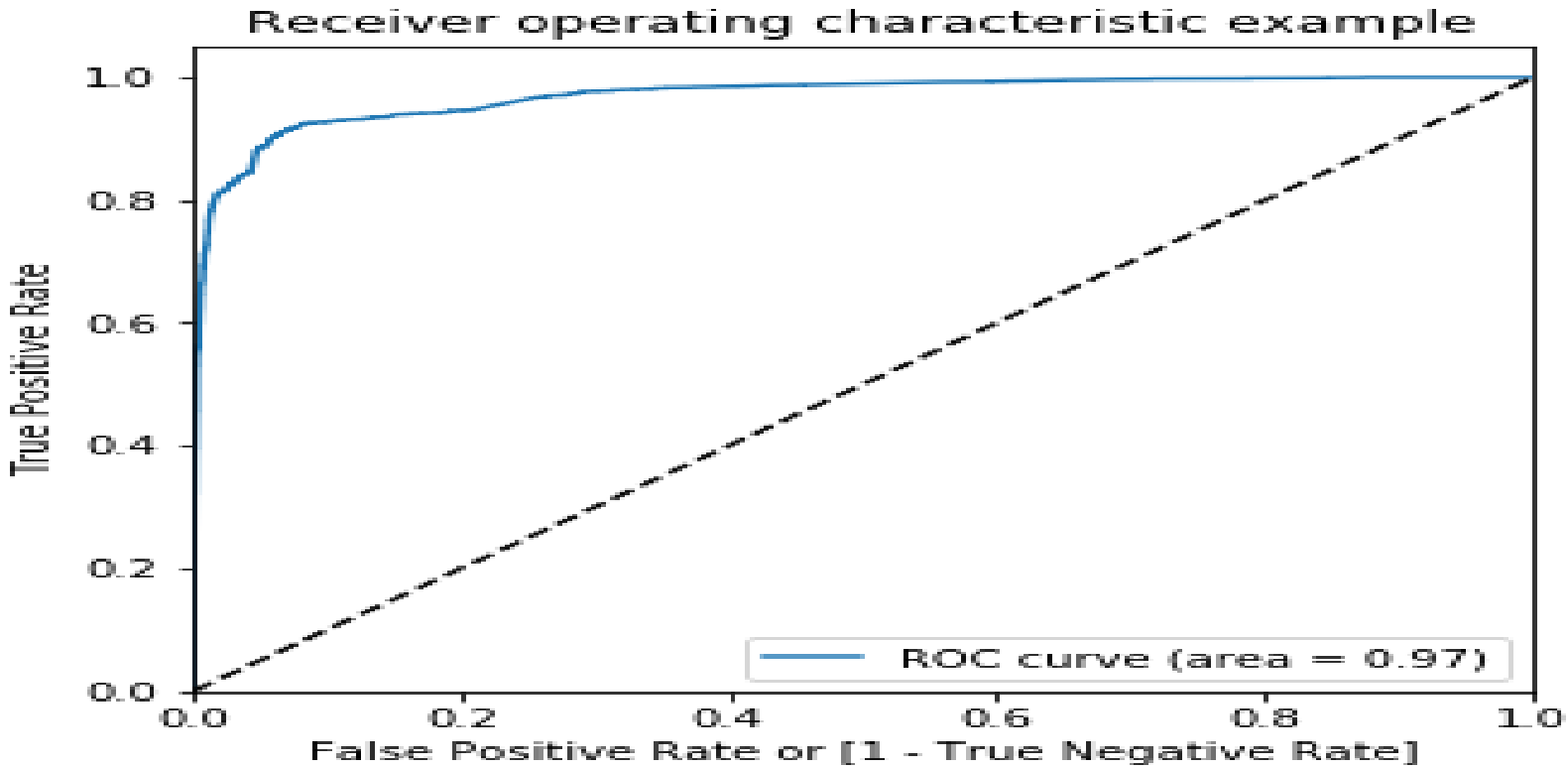
Total Time Spent on Website And Page Views Per Visit Are Positively Correlated With Correlation Of 0.74







"Total Visits" vs Converted variable, "Total Time Spent on Website" vs Converted variable & "Page Views Per Visit" vs Converted variable



So as we can see above the model seems to be performing well. The ROC curve has a value of 0.97, which is very good. We have the following values for the Train Data:

Accuracy : 92.29%

Sensitivity : 91.70%

Specificity : 92.66%

Observation

The ROC curve has a value of 0.97, which is very good. We have the following values for the Train Data :

Accuracy : 92.29%
Sensitivity : 91.70%
Specificity : 92.66%

Observation

After running the model on the Test Data these are the figures we obtain:

Accuracy : 92.78%
Sensitivity : 91.98%
Specificity : 93.26%

Final Observation

Let us compare the values obtained for Train & Test:

Train Data:
Accuracy : 92.29%
Sensitivity : 91.70%
Specificity : 92.66%

Test Data:
Accuracy : 92.78%
Sensitivity : 91.98%
Specificity : 93.26%

Conclusion

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

THANK YOU