

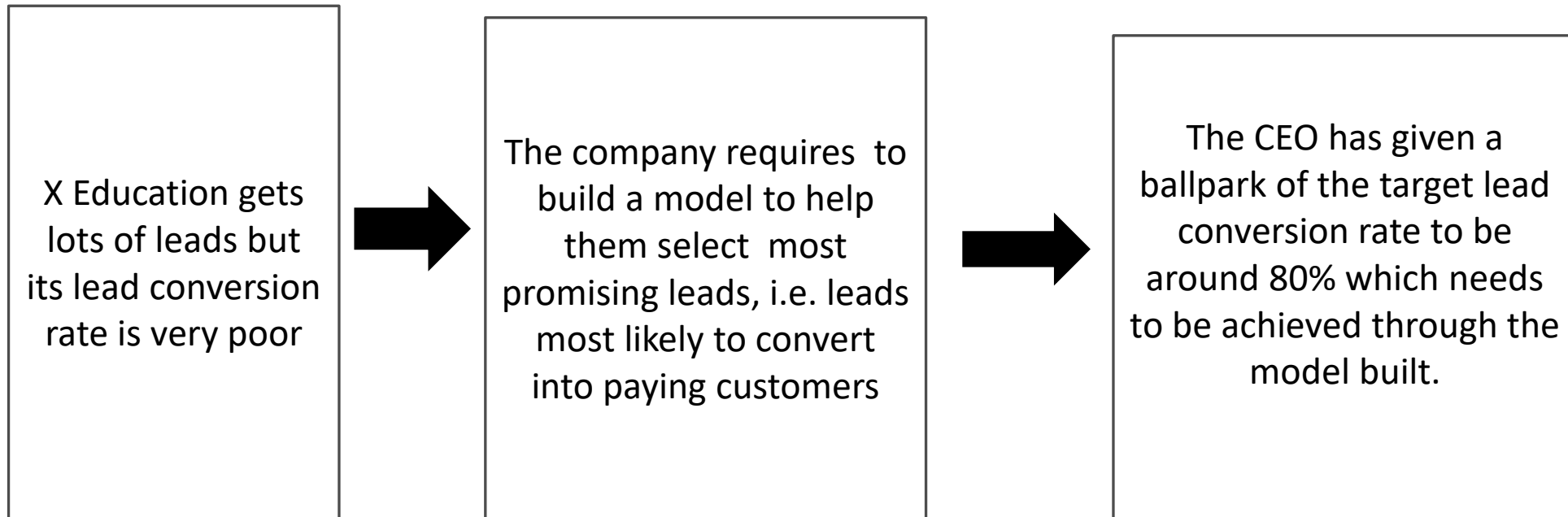
LEAD SCORE CASE STUDY ASSIGNMENT

BY—

APRAJITA SRIVASTAVA

ABHAY KUMAR SINGH

PROBLEM STATEMENT



THE GOAL OF THE CASE STUDY

To build a logistic regression model which will assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.



A higher score will mean that the lead is hot and is most likely to convert. Whereas a lower score will mean that the lead is cold and will mostly likely not get converted.

THE TARGET VARIABLE

Converted \rightarrow 1(lead successfully converted)
 \downarrow
0(lead not converted)



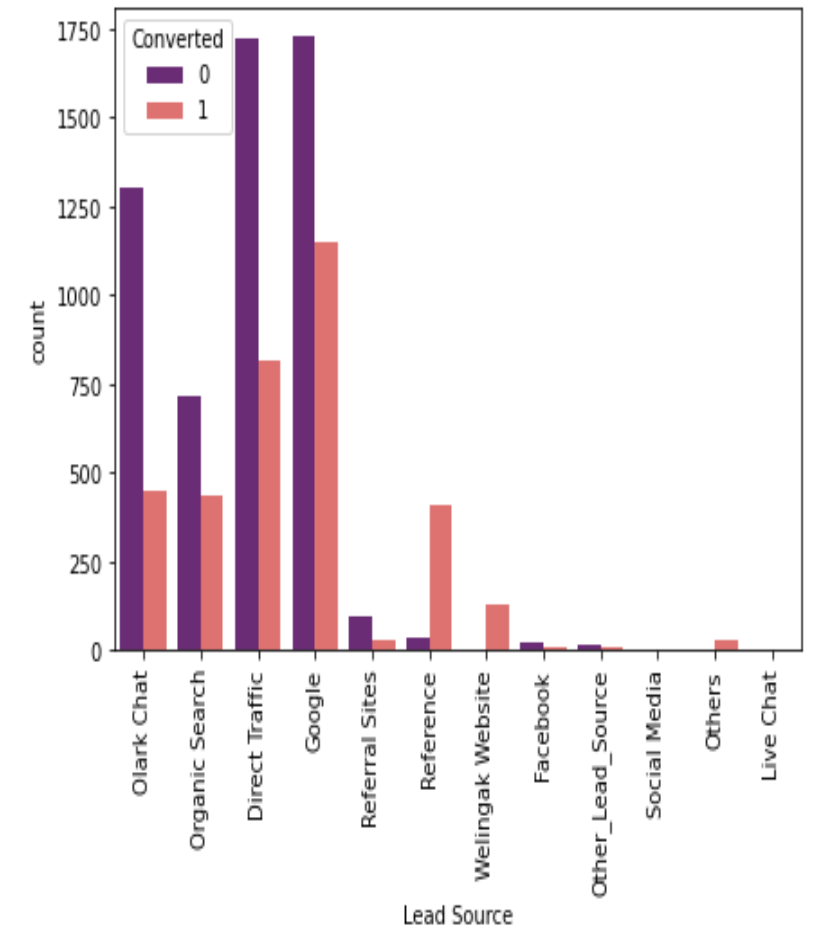
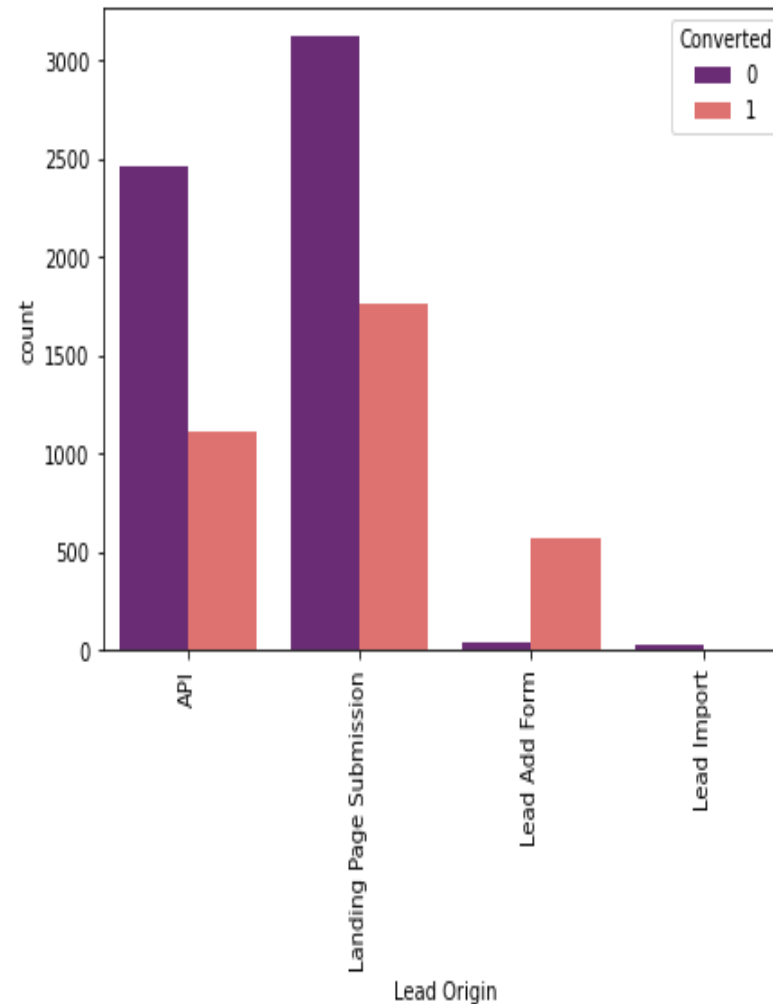
VARIABLES AFFECTING TARGET VARIABLE- CATEGORICAL

Lead Origin:

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

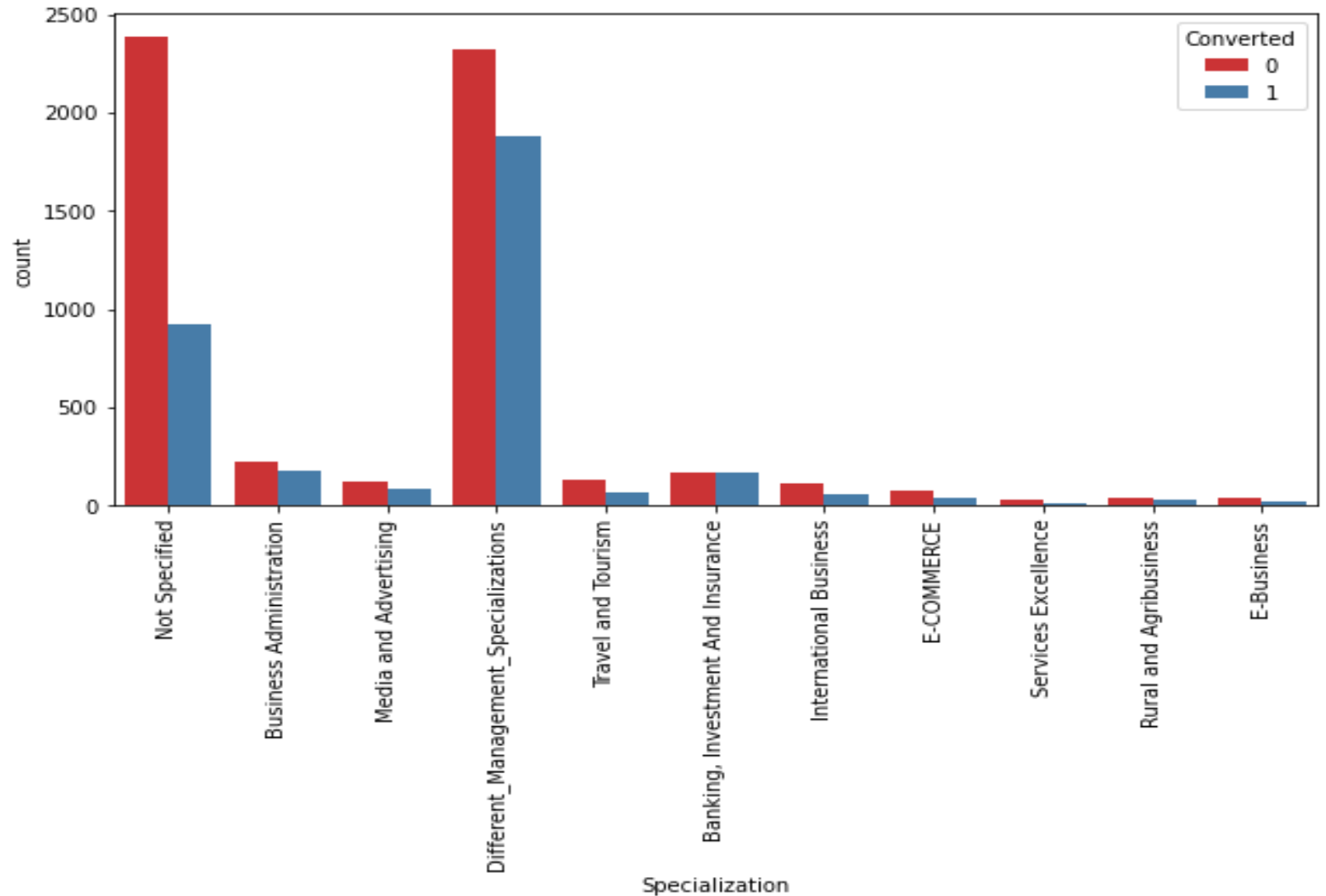
Lead source :

- Google has high number of leads with high converted number.



➤ Specialization:

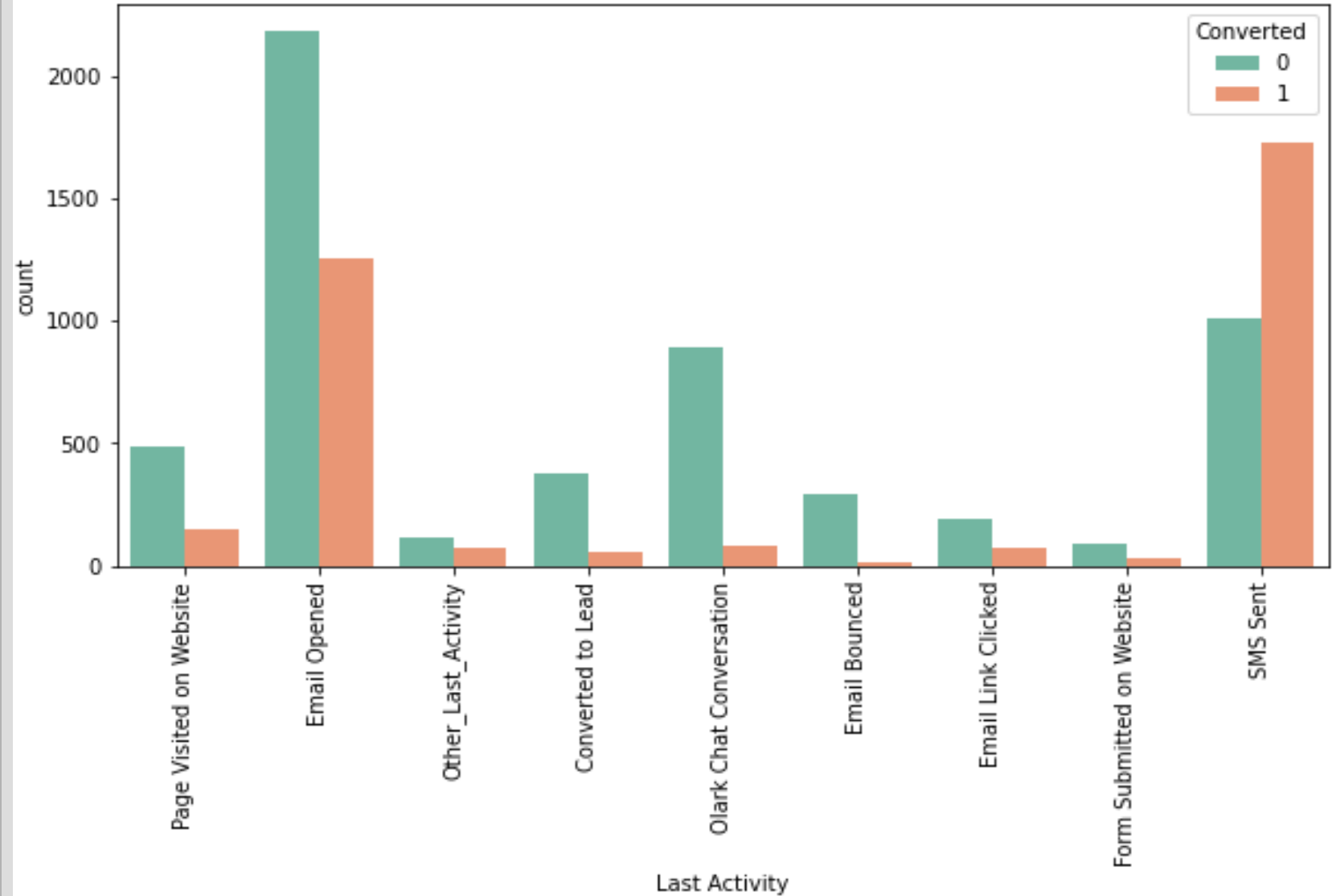
The specializations with Management in them have higher number of leads as well as leads converted. So this is definitely a significant variable .



Last Activity:

Last activity performed by the customer in which email opened bring high number of leads and converted lead.

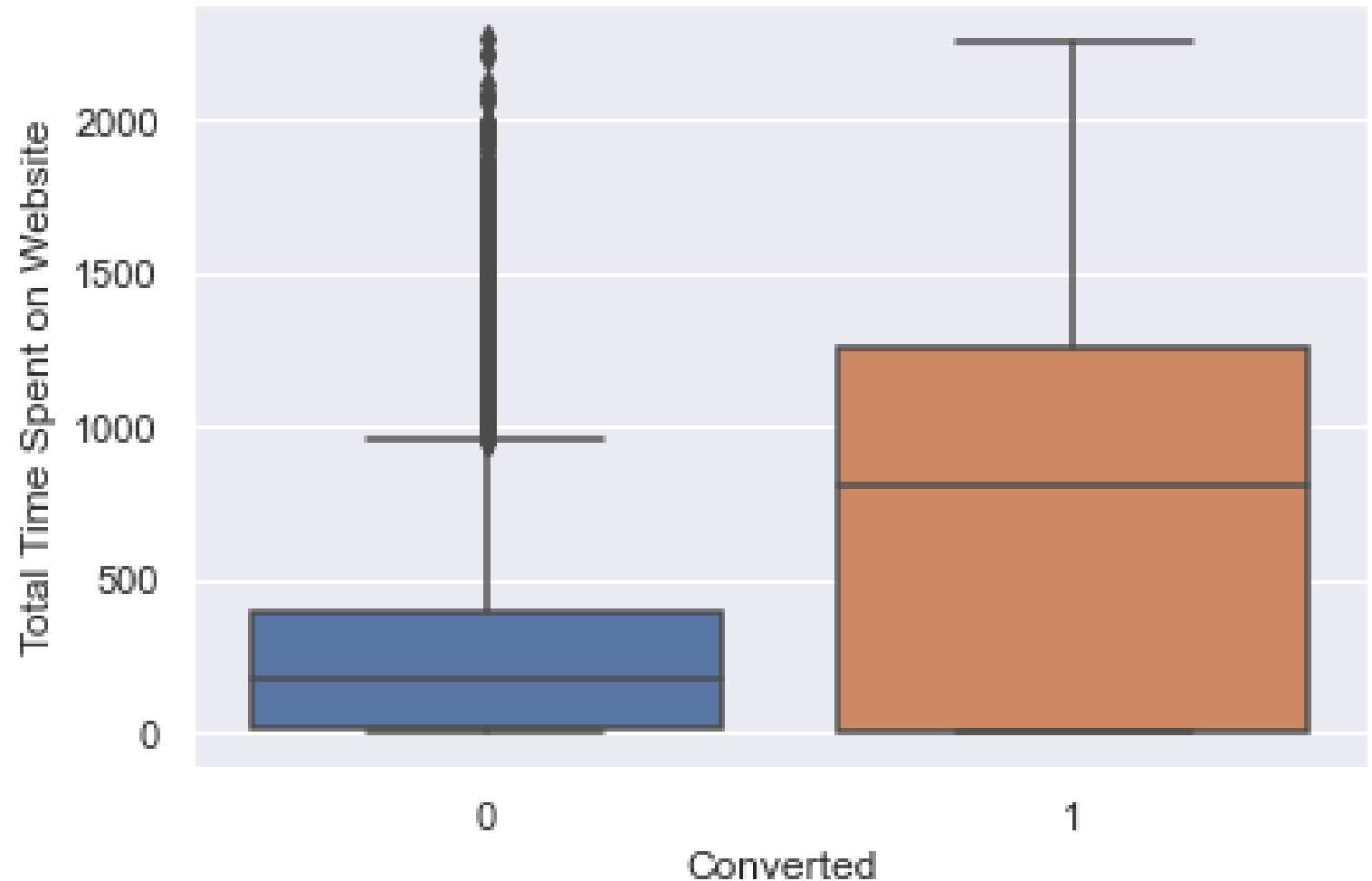
SMS sent has high converted lead.



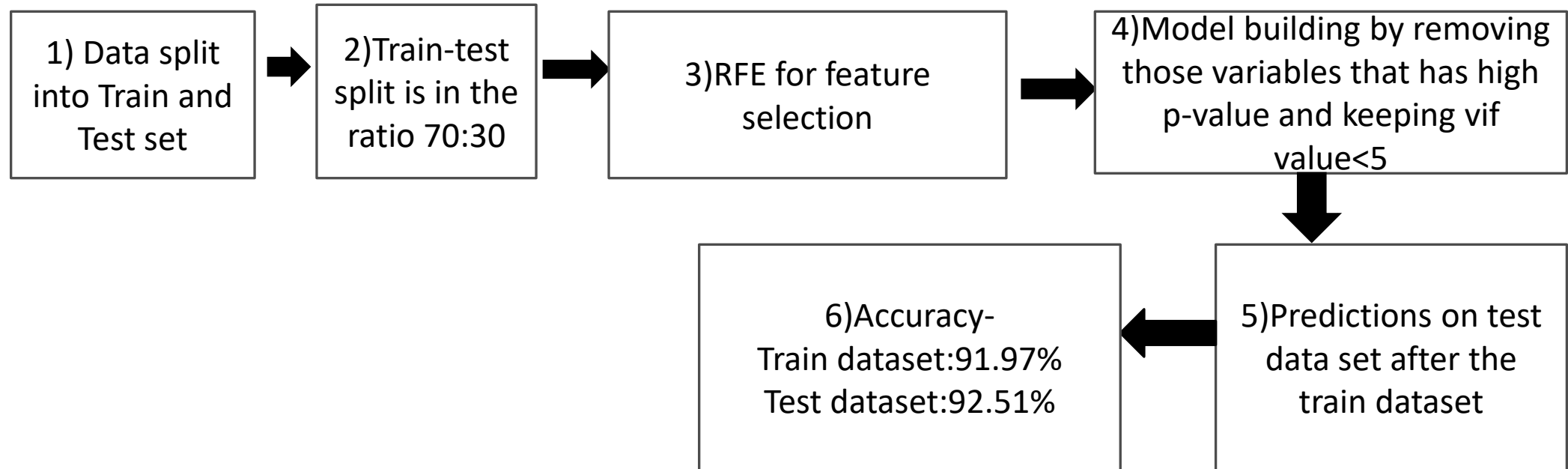
NUMERICAL VARIABLES-

Total Time Spent on Website:

- 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.

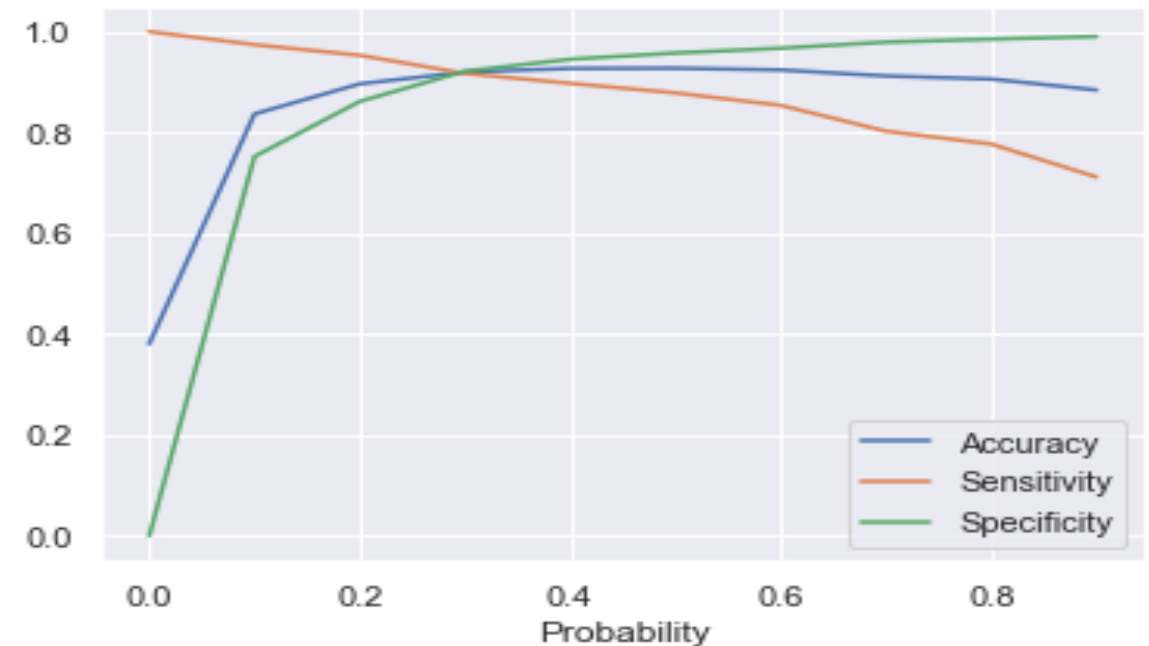
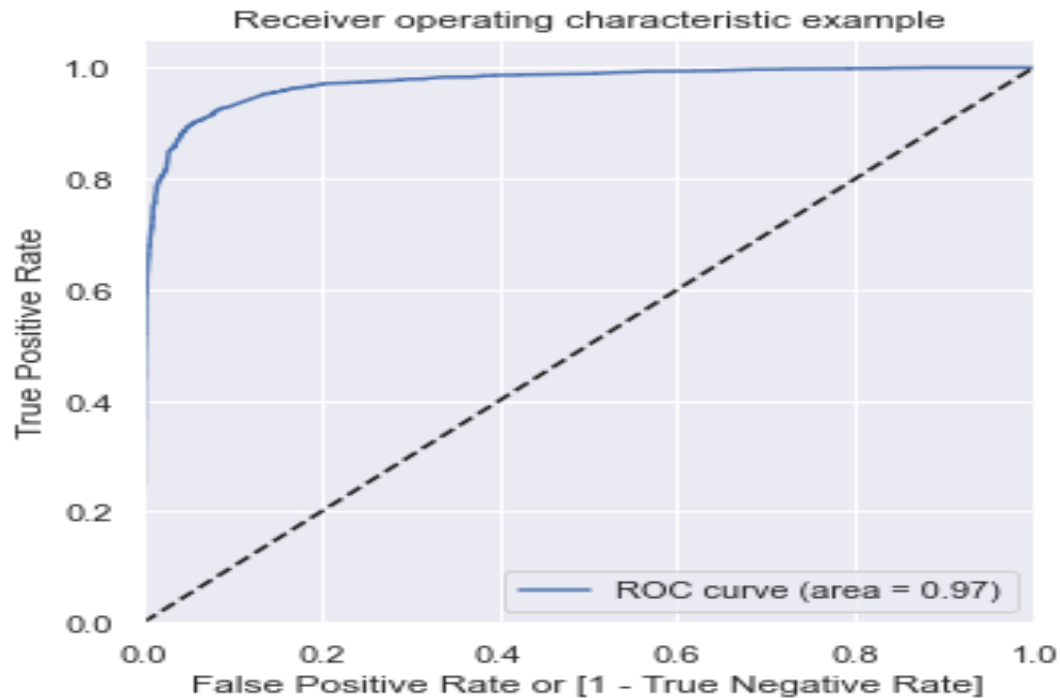


MODEL BUILDING

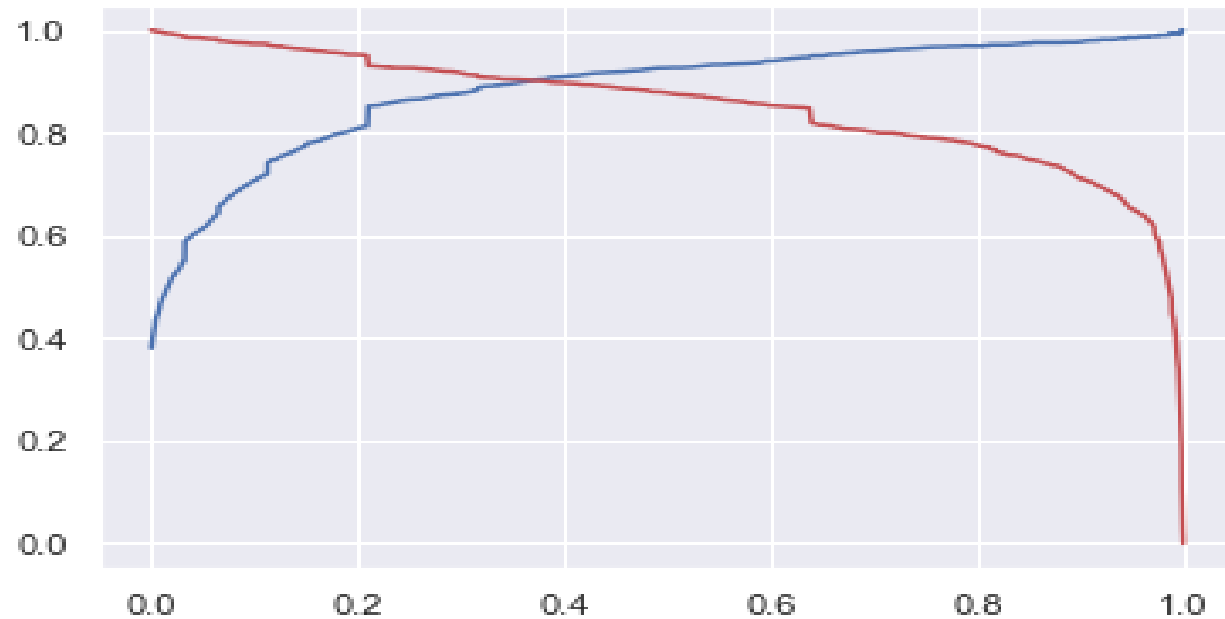


MODEL ANALYSIS

- **ROC CURVE:** ROC is a probability curve. It should have a value close to 1. The graph below shows that we are getting a good value of 0.97 indicating a good predictive model. Optimal cutoff probability is that probability where we get balanced specificity and sensitivity (right graph) and the value is 0.3.



-
- **Precision-recall curves:** The precision-recall curve is used for evaluating the performance of binary classification algorithms. It is often used in situations where classes are heavily imbalanced. The precision and Recall seem to have trade-off at .38



CONCLUSION

- The variables that affect the customers' conversion to converted lead are –
 - Total time spend on website -Leads spending more time on the website are more likely to be converted.
 - Lead source –Google and Direct traffic has high lead converted.
 - Last activity- SMS sent and Email opened
 - Specialization-Different management specialization leads to more lead converted
 - Lead origin-Landing Page Submission and API should be given attention as has high lead converted
- Through these variables the company can nurture the potential leads well so that they get a higher lead conversion.

Thank You

