LEAD SCORING CASE STUDY

BY

SAMPATHKUMAR

SNEH PATEL

HARSH KATARIA

INTRODUCTION

- Logistic Regression, a popular classification algorithm, proves to be an invaluable tool for lead scoring due to its ability to model the probability of an event occurring, in this case, the likelihood of a lead converting.
- By analyzing historical data and extracting meaningful features, Logistic Regression enables businesses to create a predictive model that scores leads on a scale, indicating the probability of conversion.

PROBLEM STATEMENT

- At X education, the lead conversion rate is approximately 30% on average. This implies that just around 30 of the leads, say, that they generate each day will actually be converted. The organization wants to find the most potential leads, or "Hot Leads," in order to streamline this procedure.
- After obtaining these leads, sales team members begin calling, emailing, and so on. Some of the leads convert during this process, but the majority do not.
- The lead conversion rate should increase if they are able to locate this group of leads because the sales staff will now be spending more time corresponding with the prospects rather than calling everyone.

BUSINESS OBJECTIVE

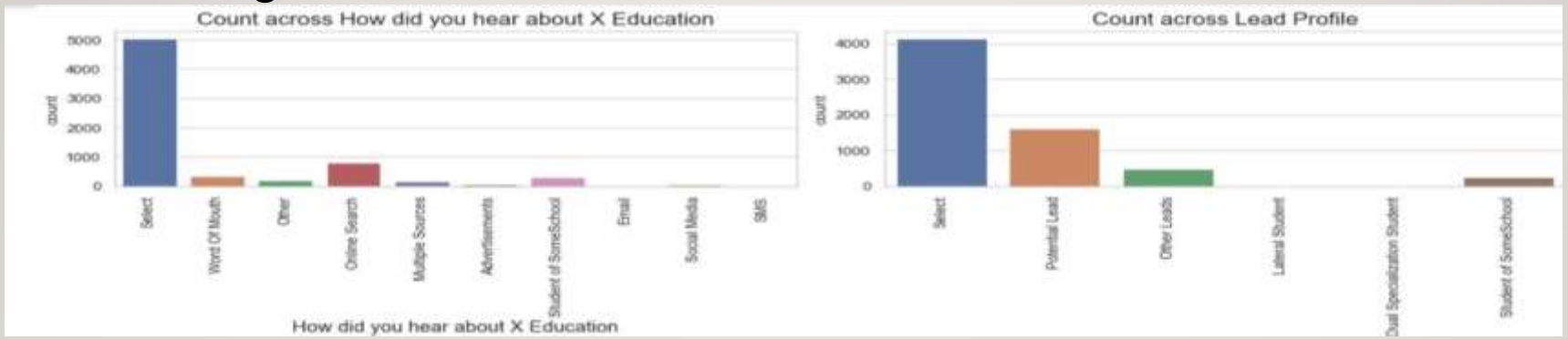
- Lead X requests that we create a model that assigns a lead score ranging from 0 to 100 to each lead. in order to detect the Hot leads and boost their conversion rate simultaneously.
- The CEO want to achieve a lead conversion rate of 80%.
- Future restrictions, such as the need for peak time actions, how to use all available manpower, and what to do after a target is reached, are other things they want the model to be able to manage.

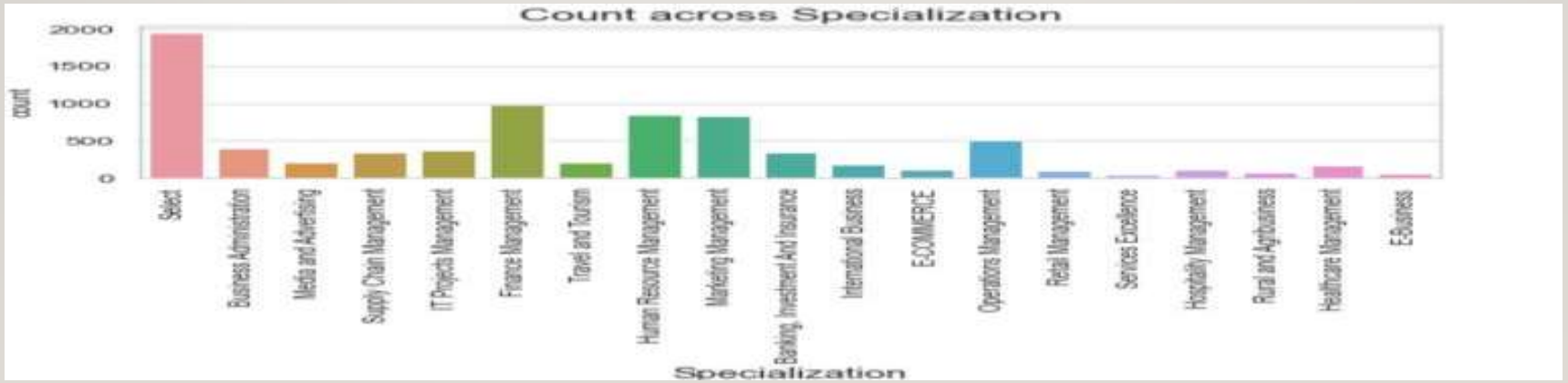
PROBLEM APPROACH

- **Define the Business Problem**
- **Set Clear Objectives**
- **Data Collection and Exploration**
- **Model Selection**
- **Data Splitting**
- **Deployment and Integration**
- **Monitoring and Iteration**
- **Feedback Loop**

GRAPHS AND INSIGHTS

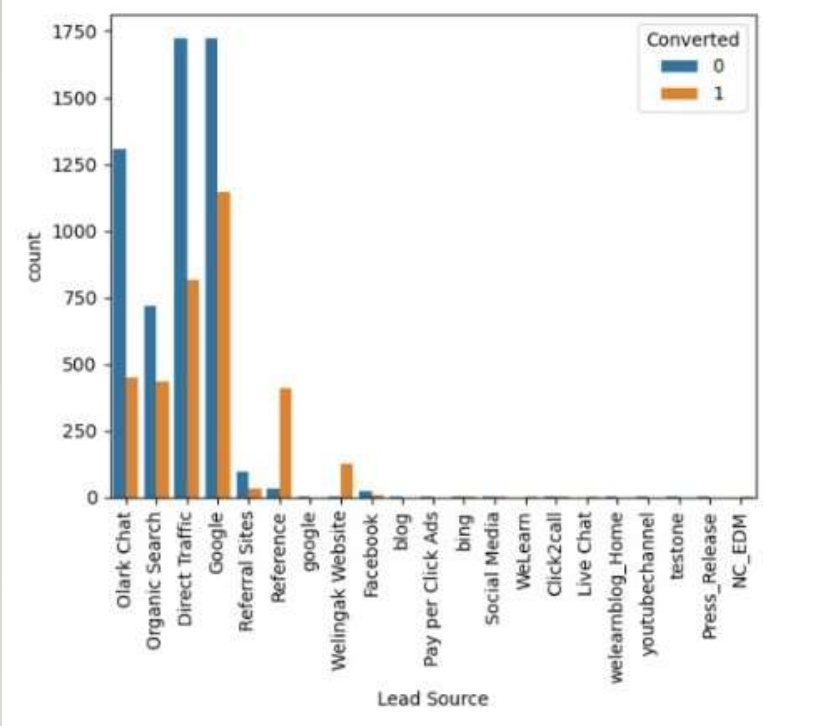
- Data Cleaning

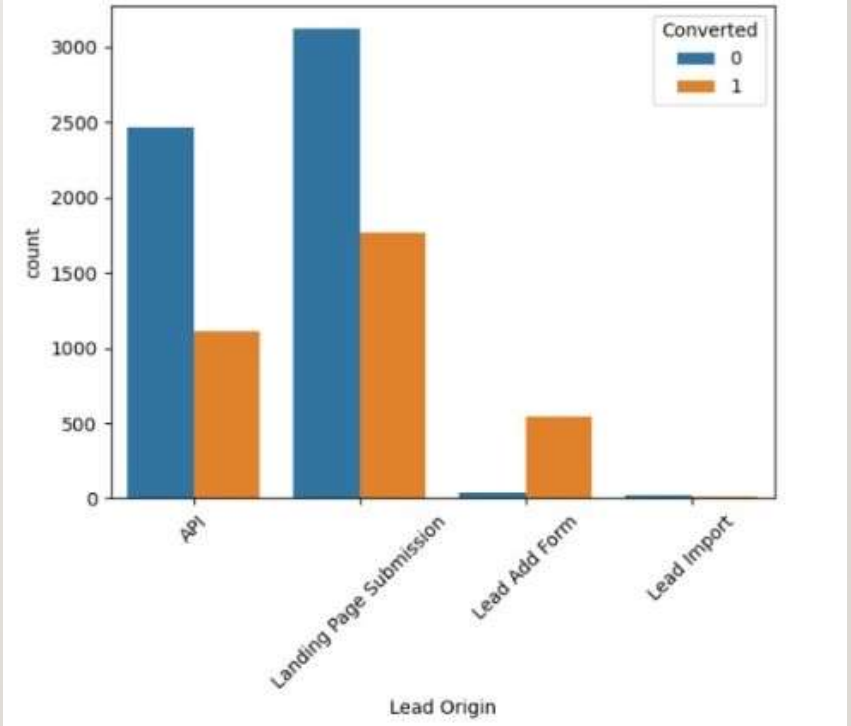




LEAD SOURCE AND LEAD ORIGIN

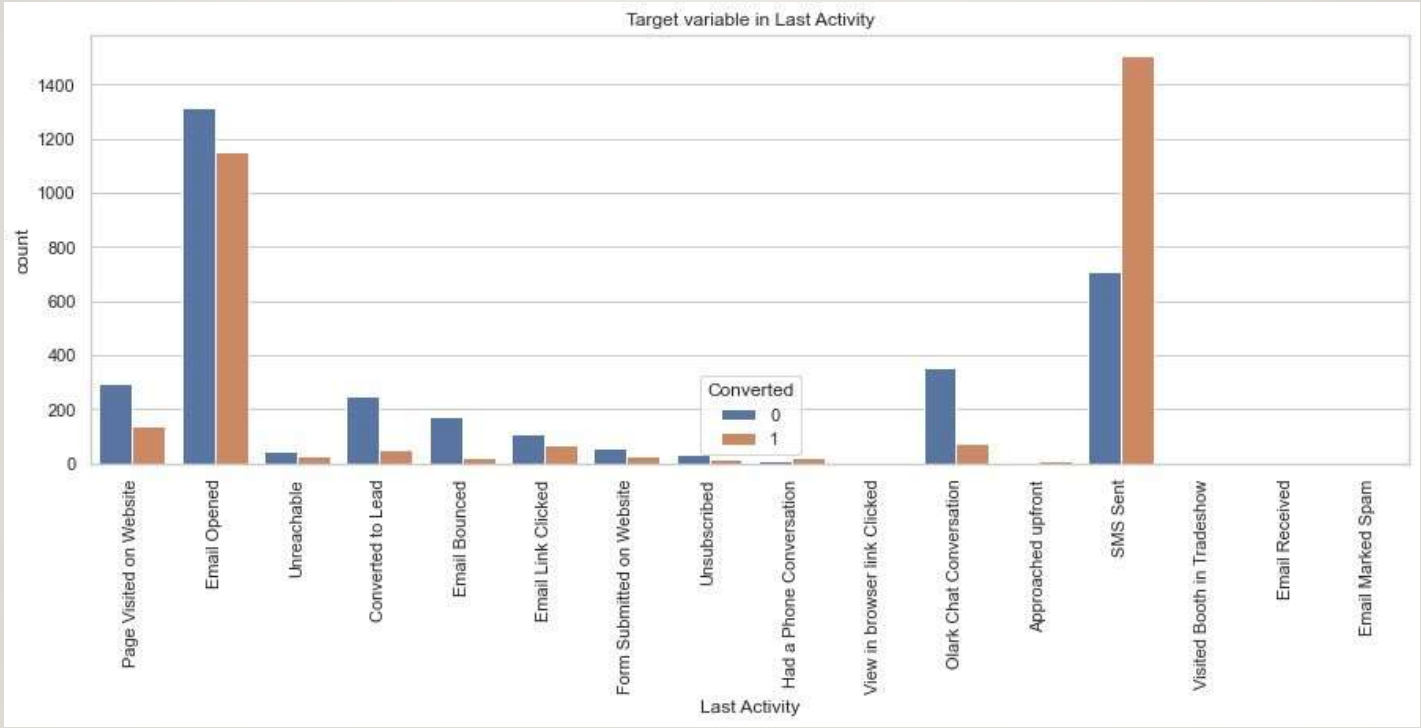
IN LEAD SOURCE THE LEADS THROUGH GOOGLE & DIRECT TRAFFIC HIGH PROBABILITY TO CONVERT WHEREAS IN LEAD ORIGIN MOST NUMBER OF LEADS ARE LANDING ON SUBMISSION





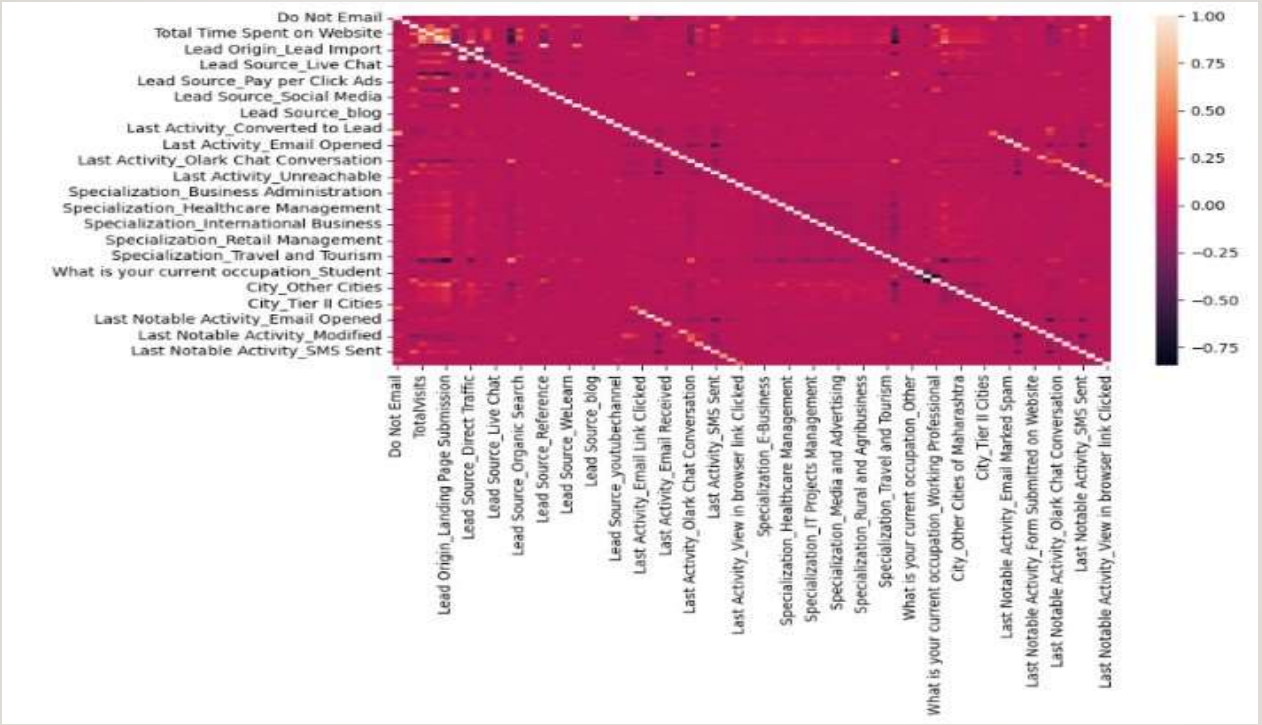
LAST LEAD ACTIVITY

LEADS WHICH ARE OPENING EMAIL HAVE HIGH PROBABILITY TO CONVERT, SAME AS SENDING SMS WILL ALSO BENEFIT.



CORELATION

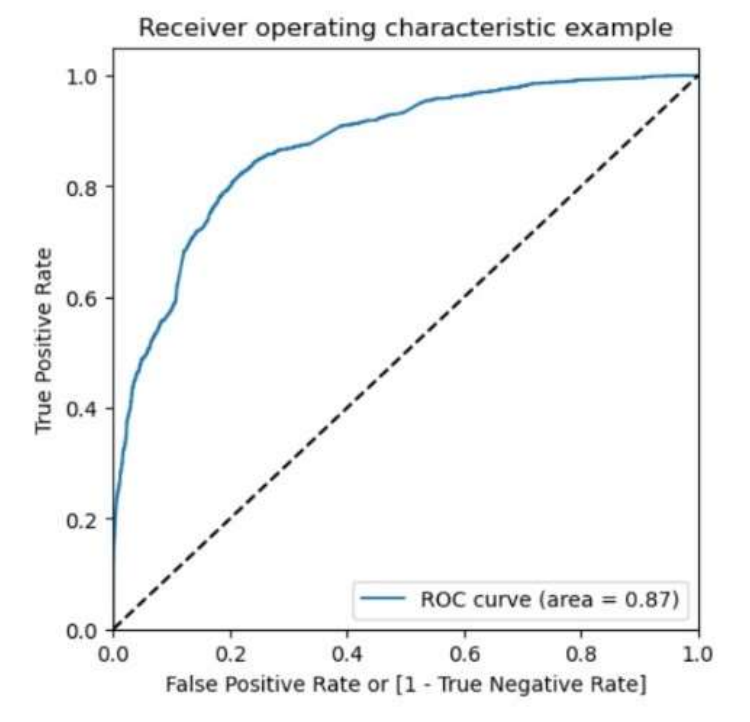
THERE IS NO CORRELATION BETWEEN THE VARIABLES

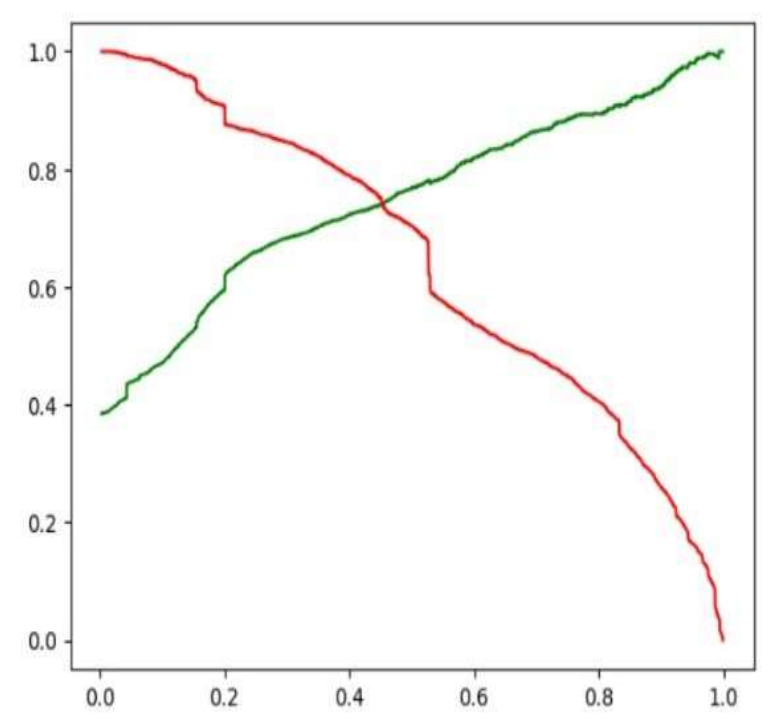


MODEL EVALUATION

ROC CURVE

0.42 IS THE TRADEOFF BETWEEN PRECISION AND RECALL -
THUS WE CAN SAFELY CHOOSE TO CONSIDER ANY PROSPECT LEAD WITH CONVERSION PROBABILITY HIGHER THAN 42 % TO BE A HOT LEAD





RECOMMENDATION OR CONCLUSION

- We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore we can intervene that we need to focus more on the leads originated from API and Landing page submission.
- We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.