K

Lead Scoring Case Study

Prepared by: Karthik

Date: 28 Sep 2025

Logistic Regression • Scoring 0–100 • Operating Playbooks

This presentation summarizes the modeling approach, key drivers of conversion, and recommended operating thresholds for aggressive vs. conservative sales modes. All insights are derived from historical lead outcomes and engagement behavior.

LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

SUBMITTED BY:- RAJAT MESHRAM

Contents

PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have a process of form filling on their website after which the company treats that individual as a lead.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.

The typical lead conversion rate at X Education is around 30%. Now, this means 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

Lead X wants us to build a model to give every lead a lead score between 0-100, so that they can identify the Hot Leads and increase their conversion rate as well.

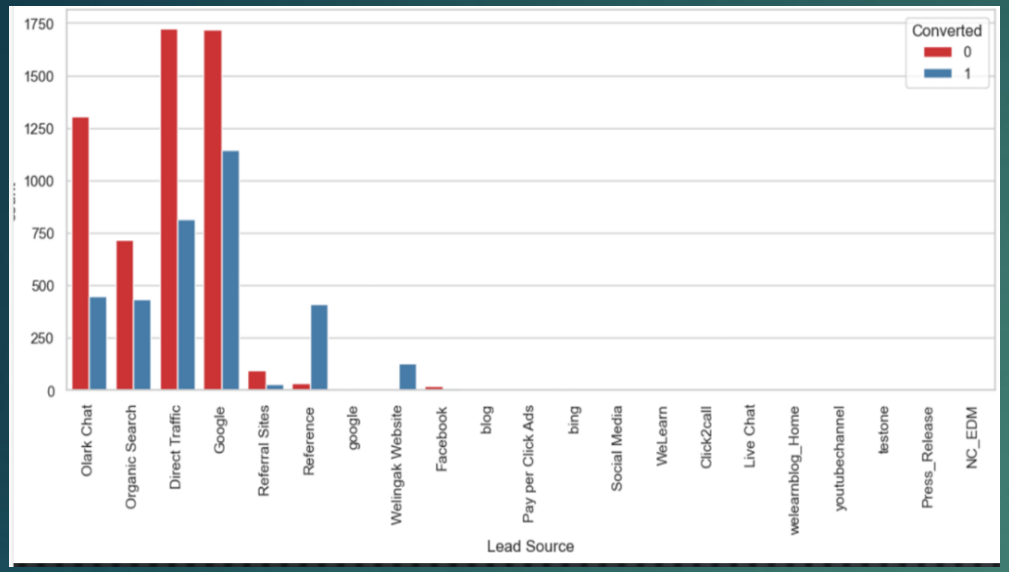
The CEO wants to achieve a lead conversion rate of 80%.

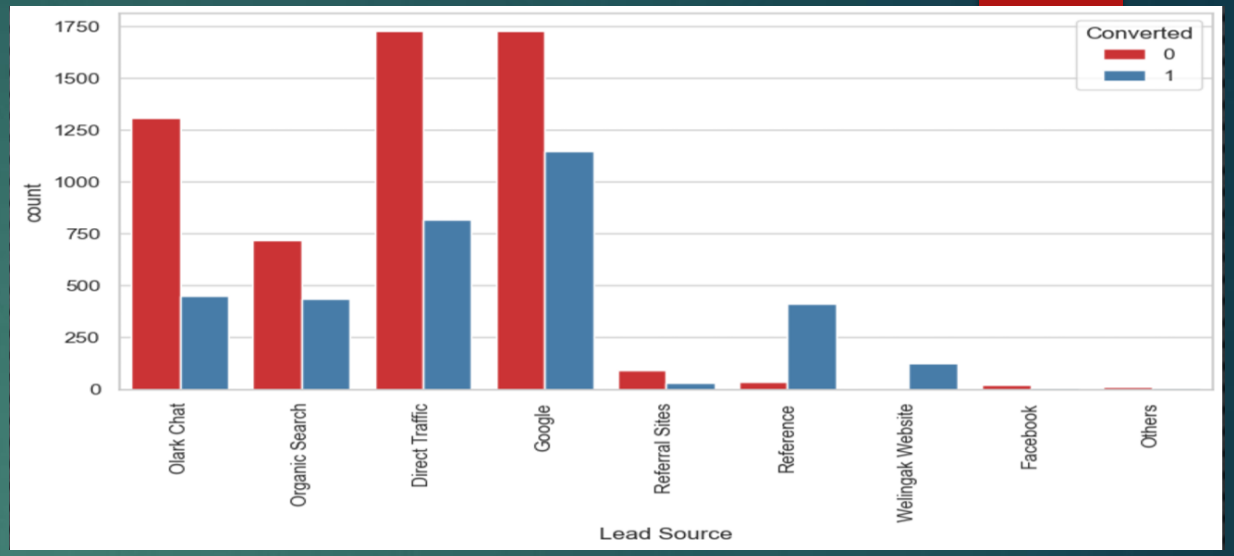
They want the model to be able to handle future constraints as well, like peak time actions required, how to utilize full manpower, and after achieving the target, what should be the approaches.

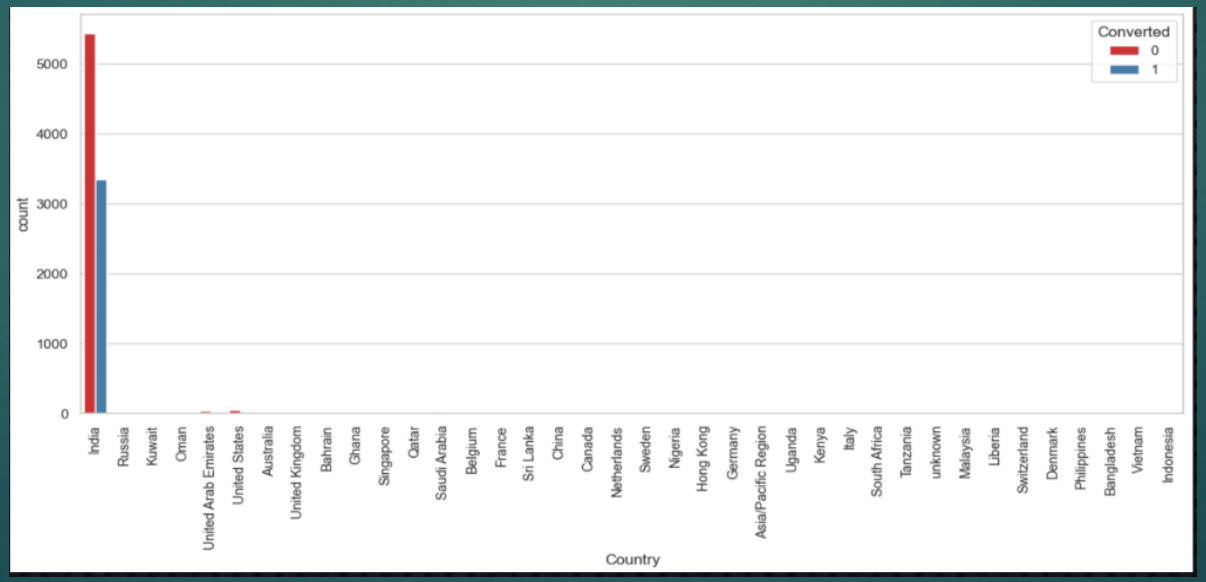
Problem Approach

- Importing the data and inspecting the data frame
- Data Preparation
- EDA
- Dummy variable creation
- Test Train Split
- Corrections
- Features Scaling
- Model Building (RFI Required VIF and PVALUES)
- Making Prediction On Test Set
- Model Evalution

EDA-Data Cleaning

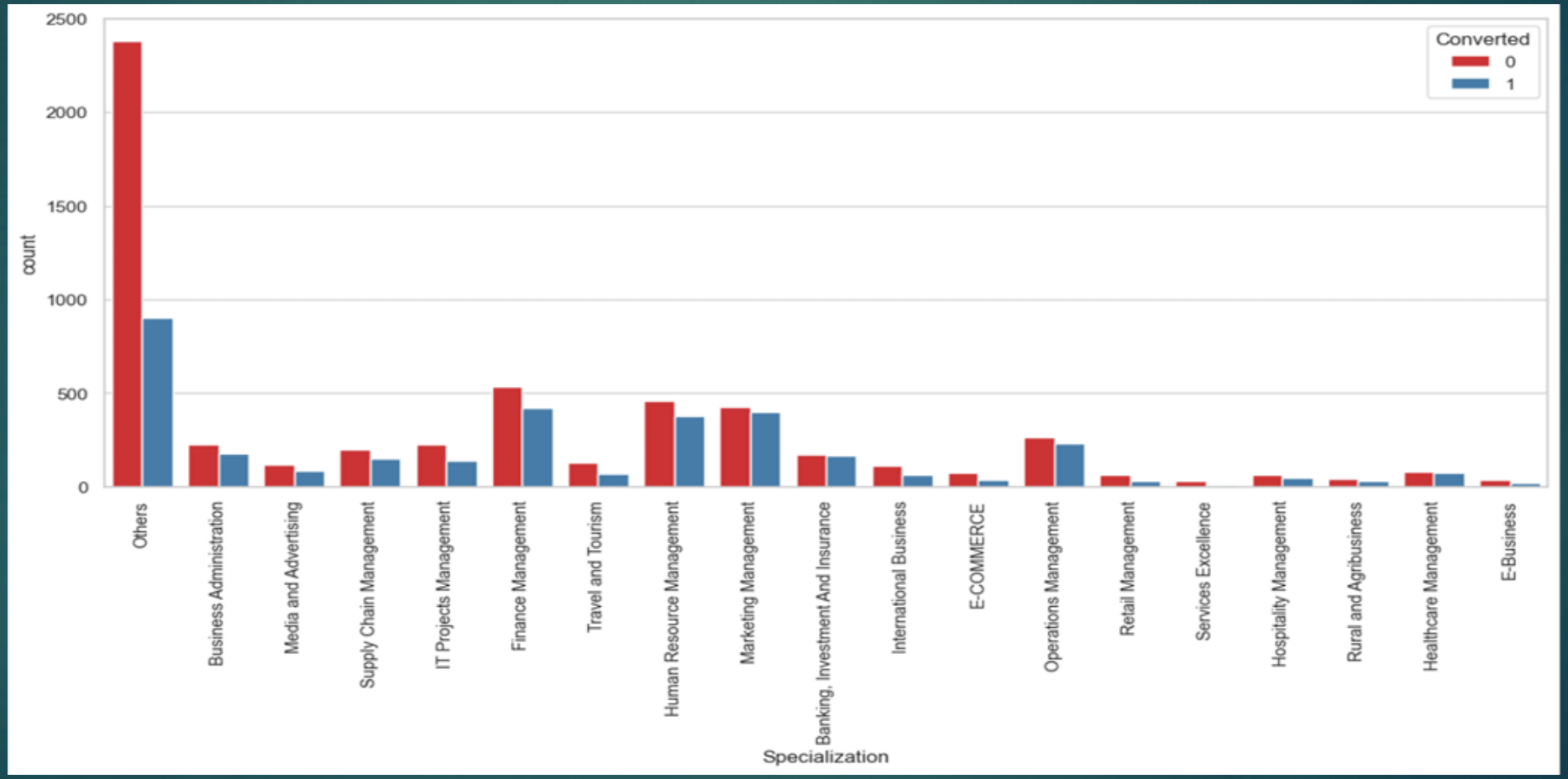






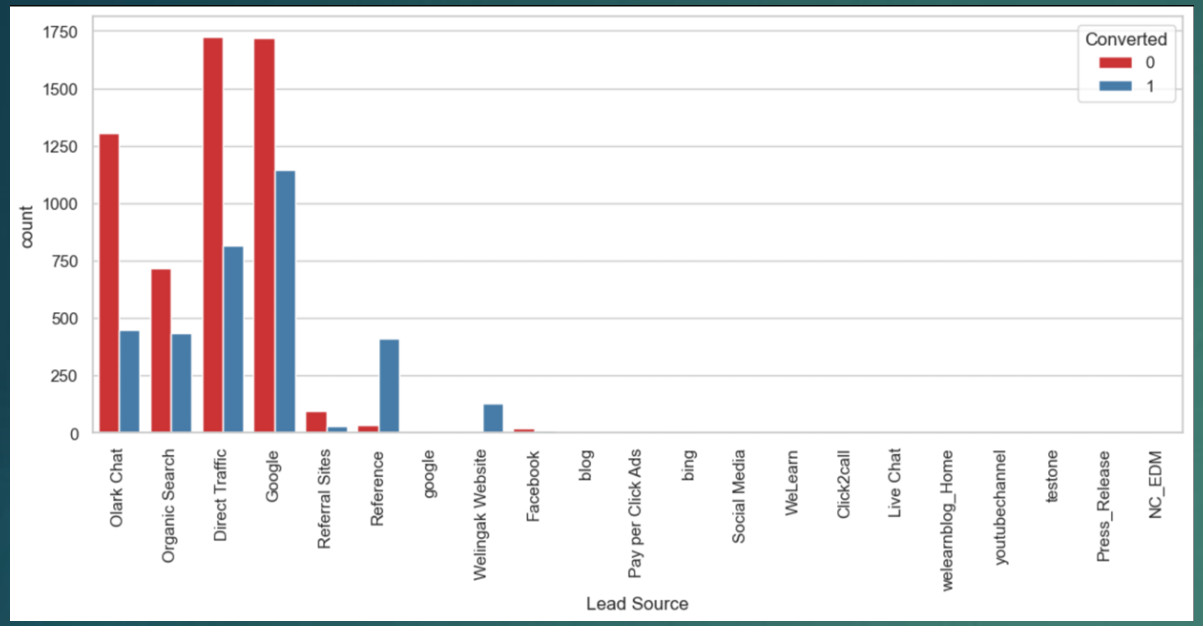
Specilization

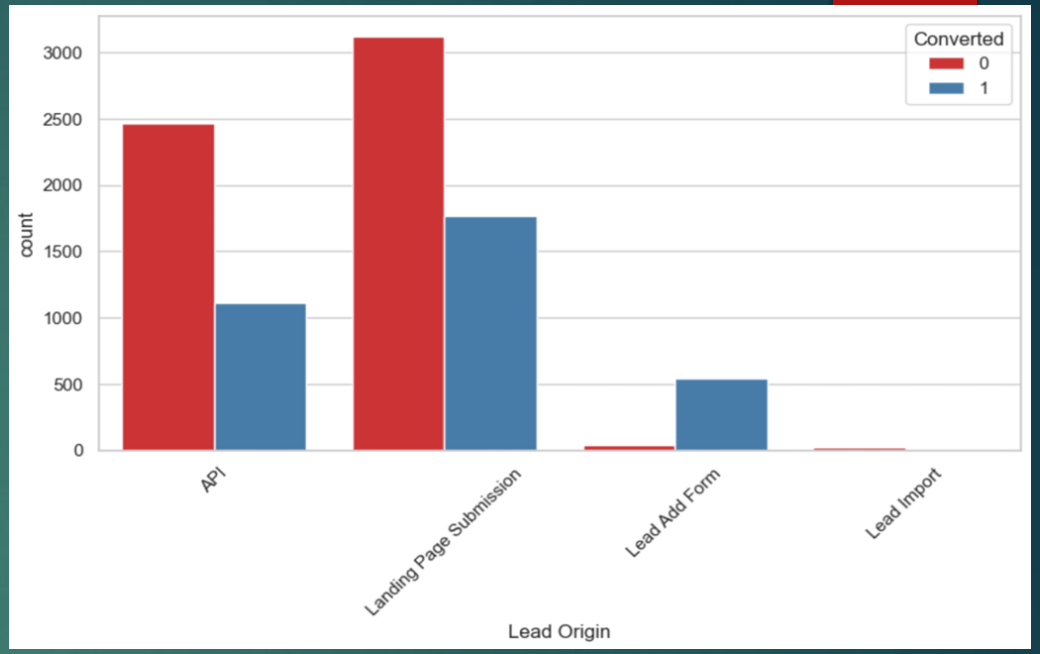
Lead From HR, Finance And Marketing Management Specilization are High Probability To Convert.

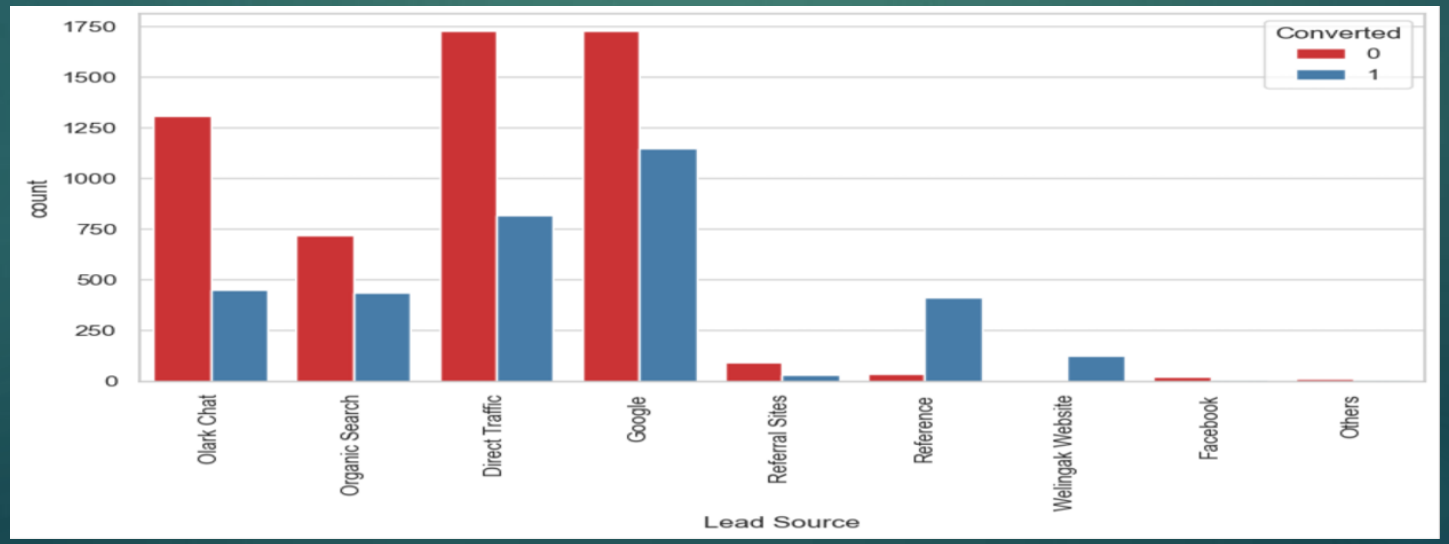


LEAD SOURCE AND LEAD

Origin in lead source the lead through Google and Direct Traffic Hhi Probability To Convert

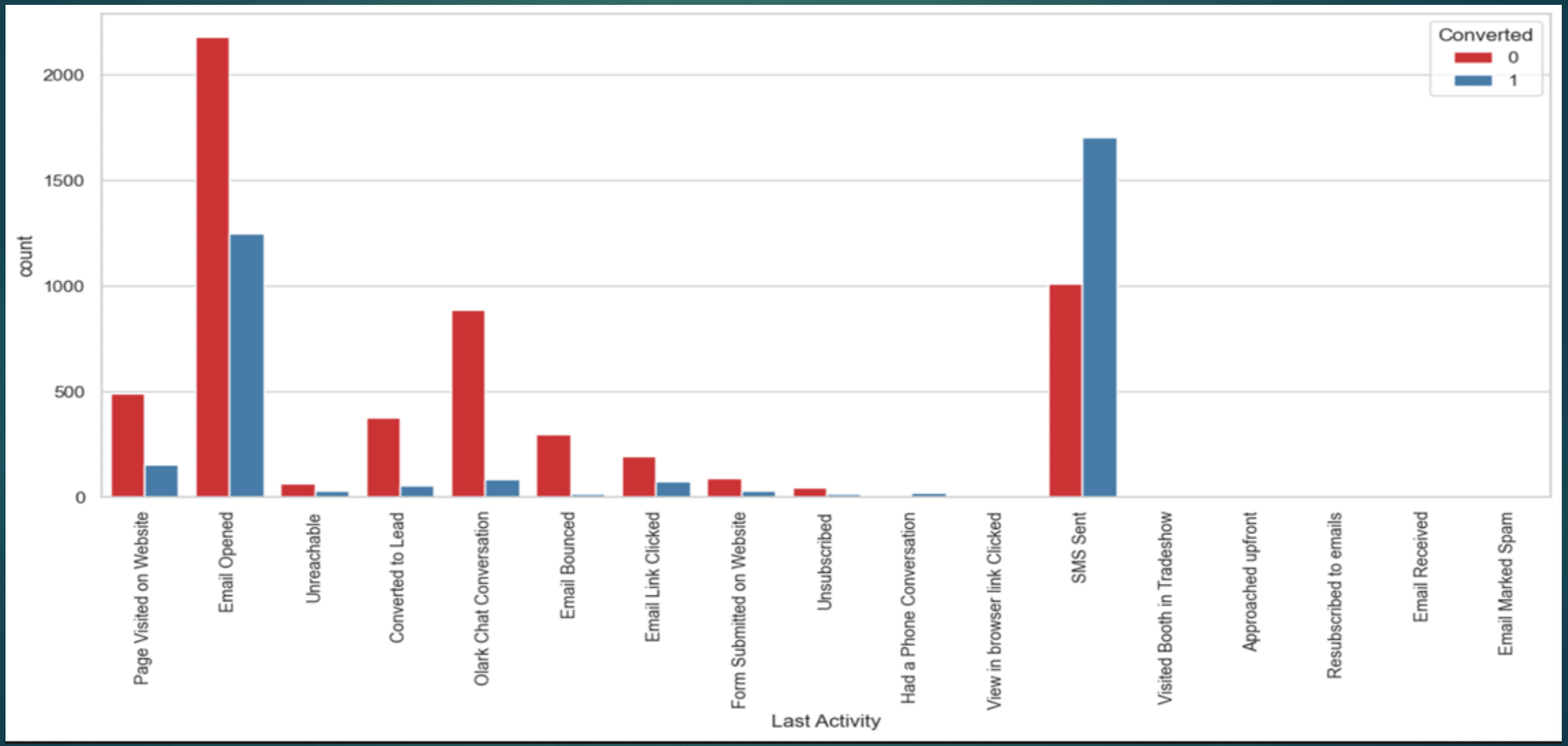






LAST LEAD ACTIVITY

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

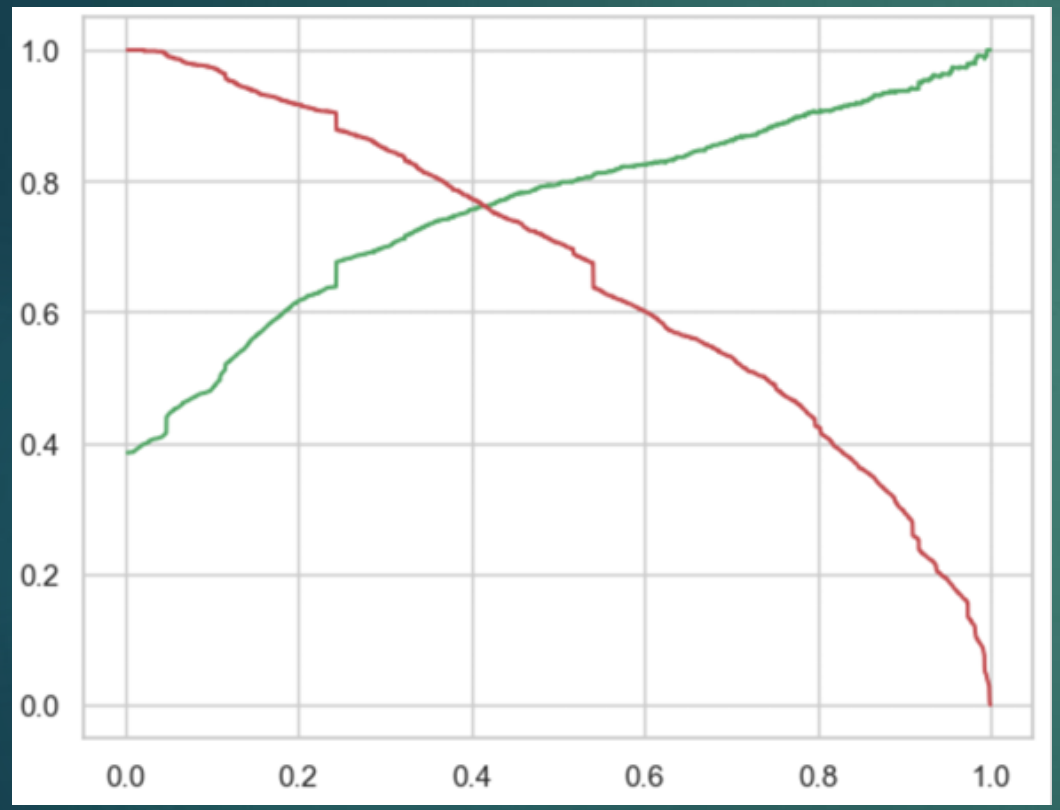


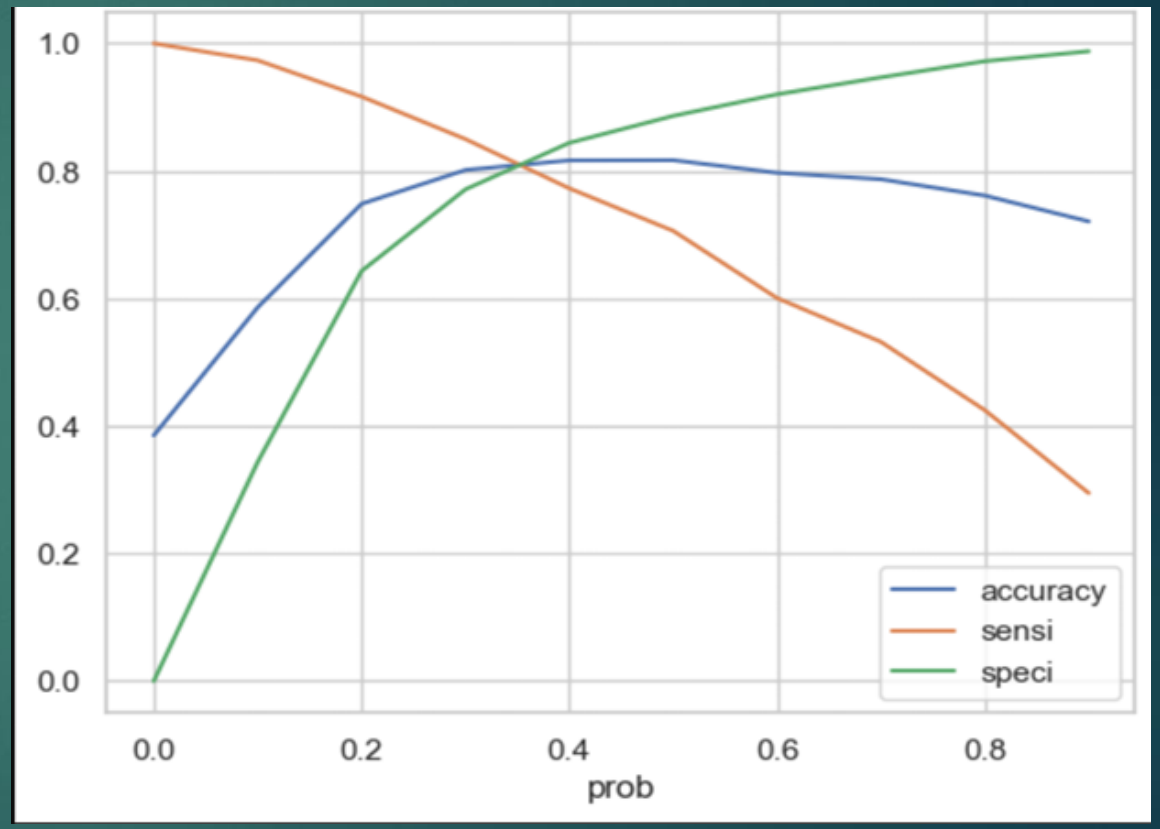
MODEL EVALUATION

ROC CURVE

0.42 IS THE TRADEOFF BETWEEN PRECISION AND RECALL THUS WE CAN TO CONSIDER ANY PROSPECT LEAD WITH CONVERSION PROBABILITY HIGHER

TO BE A HOT LEAD





OBSERVATIONS

- TRAINDATA: ACCURACY: 80%
- SENSITIVITY: 77%
- SPECIFICITY: 80%
- TESTDATA:
- ACCURACY: 80%
- SENSITIVITY: 77%
- SPECIFICITY: 80%
- FINALFEATURESLIST:
- LEADSOURCE_OLARKCHAT
- SPECIALIZATION_OTHERS
- LEADORIGIN_LEADADDFORM
- LEADSOURCE_WELINGAKWEBSITE
- TOTALTIMESPENTONWEBSITE
- LEADORIGIN_LANDINGPAGESUBMISSION
- WHAT ISYOURCURRENTOCCUPATION_WORKINGPROFESSIONALS
- DONOTEMAIL

CONCLUSION

- WESEETHATTHECONVERSIONRATEIS30-35% (CLOSETOAVERAGE) FORAPI AND LANDINGPAGESUBMISSION. BUTVERYLOWFORLEADADDFORMANDLEADIMPORT.

THEREFOREWECANINTERVENETHATWENEEDTOFOCUSMOREONTHELEADSORIGINATED FROMAPI ANDLANDINGPAGESUBMISSION.

- WESEEMAXNUMBEROFLEADSAREGENERATEDBYGOOGLE/ DIRECTTRAFFIC. MAX CONVERSIONRATIOISBYREFERENCEANDWELINGAKWEBSITE.
- LEADSWHOSPENTMORETIMEONWEBSITE, MORELIKELYTOCONVERT.
- MOSTCOMMONLASTACTIVITYISEMAILOPENED. HIGHESTRATE= SMS SENT. MAXARE UNEMPLOYED. MAXCONVERSIONWITHWORKINGPROFESSIONAL.