



Lead Scoring Casestudy.

Created by Hitesh Sundrani, Aman Dhama & Rupesh Kumar



Table of Content

- Problem Statement
- Problem Approach
- Exploratory Data Analysis
- Corelations
- Model Evaluation
- Observations
- Conclusions



Problem Statement

- 1) 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 process of form filling on their website after which the company tags individual as a lead.
- 2) 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.
- 3) 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.
- 4) 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



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 want to achieve a lead conversion rate of 80%.
- They want the model that can handle future constraints like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.

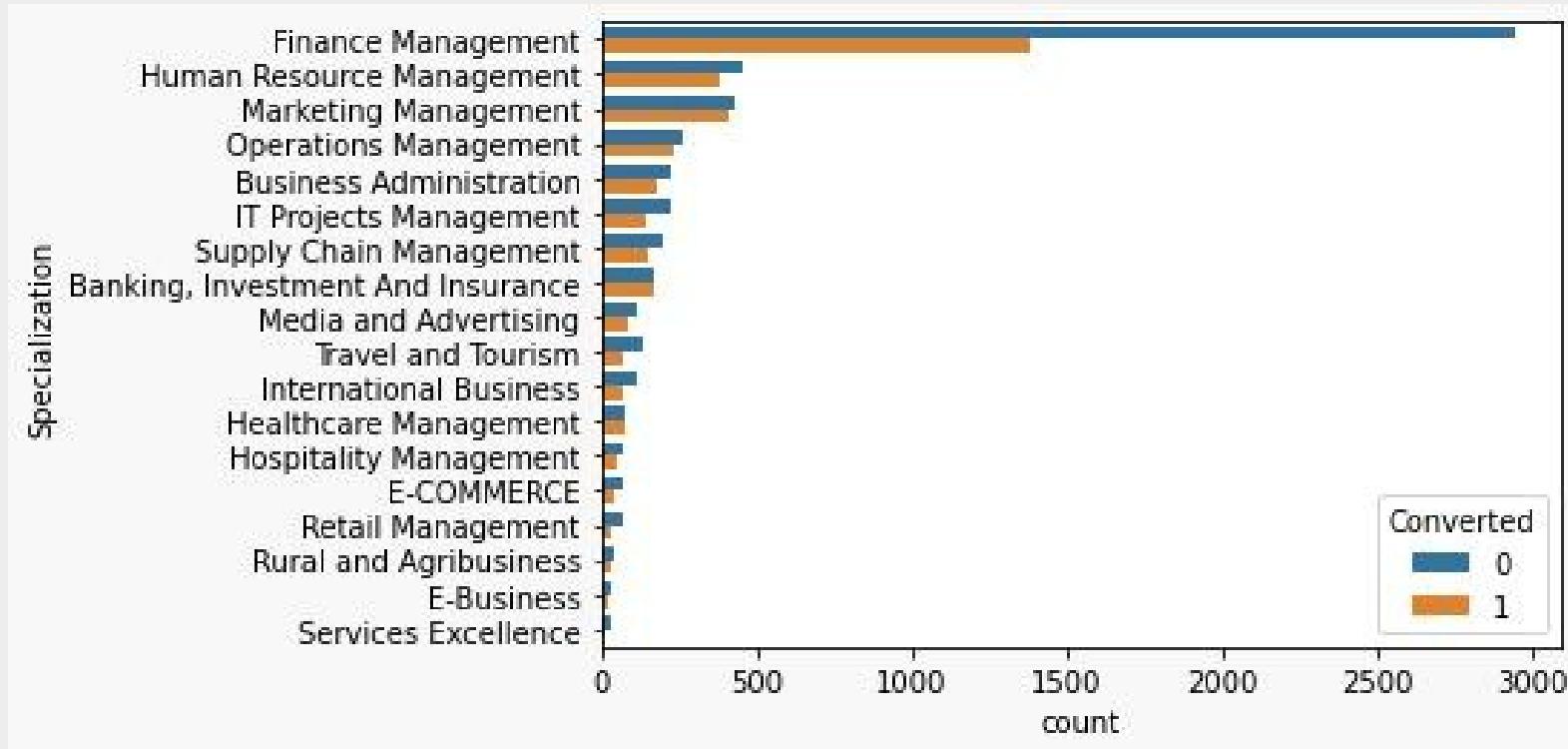
Approaching the Problem



1. Importing data and libraries
2. Data Preparation
3. Exploratory Data Analysis(EDA)
4. Dummy variable creation
5. Test-Train split
6. Feature scaling
7. Correlations
8. Model Building (RFE R-squared VIF and p-values)
9. Model Evaluation
10. Making predictions on test set

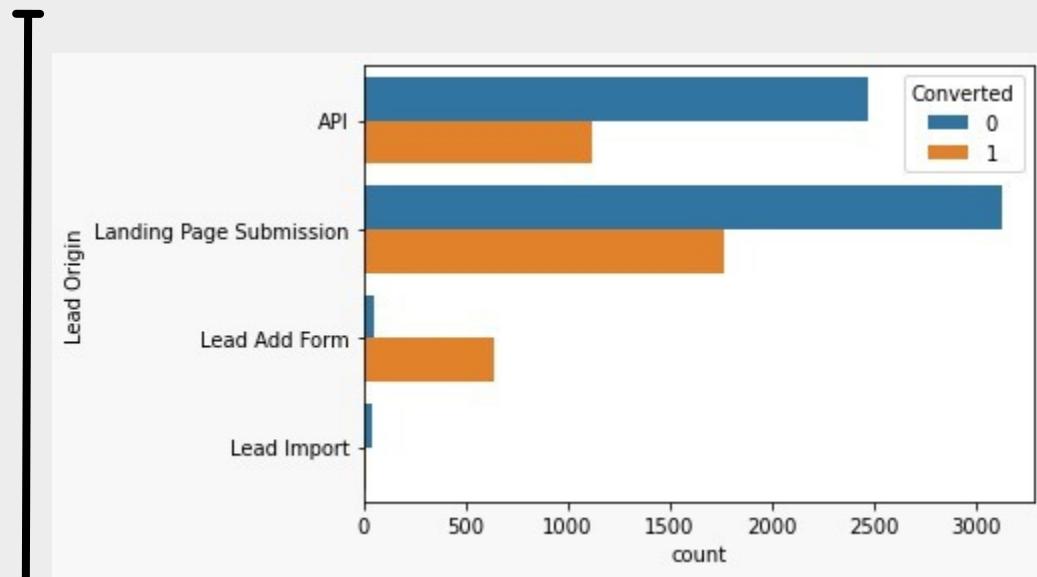
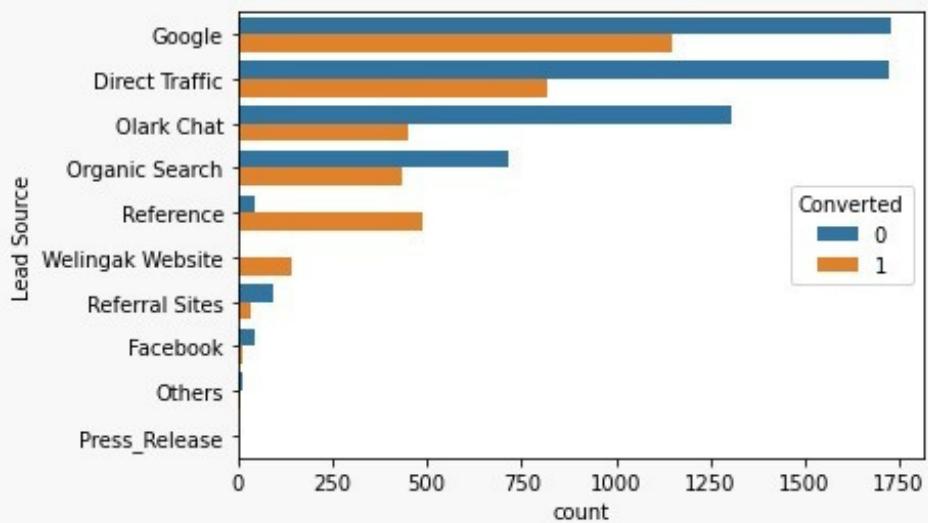


Conversion Probability



Leads from Finance, Human Resources(HR) and Marketing specialization have the highest probability of converting

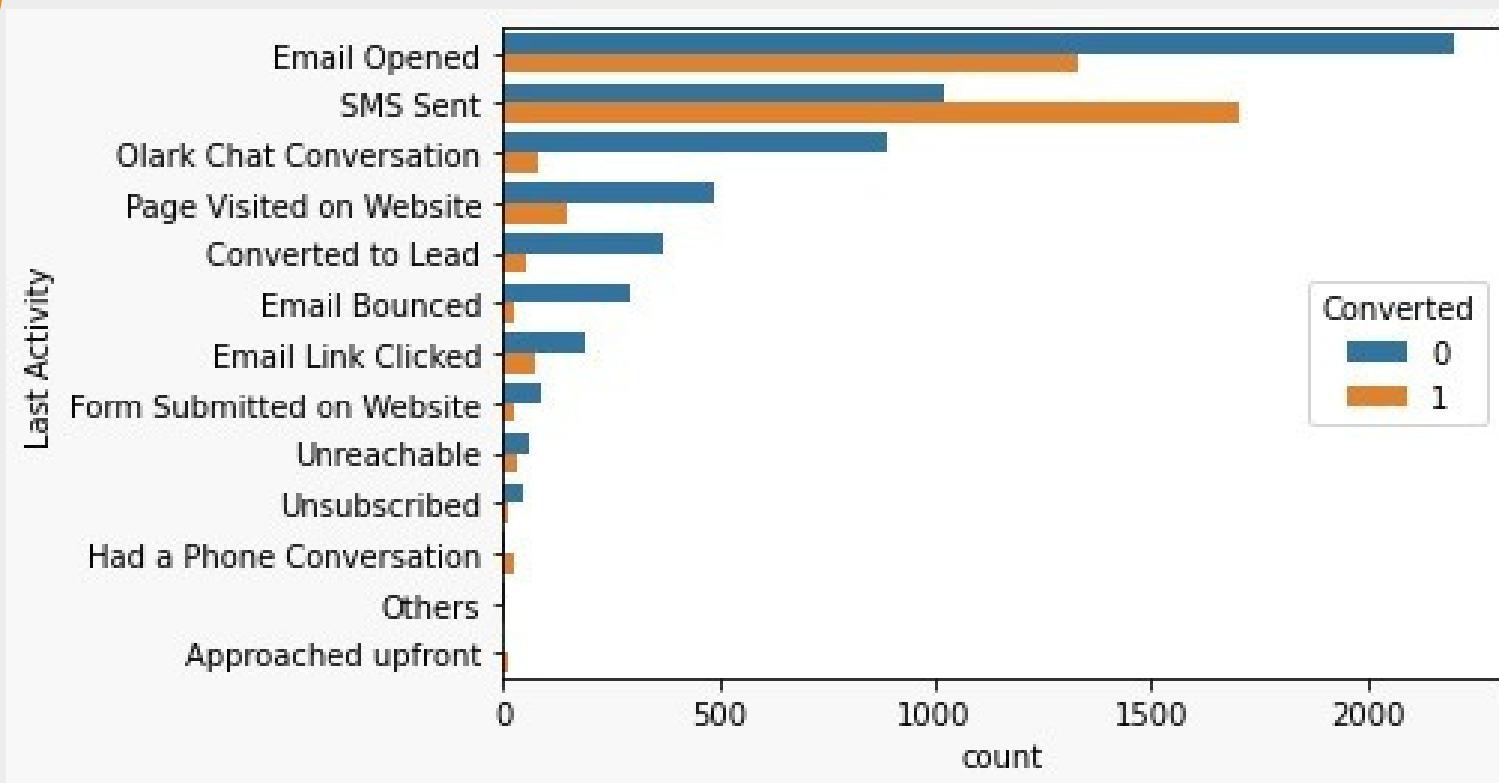
Lead Source and Origin



- The leads through Google and direct traffic have the highest probability of getting converted

- The leads through Landing Page Submission have the highest probability of getting converted

Lead Activity

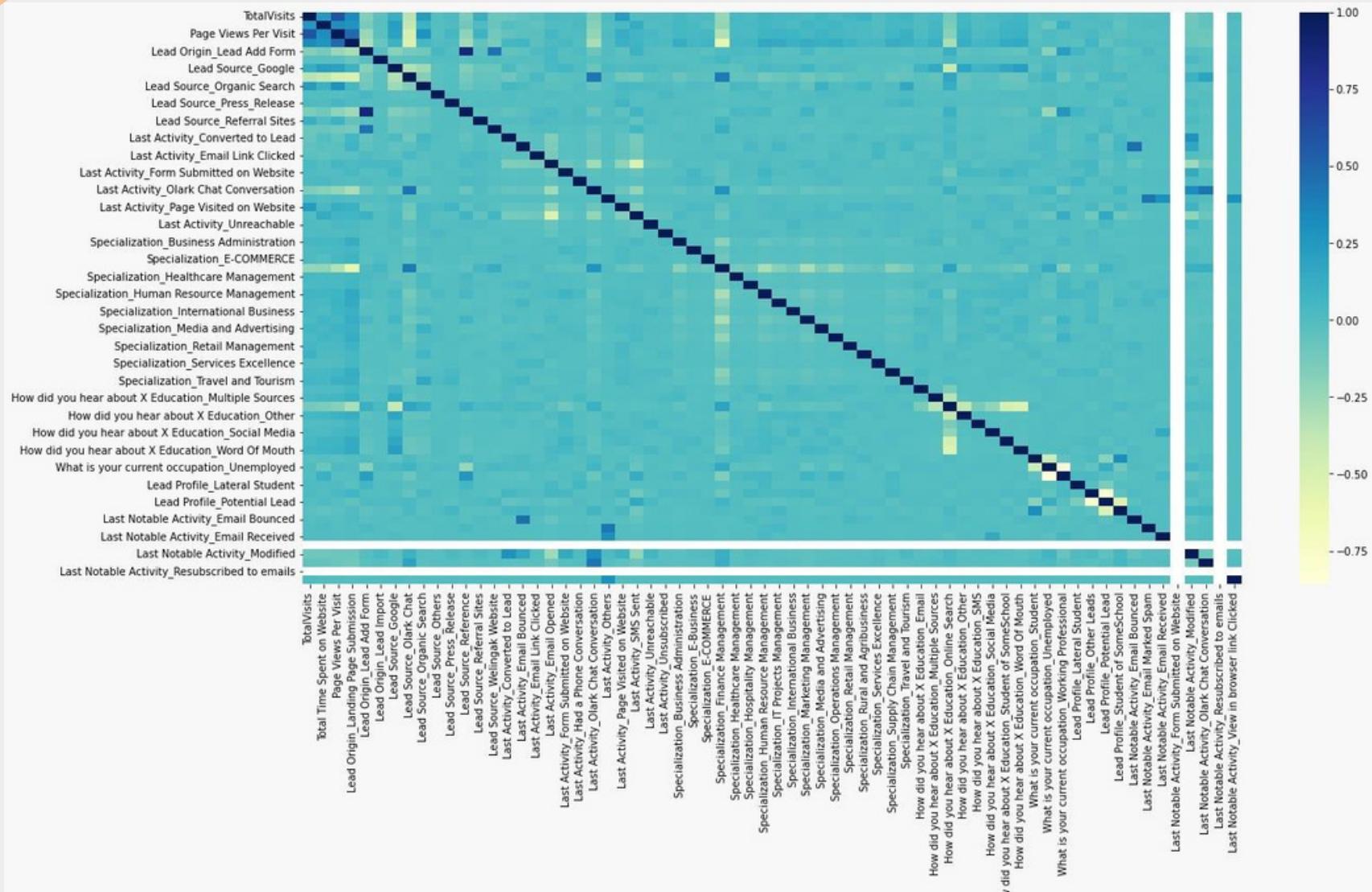


Email Opened and SMS sent have the highest rate
of conversion based on the lead activity



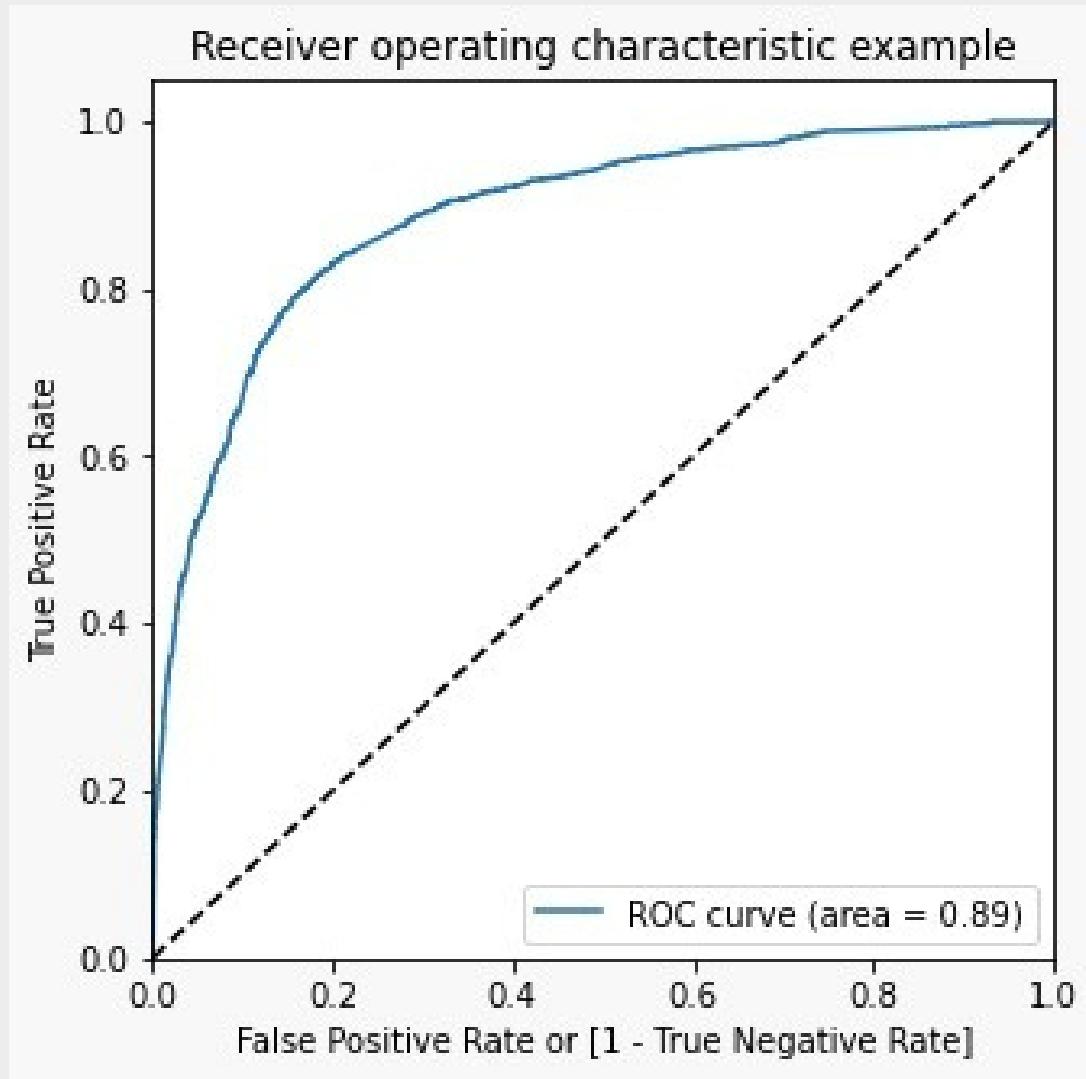


Correlation



No Correlations can be observed

MODEL EVALUATION



The tradeoff between precision and recall is 0.42, hence we can say that any prospect with probability higher than 42% is a hot lead

OBSERVATIONS



Train Data:

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

Test Data:

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%



Final Features List:

*Lead Source_ Olark chat

*Lead Origin_ Lead Add Form

*Specialization_ Others

*Lead Source_ Welingak Website

*Lead Origin_ Landing Page submission

*Total Time Spent on website

*What is your current occupation

*Do Not Email

*Working professionals

CONCLUSION



- We see that the conversion rate is 30-35% on an average for API and Landing page submission.
- The conversion rate is very low for Lead Add form and Lead import. Therefore we can deem that we need to focus more on the leads originated from API and Landing page submission.
- We can see maximum numbers of leads are generated by Google/direct traffic. Maximum conversion ratio is by reference and Welingak website.
- People who spent more time on website, more likely to convert to a lead
- Most common last activity is email opened.
- Maximum conversion occur with working professional.