Lead Scoring Case Study

Submitted By:

Garima Rai
Rosy Sharma
Sachin Jha

Business Objective

- In order to facilitate X Education in identifying the most prospective leads—those with the highest likelihood of transitioning into paying customers—we aim to optimize the lead selection process.
- Aim is to develop a logistic regression model for assigning lead scores ranging from 0 to 100 to each potential lead. This model will serve as a valuable tool for the company in strategically targeting and prioritizing potential leads for more effective engagement.

Problem Statement

Company

X Education is an online education company that sells online courses to industry professionals.

Context

The company does marketing of its courses on various websites and search engines such as Google. Once the visitors land on the website, they browse through different courses or fill up a form for the course or watch related videos. When the visitors fill up a form and provide their contact details, they are classified as leads.

Problem statement

X Education gets a lot of leads, the lead conversion rate is low i.e., about 38%. The company wants to identify the best leads, i.e. the leads that have high conversion probability. The company wants to assign a lead score to each lead. The customers with a higher lead score have a higher conversion probability and the customers with a lower lead score have a lower conversion probability.

Data Cleaning

Handling Null values

The columns with more than 40% of null values were dropped

- How did you hear about X Education (76%)
- Lead Profile(74%)
- Lead Quality (52%)
- Asymmetrique Profile Score, Asymmetrique Activity Score, Asymmetrique Activity Index, Asymmetrique Profile Index (46% for all four)

Dropping redundant columns

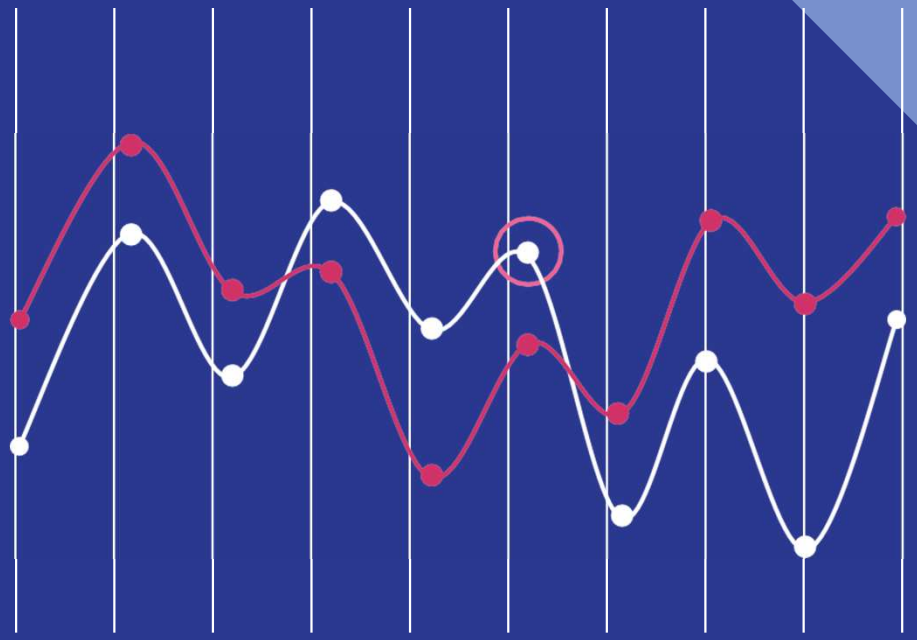
The columns with redundant informations were dropped

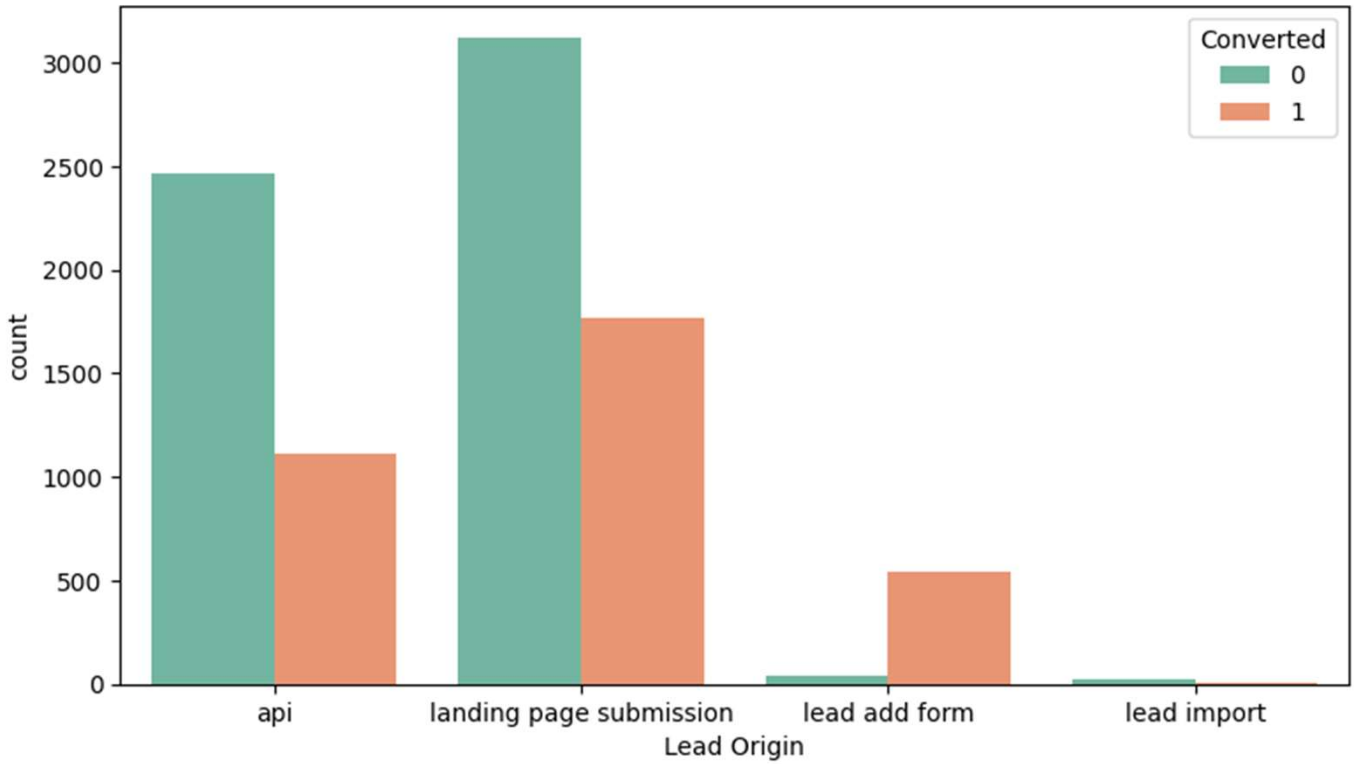
- Updates
- Last_activity
- Prospect_ID

Imputing missing values

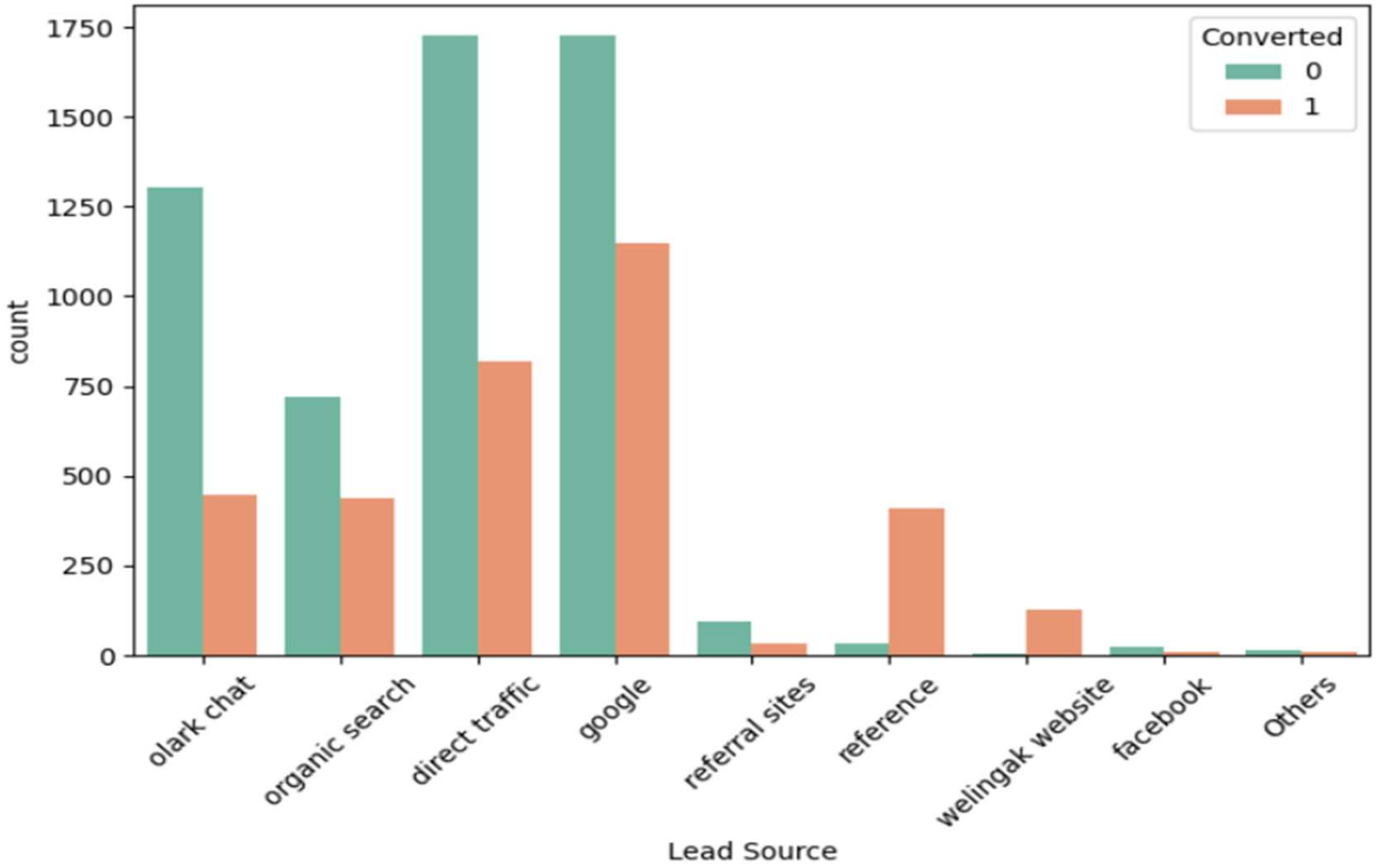
Missing Values are imputed with median values.

EXPLORATORY DATA ANALYSIS

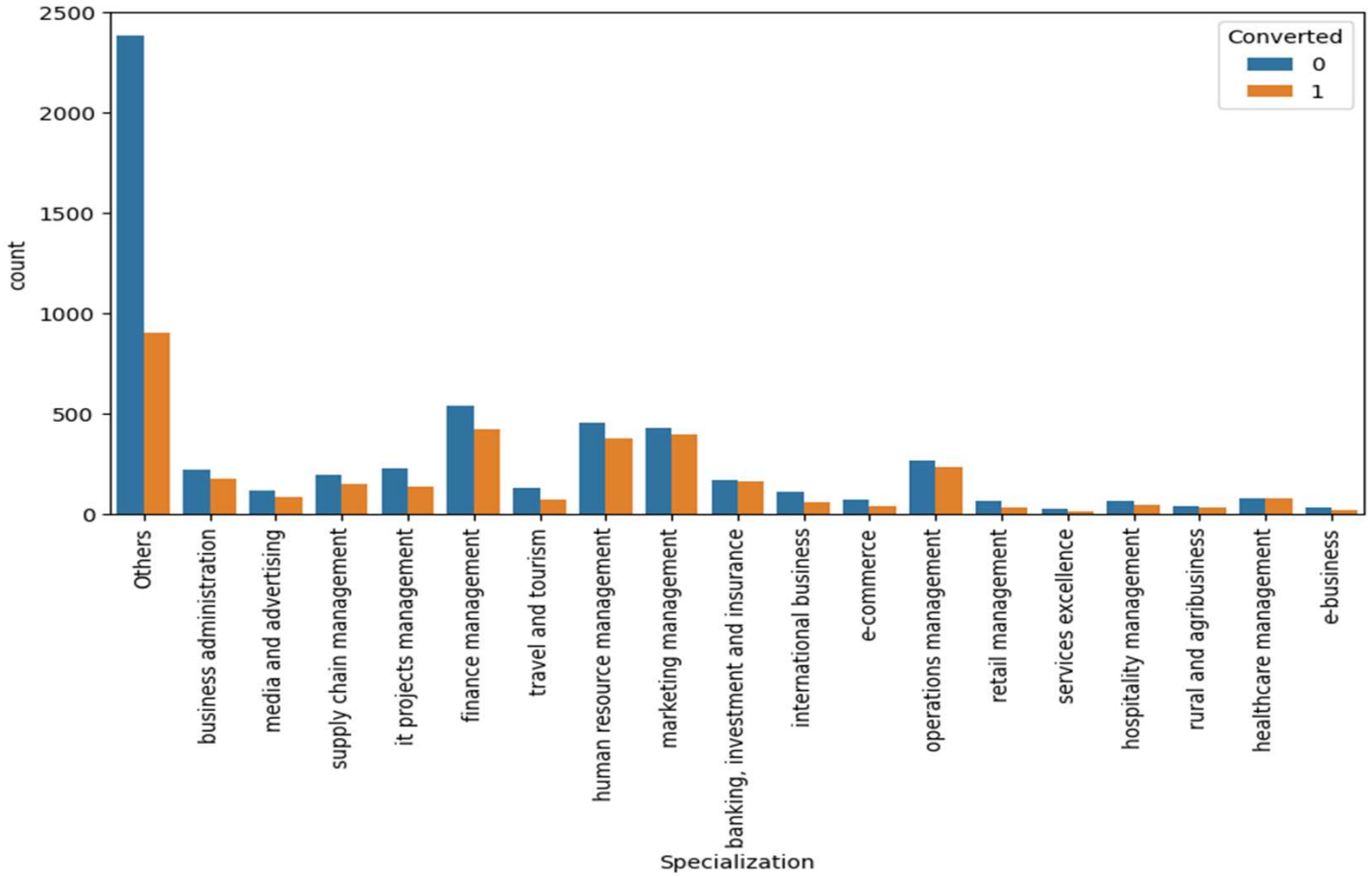




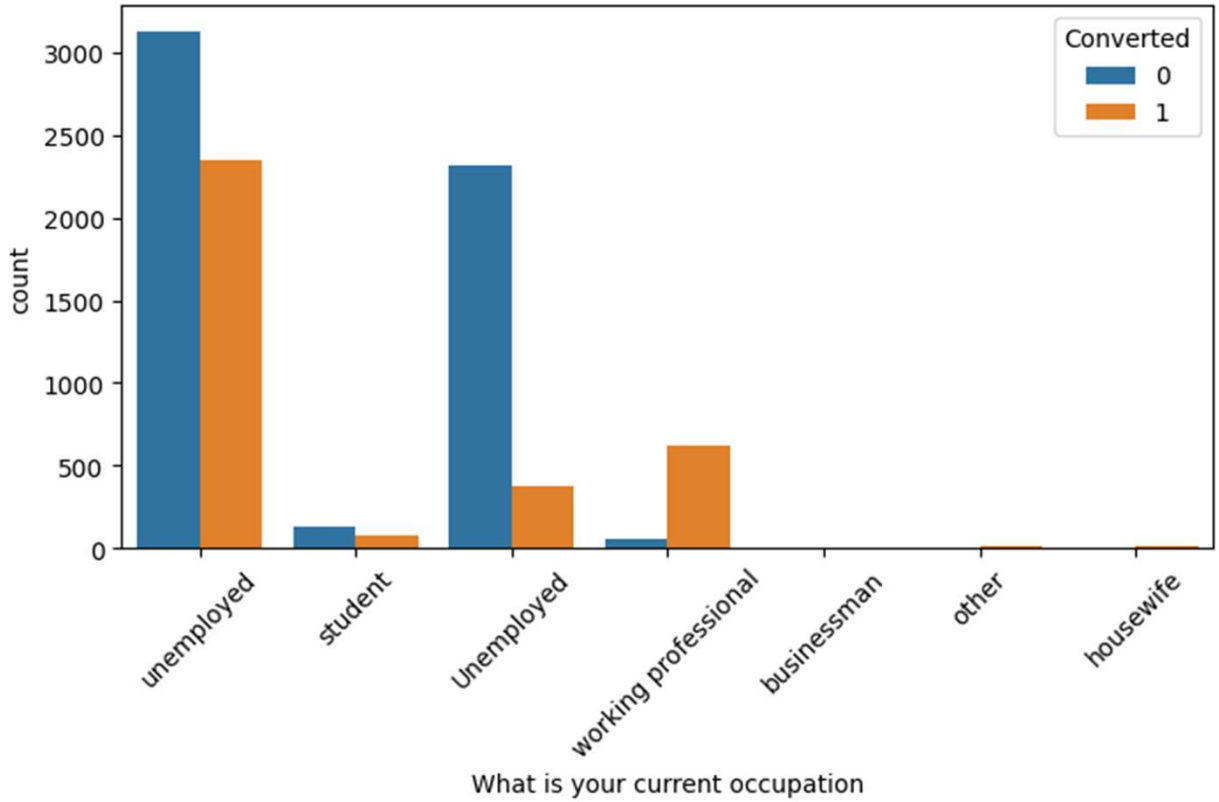
1. API and Landing Page Submission have more conversion rate.
2. Lead Add Form has more conversion rate but count of lead are not very high Lead Import are very less in count.



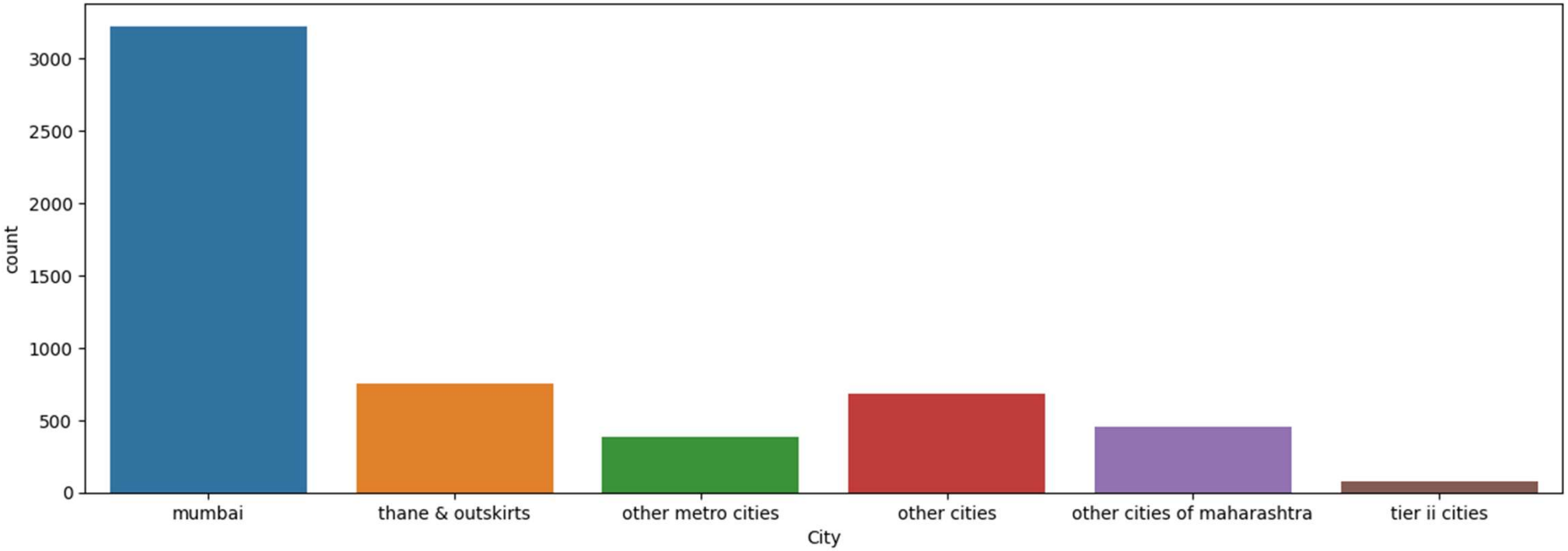
Leads sourced from Google or Direct Traffic are having high conversion rate



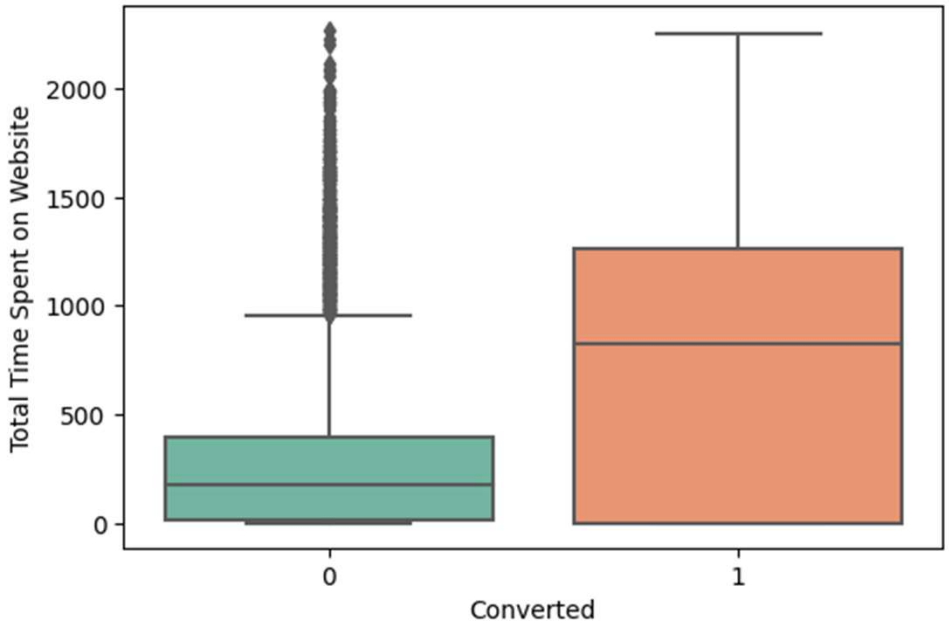
Lead from Finance management specialization having highest conversion rate.



- Leads with occupation type as Working professional will have high conversion rate.
- We are having the highest converted leads from the unemployed occupation type but the conversion ratio is less than 50%



Most of the leads are from the Mumbai city and having the conversion ratio more than 50%.



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

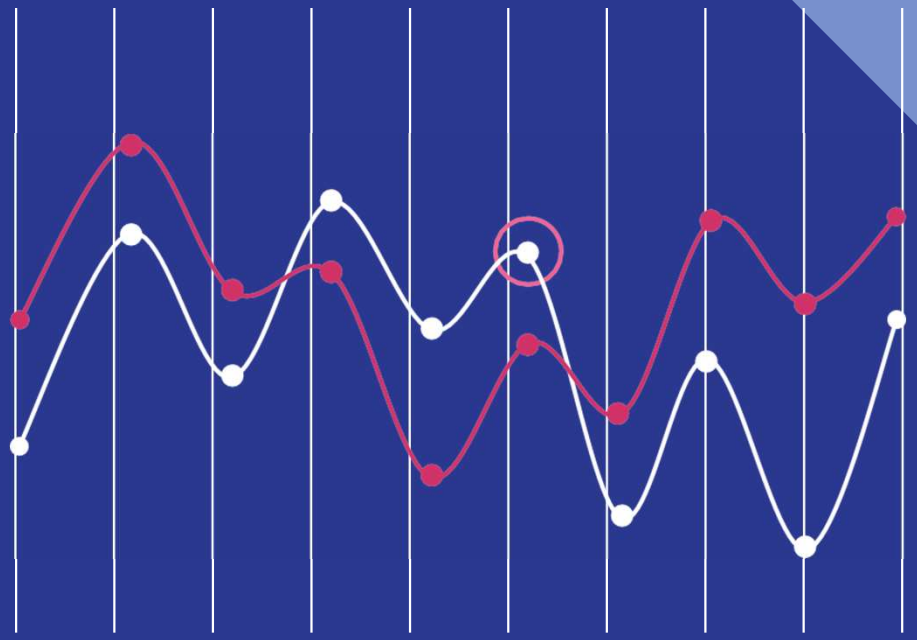
Analysis – Inferences I

- Lead origin
  - Maximum leads are originated from Landing Page Submission and API
  - Conversion rate is also high from these
- Lead Source
  - Maximum leads are sourced from Google and Direct Traffic
  - Conversion rate is also high for these
- Country
  - ~96% of leads are from India
- Specialization
  - Most of the leads have `Finance and Human Resource Management` specialization
- Occupation
  - 85% of leads are unemployed with high conversion rate
- Reason_for_Choosing
  - ~99% of leads are choosing the platform for better career prospects
- City
  - Most of the leads are from Mumbai with high conversion rate

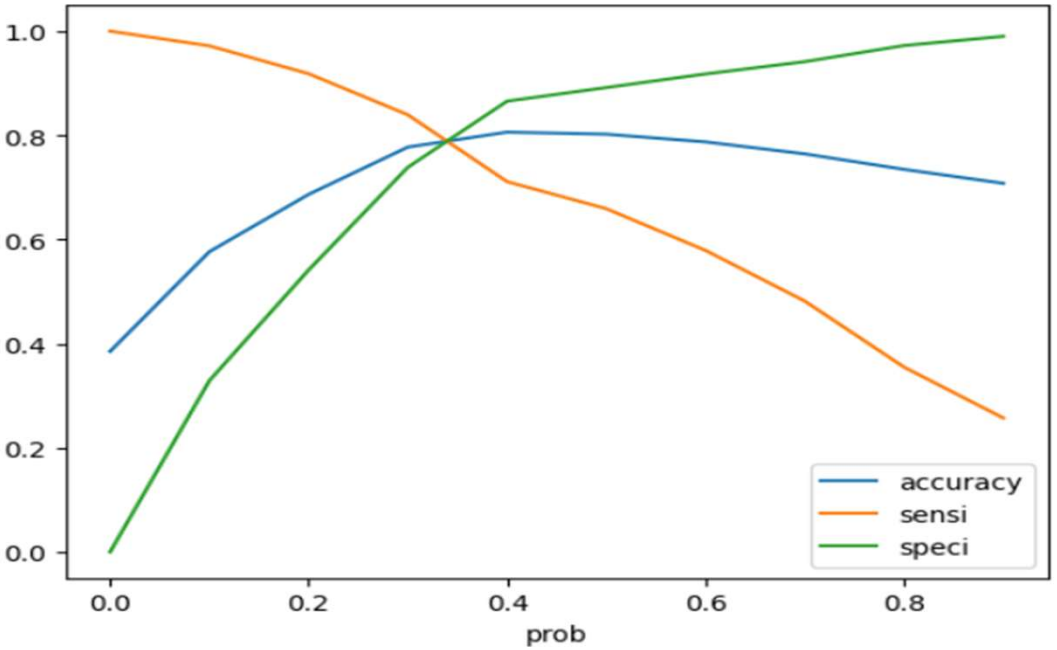
Analysis – Inferences II

- Leads who are sourced from organic search have high page views per visit and have the highest conversion, followed by Google and Direct Traffic
- Leads who are sourced from organic search have spent most of the time on website and have the highest conversion, followed by Google and Direct Traffic
- Leads who are sourced from organic search have most number of visits on the website and have the highest conversion, followed by Google and Direct Traffic

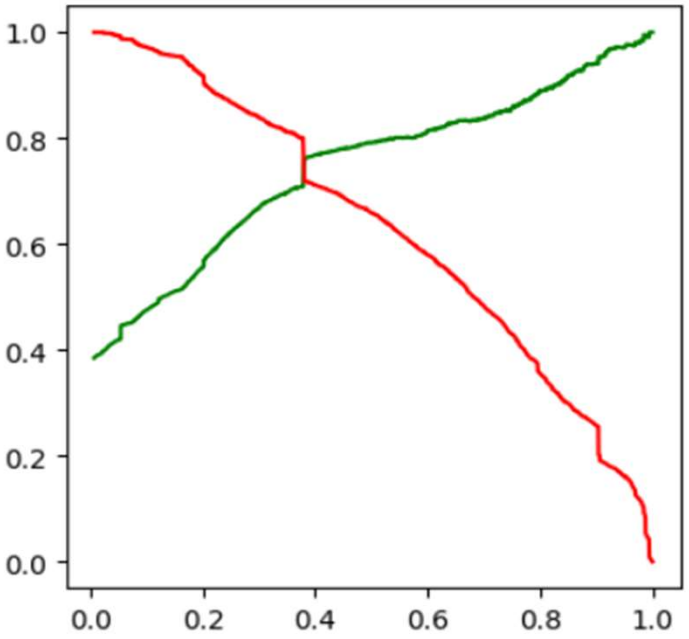
Model Details



Model Evaluation -Train Dataset



The graph shows the optimal cut off of 0.3 based on Accuracy, Sensitivity and Specificity.



The graph shows an optimal cut off of 0.4 based on Precision and Recall

Model Evaluation -Test Dataset

Confusion Matrix	
1338 (TN)	396 (FP)
188 (FN)	801 (TP)

Accuracy
79%

Sensitivity
81%

Specificity
79%

Precision
80%

Recall
66%

Conclusion

- The leads having high ‘Lead Score’ can be focused on more for better conversion rate.
- Marketing on Google, since the conversion rate from the traffic from Google is high.
- Encouraging existing converted leads for referrals by providing some incentives for the referrals.
- Since the number of leads is high in Mumbai as compared to other major cities, the company can increase marketing in the other cities as well to achieve more leads.
- The unemployed category can be focused on more and also individuals having Finance Management as specialization.
- Focus on the students can be minimized since the conversion rate is significantly low.