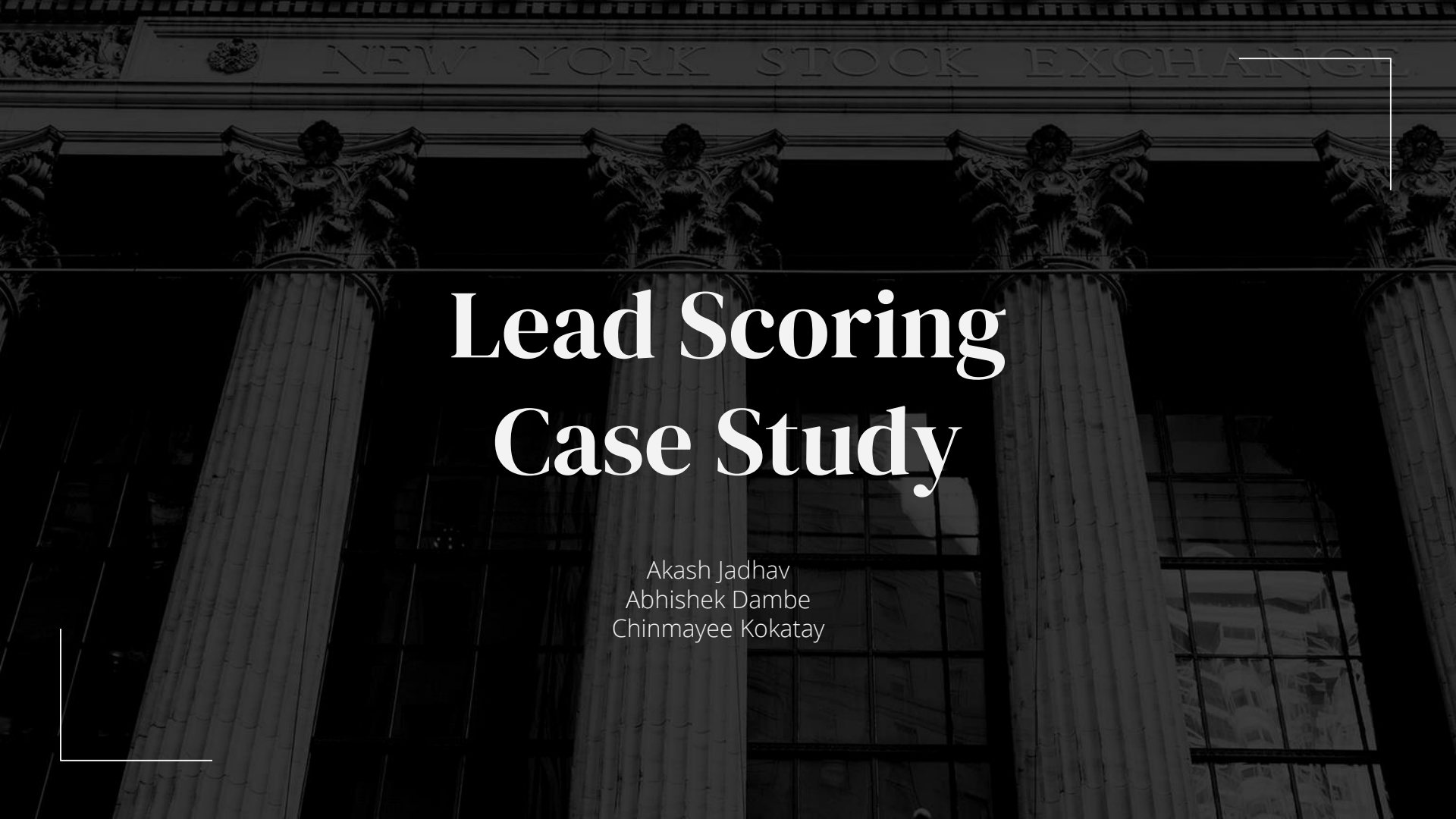



# Lead Scoring Case Study

Akash Jadhav
Abhishek Dambe
Chinmayee Kokatay

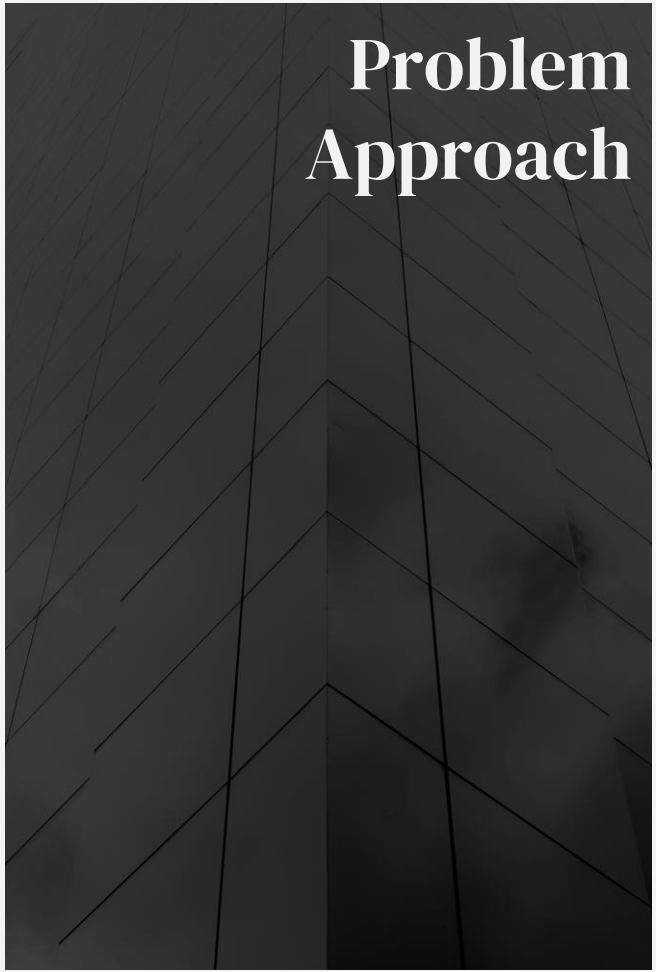
# Table of contents

# Problem statement

X Education company wants to select the most promising leads. The company wanted to build a model wherein it is needed to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, has given a ballpark of the target lead conversion rate to be around 80%.

# Business Objective

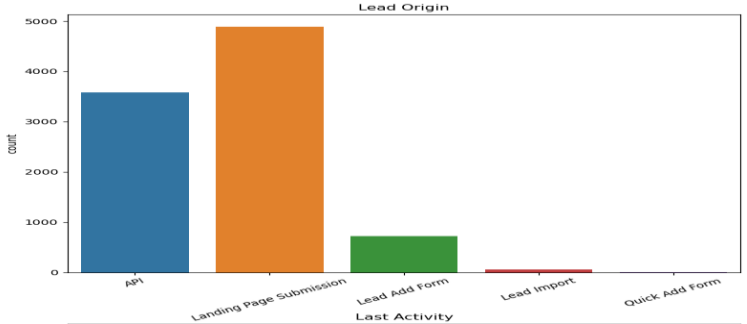
Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the hot leads and increase their conversion rate as well.
The CEO want to achieve a lead conversion rate of 80%.
They want the model to be able to handle future constraints.

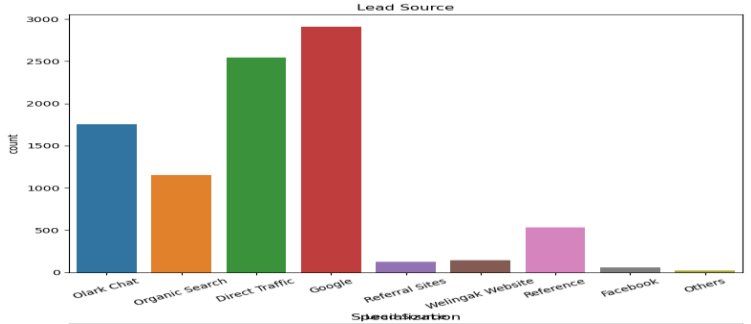


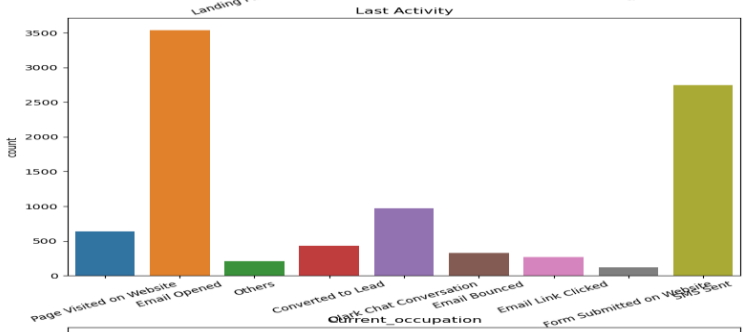
# Problem Approach

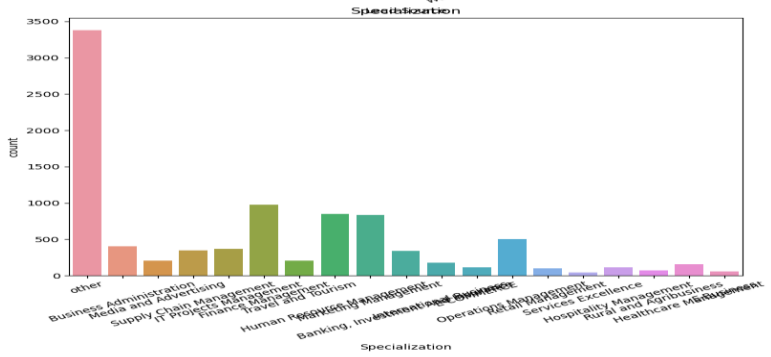
- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and pvalues)
- Model Evaluation
- Making predictions on test set

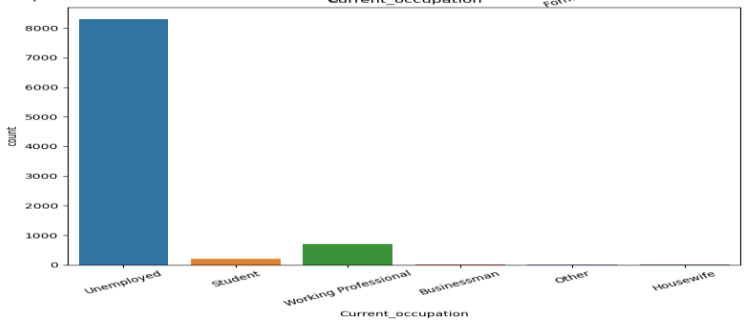
# EDA


Lead Origin
count
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form

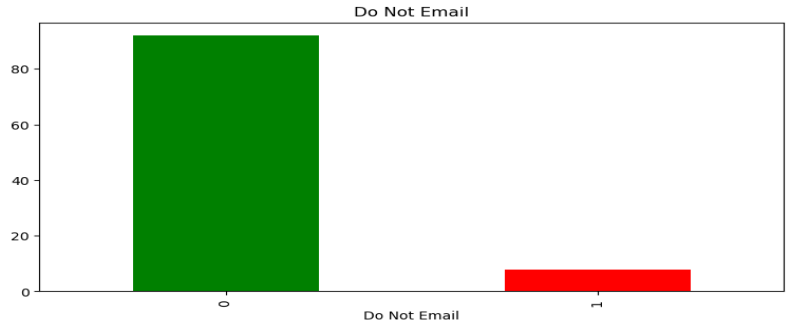

Lead Source
count
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Welingak Website
Reference
Facebook
Others

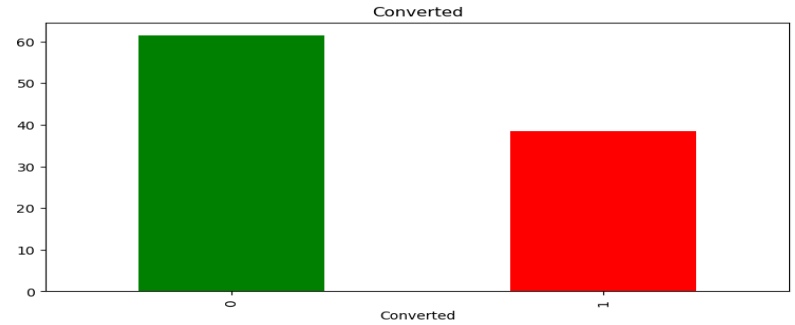

Last Activity
count
Page Visited on Website
Email Opened
Others
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
SMS Sent

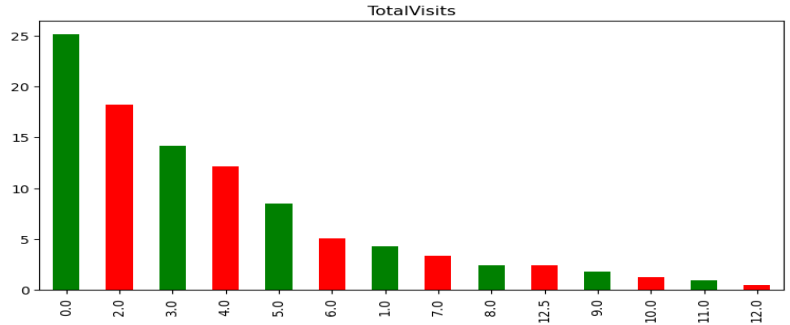

count
other
Business Administration
Media and Advertising
Supply Chain Management
Finance Management
Human Resource Management
Marketing Management
Operations Management
Services Excellence
Hospitality Management
Healthcare Management
Specialization

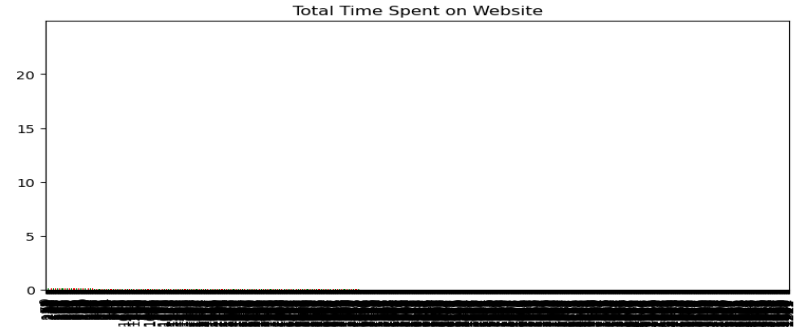

Current_occupation
count
Unemployed
Student
Working Professional
Businessman
Other
Housewife
Current_occupation

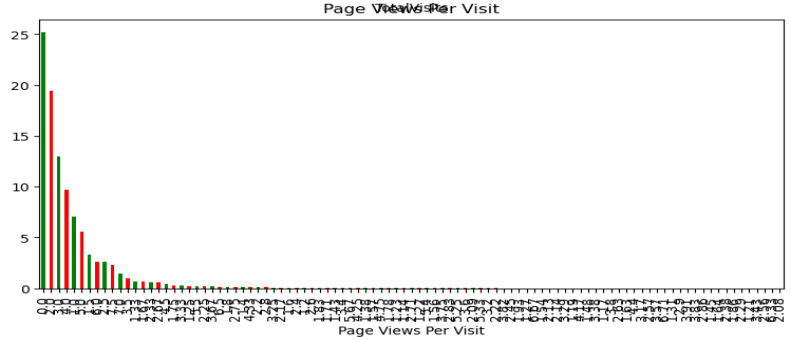
# Univariate Analysis for numerical variable



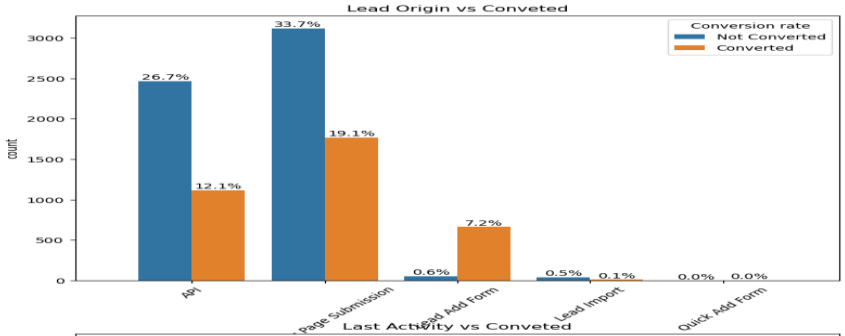



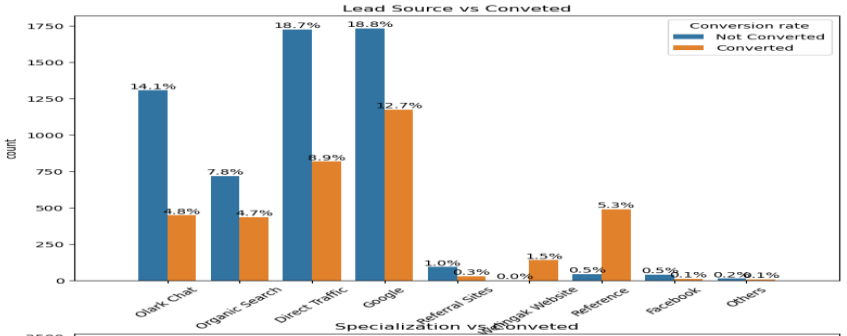



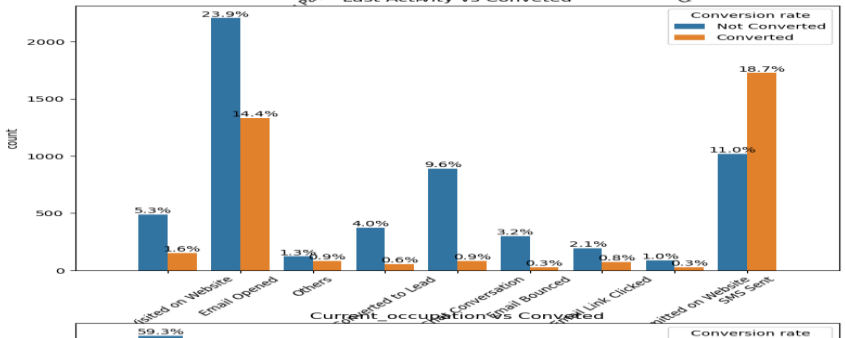



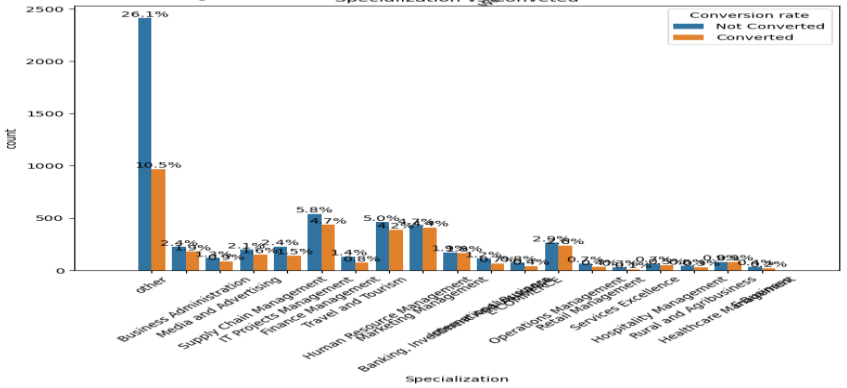



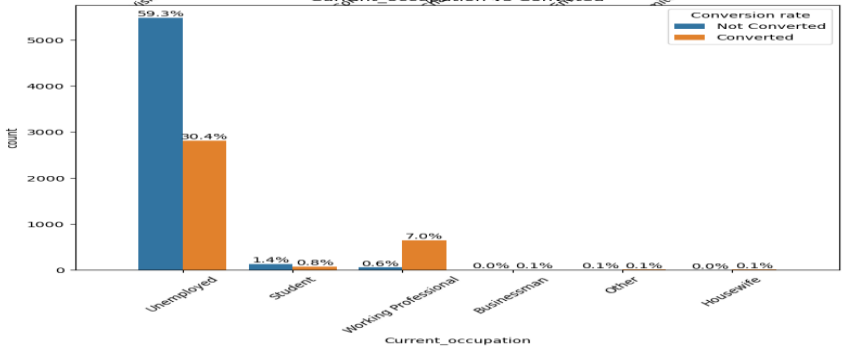

# Bivariate Analysis for categorical variable



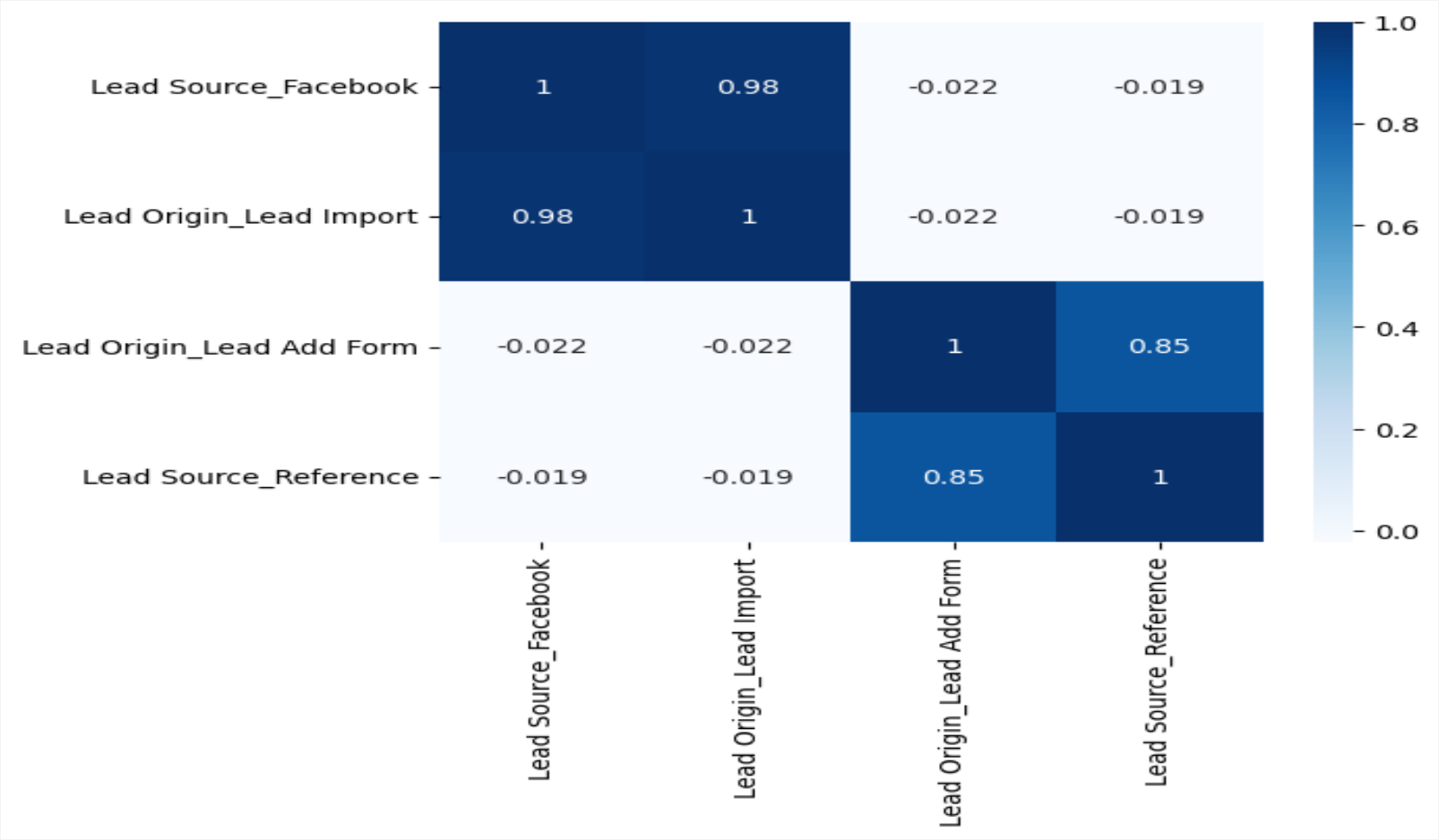



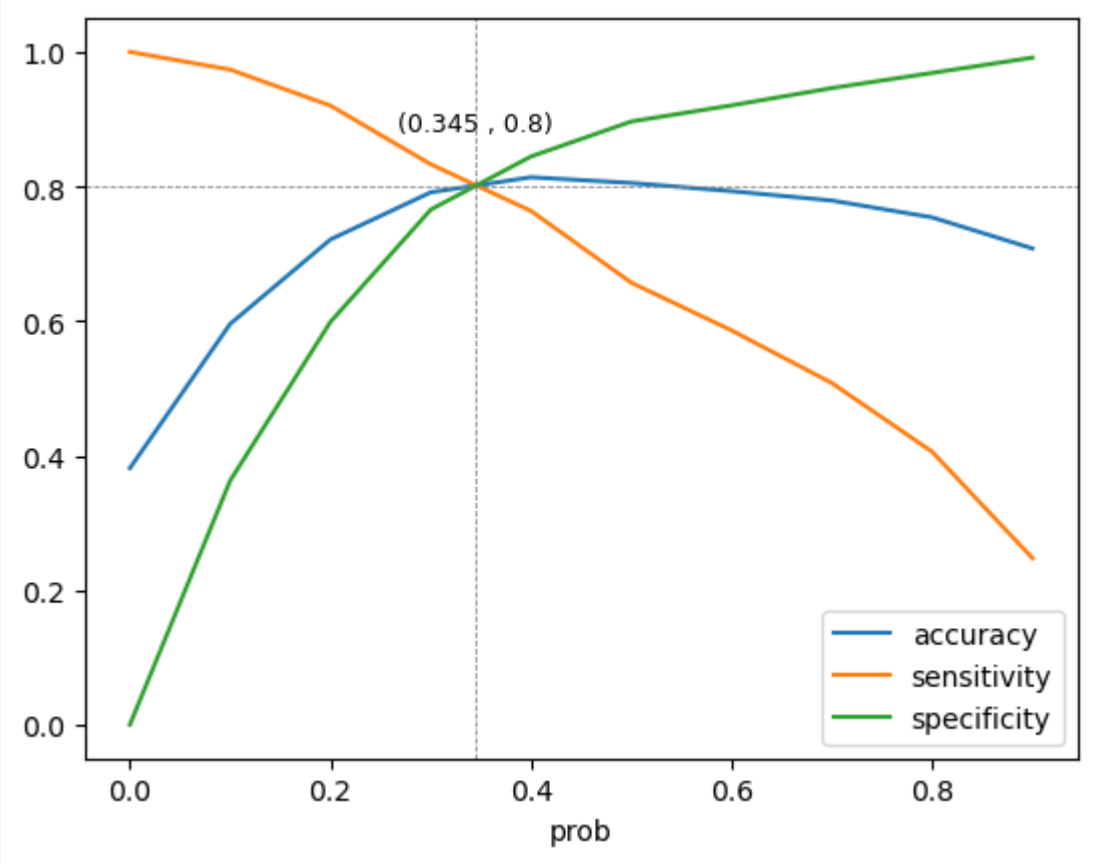



# Correlations



# ROC Curve


1.0
0.8
0.6
0.4
0.2
0.0
(0.345 , 0.8)
accuracy
sensitivity
specificity
0.0
0.2
0.4
0.6
0.8
prob

# Observations

## Train Data Set

- Accuracy: 80.46%

- Sensitivity: 80.05%

- Specificity: 80.71%

## Test Data Set

- Accuracy: 80.34%

- Sensitivity: 79.82% ≈ 80%

- Specificity: 80.68%

# Conclusions

The maximum number of leads are generated by google / direct traffic also the
conversion ratio  by reference and welingak website is also high.

The top three features contributing to hot leads are  Welingak Website, Reference and Working Professionals.

Leads who spend more time on website more likely to convert.