Lead Scoring Case study

Manas Prasun Pandey

Mahima Sharma

Manibalan Baskaran

Contents

- Problem Statement
- Objectives
- Exploratory Data Analysis
- Model Building
- Model Evaluation
- Observations
- Conclusion

Problem Statement

- X Education, an educational company, specializes in offering online courses tailored for professionals in various industries.

- Each day, a considerable number of professionals interested in these courses visit the company's website and explore the available courses. The company has an online form for these individuals to complete, at which point they are classified as leads.

- After acquiring these leads, the sales team engages in various activities such as making phone calls and sending emails to them. This process results in the conversion of some leads, while the majority remain unconverted.

- X Education typically experiences a lead conversion rate of approximately 30%. In practical terms, this means that if they obtain 100 leads in a single day, only around 30 of them ultimately become customers. To enhance the efficiency of this process, the company aims to identify the most promising leads, often referred to as "Hot Leads."

- By successfully identifying this subset of leads, the company anticipates an increase in the lead conversion rate. This improvement stems from the sales team's ability to concentrate their efforts on communicating with potential leads rather than contacting every lead indiscriminately.

Objective

- Lead X has requested the development of a model capable of assigning a lead score ranging from 0 to 100 to each lead. This scoring system will enable them to distinguish and prioritize Hot Leads, ultimately leading to an enhanced lead conversion rate.

- The CEO is determined to attain a lead conversion rate of 80%, reflecting a strong commitment to boosting the company's conversion performance.

- Additionally, the company seeks a model that can adapt to future challenges and constraints, including managing peak-time demands, optimizing workforce utilization, and defining strategies once the 80% conversion rate target is reached.

Exploratory Data Analysis

- Impute or remove missing values
- Handle outliers
- Standardise values
- Filter Data
- Univariate Analysis
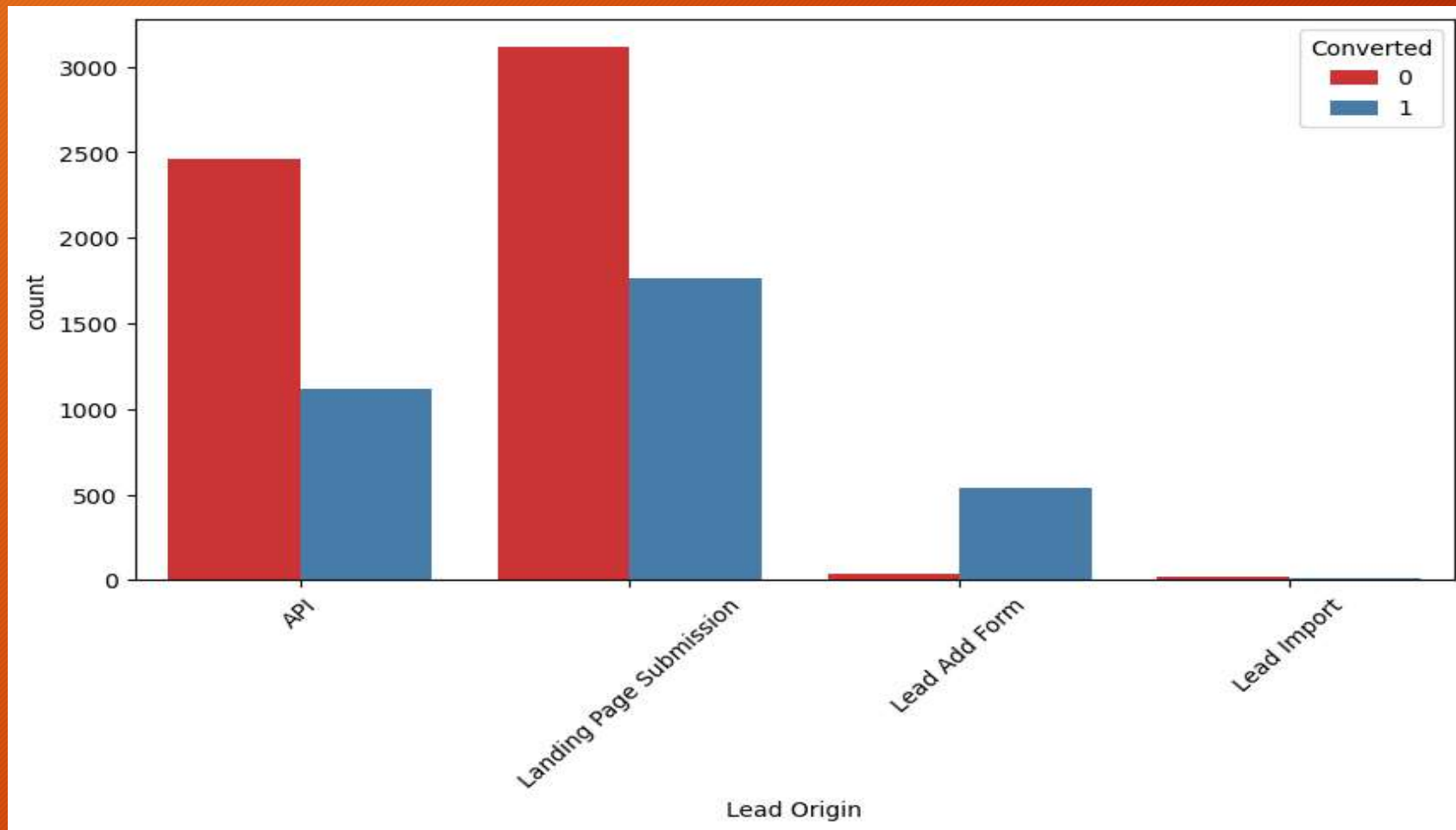- Multivariate Analysis

Specialization
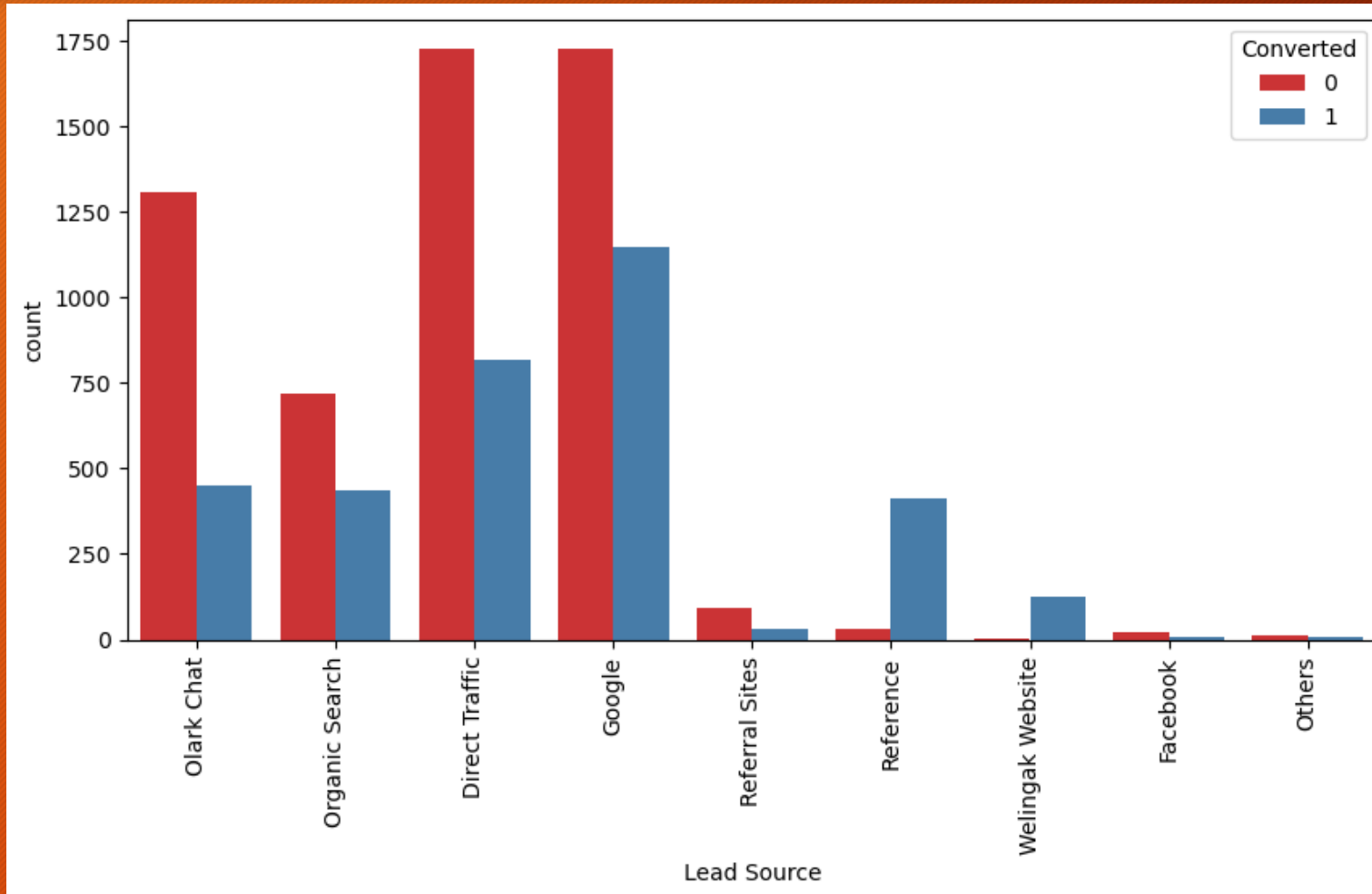
Imputing null values to Others
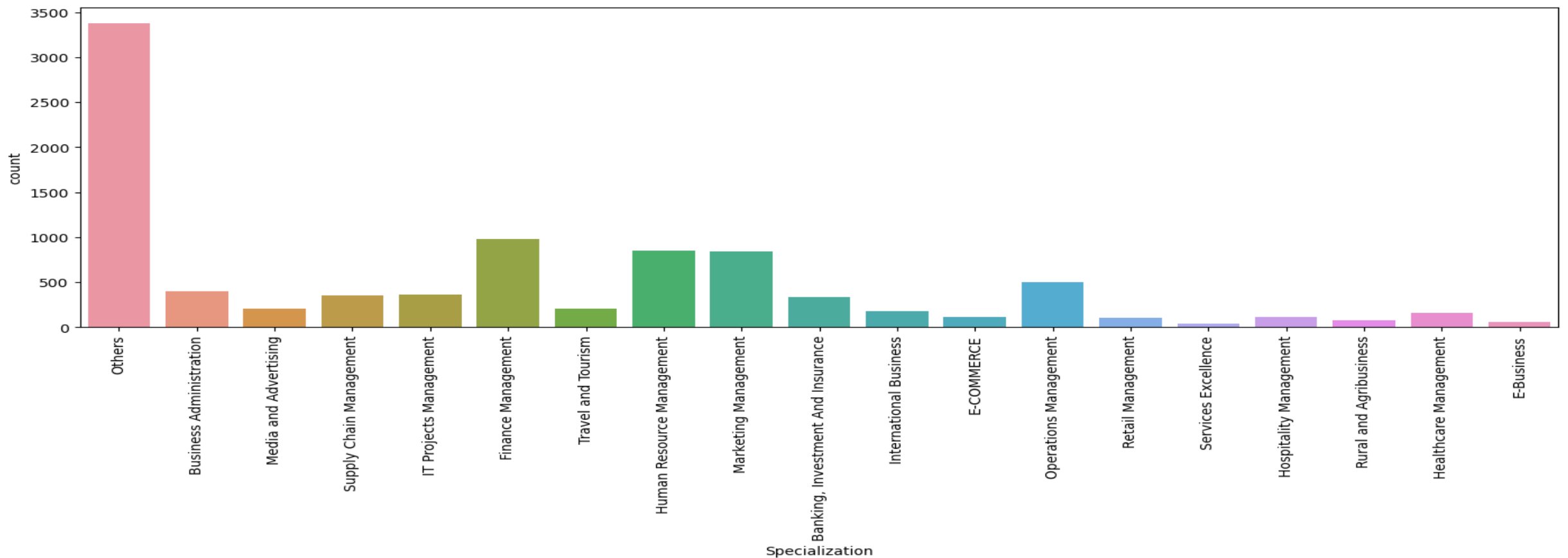
Other columns to Impute

- Tags
- What matters most to you in choosing a course
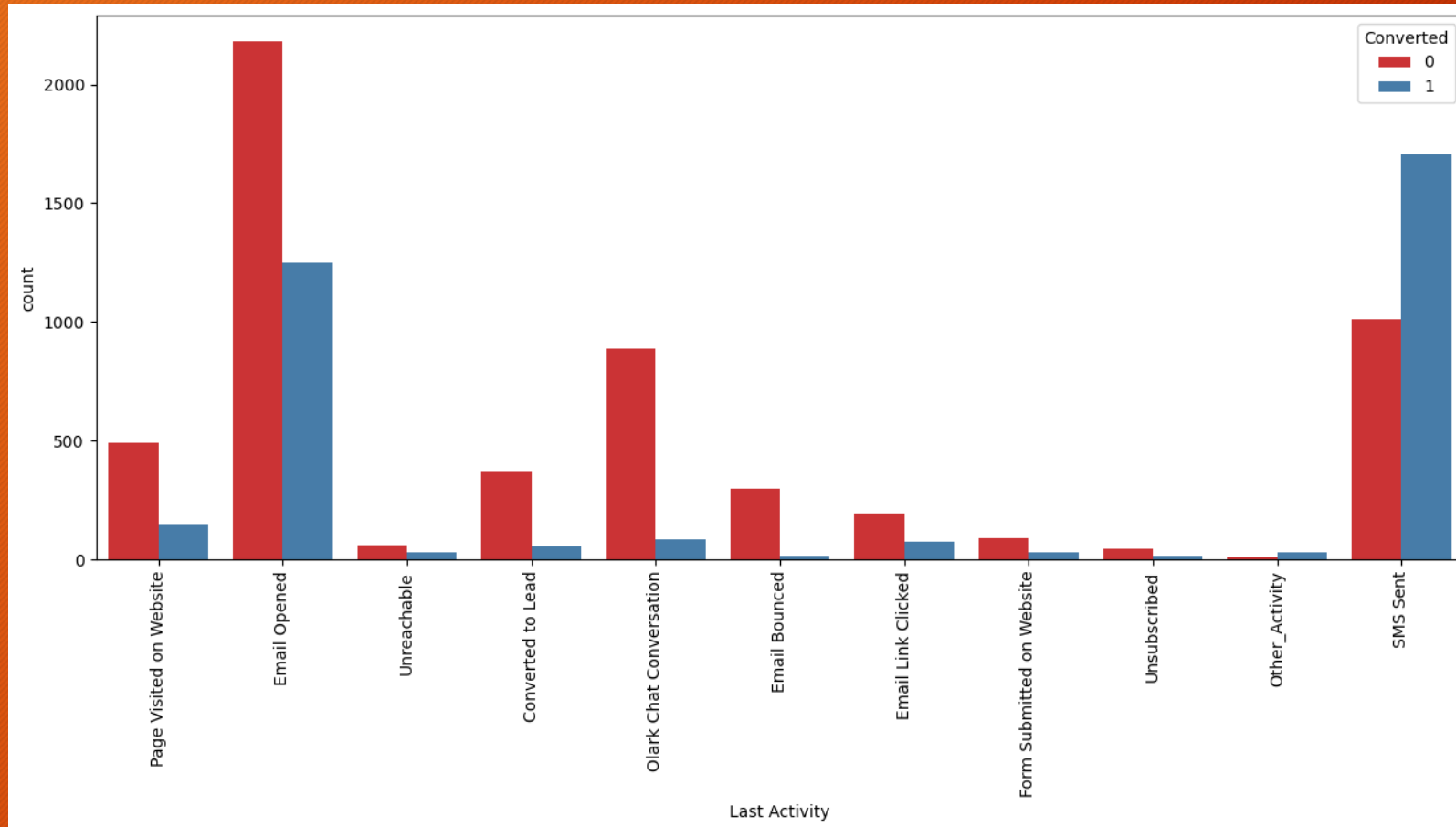- What is your current occupation
- Country
- City

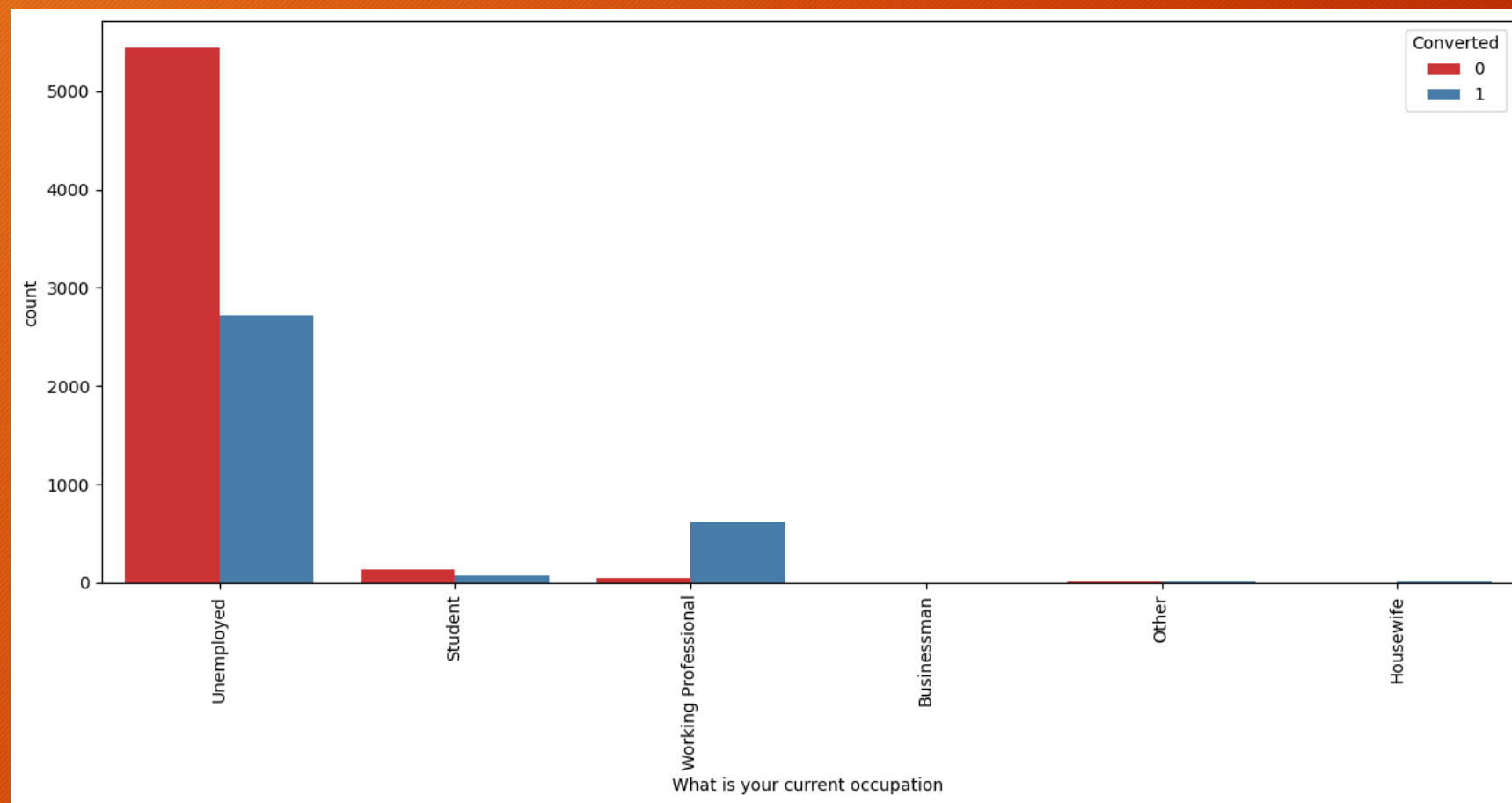# Distribution of Converted with Lead Origin
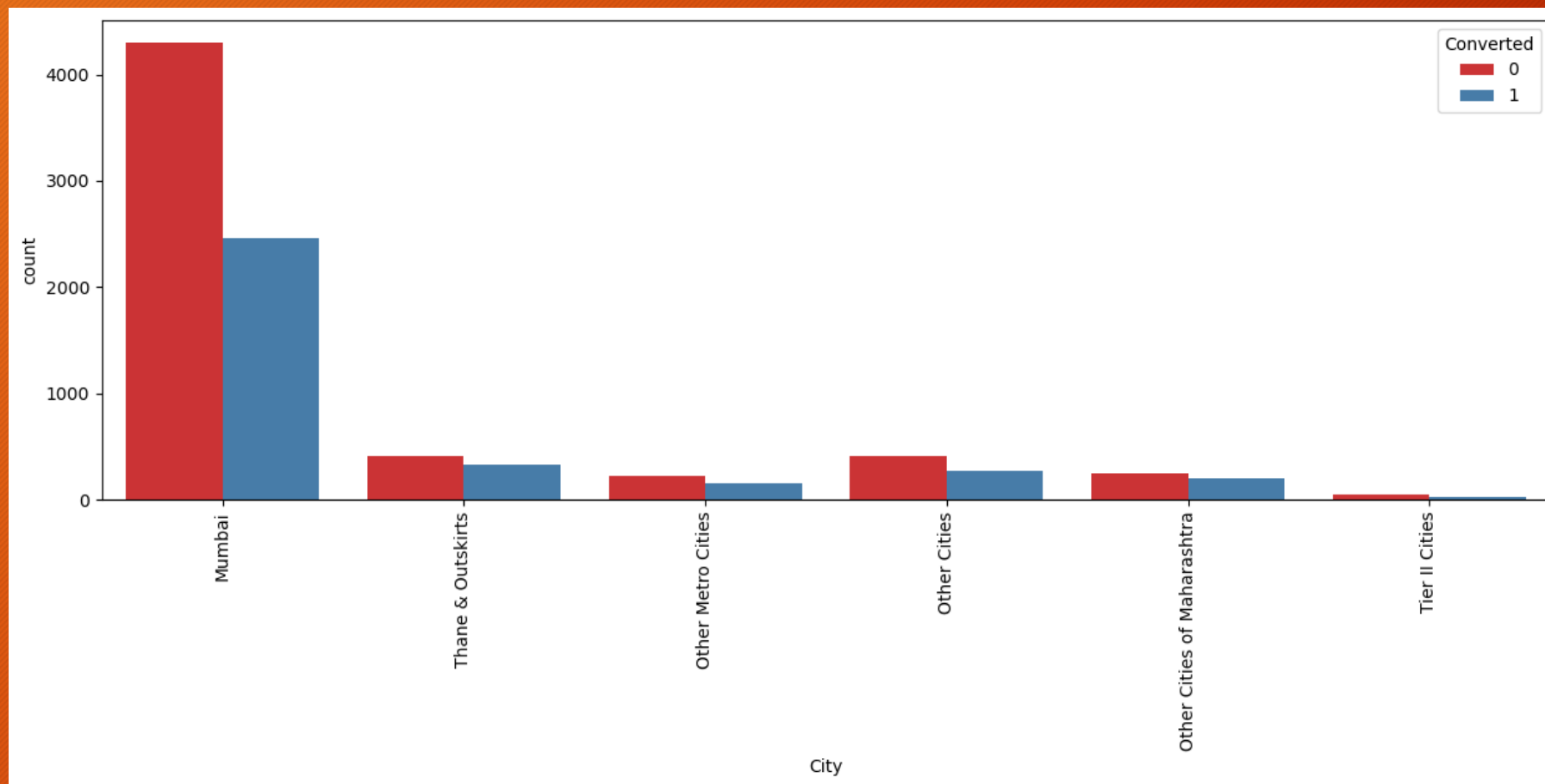
Distribution of Converted with Lead source
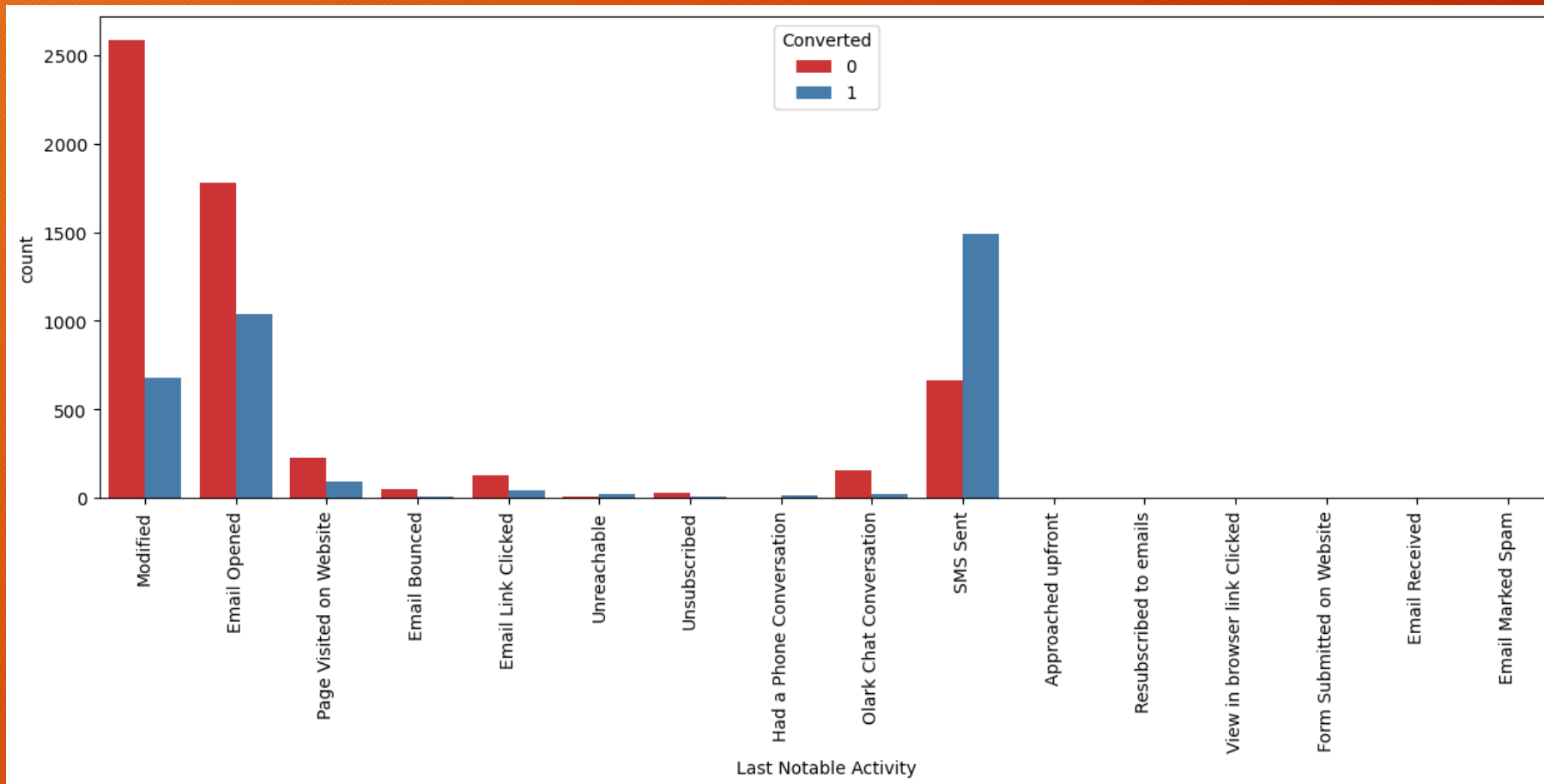
Distribution of converted with Last Activity

Distribution of Converted with Occupation
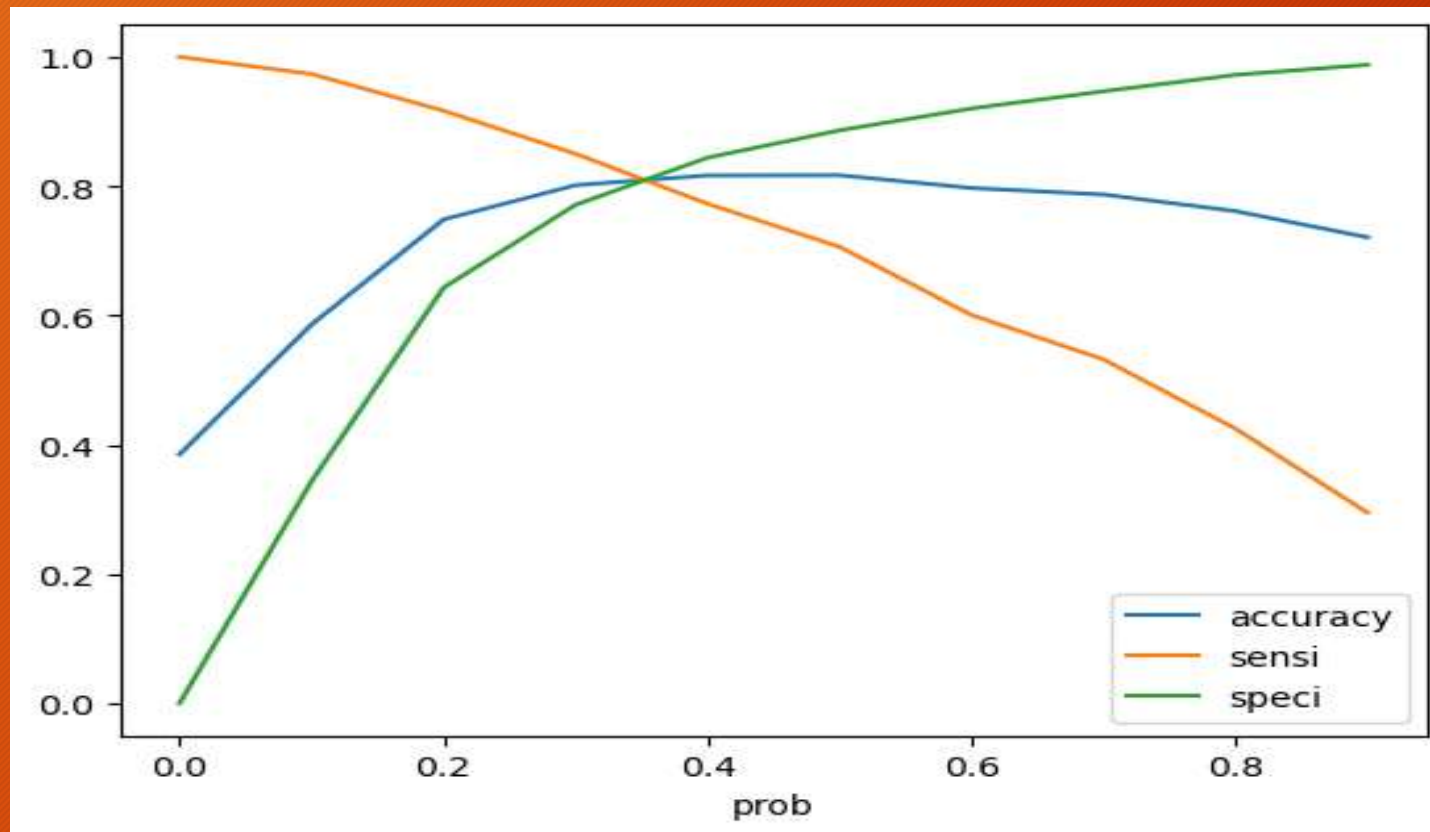
Distribution of Converted with City

Distribution of Converted with Last Notable Activity
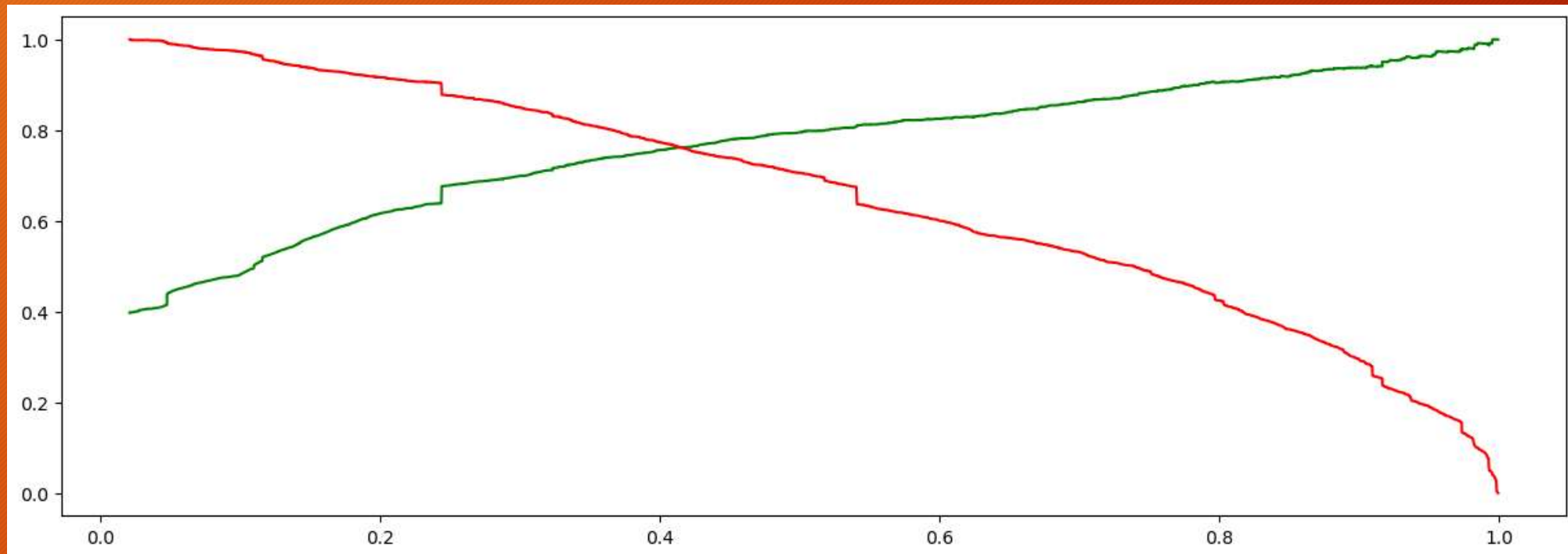
Model Evaluation

Accuracy, Sensitivity and Specificity for various probabilities

ROC Curve

0.42 is the tradeoff and so any lead with conversion probability higher than 42% is a hot lead

Observations

Important Features

- Lead Source_Welingak website
- Lead Source_Reference
- What is your current occupation_Working Professional
- Last Activity_Other Activity
- Last Activity_SMS Sent
- Total Time spend on website
- Lead Source_Olark Chat
- Specialization_Others
- Do Not Email

Train Data

Accuracy: 81%
Sensitivity: 81%
Specificity: 80%

Test Data

Accuracy: 80%
Sensitivity: 80%
Specificity: 80%

Conclusion

- The observed conversion rate for API and Landing Page submissions falls within the range of 30-35%, which is close to the average. However, the conversion rate is notably lower for Lead Add Form and Lead Import. Consequently, it's evident that there is a need to prioritize leads generated from API and Landing Page submissions for further attention and improvement.

- The majority of leads originate from sources like Google and direct traffic. The highest conversion ratios are associated with leads referred from other sources and the Welingak website.

- There is a positive correlation between the time spent on the website and the likelihood of lead conversion. Leads who spend more time on the platform are more likely to convert.

- Among the last recorded activities of leads, the most common one is "email opened," followed by "SMS Sent." Additionally, a significant portion of leads is currently unemployed, with the highest conversion rates observed among working professionals.