Lead Scoring Case Study

Ganesh Balaji, Siva Sankar, Anwita Ghosh

Problem Statement

- An education company named X Education sells online courses to industry professionals. The company markets its courses on several platforms and websites. While surfing through the website the customers are prompted to enter their credentials which are classified as leads. Once the company acquires enough leads it deploys its Sales team to acquire them. Over the past few years, X Education has observed a decline in their lead conversion.

- X Education has appointed us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance

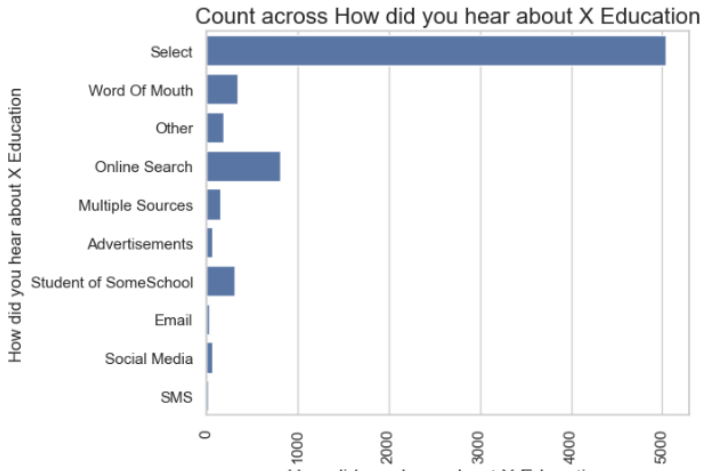
Solution Methodology

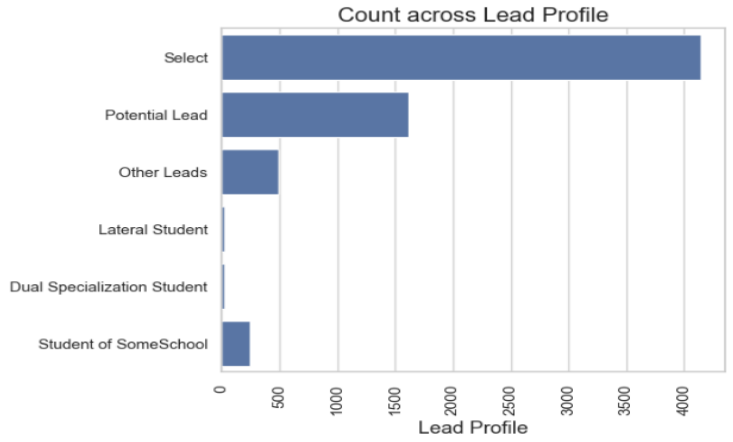
Workflow Overview

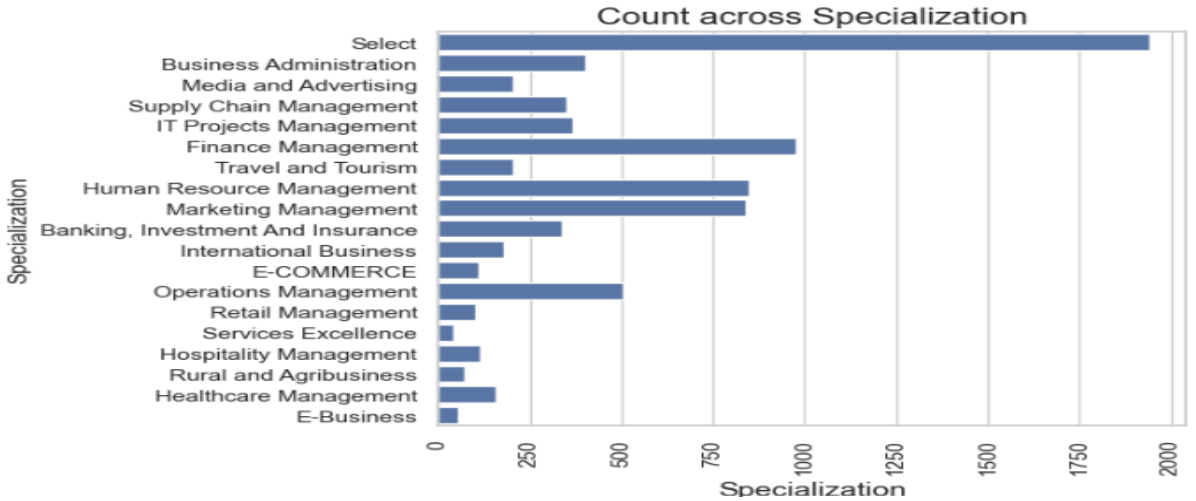
- Load the dataset and explore the structure of the DataFrame
- Prepare the data for analysis
- Perform Exploratory Data Analysis (EDA)
- Create dummy variables for categorical features
- Split the dataset into training and testing sets
- Apply feature scaling to standardize the data
- Analyze feature correlations
- Build the model using techniques such as RFE, R- squared, VIF, and p-values
- Evaluate the performance of the model
- Generate predictions on the test dataset

Exploratory Data Analysis

We've seen some columns with 'Select' as a category



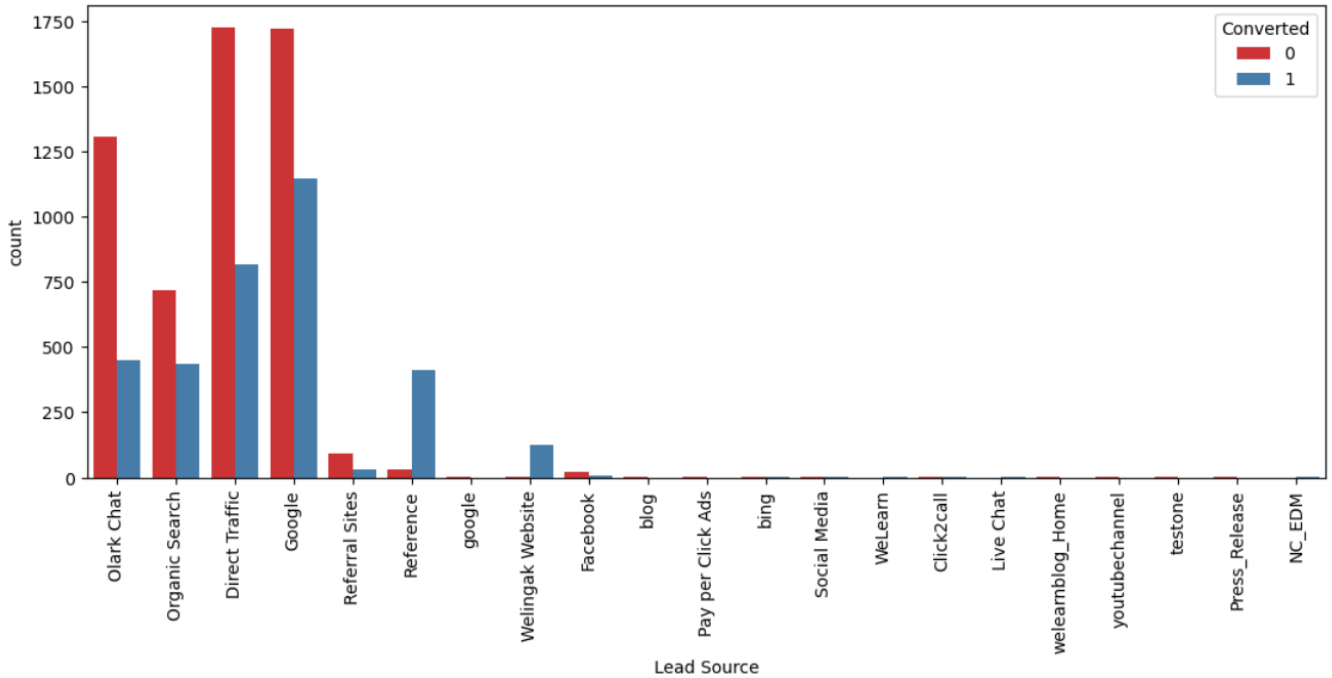


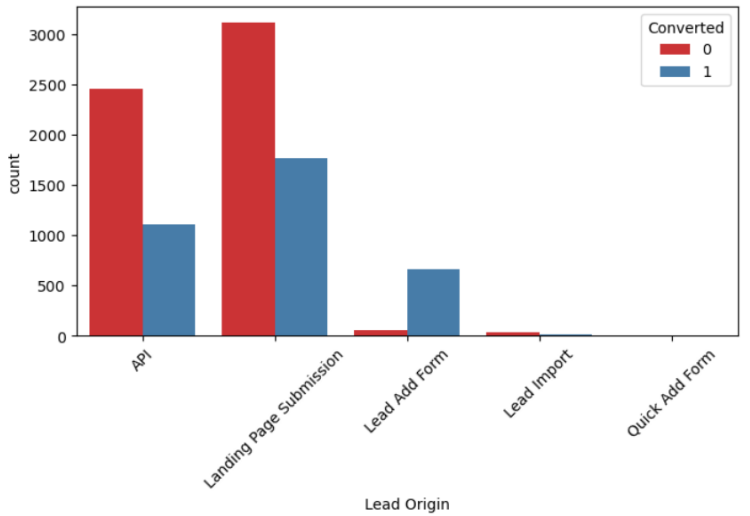


Exploratory Data Analysis

- Lead Source has high volume of conversion from Direct Traffic and Google

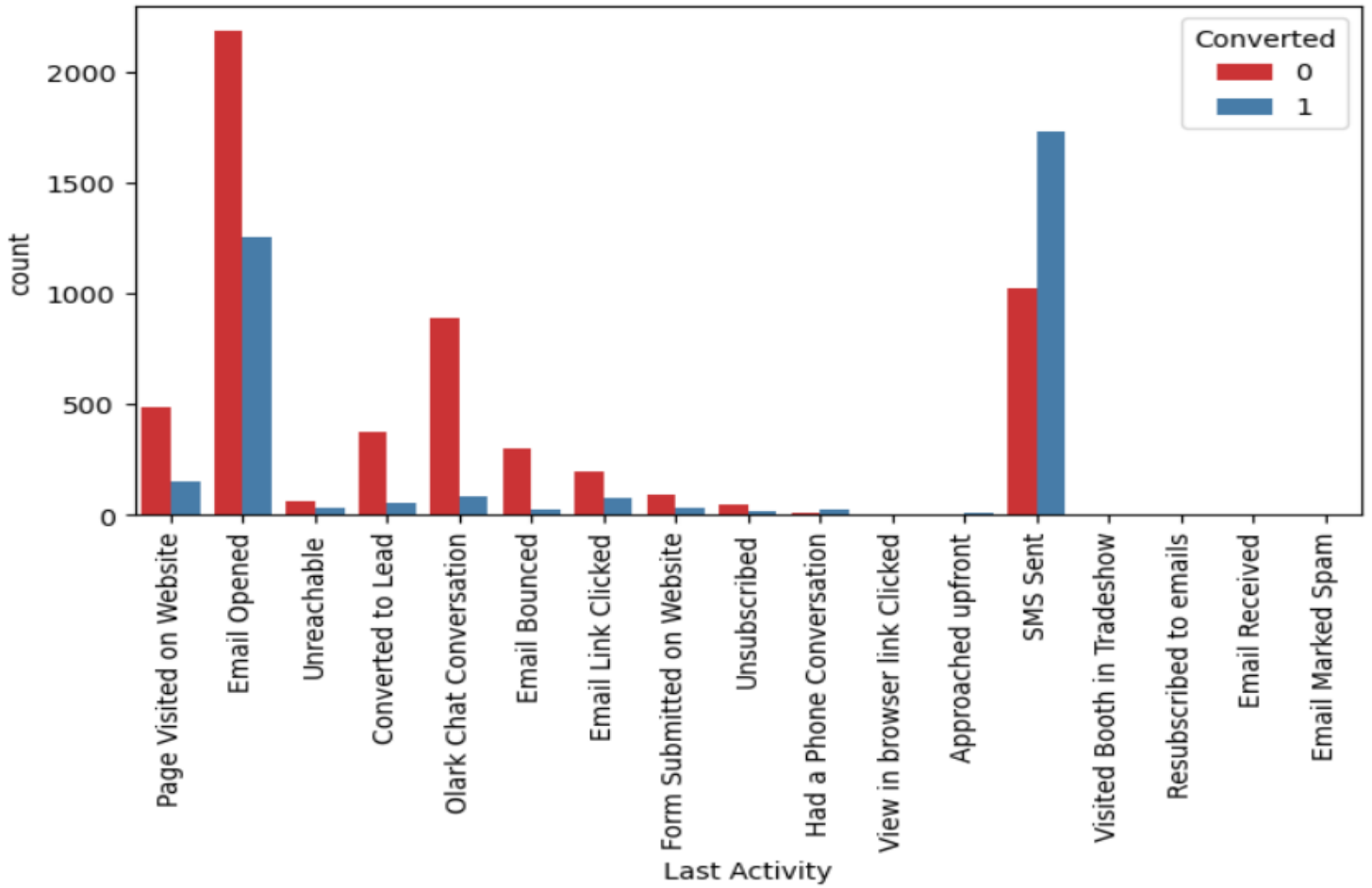
- Lead Source has most number of leads originating from 'Landing Page Submission'





Exploratory Data Analysis

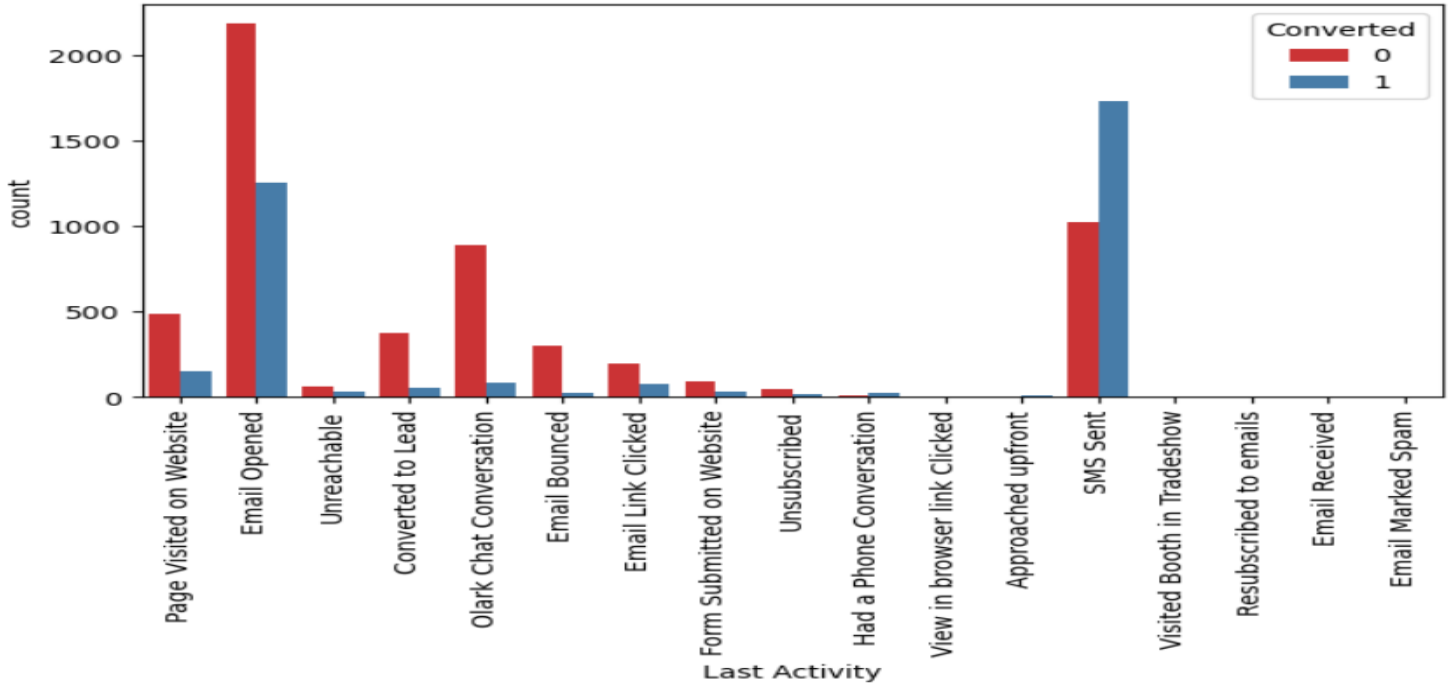
- Leads who have 'Email Opened' and 'SMS Sent' have had highest conversion rate in terms of volume

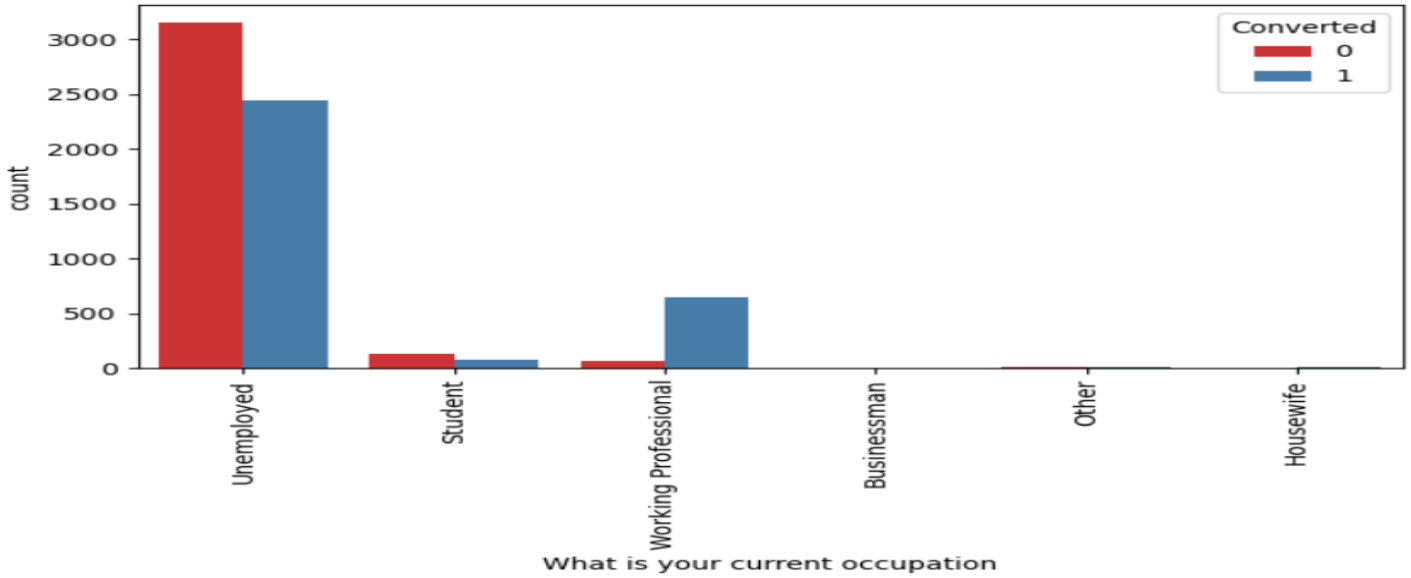


Exploratory Data Analysis

- Leads who have ‘Email Opened’ and ‘SMS Sent’ have had highest conversion rate in terms of volume

- Leads who are ‘Unemployed’ have a high conversion rate possible to increase their marketability as prospective emplooyees





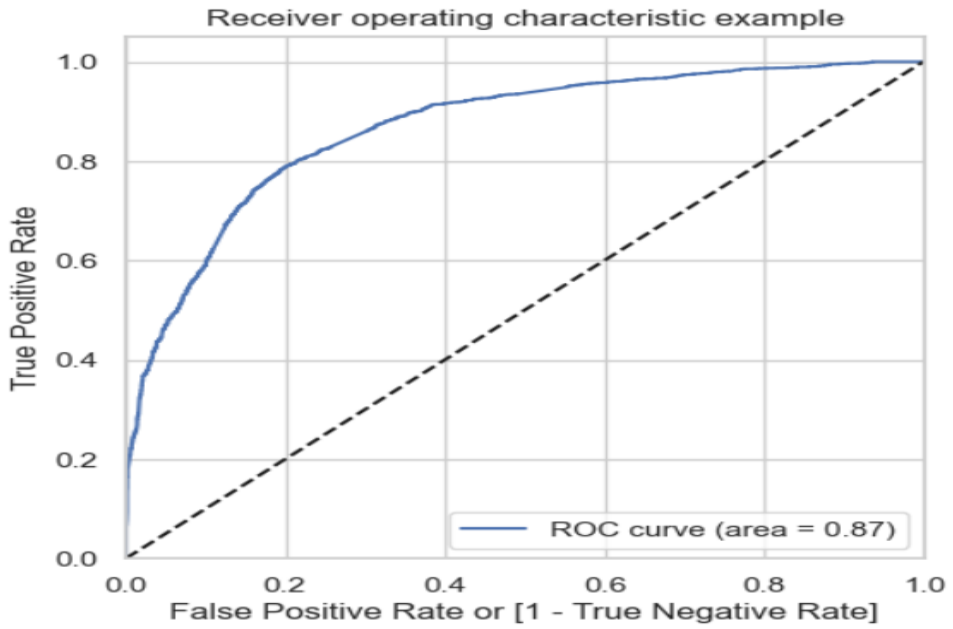
Exploratory Data Analysis

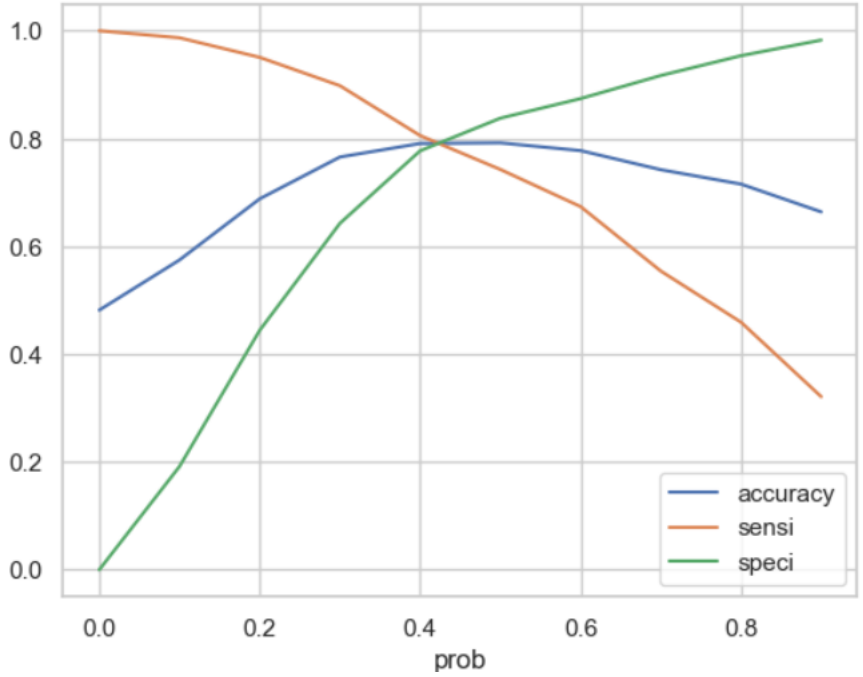
- We wanted to check for any relationships between variables using a heatmap and observed that 'Total Visits ' and 'Page Views per Visit' have a relatively high correlation between them. However, for analysis purpose we'll use either of the two variables

Model Evaluation

- The area under curve of the ROC is at 0.87 which is above 80% and model fit has been decent so far.

- 0.42 is the trade-off between Precision and Recall which can be alternatively stated as: consider any Prospect Lead Conversion Probability higher than 42% to be a hit Lead





Observations

Train data: Accuracy – ~80% (0.7983), Sensitivity – 79%, Specificity – 80%

Test data: Accuracy – 80% , Sensitivity – 79%, Specificity – 80%

List of Features selected

- TotalVisits
- Total Time Spent on Website
- Lead Origin_Lead Add Form
- Lead Source_Olark Chat
- Lead Source_Welingak Website
- Do Not Email_Yes
- Last Activity_Had a Phone Conversation
- Last Activity_SMS Sent
- What is your current occupation_Housewife
- What is your current occupation_Student
- What is your current occupation_Unemployed
- What is your current occupation_Working Professional
- Last Notable Activity_Had a Phone Conversation
- Last Notable Activity_Modified
- Last Notable Activity_Unreachable

Conclusion

- The conversion rate for leads generated via API and landing page submissions is around 30–35%, which is considered average. However, the rate is significantly lower for leads obtained through Lead Add forms and Lead imports. This indicates a need to prioritize and enhance lead generation efforts through APIs and landing pages.

- The majority of leads come from Google or direct traffic sources. However, the highest conversion rates are observed in leads referred by others or originating from the Welingak website.

- Users who spend more time browsing the website show a higher likelihood of conversion.

- The most frequent final activity recorded is "Email Opened," while the highest activity rate is observed for "SMS Sent." While most leads are unemployed, the highest conversion rates are seen among working professionals.