Lead Scoring Case Study using logistic regression

By kajal Mishra

Contents

Problem Statement

- X Education is an online learning company that offers courses to working professionals.
- Each day, many interested professionals visit the website and fill out a form, after which they are marked as leads.
- Once leads are captured, the sales team contacts them via calls and emails. Some leads convert, but most do not.
- 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.
- To improve this, the company aims to identify **Hot Leads**—those most likely to convert—so the sales team can focus their efforts more efficiently and boost conversions.

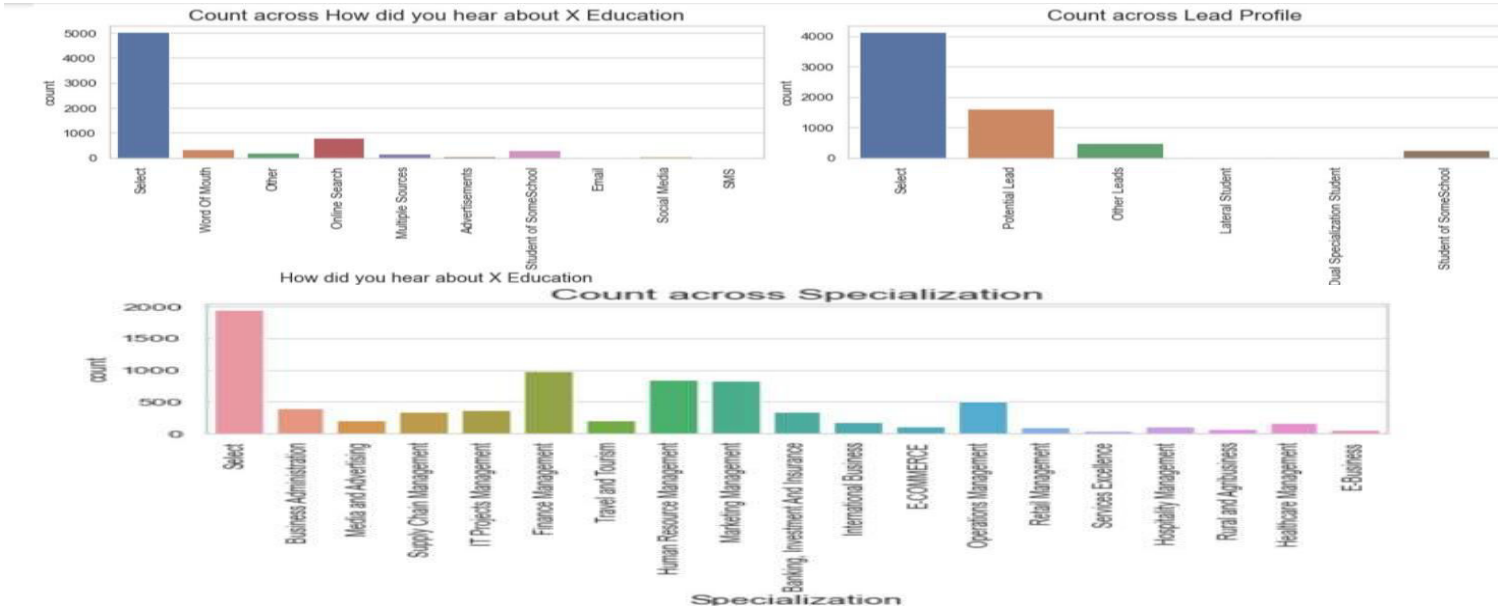
Business Objective

- Lead X has requested a model that assigns each lead a score between 0–100, helping them identify Hot Leads and improve their conversion rate.
- The CEO’s goal is to achieve a lead conversion rate of 80%.
- The model should also be designed to handle future scenarios, such as actions during peak periods, optimal use of manpower, and strategies after meeting targets.

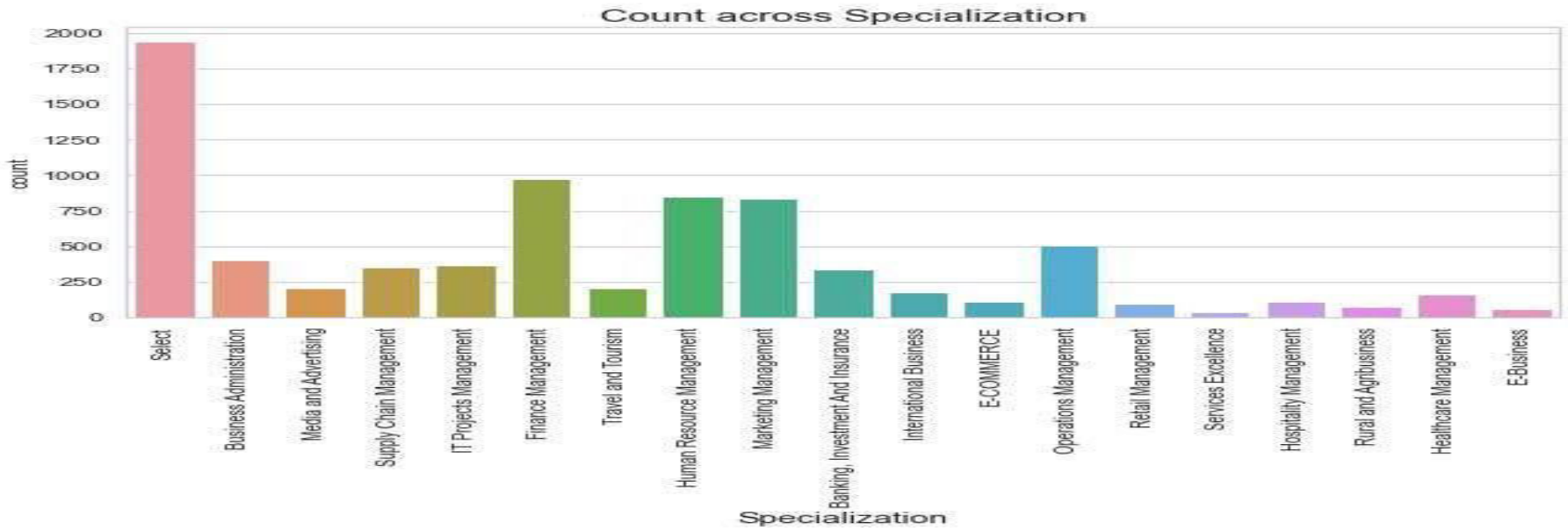
Approach to the Problem

- Data Import and Initial Inspection
- Data Cleaning and Preparation
- Exploratory Data Analysis (EDA)
- Creation of Dummy Variables
- Splitting Data into Training and Testing Sets
- Applying Feature Scaling
- Analyzing Correlations
- Model Building using RFE, R-squared, VIF, and p-values
- Evaluating Model Performance
- Making Predictions on the Test Set

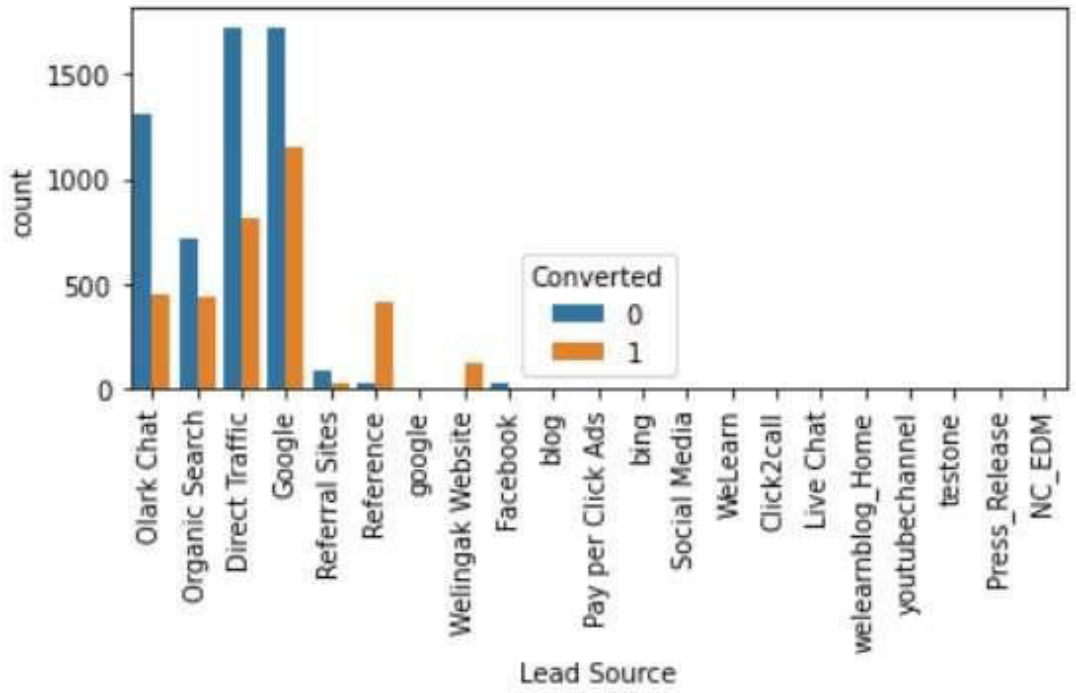
Exploratory Data Analysis EDA – Data Cleaning

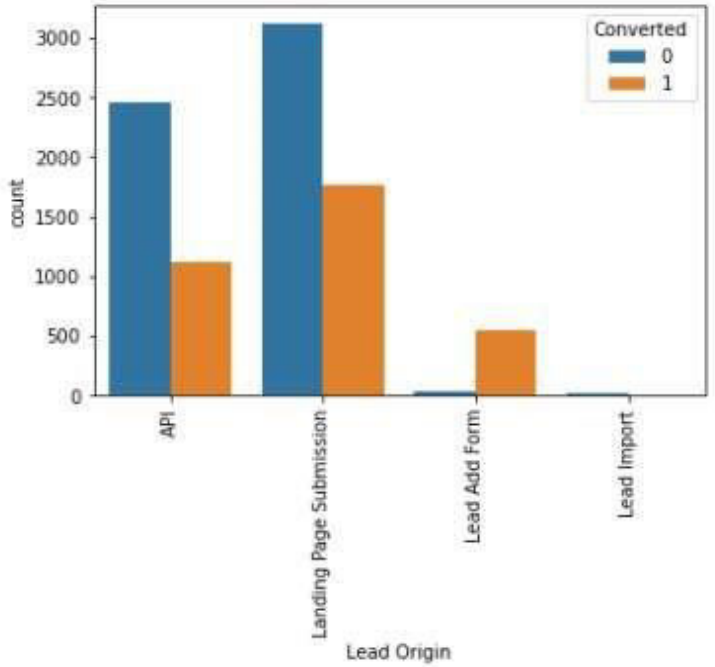


Specialization

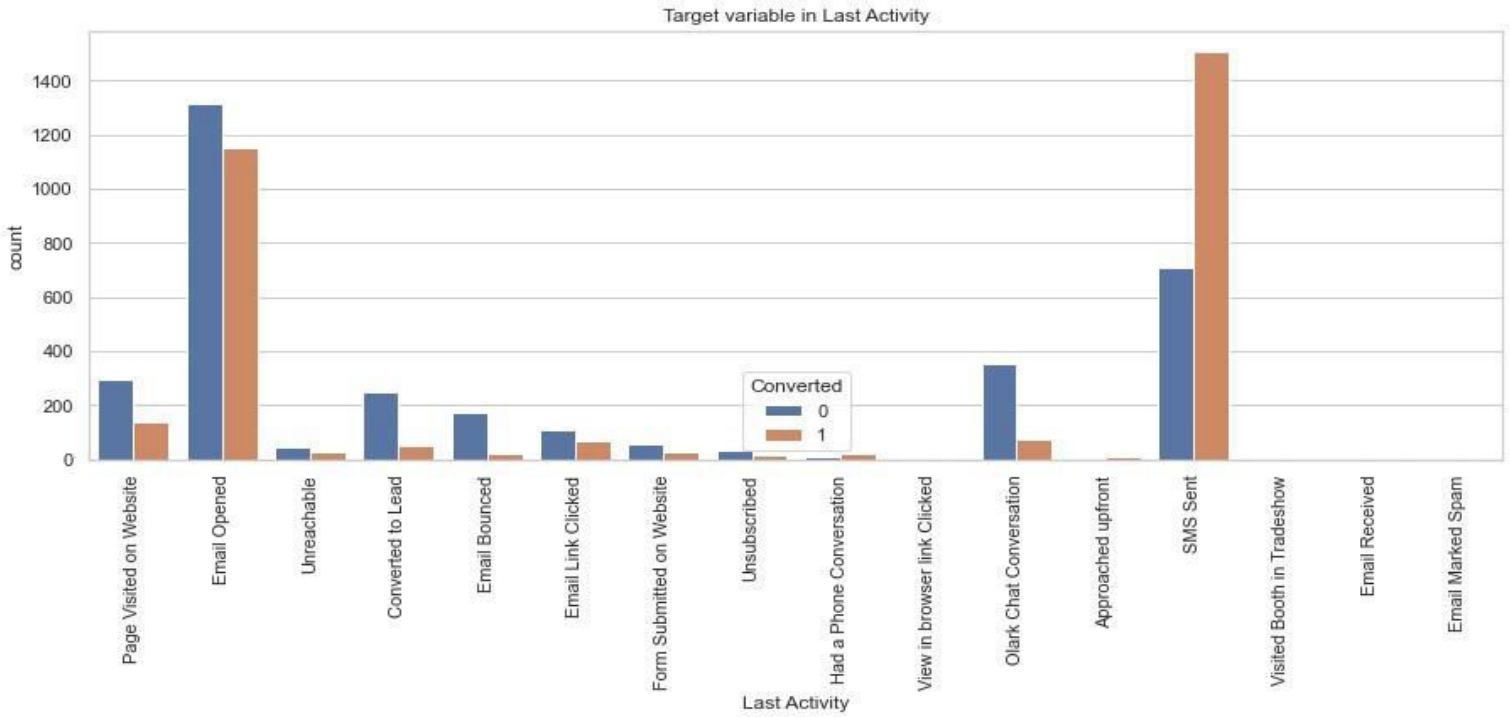


Lead Source and Lead origin

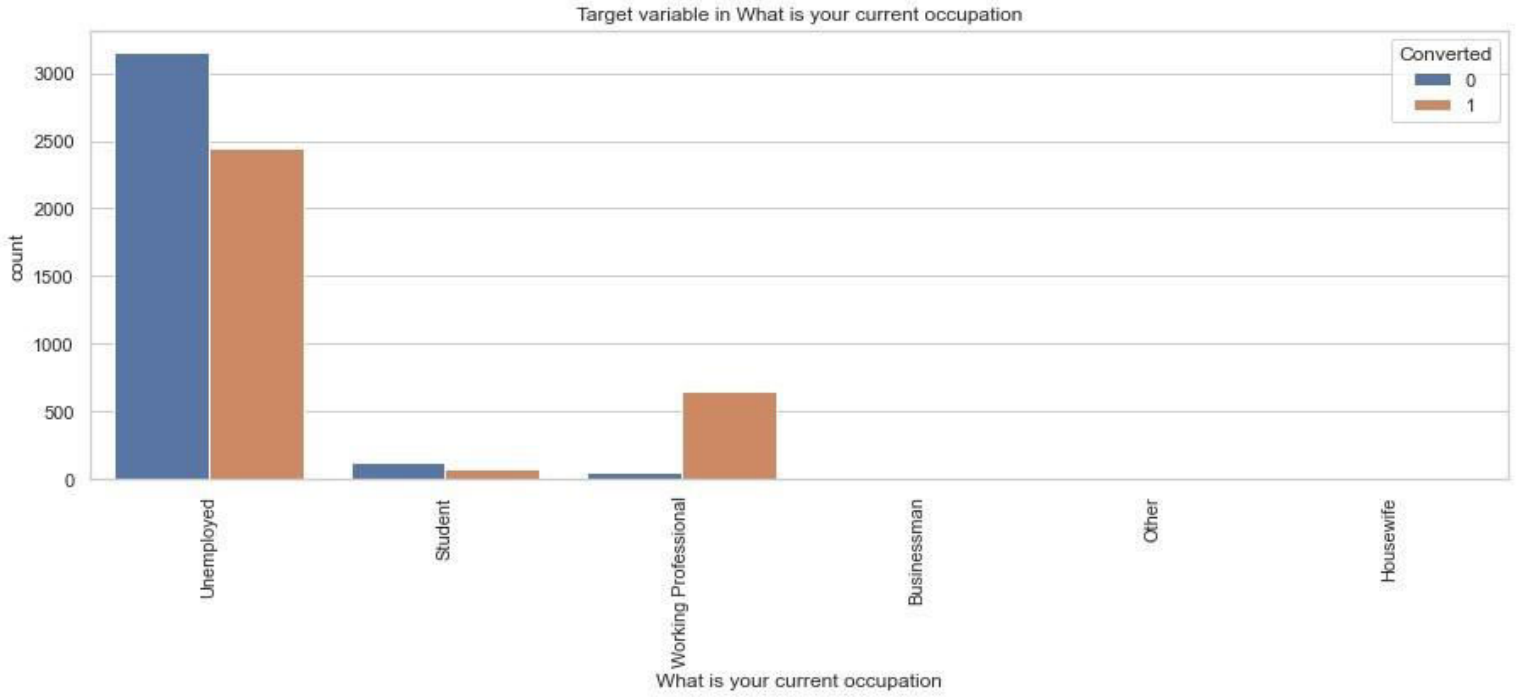




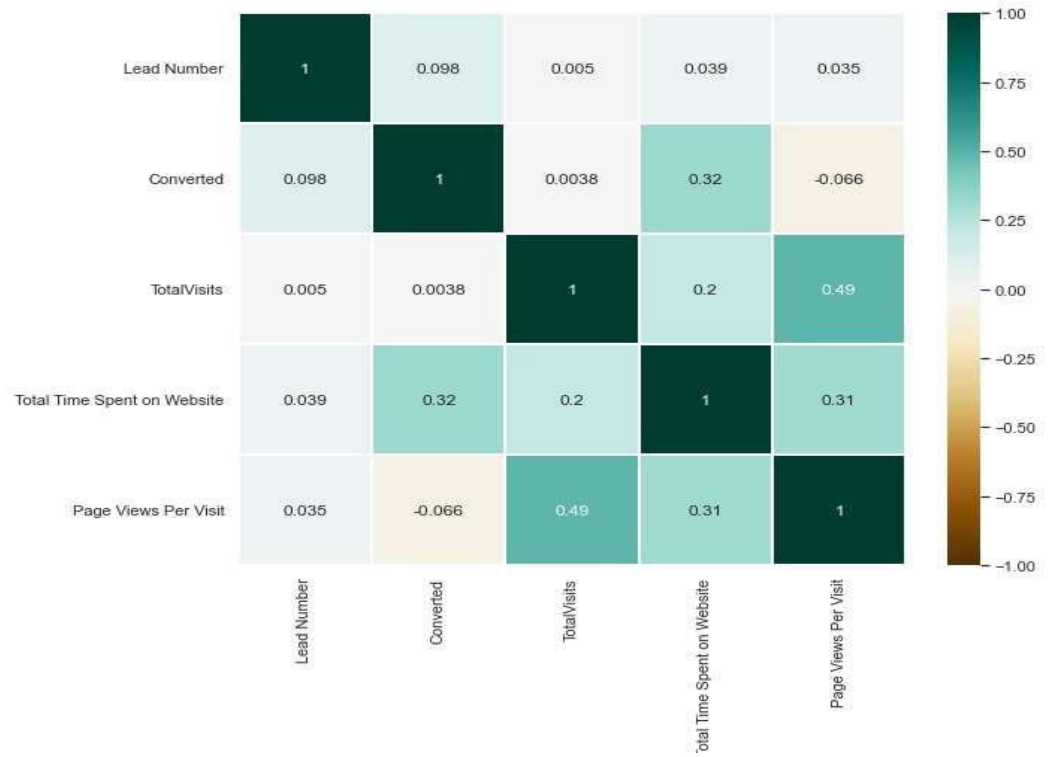
Last lead Activity



What is your current occupation?

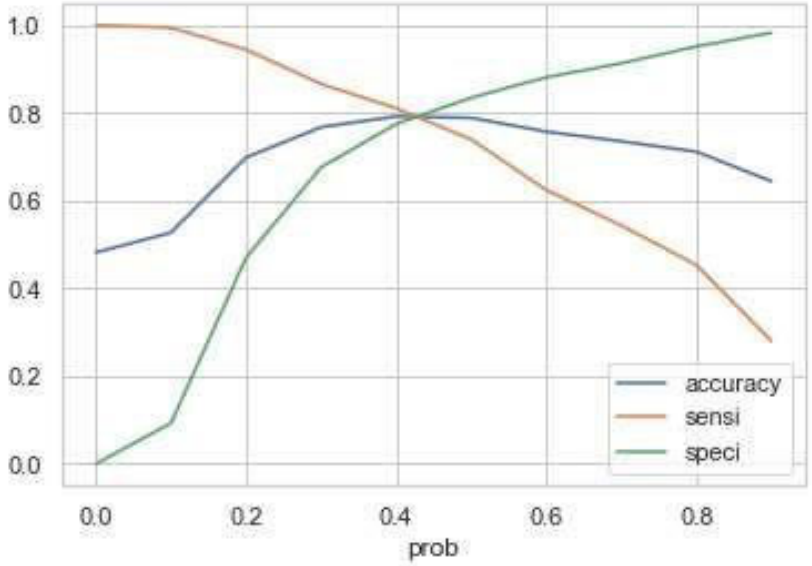


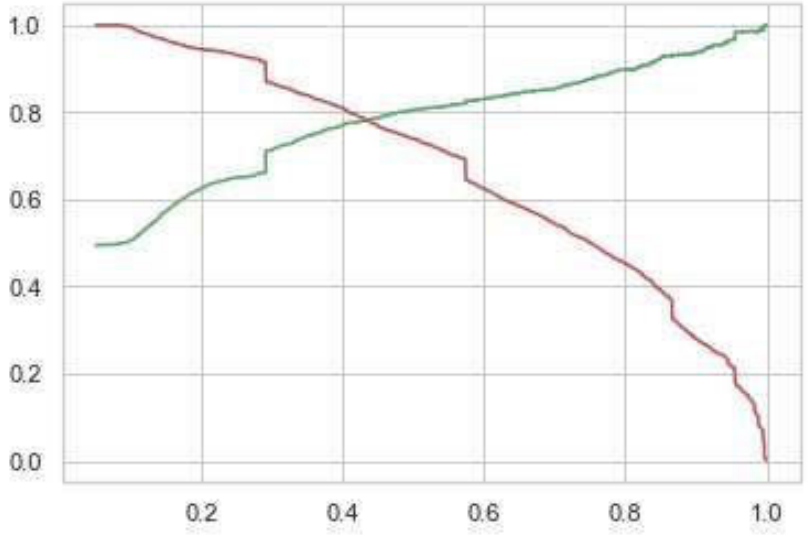
Correlation Analysis



Model Evaluation

A threshold of **0.42** represents the balance between **Precision and Recall**. Therefore, any prospect lead with a **conversion probability greater than 42%** can be confidently classified as a **Hot Lead**.





Key Findings

Final Features list:

- Lead Source_Olark Chat
- Specialization_Others
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- What is your current occupation_Working Professionals
- Do Not Email

Train Data:
Accuracy : 80%
Sensitivity : 77%
Specificity : 80%

Test Data:
Accuracy : 80%
Sensitivity : 77%
Specificity : 80%

Summary

- The conversion rate for leads from API and Landing Page Submission is around 30-35%, which is close to the average. However, the conversion rate is significantly lower for Lead Add Form and Lead Import. This indicates the need to focus more on leads originating from API and Landing Page Submission.
- A majority of leads come from Google/Direct Traffic, but the highest conversion rates are seen from Referrals and the Welingak Website.
- Leads that spend more time on the website are more likely to convert.
- The most common last activity is email opened, while the highest conversion rate is associated with SMS sent. The majority of leads are unemployed, with the highest conversion rate coming from working professionals.