LEAD SCORING USING LOGISTIC REGRESSION

Meenakshi Sharma

19 June 2024

BACKGROUND

- An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.

- X Education wants to select most promising leads that can be converted to paying customers.

- although X Education gets a lot of leads, its lead conversion rate is very poor. Leads come through numerous modes like email, advertisements on websites, google searches etc.

- The typical lead conversion rate at X education is around 30%. To make this process more efficient, the company wishes to identify the most potential leads. the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Objective:

- The company requires a model to be built for selecting most promising leads.

- Lead score to be given to each 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 model to be built in lead conversion rate around 80% and more.

STEPS REQUIRED FOR ANALYSIS

- Data Inspection and Data Understanding
- Data Preparation
- Exploratory Data Analysis
- Dummy Variable Creation
- Train-Test Split
- Feature Scaling
- Logistic Regression Model Building
- Model Evaluation by different measures and metrics
- Making Predictions on Test Set

SPECIALIZATION

- Leads from HR, Finance and Marketing Management specialization have high probability to convert.
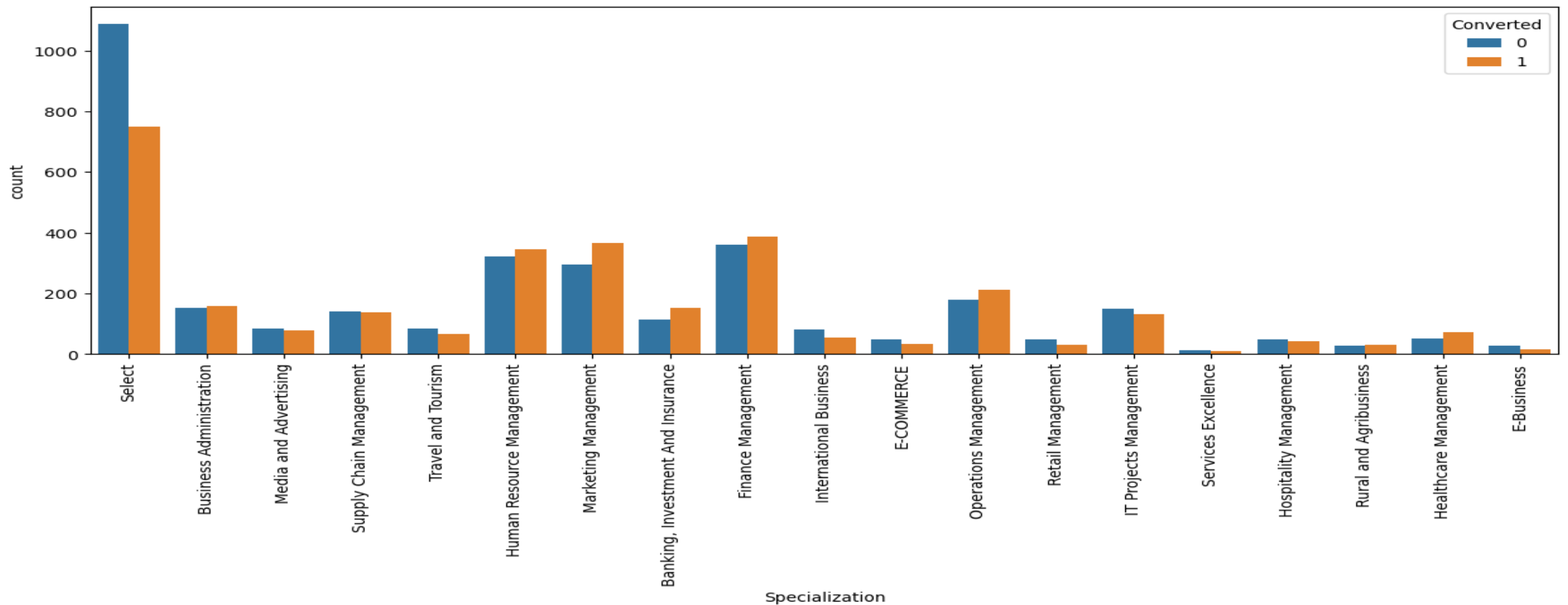
WHAT IS YOUR CURRENT OCCUPATION

- Though the lead count is less, working professionals have a highest conversion rate.
- Lead count for unemployed segment is the highest and conversion rate is also good.
- While the lead count under Not Specified category is very good, conversion rate is comparatively low against the previous two.
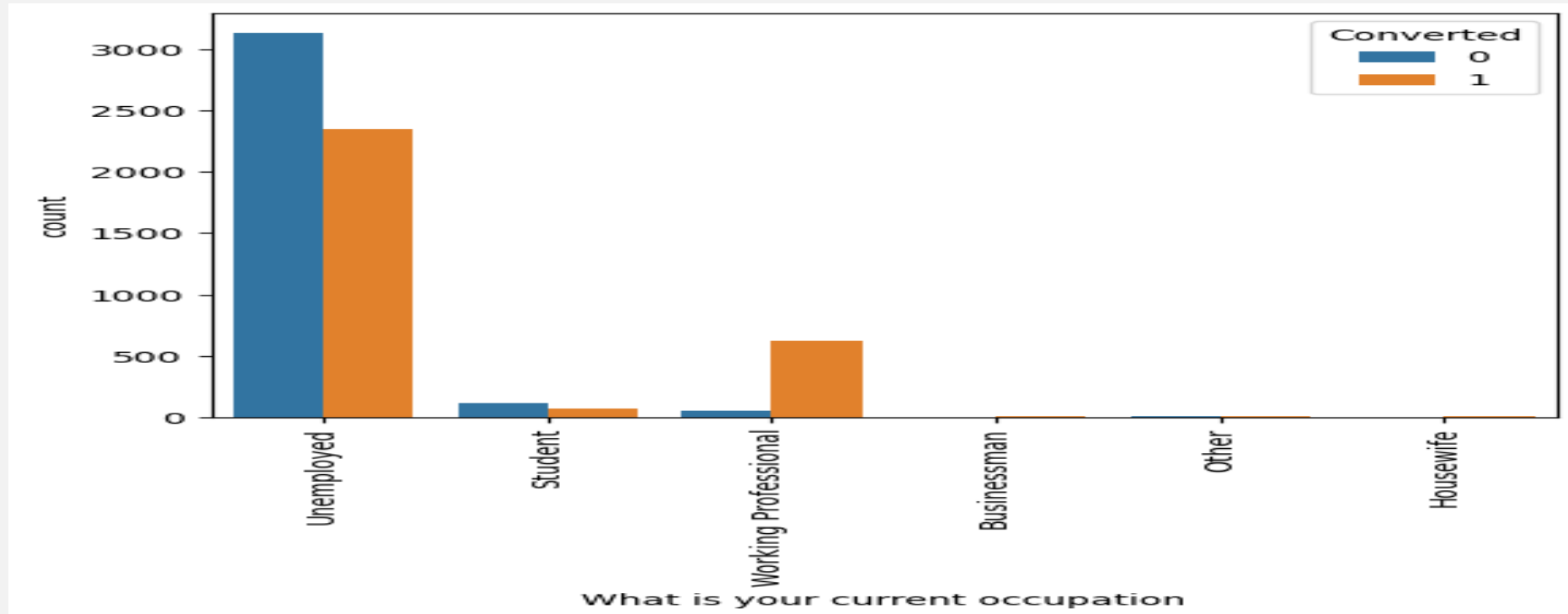
LEAD SOURCE

- Maximum leads are generated from Google and Direct Traffic and conversion rate is also good.
- The conversion rate is very high for leads generated from Welingak Website and Reference.
- The company should work on increasing their leads from Welingak Website and Reference.
- While the lead count from Organic Search is comparalively low, conversion rate is very good.
- The company should work on introducing better strategies for lead conversion from Olark Chat, Organic Search, Direct Traffic and Google.
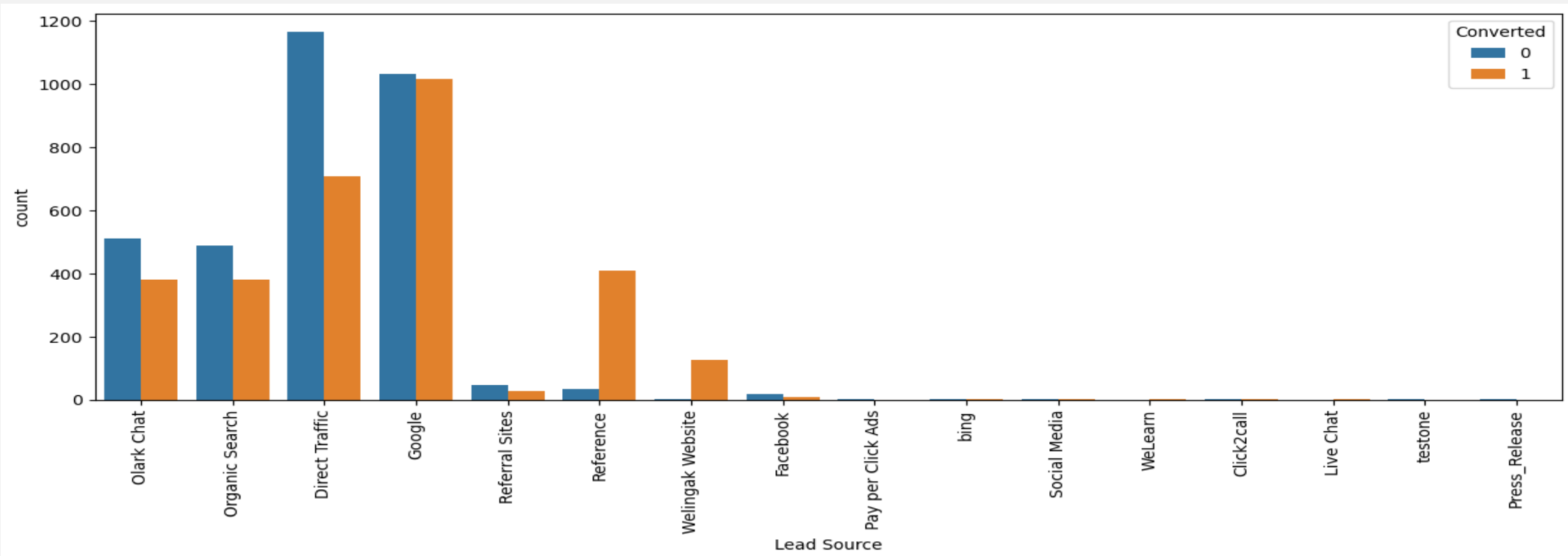
LEAD ORIGIN

- Landing Page Submission and API have the highest number of leads as well as conversion.
- Lead Add Form has a very high conversion rate but count of leads generated is not very high.
- Lead Import has negligible leads.
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
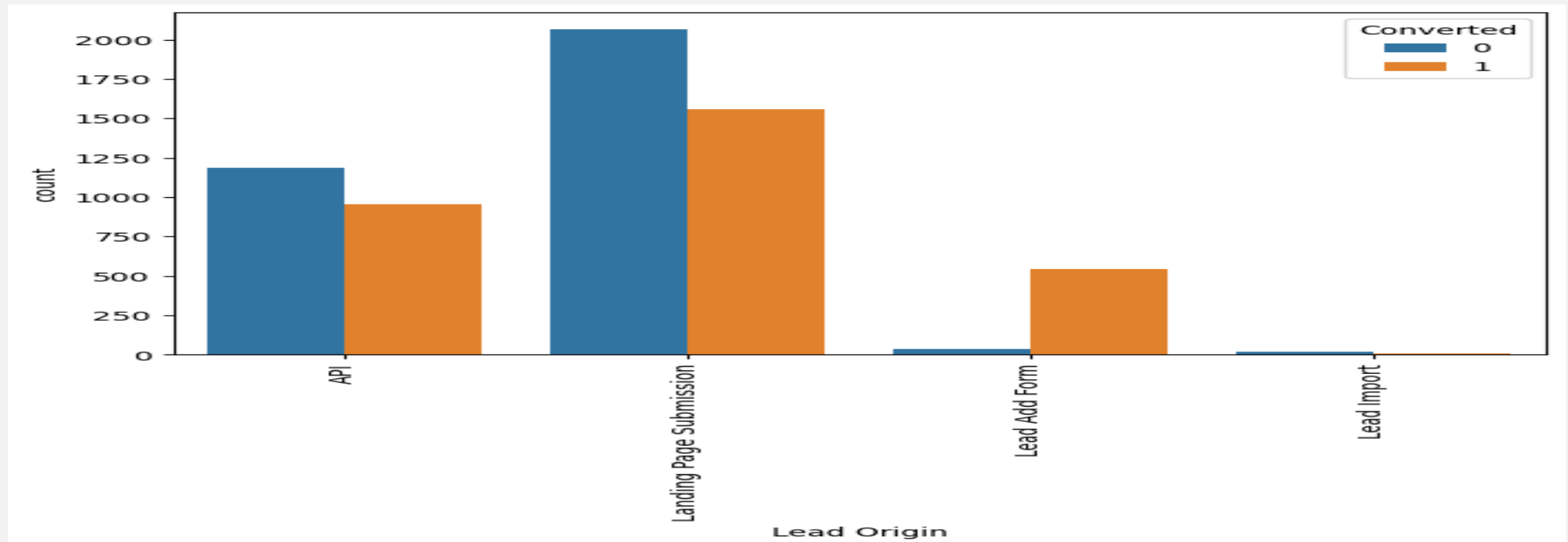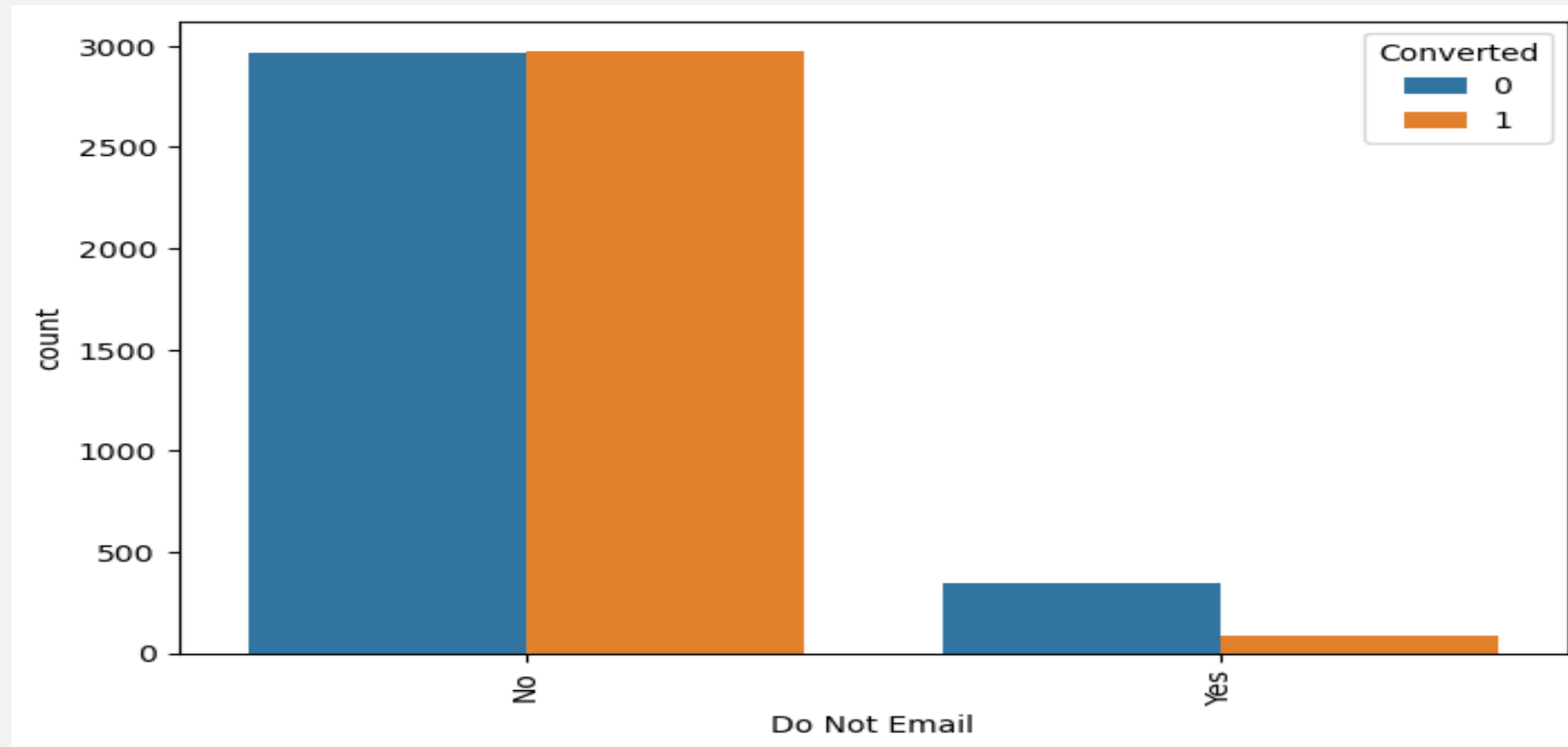
LAST ACTIVITY

- Most common last activity is email opened. SMS sent has the highest conversion rate.
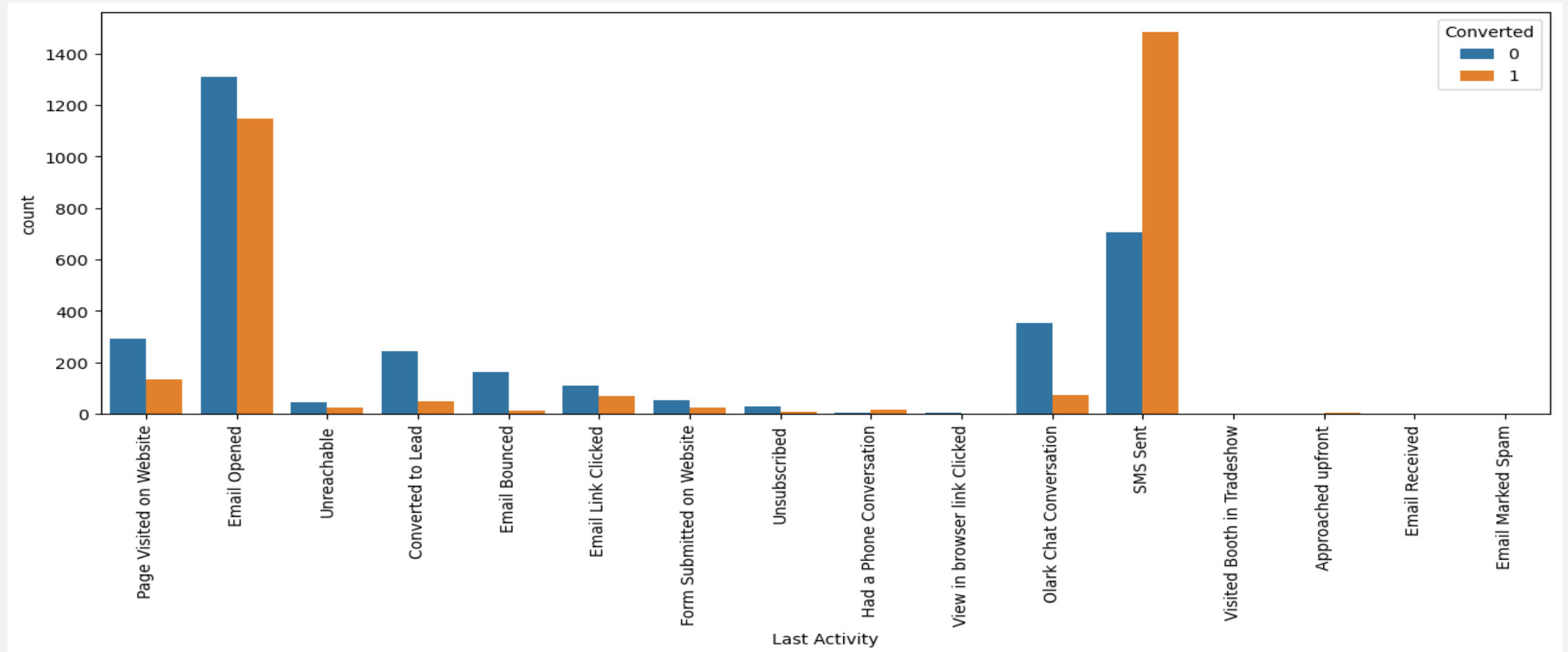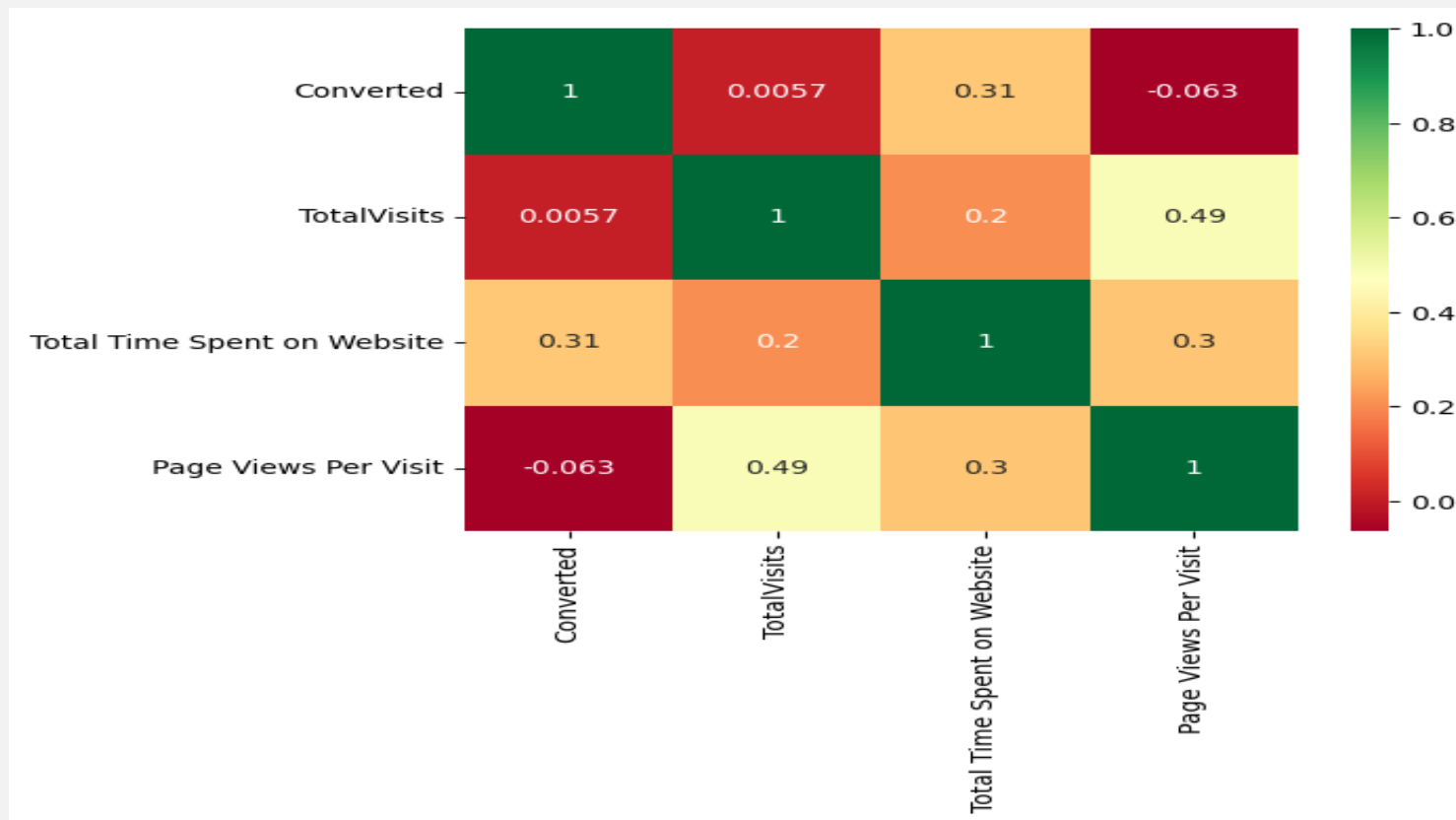
DO NOT EMAIL

- Contacting individuals via email plays an important role in conversion.
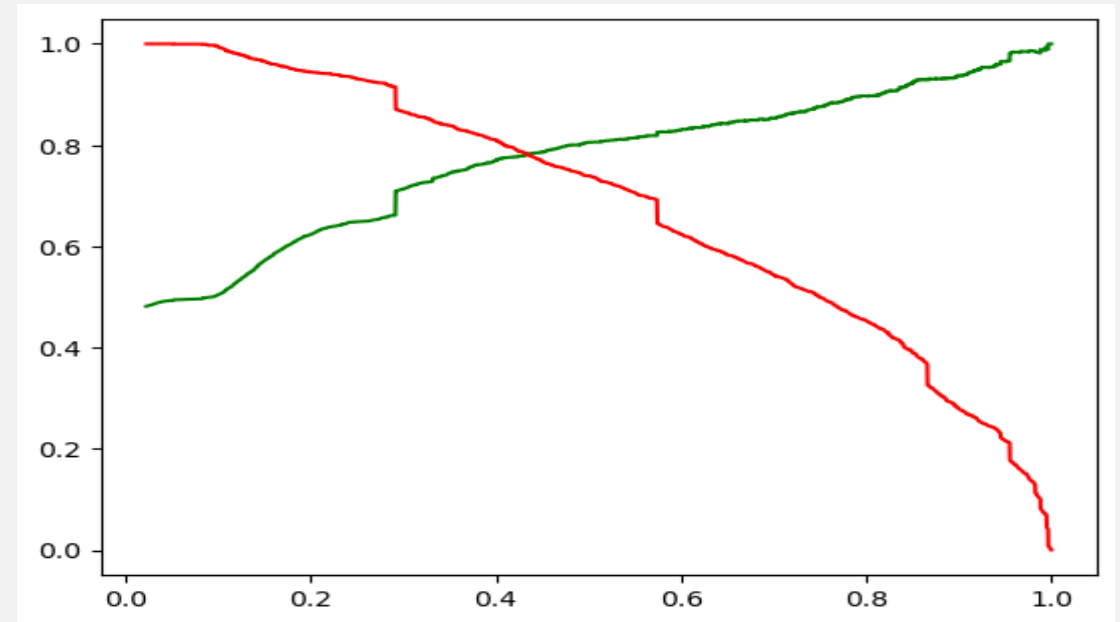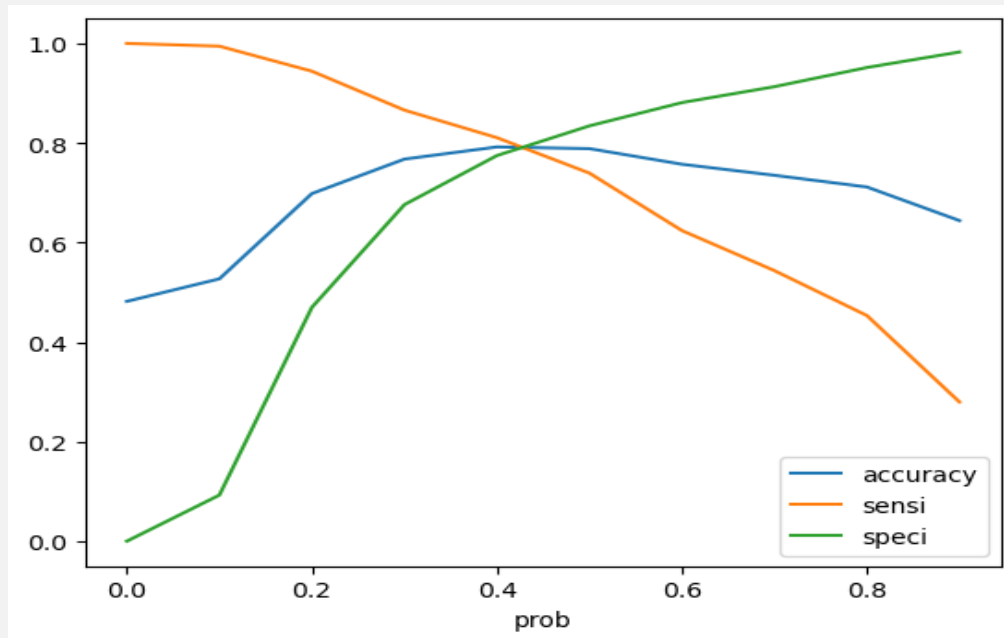
CORRELATION

- There is no correlation between the variables.

MODEL EVALUATION

- 0.42 is the tradeoff between precision and recall.
- Thus we can safely choose to consider any prospect lead with conversion probability higher than 42% to be a hot lead.

OBSERVATIONS

Final Features list –

- 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_Student
- What is your current occupation_Unemployed
- Last Notable Activity_Unreachable

Train Data
Accuracy : 79%
Sensitivity : 79%
Specificity : 79%

Test Data
Accuracy : 79%
Sensitivity : 78%
Specificity : 79%

CONCLUSION

- We see that the conversion rate is 30-35% for API and landing page submission but very low for lead add form and Lead import, therefore we can intervene that we need to focus more on the leads originated from API and landing page submission.
- We see max number of leads are generated by google/direct traffic. Max conversion ratio is by reference and welingak website.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. SMS sent has the highest conversion rate.
- Lead count for unemployed segment is the highest and conversion rate is also good. Though the lead count is less, working professional have a highest conversion rate.

THANK YOU