

Lead Score Case Study

Presented by:
Lalita Singh
Nikhil Swarnkar
Garvit Luthra



Agenda

- Problem Statement
- Assumptions
- Approach
- EDA Findings
- Conclusions
- Recommendations

Problem Statement

- X Education gets a lot of leads however the conversion rate is extremely poor.
- X Education wants the target lead conversion rate to be around 80%.
- Identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads.
- Once hot leads are found, conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Approach

- Data Cleaning and Data Manipulation

1. Inspecting the Dataframe

2. Check and handle NA values and missing values.

3. Drop columns, if it contains large amount of missing values and not useful .

4. Imputation of the values, if necessary.

5. Check and handle outliers in data.

- EDA

1. Univariate data analysis: value count, distribution of variable etc.

2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

- Data preparation for Modelling

- Splitting the Data into Test-Train Set

- Feature Scaling

- Model Building

- Preparing the model

- Conclusions and recommendations

Data Manipulation

This dataset has: 9240 rows, 37 columns

- As there are column with null values, we have dropped all column with 45% null values

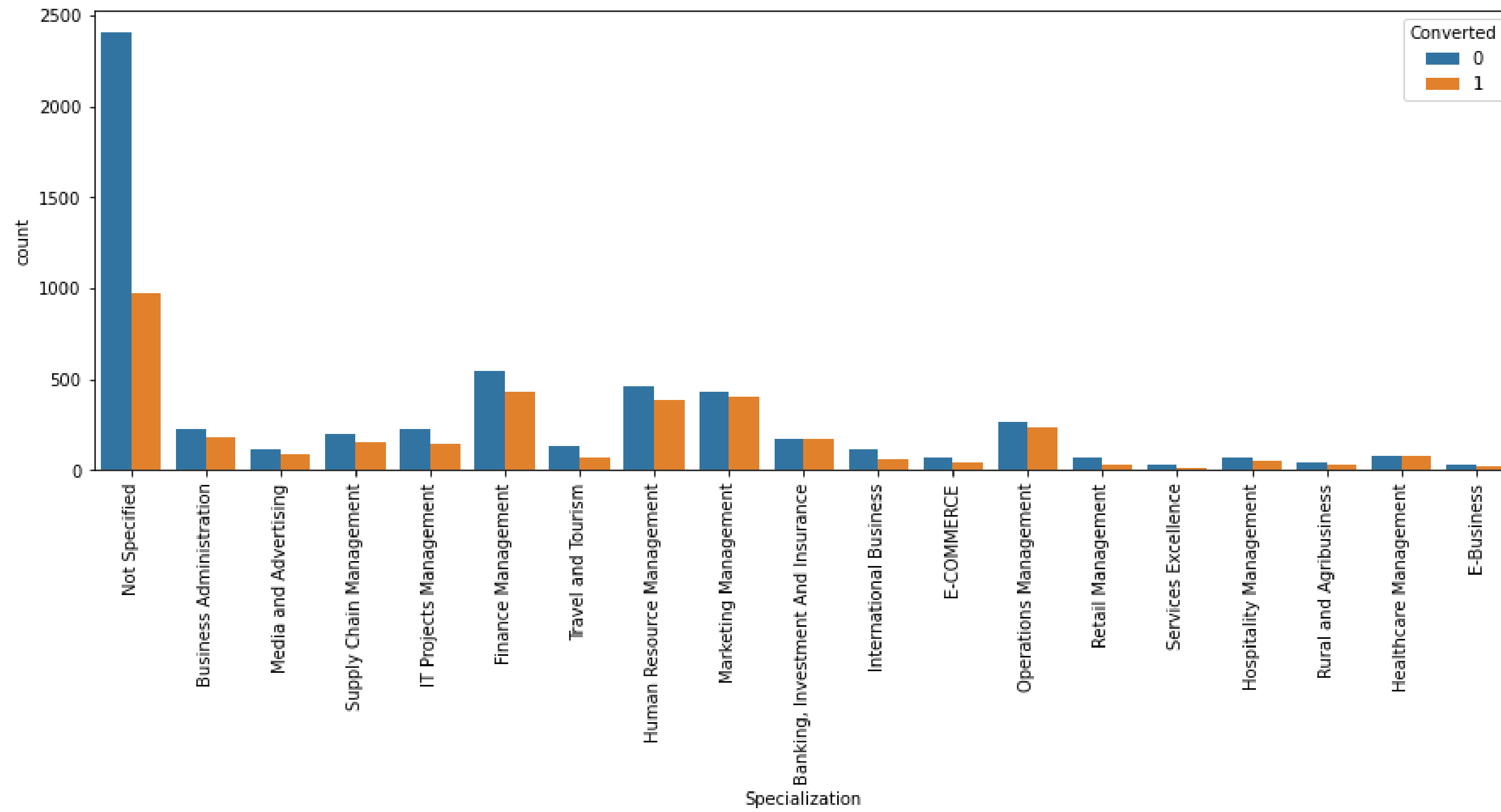
This dataset has: 9240 rows, 28 columns

- As India's occurrence in the country column is most, imputing 28% of missing values as India.
- In specialization, replacing NaN with Not specified as they can be students or their specialization might not be mentioned.
- In occupation, Nan are imputed to unknown

EDA

Insights:

Specialization with Management have higher number of leads as well as leads converted.

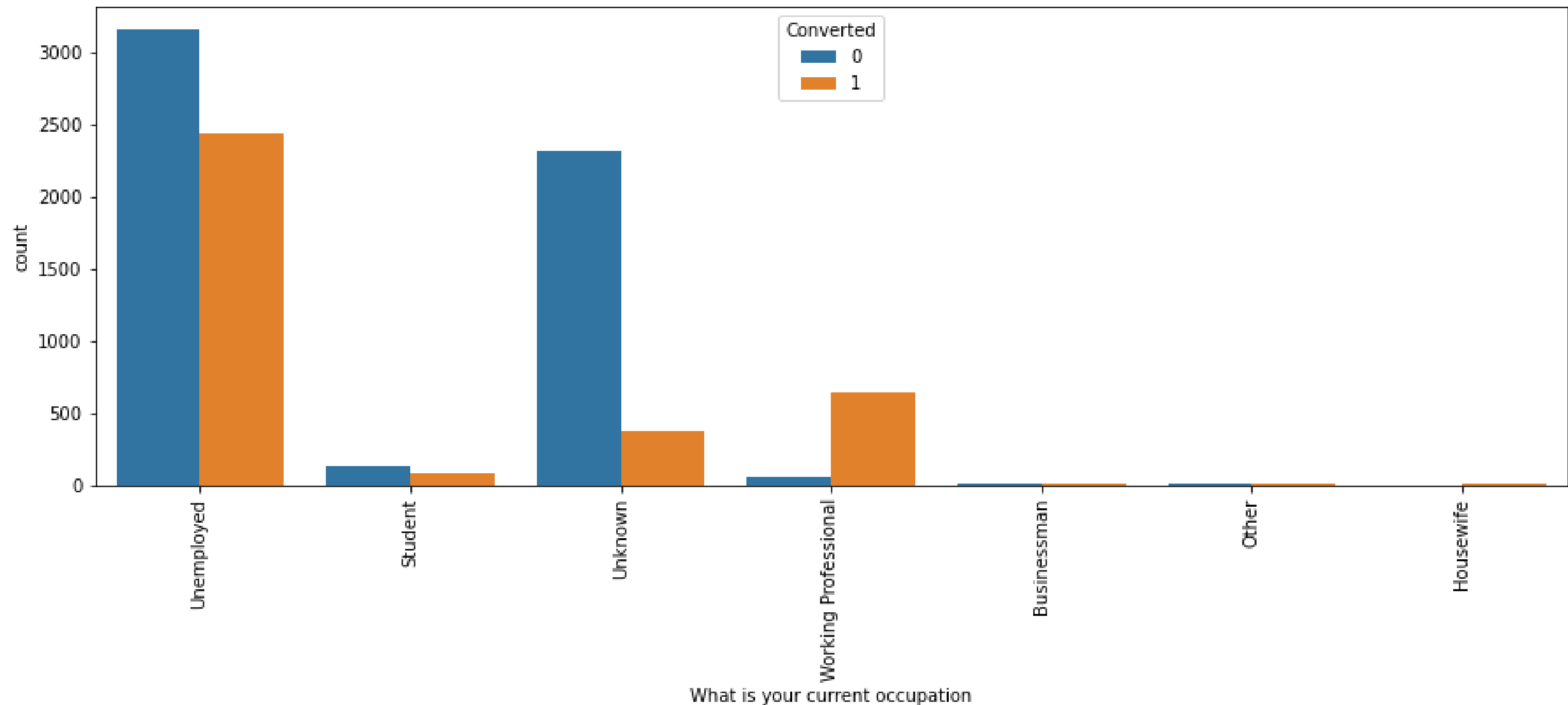


EDA

Insights:

Leads that are working professionals have a higher chance of conversion.

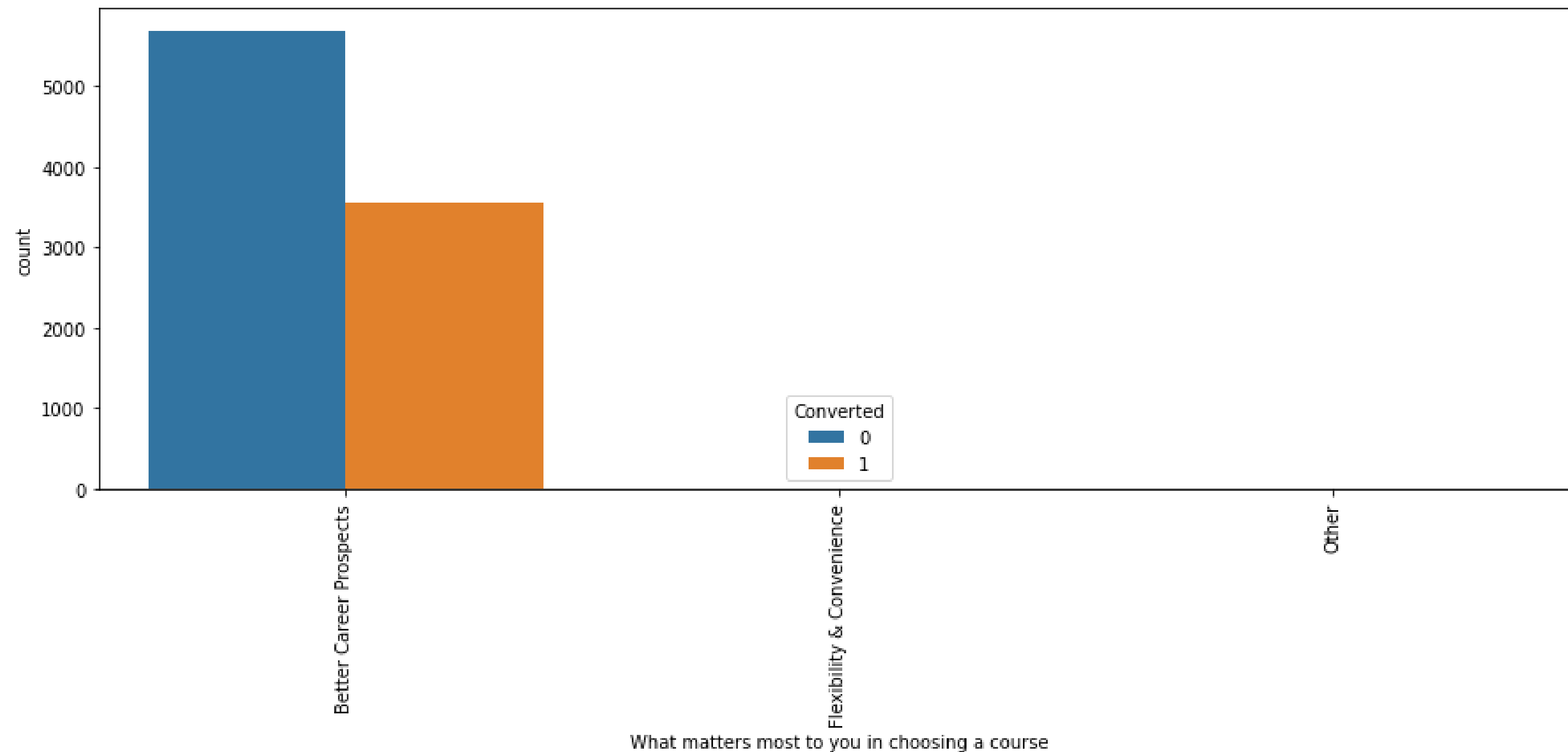
Unemployed leads are the most in terms of absolute numbers.



EDA

Insights:

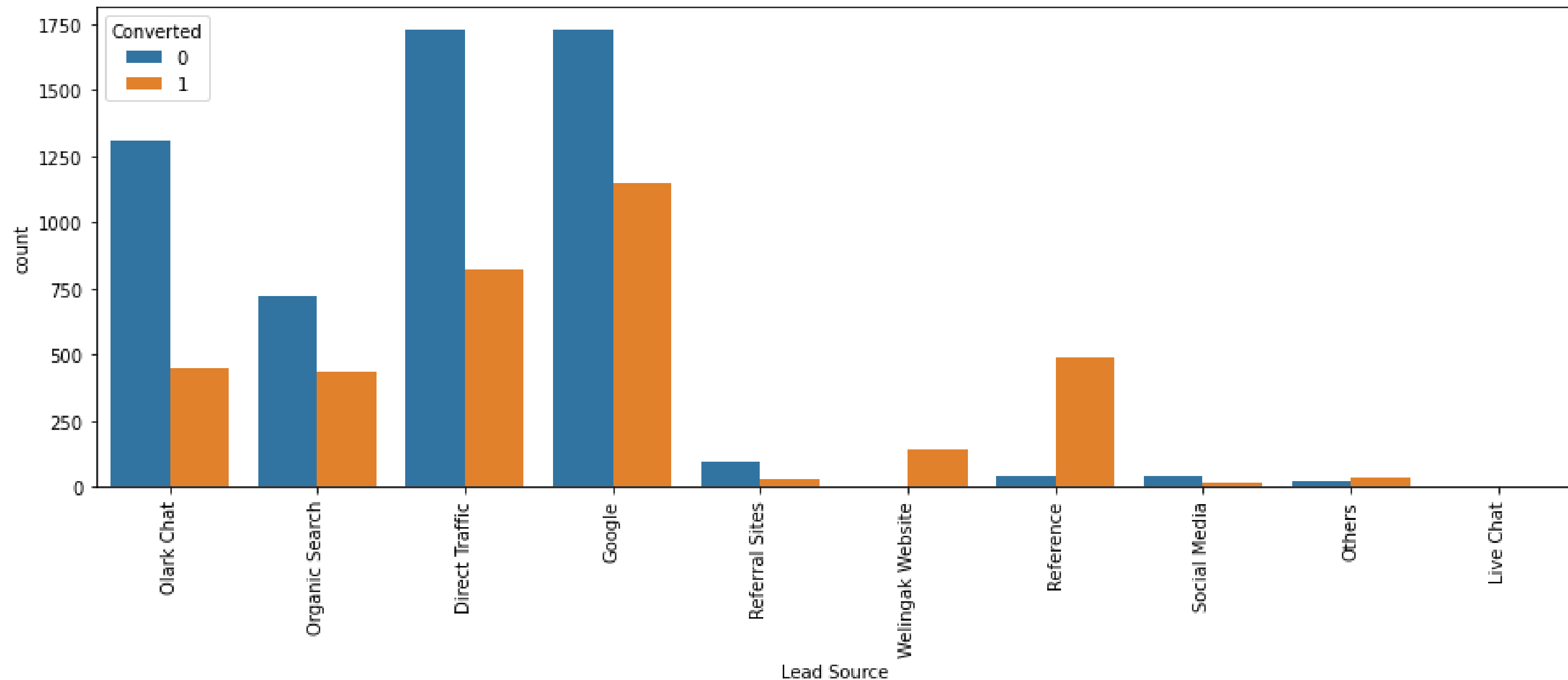
1. The reason to choose the course for 99% leads was 'Better Career Prospects'.



EDA

Insights:

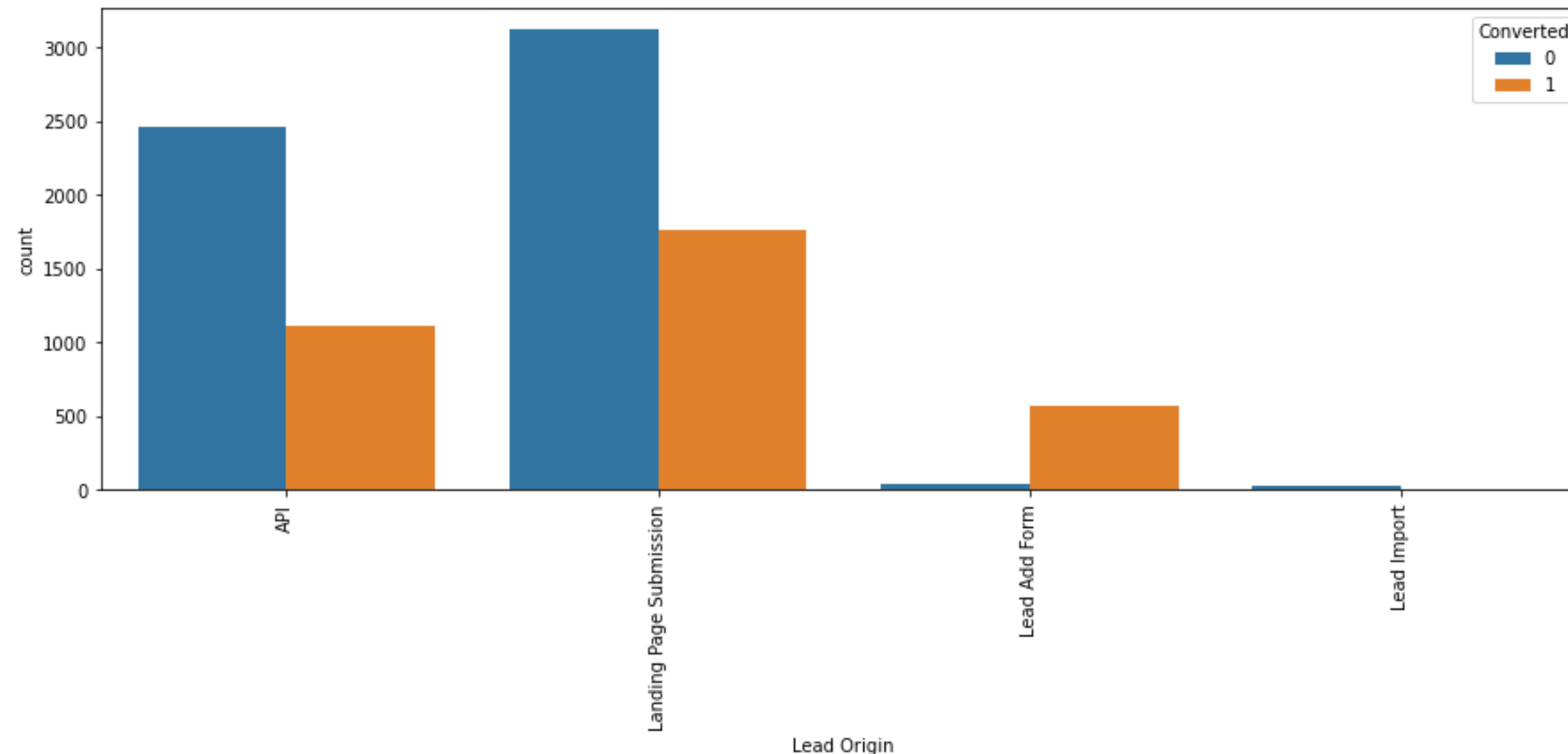
1. Maximum number of leads generated by Google and Direct Traffic.
2. Conversion rate is higher of leads from Reference and Wellingak Website
3. We should focus more on OlarkChat as we get a lot of leads but they are not converted.



EDA

Insights:

- 1.API and Landing Page Submission has higher number of leads as well as conversion.
- 2.Lead Add Form has a very high conversion rate but count of leads are not very high.
- 3.Lead Import and Quick Add Form get very few leads.
- 4.It's recommended 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.



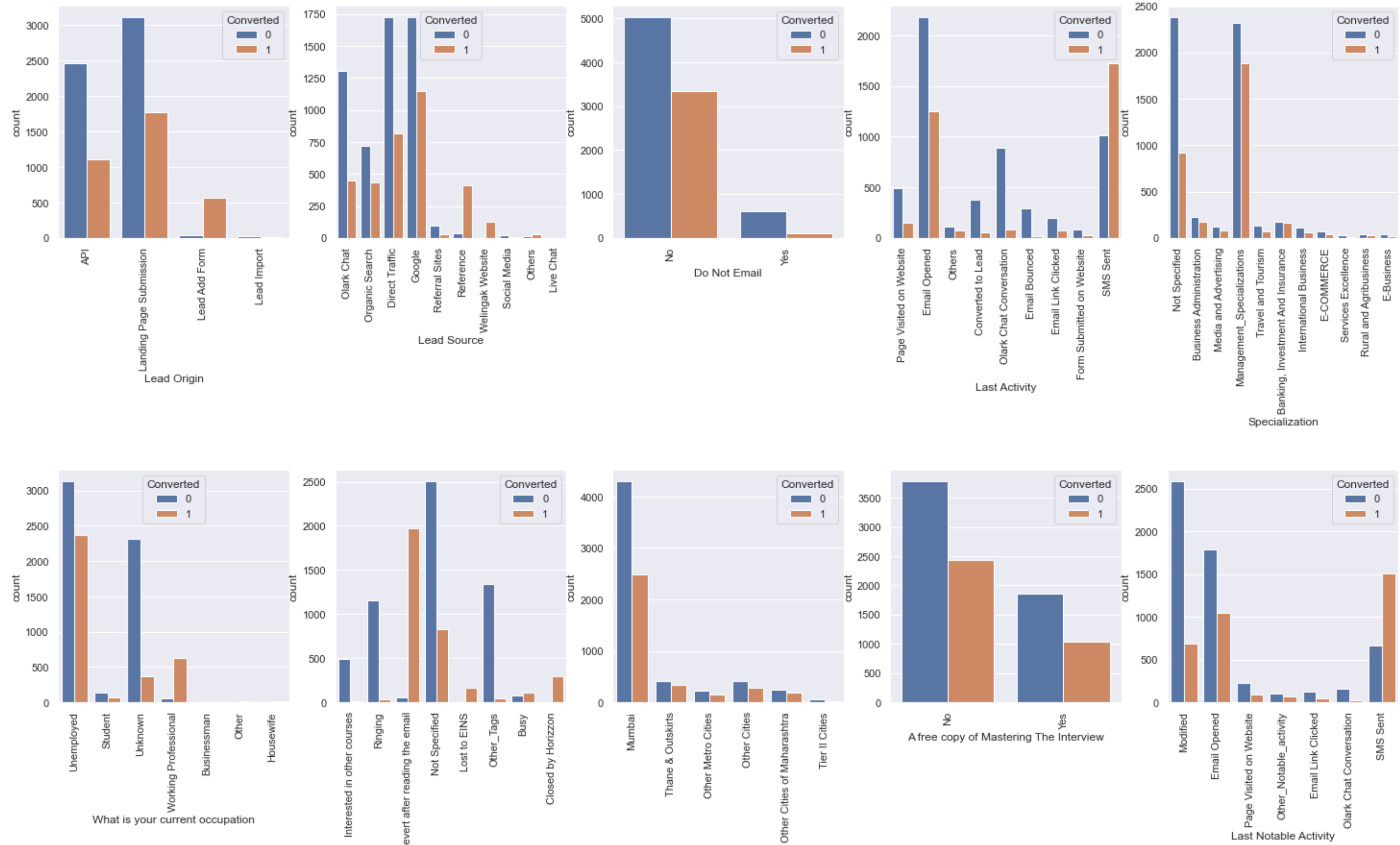
EDA

Insights:

1. Page Views Per Visit & TotalVisits are highly correlated with each other
2. Converted has the highest correlation with Total Time Spent on Website
3. Total Time Spent on Website and TotalVisits are also correlated with each other
4. Page Views Per Visit and Converted are negatively correlated



EDA



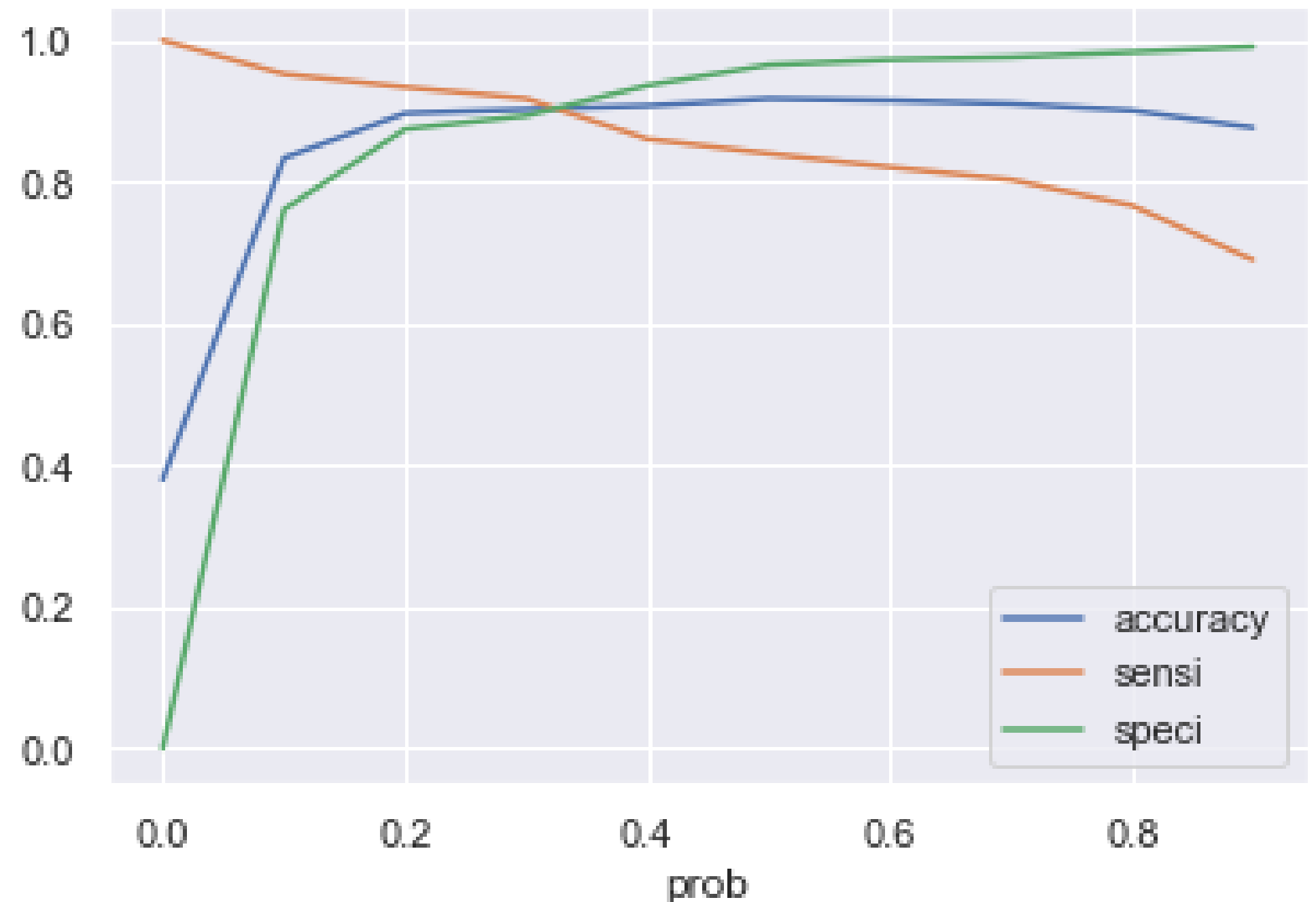
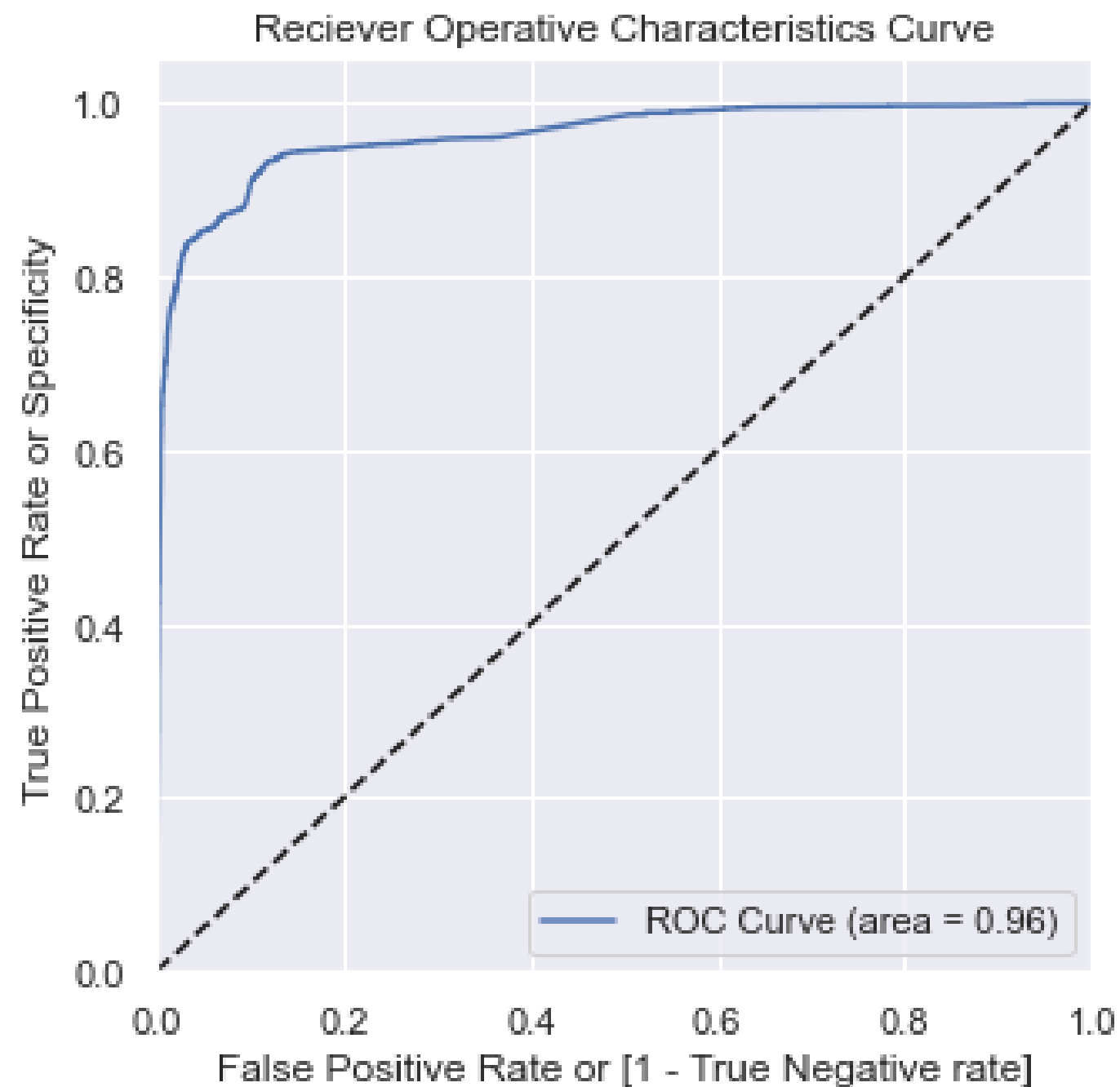
Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a test split, we have chosen 70:30 ratio.
- Feature Selection Using RFE
- Since 'All' the p-values are less we checked the Variance Inflation Factor to see if there is any correlation between the variables.
- Accuracy- Overall accuracy of the model: 90.27%
- Sensitivity – Proportion of converted predicted correctly: 91.77%
- Specificity – Proportion of not – converted predicted correctly: 89.34%

ROC CURVE

Insights

- ROC Curve should be a value close to 1. We are getting a good value of 0.96 indicating a good predictive model.
- From the curve below, 0.3 is the optimum point to take it as a cutoff probability.



Conclusion

To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google

The total time spent on the Website has a great influence on lead conversion.

Leads who are working professionals and are from the management domain have a high chance of conversion

As there are so many unemployed leads, providing internship support/Job support may help in conversion.