

Lead Scoring Case Study

INCREASING LEAD CONVERSION RATE FOR X EDUCATION

BY, AKILA NAGARAJAN

AKSHAY JADHAV

AKANKSHA GUPTA

Problem Statement

X Education, an online course provider for professionals, drives potential customers to its site through various marketing channels, including ads on search engines like Google. When visitors provide contact details, they are marked as leads. These leads also come from previous referrals. The sales team then reaches out to these leads, resulting in a 30% conversion rate. However, despite acquiring a large number of leads daily, not all are quality leads, leading to a low conversion rate. For instance, of 100 daily leads, only 30 typically convert.

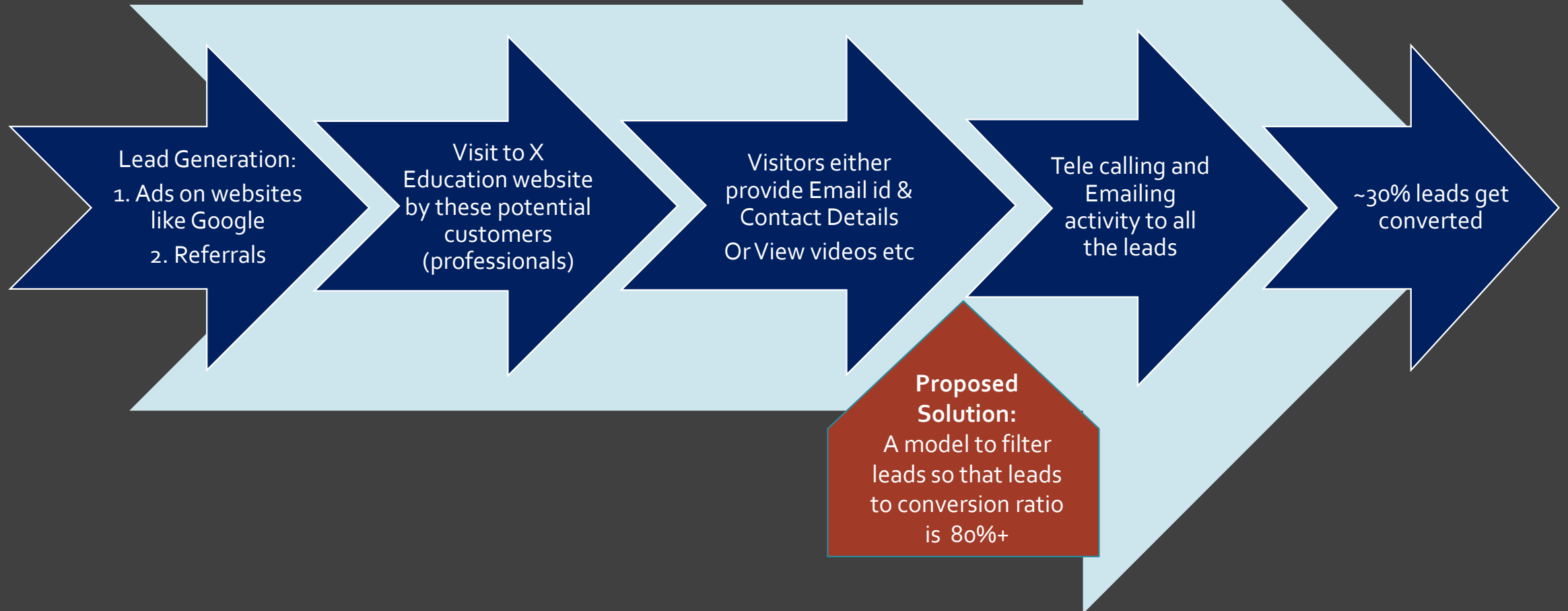
The company's goal is to enhance efficiency by pinpointing 'Hot Leads' or those most likely to become paying customers.

Your task is to assist X Education in identifying these promising leads. You're required to develop a model that assigns a lead score to each prospect. A higher score indicates a greater likelihood of conversion, and vice versa.

Regarding data, you have access to a historical dataset containing roughly 9,000 entries, featuring various attributes like 'Lead Source', 'Total Time Spent on Website', and 'Last Activity', among others. The 'Converted' column indicates whether a lead was successfully converted (1) or not (0).

Lead – Conversion Process

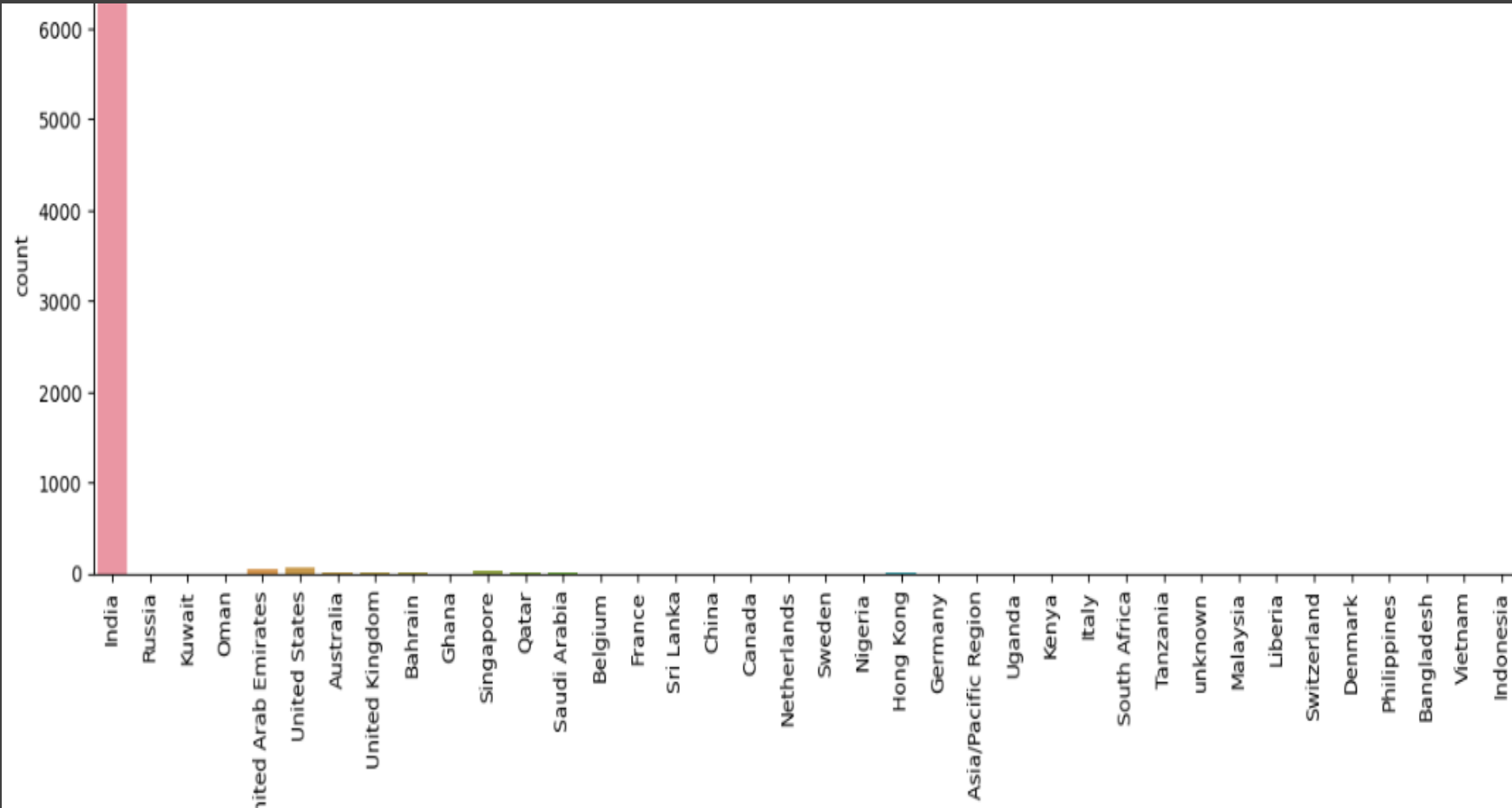
Lead to Conversion process



Exploratory Data Analysis

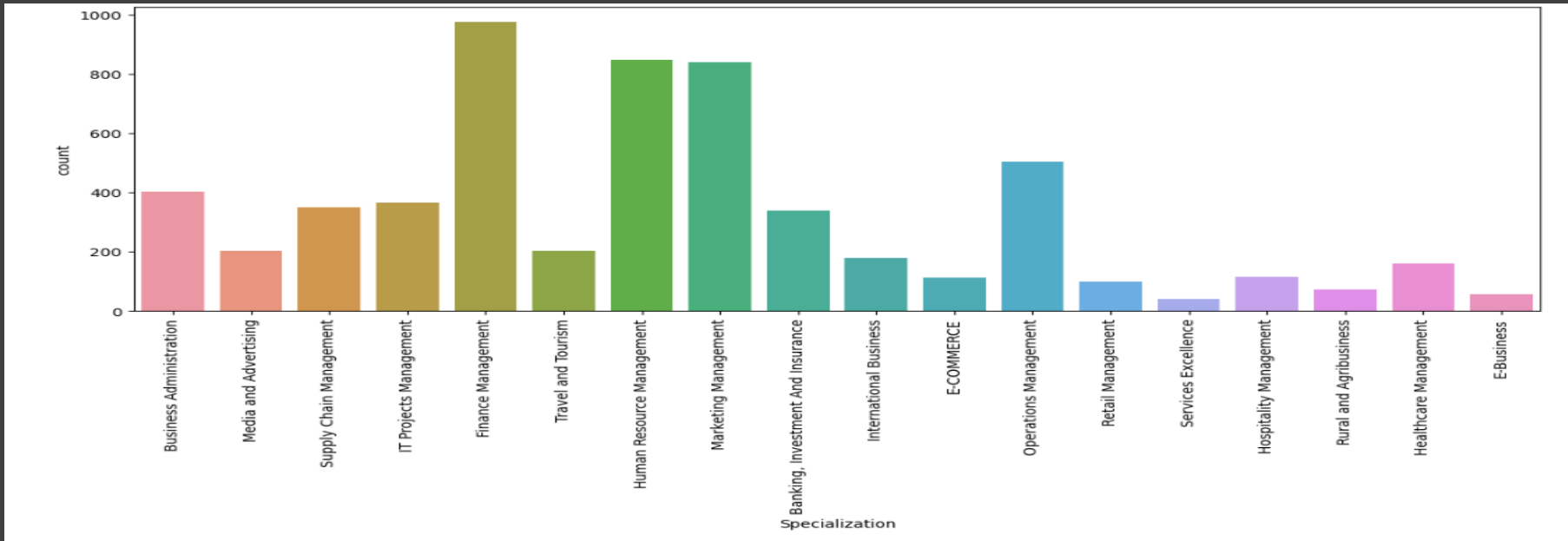
DATA VISUALIZATIONS

Missing values



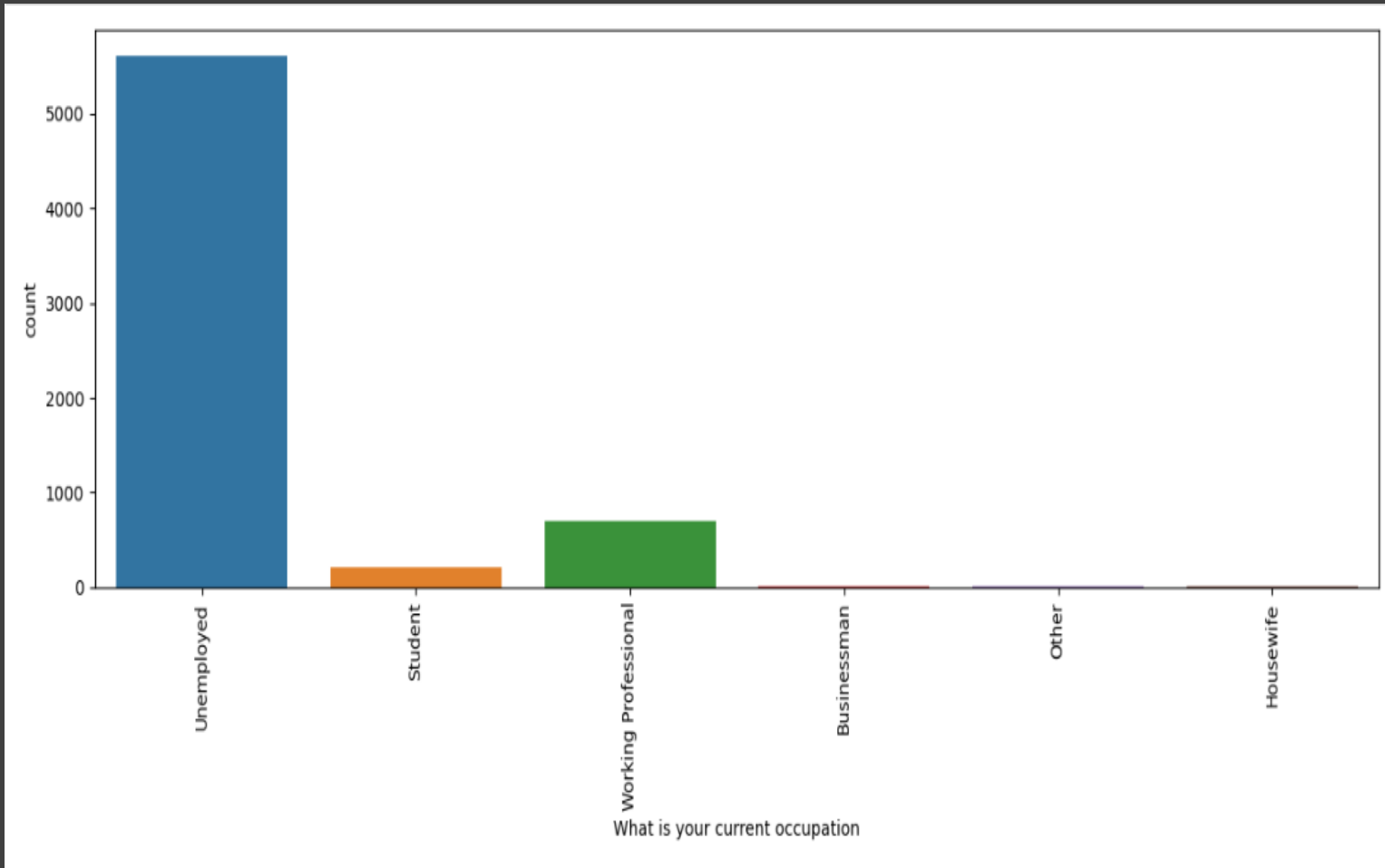
Since the column is skewed towards 'India', we can impute the missing values with India.

Missing values



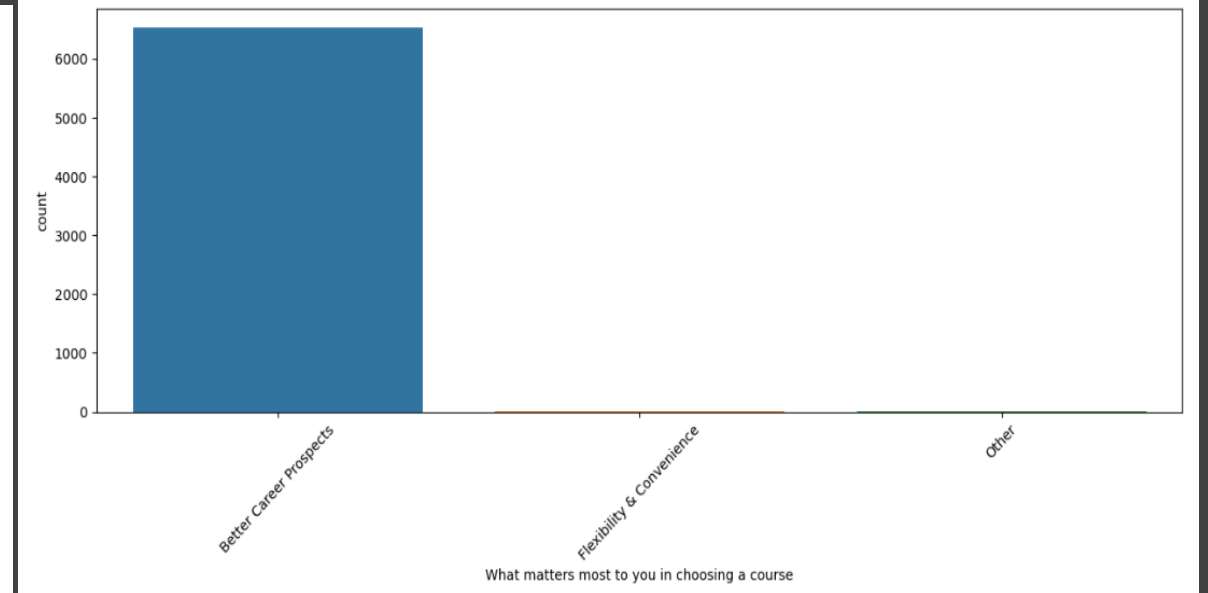
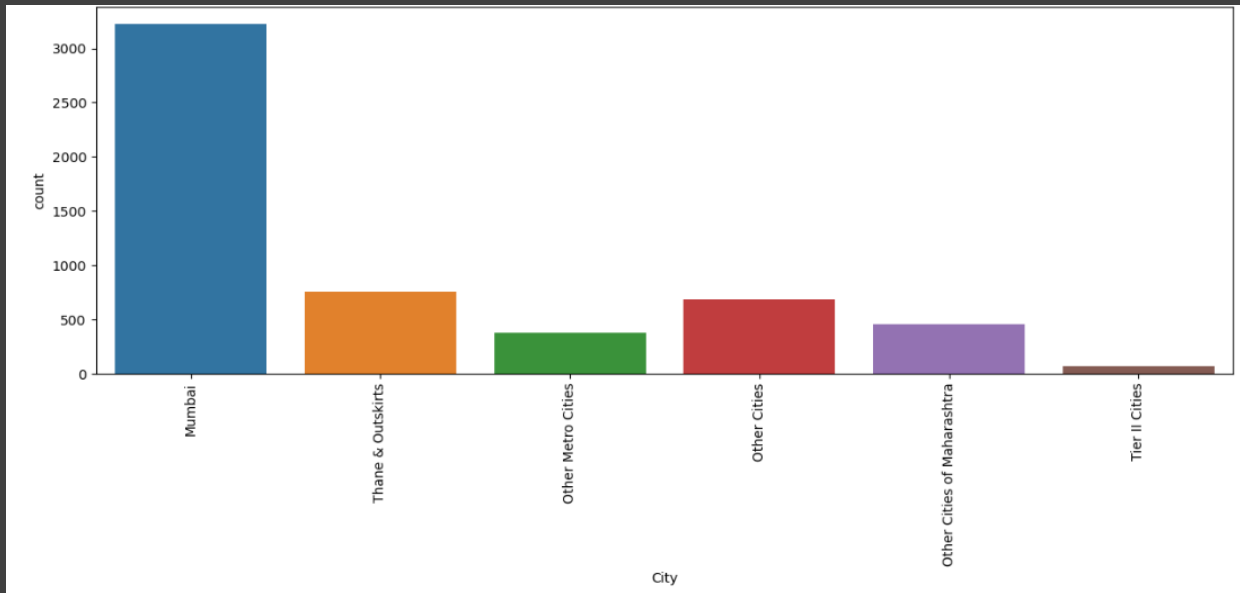
A new category called 'Others' is created for the missing values.

Missing values



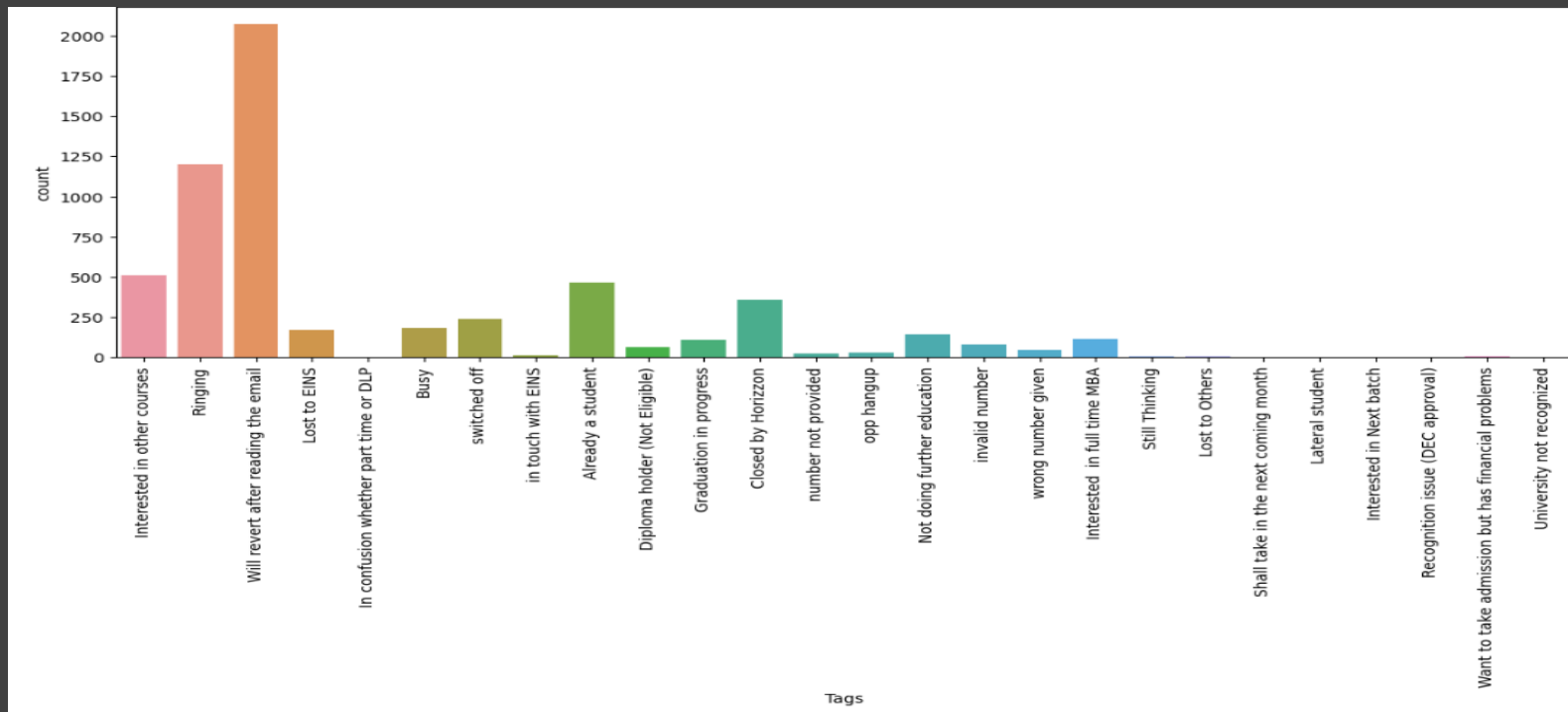
Since the column is skewed towards 'Unemployed', we can impute the missing values with 'Unemployed'.

Missing values

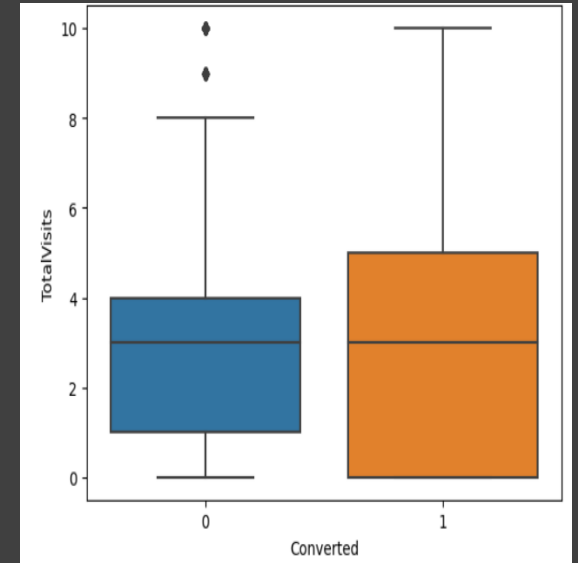
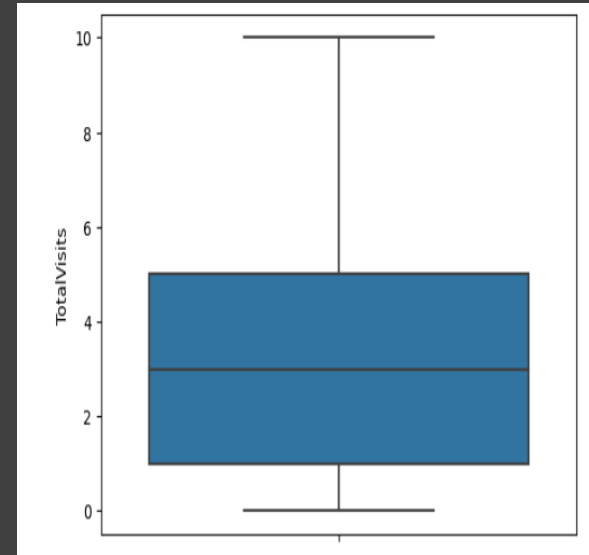
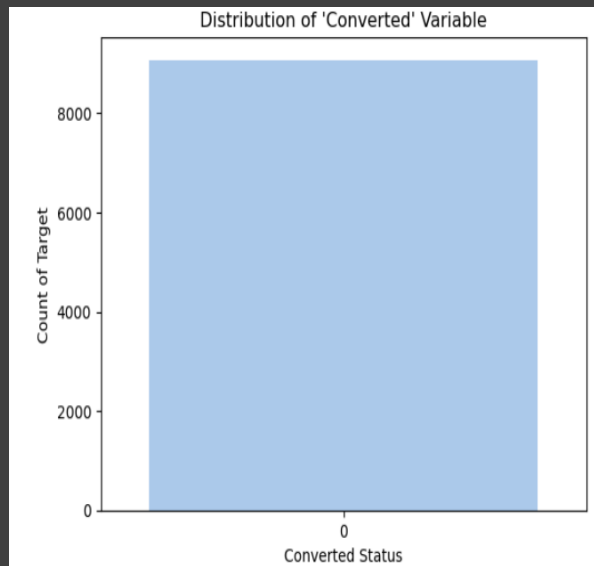
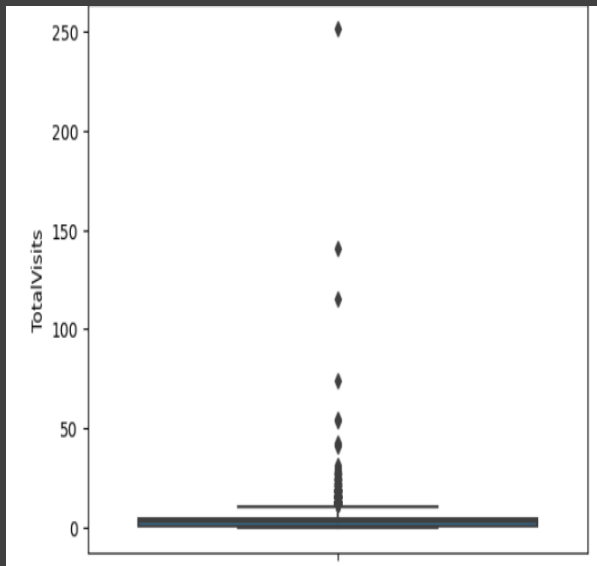


Highly skewed columns were removed
Otherwise only the rows with NULL values were removed

Missing values



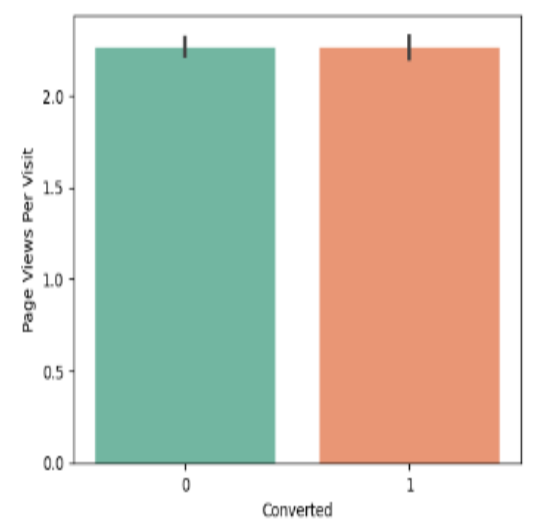
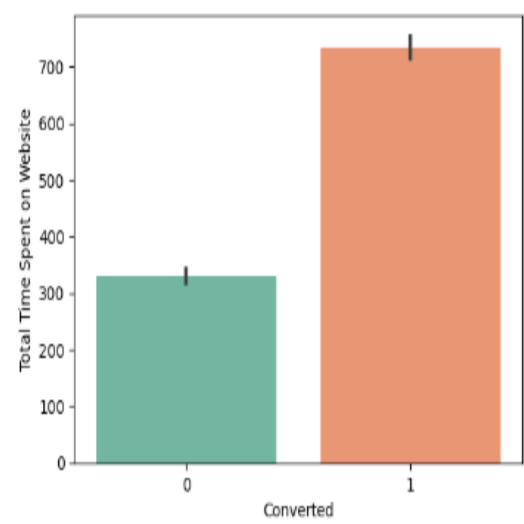
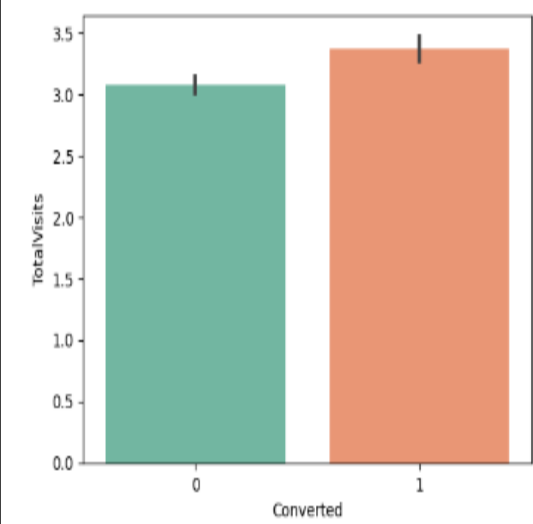
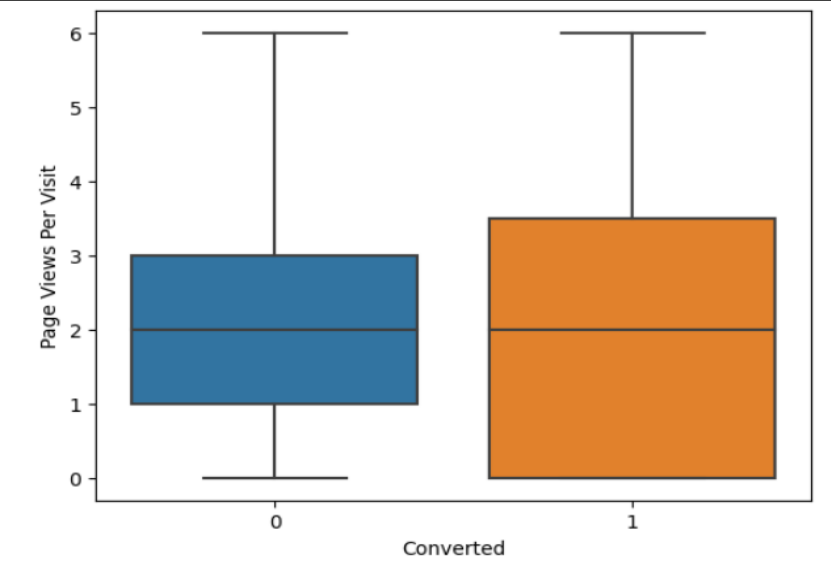
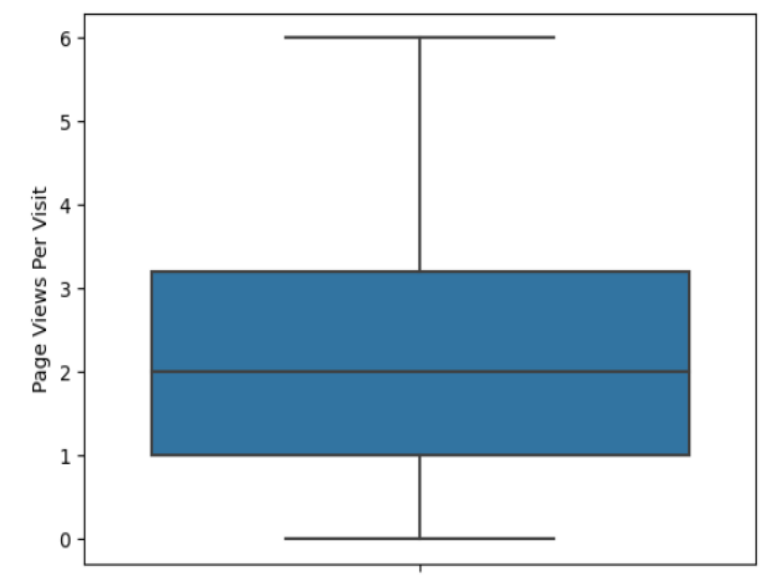
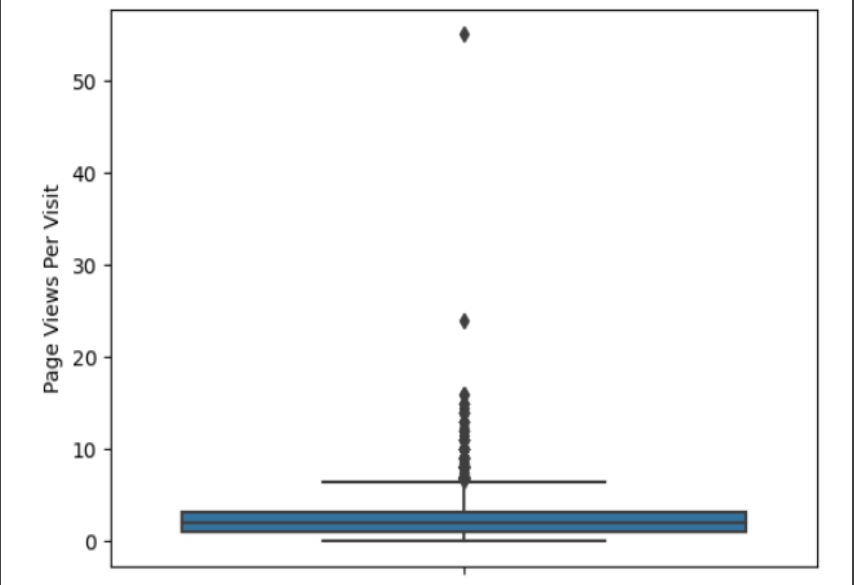
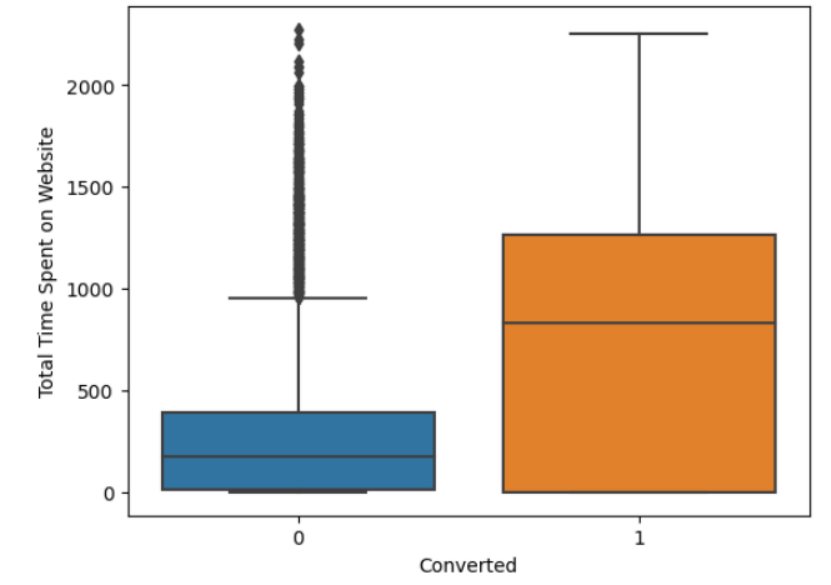
Highly skewed columns were removed
Otherwise only the rows with NULL values were removed



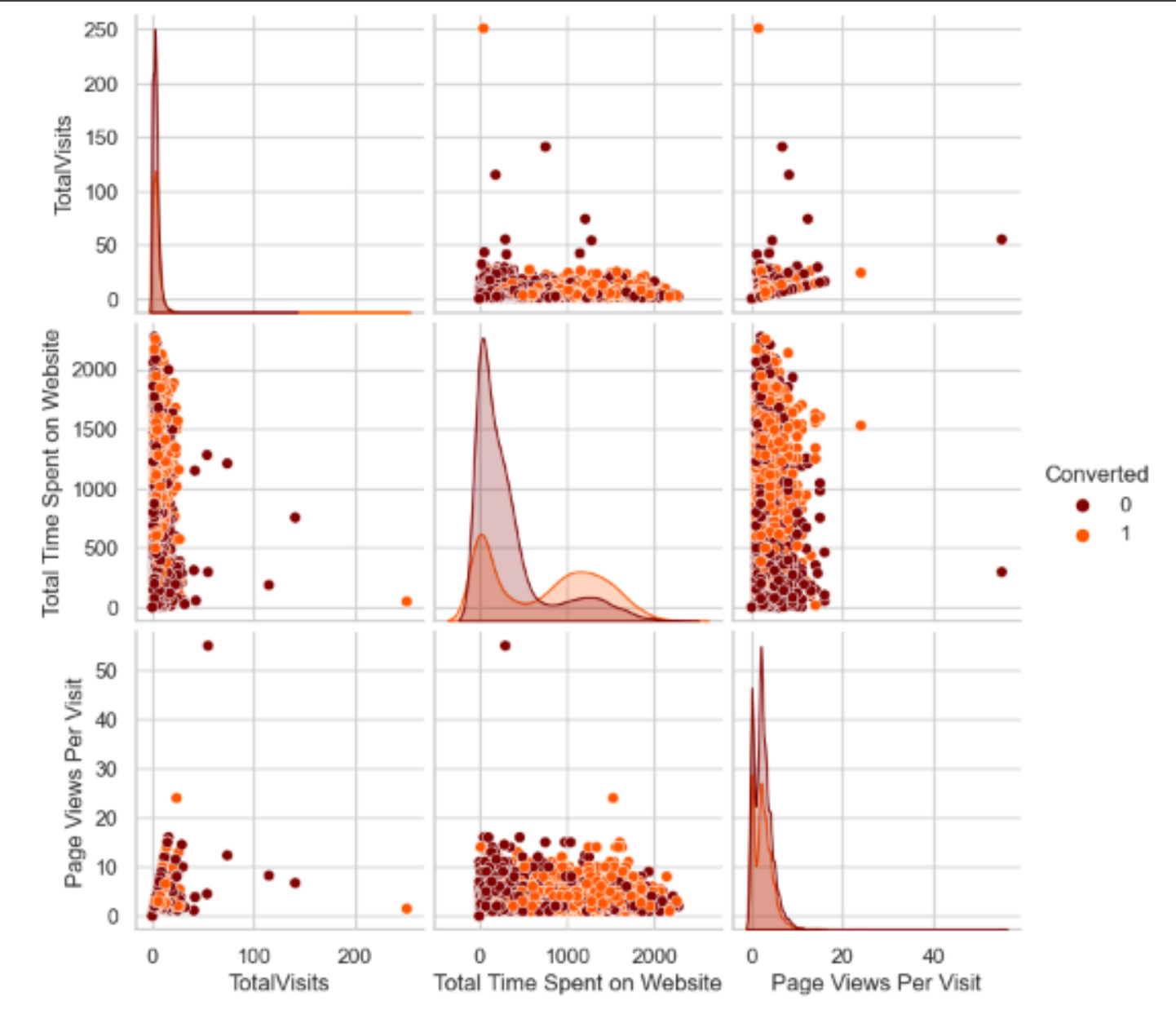
Dealing with categorical Variables

Bivariate & Univariate Analysis

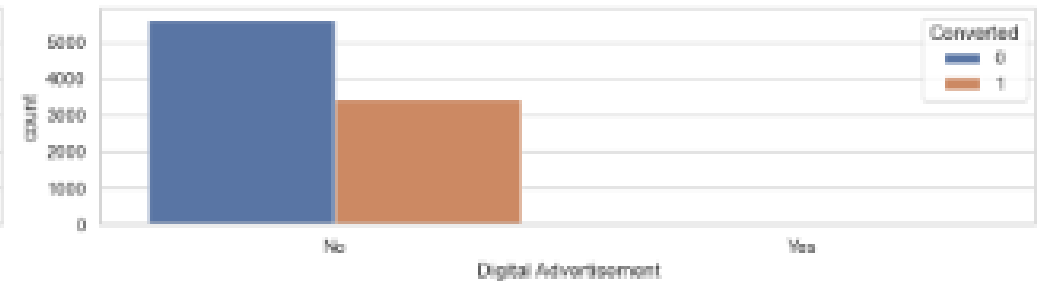
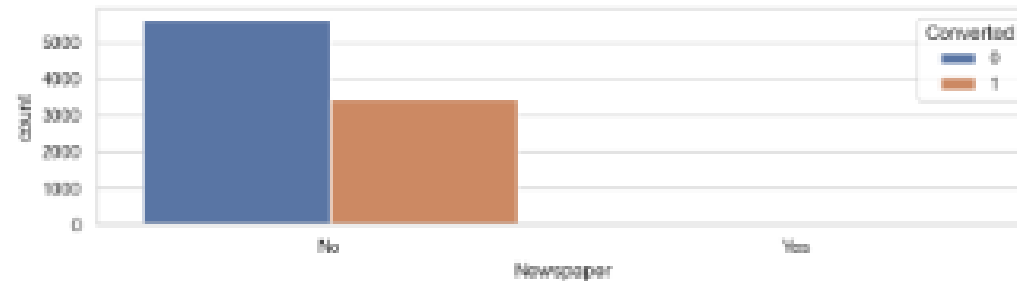
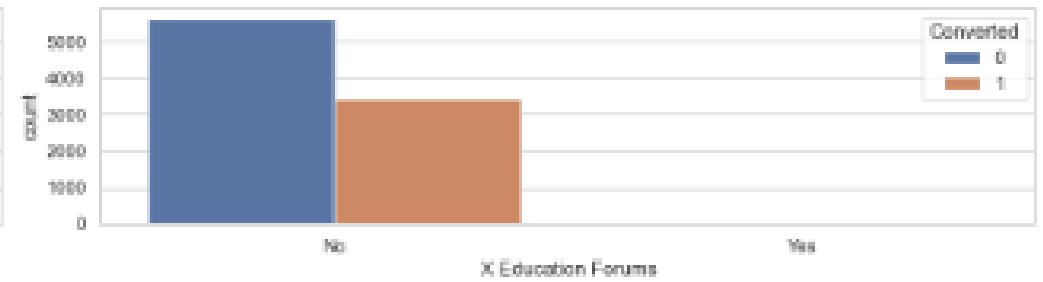
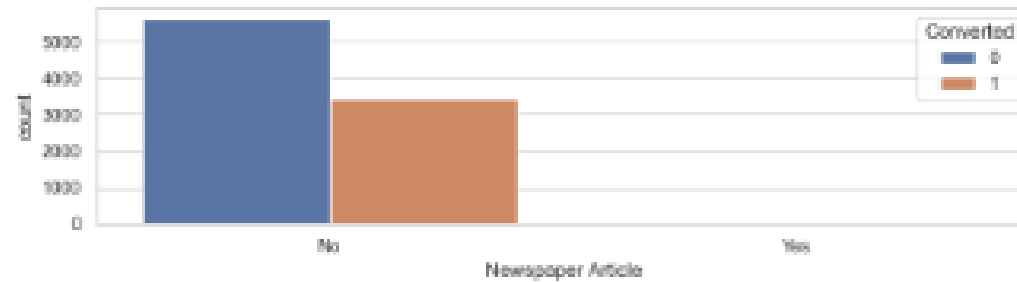
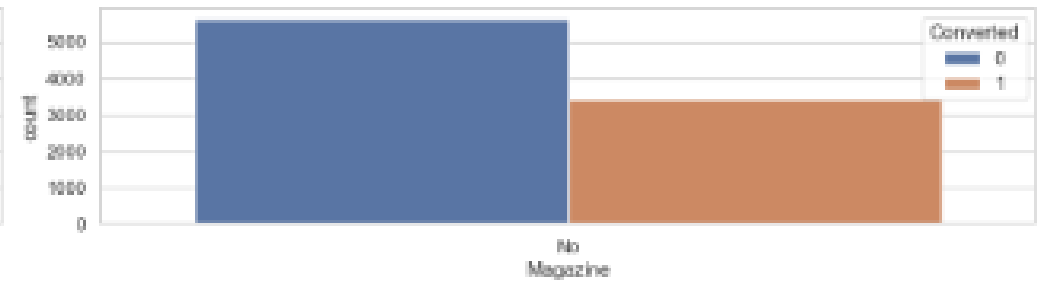
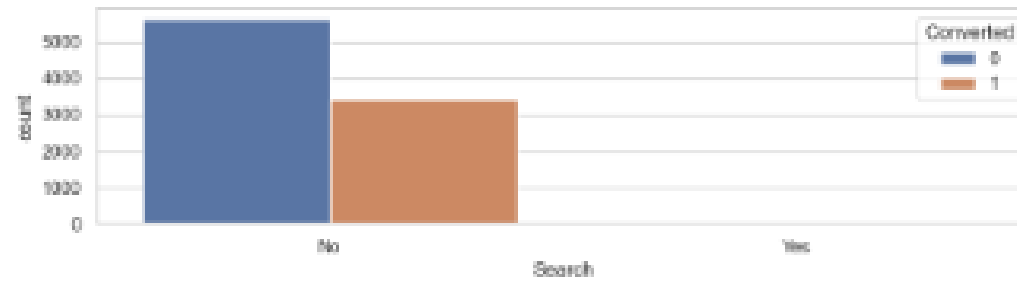
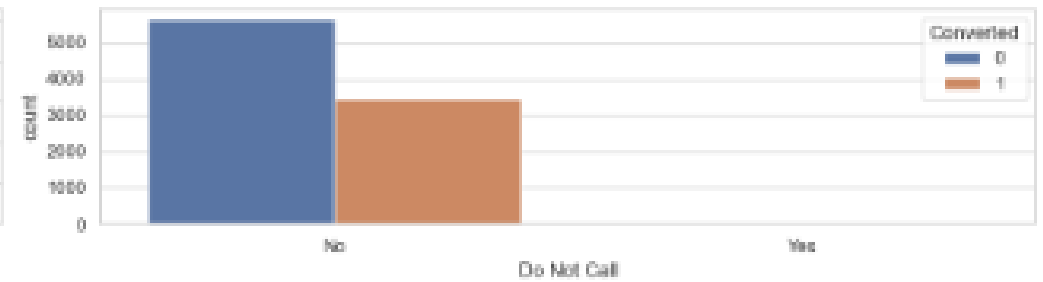
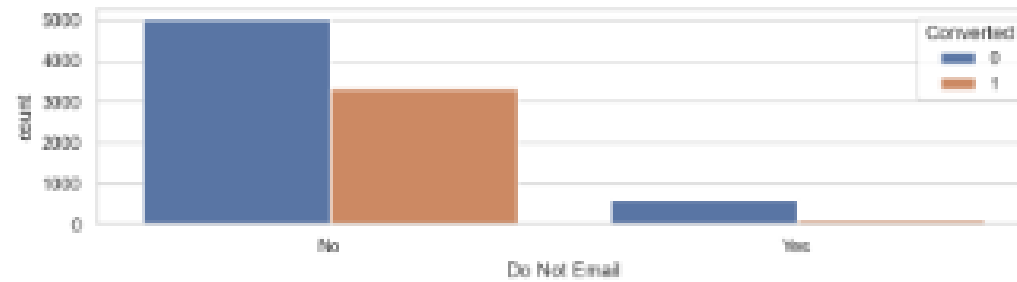
Bivariate & Univariate Analysis



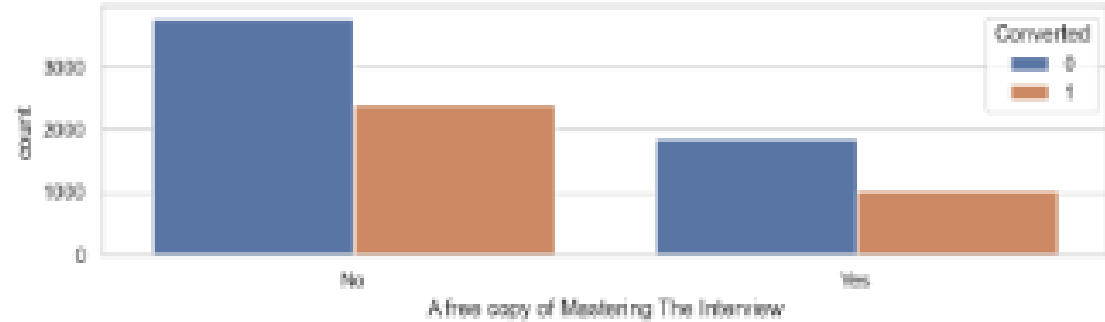
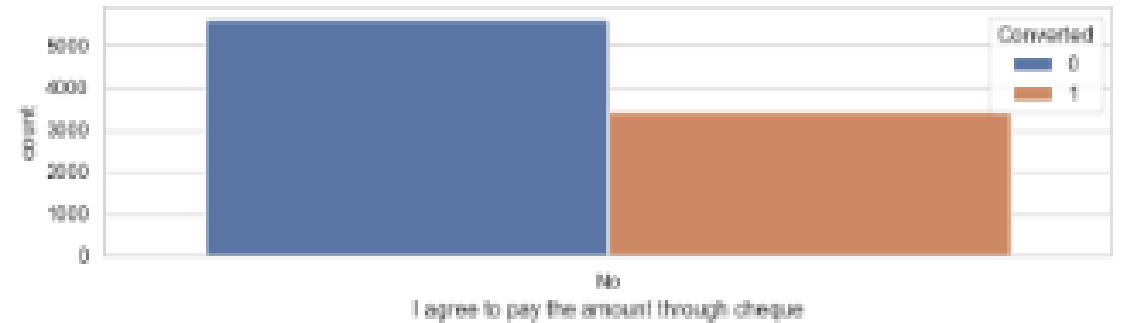
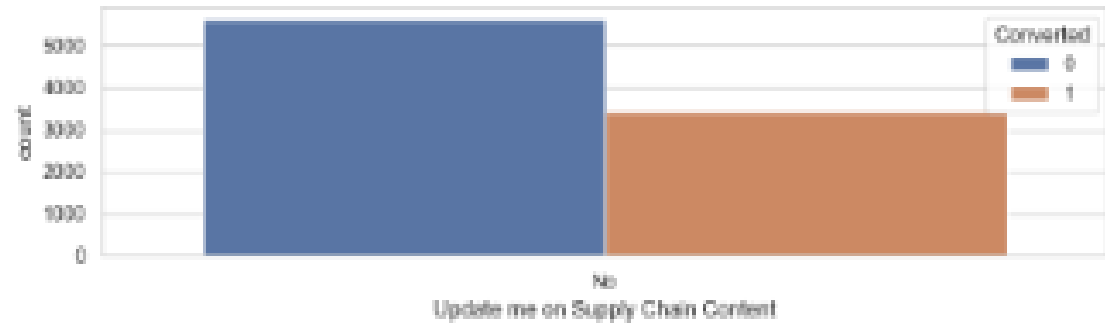
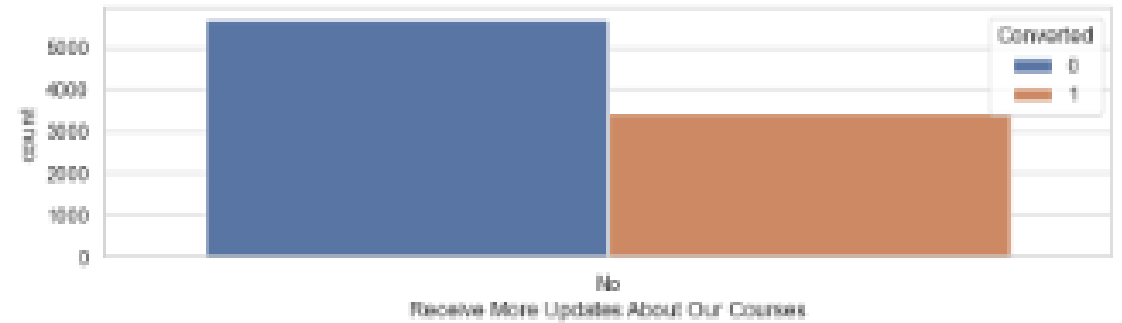
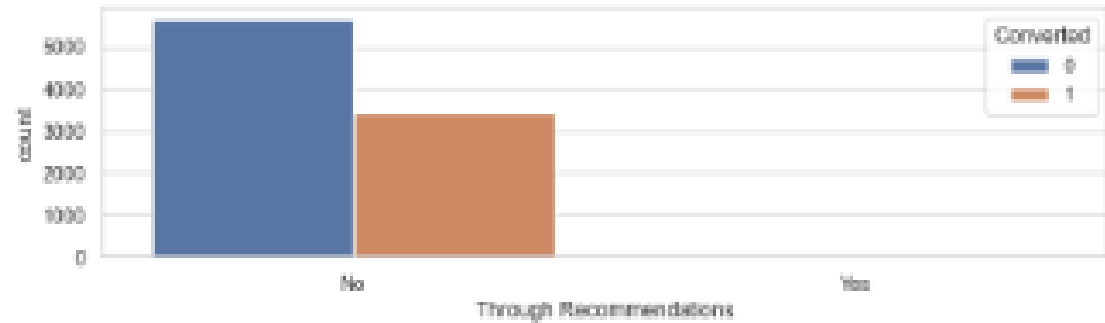
Bivariate &
Univariate Analysis



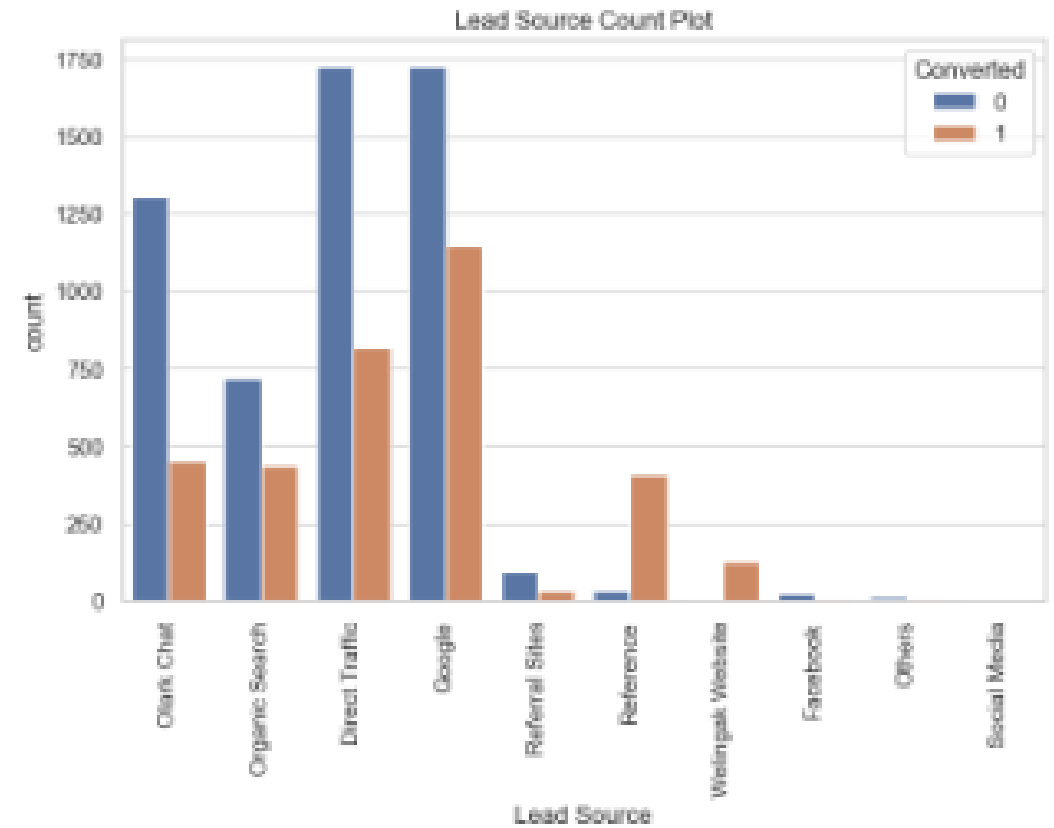
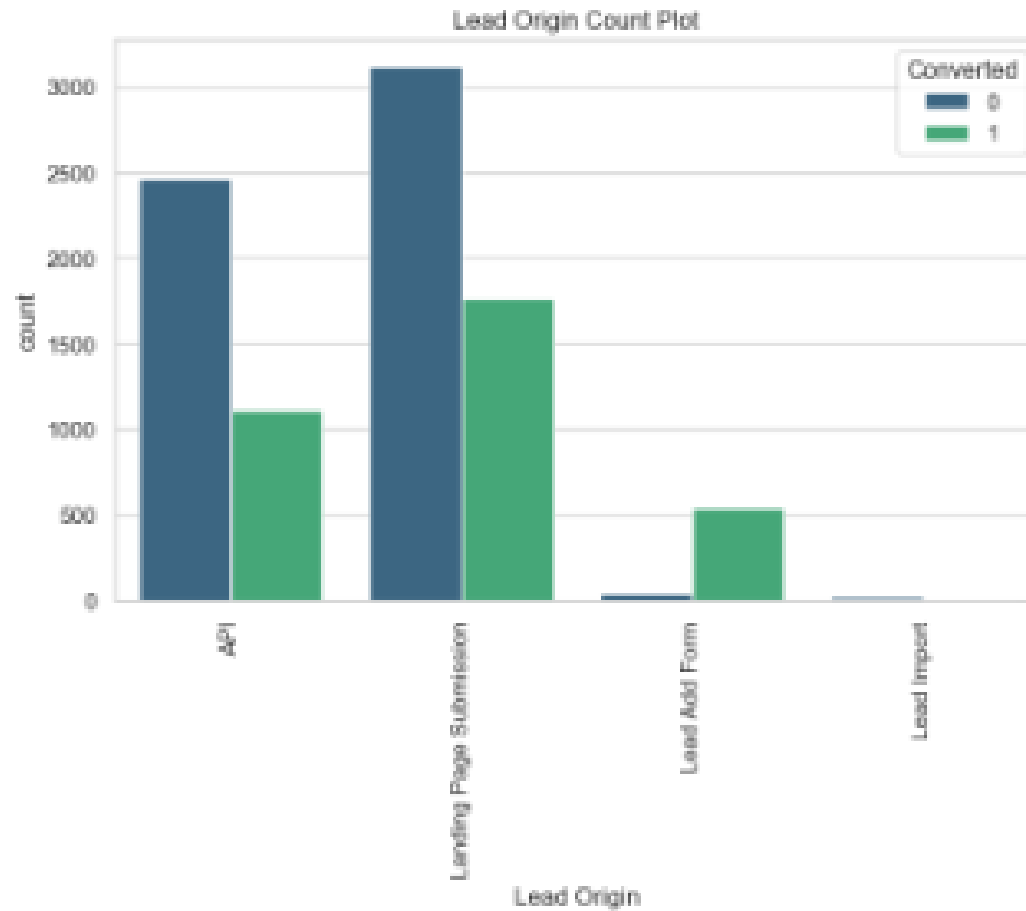
Bivariate & Univariate Analysis



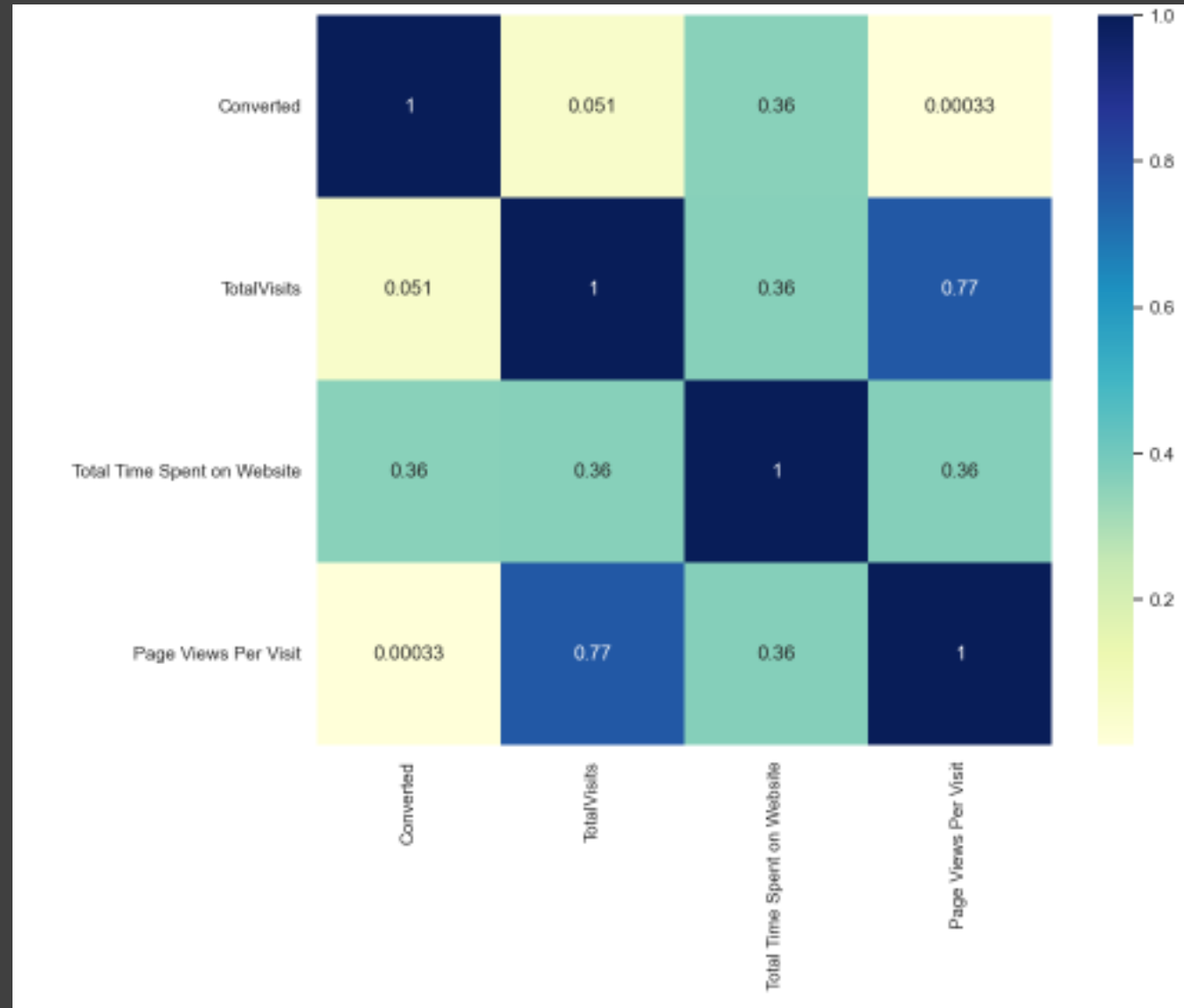
Bivariate & Univariate Analysis



Bivariate & Univariate Analysis

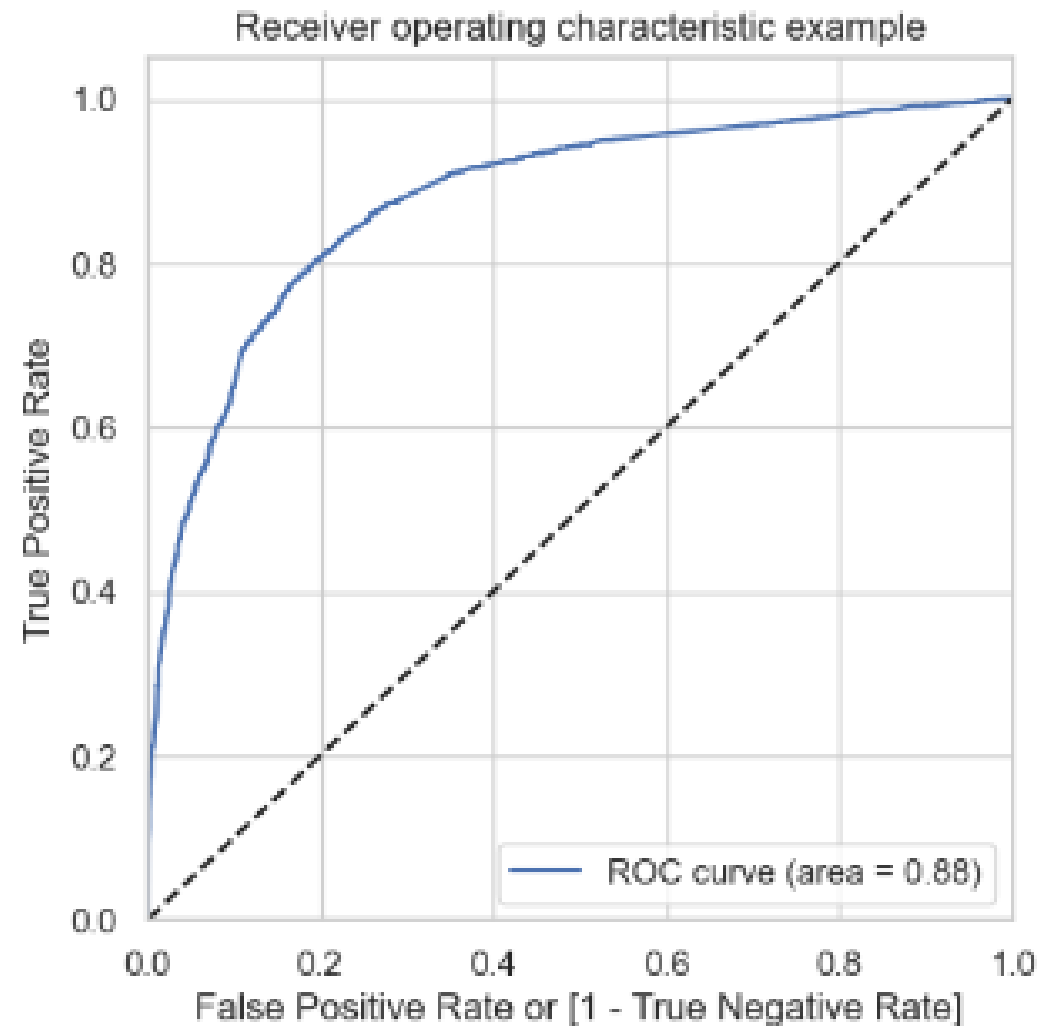


Heat map for numerical variables



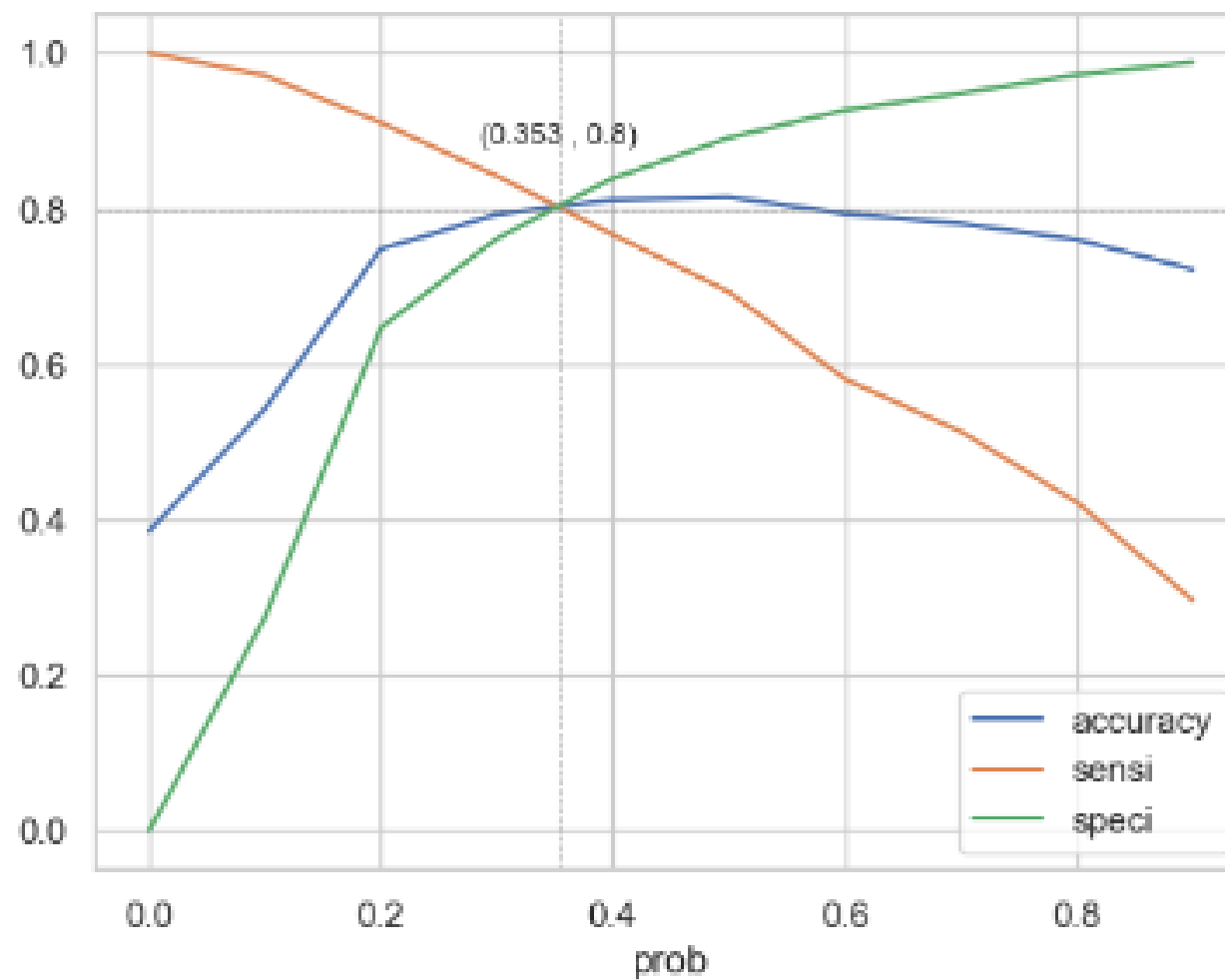
Model building and Evaluation

ROC CURVE



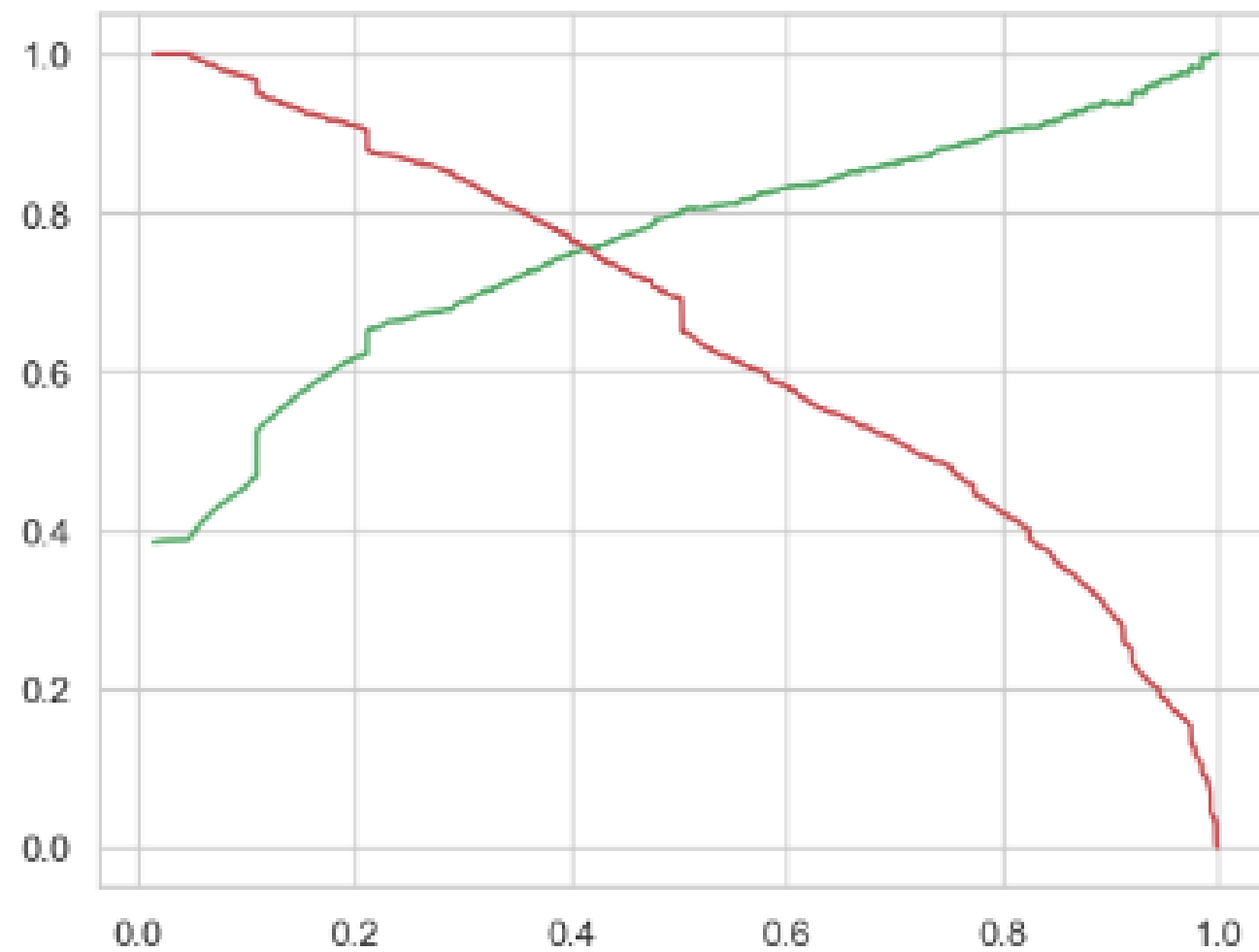
The area under the ROC curve is 0.88. Hence, this is a good model.

CUT OFF POINT FOR TRAIN DATA



From the above graph a cut off point is 0.353

CUT OFF POINT FOR TEST DATA



The above cut off point is 0.42

Conclusion

Evaluated the model's performance on the test data by calculating accuracy, generating a classification report, and creating a confusion matrix. The values are as follows:

- Accuracy : 80.36%
- Sensitivity :80.29%
- Specificity : 80.41%

Once the model's performance becomes satisfactory, we used it to make predictions on Test data. The values are as follows:

- Accuracy: 81.08%
- Sensitivity: 75.12%
- Specificity: 84.48%

Conclusion

Insights gathered from this Logistic regression model:

The important features that contribute to the probability of a lead being converted are:

- ❖ Lead Origin: Lead Add Form
- ❖ What is your current occupation: Working Professional
- ❖ Lead Source_Welingak Website

These features can be used to target marketing campaigns more effectively. For example, if a company is trying to sell a product or service to working professionals, they may want to focus their marketing efforts on leads that have those characteristics.

- The Sales team should focus on leads originating from 'Add Forms'.
- They should focus on leads sourced from Welingak Website.

Conclusion

- They should focus on targeting Working Professionals
- Students can be approached, but they will have a lower probability of converting since they are already studying and would not be willing to enroll into a course specially designed for working professionals, so early in the tenure. However, this can also be a motivating factor to ensure industry readiness by the time they complete their education.
- They should focus on targeting the people who spend more Time Spent on the Website.
- They should not focus on unemployed leads. They might not have a budget to spend on the course
- Do not focus on leads who are on 'Do not email' lists as they have a very low probability of converting.