LEAD SCORE CASE STUDY

GROUP MEMBER

SHREYA ARON
SHUBHRANGI SHARMA
SIDDHANT SINGH

TABLE OF CONTENT

PROBLEM STATEMENT

X Education, an online course provider for professionals, faces a low lead conversion rate of around 30%. Despite acquiring numerous leads through marketing campaigns and referrals, only a small percentage of these leads turn into paying customers. The company aims to improve its conversion rate to 80% by identifying "Hot Leads"—prospects with the highest potential to convert. This requires building a predictive model to assign a lead score to each prospect, enabling the sales team to prioritize and focus their efforts on nurturing the most promising leads through targeted communication and education.

SOLUTION METHODOLOGY

1. Importing Libraries and Data
   a) Import Libraries, Suppress Warnings and Set Display
   b) Reading the Data
2. Data Understanding and Inspection
3. Data Cleaning & Preparation
   a) Treatment for 'Select' values
   b) Handling Missing Values ,Outliers and unwanted data
4. Data Analysis (EDA)
   a) Univariate Analysis
   b) Bivariate Analysis
5. Data Preparation
   a) Dummy Variables

SOLUTION METHODOLOGY

6 Test-Train Split

7 Feature Scaling
a) Looking at Correlations

8 Model Building
a) Feature Selection Using RFE

9 Model Evaluation

10 Making Predictions on test set

DATA MANIPULATION

Total Number of Rows =37, Total Number of Columns =9240

The value 'Select' is treated as null. Columns with over 40% missing values were dropped, totaling 7.

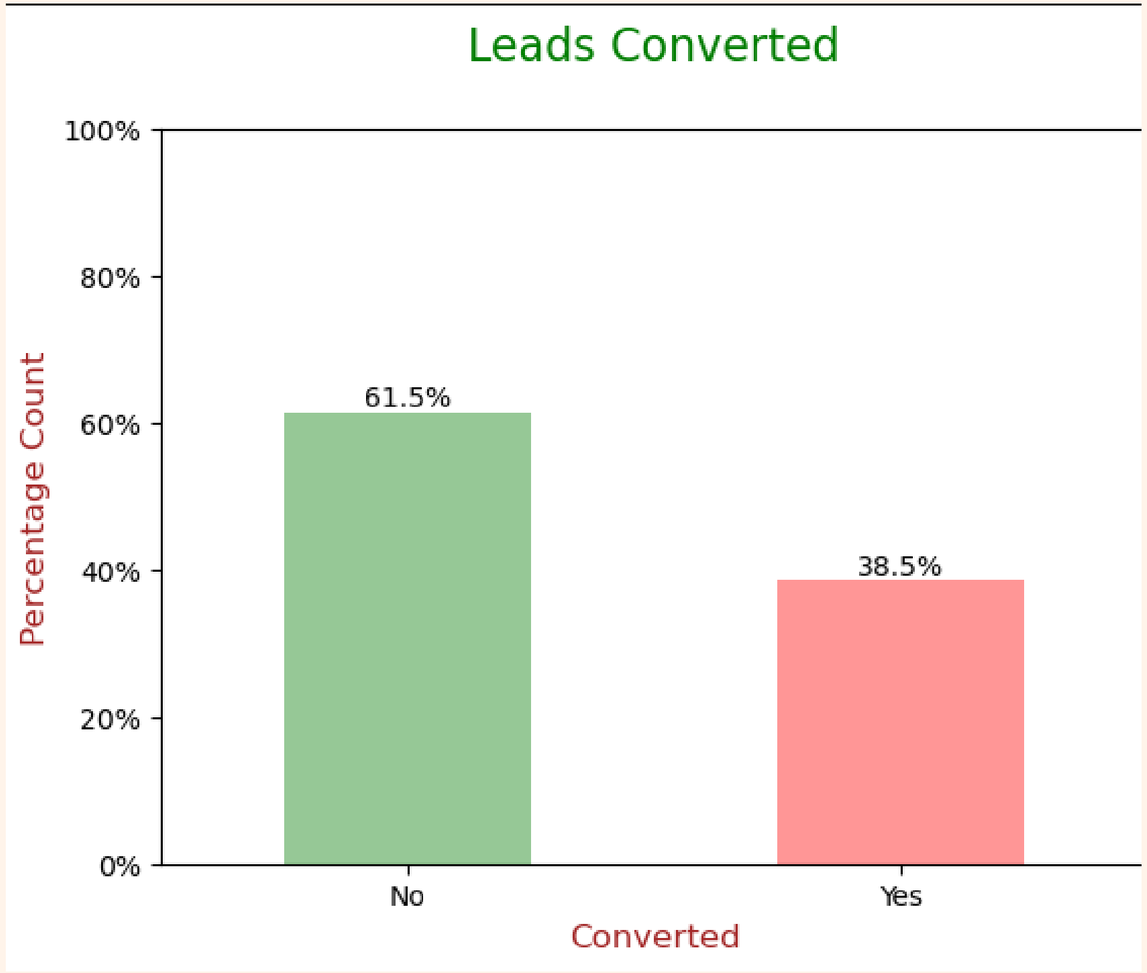
Checked value counts in each categorical column to decide whether to impute or drop missing values.

Removing columns with one unique value, irrelevant columns, and skewed categorical columns.

Also handle the outliers in columns with numerical values and Fixing Invalid values and Standardising Data in columns

EDA

UNIVARIATE ANALYSIS



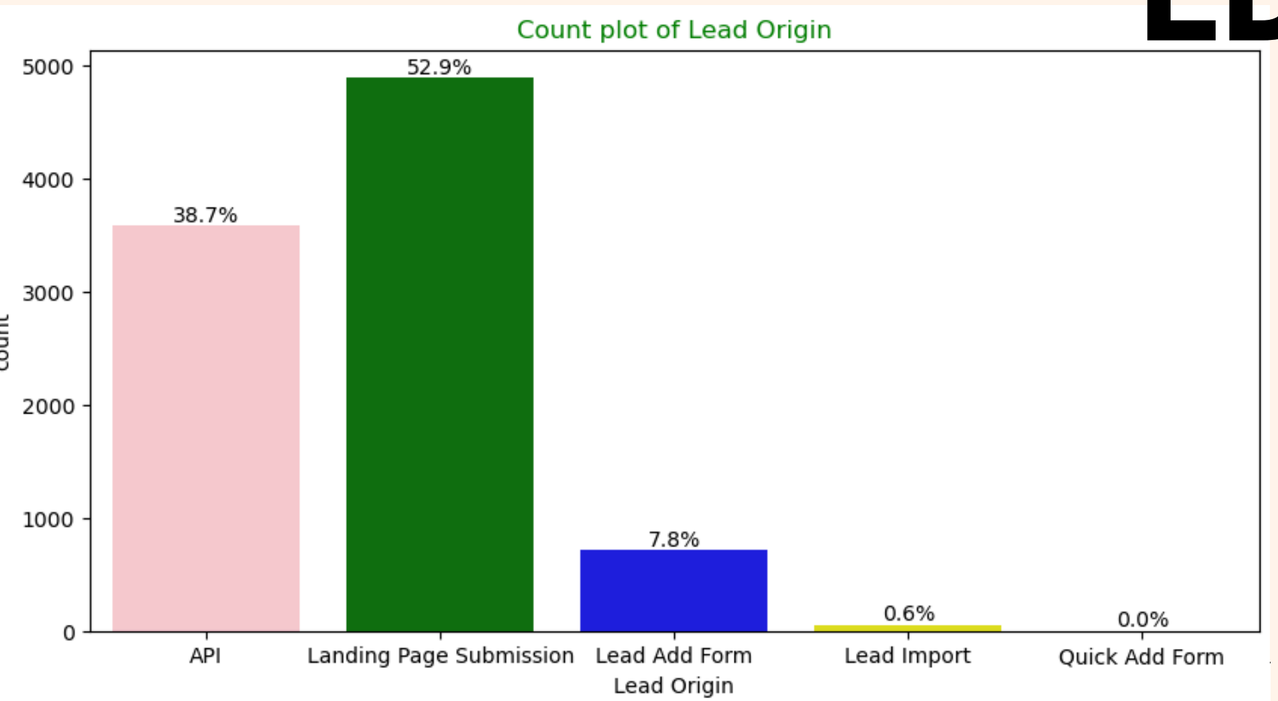
1. Lead Conversion Rate: 38.5% of leads convert, aligning with the company's average.

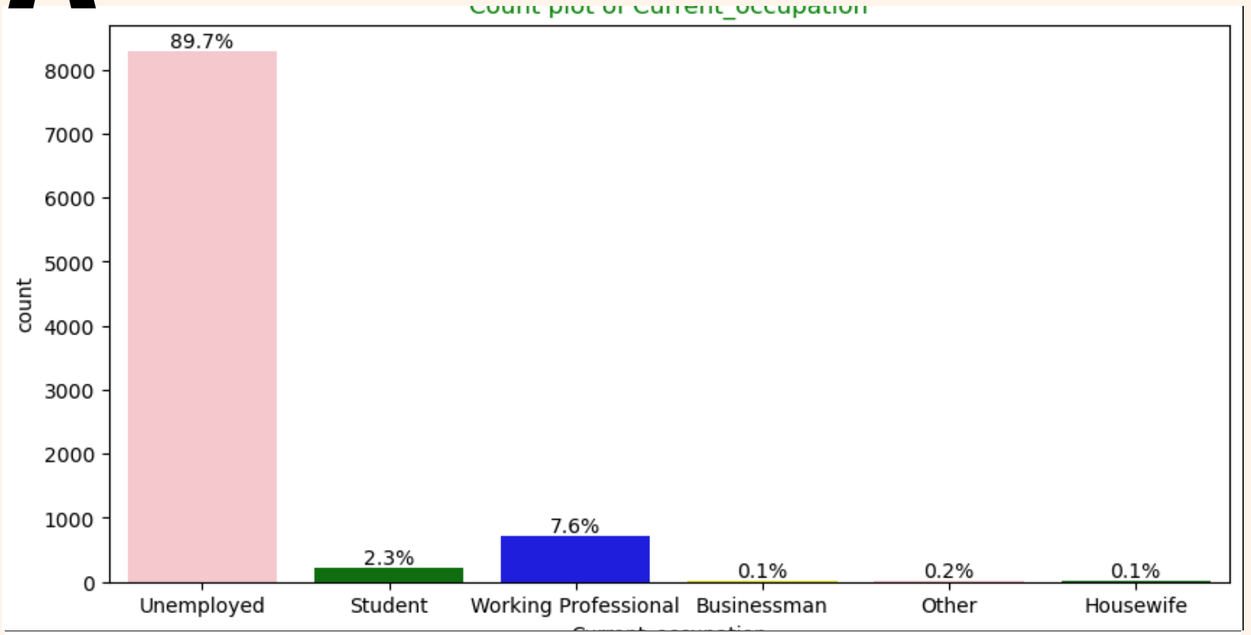
2. Non-converted Leads: 61.5% remain unconverted, revealing process inefficiencies.

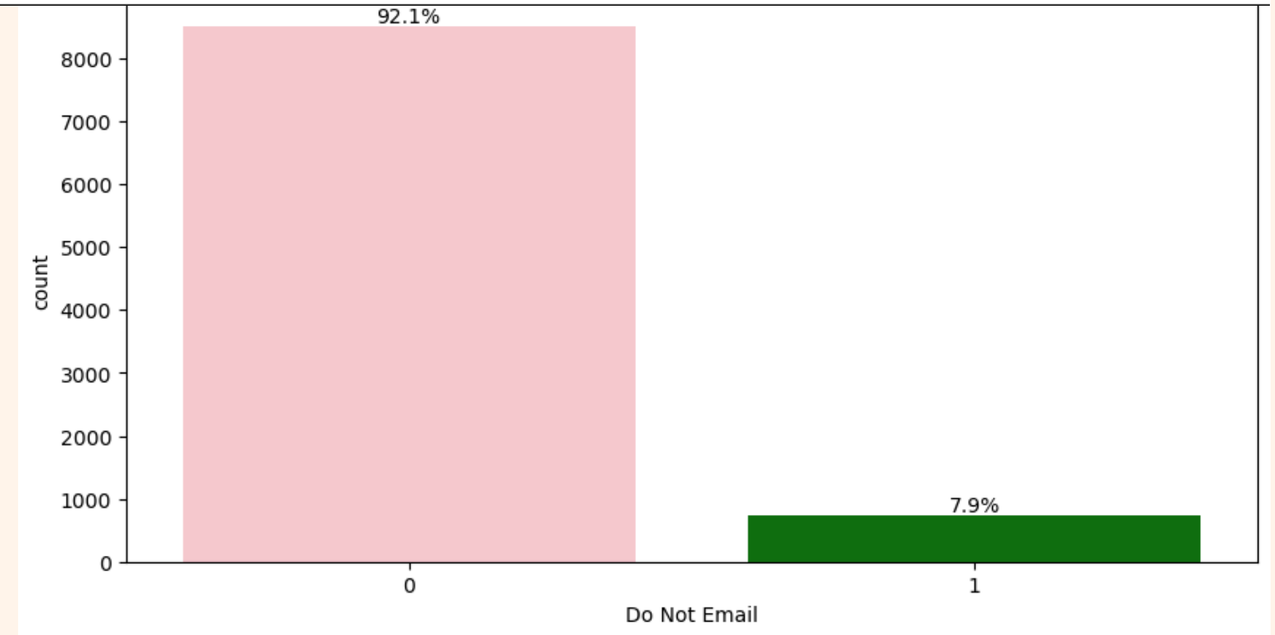
3. Clear Visuals: Color coding and percentage annotations aid intuitive understanding.

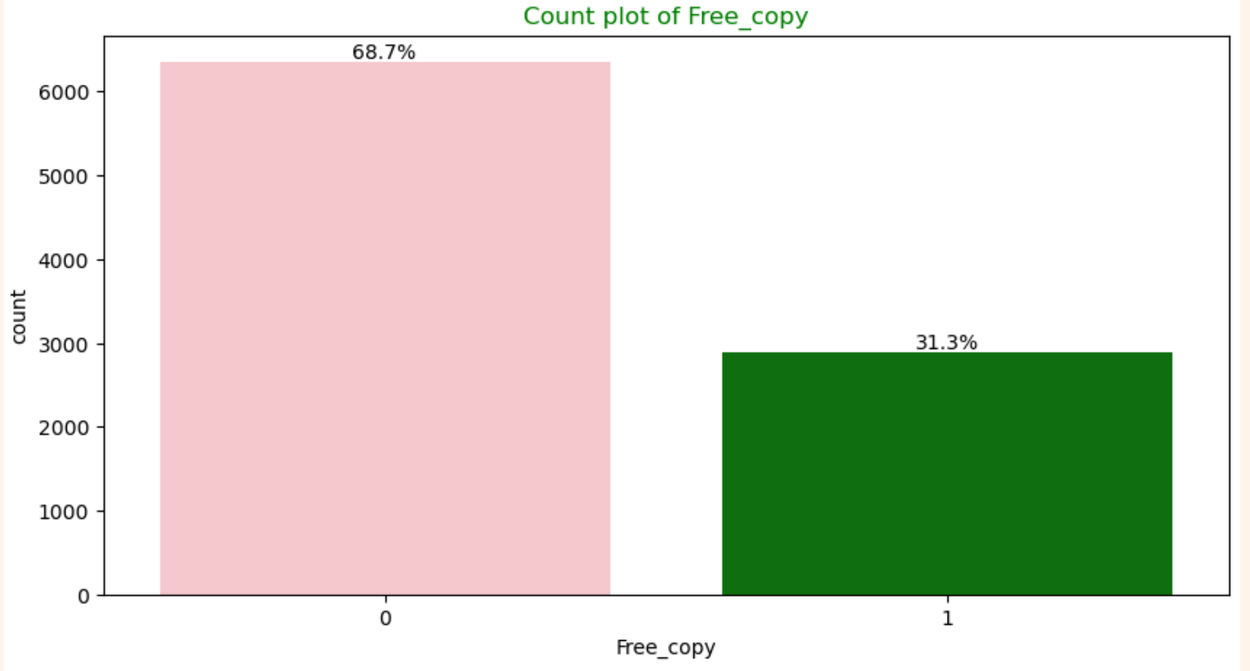
4. Improvement Opportunity: The conversion gap shows potential for focused efforts.

EDA

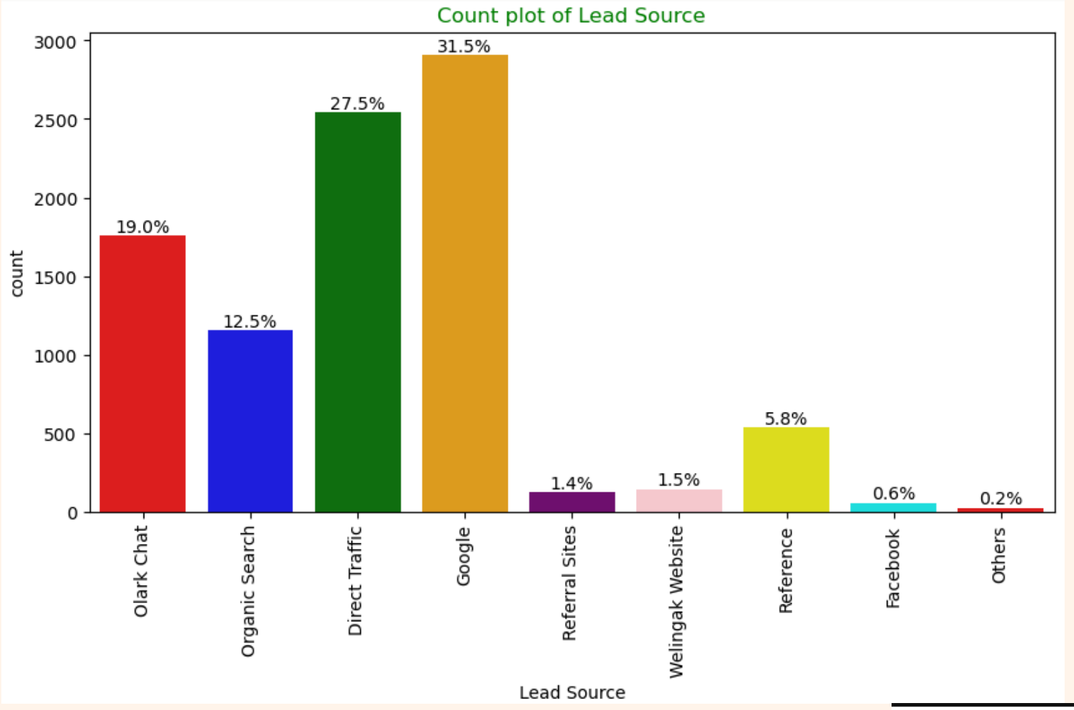

Count plot of Lead Origin
count
Lead Origin
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form
38.7%
52.9%
7.8%
0.6%
0.0%
0
1000
2000
3000
4000
5000

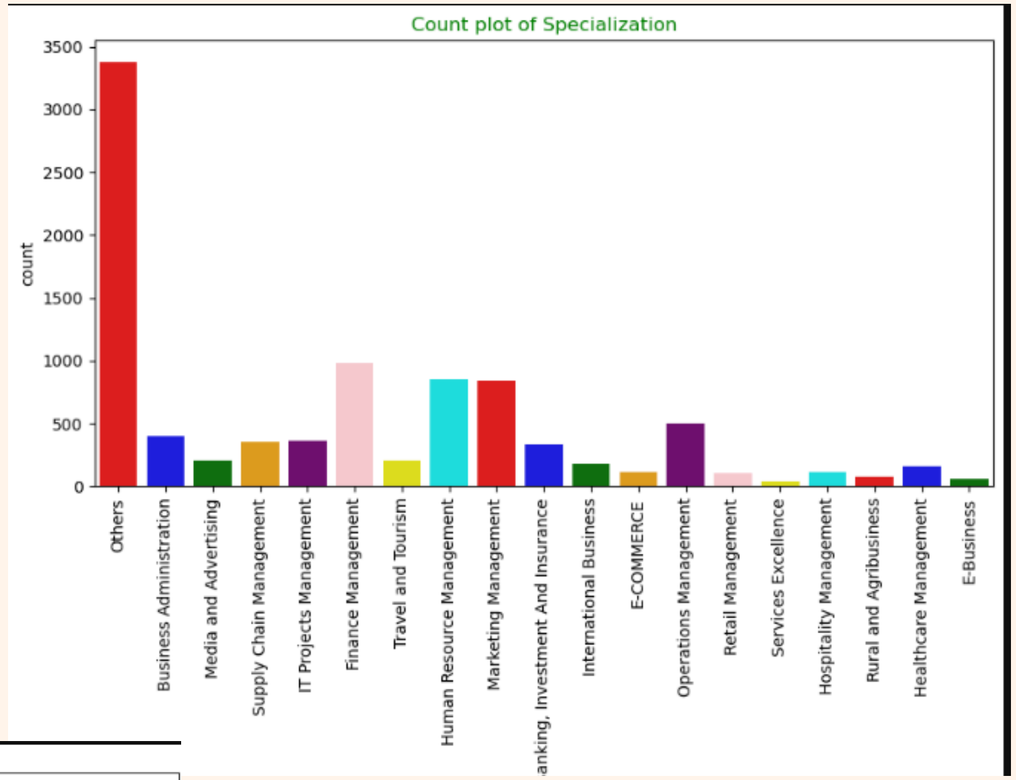

count
Unemployed
Student
Working Professional
Businessman
Other
Housewife
89.7%
2.3%
7.6%
0.1%
0.2%
0.1%
0
1000
2000
3000
4000
5000
6000
7000
8000

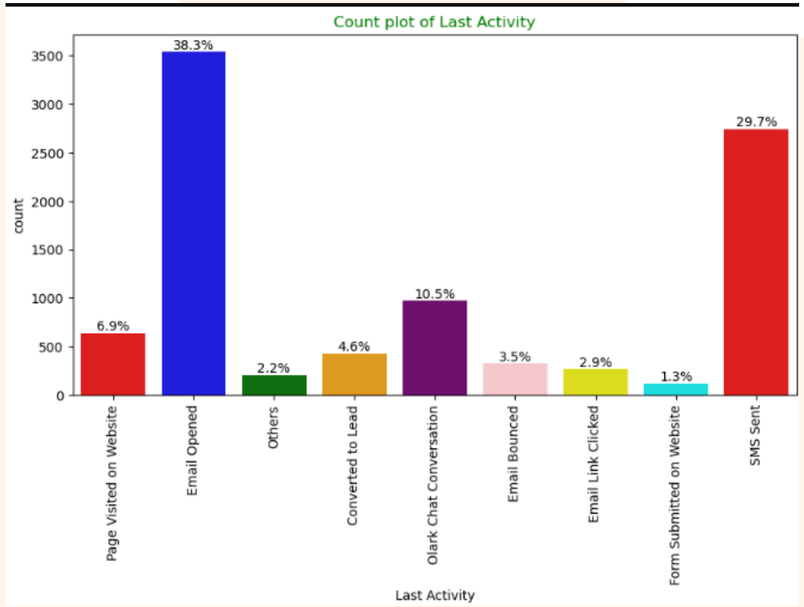

count
Do Not Email
0
1
92.1%
7.9%
0
1000
2000
3000
4000
5000
6000
7000
8000


Count plot of Free_copy
count
Free_copy
0
1
68.7%
31.3%
0
1000
2000
3000
4000
5000
6000

EDA







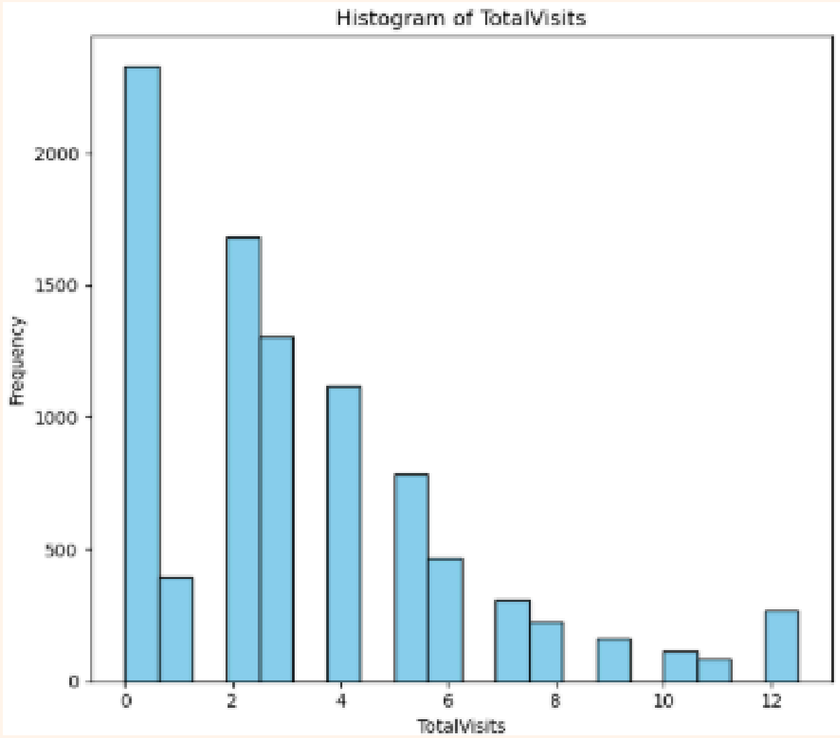
1. Distribution Diversity: Categorical columns show varied distributions, with dominant categories highlighting key trends.

2. Lead Source Insights: 'Lead Source' identifies top acquisition channels, aiding marketing focus.

3. Specialization Skew: 'Specialization' shows a strong skew, with some specializations dominating, while others are less represented.

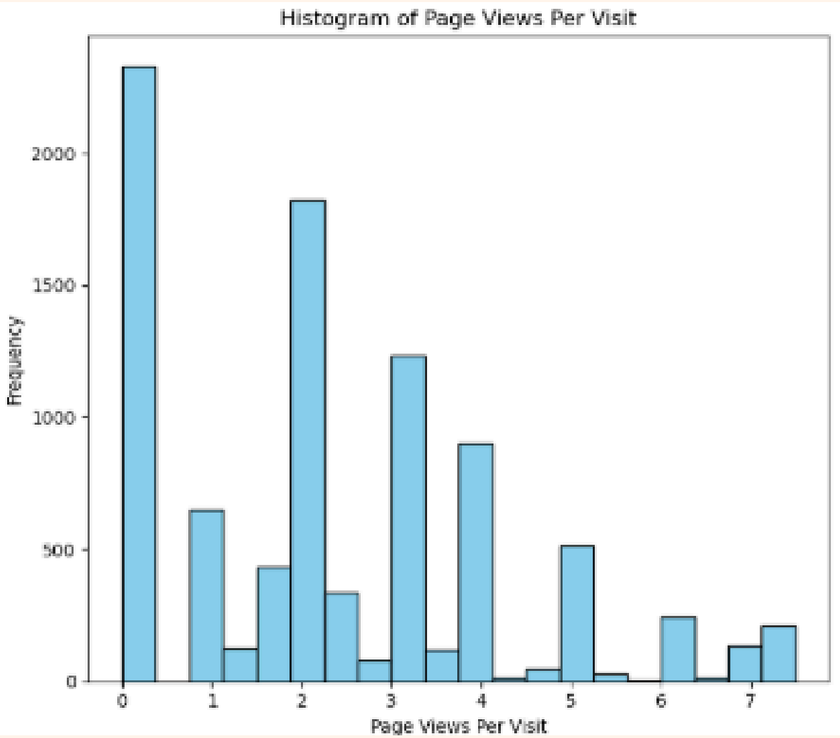
4. Clear Visuals: Custom colors and percentage annotations enhance clarity and interpretation.

EDA



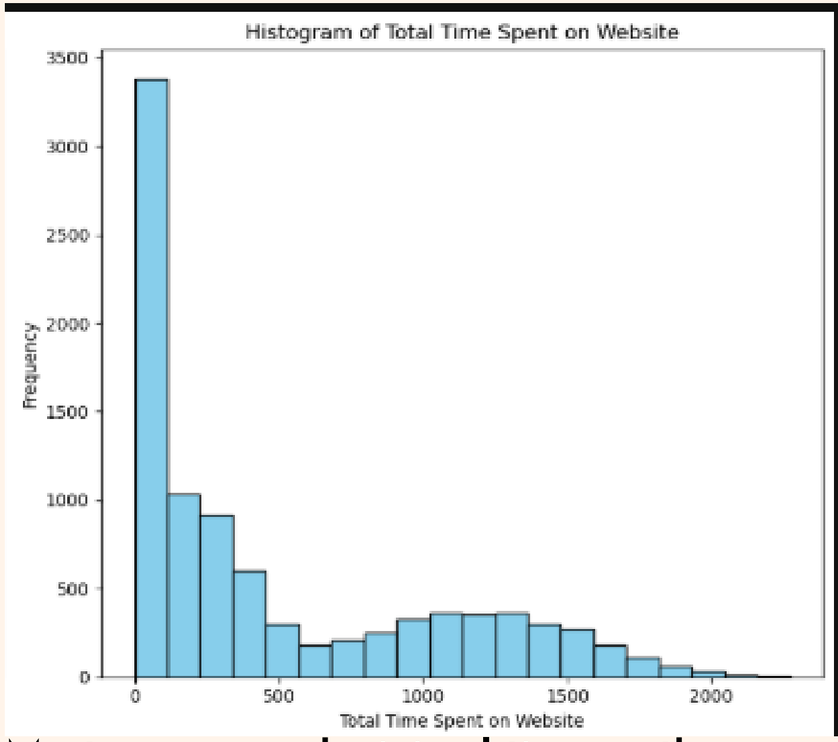
Most users have very few visits, with a significant spike at 0 or 1 visit.

The frequency decreases rapidly as the number of visits increases, indicating only a few users revisit the website frequently.



A majority of users view only 1 to 2 pages per visit, suggesting limited engagement with the website content.

There's a small subset of users who view multiple pages, indicating potential interest or intent.

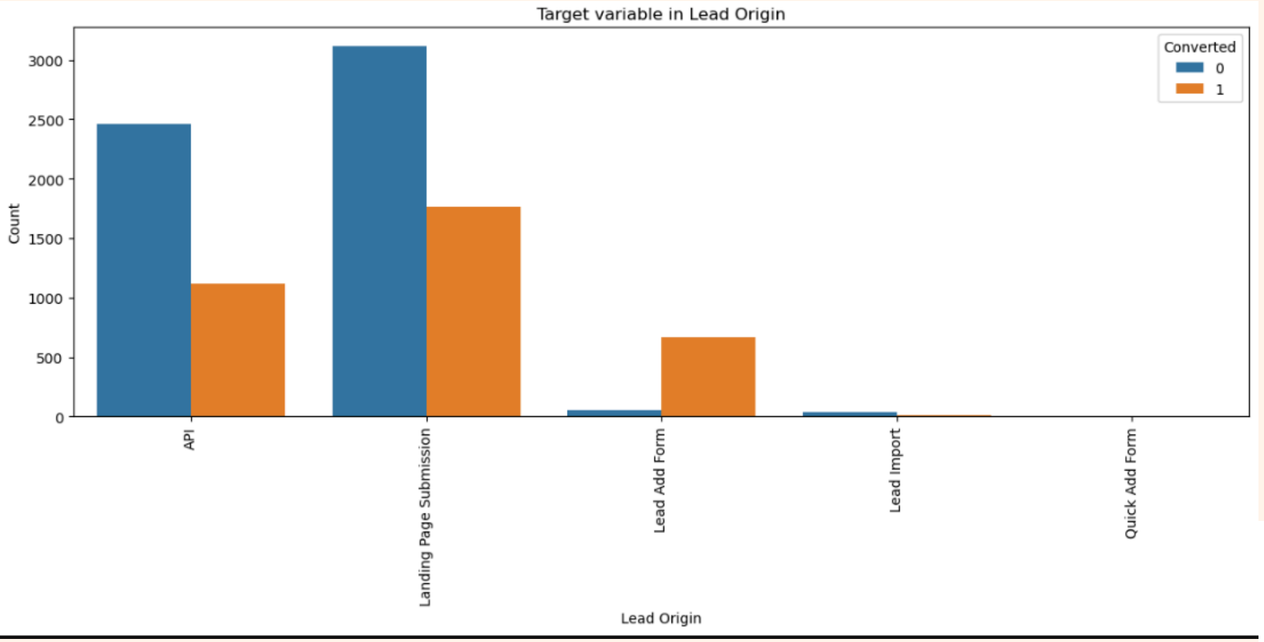


Most users spend minimal time on the website (0-100 seconds), reflecting quick exits or disinterest.

A long tail distribution suggests some users spend significantly more time, likely indicating higher engagement or exploration.

EDA

BIVARIATE ANALYSIS

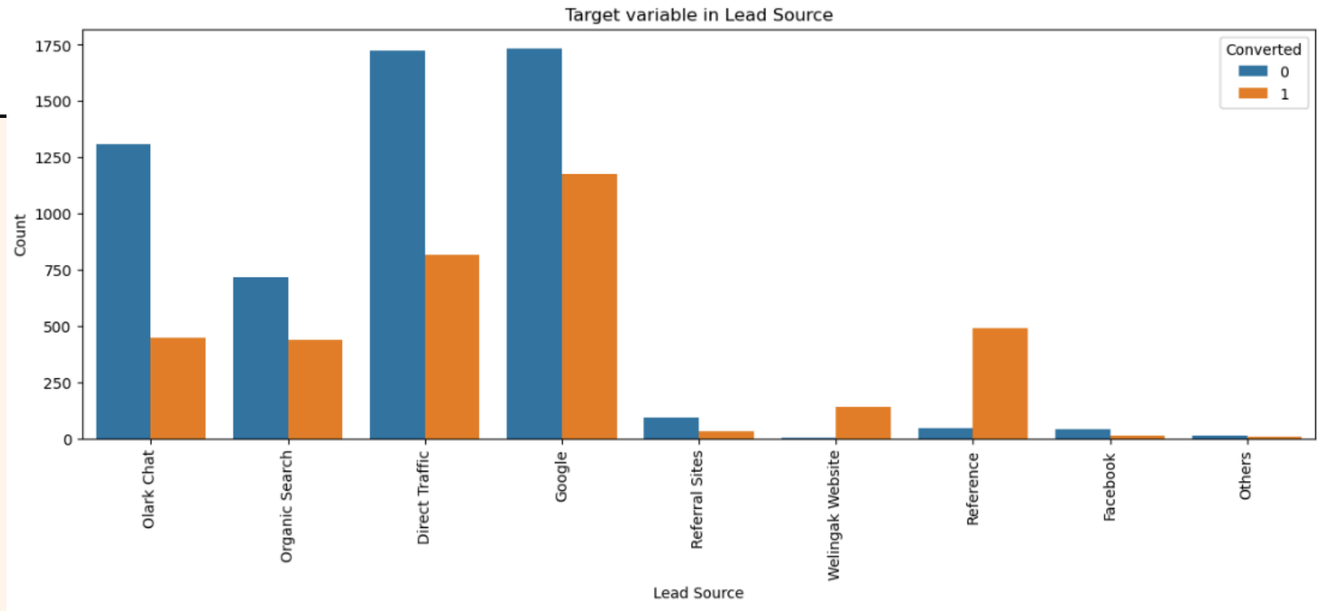


Effectiveness of Landing Page Submissions: "Landing Page Submission" is the top lead origin, generating high leads and strong conversions, proving its effectiveness.

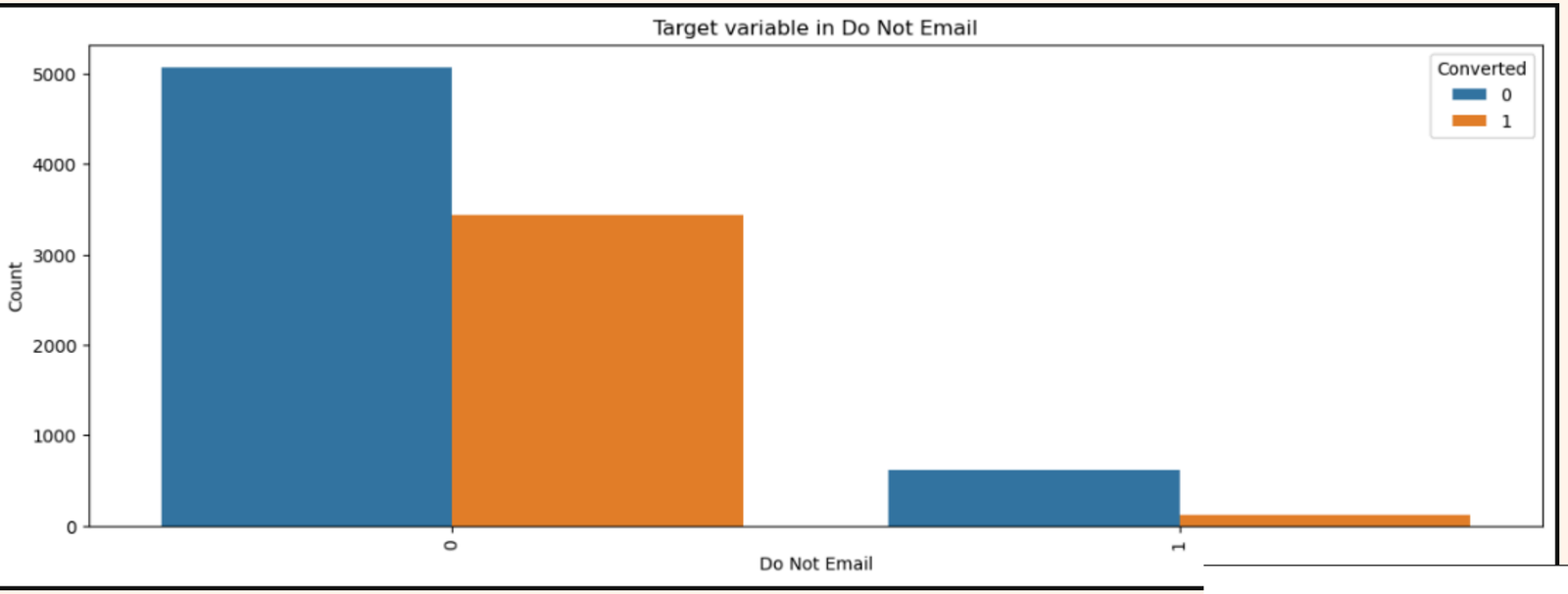
Lead Add Form Efficiency: Fewer leads but higher conversion rates make it an efficient channel.

Google's Impact: Google drives significant leads and conversions, dominating the funnel.

Efficient References: References have fewer leads but a higher conversion rate, proving highly effective.



EDA

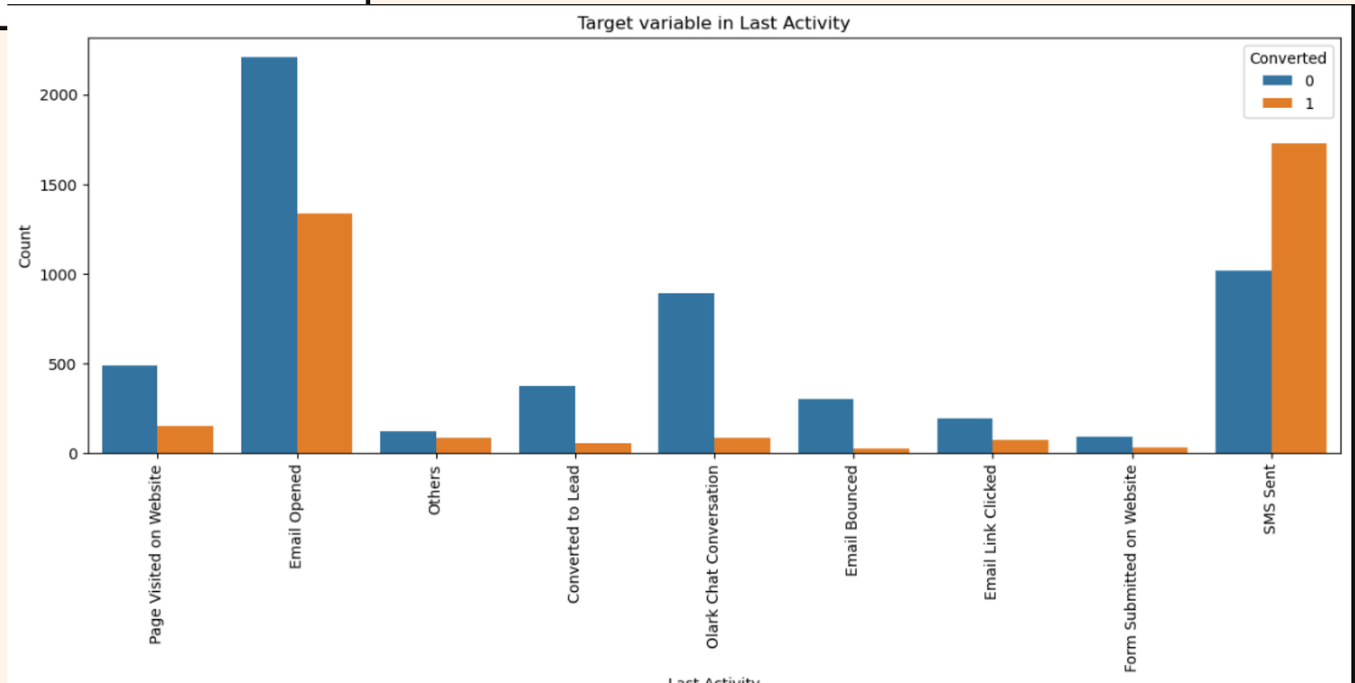


Email Preference: Most individuals in “Do Not Email = 0” don't convert, showing low effectiveness.

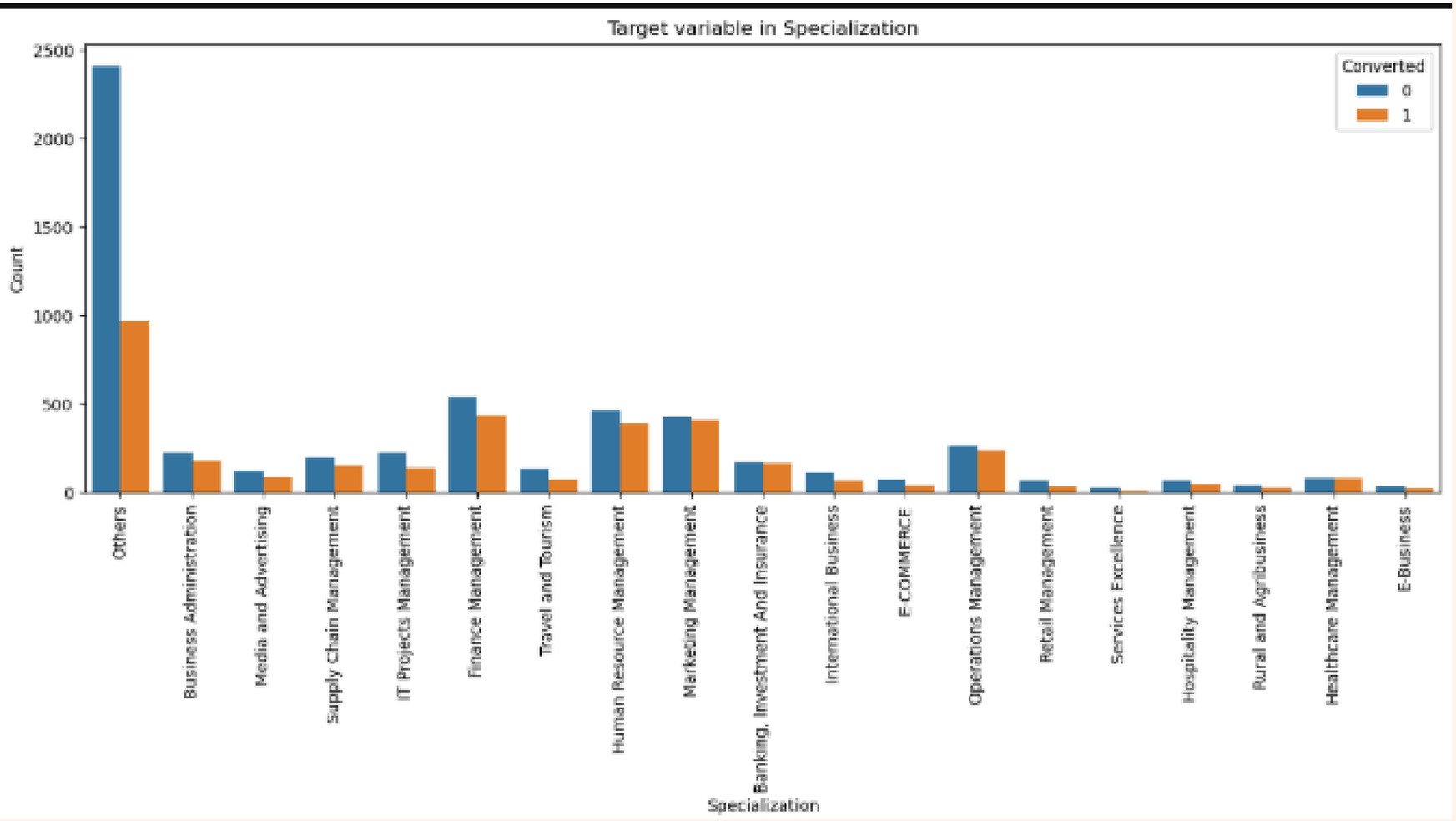
Limited Impact: Few leads opt for “Do Not Email = 1,” making it less influential on conversions.

Email Opened: Highest activity count, but most leads don't convert, showing low correlation.

SMS Sent: Higher conversions, proving it more effective than other activities.



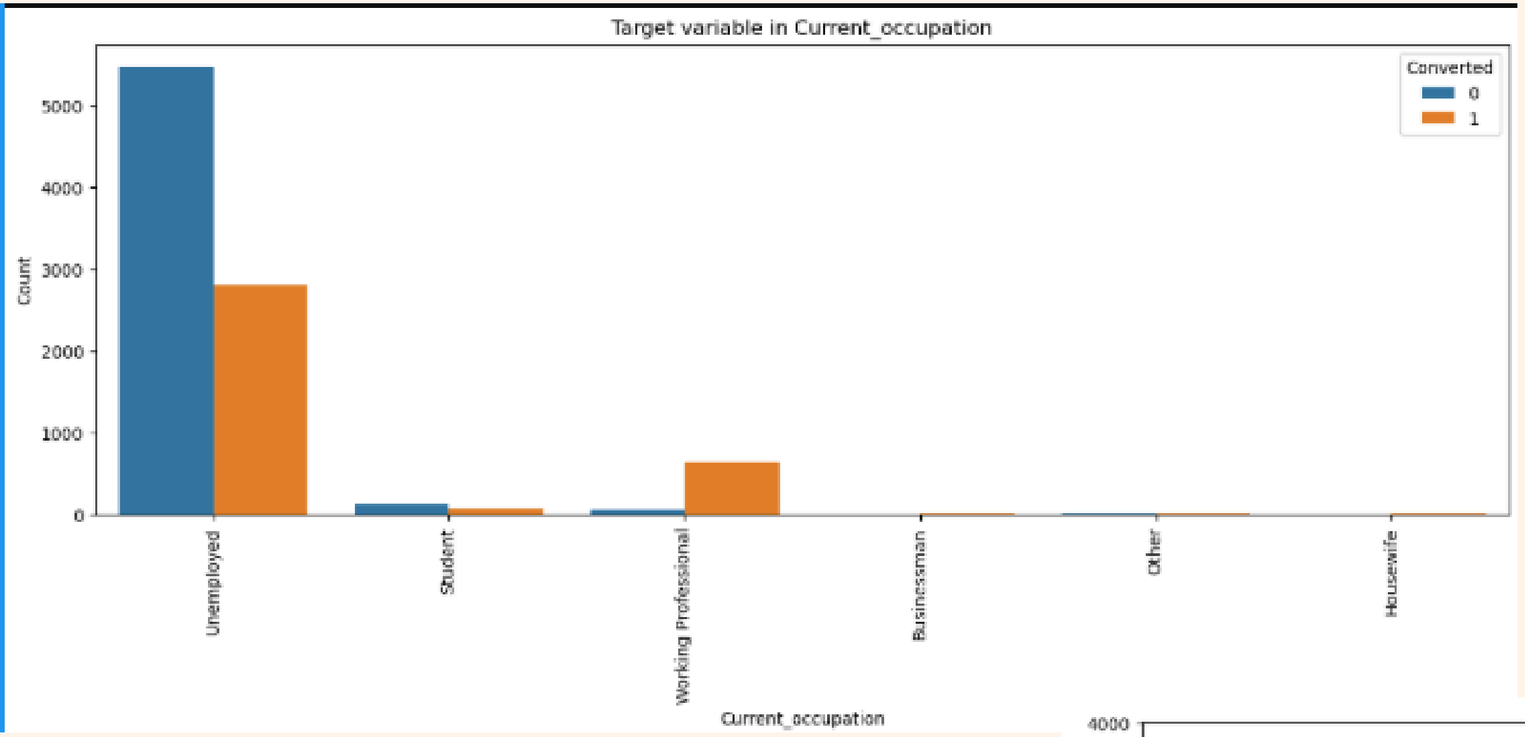
EDA



"Others" Specialization: Broad representation with the highest count, offering key trend insights.

Balanced Conversions: Specializations like Finance, HR, and Marketing show stable conversion rates.

EDA

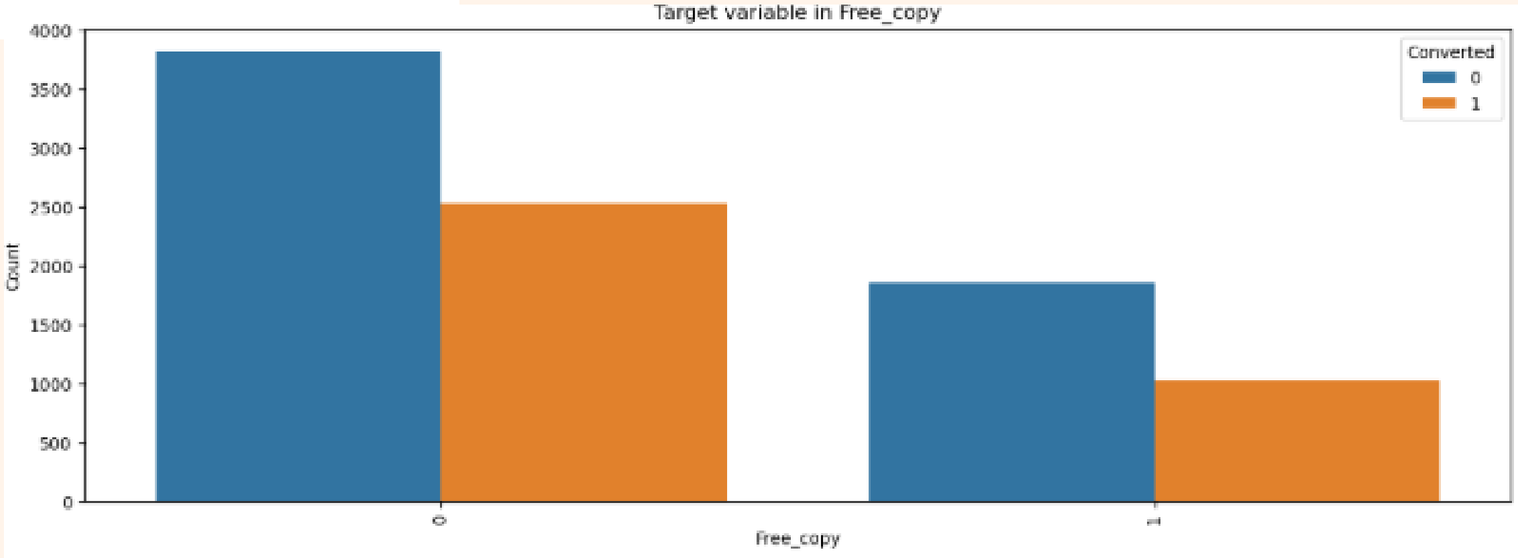


Unemployed Dominance: Unemployed individuals form the majority but show low conversion rates.

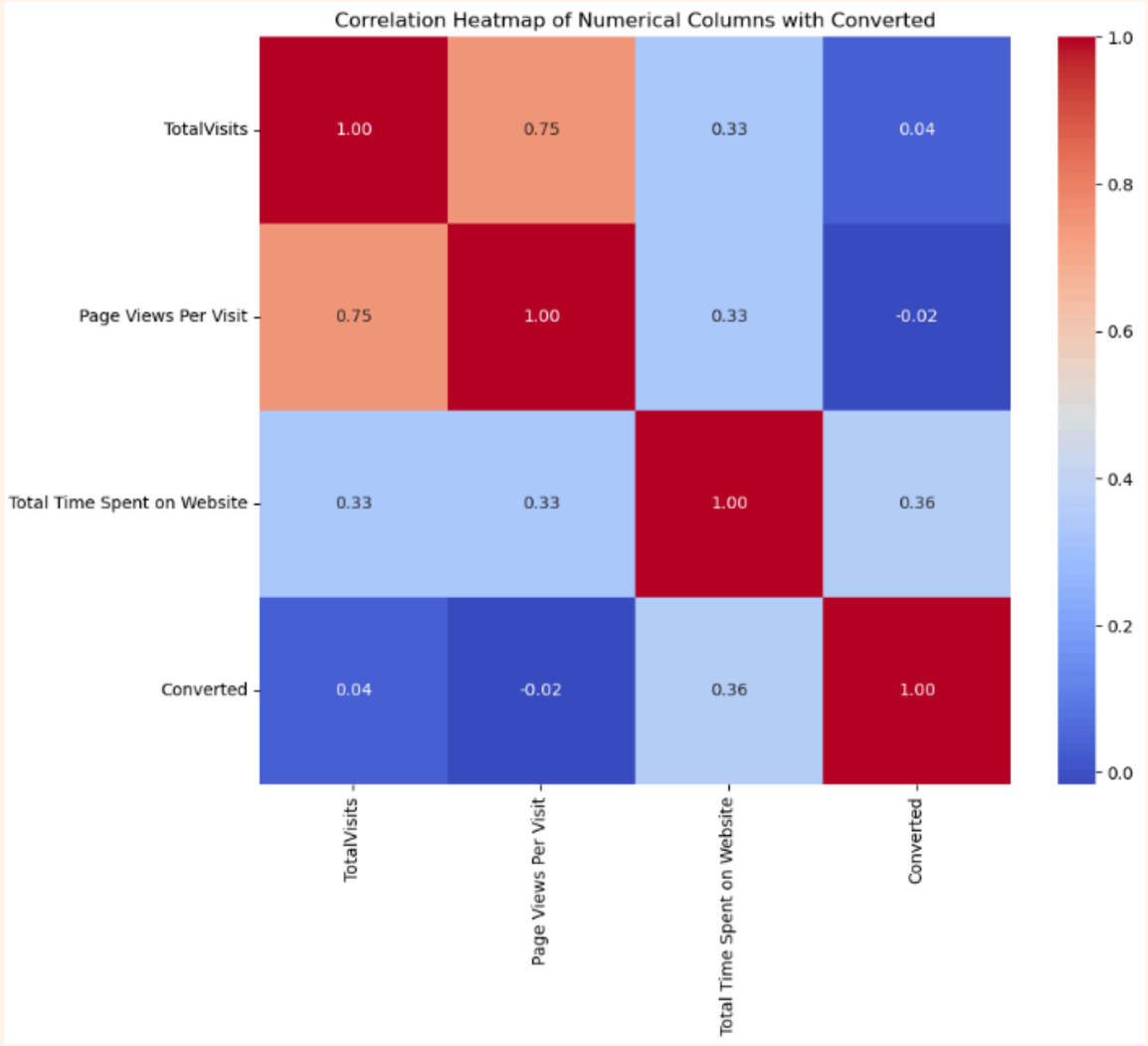
Working Professionals Convert: Higher conversions in professionals highlight better engagement compared to students or housewives.

No Free Copy: Higher non-conversions indicate its absence lowers conversion rates.

Free Copy Provided: Boosts conversions, showing a positive influence despite non-conversions dominating.



EDA

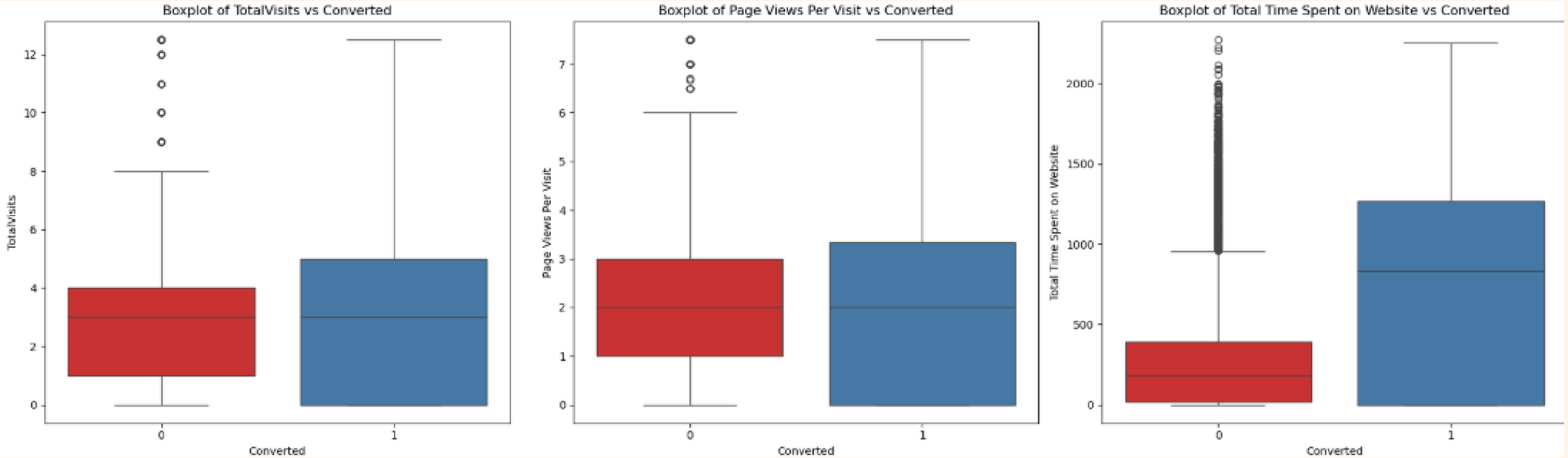


Time Spent and Conversion: Moderate positive correlation (0.36) shows higher time on-site increases conversion likelihood.

Visits and Page Views: Strong positive correlation (0.75) indicates more visits lead to higher engagement.

Page Views and Conversion: Very weak negative correlation (-0.02) suggests minimal impact on conversion.

EDA



'Total Visits': Median for converted and unconverted leads are same.

‘**'Page Views Per Visit':** Median for converted and unconverted leads are same.

'Total Time Spent on Website': Leads spending more time on the website are more likely to be converted. Therefore, a engaging Website might help in making more leads.

DATA CONVERSION

Numerical Variables are Normalised

Dummy Variables are created for object type variables

Total Rows for Analysis: 9240

Total Columns for Analysis: 49

MODEL BUILDING

Splitting the Data into Training and Testing Sets

The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.

Use RFE for Feature Selection

Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5

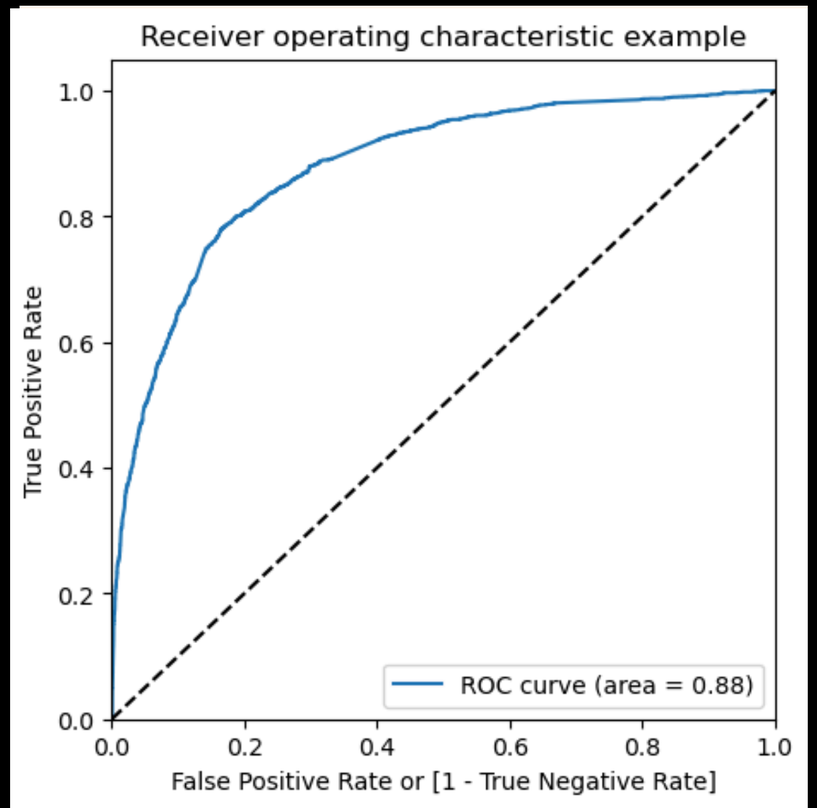
Predictions on test data set
&
Overall accuracy 80.47%

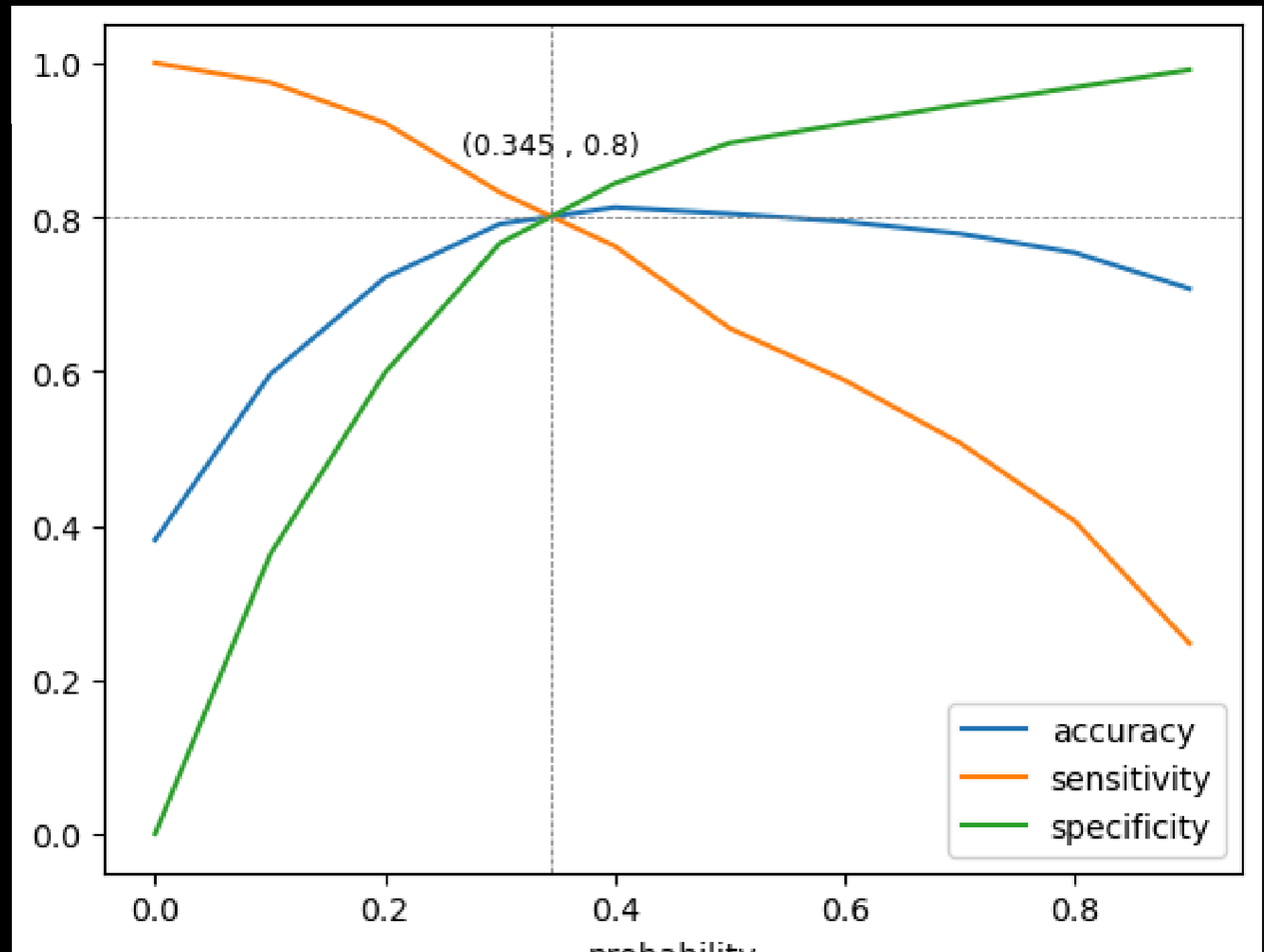
ROC CURVE

Model Performance: AUC of 0.88 indicates excellent discriminative power.

Optimal Threshold: At 0.345, accuracy, sensitivity, and specificity reach ~0.8 for balanced outcomes.

Trade-off: High sensitivity and specificity (~0.8) minimize false negatives and positives effectively.





CONCLUSION

Train Data Set

- Accuracy: 80.47%
- Sensitivity: 80.13%
- Specificity: 80.68%

Test Data Set

- Accuracy: 80.16%
- Sensitivity: 79.82% ≈ 80%
- Specificity: 80.38%

Notes

- The model achieved a sensitivity of 80.13% in the train set and 79.82% in the test set, using a cut-off value of 0.345.
- Sensitivity in this case indicates how many leads the model identify correctly out of all potential leads which are converting.
- The CEO of X Education had set a target sensitivity of around 80%.
- The model also achieved an accuracy of 80.47%, which is in line with the study's objectives.

RECOMMENDATIONS

To increase our Lead Conversion Rates:

- Focus on features with positive coefficients to refine targeted marketing strategies.
- Devise plans to attract high-quality leads from top-performing sources.
- Engage working professionals with personalized messaging.
- Optimize communication channels based on their impact on lead engagement.
- Allocate additional budget to the Welingak Website for advertising and promotions.
- Offer incentives or discounts for references that convert, encouraging more referrals.
- Aggressively target working professionals, given their higher conversion rates and better financial capacity to afford higher fees.

To identify areas of improvement:

- Analyze negative coefficients in specialization offerings.
- Review landing page submission process for areas of improvement.


THANK
YOU