

LEAD SCORING CASE STUDY

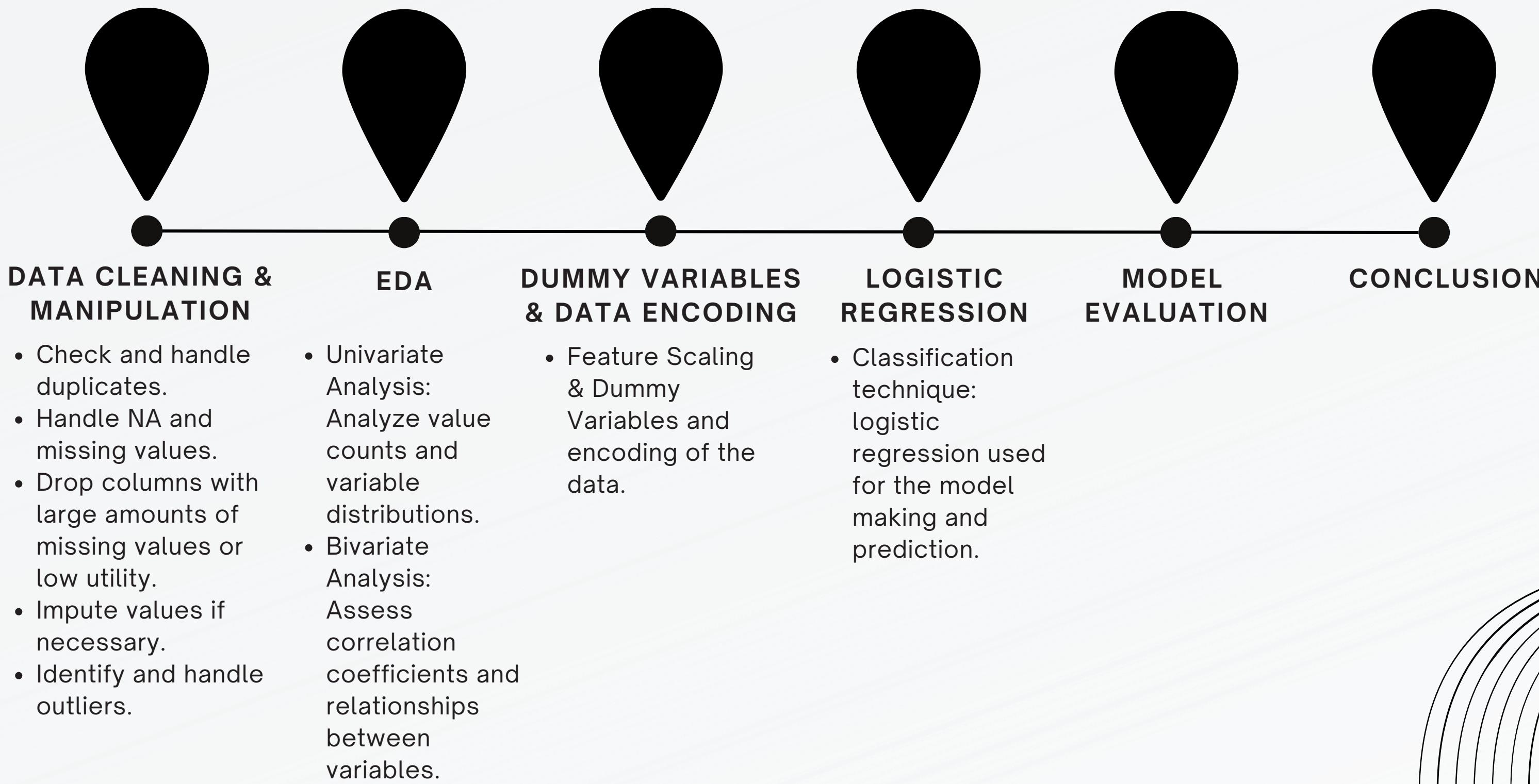
X Education

P. JAYA VINEELA, PALAK PALIWAL & PARIMAL MANOJ

PROBLEM STATEMENT

- **Company** - X Education sells online courses to industry professionals.
- **Challenge** - Low lead conversion rate (30%) despite acquiring many leads daily.
- **Goal** - Identify 'Hot Leads' to improve conversion rates to around 80%.
- **Data** - A dataset of approximately 9,000 leads with various attributes, including a target variable 'Converted' (1 for converted, 0 for not).
- **Objective** - Build a logistic regression model to assign lead scores, focusing on key variables and categorical factors for better targeting.

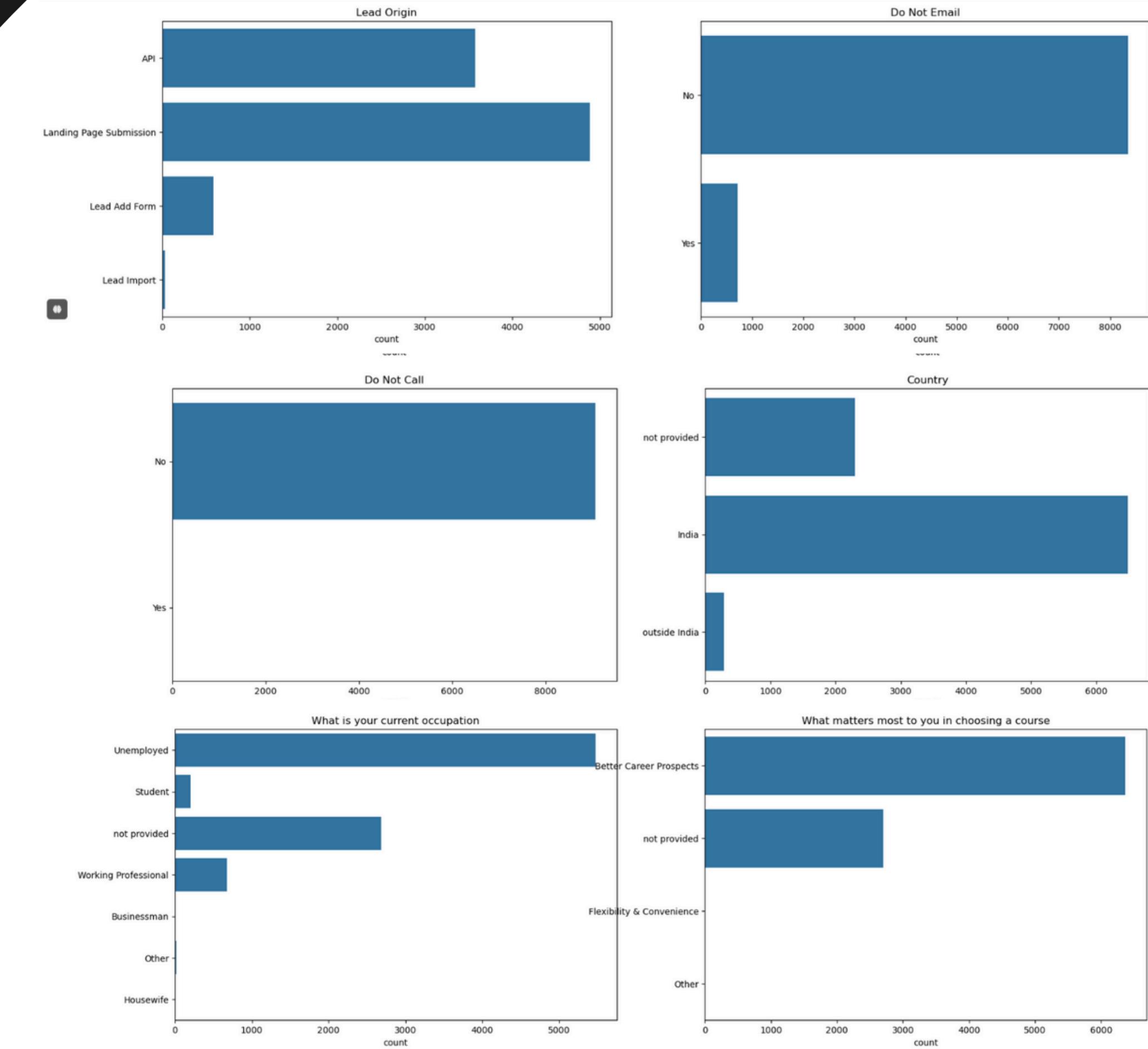
SOLUTION METHODOLOGY



EDA

EXPLORATORY DATA ANALYSIS

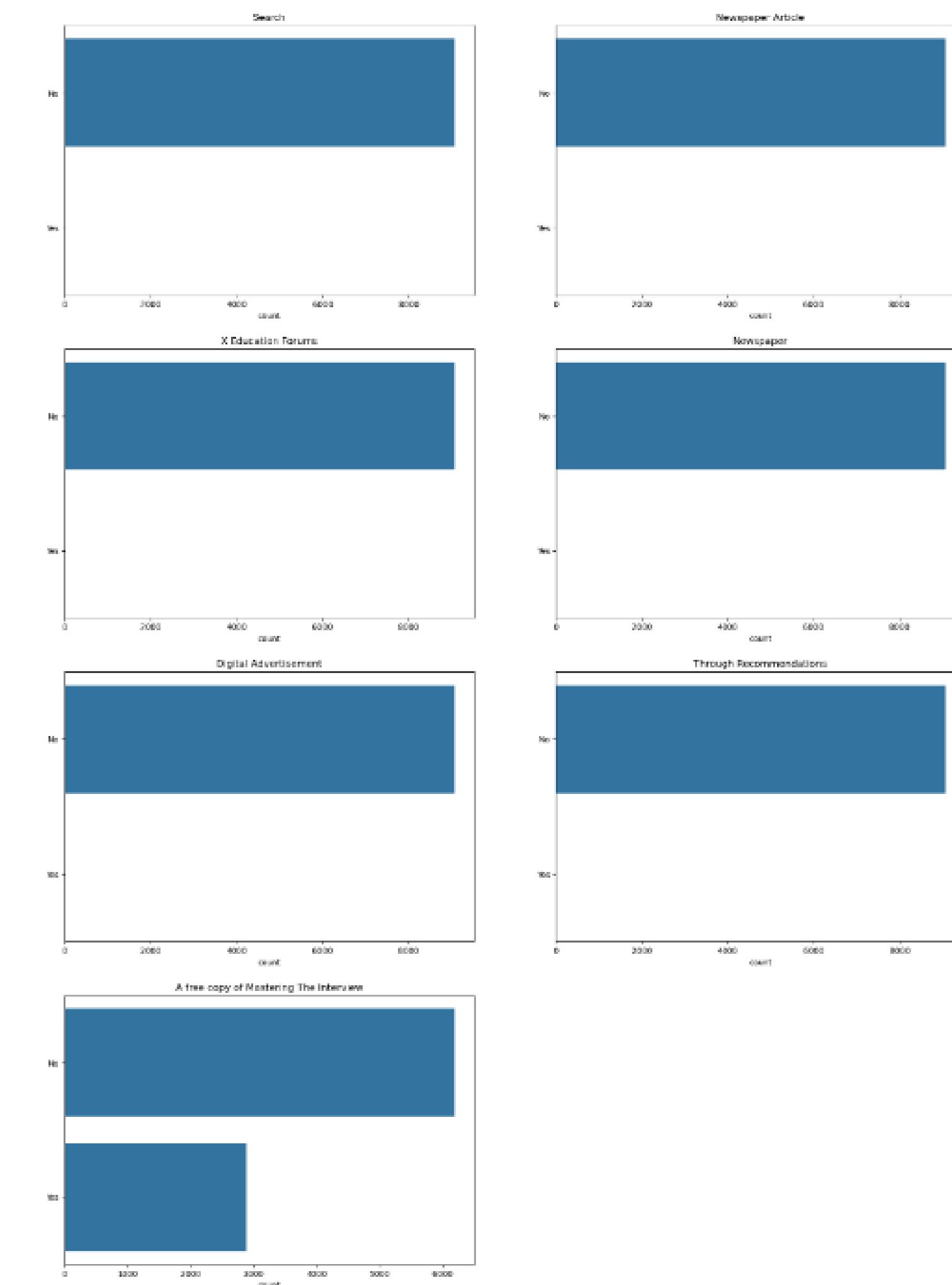
- UNIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

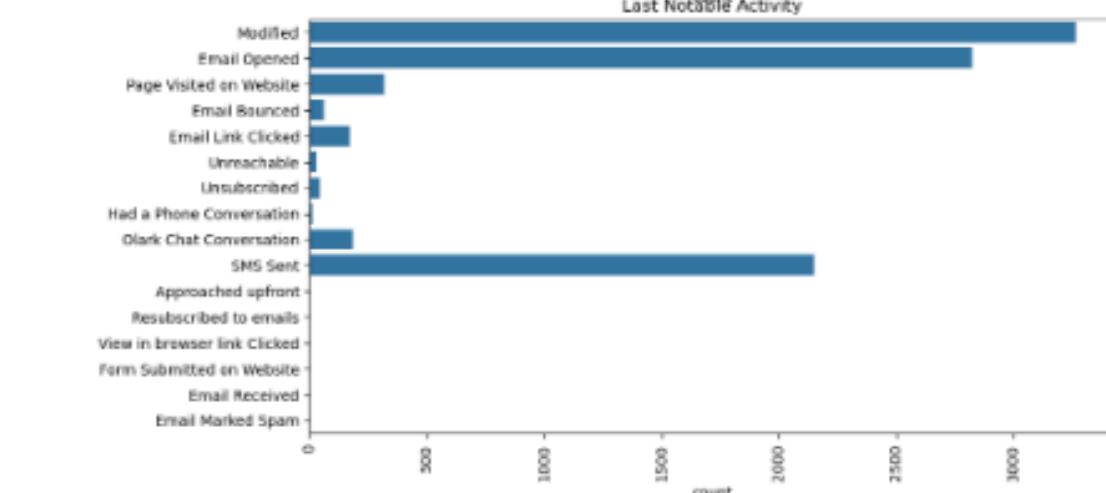
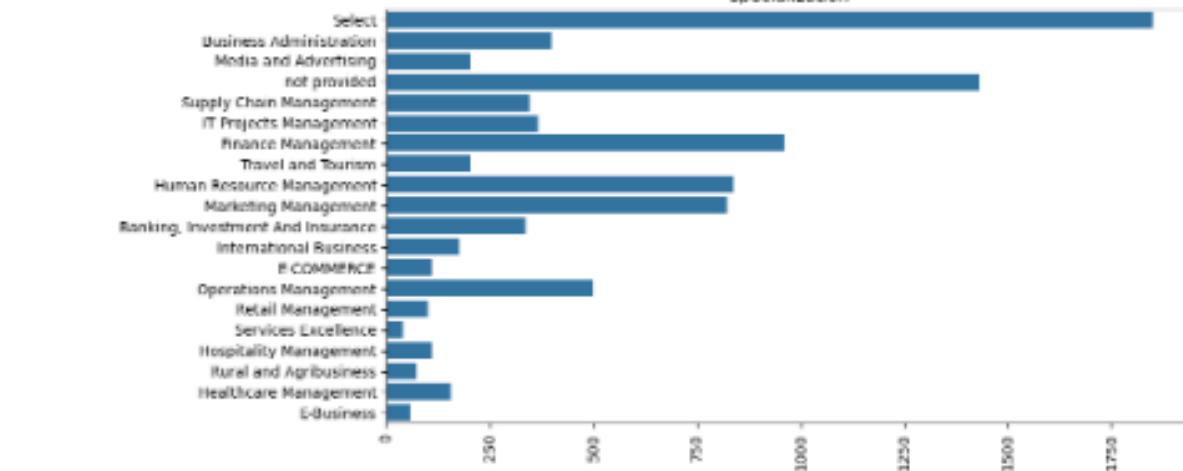
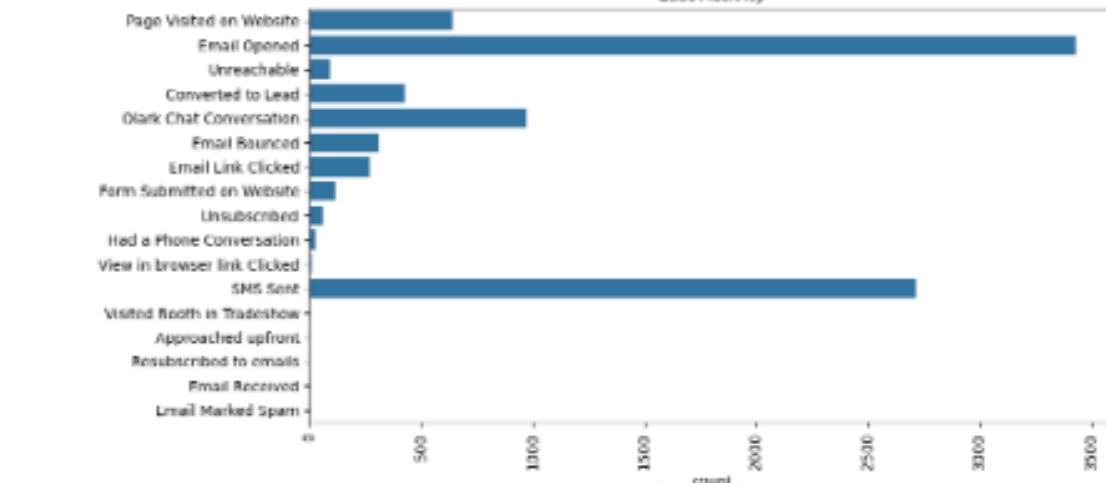
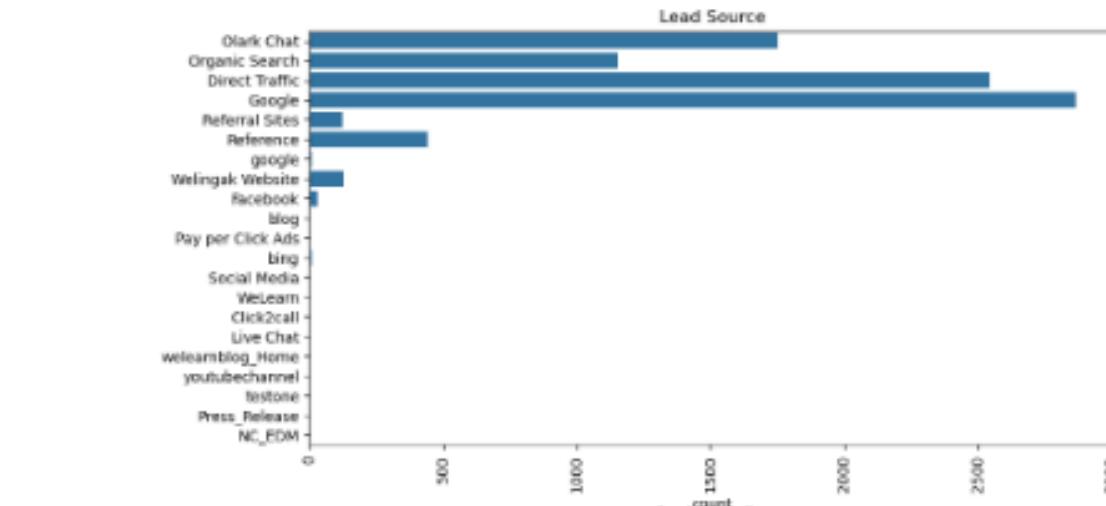
- UNIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

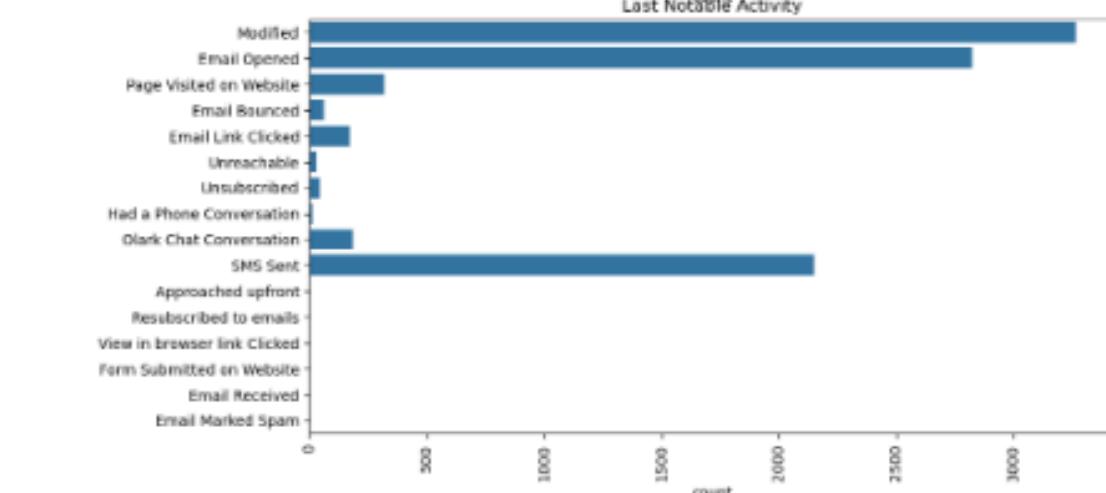
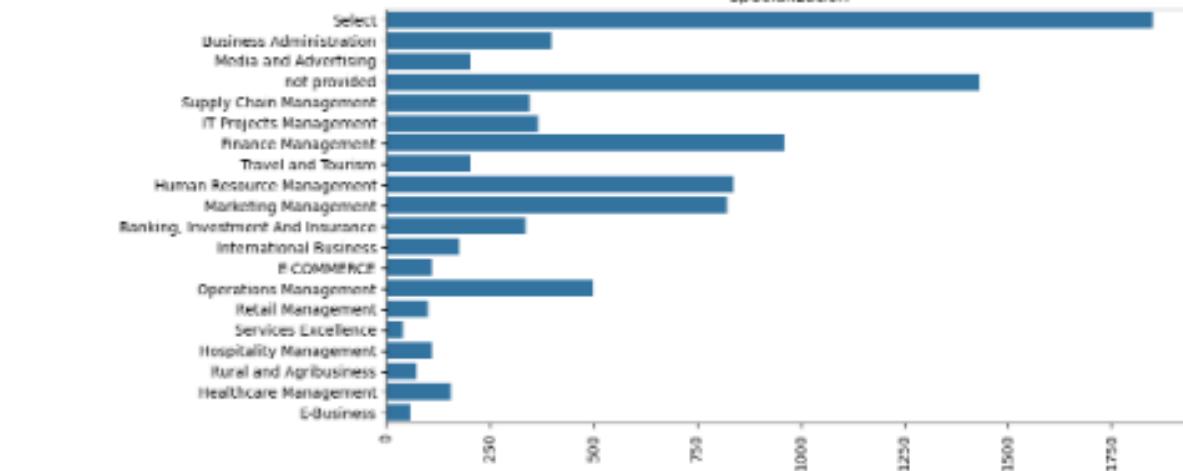
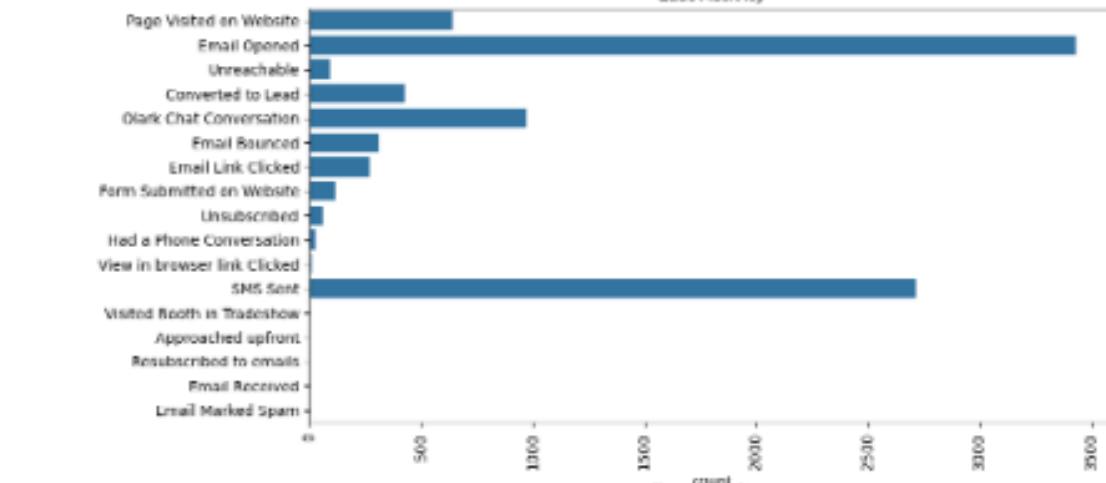
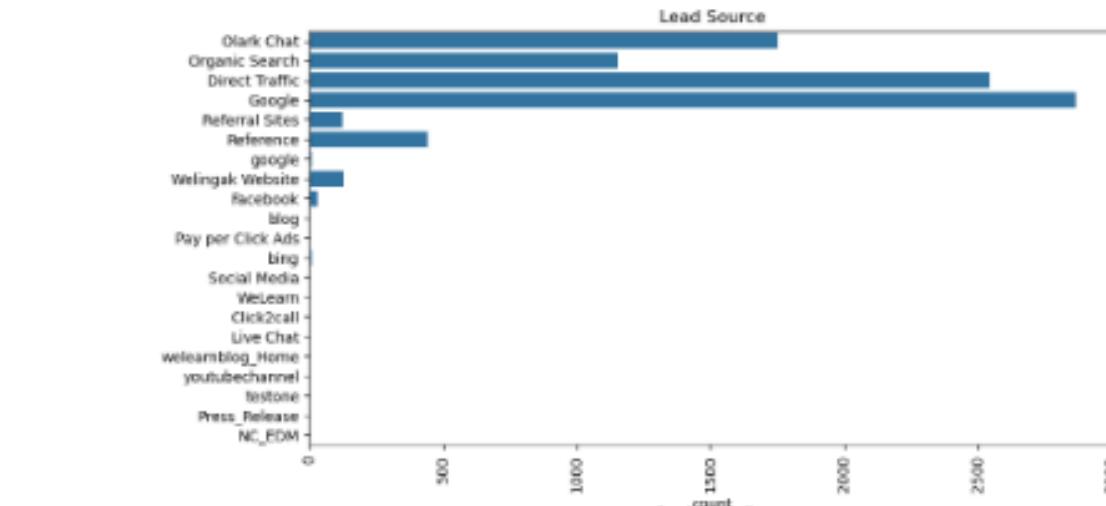
- UNIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

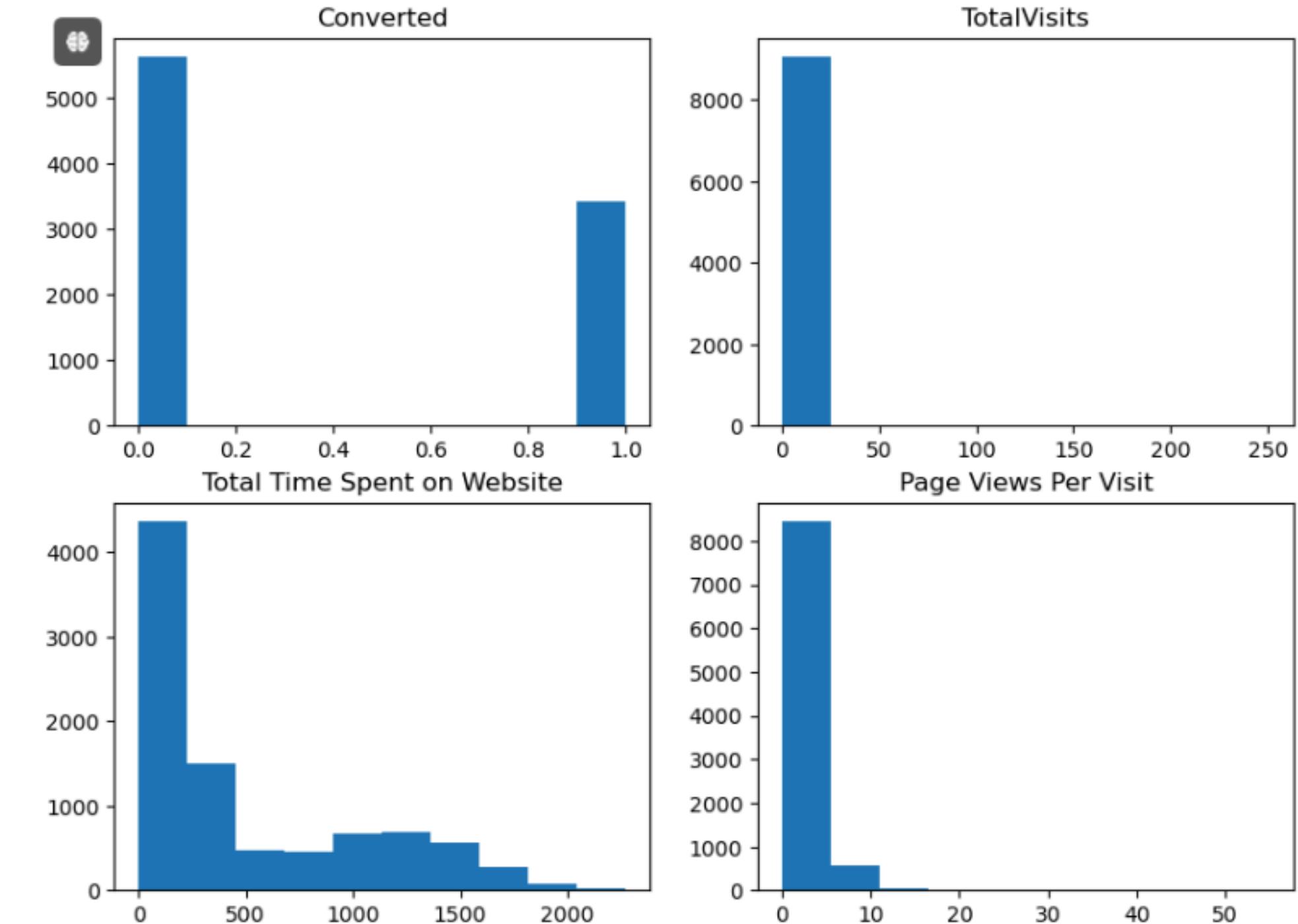
- UNIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

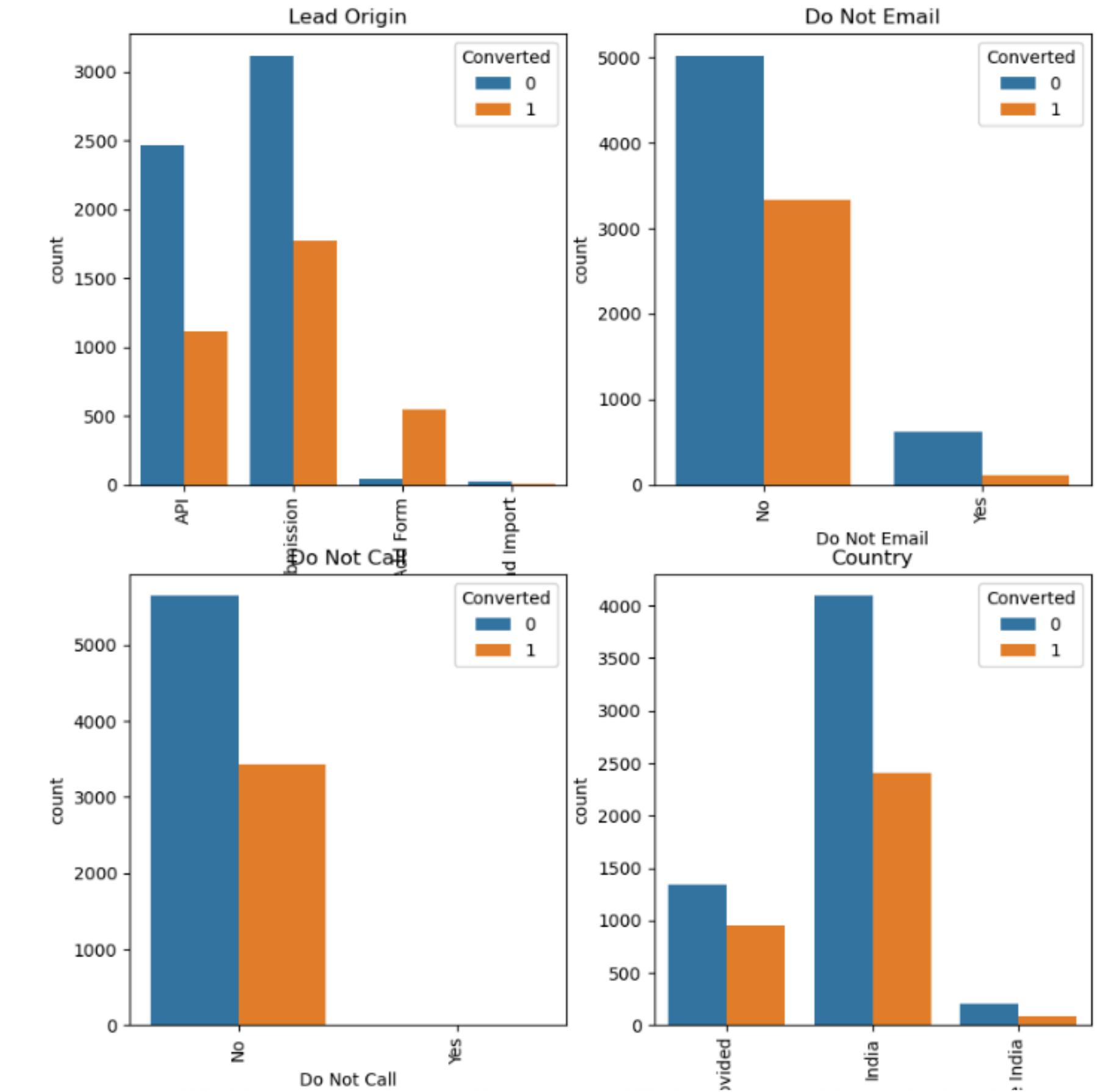
- UNIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

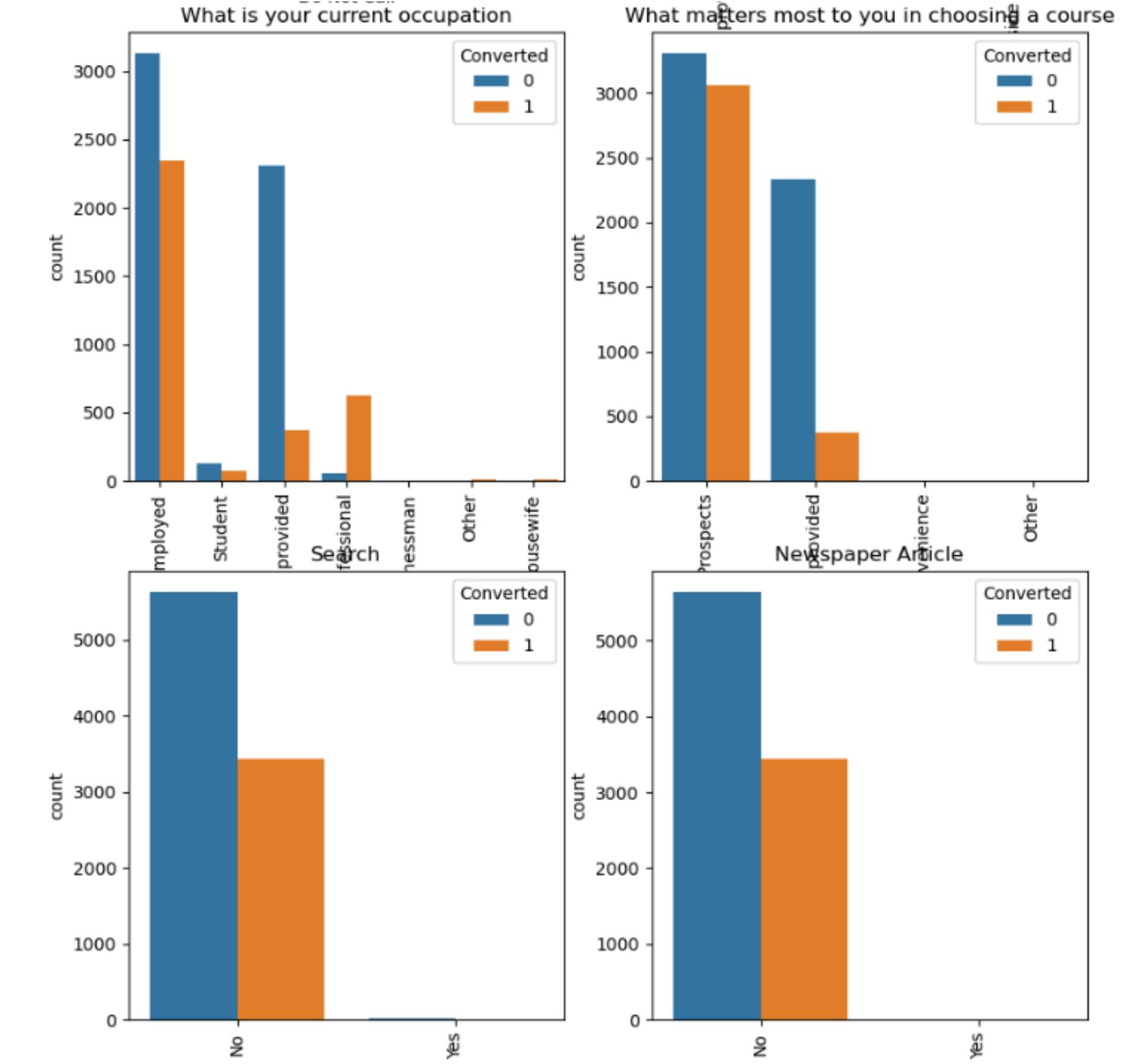
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

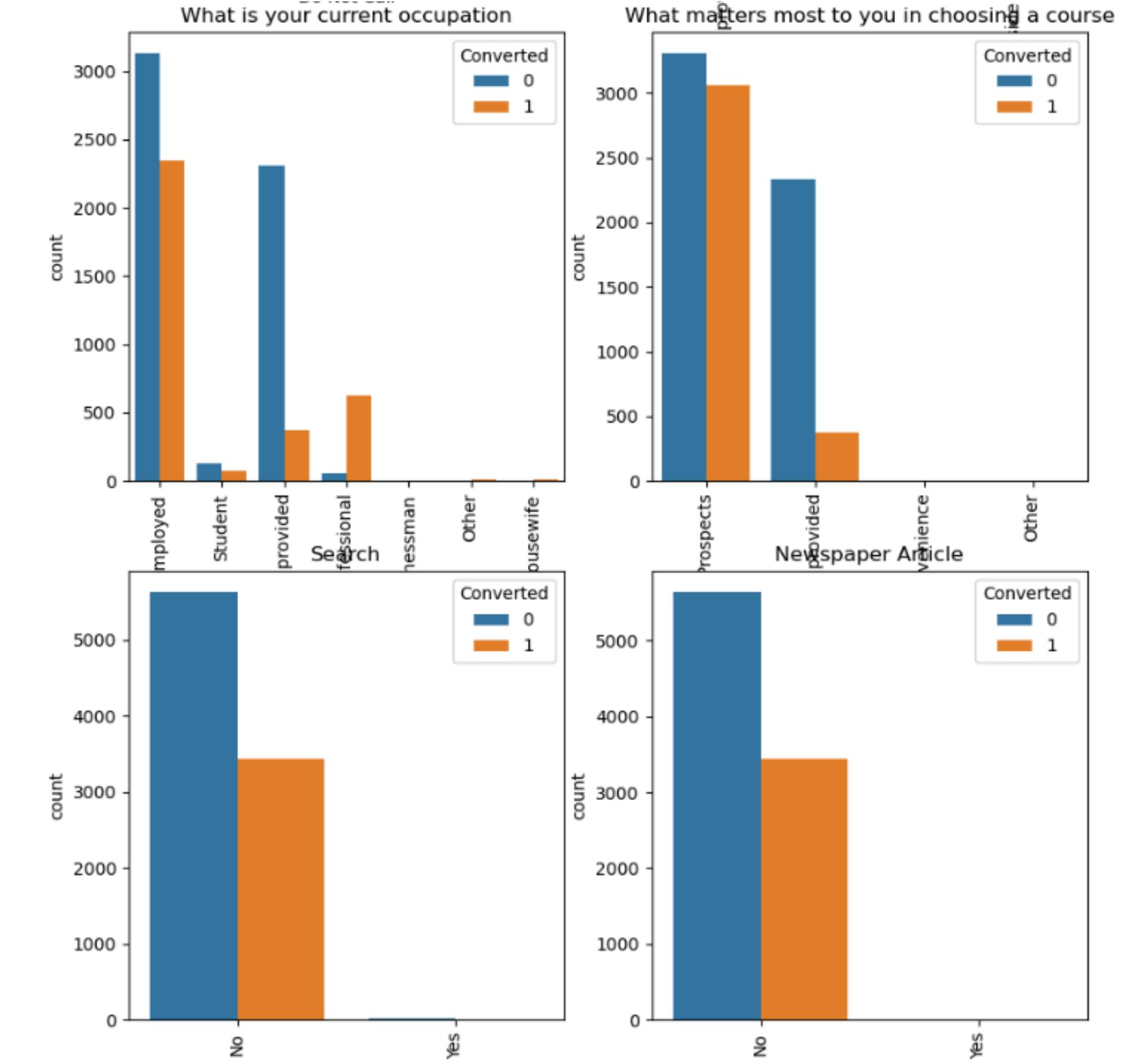
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

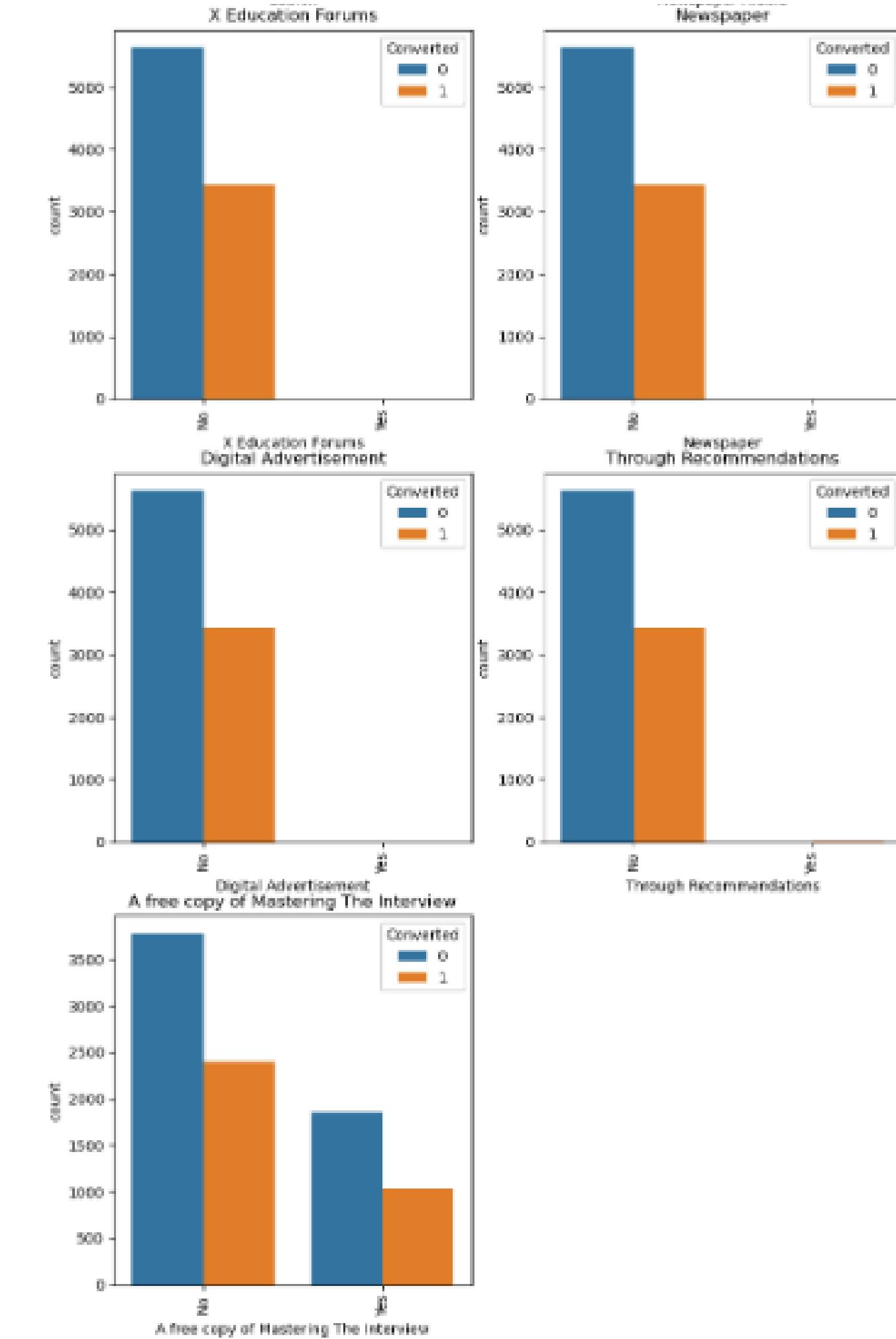
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

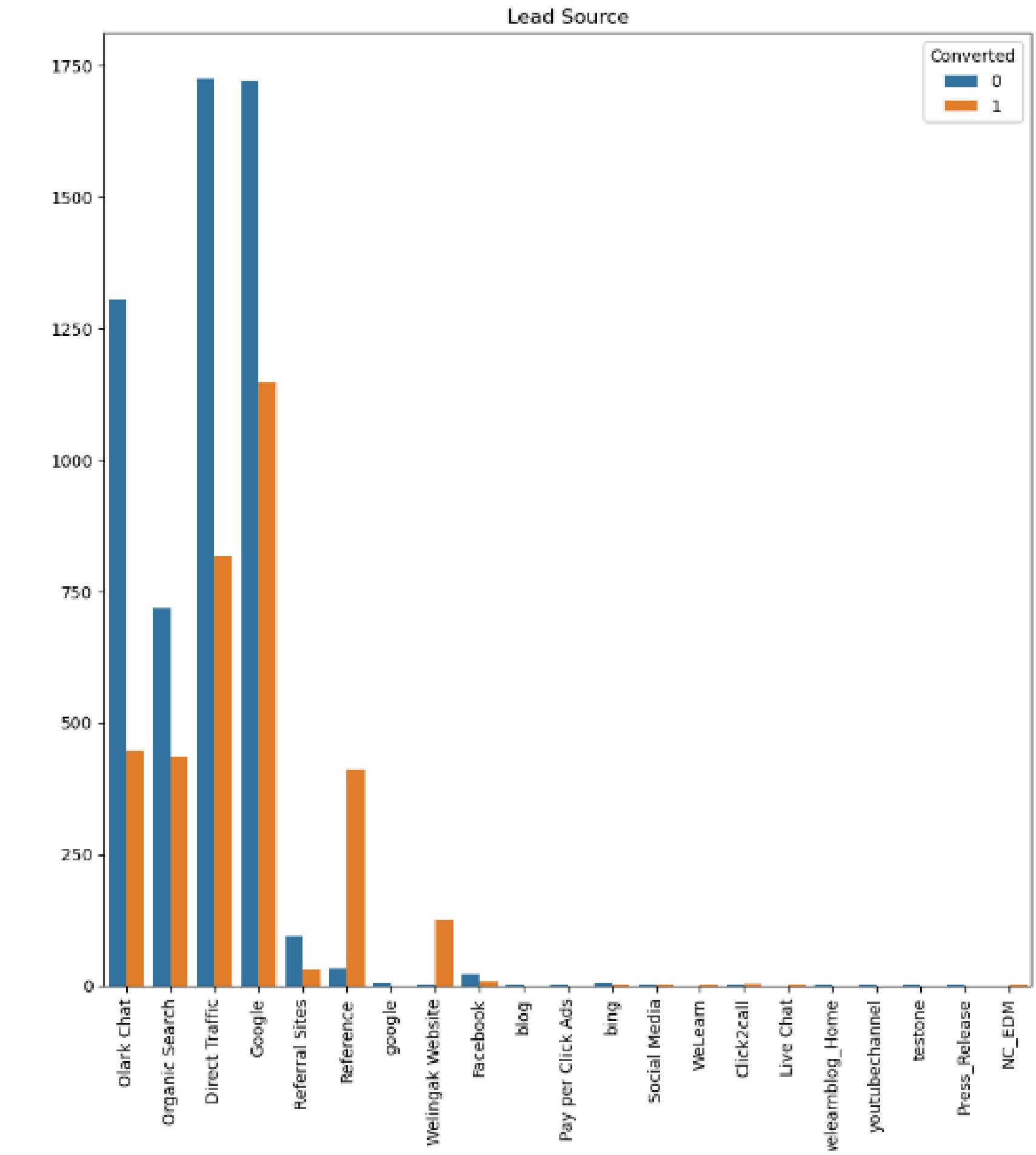
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

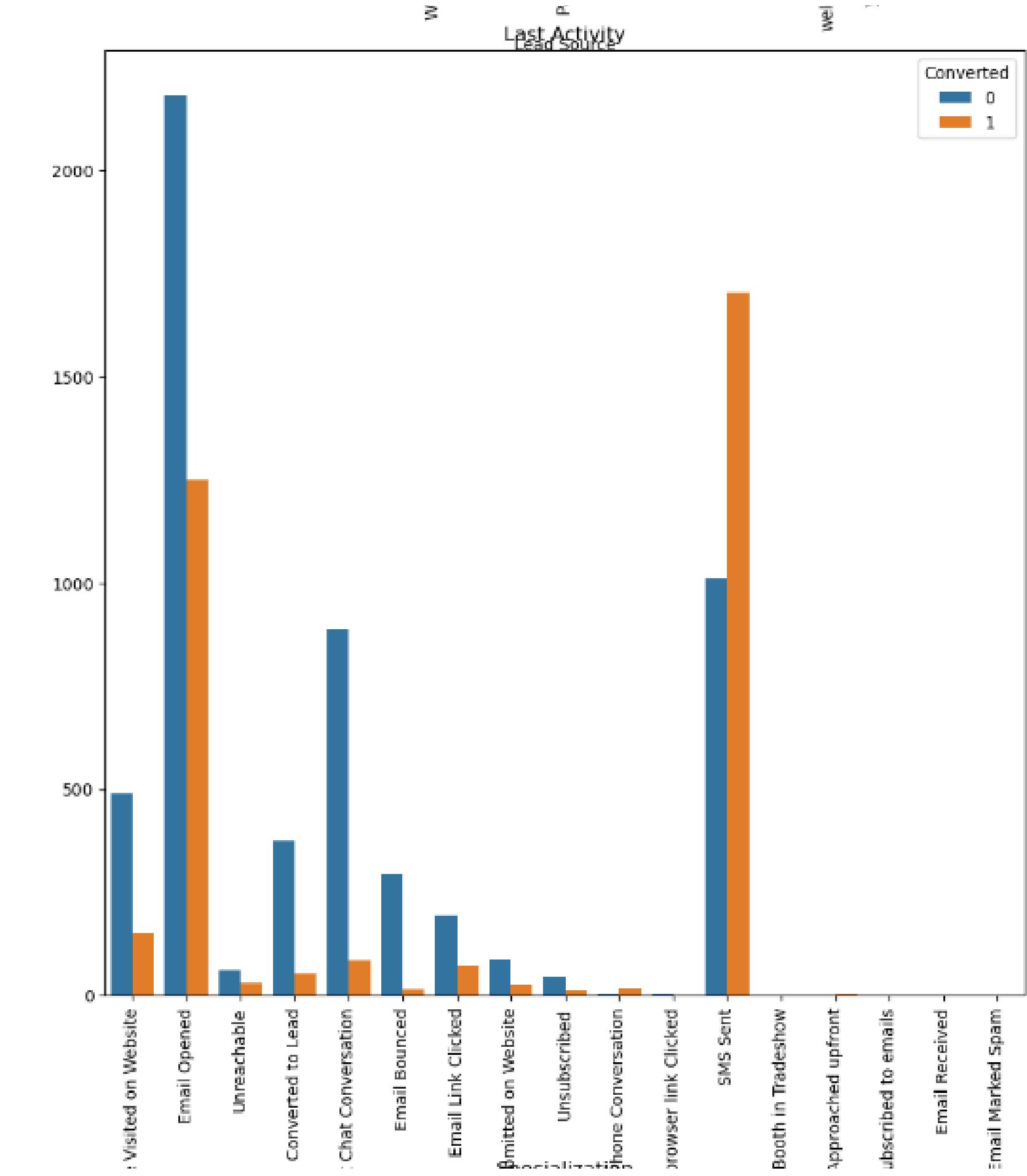
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

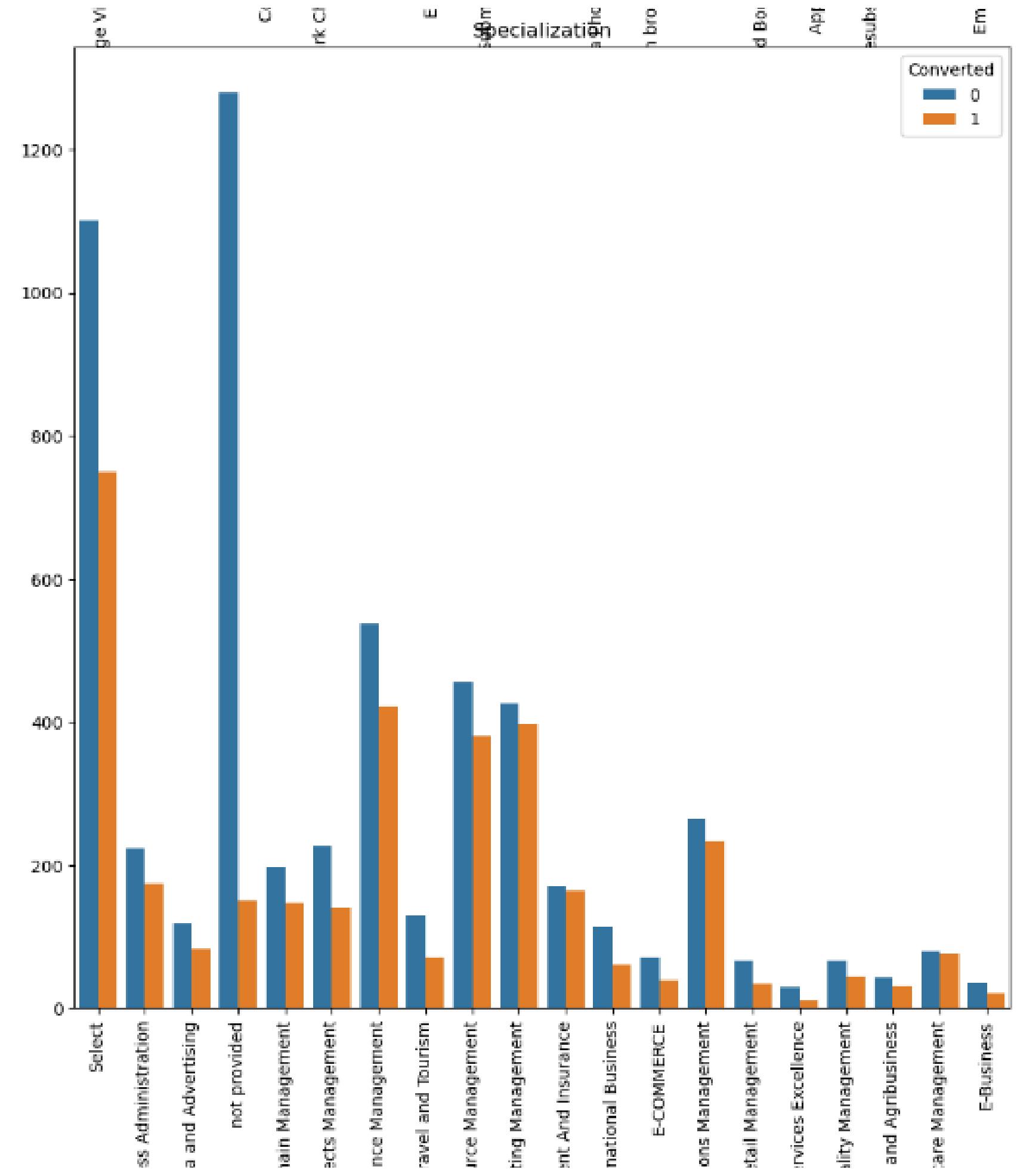
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

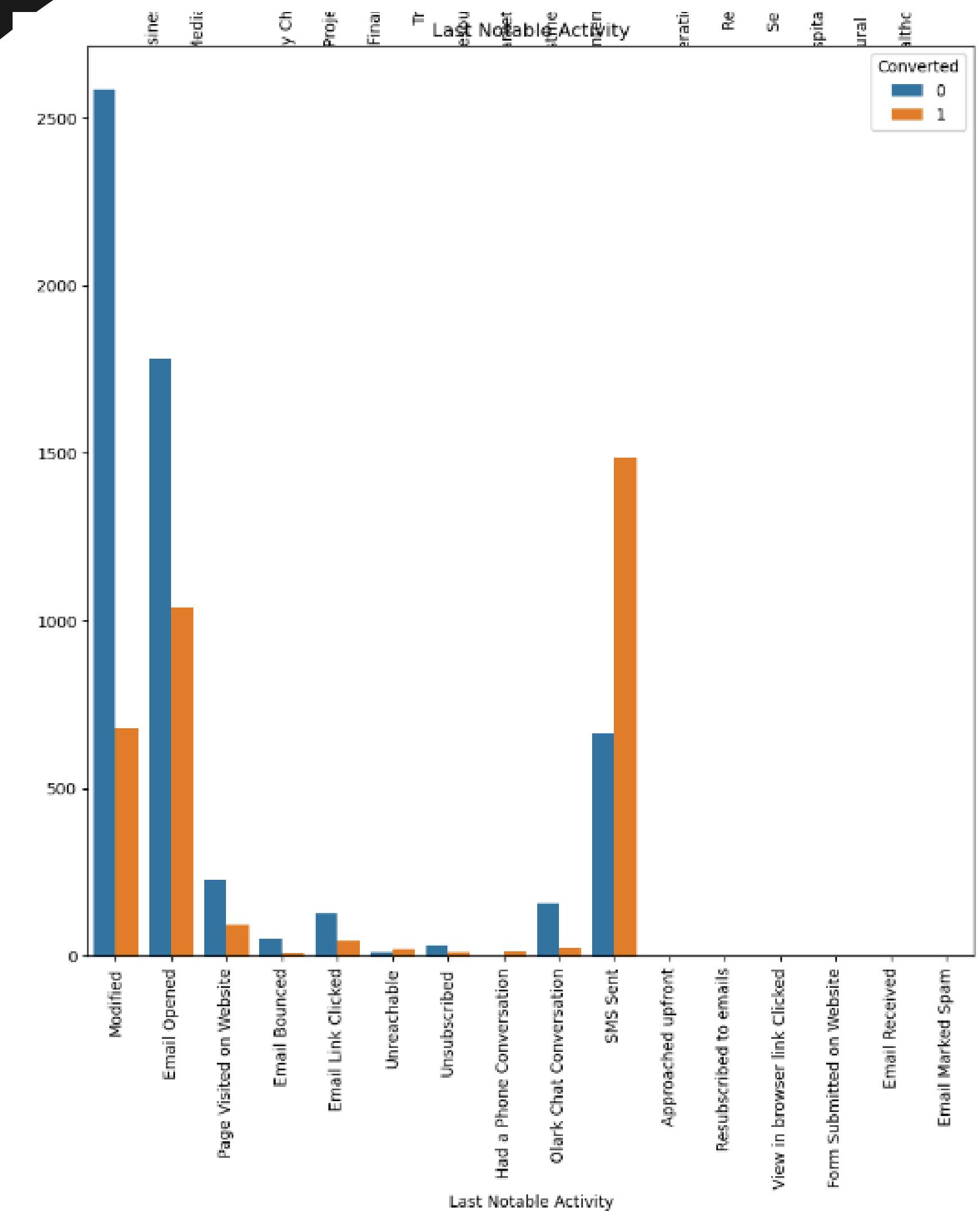
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

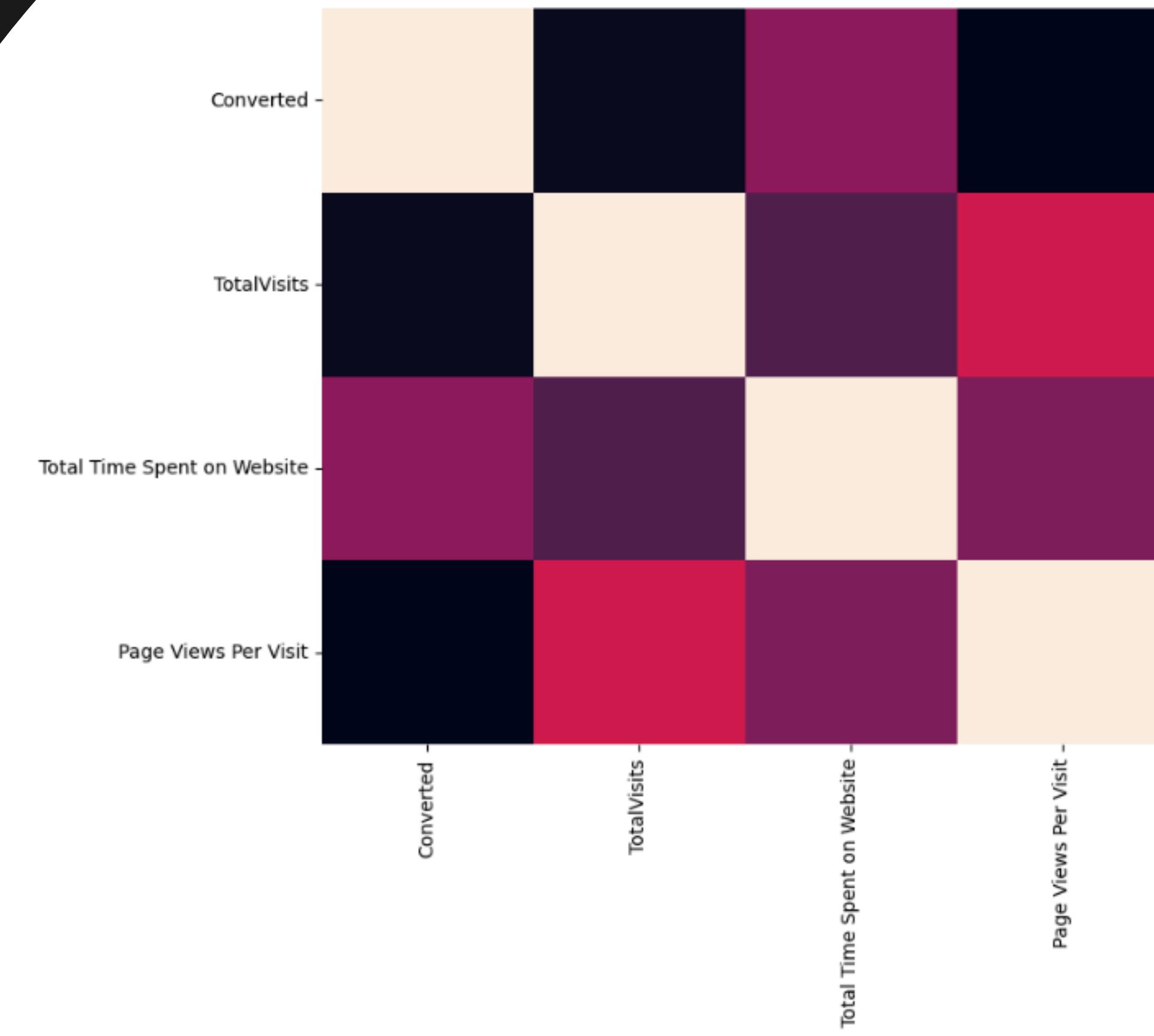
- BIVARIATE ANALYSIS



EDA

EXPLORATORY DATA ANALYSIS

- There is less correlation between numerical columns.

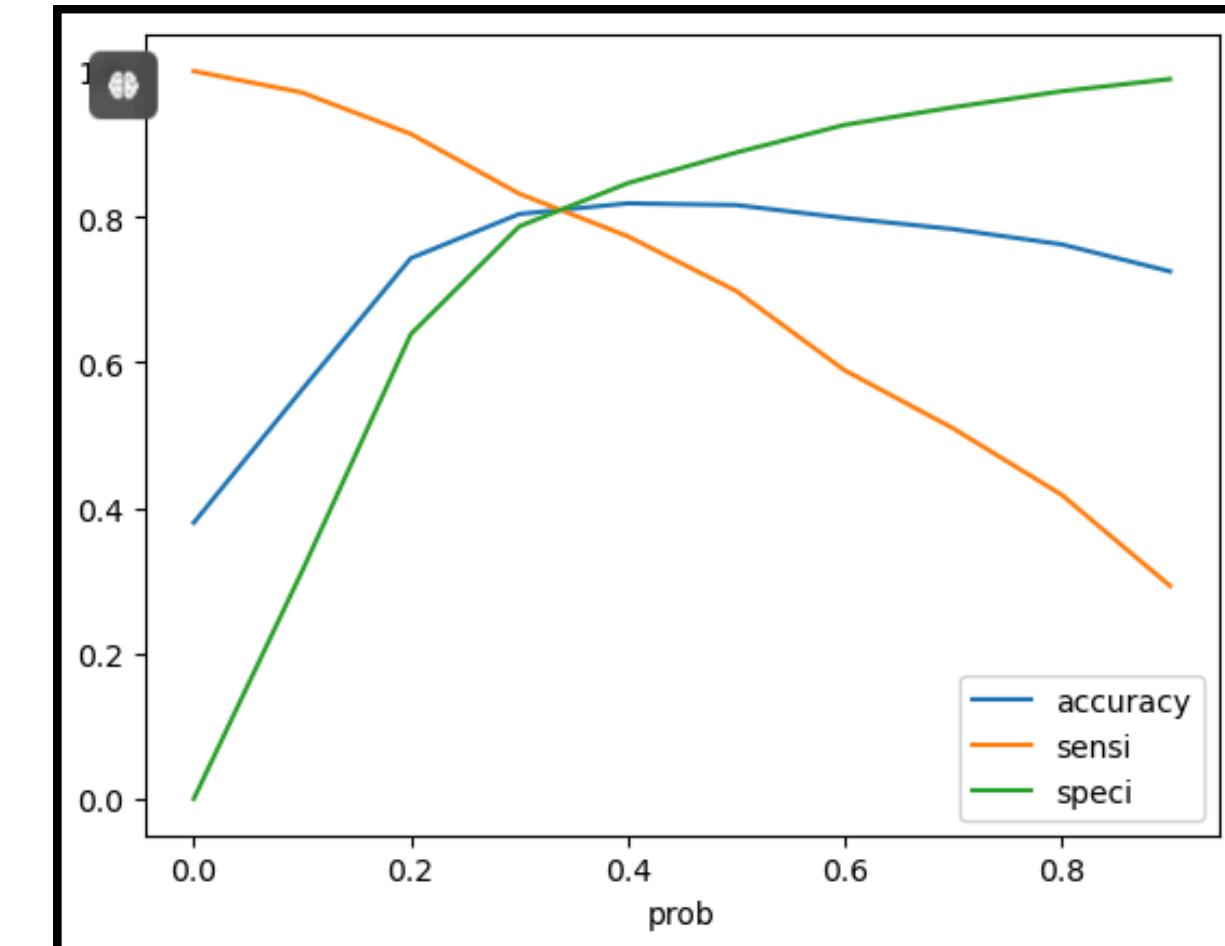
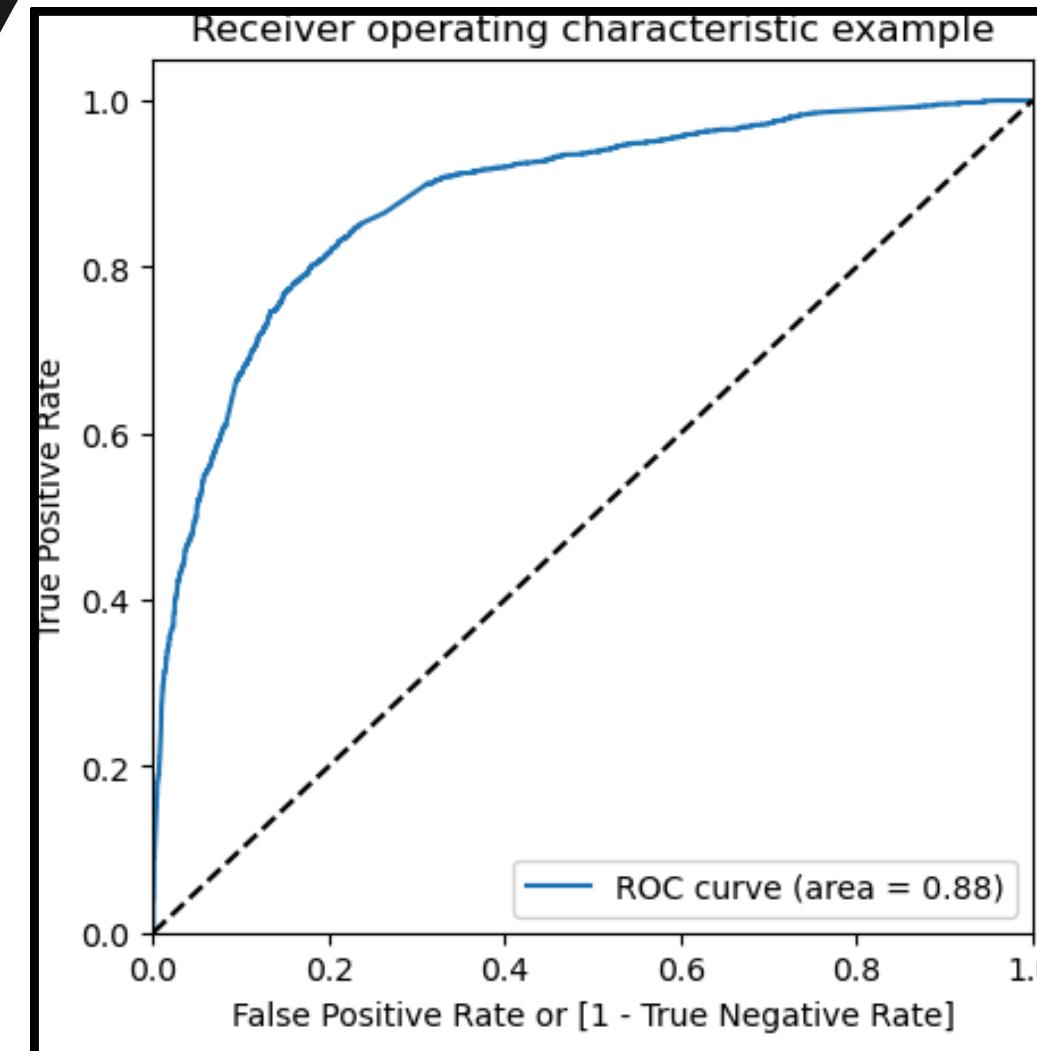


MODEL BUILDING

- After the EDA we have separated the categorical and numerical columns and then we have created the dummy variables from the categorical columns.
- Then we have split the data into training and validation dataset, for which we have used 70-30 ratio. Then we have applied MinMax scaling.
- Next we have checked the correlation among the variables again, since its a large pool of data we could not find any valuable insights. Hence, we are proceeding with RFE.
- Next we have chosen the 15 features in the RFE model as a base and started building the logistic regression model.
- Lastly by iteratively checking the VIF and p values we have eliminated some columns and came with final model.

ROC CURVE

- Area under ROC curve is 0.88
- The optimal cut off is at 0.35.



CONCLUSION

The following variables significantly impact the likelihood of potential buyers converting into customers, ranked in descending order of importance:

- **Total Time Spent on Website:** Longer engagement indicates higher interest.
- **Total Number of Visits:** Frequent visits suggest a strong intent to purchase.
- **Recency of Last Activity:** Recent interactions (e.g., last modified date) reflect current interest.
- **Specific Pages Visited:** Engagement with key content can signal readiness to buy.
- **Olark Chat Conversations:** Direct interactions can enhance buyer confidence and clarify doubts.
- **Email Engagement:** Metrics such as opened emails and clicked links show responsiveness to marketing efforts.
- **Current Occupation:** Working professionals may be more inclined to invest in courses for career advancement.

By focusing on these variables, X Education can strategically target and convert potential buyers, significantly increasing their chances of successful sales.