

LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

CONTENTS

- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- □ Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.
- □ The typical lead conversion rate at X education is around 30%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- □ If they successfully identify this set of leads, the lead 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

BUSINESS OBJECTIVE

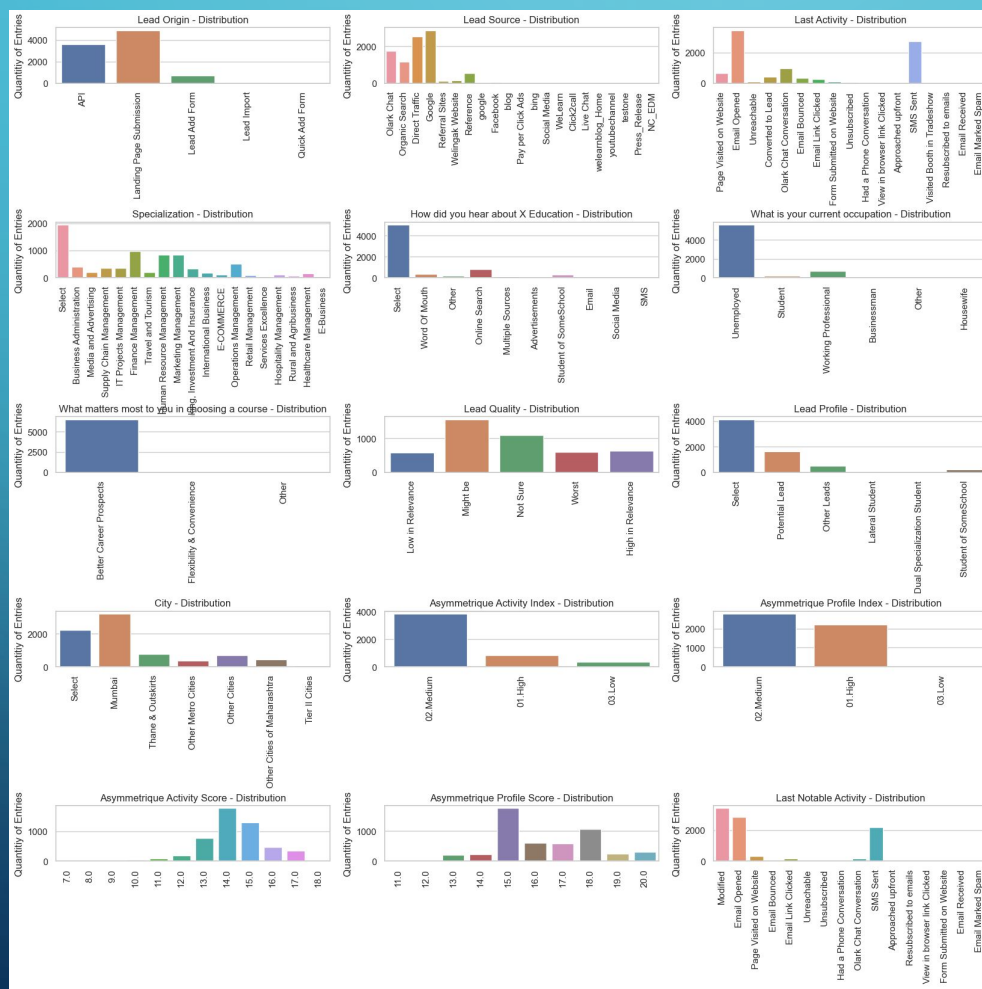
- ❑ Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.
- ❑ The CEO want to achieve a lead conversion rate of 80%.
- ❑ They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches

PROBLEM APPROACH

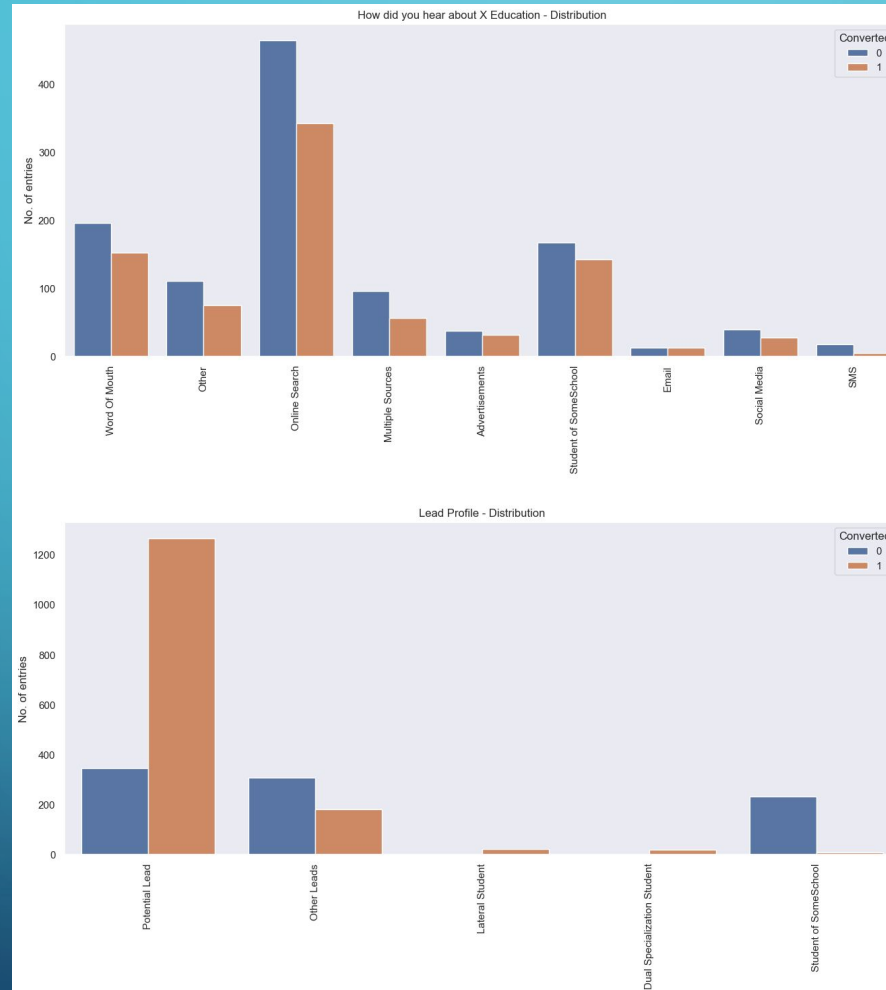
- ☐ Data Analysis
- ☐ Data Cleaning
- ☐ Deleting Columns
- ☐ Detection of Outliers
- ☐ Analyze the Data
- ☐ Univariate Analysis
- ☐ Bivariate Analysis
- ☐ View Points
- ☐ Dummy variable creation Test-Train split

EDA – DATA CLEANING

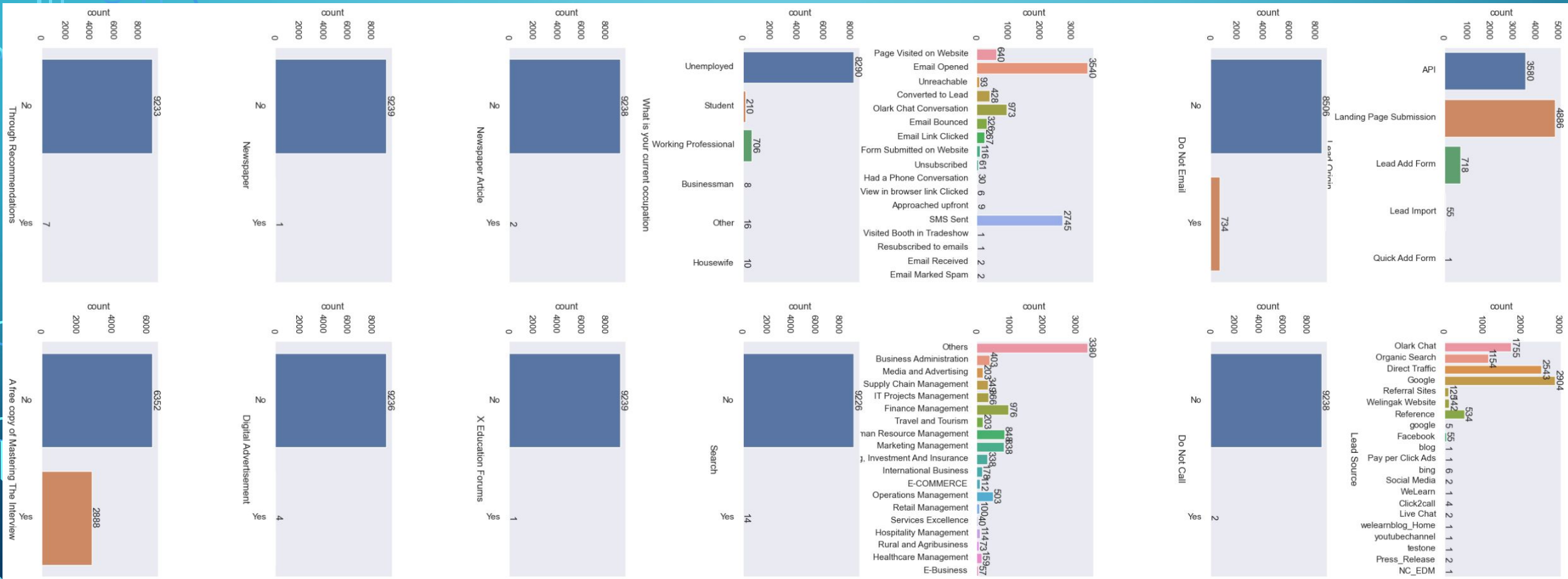
Data Cleaning & Handling duplicate values



THERE ARE A FEW COLUMNS IN WHICH THERE IS A LEVEL CALLED 'SELECT' WHICH IS ASKING

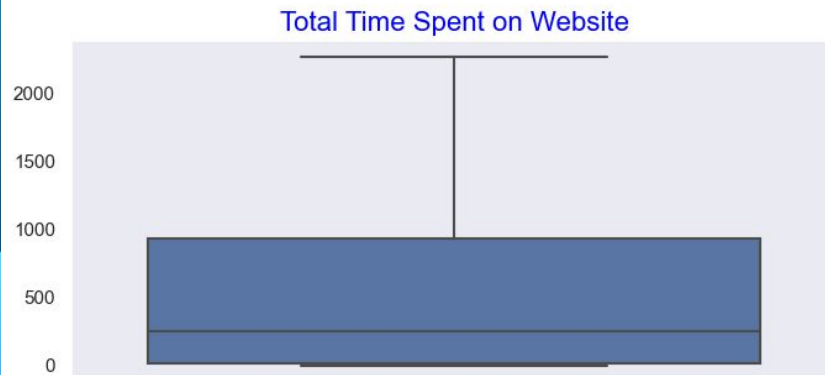
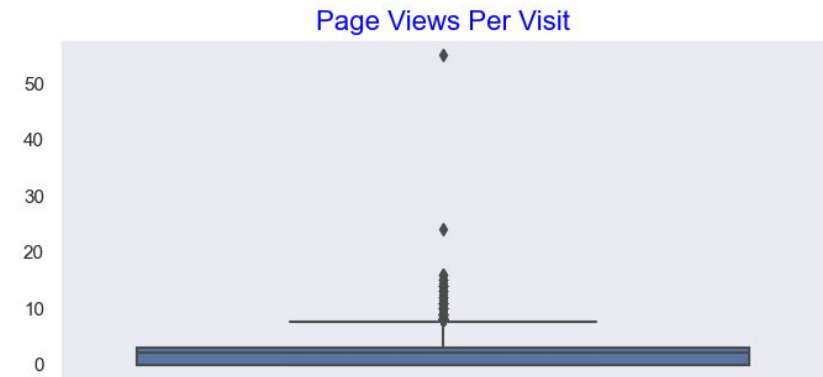
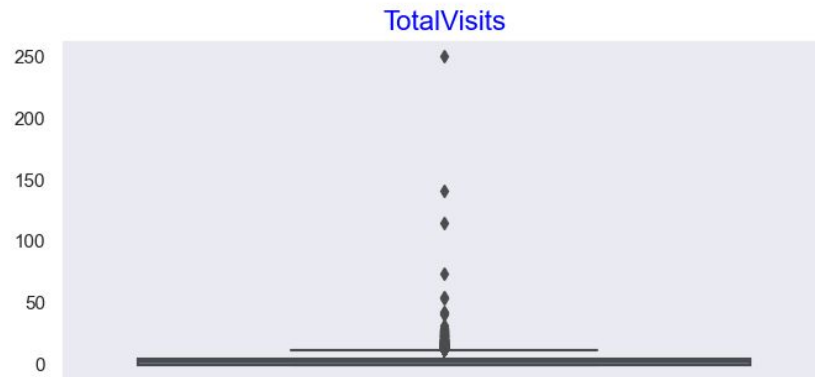


ANALYZING CATEGORICAL COLUMNS

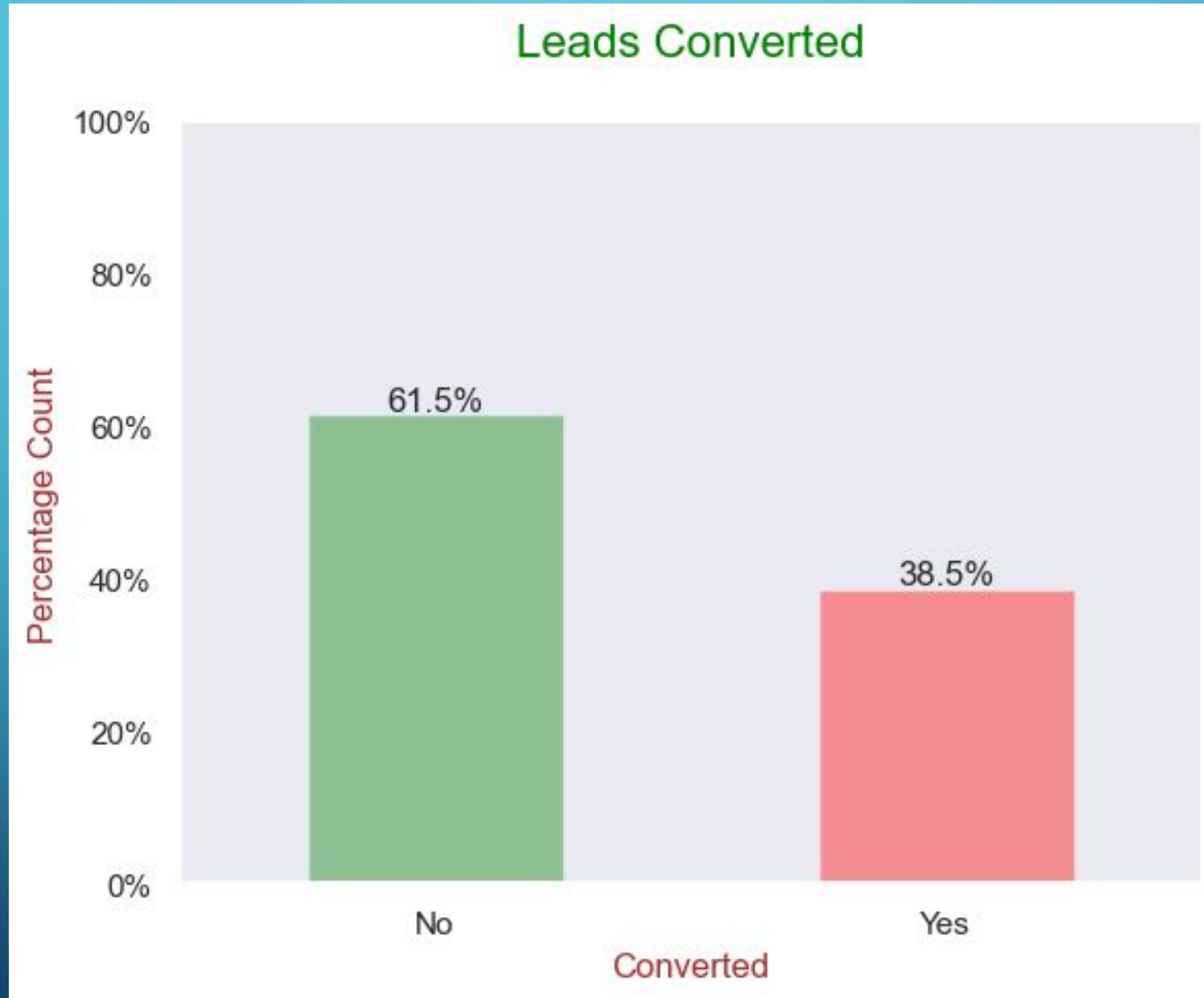


DETECTION OF OUTLIERS

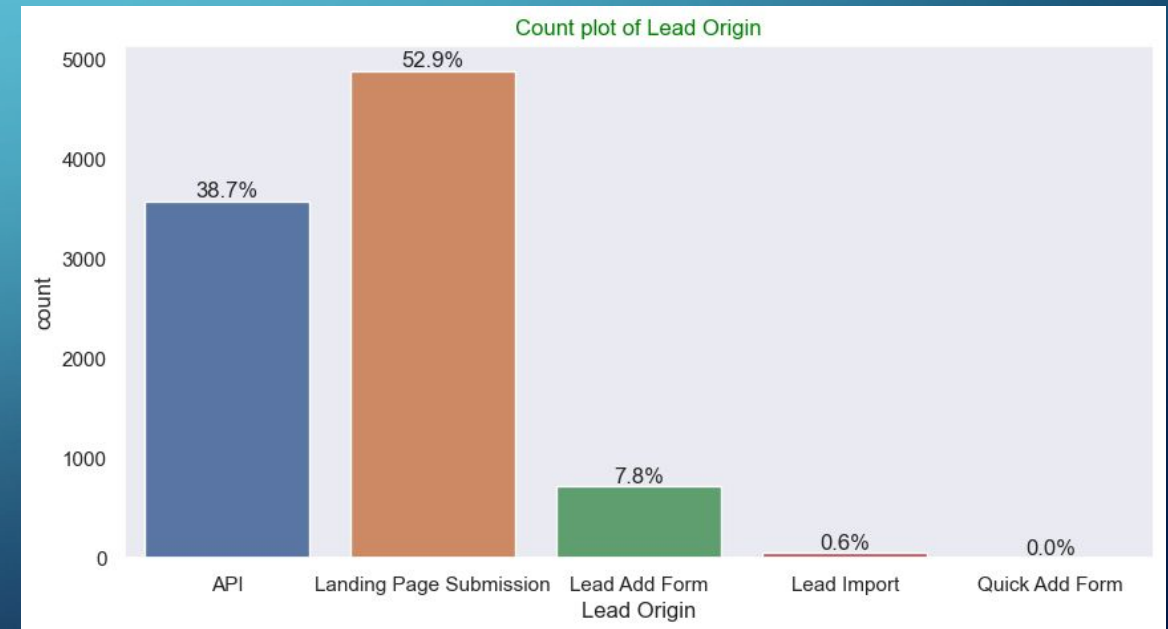
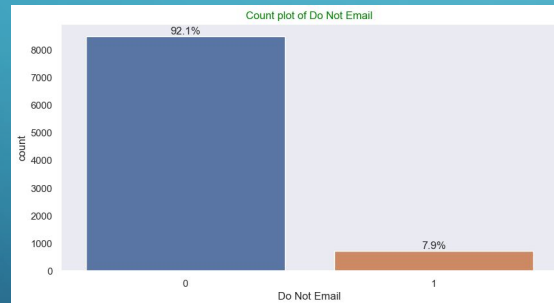
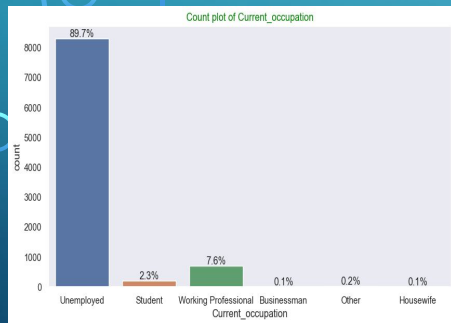
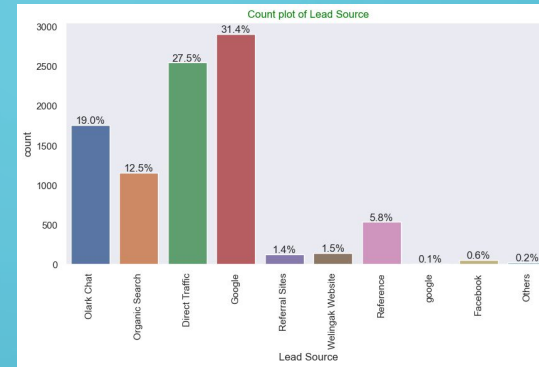
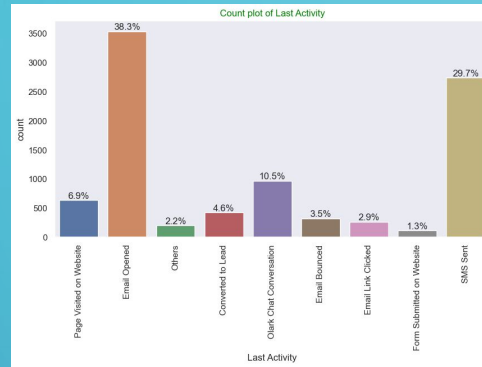
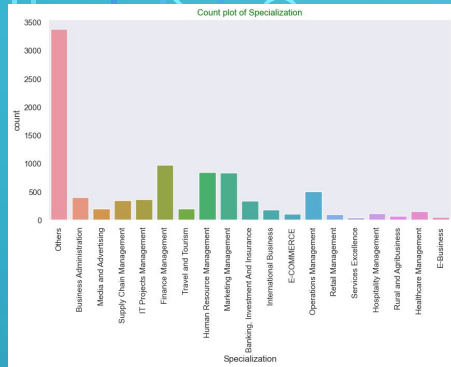
Checking Outliers using Boxplot



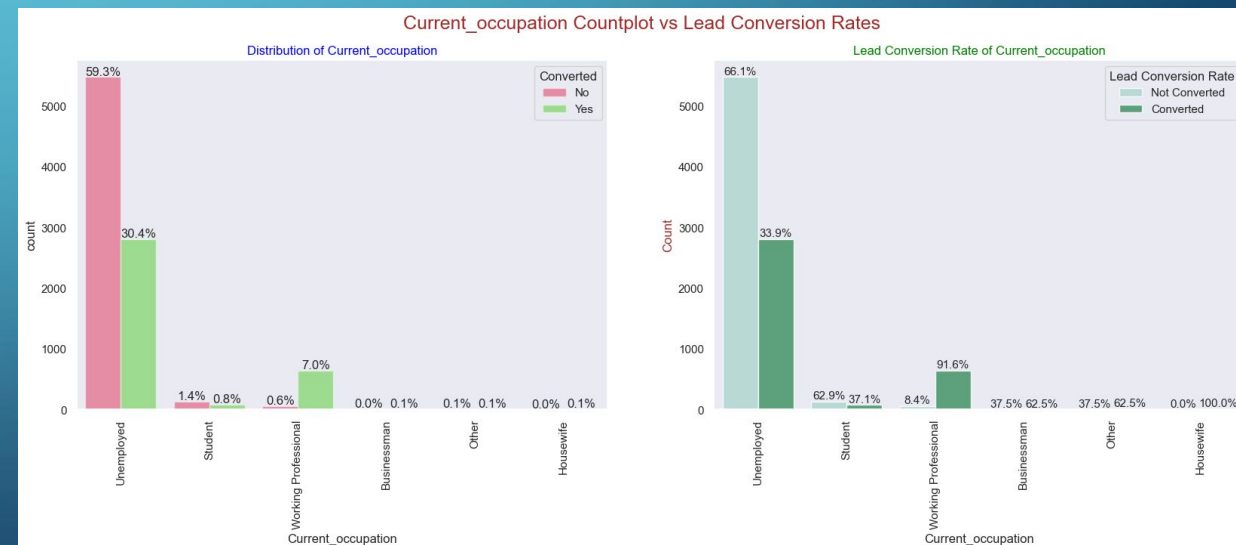
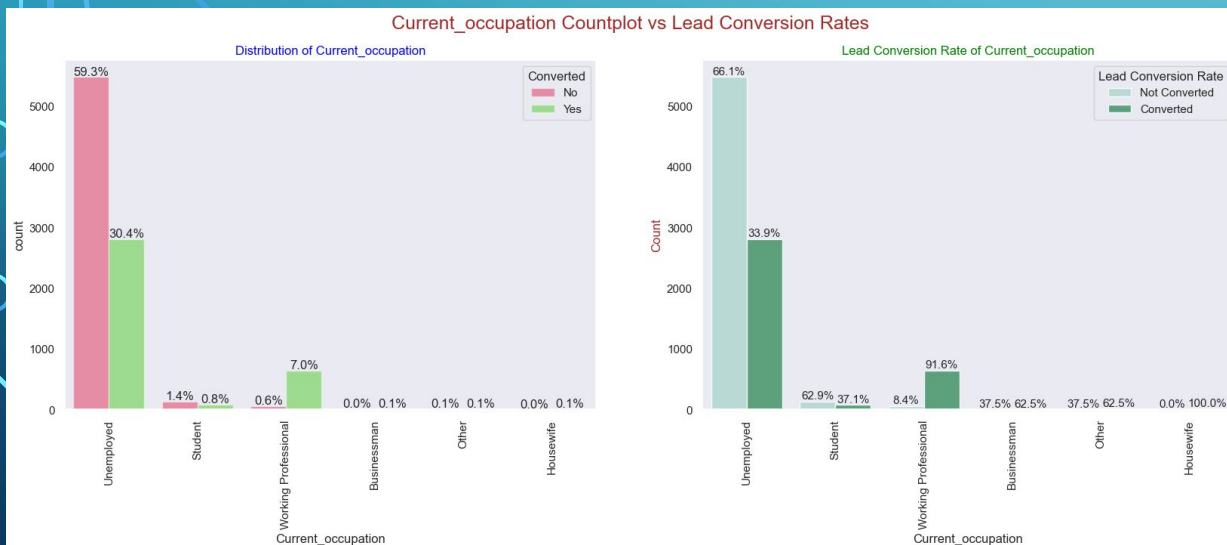
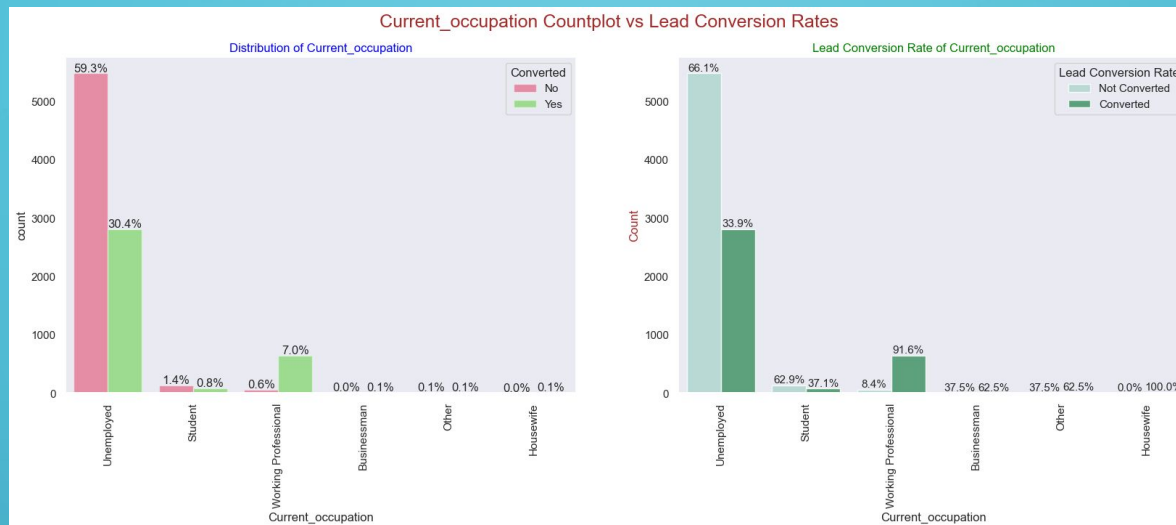
ANALYSE THE DATA



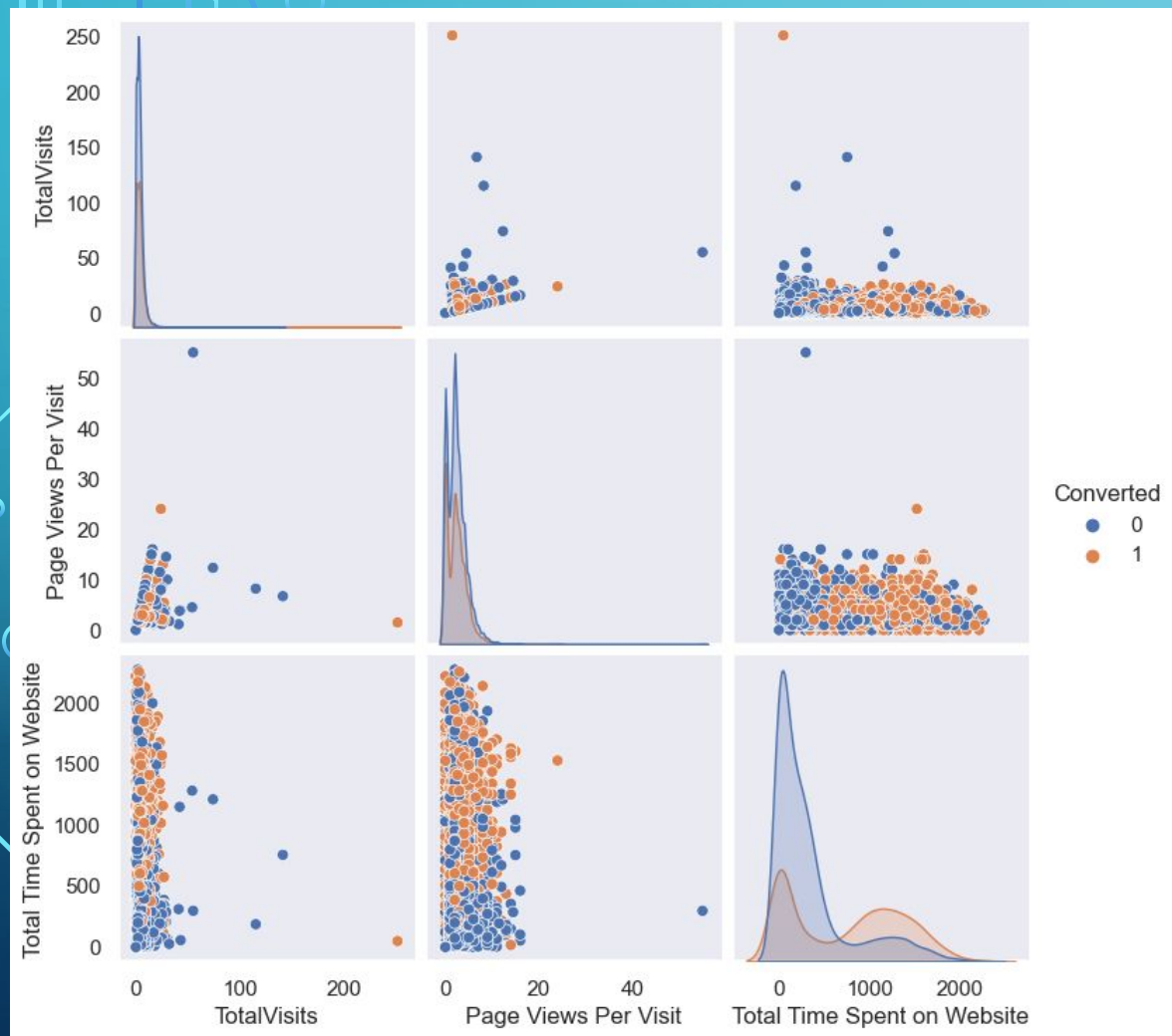
UNIVARIATE ANALYSIS



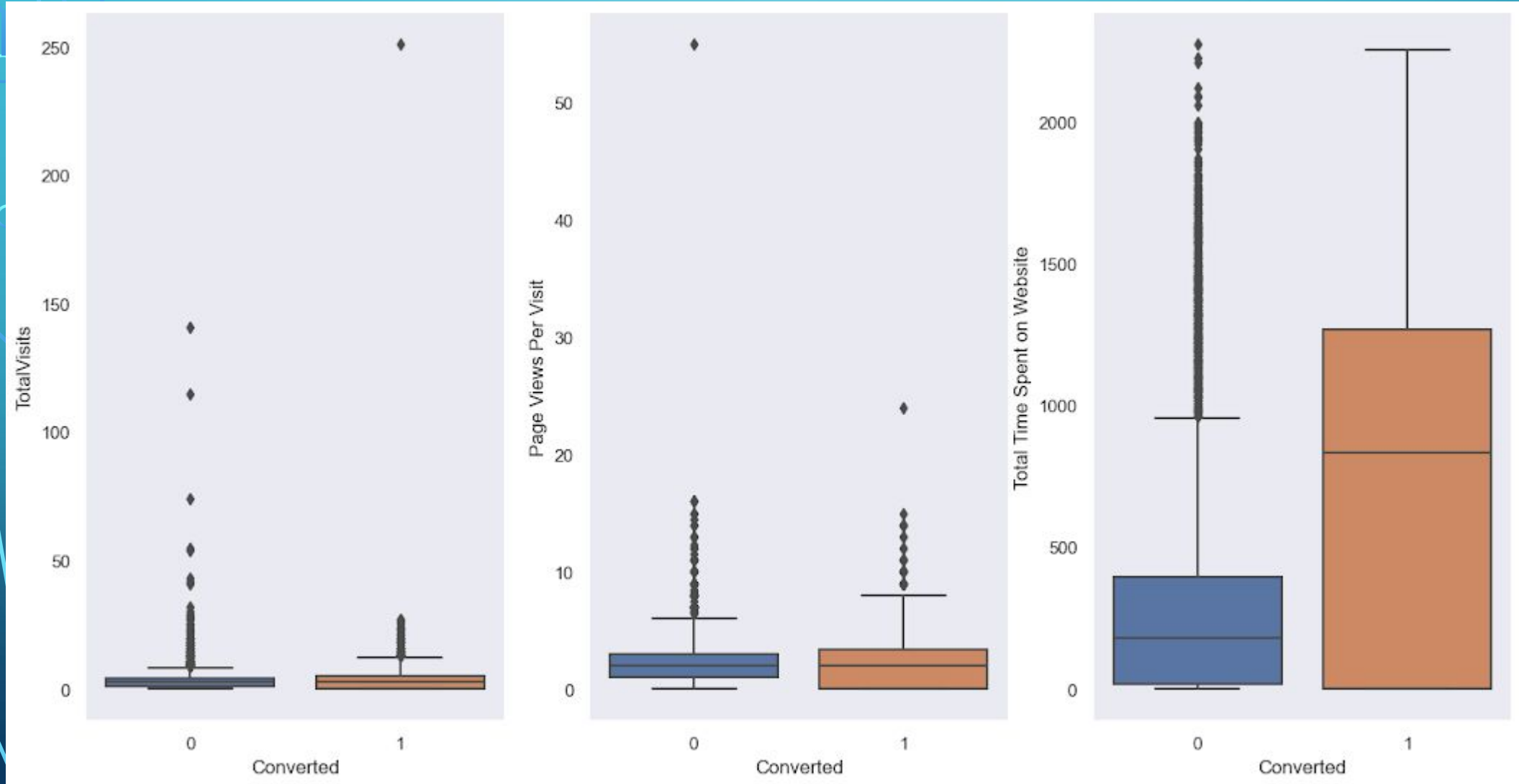
BIVARIATE ANALYSIS



VIEWPOINTS:

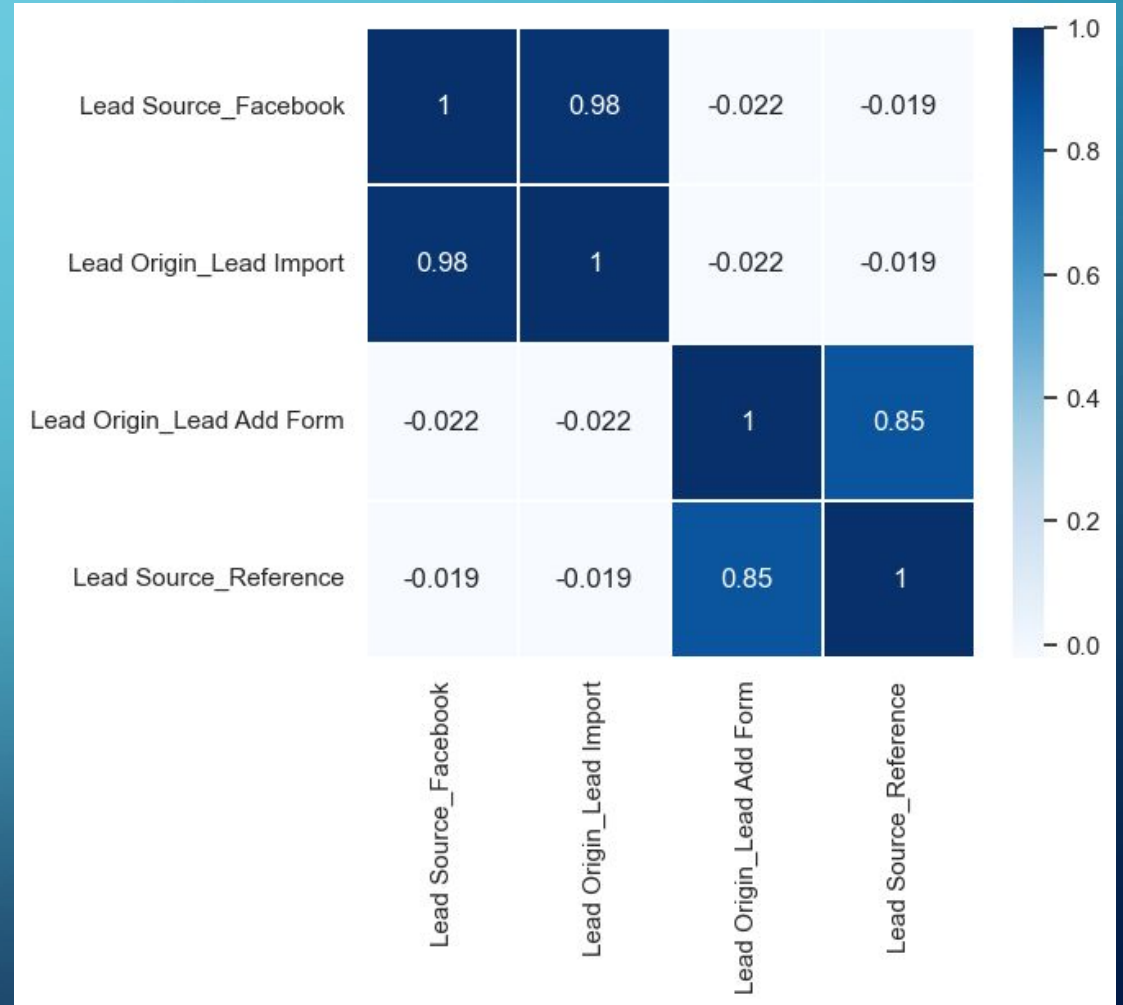
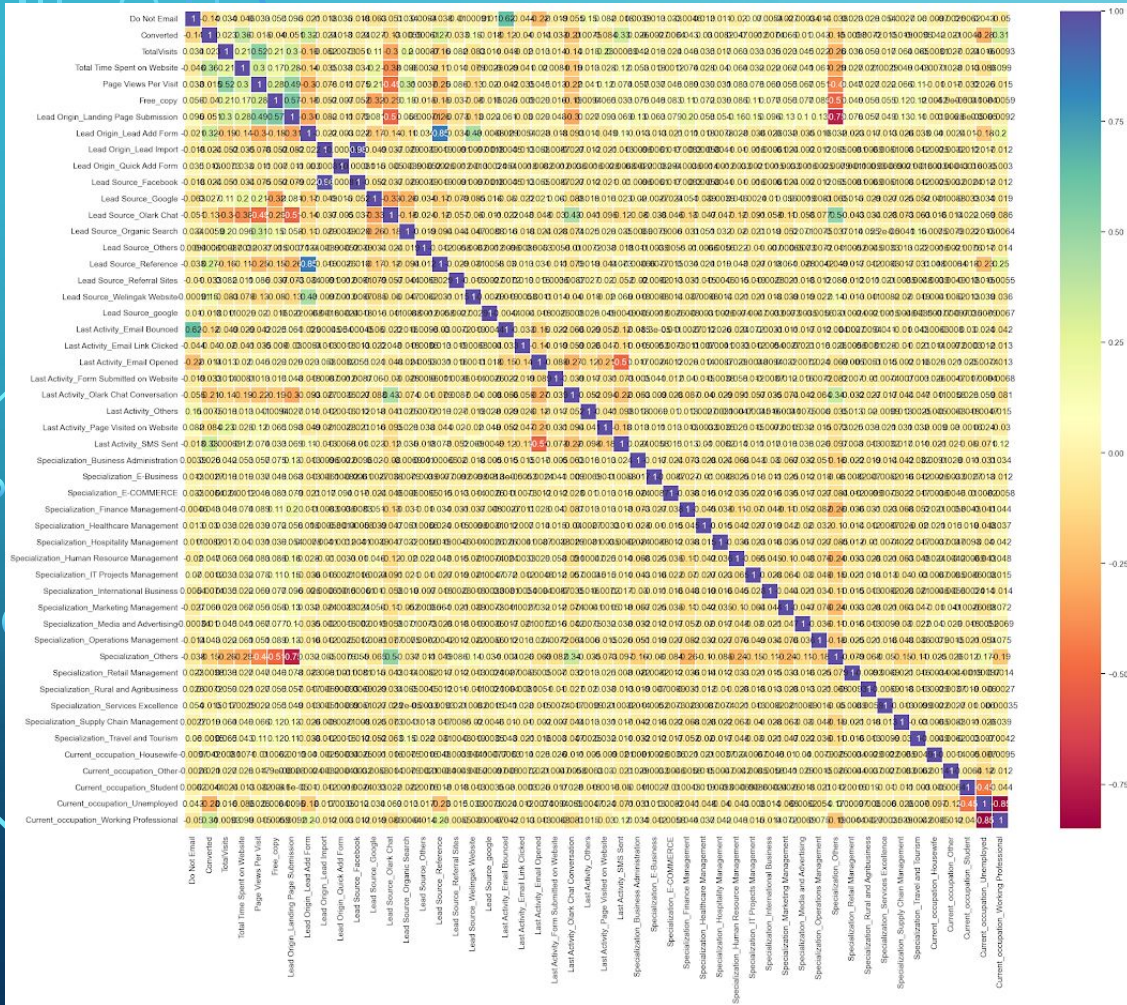


BOX PLOT



DATA PREPARATION

Creation of Dummy Variables – Test Train Split



FINAL FINDINGS:

Train Data Set:

- Precision: 80.57%
- 80.41% sensitivity
- Particularity: 80.66%

Test Data Set:

- 80.41% accuracy
- 79.82% sensitivity
- Particularity: 80.80%

- The model is performing consistently across numerous evaluation criteria in both the test and train datasets, as shown by the evaluation matrices' proximity to one another.

- Using a cut-off value of 0.347, the model attained a sensitivity of 80.05% in the train set and 79.82% in the test set. Sensitivity here refers to the proportion of leads that the model correctly predicts among all possible leads that convert. A target sensitivity of about 80% had been established by the CEO of X Education.

Additionally, the model's accuracy, which was 80.46%, met the goals of the study.