

# LEADS SCORING CASE STUDY

SUBMITTED BY :

1. Anadi Mitra
2. Anju Gautam
3. Anmol Bhandari

# PROBLEM STATEMENT:

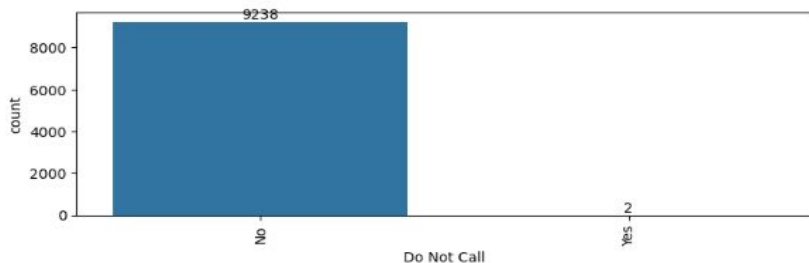
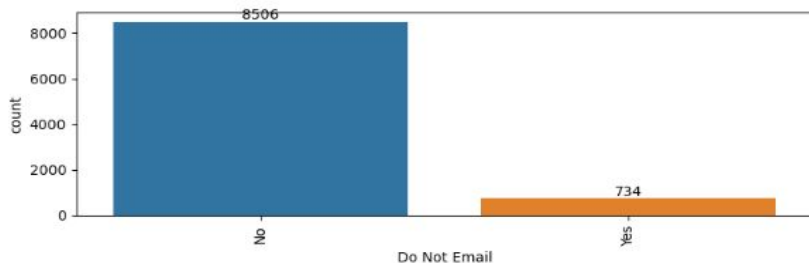
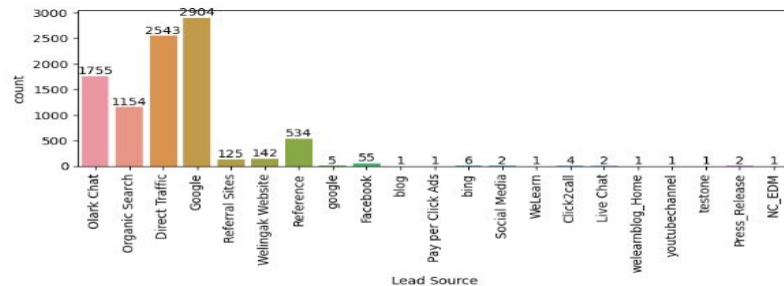
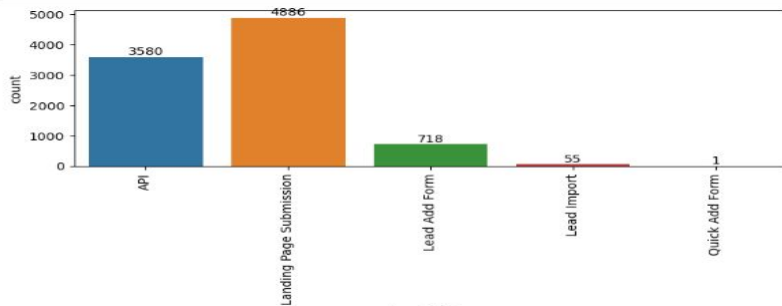
- X Education sells online courses to industry professionals.
- X Education gets a lot of leads on a daily basis, its lead conversion rate is very poor at 30% as every single person is being contacted and very few people are actually buying the courses.
- To make the process more efficient the company wishes to identify most potential leads known as 'Hot Leads' and pursue them.
- This will can inturn take the conversion rate to almost 80%.

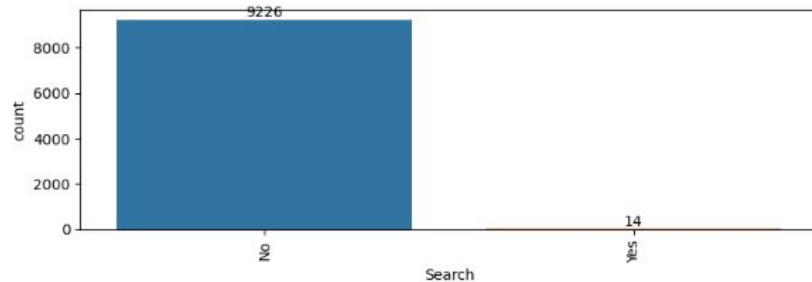
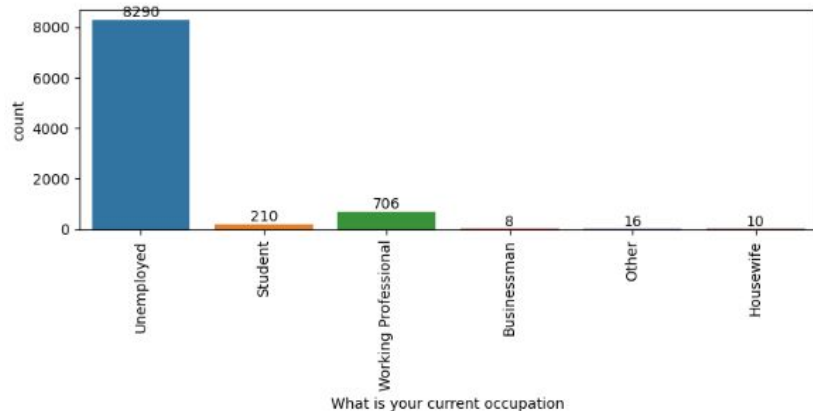
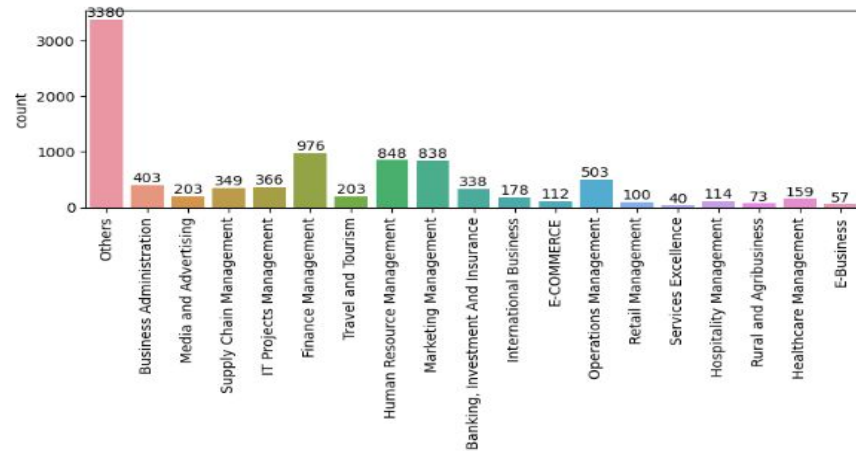
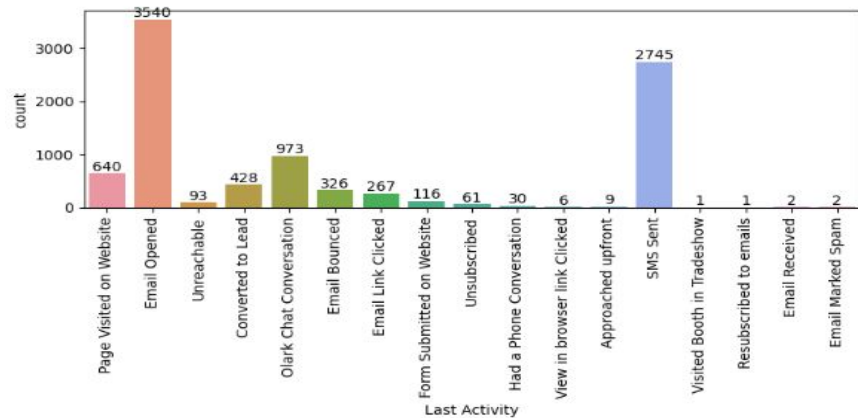
# PROBLEM APPROACH:

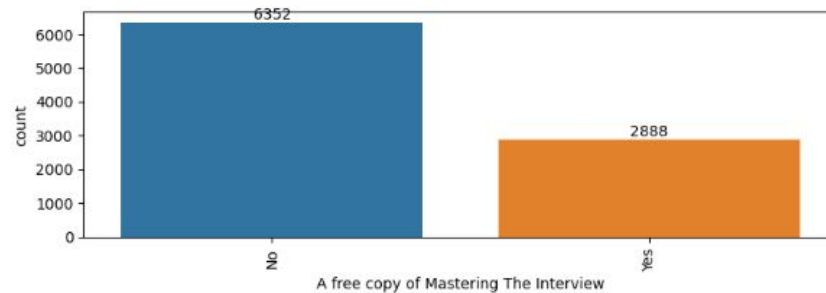
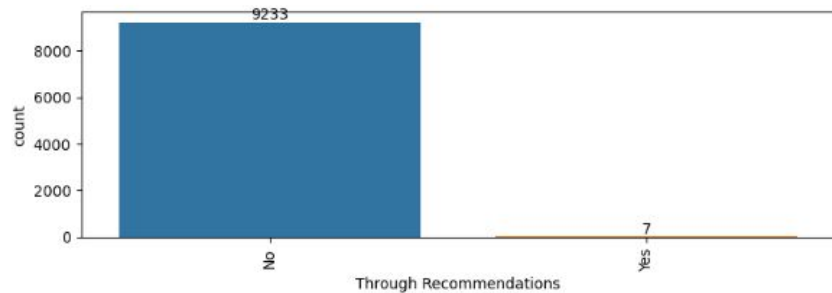
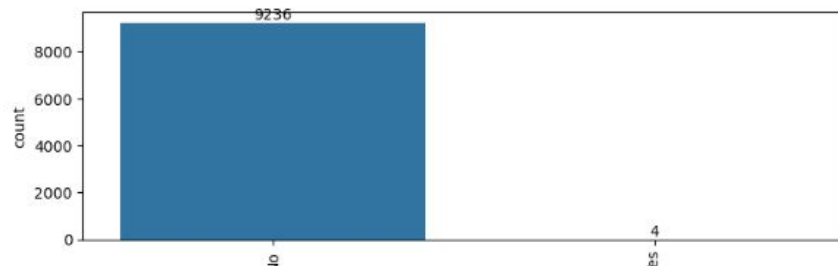
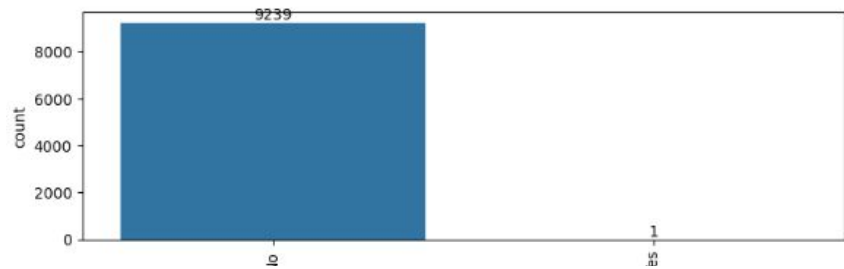
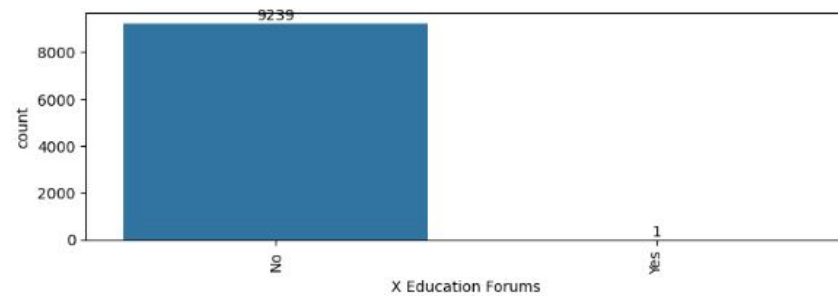
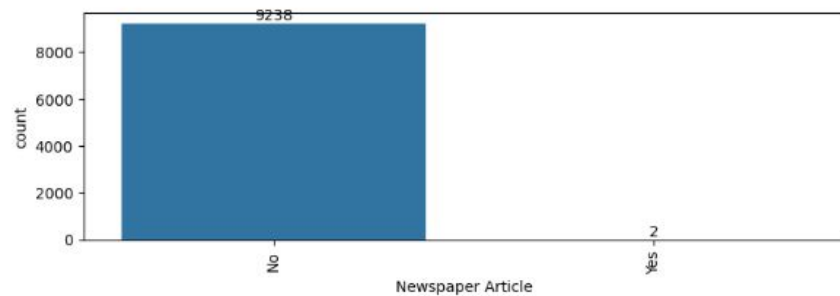
- Importing the data and inspecting the data frame
- Data Understanding and Inspection
- Data Cleaning
- Data Analysis (EDA)
- Data Preparation
- Test-Train Split
- Feature Scaling
- Model Building
- Model Evaluation
- Making Predictions on test set

# Data Analysis (EDA)

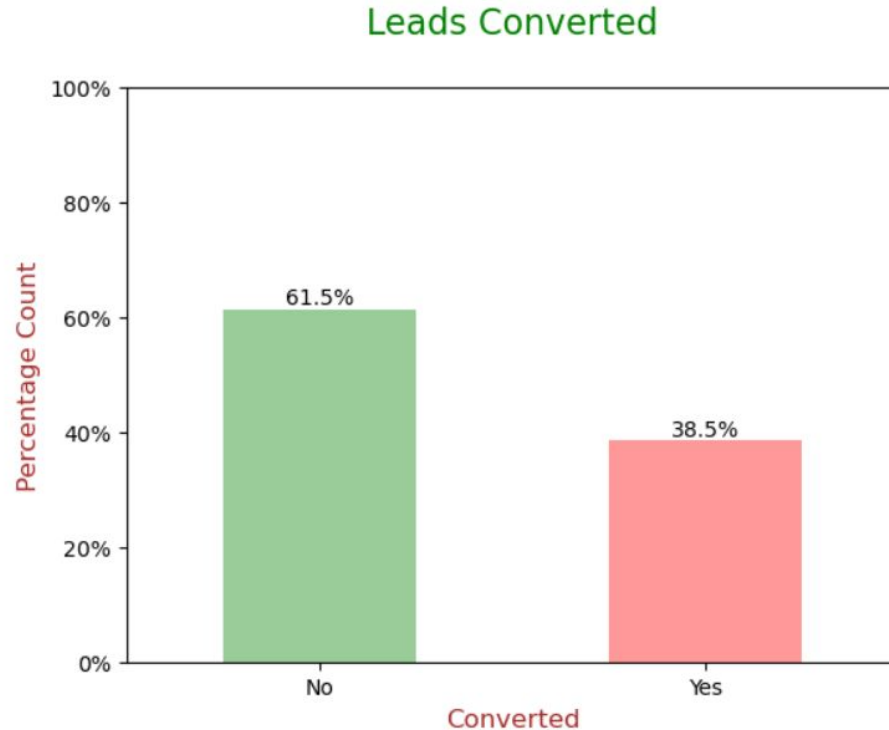
## Checking & Dropping Category Columns that are Skewed



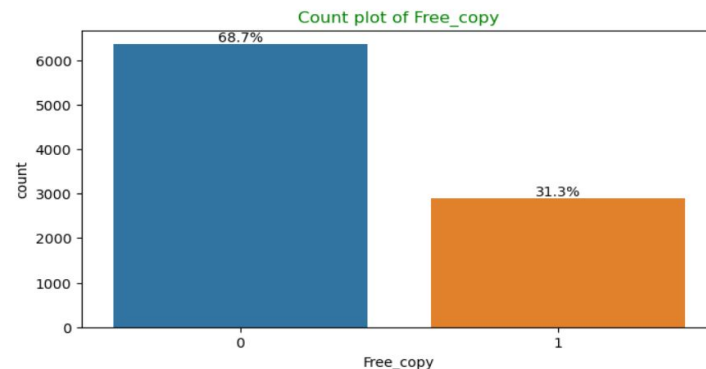
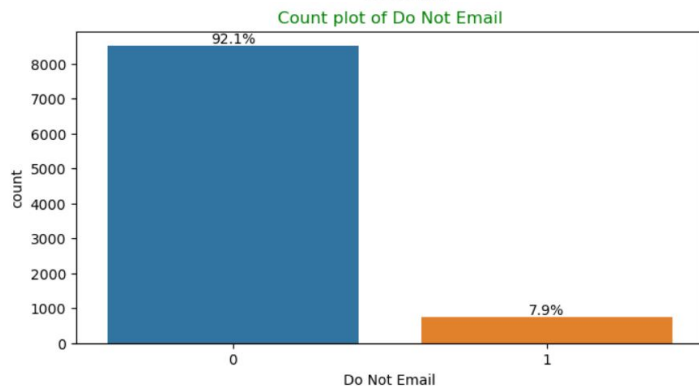
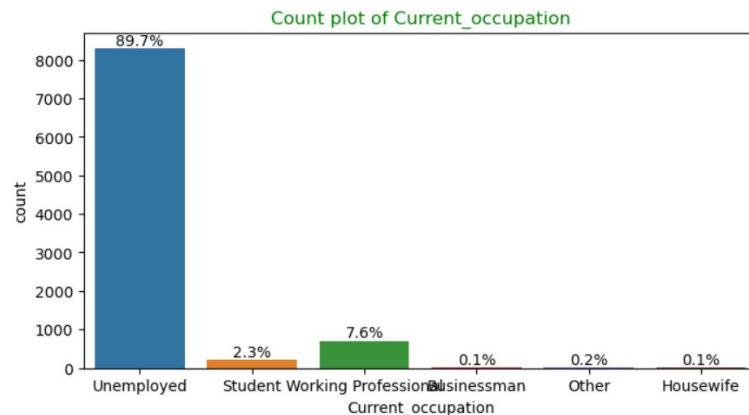
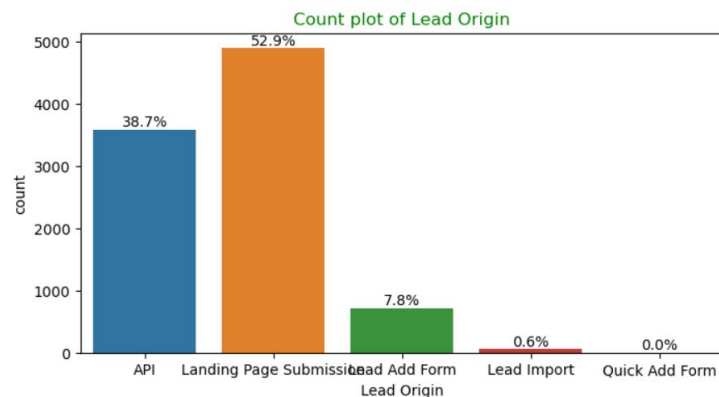




# Checking the imbalance in the target variable



# Univariate Analysis for Categorical Variables



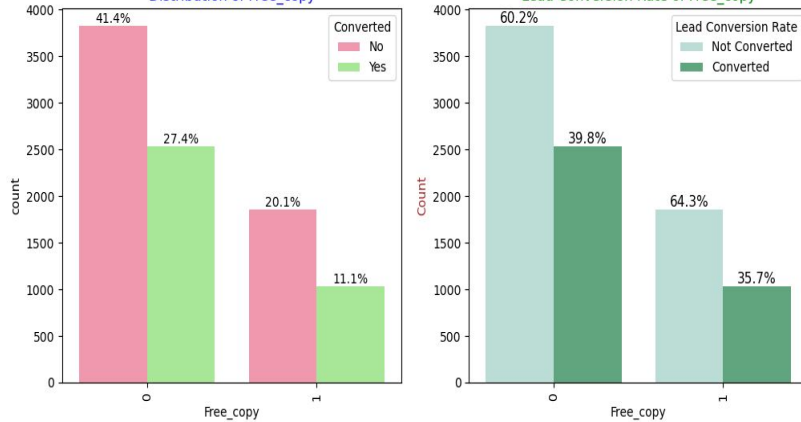


# Bivariate Analysis

Free\_copy Countplot vs Lead Conversion Rates

Distribution of Free\_copy

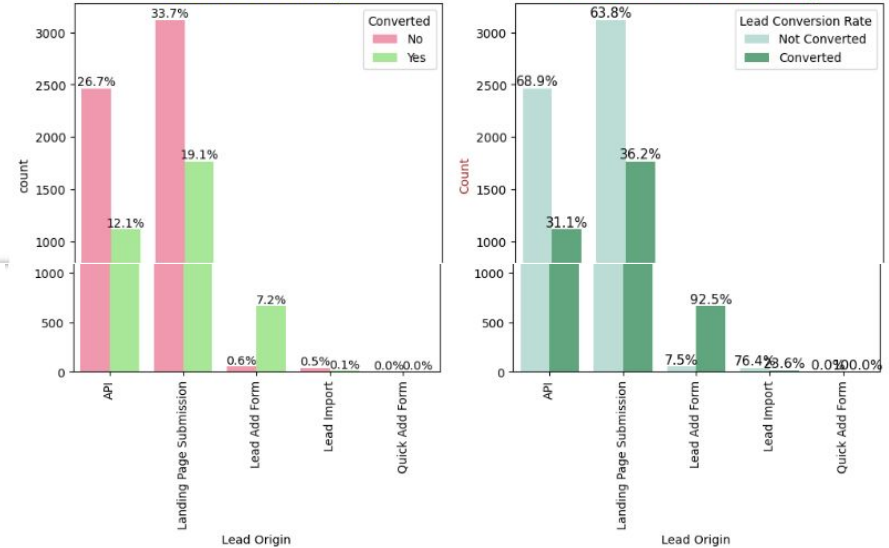
Lead Conversion Rate of Free\_copy



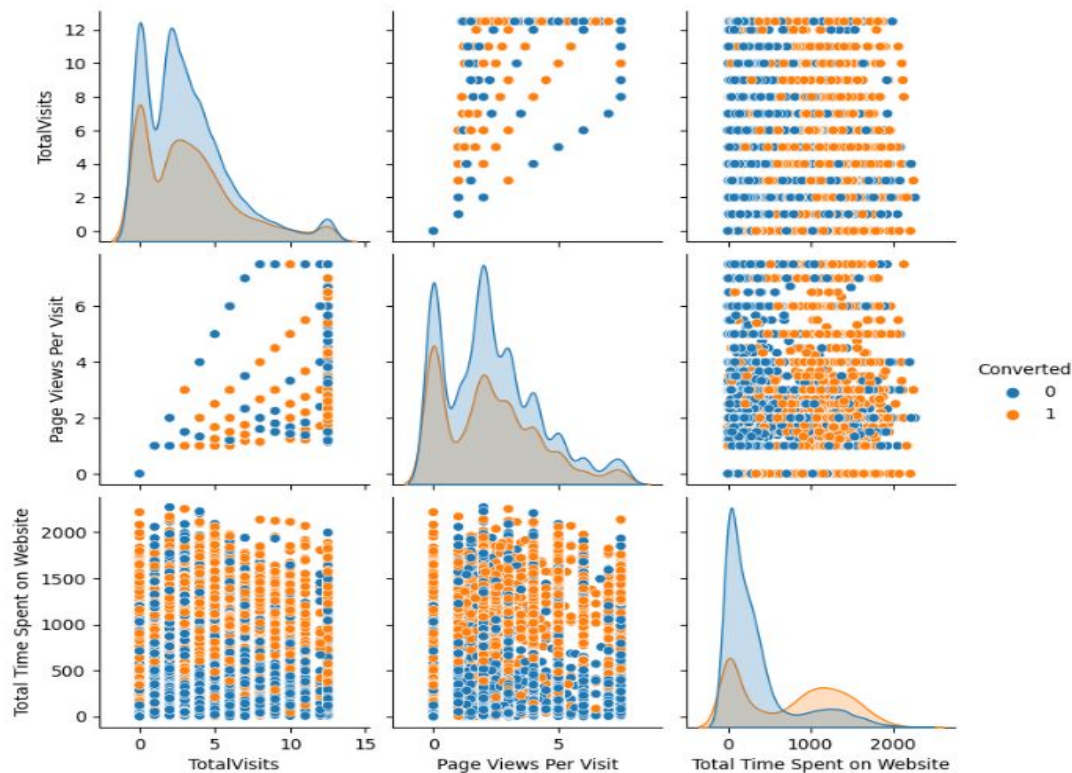
Lead Origin Countplot vs Lead Conversion Rates

Distribution of Lead Origin

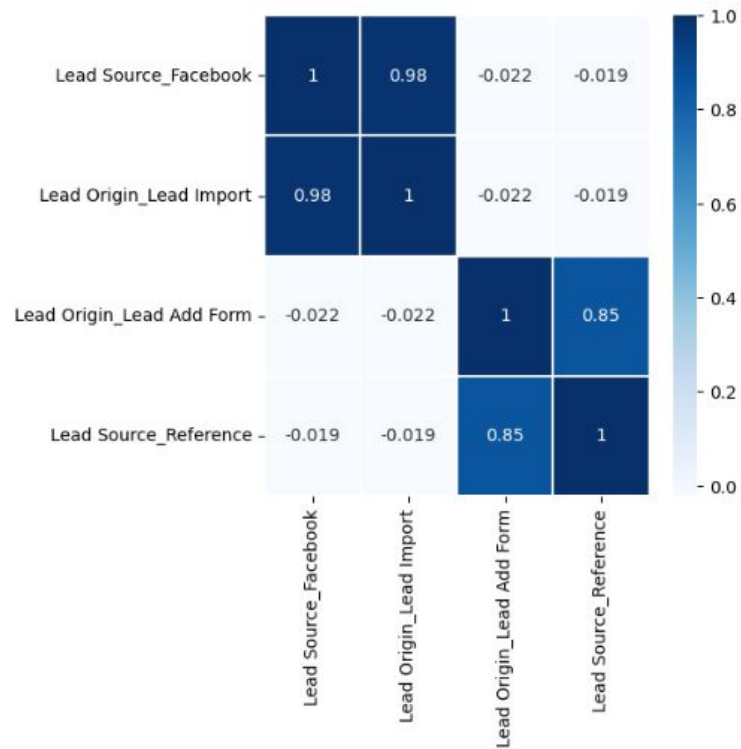
Lead Conversion Rate of Lead Origin

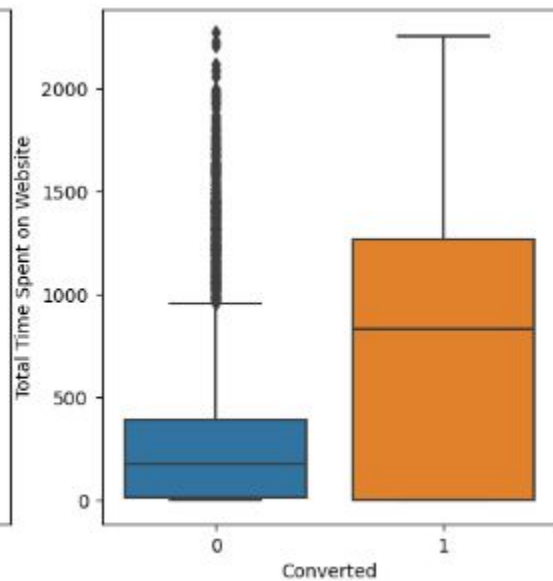
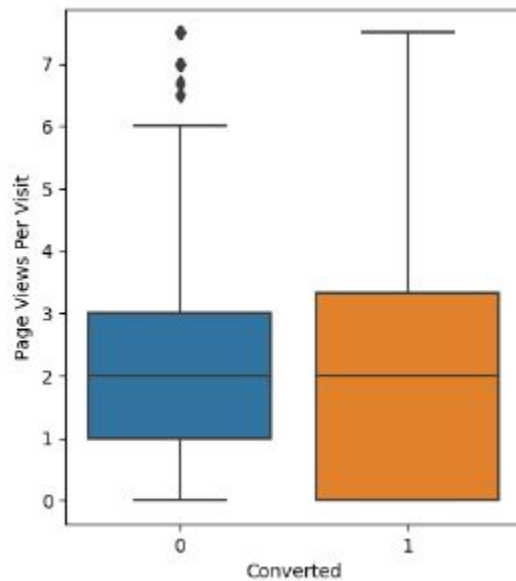
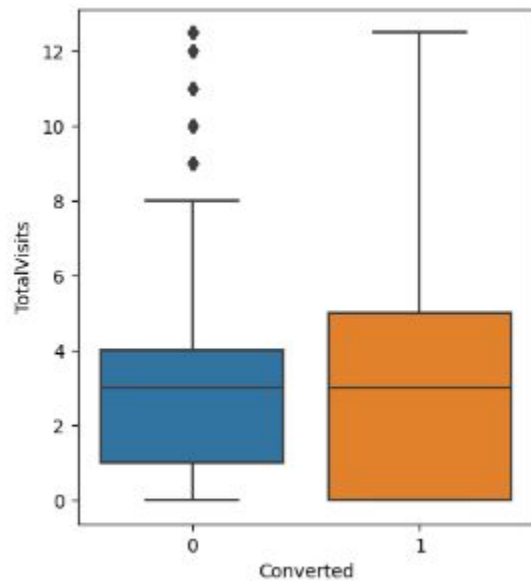


# Bivariate Analysis for Numerical Variables

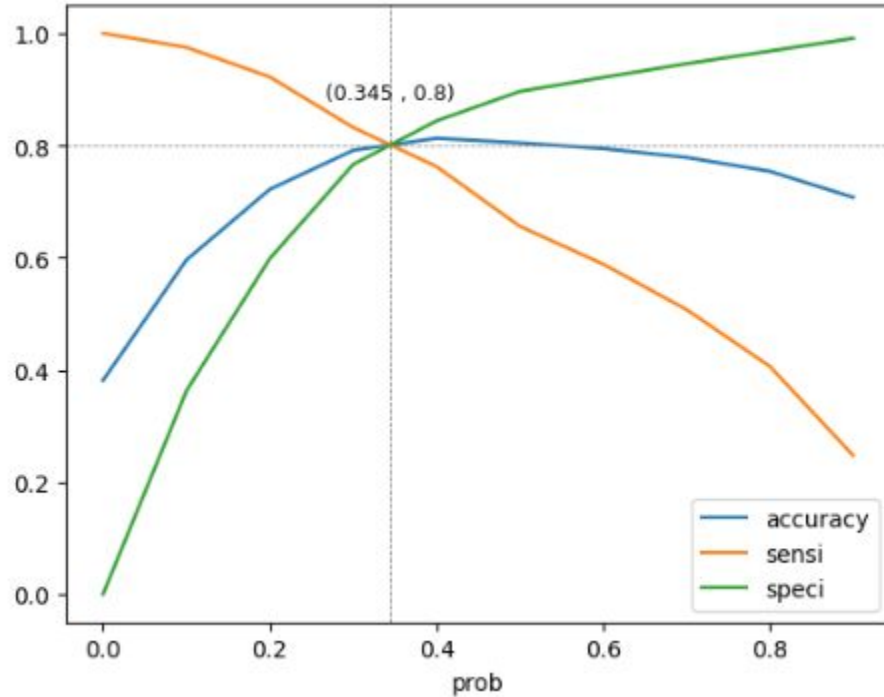


# Correlation Matrix





# Finding Optimal Cutoff Point/ Probability



# Precision and recall tradeoff

