

# LEAD SCORING CASE STUDY

By Parag Soni

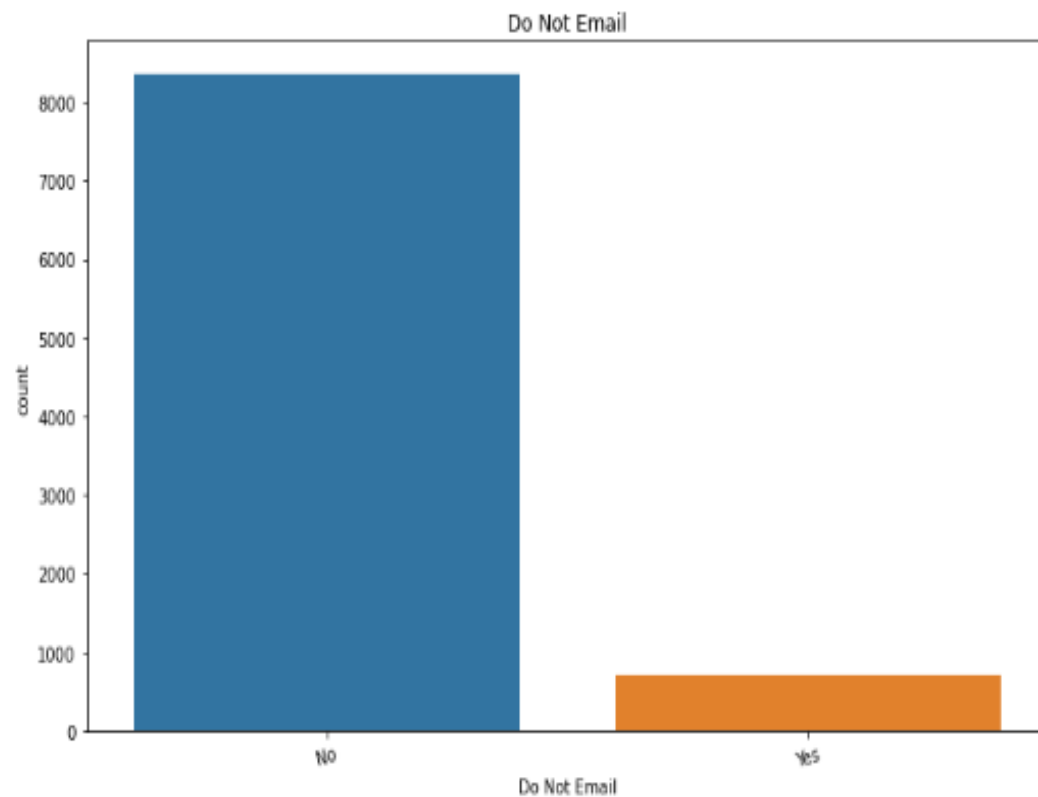
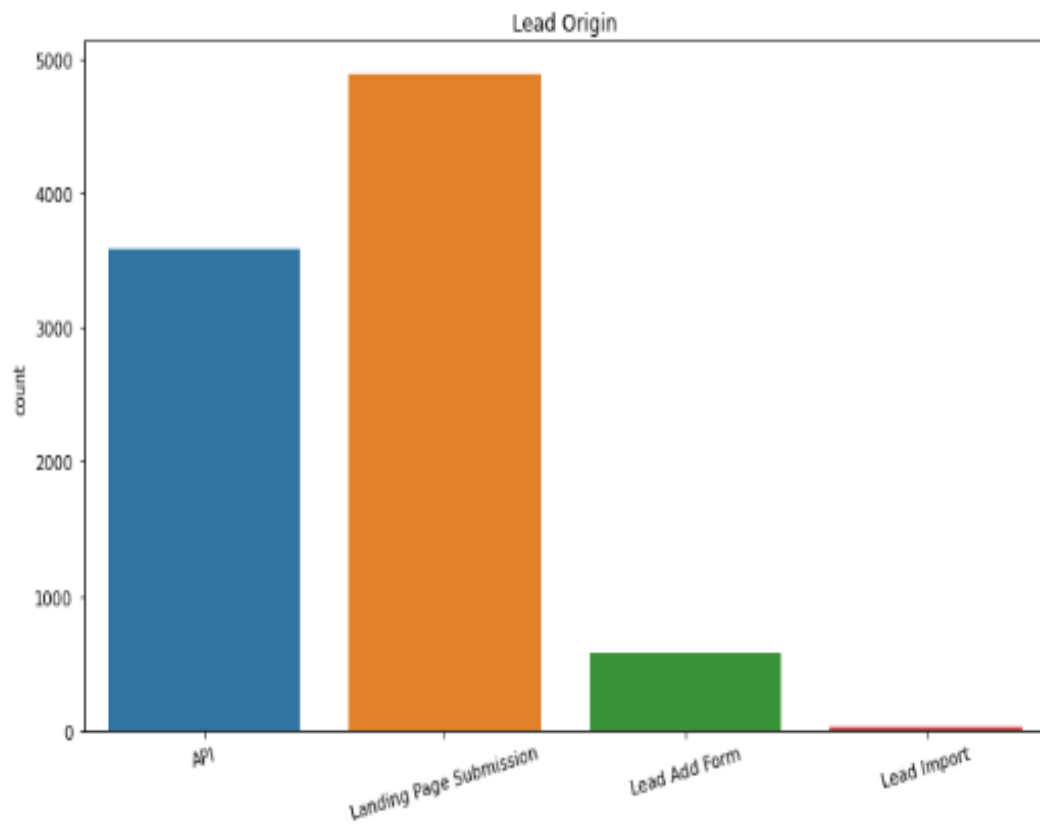
# Methodology

- Data Cleaning and Preparation
- Model Building
- Model Evaluation
- Predictions on the Test Set

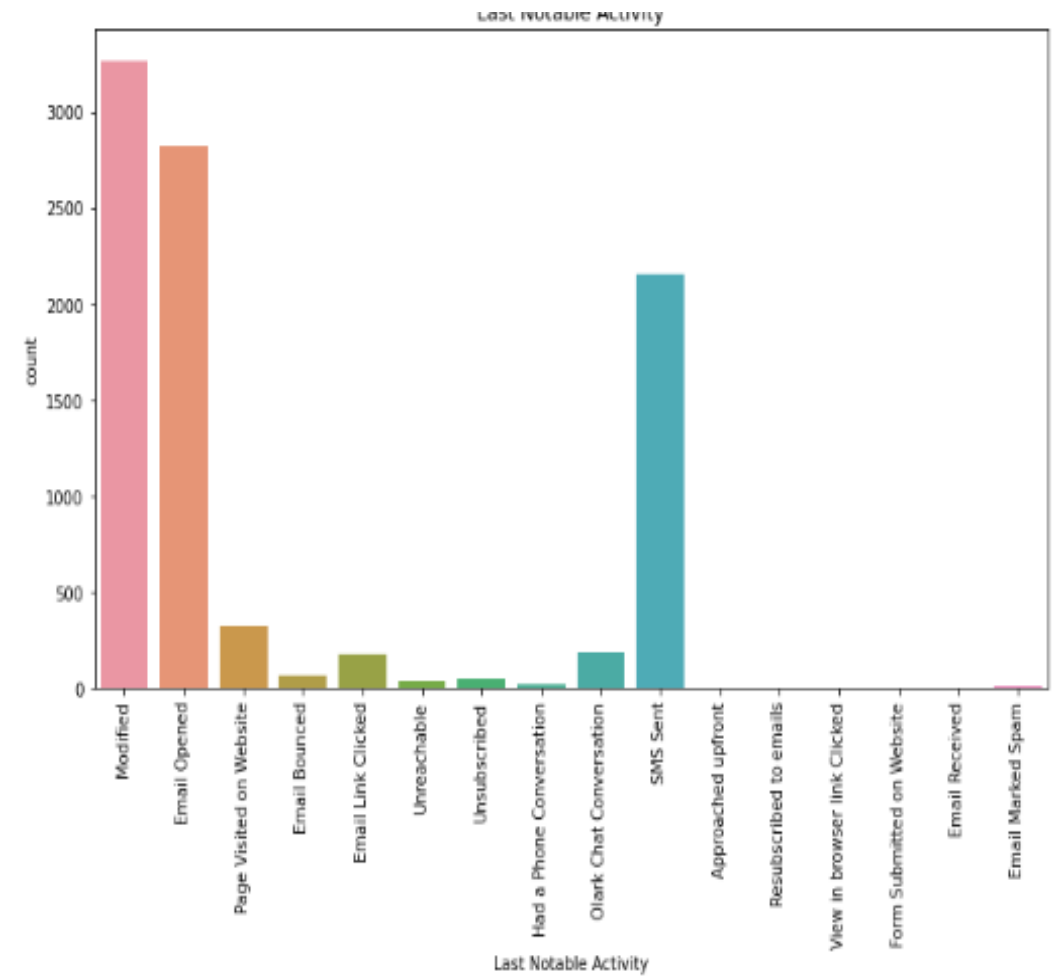
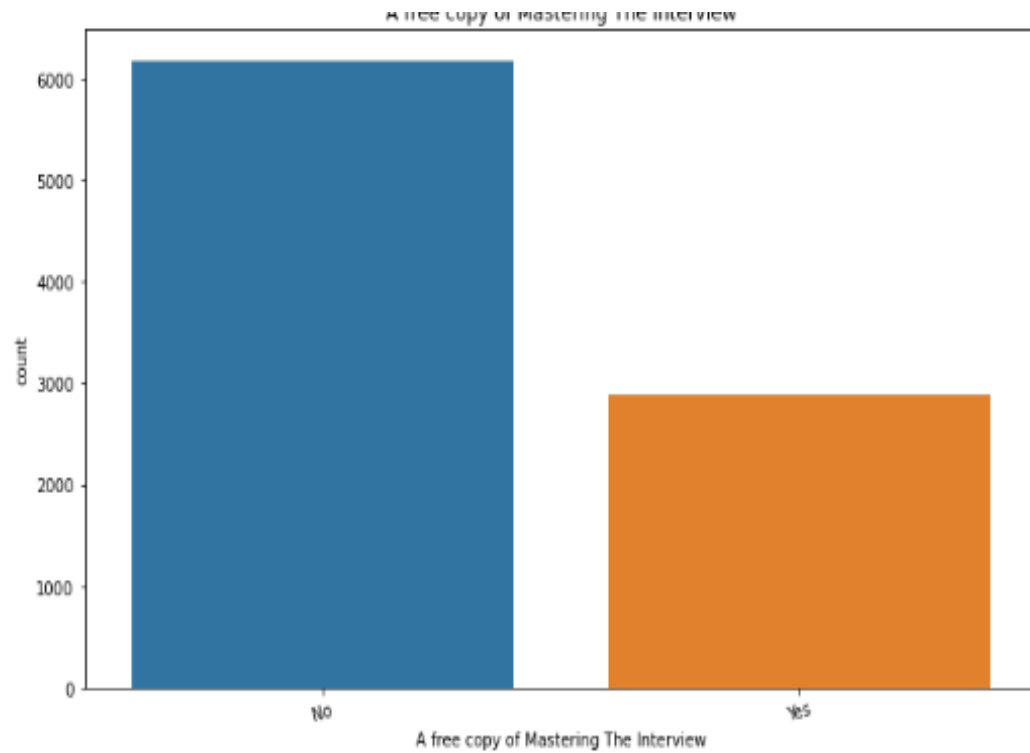
# Data Cleaning and Preparation

- Dataset contains rows: 9240 columns: 37
- Remove Columns which contains null values greater than 35 % and which contains single value for all datapoints.
- Visualization of Cleaned Data.
  - Univariate Analysis
  - Bivariate Analysis
- Dummy Variable are created for object type variable
- Split data in Train and Test ratio used is 70:30.
- Numeric variables are Normalized

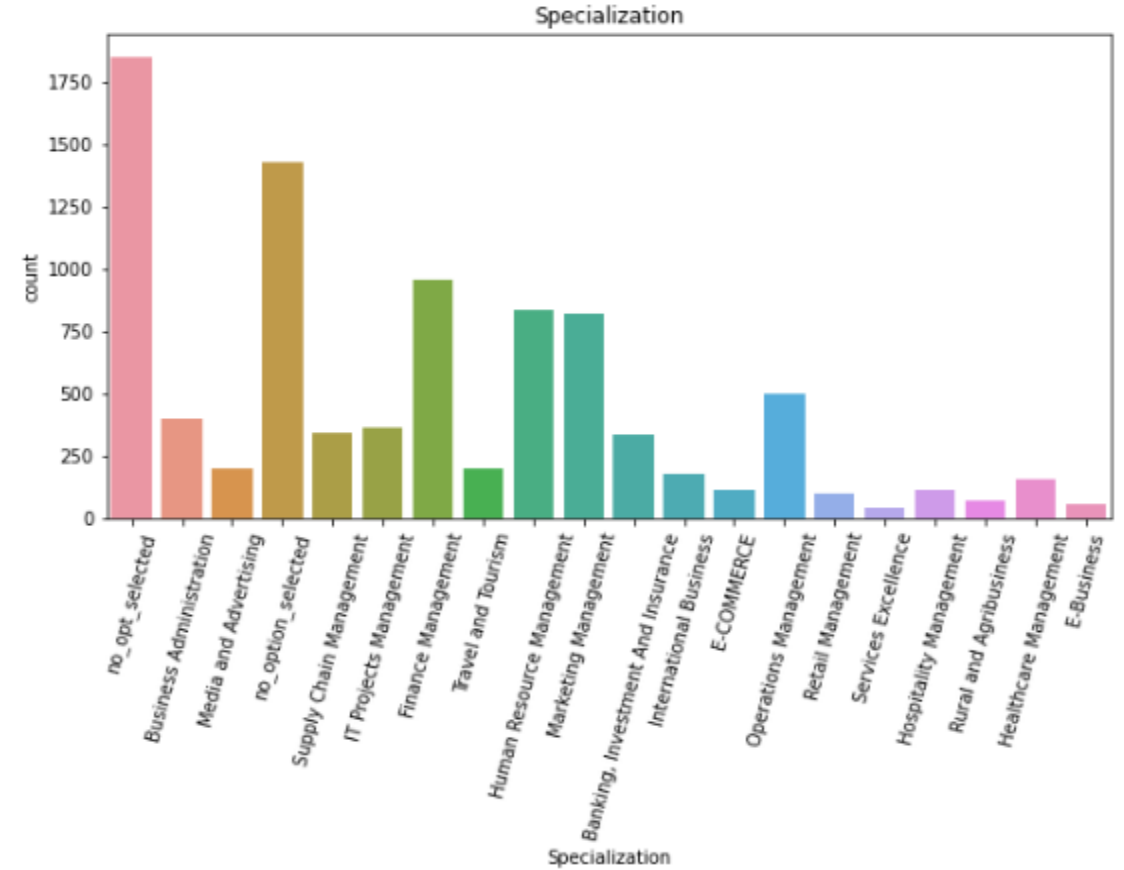
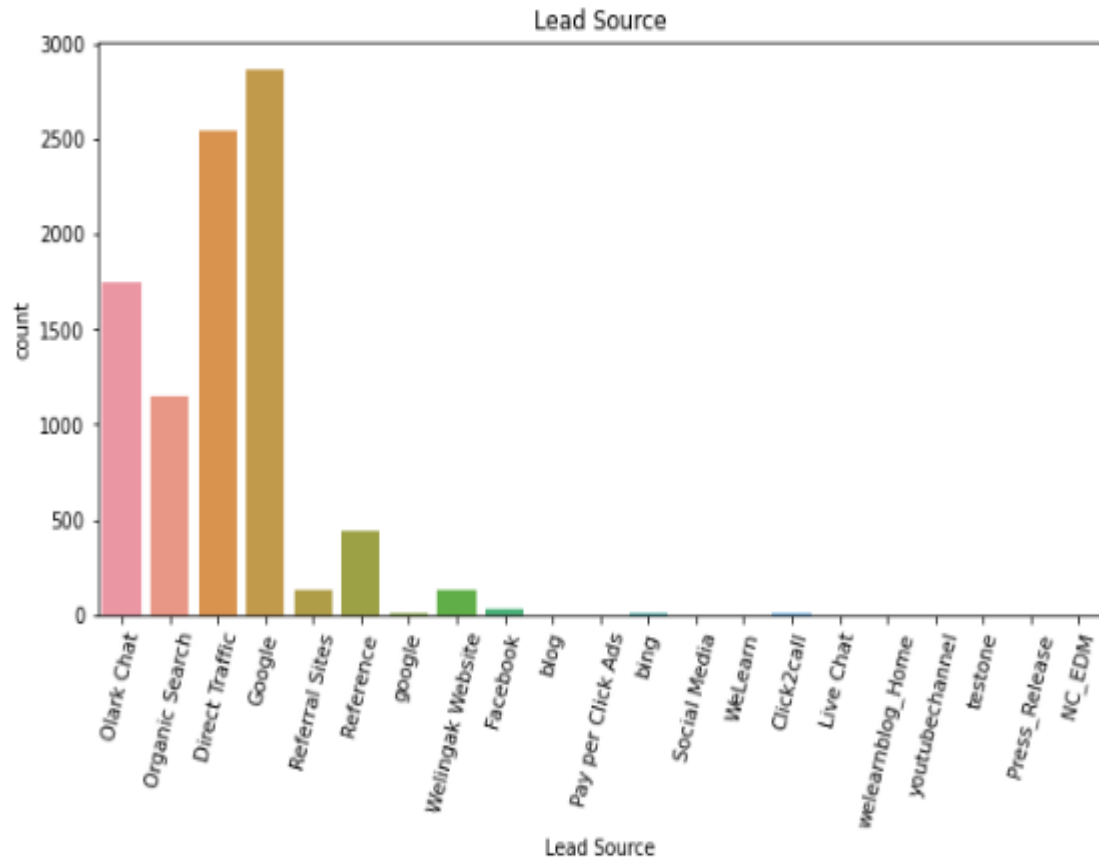
# Univariant Analysis



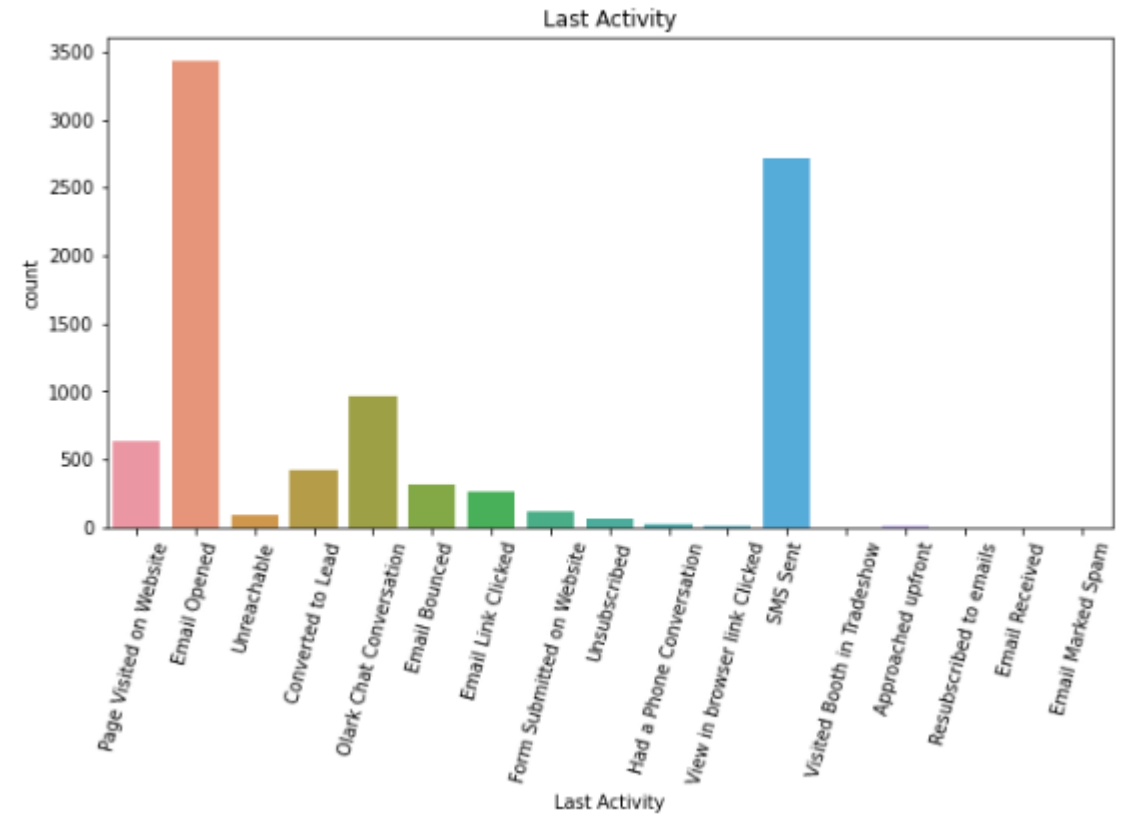
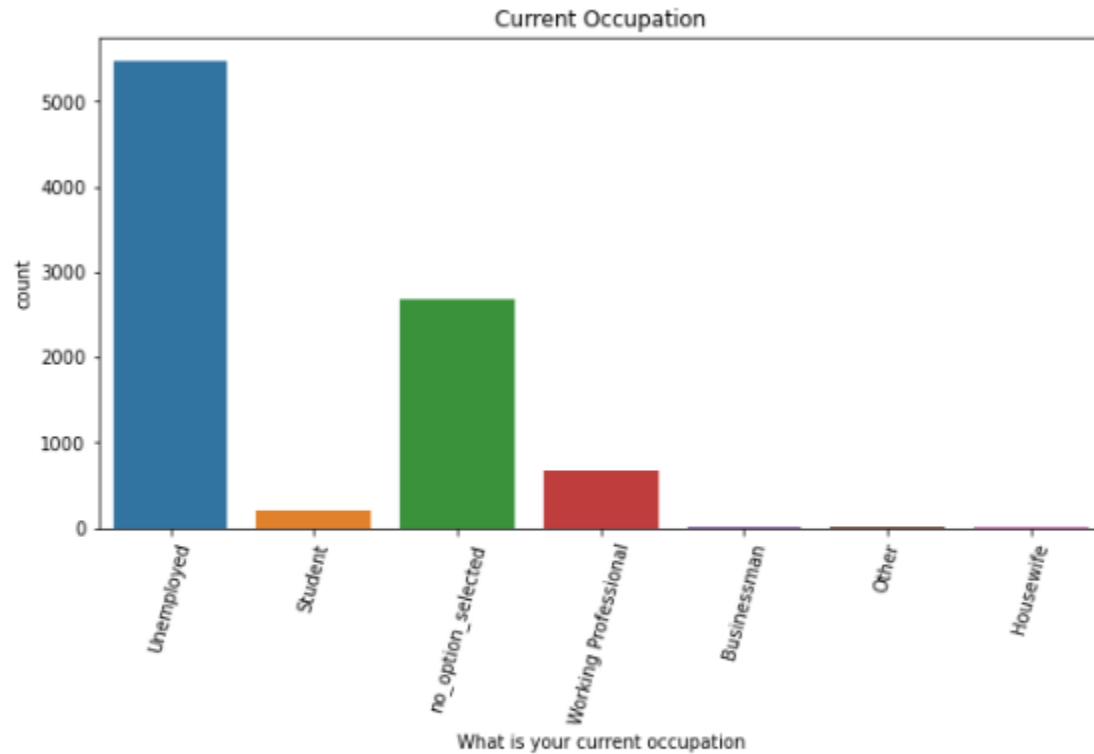
# Univariate Analysis



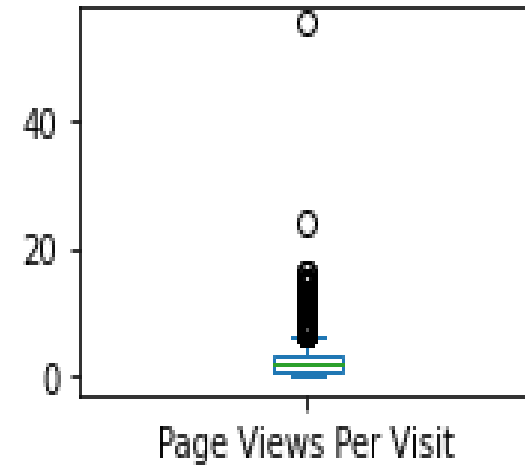
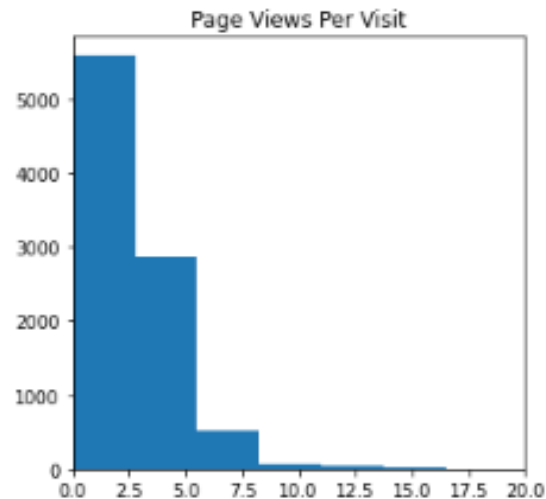
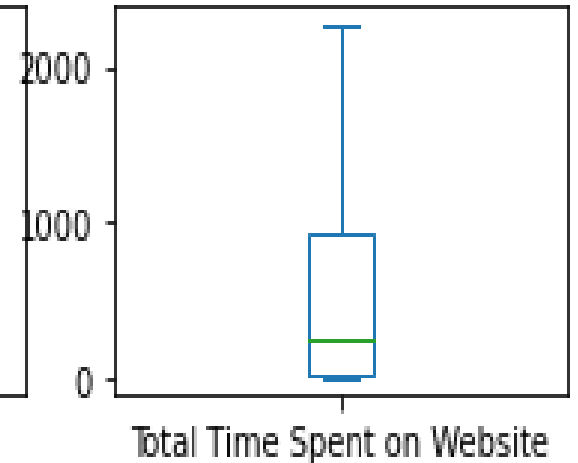
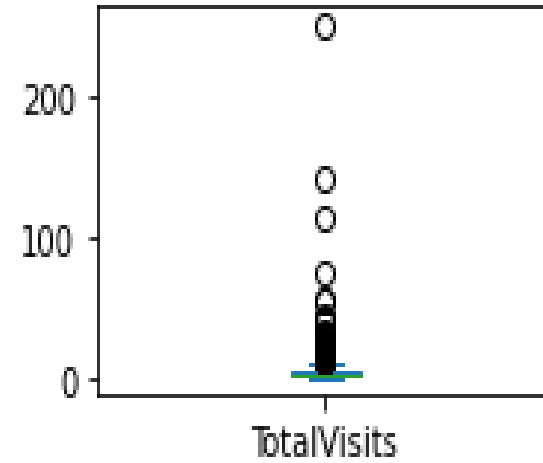
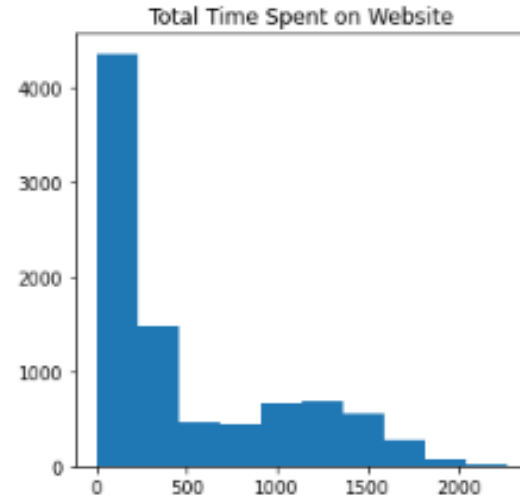
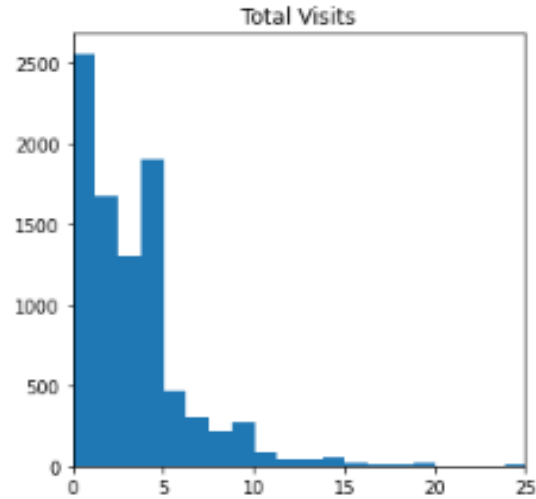
# Univariant Analysis



# Univariant Analysis

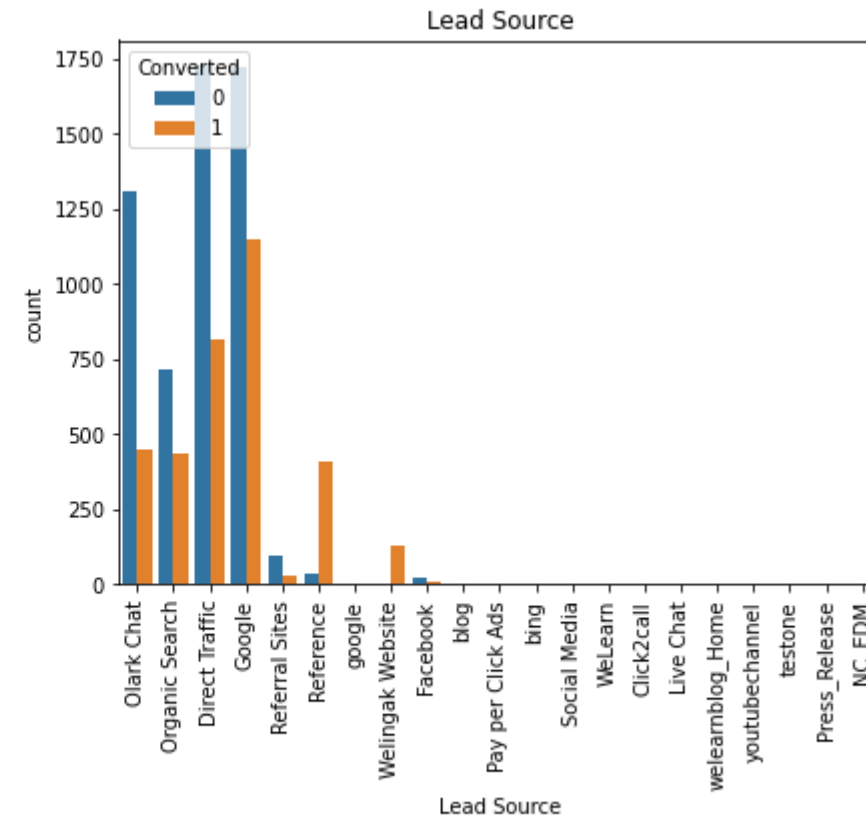
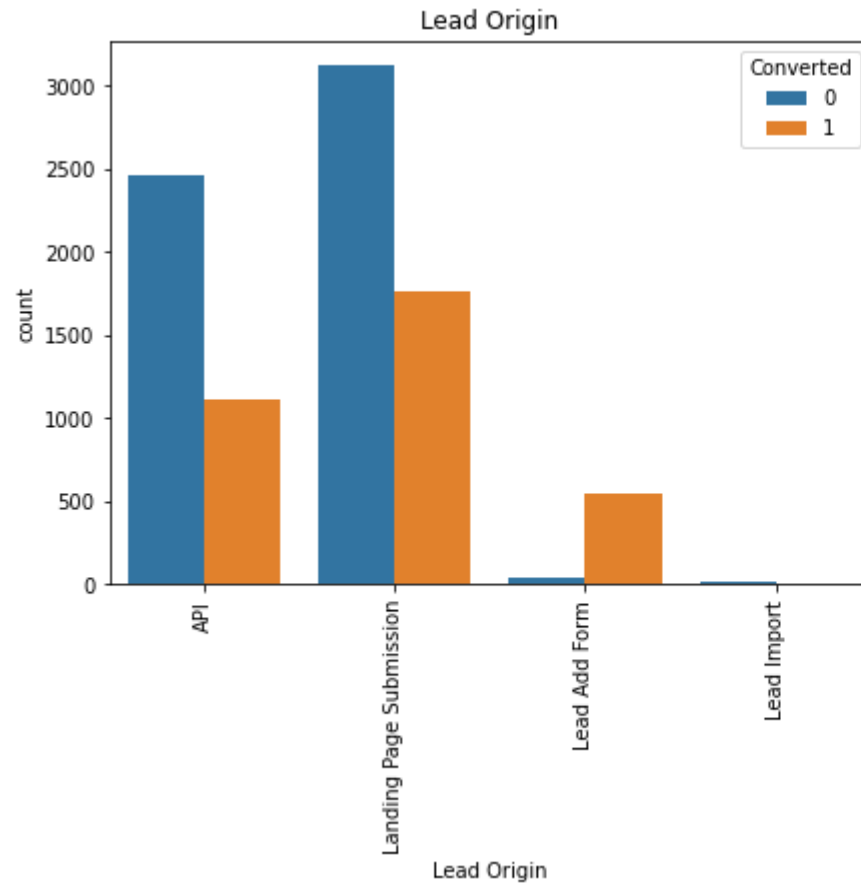


# Numerical Variable Univariant Analysis & Outliers

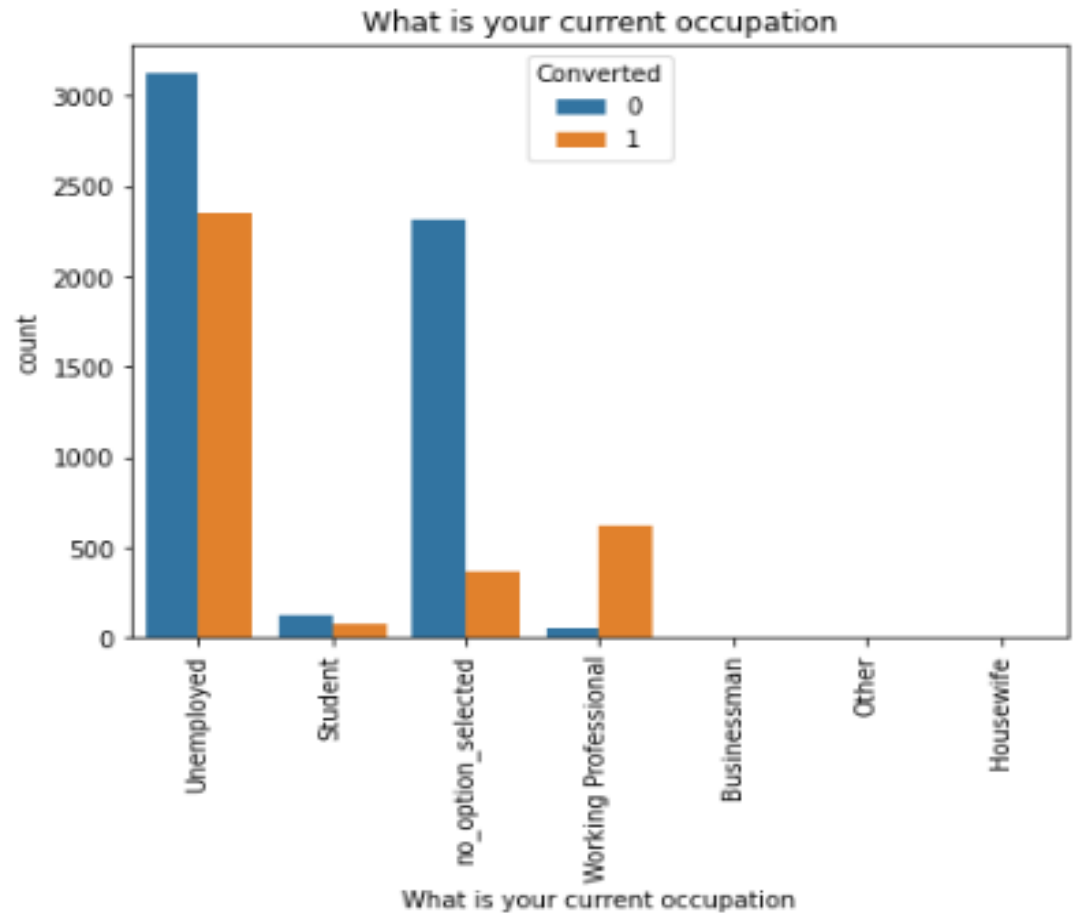
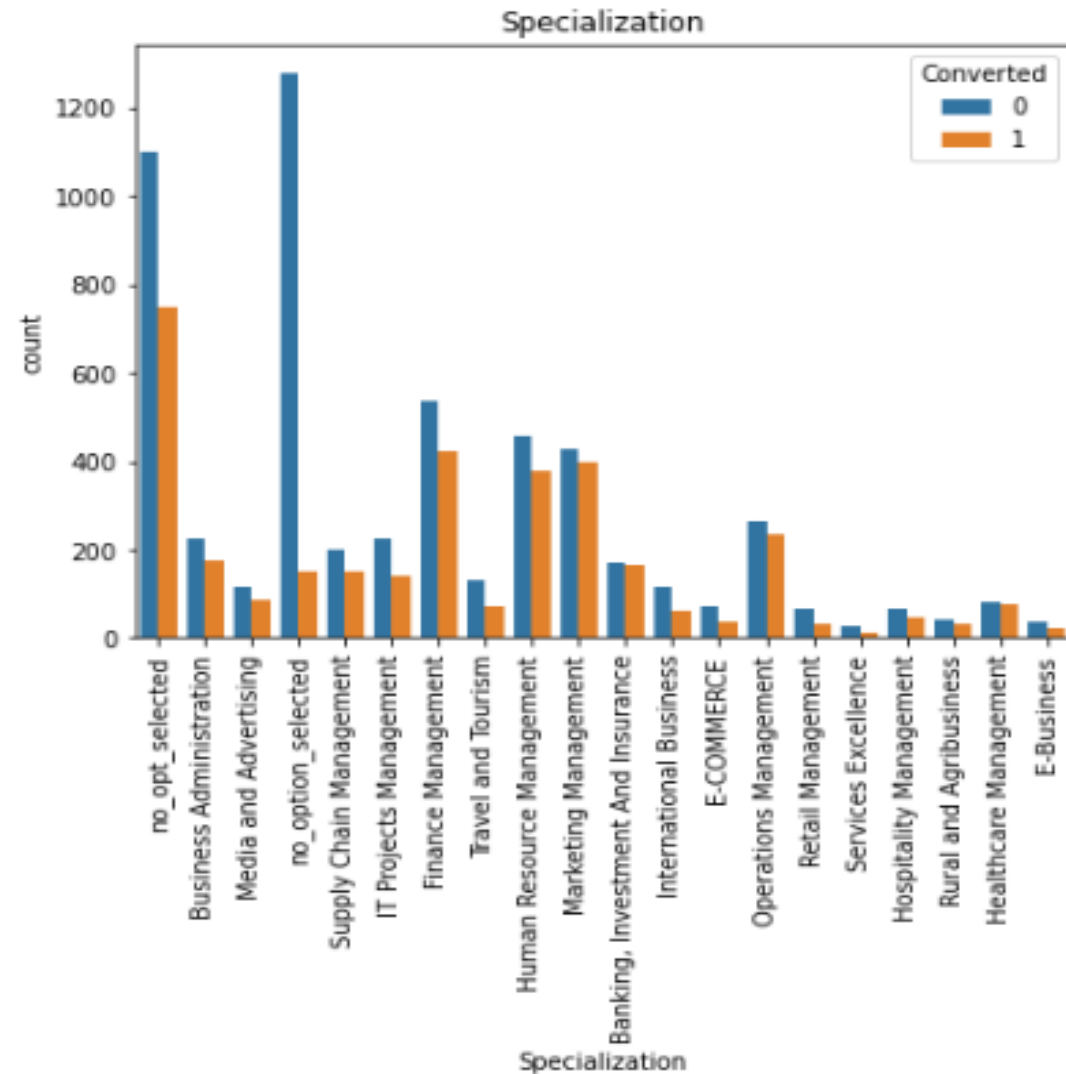




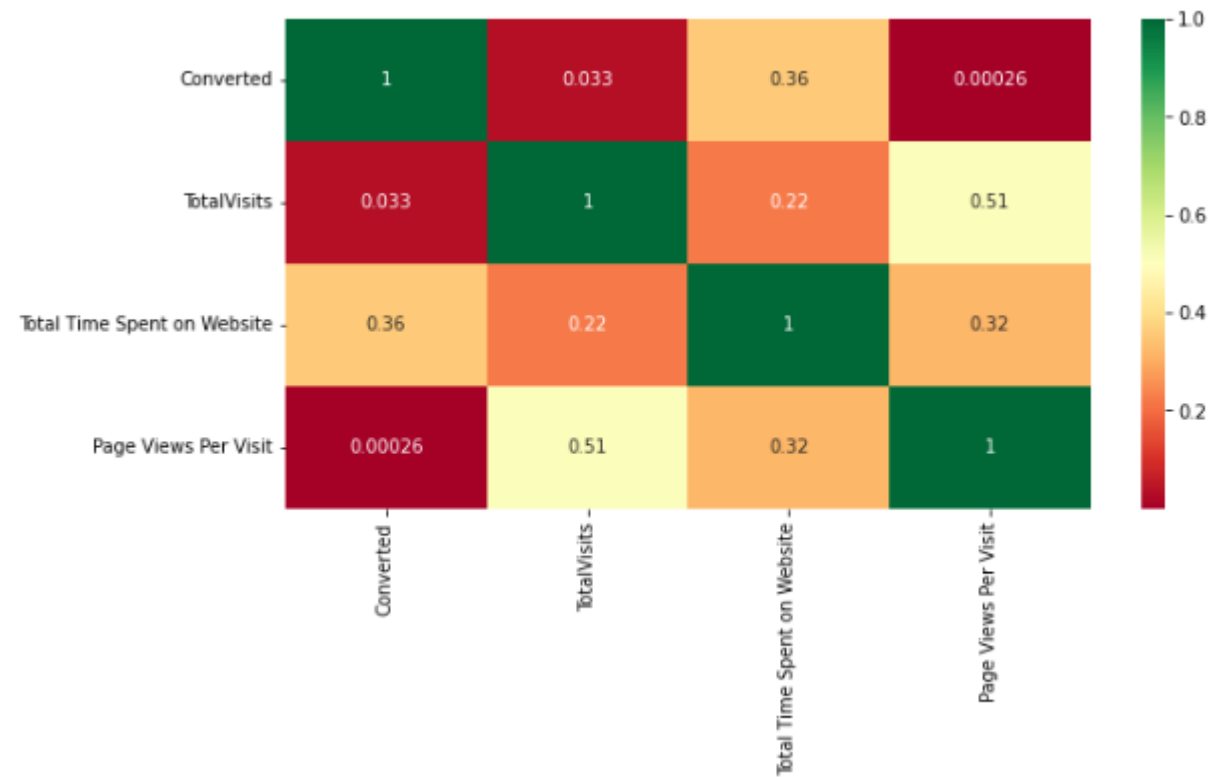
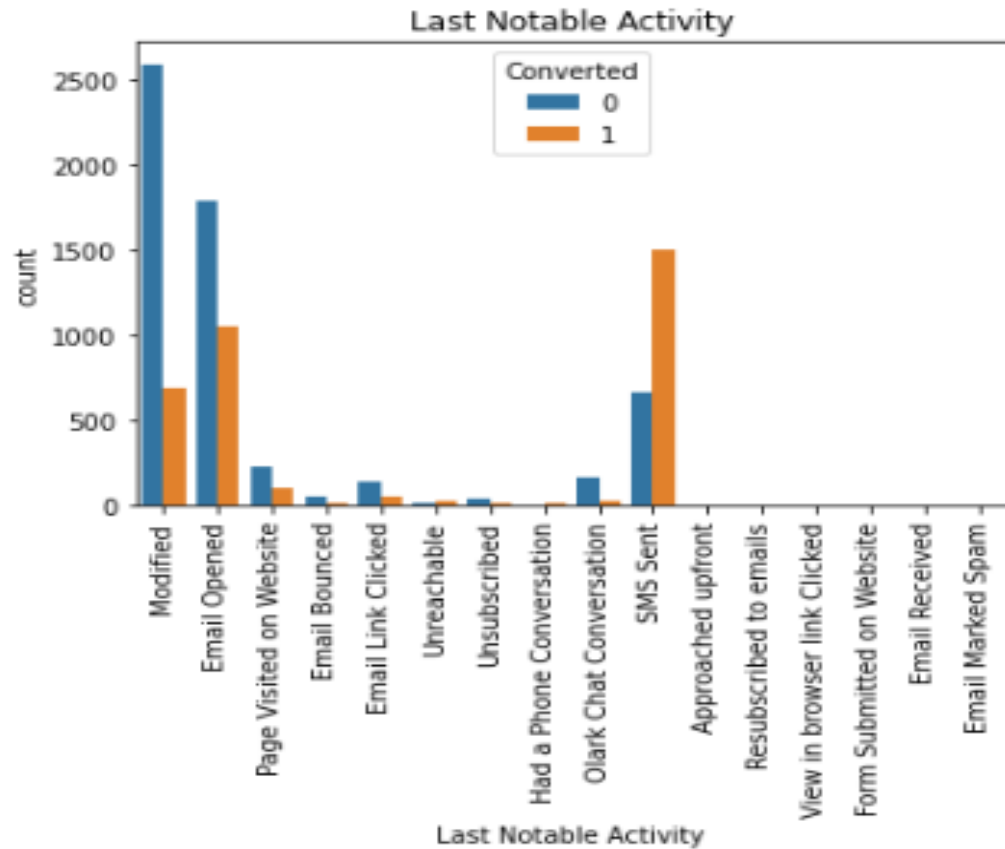
# Bivariant Analysis



# Bivariant Analysis



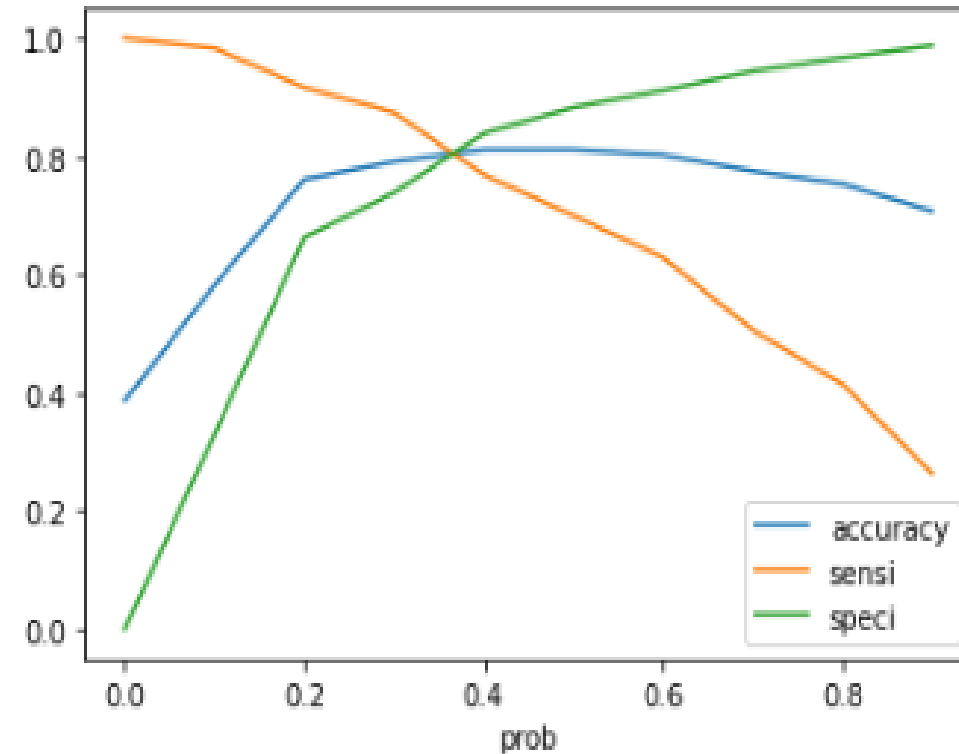
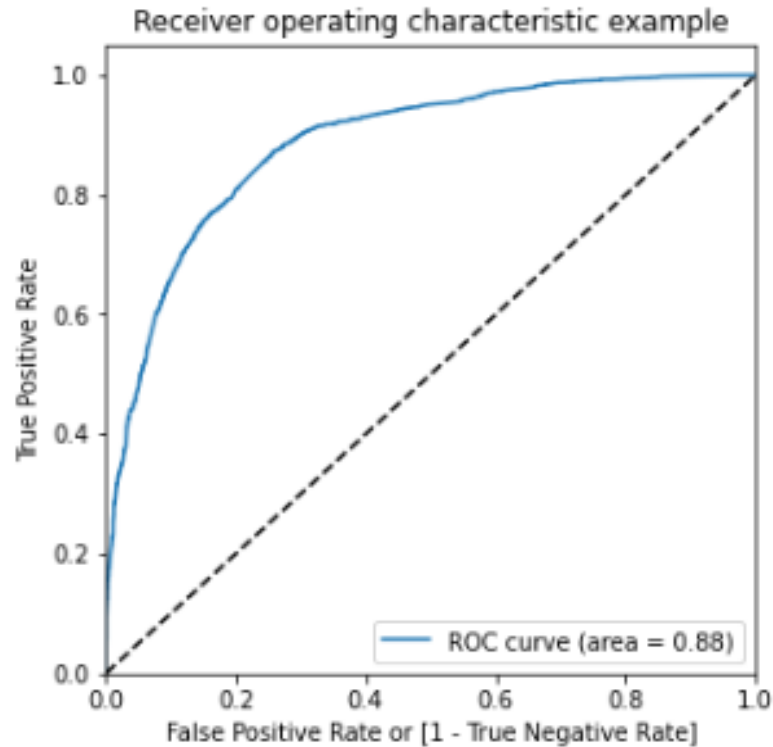
# Bivariant Analysis and Correlation Matrix



# Model Building

- 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 81%

# ROC Curve and Optimal Cutoff



- From second graph it is visible that the optimal cut off is at 0.35.