

Leads scoring - Case study

- ▶ Group Members
 - ▶ - Vamsi Sure
 - ▶ - Arpan Majumder
 - ▶ - Mohit Tanga

Problem Statement

Solution Methodology

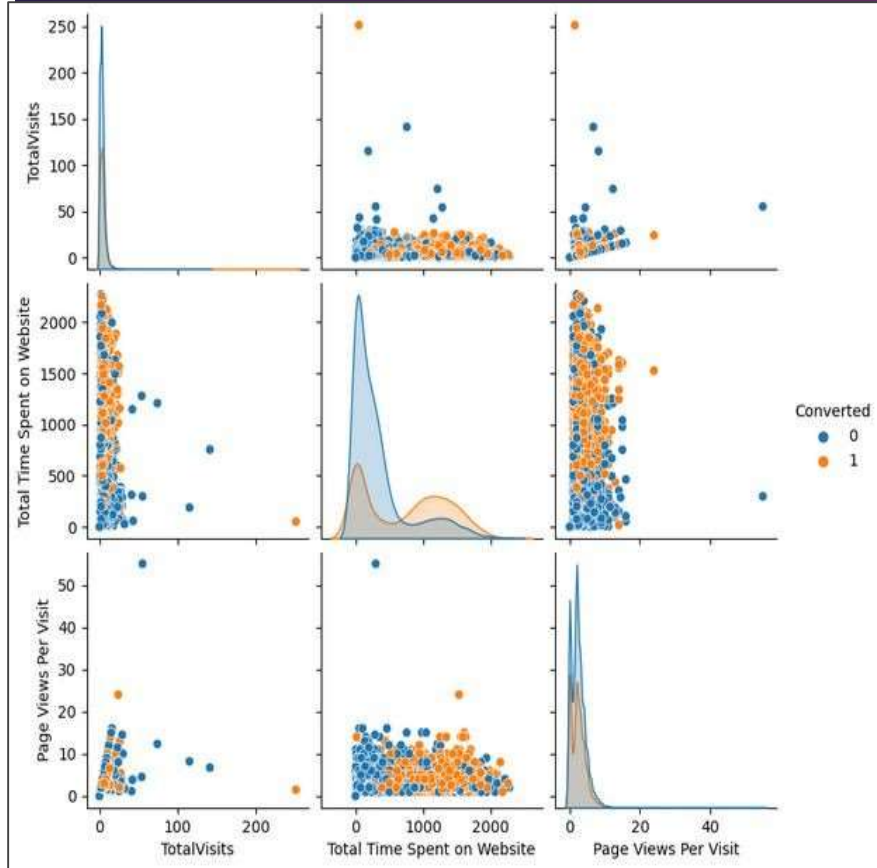
Data cleaning

Data imputation/manipulation

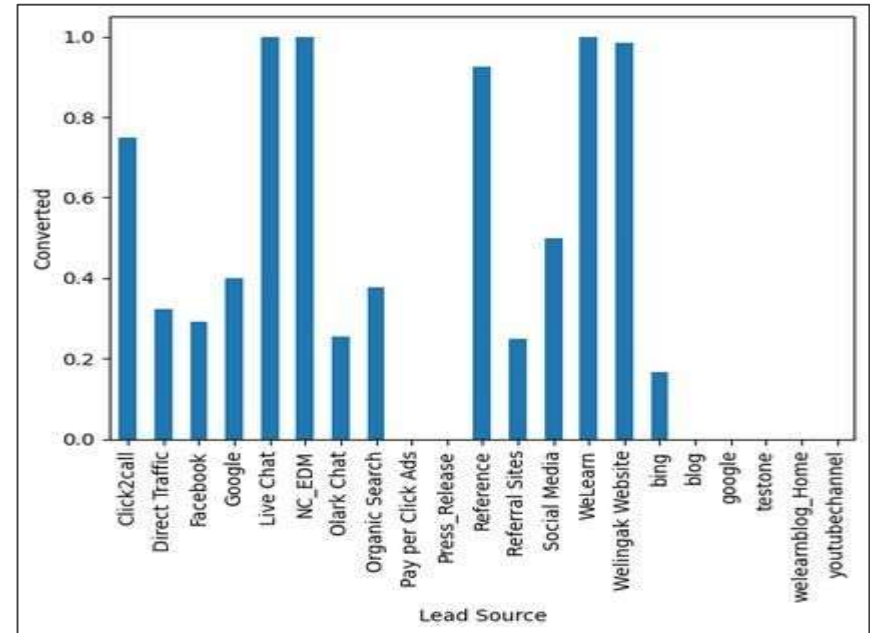
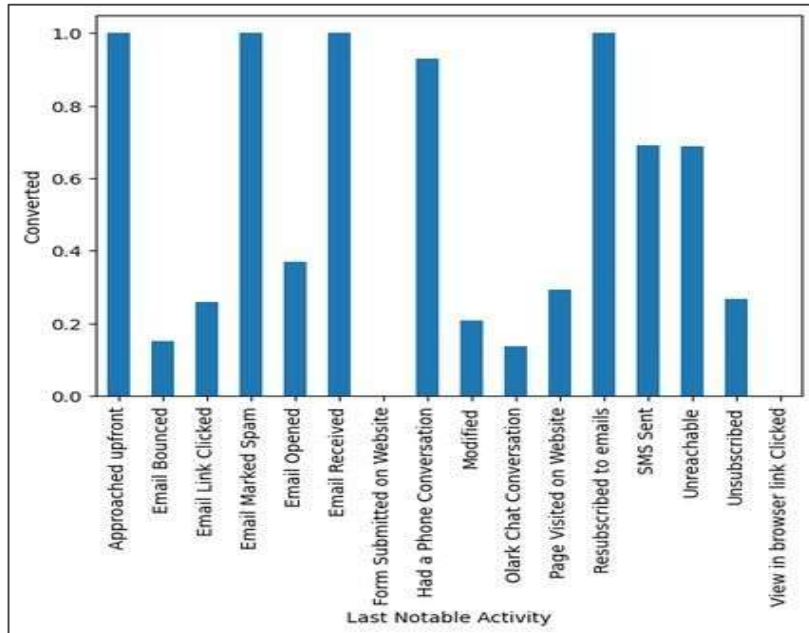
Imputing the missing values in 'Lead Quality', 'Country', 'Specialization', 'How did you hear about us'

EDA

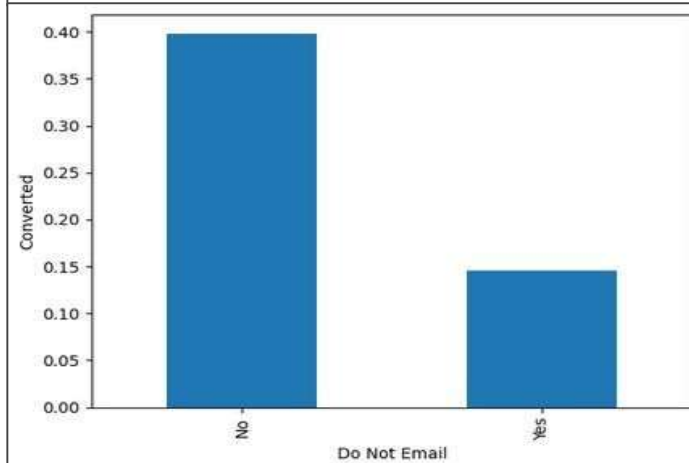
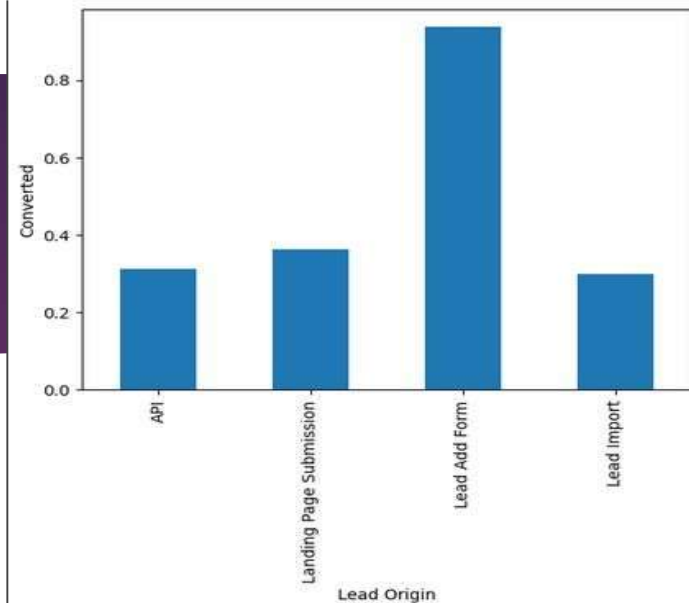
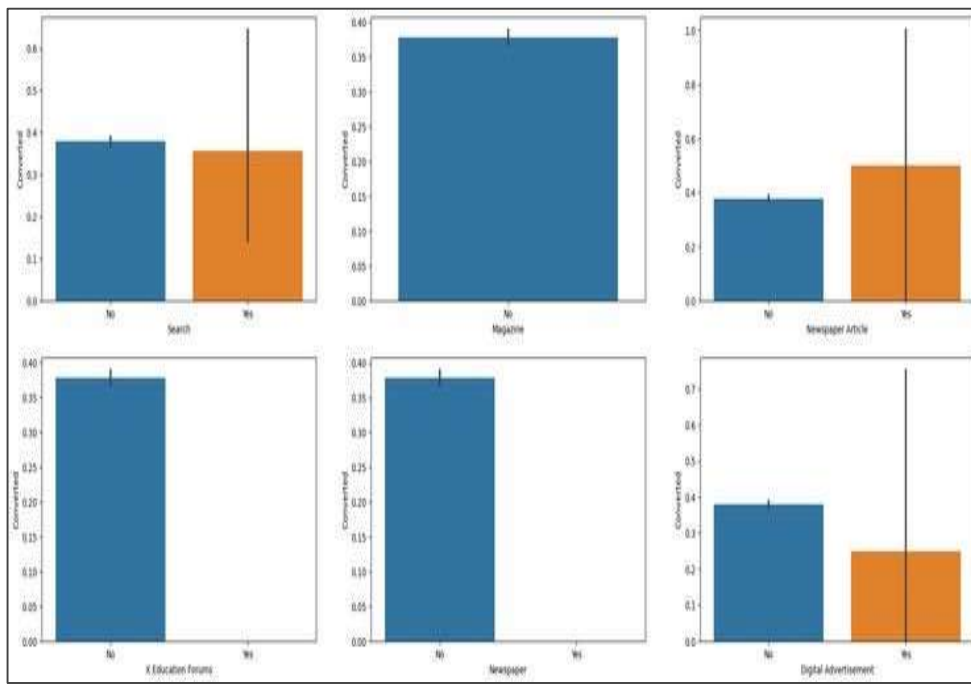
EDA on numerical



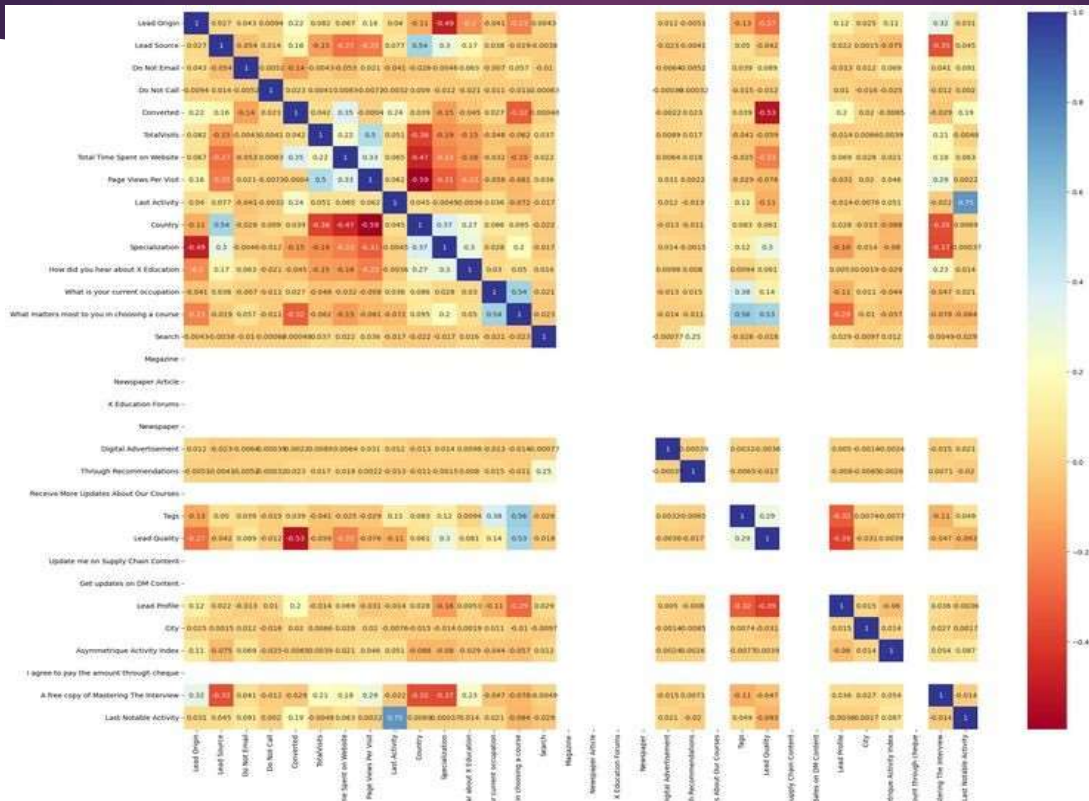
EDA on categorical variables #1



EDA Categorical variable - #2



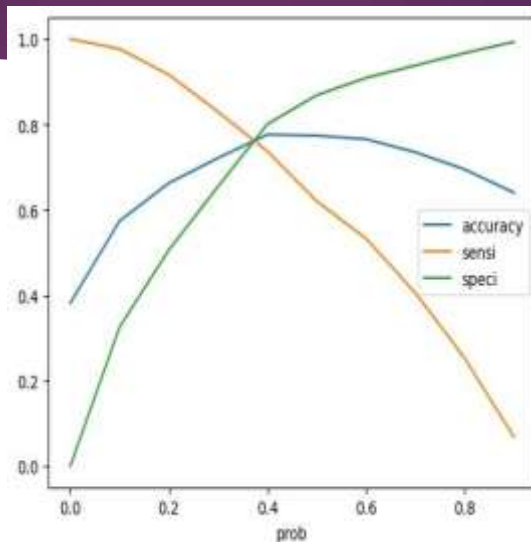
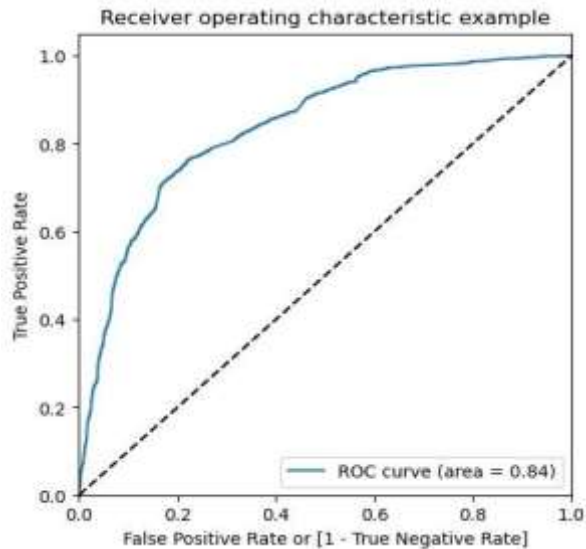
Correlation Matrix



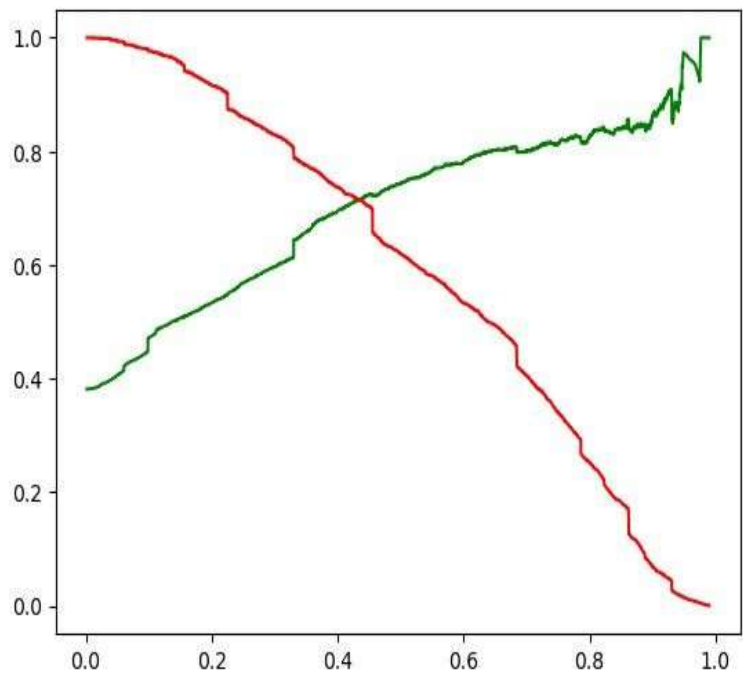
Model building

- Splitting the dataset into train and test sets
- Splitting the train data into X and y variables, where x contains all variables and y contains the target variable
- Conducting RFE on the X_train and y_train dataframes - RFE has selected 15 columns
- Feature elimination based on high VIFs and P-values - Cutoff 0.5
- With the current cut off of 0.5 VIF and 77% Accuracy. The sensitivity stands at 62% and specificity metric at 87%

ROC curve



Precision-Recall Tradeoff



Conclusion

