Lead Scoring Case Study

Business Objective

- To help X Education to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- To build a logistic regression model to assign a lead score value between 0 to 100 to each of the leads which can be used by the company to target potential leads.

Data Cleaning

Handling Null Values :

The columns with more than 40% of null values were dropped

- How did you hear about X Education (78.46%)
- Lead Profile(74.18%)
- Lead Quality (51.60%)
- Asymmetrique Profile Score, Asymmetrique Activity Score, Asymmetrique Activity Index, Asymmetrique Profile Index (45% for all four)

Data Cleaning

Dropping redundant columns :

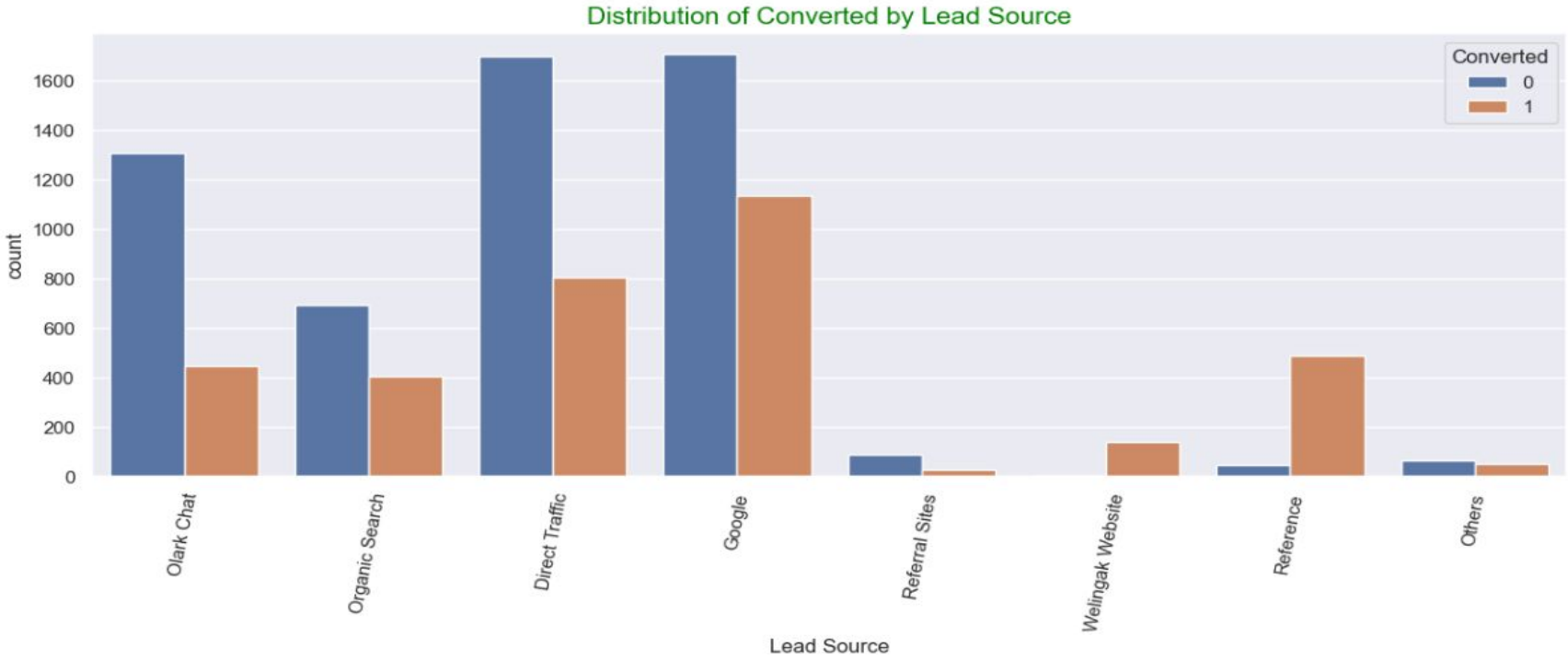
The columns with redundant informations were dropped

● Updates ● Last_activity ● Prospect_ID

Imputing missing values :

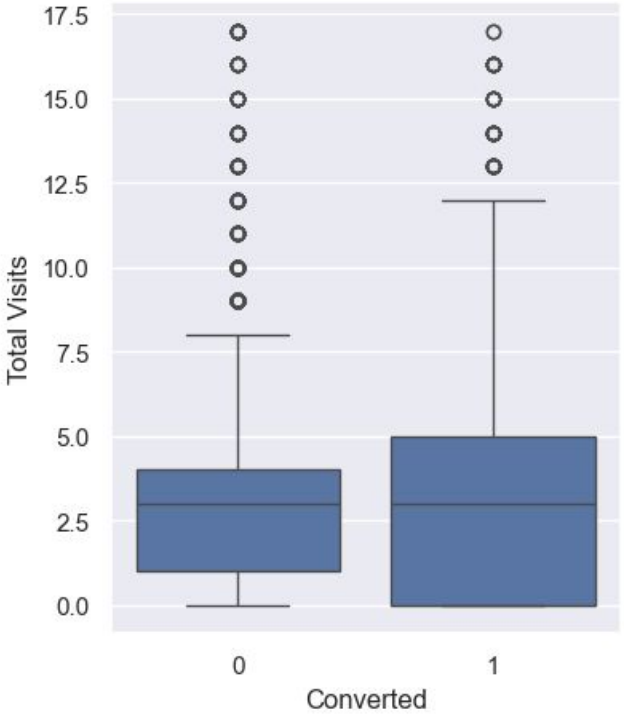
- The numerical columns are imputed with median and categorical columns are imputed with mode

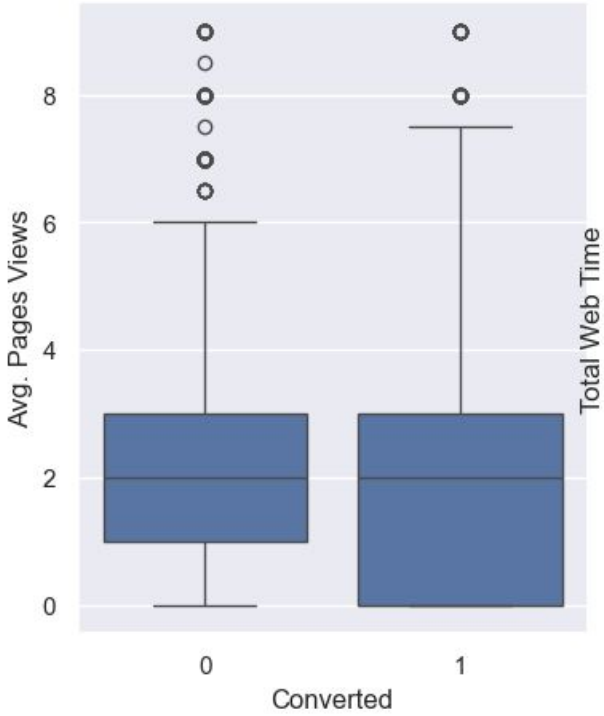
Univariate Analysis

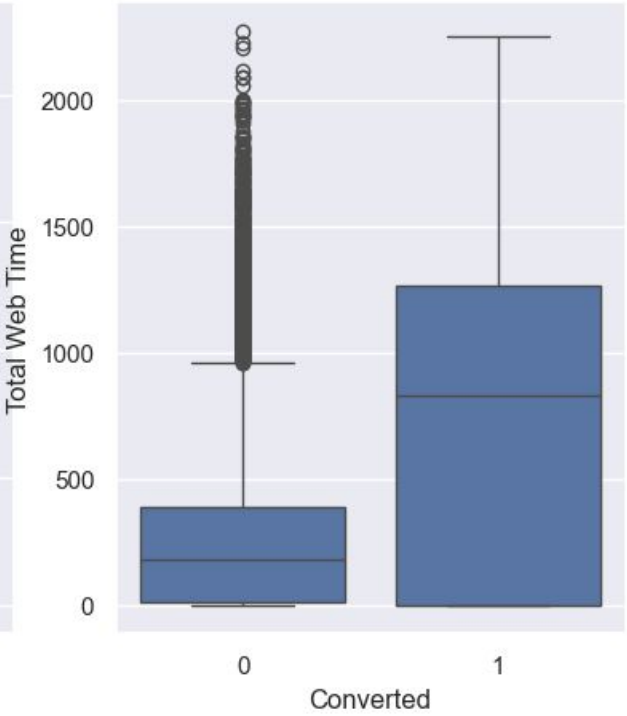


Leads sourced from Google or Direct Traffic will have high conversion rate

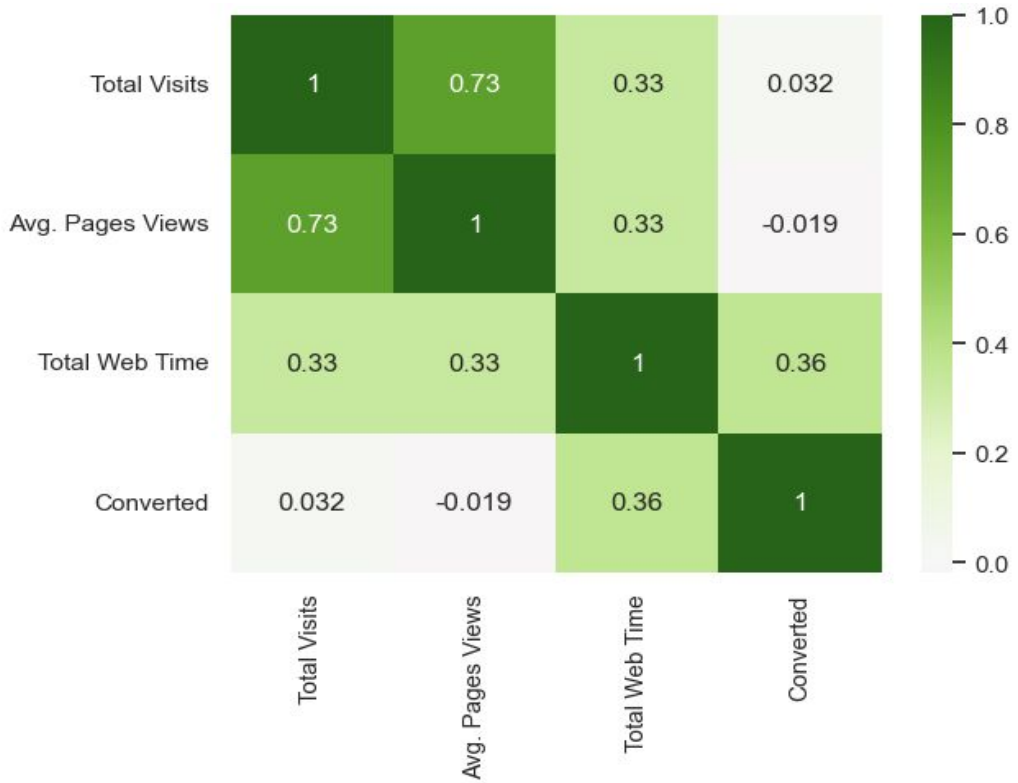
Bi-variate Analysis





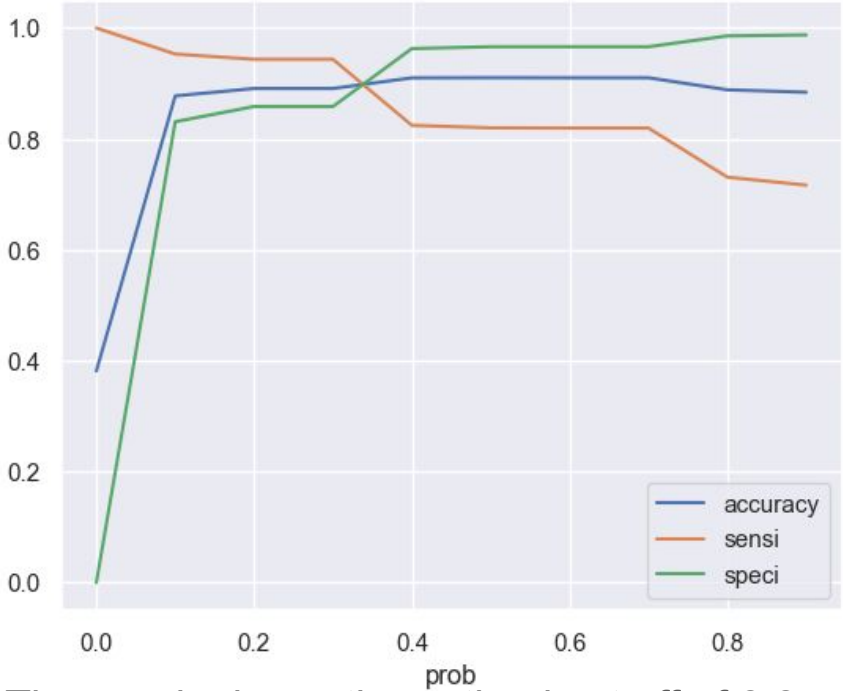


Correlation

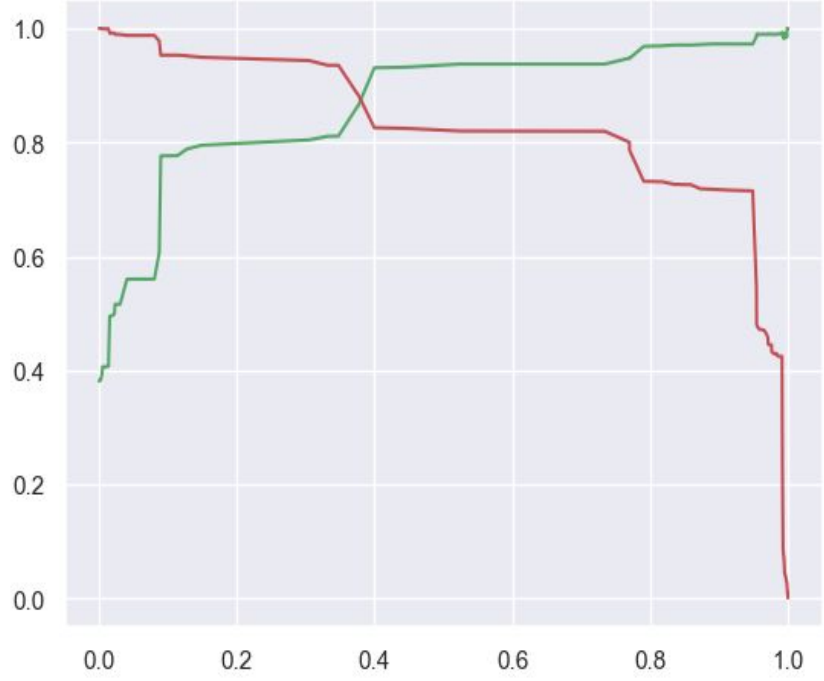


- There is high positive correlation between Total Vists and Pages viewed per visit.
- Page views is negatively correlated with Converted

Model Evaluation - Train Dataset



The graph shows the optimal cut off of 0.3 based on Accuracy, Sensitivity and Specificity.



The graph shows an optimal cut off of 0.4 based on Precision and Recall