Project Title:
Lead Conversion Rate Optimization using Logistic Regression Model

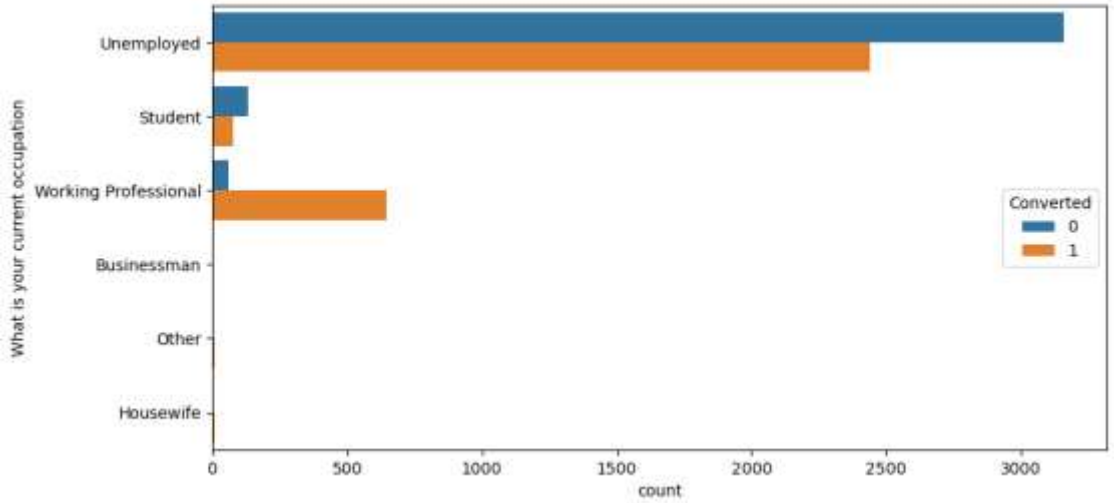
Problem Statement

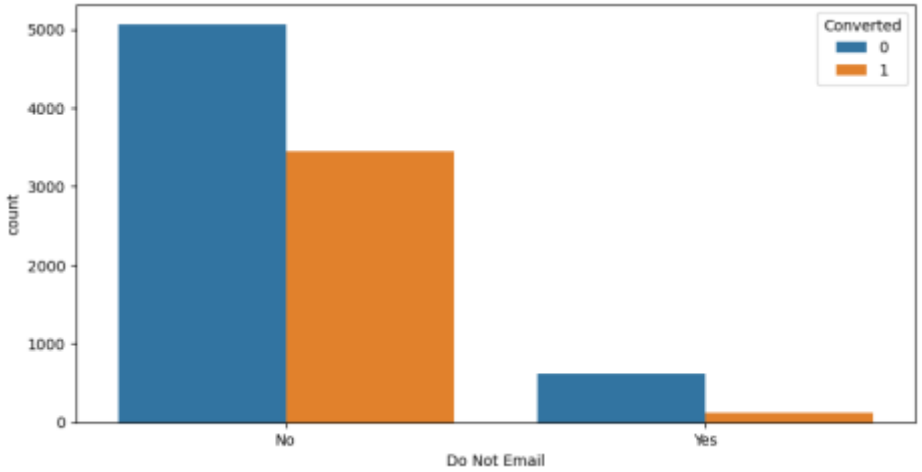
- X Education, an online education company, aims to improve its lead conversion rate. Despite receiving a significant number of leads from professionals interested in their courses through website visits, form submissions, and referrals, the company faces a low conversion rate of around 30%. To address this issue, X Education wants to identify the most promising leads, known as 'Hot Leads,' to allocate resources more effectively. By distinguishing these high-potential leads from the rest, the company expects to increase its lead conversion rate. This will enable the sales team to concentrate on communicating with the potential leads, resulting in a more efficient and targeted conversion process.

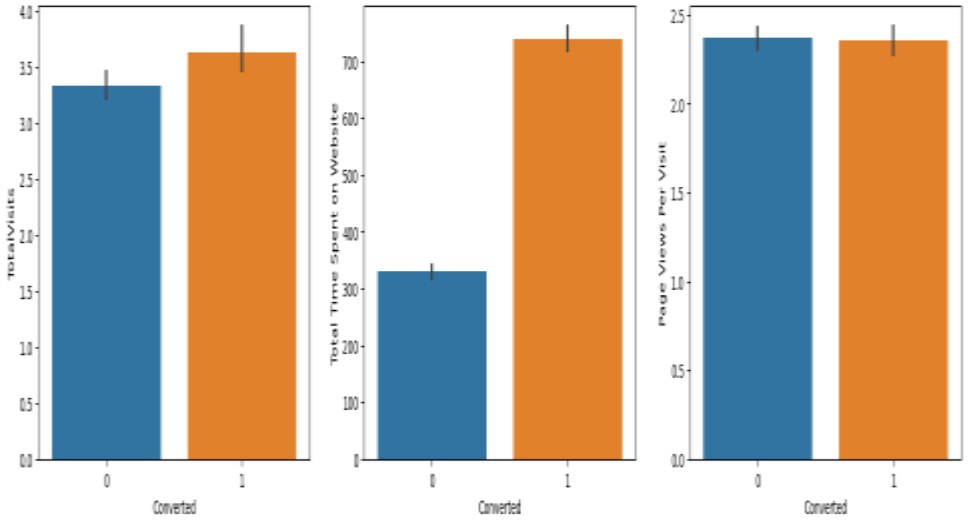
Approach

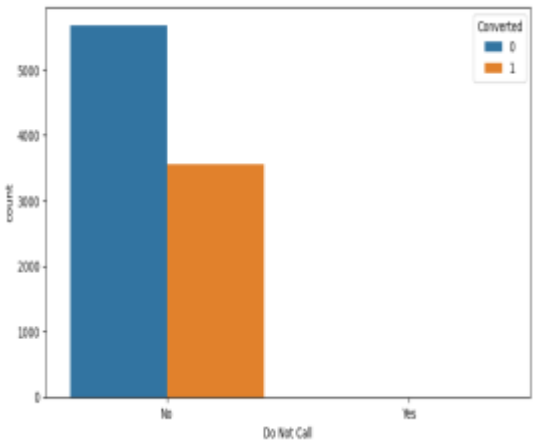
1. Source and clean data for Analysis
2. Data Cleaning and Preparation
3. EDA
4. Feature Scaling
5. Splitting data into train and test sets
6. Building a Logistic Regression model
7. Evaluating the model’s efficacy using metrics like Specificity and Sensitivity or Precision and Recall
8. Based on metrics choosing the best model for production

Initial Findings









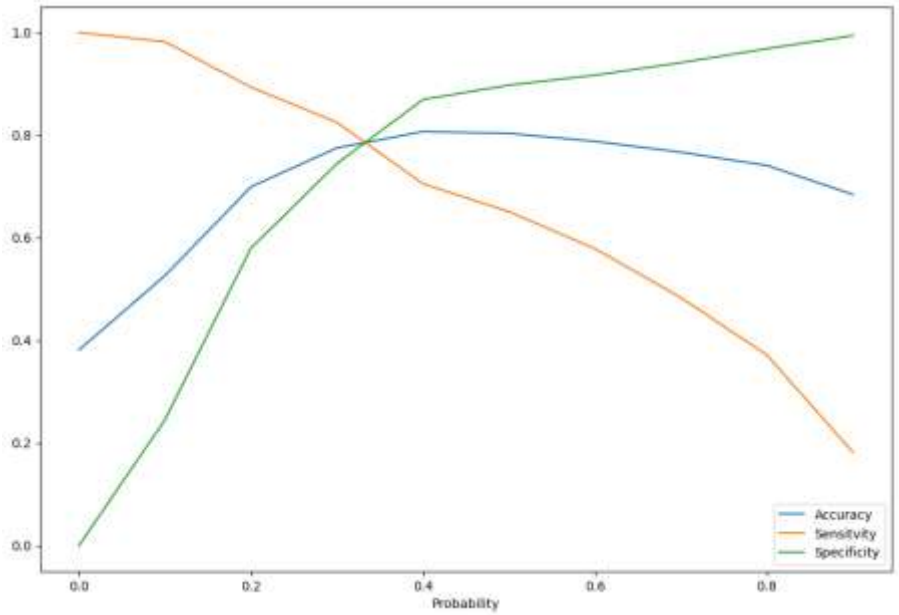
- Maximum lead conversion happened from Landing Page Submission.
- Major lead conversion in the lead source is from 'Google'
- Major lead conversion is from the Unemployed Group
- Major lead conversion from TotalVisits, Total Time Spent on Website, Page Views Per Visit
- Major conversion has happened from the emails that have been sent

Factors that Impact the conversion rate

The following features have a high tendency to estimate the conversion of a lead

- Total Visits
- Time Spent on Websites
- Lead Origin
- Lead Source
- Last Activity
- Current Occupation
- Last Notable Activity

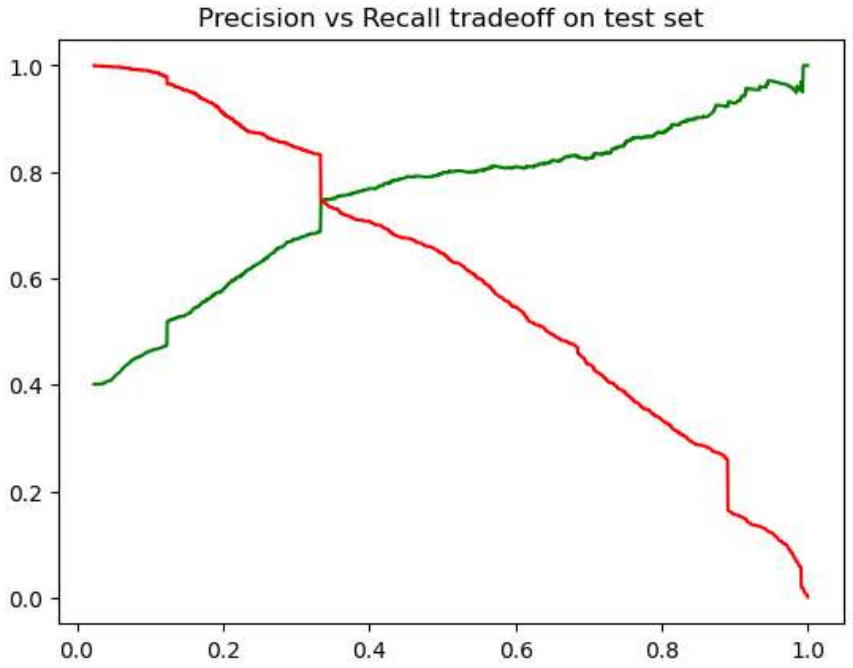
Model Evaluation – Sensitivity and Specificity on train set



1228	449
169	926

Sensitivity: 82.48
Specificity: 74.46
Precision: 66.56
Recall: 82.48
Accuracy: 77.52

Model Evaluation – Precision and Recall on test set



Sensitivity: 84.57
Specificity: 73.23
Precision: 67.35
Recall: 84.57
Accuracy: 77.71

Lead Scoring based on Conversion Probability

- Based on the probability of conversion the leads are ranked
- Higher the score, hotter the leads
- The business is recommended to approach the leads with higher score to attain better conversions

Conclusions

- The Sensitivity and Specificity, Accuracy, Precision and Recall score we got from test set are similar to the train set
- We have high recall score than precision score which is a sign of good model.
- In business terms, this model has an ability to adjust with the company’s requirements in coming future.
- This concludes that the model is in stable state.
- Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :
    - Lead Origin_Lead Add Form
    - Total Time Spent on Website
    - What is your current occupation_Working Professional