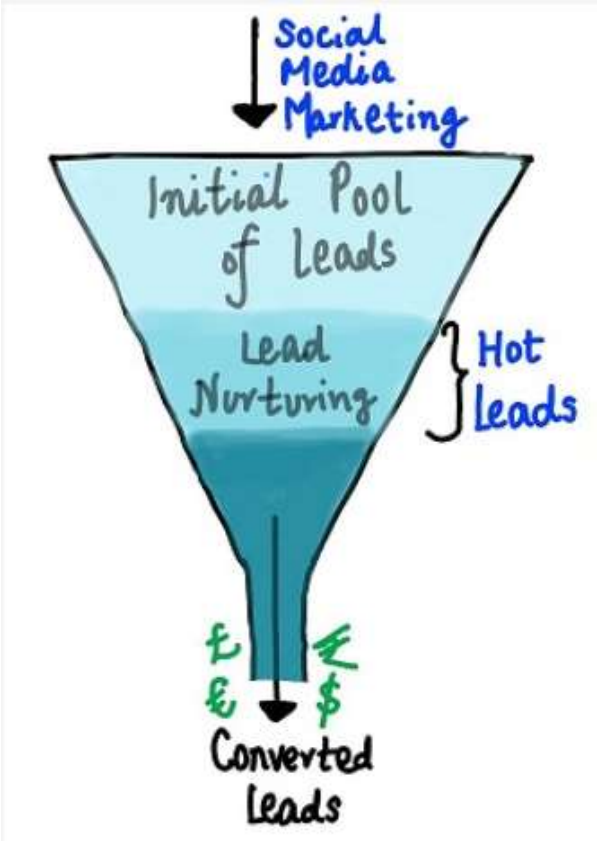
Lead Scoring Case Study

Presentation By :
Bharat Gupta Paluri



Lead Conversion Process - Demonstrated as a funnel

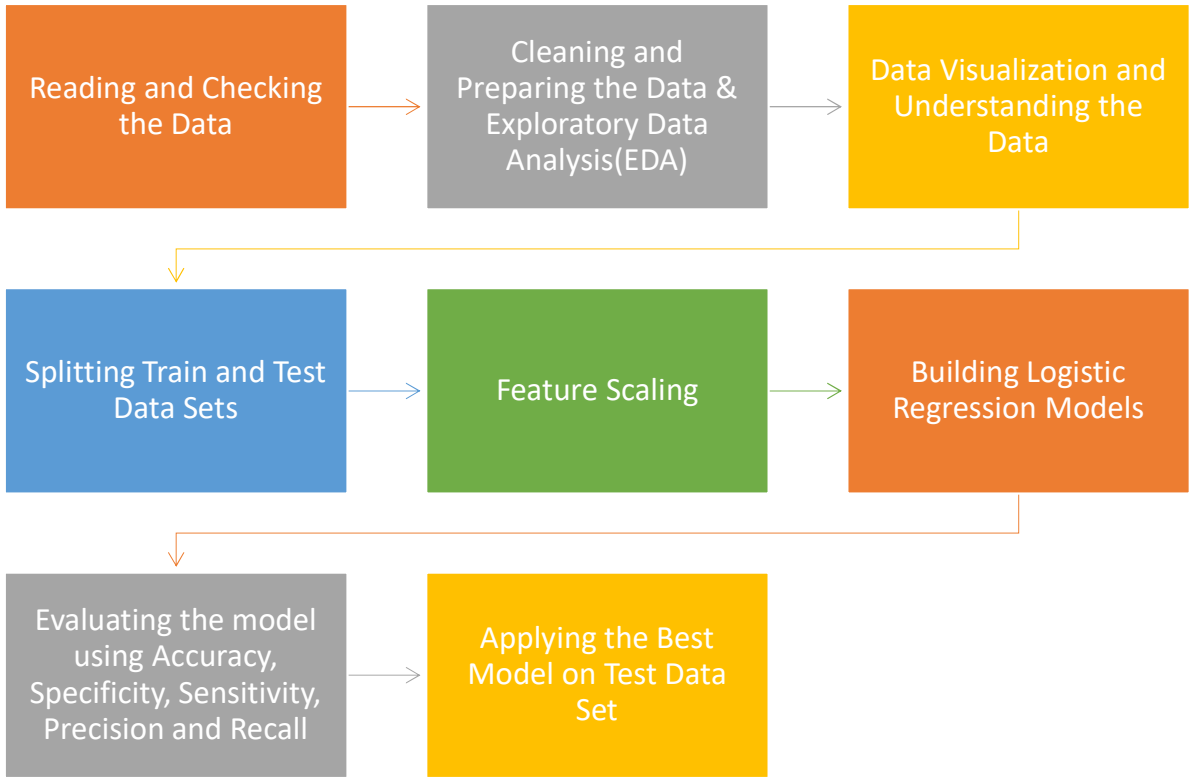
Problem Statement :

- X Education sells online courses to industry professionals, the company wishes to identify the most potential leads.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

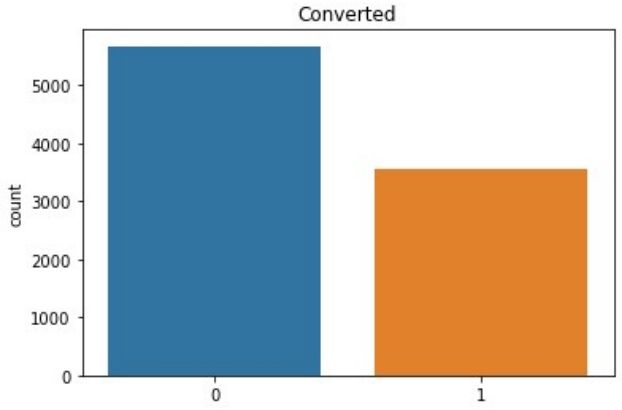
Business Goal :

- The company wants to build a model with lead scores assigned to it, the Higher the lead score means higher chances of conversion and vice versa.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%.

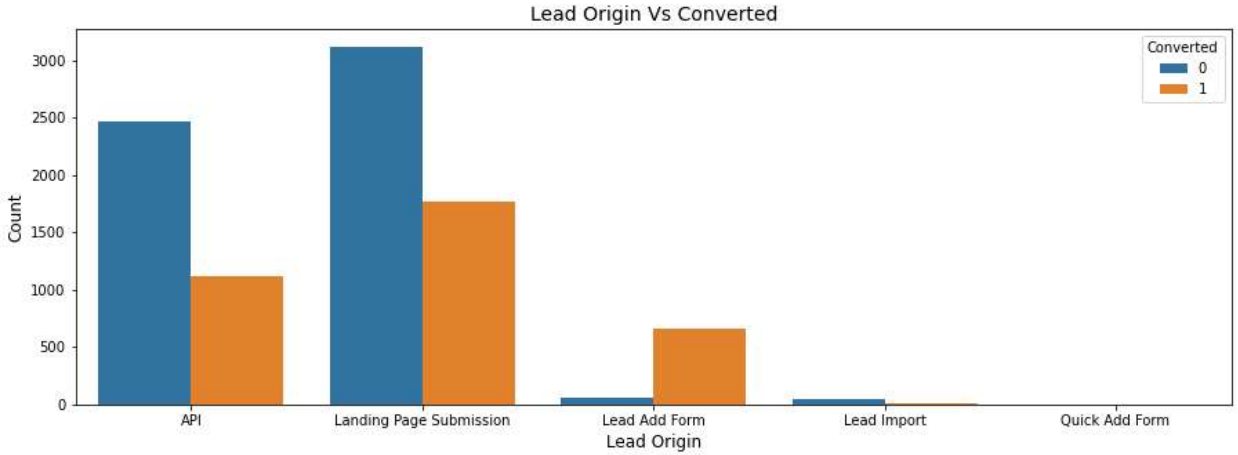
Strategy


Reading and Checking the Data
Cleaning and Preparing the Data & Exploratory Data Analysis(EDA)
Data Visualization and Understanding the Data
Splitting Train and Test Data Sets
Feature Scaling
Building Logistic Regression Models
Evaluating the model using Accuracy, Specificity, Sensitivity, Precision and Recall
Applying the Best Model on Test Data Set

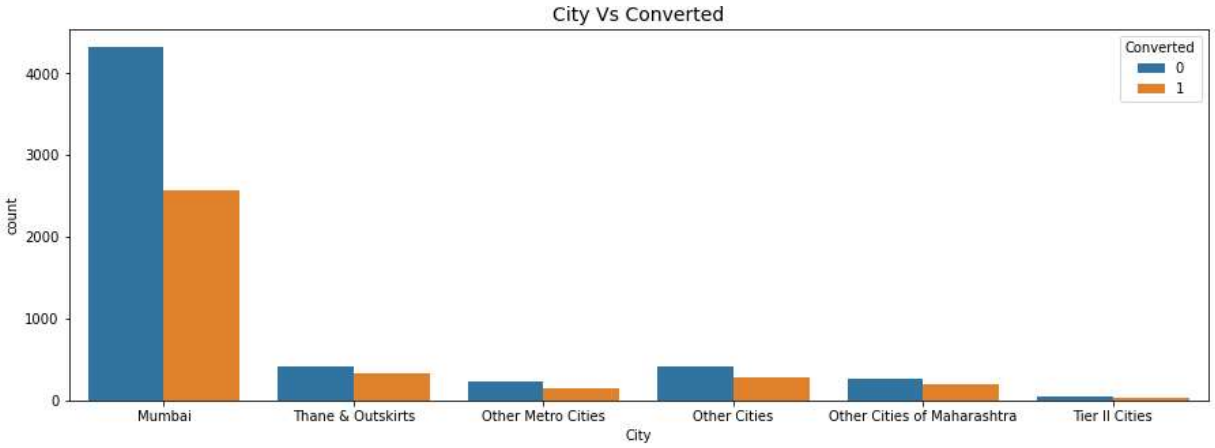
Exploratory Data Analysis



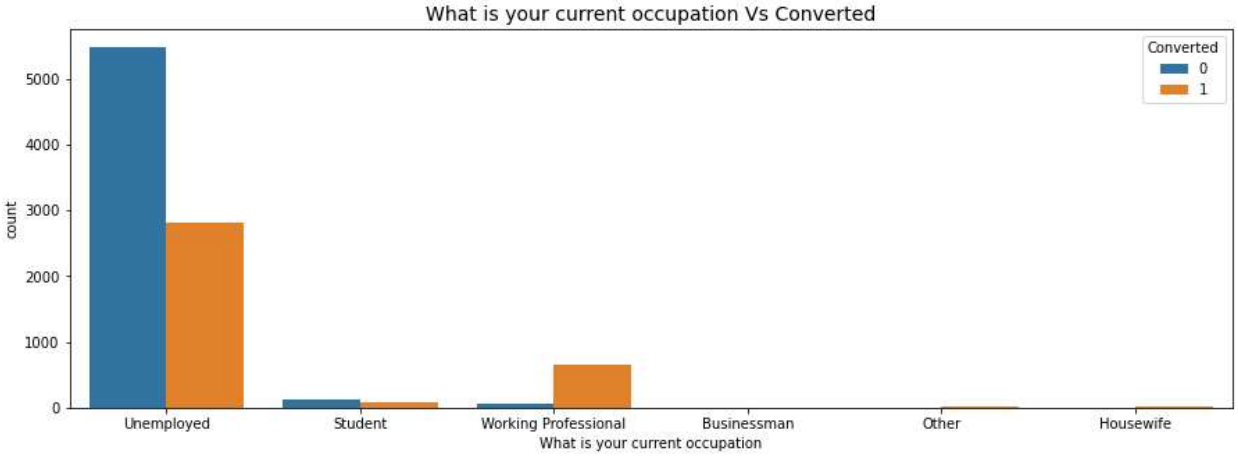
We have the conversation rate around 38.5%



The plot of 'Lead Origin' clearly shows that leads from category 'Lead Add Form' have greater rate of positive conversion as compared to other Lead Origin. However the highest contributor are leads 'Landing Page Submission'.



The highest contributor to conversion rate are leads from city of 'Mumbai'.



The plot of 'What is your current occupation' clearly shows that leads of 'Working Professional' have higher rate of positive conversion as compared to other occupation. However, the highest contributor to conversion rate are leads which are 'Unemployed'.

Heat Map

The heatmap shows weak correlation between the numeric variables except 'Total Visits' and 'Page Views per Visit' which have a moderate correlation with each other.

Model Evaluation – Train Data Set

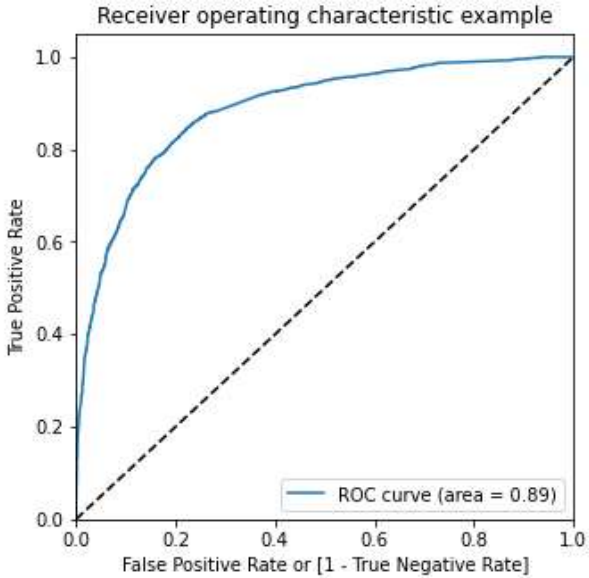
Confusion Matrix

Initial Cut-off point of 0.5

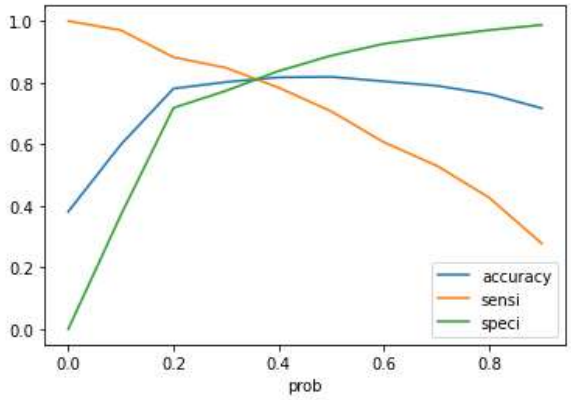
3552	450
723	1743

Accuracy – 81.86%
Precision – 79.48%
Recall – 70.68%

ROC Curve



We plotted the ROC curve for the remaining features and found the curve is decent with 89% area coverage.



The graph depicts the optimal cutoff is 0.41 based on Precision and Recall

Confusion Matrix

Cut-off point of 0.38

3294	**708**
509	1957

Precision – 73.43%
Recall – 79.35%

Model Evaluation on Test Data Set

Confusion Matrix

1290	**387**
160	935

- Accuracy 80.26%
- Sensitivity 85.38%
- Specificity 76.92%
- Precision 70.72%
- Recall 85.38%

Summary

- The top three variables in the model which contribute most towards the probability of a lead getting converted are 'Last Notable Activity_Had a Phone Conversation', 'Lead Origin_Lead Add Form' and 'What is your current occupation_Working Professional’.
- Then finally we implemented the Model on Test Data set and found out that the accuracy value is 80.26%, sensitivity 85.38% and specificity is 76.92%.
- We calculated precision and recall values and found out to be 73%, 79% respectively
- Based on the precision and recall curve, we got the cut-off value at 0.41

Thank You