Lead Scoring Case Study

Abhishek Chattopadhyay

Shreya Sunil Phade

Devika Babu

Objective: Lead Generation analysis

X into paying customers. The company wants us to build a model wherein we need to Education expects to help them select the most promising leads, i.e. the leads that are most likely to convert assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Goals of the Case Study

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

- A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

- Analyzing the Data set and cleaning the data by removing the unwanted columns.

- Building a model(mixed method) by eliminating the columns basing on RFE, P>Z and VIF.

- Finding the optimum probability and checking the Sensitivity, Specificity and Accuracy for both train and test data.

Categorical Variables
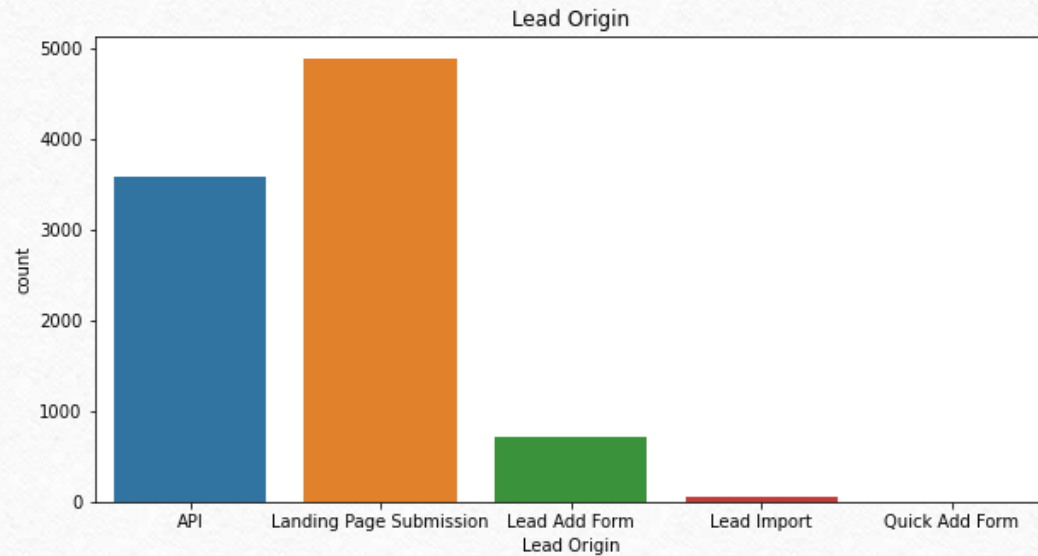
Landing Page Submission has highest count followed by API.
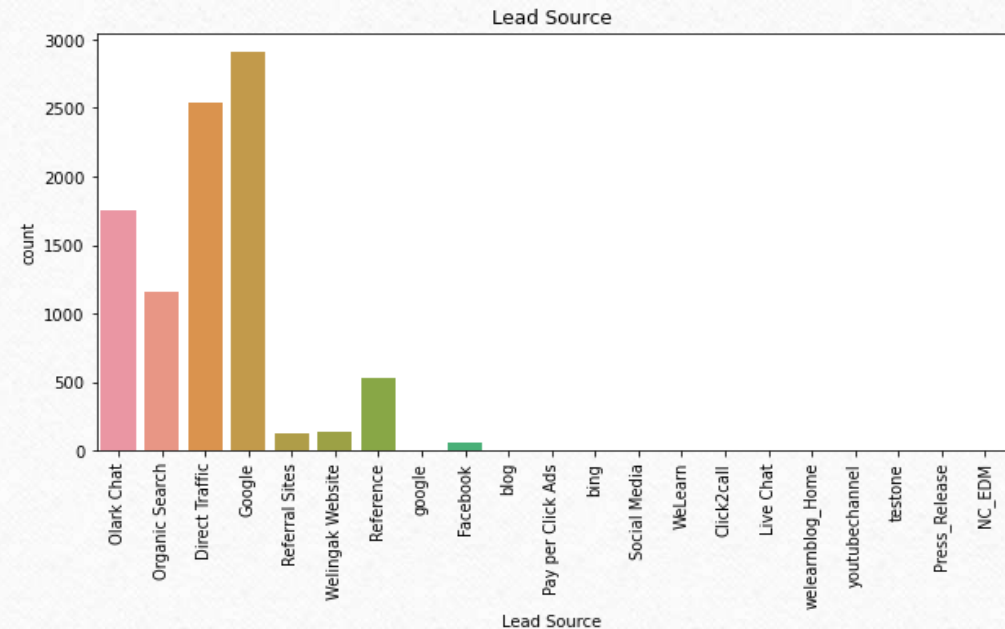
Most of the people came via Google followed by Direct Traffic source.

•Numerical Variables

- inference 1 = Added form is more effective way to convert people but it is significantly less in count.
- inference 2 = Landing Page Submission has highest count of people who didn't convert. Still it is second best effective way to convert people.
- inference 3 = Reference helps most in converting people followed by Google.
- inference 4 = Olark chat and referral sites perform lowest in conversion of people.
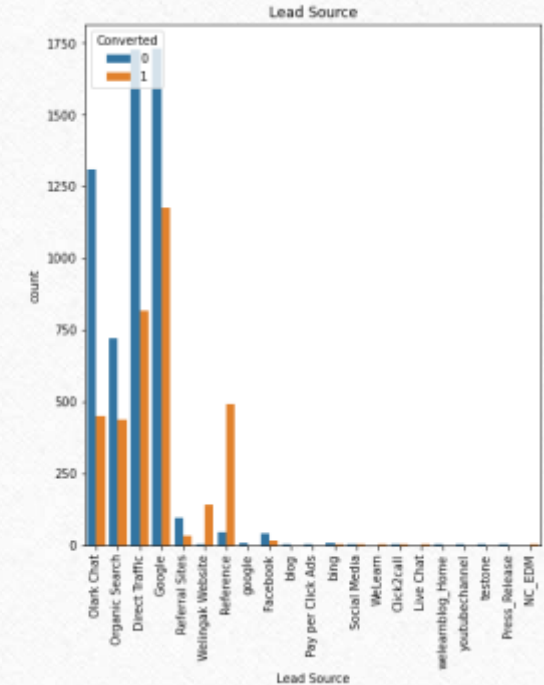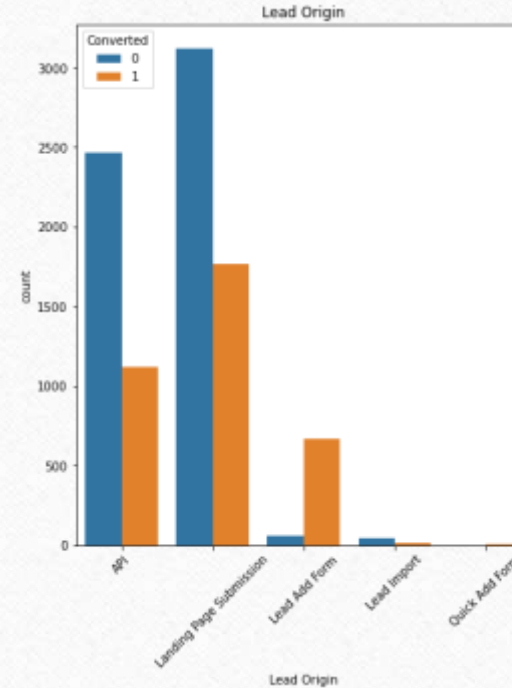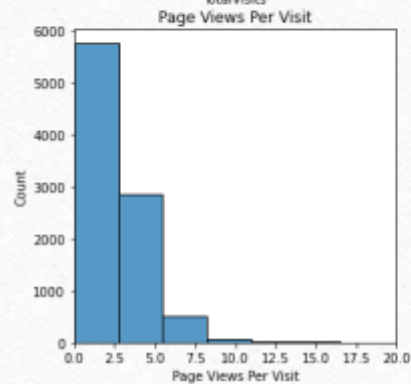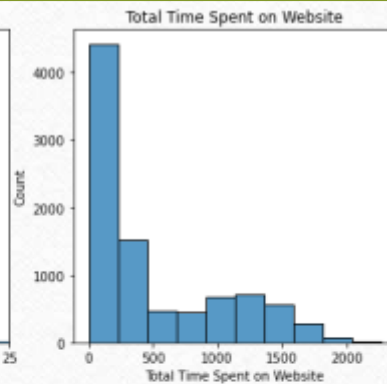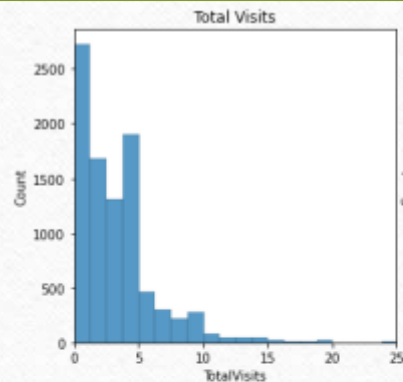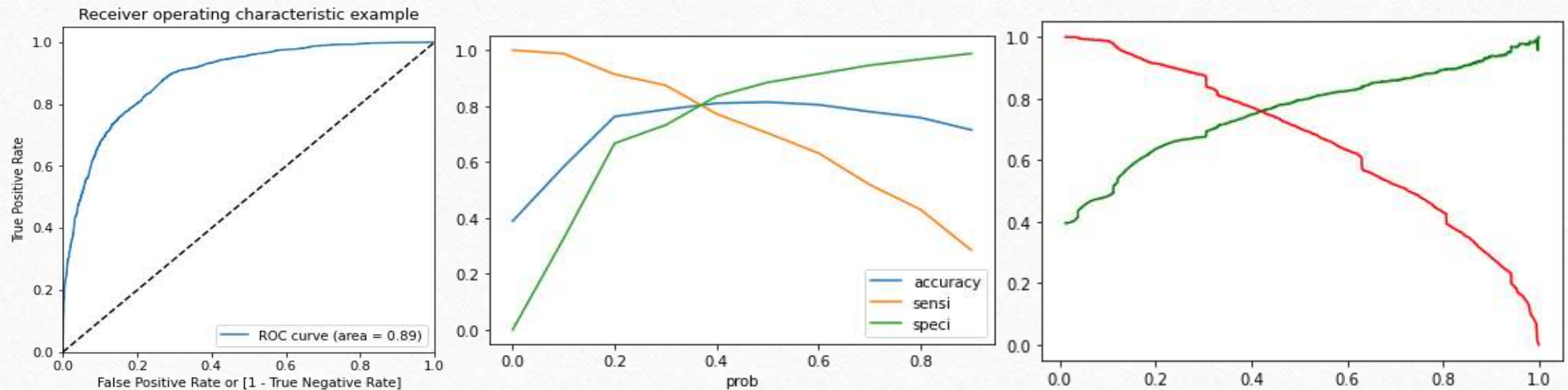
Multi Variate Analysis
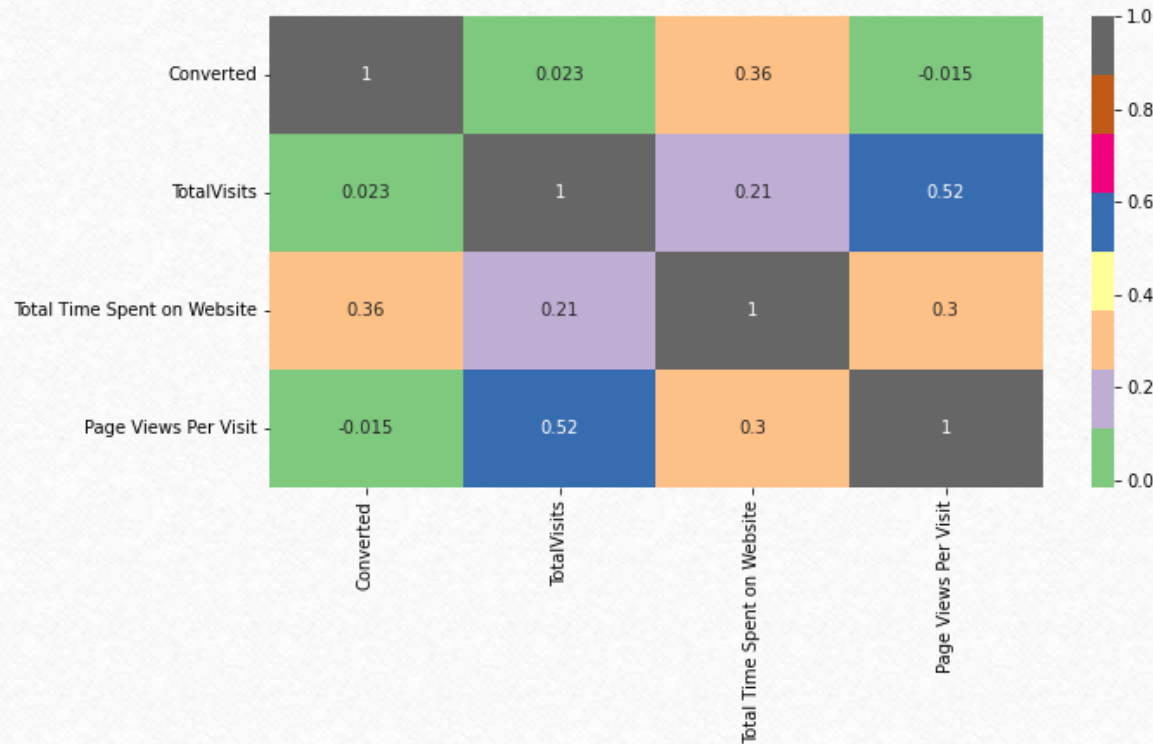
•Inference 1 = There is 0.36 correlation of "Total Time Spent on Website" with target variable "Converted".

•Inference 2 = "Page Views Per Visit" have -0.015 correlation with target variable.

Correlation among variables

Model Evaluation

Area under the ROC curve was found 0.95 .
Accuracy, Sensitivity and Specificity met at 0.2. This value was used in the model

Final Observation

Train Set:
- Accuracy = 81.03%
- Sensitivity = 77.17%
- Specificity = 83.49%

Test Set:
- Accuracy = 81.02%
- Sensitivity = 76.31%
- Specificity = 83.88%

After running the model on the Test Data these are the figures we obtain:

- Accuracy = 81.10%
- Sensitivity = 77.08%
- Specificity = 83.54%

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

THANK YOU