

Lead Scoring Case Study

Swaroopini T
Harshita Kapoor

Problem statement

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to 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 and Objectives

- There are quite a few goals for this 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.

Steps Followed

We have performed the Exploratory Data Analysis on the dataset provided by X company and created models for the same to provide a detailed analysis of the trend and major features that are affecting the leads.

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.

Model Development & Evaluation

❑ Model Building:

Firstly, RFE was done to attain the top 16 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with $VIF < 5$ and $p\text{-value} < 0.05$ were kept).

❑ Model Evaluation:

With the optimal cut off as 0.36, we have around 79% accuracy, sensitivity of around 74% and specificity of around 83%.

Prediction and Recall

- ❑ Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.32 with Accuracy 68%, sensitivity 68% and specificity of 85%.

- ❑ Precision – Recall:

This method was also used to recheck and a cut off of 0.39 was found with Precision around 74% and recall around 71% on the test data frame.

Conclusion:

It was found that the variables that mattered the most in the potential buyers(Converted leads) are (In descending order) :

1. Total Time Spent on Website.
2. Lead Origin_others:
 - a. lead add form
 - b. lead import
 - c. quick add form
3. When their current occupation is as a working professional.
4. When the lead source was: Olark chat conversation
5. When their current occupation is unemployed Keeping these in mind the X

Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Few Visualizations

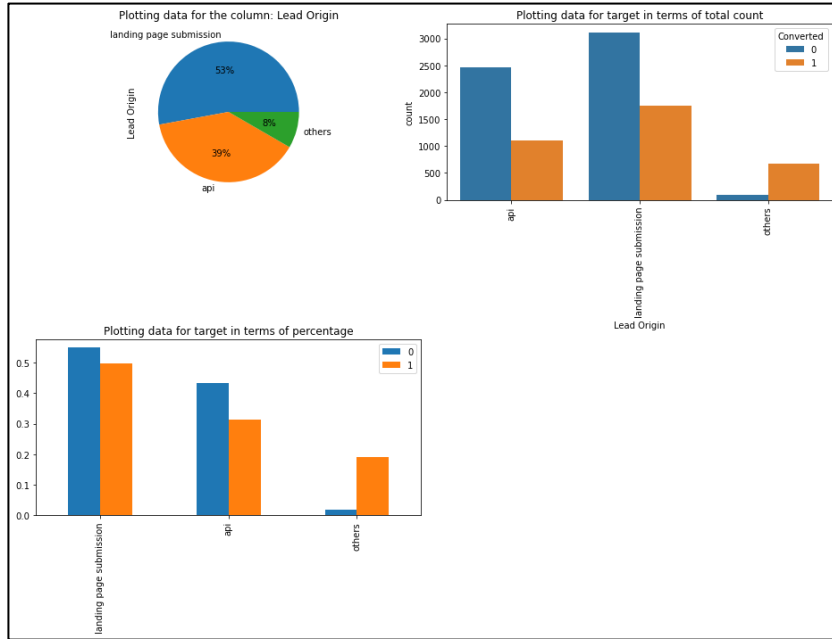


Fig: Lead Origin

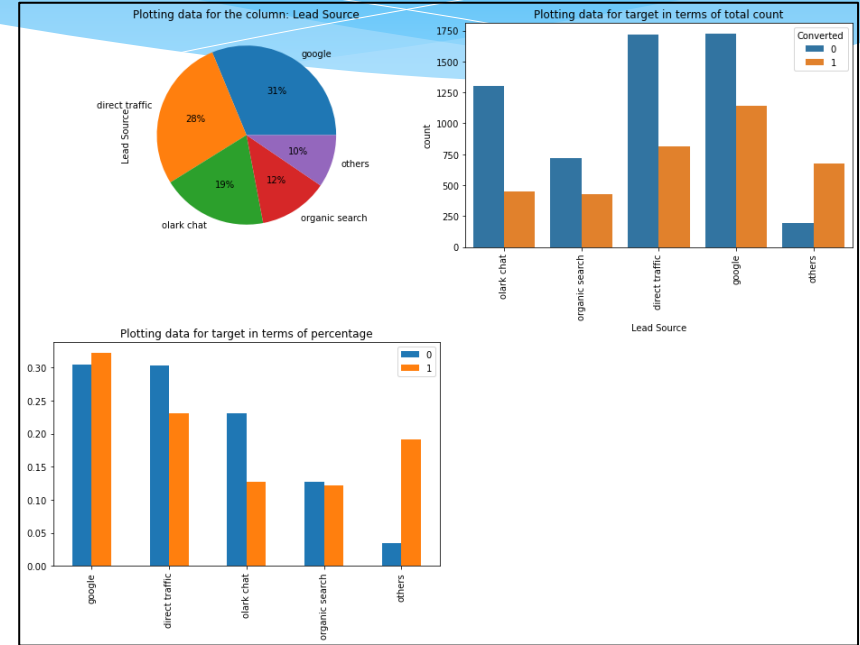


Fig: Lead Source

Visualizations

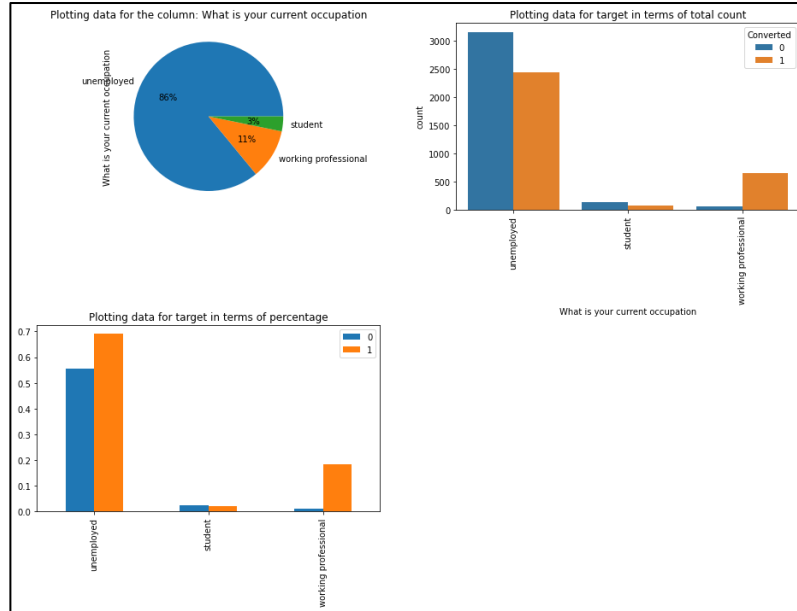


Fig: Lead Occupation

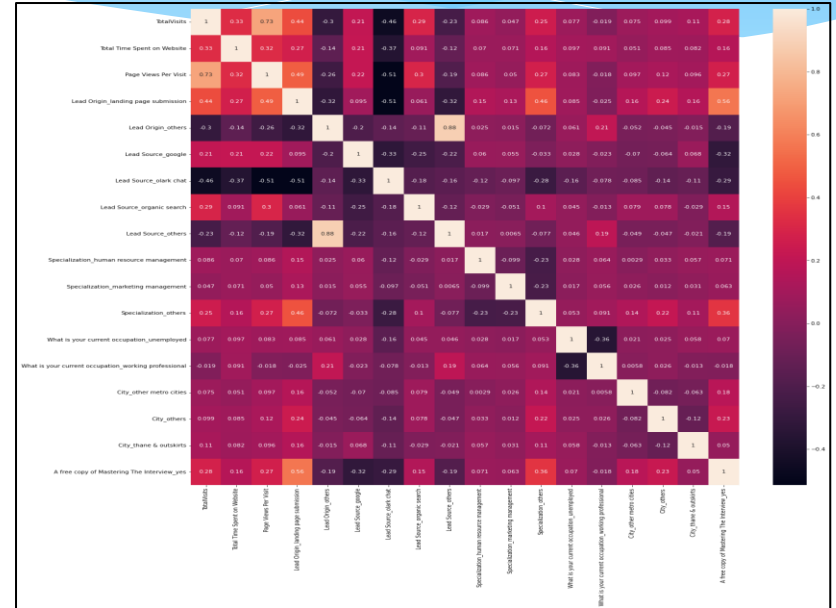


Fig: Correlation b/w all the feature after creation of dummies

THANK YOU!