Lead Scoring Case Study

By Soumya Ninan & Amita Parmar

Problem Statement

➔   X Education sells online courses to industry professionals.

➔   X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

➔   To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

➔   If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective

- X education wants to know most promising leads.

- For that they want to build a Model which identifies the hot leads.

- Deployment of the model for the future use.

Solution Methodology

- ❏ **Data cleaning and data manipulation.**
    - ❏ Check and handle duplicate data.
    - ❏ Check and handle NA values and missing values.
    - ❏ Drop columns, if contains large amount of missing values and not useful for analysis.
    - ❏ Imputation of the values, if necessary.
    - ❏ Check and handle outliers in data.
- ❏ **EDA**
    - ❏ Univariate data analysis: value count, distribution of variable etc.
    - ❏ Bivariate data analysis: correlation coefficients and pattern between the variables etc.
- ❏ **Feature Scaling & Dummy Variables and encoding of the data.**
- ❏ **Classification technique: logistic regression used for the model making and prediction.**
- ❏ **Validation of the model.**
- ❏ **Model presentation.**
- ❏ **Conclusions and recommendations.**

Understanding Dataset

❏ We got a file named "Leads.csv" provided with a leads dataset from the past with around 9000 data points.
❏ This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not.
❏ To learn more about the dataset we got the data dictionary.
❏ The target variable, in this case, is the column 'Converted' which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn't converted.
❏ Another thing that to check out for are the levels present in the categorical variables. Many of the categorical variables have a level called 'Select' which needs to be handled because it is as good as a null value.

EDA
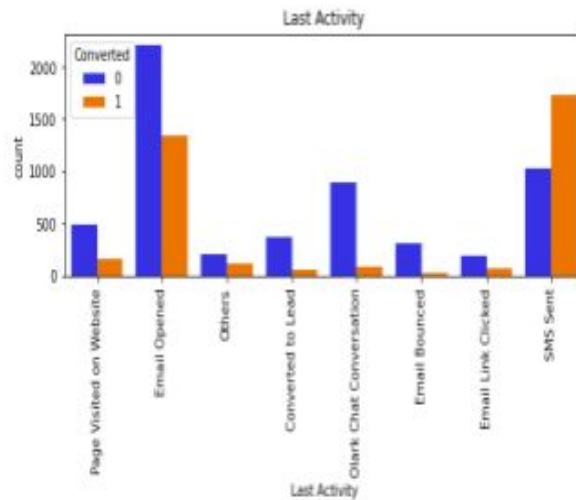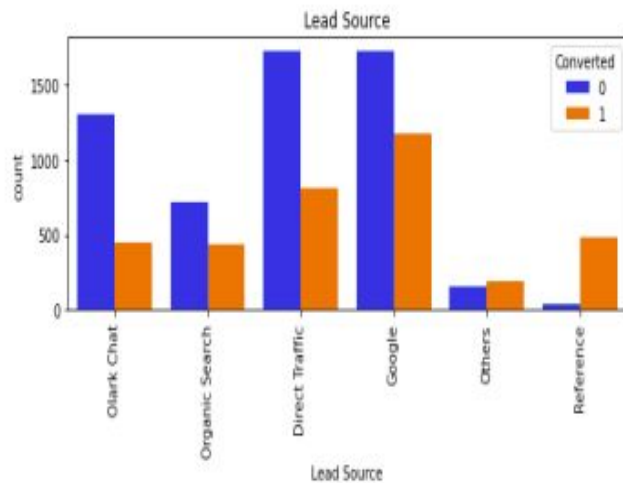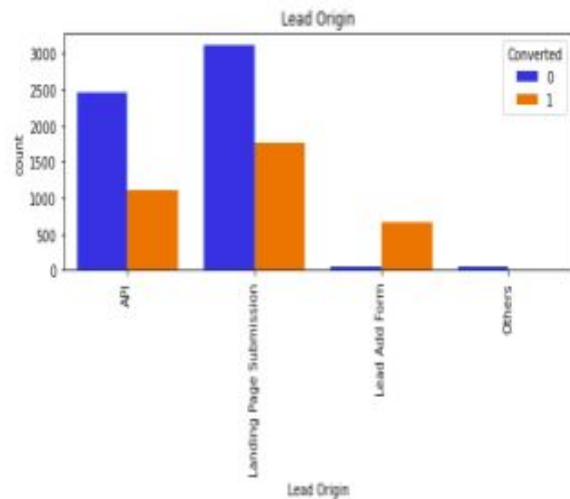
Model Performance Test

Train Set

- ACCURACY - 81.19%
- SENSITIVITY - 80.45%
- SPECIFICITY - 81.7%

Test Set

- ACCURACY - 80.08%
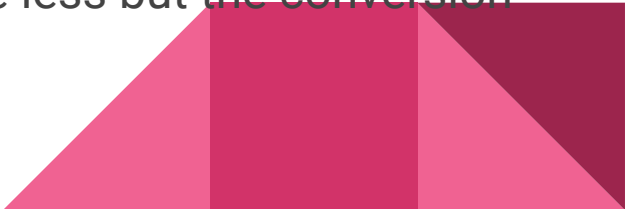- SENSITIVITY - 80.0%
- SPECIFICITY - 80.3%

Hot Leads

❖ Hot leads are people who have a high probability to be converted as a Lead and thus needs to be identified. They have a higher conversion rate.

❖ The leads whose lead score is greater than 35% are considered as potential leads. The conversion rate is around 73%. When we increase this threshold from 35% to 95% we get Hot Leads.

❖ Conversion Rate for hot leads is increases from 73% to 96%. This means they have a 96% probability of getting converted to a lead.

❖ Focusing on Hot Leads will increase the chances of obtaining more value to the business as the number of people we contact are less but the conversion rate is high.

Conclusion Points

From our model, we can conclude following points :

- The customer/leads who fills the form are the potential leads.
- We must majorly focus on working professionals.
- We must majorly focus on leads whose last activity is SMS sent or Email opened.
- It's always good to focus on customers, who have spent significant time on our website.
- It's better to focus least on customers to whom the sent mail is bounced back.
- If the lead source is referral, he/she may not be the potential lead.
- If the lead didn't fill specialization, he/she may not know what to study and are not right people to target.
- So, it's better to focus less on such cases.

Recommendations

❖ It's good to collect data often and run the model and get updated with the potential leads. There is a belief that the best time to call your potential leads is within few hours after the lead shows interest in the courses.

❖ Along with phone calls, it's good to mail the leads also to keep them reminding as email is as powerful as cold calling.

❖ Reducing the number of call attempts to 2-4 and increasing the frequency of usage of other media like advertisements in Google, or via emails to keep in touch with the lead will save a lot of time.

❖ Focusing on Hot Leads will increase the chances of obtaining more value to the business as the number of people we contact are less but the conversion rate is high.

Thank You