LEAD SCORING CASE STUDY

Problem Statement An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. The company markets its courses on several websites and search engines like Google. 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%. Now, although 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.

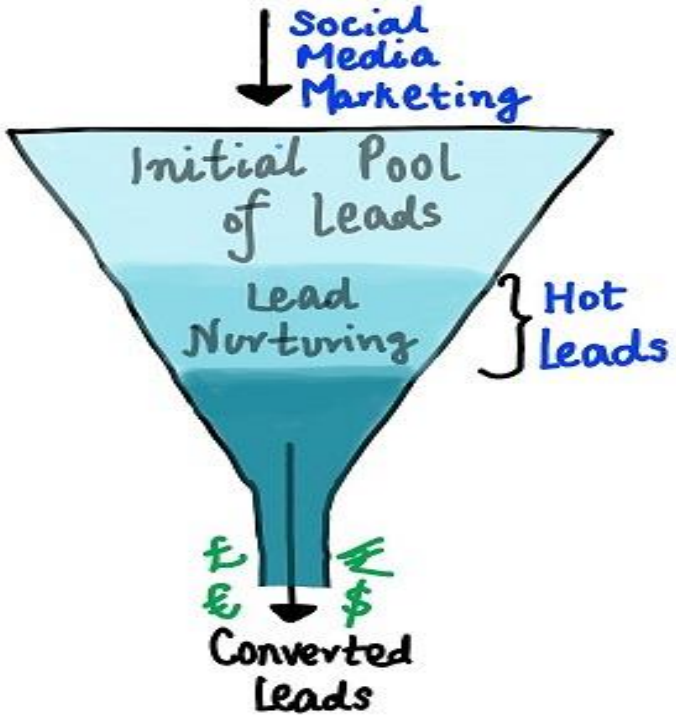


Fig. Lead Conversion Process

Objectives

- To help the company in selecting the most potential leads, also known as ‘Hot Leads’ whose lead conversion rate is around 80%.

- To build a model wherein a lead score is assigned 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.

- Help the sales team to divert their focus on potential leads & avoid them from making useless phone calls.

Approach

- Analysing Patterns: - Using Exploratory Data Analysis, we have analysed the patterns present in the Dataset which will provide us intuition that the which features will help in driving the lead conversion.

- Driving Factors: - Looking at the below data we get an intuition that how the variables are distributed.

- Correlations: ¬ Identifying correlations amongst variables to identify the variability in data and identify most important features that can help in driving the conversion of leads.

- Recommendations:

    - Focus on features that can expedite the conversion of leads.:

Data Insights

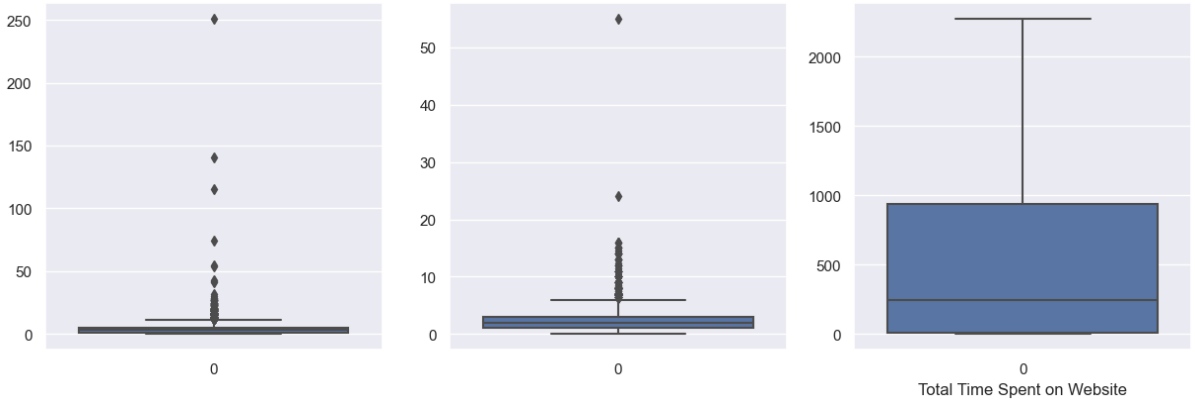
1. We have total 9240 entries of unique customers and we needs to identify out of these which have the highest probability of getting converted.

   Decision Criteria:

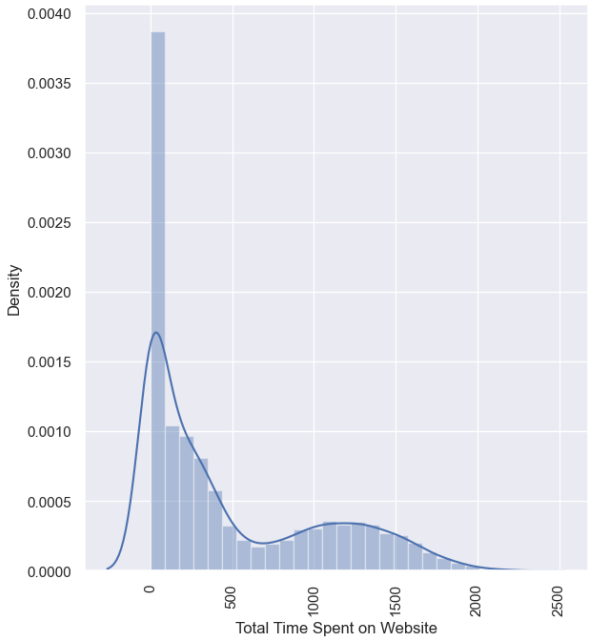
   - Potential Leads can be bifurcated on the basis of Leads Score (which is probability of getting converted).

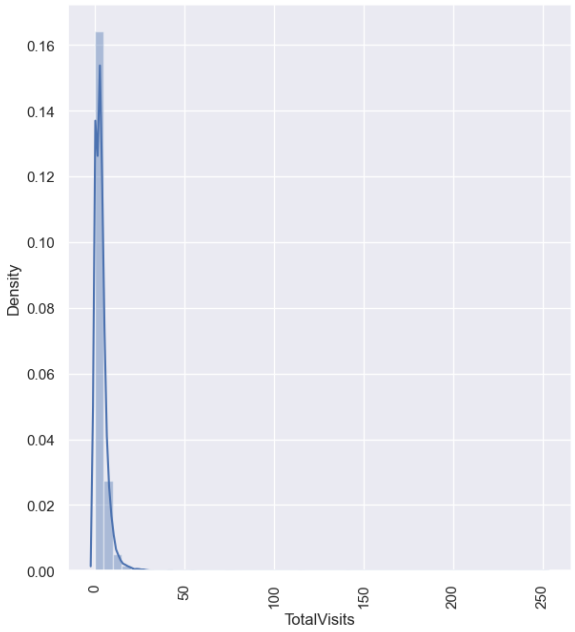
   - Out of 9240 entries we see that around 37% of leads are converted and 73% of leads are not converted.

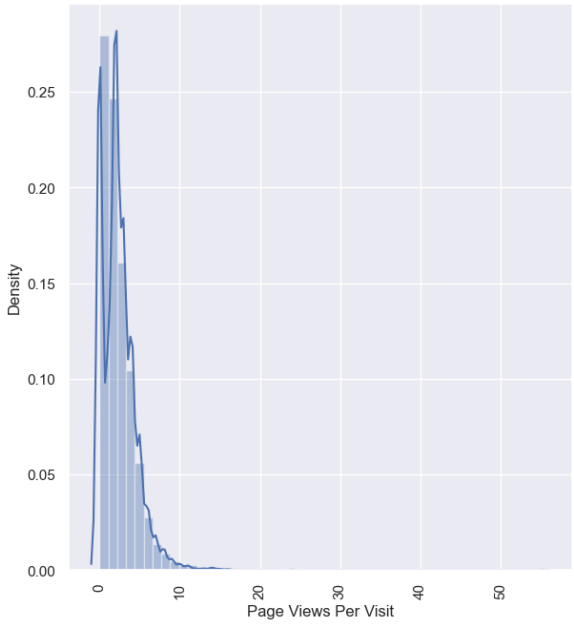
2. Outliers are Present in both the Variables TotalVisits and Page Views Per Visit it should be treated and the value spreaded above median highly in Total Time Spent on Website



3. Let us see the spread of numerical columns



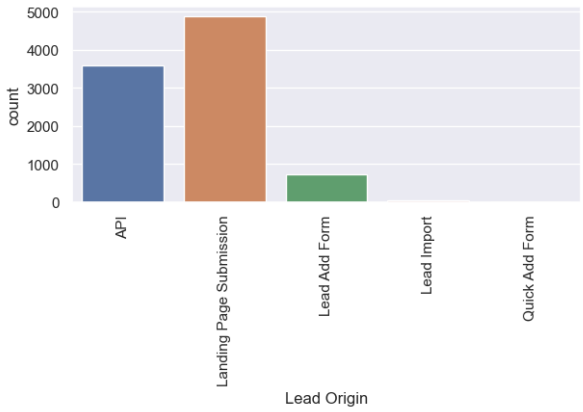


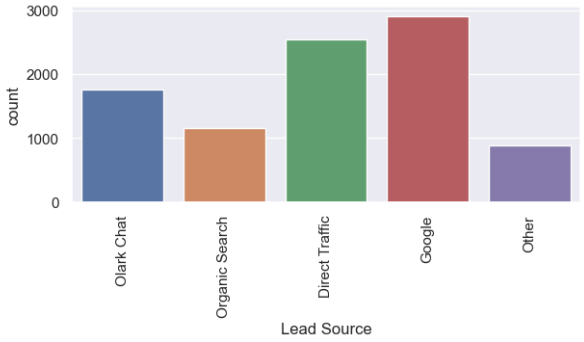


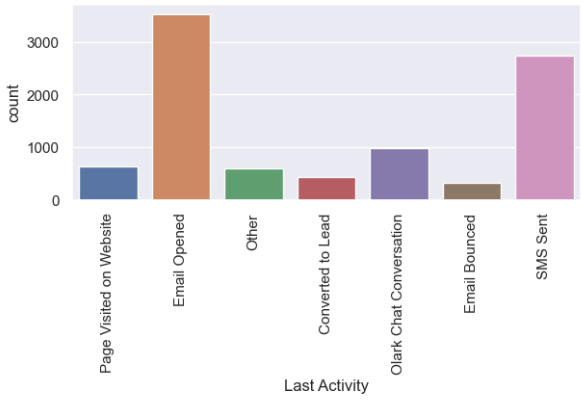
Observations:

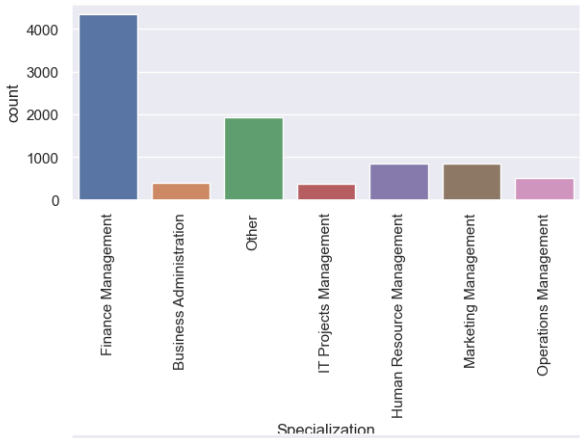
We observe that our data is skewed.

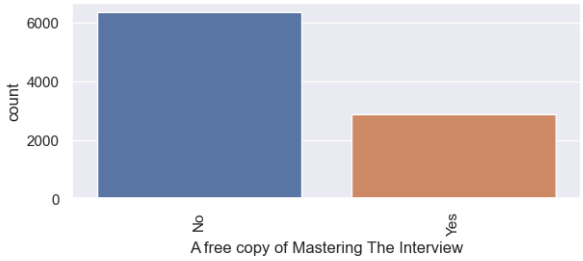
4. Lets see the spread of Categorical Columns w.r.t Converted Columns.

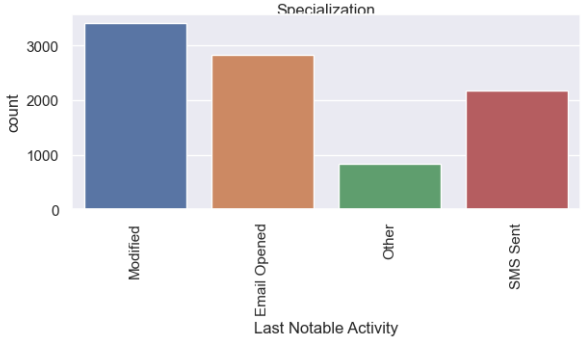












To improve overall lead conversion rate, we need to focus more on improving lead conversion rate of Customers whose last activity was Email Opened and generate more leads from the ones whose last activity was SMS Sent.

Terminologies Required

Before proceeding ahead, we need to understand few terminologies

- Conversion of categorical columns to numerical. This step is done as our algorithm runs only on numerical data.
- Feature Scaling. This is done to bring our data into same scale.
- Data Splitting: We have split the data into 80:20 and named it as train data and test data. We run model on train data and validate our model on test data.

Conclusion Focus:

Company should focus on following features to increase the leads

- Tags_Closed by Horizzon: Leads that have been assigned Tags as ‘closed by horizon’ have the highest probability of conversion.
- Tags_Lost: Leads that have been tagged as ‘Lost ‘also contribute to the conversion to a considerable extent.
- Tags_Will revert after reading the email: Leads that have been tagged as ‘will revert after reading the mail’ also have significant correlation with the conversion.

Expansion:

Company should also focus on Lead Score (which are the probabilities obtained via algorithm) which are greater than 80% to expedite the conversion rate.