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'.

As you can see, there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you 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%.

Analysis Methodology

Steps:

Data Understanding – Data Cleaning, formatting, Identifying Patterns and Outliers in the data. Treating outliers and missing Values.

Data Analysis: Understanding Numerical and Categorical Features. Understanding its influence on the dependent variable.

Data Modelling: Creating Dummy Variables, Logistic Regression Predictive Modelling for Lead Conversion

Model Evaluation: After Training the model with the train data, testing the model to determine its significance.

Conversion Rate:

As we can see in the graph below that the conversion rate of customer is low at 37.89% which means that more than 62% of the customers that visit this site have not joined any of the courses offered by X Education i.e, some of the potential leads might have been missed. Our job is to identify such potential leads

Conversion Rate

Converted (group)

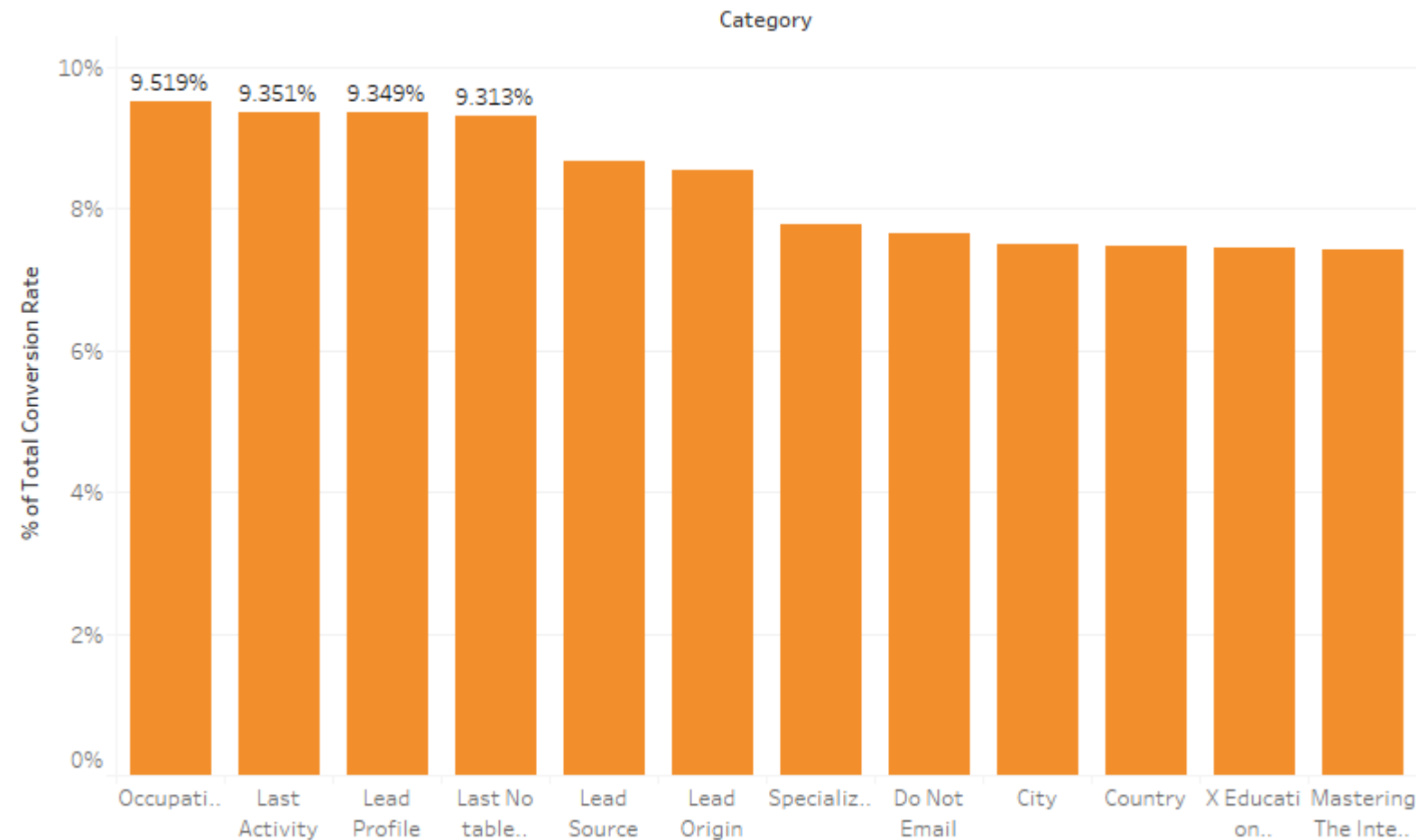%of Conversion

62.11%

37.89%
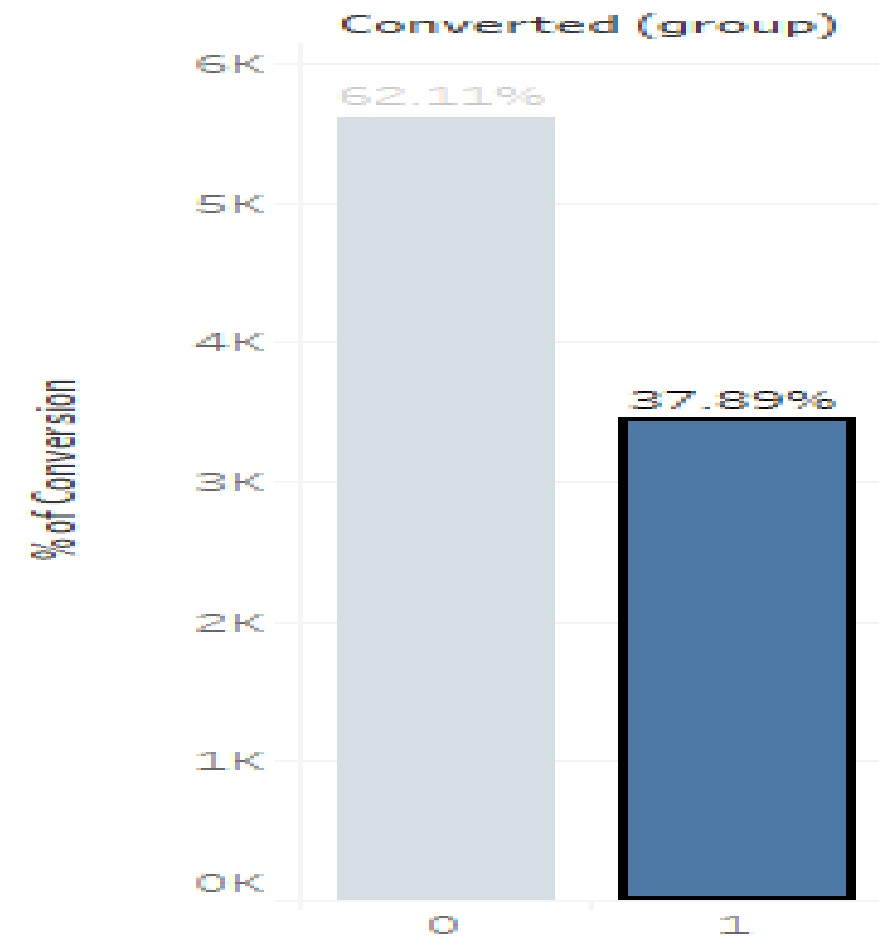
Count of Converted for each Converted (group). The marks are labeled by % of Total Count of Converted.

Significant Factors:

1. Occupation
2. Last Activity
3. Lead Profile

These features have higher significance on conversion rate. So we will use these along with some other features to build the model and assign a Lead Score


Significant Factors

Category

% of Total Conversion Rate for each Category. The marks are labeled by % of Total Conversion Rate.

Identifying how customer know about the X Education should be the first task. We can see from the chart that majority of sources have known about X-Education from Google and Online Searches. So advertising here could increase the number of potential leads.

Lead Source
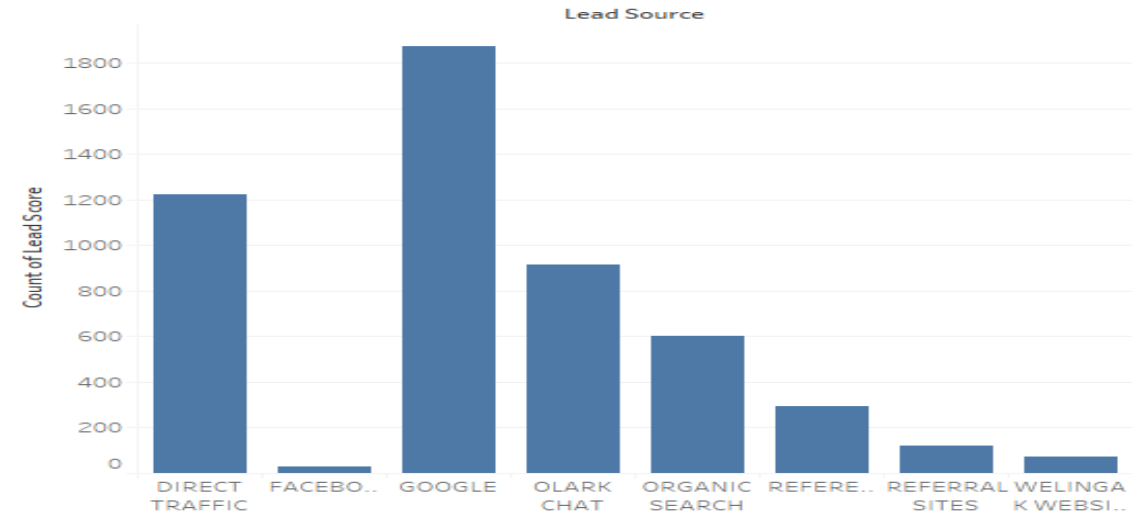
Lead Source

Count of Lead Score for each Lead Source. The view is filtered on Lead Source, which keeps 8 of 20 members.

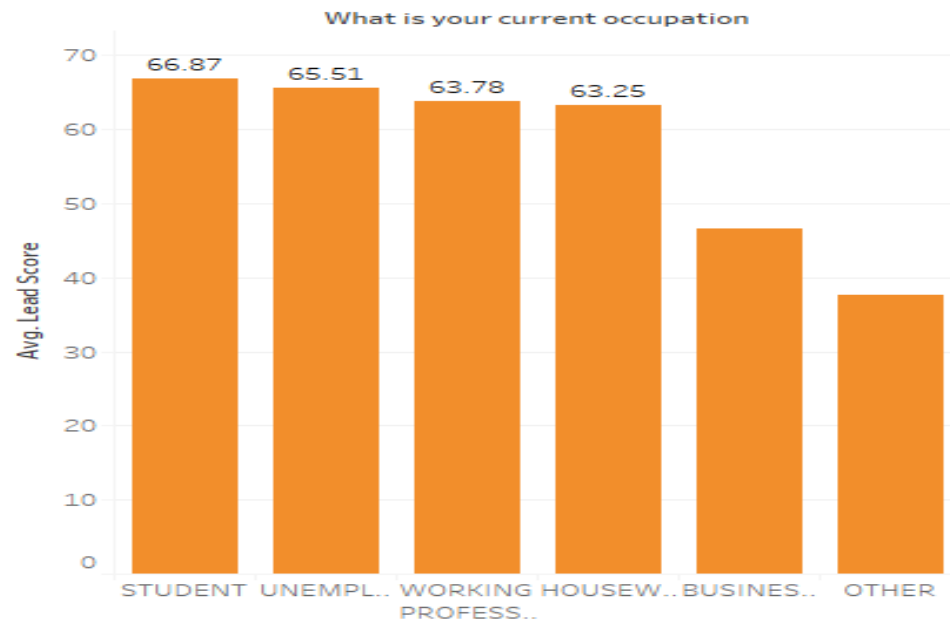## X-Education Source

How did you hear about X Education
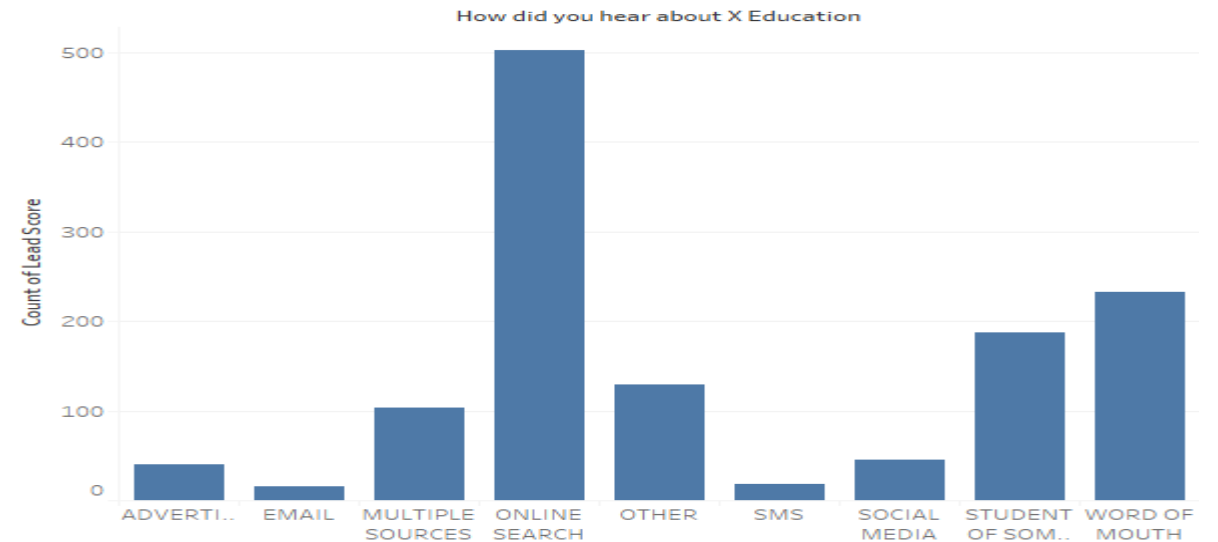
Count of Lead Score for each How did you hear about X Education. The view is filtered on How did you hear about X Education, which excludes SELECT.

Lead Score Explanation:

Lead score was calculated based on the optimum probability found when building the model which had the sensitivity of 85% and Specificity of 80%. So the customer having a lead score of 80 or more have higher chance of conversion. Hence, for new customers lead score should be calculated using the model and respective decision should be made based on this score.
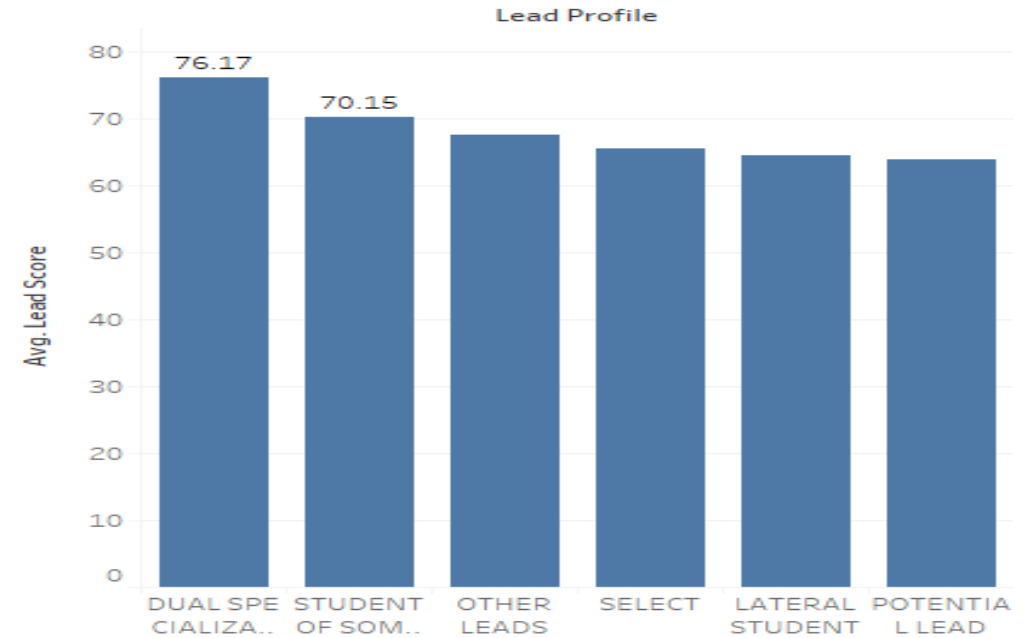
Students, Unemployed and Working professionals have higher rate of conversion while Customer in Business have lower conversion rate. So the X-Education should focus on the former to improve conversion rate

Lead Score based on Lead Profile

Lead Profile

Average of Lead Score for each Lead Profile. The marks are labeled by average of Lead Score.

Lead Score based on Occupation

What is your current occupation

Average of Lead Score for each What is your current occupation. The marks are labeled by average of Lead Score. The view is filtered on What is your current occupation, which excludes SELECT.

Lead Profile is one of the significant factors that drives the conversion rate. Dual Specialization Student and other students have high lead score. So focusing on this would increase the conversion rate.

Customer who have approached upfront, submitted forms online and re-subscribed to e-mail have highest lead score. So when this happens, the employees should focus on these customers.

Lead Score based on Occupation

Last Notable Activity
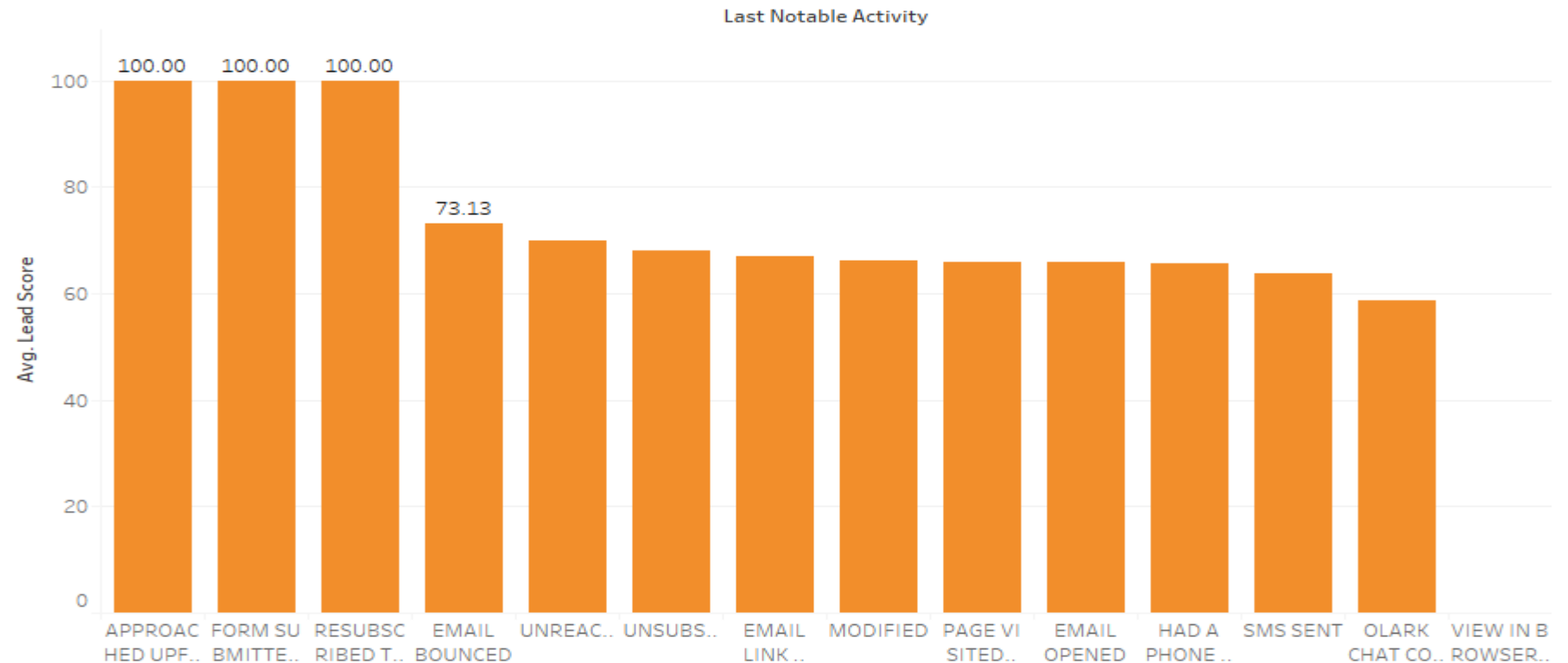
Average of Lead Score for each Last Notable Activity. The marks are labeled by average of Lead Score. The view is filtered on Last Notable Activity, which excludes EMAIL MARKED SPAM and EMAIL RECEIVED.

THANK YOU!