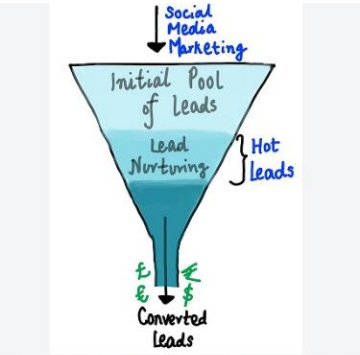
Summary

Objective: An Ed-tech company named X Education sells courses in online which are specially made for working professionals. The company market there courses via Google. Though which company got many aspirants to their list in which there are unemployed, students and working professionals. Making calls and SMS repetitively consumes a lot of time and 100 leads in a day, only about 30 of them are converted.



Lead Conversion Process - Demonstrated as a funnel

So company want to filter out the leads who are not interested so that they can focus in 'HOT LEADS'. This way they can target lead conversion rate to be around 80%.

1. A Leads dataset is provided. 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. And another thing to focus on the categorical feature which is having level 'select' as imported as NULL value.

2. Lead scoring case study has been done using logistic regression model to assign the lead score from 0 to 100; a higher value would mean that the lead is hot vice versa.

3. '**INDIA'** recorded as the highest leads, whereas in terms of city '**MUMBAI'** has the maximum leads.

4.There are a few columns in which there is a level called `'Select'` which basically means that the student had not selected the option for that particular column which is why it shows 'Select'. These values are as good as missing values and hence we

need to identify the value counts of the level 'Select' in all the columns that it is present.

‘Lead Profile, How did you hear about X Education and Specialization’ are the columns where level SELECT is present. In all three categories the majority of the students are not selected the Specialization.

5. The Rows of the columns are dropped which includes NULL values in Majority.

6. After removing all NULLS next step is to creating Dummies for categorical columns.

7. Splitting the dataset into 70% train and 30% test. We on observed that the values present in the columns are having highest values so opting standardizing process i.e. minmax scaling.

In first model we found that few variables are having greater values, the values of VIF are decent except three variables, and they are dropped one by one. This process of checking the p-values and VIF is performing until the desired values is found.

After the necessary steps are perform.

7. The high number of total visits & Total time spent on platform is directly proportional to the chances of lead to be converted.

8. Most of the leads joined course for their better career path, and majority of them are from Finance Management background. Leads from HR, Finance & Marketing management specializations are said to be high probability to convert.

9. One thing that can convert the leads to HOT LEADS is through repetitive calls and engagement through emails. If leads show interest to the emails or SMS’s it show that the lead is interest in taking up the course.

10. Almost all of leads current occupation is Unemployed, which means gave more focus on unemployed leads