LEAD SCORE CASE STUDY

The company markets its courses on several websites and search engines like Google. Once potential customers land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these potential customers fill up a form providing their email address or phone number, they are classified as 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%.

Our task was to help X Education identify the most potential leads, also known as ‘Hot Leads’. The company wanted to build a model 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, had given a ballpark of the target lead conversion rate to be around 80%.

We were 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. 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.

The first step we took was to analyze the dataset to determine which variables were. We found that many of the categorical variables had a level called 'Select' which needed to be handled because it was as good as a null value. We also discovered that some variables, such as 'Lead Source', had a large number of unique categories, which could potentially affect the model's performance. Therefore, we decided to create a new category for all values with a frequency less than 100.

This analysis helped to find hot lead will gets converted.