This report summarises the approcach taken to predict the lead conversion rate for aan education company named X Education which sells online courses to industry professionals.

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%.

We have build a classification model based on logistic regression to identify a candidate as a potential lead or not.

We start off by cleaning the data , treating null values , dropping imbalanced columns, treating outliers by binning , standardizing the variables so as to get a correct prediction.

The dataset is then split into a train and test data set. A logistic regression model is built using RFE which selects top 19 features to train our model.

Then we have proceeded with the manual feature elimination based on p-value and VIF. Variables with high p-value and VIF are dropped one by one.

The test data is used to predict the target variable i.e. Converted .The accuracy , precision and recall on the train and test data model is pretty good ,which means our model will be able to predict conversions correctly in future even when the company's requirement changes.

Concentrate more on working professionals as they can spend money on course and people who have had phone conversation earlier who seem to be more interested .Here we will only be checking the hot leads having conversion score above 90 , so that we can minimize the rate of useless phone calls