BRIEF SUMMARY

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. 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. After acquiring the leads, company start making calls to the leads. The typical lead conversion rate is around 30%.

The company requires a model wherein score should be assigned to each of the lead 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 so that the right customer can be targeted.

Approach to solve the problem was be to identify various factors (or columns) which depicts whether customer is a potential lead or not. Thorough study of the data dictionary was done so as to understand the domain and find appropriate columns to use in the analysis. Exploratory Data Analysis was performed on dataframe for which Logistic Regression model was to be built. Feature scaling, splitting the data into train and test & preparing the data for modeling were the next steps to be performed. A logistic Regression model was build after the preparation. Model evaluation was done and predictions were made thereafter.

Model Evaluation factors of the Train data set were as follows:

1. Accuracy : 80.8%
2. Sensitivity : 82.5%
3. Specificity : 79.7%

Model Evaluation factors of the Test data set were as follows:

1. Accuracy : 80.4%
2. Sensitivity : 80.7%
3. Specificity : 80.2%

Following observations were made from the model:

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference".
- The company should make calls to the leads who are the "working professionals".
- The company should make calls to the leads who spent "more time on the websites".

- The company should make calls to the leads coming from the lead sources "Olark Chat".
- The company should not make calls to the leads whose last notable activity was "Email Link Clicked".
- The company should not make calls to the leads whose last notable activity was "Page Visited on Website".
- The company should not make calls to the leads whose lead origin is "Landing Page Submission"
- The company should not make calls to the leads whose Specialization was "Not Specified"
- The company should not make calls to the leads whose last activity was "Email Bounced".
- The company should not make calls to the leads whose last activity was "Olark Chat Conversation".
- The company should not make calls to the leads who chose the option of "Do not Email" as "yes"