Summary

An education company named as X Education is a company whose business is to offer online courses to the industry professionals. Currently, company is strategizing to improve their conversion rate of leads to consumers. With the time of 2 months along with the ample resources (10 interns), company would like to leverage the resources efficiently and plan to strategize to achieve its goal. In the past, X education had poor conversion rate of approx. 30% hence the company requires us to build a model which can help improve their business by focussing on achieving the conversion rate to 80%.

Steps taken to carry out this analysis are:

- ✓ Read and understand the data given
- ✓ Date cleaning
- ✓ Preparing the data for Model building
- ✓ Model Building
- ✓ Model Evaluation
- ✓ Testing and predictions

- **Data cleaning**: Data was cleaned partly only for few Null values and option select was replaced with null values as it did not provide much information. Other few null values in specialization were changed to 'Others' and all the columns having missing values were changed to 'remaining_missing'. Since India had the highest count, we replaced the missing values with 'India'
- **EDA**: this step was done to check the condition of our data. Many elements under categorical variables were not related. The numeric values didn't have any outliers.
- **Dummy variables**: Dummy variables were created for some of the categorical variables like Lead Origin, Lead Source, Last Activity, Specialization and What is your current occupation as 'dummy'
- **Test Split**: Train-test split was done at 70 and 30% ratio for train and test data respectively
- **Model building:** RFE was done to obtain the top 15 relevant variables. Then the rest of the variables were removed manually considering the VIF and p-values. The variable with VIF<5 and p-value <0.5 were kept
- **Model Evaluation**: The optimum cut off value using ROC curve was leverage to find the sensitivity, specificity and accuracy which came around 82%
- **Prediction:** Prediction was carried out on test data with an optimum cut off of 0.35 with sensitivity, specificity and accuracy of 82%

With the provided data, we understand that the company has high number of leads reaching in the initial stage but number reduces drastically at the last stage. So, in the middle stage is where we need to strategize based on data, variables in order to fostering the potential leads to get higher conversion rate.

The important variables that were found which potential customers are:

- Time spent on the company's website
- When the lead origin is Lead add form
- When the Lead source is Wellingak Website
- When Last activity is SMS sent

With the above-mentioned points, X education can strategies to achieve 80% conversion rate in the upcoming quarter