Summary

The analysis presented is meant for X Education to dig into the reason to increase the leads for enrolling into the courses offered. We were provided with the leads data that gave us a lot of insights about the leads behaviour like how much time they have spent on the website, how they got to know about the company, how they landed on the site etc. Based on this we did a detailed analysis and also came up with the logistic regression model so as to predict the probability for a lead to get converted.

We went through below steps in order to reach the solution:

1. Data Understanding
    - We started of with importing the data and going through it like the shape of data , describing data and getting the data info which gave us and insight on how the data look like, how many null values are present in different columns and how the numeric data looks.
2. Data Cleaning
    - After looking into the data and understanding it we move into cleaning the data and filling in the blank values or dropping in columns for which we cannot determine which value to fill in.

3. Exploratory Data Analytics
    - After cleaning the data we moved forward with doing the EDA on data to get insights into the data and pull out hidden information using graphs. We used bar graph graphs and boxplot to get the insights into the given data.
4. Data Preparation
    - We then moved on to preparing the data to build our model. We started of with generating dummy variables for categorical data and splitting the data into train and test set and then doing feature scaling on training data.
5. Model Building
    - After preparing the data and we started of with building our 1st model using logistic regression. After doing so we reduced the number of parameter to 14 and dropped other parameter using RFE.
    - After doing the coarse tuning we moved on to tune are model mauanlly by comparing p-values of all the parameter and dropping the parameters with high p-values and building model again.
    - Once done we looked at VIFs of all the parameter which came out to be in range and so we didn't drop any parameter based on VIFs.
    - After we reached our final model we then derived different evaluation metrics like accuracy , sensitivity and specificity

for different probability and derived the optimal cutoff probability for our model.
    - We then move on to derive the lead score for all the leads based on the probability.
    - We then also derived other evaluation metrics confusion matrix and Precision and Recall.
    - We also build the ROC curve as well.
6. Model Testing
    - After successfully building the model we then tested the model on the test set and pulled all the evalution matrix data and compared it with the training set data.
    - Data generated from test set and training set came out to be accuraltelty close to each other thus showing us that our model was good model for test set as well.

Following the above steps we thus did our EDA and model building exercise for X Education giving us good insight into company and its lead and their data.