Summary Report

Objective of this case study:

The Data given is of X Education who needs to find out the customers who will enrol for the course with the company based on certain details about the customer behaviour and customer demographics. The dataset includes 37 factors based on which we need to clean the data and create a model to make it easier for the sales team to optimise their calls only to the potential customers.

Tasks undertaken and the observations made:

1) **Data Cleaning:**
   Deleting columns with missing values greater than 40% and imputing columns and also dropping unnecessary columns which are not helping in analysis.
   Checking for skewed data.
   Observations made: Most leads belong to Mumbai. Team should not spend on advertising through newspapers or magazines or forums or digital platforms as the data shows that the leads have not come through these platforms.
2) **Checking the demographics of the customers**
   The average number of pages viewed per visit is 2. The total time spent is higher for the leads who have converted. The leads have higher conversion when the last notable activity is SMS sent, email opened and modified.
3) Combining the low frequency observations under one observation so that the dataset is easier to asses from the business perspective.
4) Dummy variables were created to observe the effect of each variable on conversion rate.
5) Split the data into train and test set. 70% - 30% split was done.
6) Created a logistic regression model by first performing RFE for obtaining 15 variables. Deleted the columns with high p values and high VIFs. We finally obtained 11 variables in total.
7) The prediction was performed keeping the threshold as 0.5 by performing a trade-off between sensitivity, specificity and accuracy.

The potential leads should not be missed at all hence we need to aim for higher sensitivity. This model has been successful in achieving a sensitivity of 82%, accuracy of 76% and specificity of 70%.

Recommendations for Business:

The sales team should focus on increasing the referrals and hence should encourage existing students to spread the word about the program. Since most of the people opt for the programs to get better career prospects and the conversions are higher for working professionals and housewives, it is important to bring out strategies to make them aware of the programs offered by X Education. The conversions are higher through direct traffic and hence it is important to devise strategies to increase that. Since the conversion are higher for people spending more time on the website, it is important to make the website more informative with easy access to FAQs so that people can opt for the program they are looking for.