1. a succinct summary report of no more than 500 words outlining your approach to the task and the lessons you learned.

We have been given information about how the business X Education pursues customer leads from different sources and tries to convert them into potential customers as part of the Lead Scoring case study.

At 30%, the convert rate is currently quite low. As a result, we were given the assignment to examine the data and develop a model that could forecast a lead conversion rate of 80% or more.

For this, we have proceeded with the basic analysis of the given data set.

1. Using the Data Dictionary to identify the categories.
2. Removing records with > 30% missing data;
3. Imputing a small number of columns with missing data;
4. Finding potential data columns that can contribute to an accurate prediction;
5. Determining the relationship and distribution of column data using graphs.
6. Eliminating anomalies from numerical factors;
7. Creating a heat map to visualize correlations

- In order to quickly transform the categorical data into features that could be put into a model used for predictions, we then went about encoding the categorical data into dummy variables.
- Other people, The Dummy columns lose any unknown numbers that were converted into columns.
- Following that, the data are divided 70:30 into training and test sets.
- To prevent any disparities in the magnitude of the data values from affecting the model forecast, the training data is scaled.
- A general linear model is given the training data. (GLM).
- The ineffective variables are eliminated using RFE and VIF.

- We are then left with 14 variables + 1 constant which has been able to predict the training data set at more than 80% accuracy and precision.
- The same model has been applied to test data set after the test data has been scaled. And we have also observed more than 80% accuracy & precision there as well.
- The final formula for this Log Reg model is:

----

ln (p/(1-p)) = -1.9079 + 5.7010 * Tags_Closed by Horizzon + 4.3909 * Lead Source_Welingak Website + 4.3704 * Tags_Lost to EINS + 3.6220 * Tags_Will revert after reading the email + 1.9865 * What is your

current occupation_Working Professional + 1.8385 * What is your current occupation_Unemployed - 4.0378 * Tags_Already a student - 3.9105 * Tags_switched off - 3.6396 * Tags_Not doing further education - 3.5416 * Lead Quality_Worst - 3.3832 * Tags_Diploma holder (Not Eligible)-3.3131 * Tags_Ringing - 3.0180 * Tags_Interested in other courses - 2.8539 * Tags_Interested in full time MBA