Summary

-X Education company had presented us with a problem of identifying the most promising leads on their website that would get converted when contacted.
-They had shared their past data with us containing information of 9000 leads.
-We were supposed to build a model based on this data, that would predict future lead conversions.

-Earlier, the conversion rate the company had was 30%. The CEO was expecting the conversion rate of leads to come upto around 80%.
-We proceeded to build a model that would classify their leads into ‘hot’ and ‘cold’, with ‘hot’ meaning high chances of conversion so that the company’s sales team can only focus on ‘hot’ leads.

- To build the model, we first imported all necessary libraries and loaded the data.
- We then analysed the data and found many missing values in it. We treated it by dropping the columns or imputing values.
- Once done, we dropped some useless columns and those columns which had imbalanced data classes as it would impact the performance of the model.
- We then did exploratory data analysis.
- We then ran the logistic regression algorithm after splitting the data into training and testing sets, and making the data suitable to fit into the model.
- We then analyzed which variables are top most in contributing to lead conversion and selected those variables.

- Using those variables, we predicted the outcome on test set, and analyzed difference between predicted values and actual values of target variable.
- We set a probability score to each lead indicating the probability the lead will get converted.
- Based on this, we also assigned a leadscore to each lead.

- We iterated the model 2 more times, and found that the first model was performing best. Therefore we chose to proceed with it.
- In the first model, we achieved the following metrics-
- Accuracy: 0.78 Precision: 0.74 Recall: 0.66 F1 Score: 0.70 ROC AUC Score: 0.76

- Threshold: 50, Conversion Rate: 74.19%
- Threshold: 60, Conversion Rate: 78.88%
- Threshold: 66, Conversion Rate: 79.83%
- Threshold: 70, Conversion Rate: 81.27%
- Threshold: 80, Conversion Rate: 83.01%
- Threshold: 90, Conversion Rate: 90.46%

- Above, threshold refers to lead score.

So, we finally came to the conclusion that if the company sales team calls only those leads whose lead score assigned is above 66, then they will definitely achieve a conversion rate of around 80%.

The company also presented us with some other problems.

- They wanted us to suggest a strategy to make the lead conversion more aggressive during some periods.
- We advised them to segment all leads into 3 segments based on their lead score.
- And then each segment should be dealt with accordingly, i.e. segment with top lead scores should be focused more, and segment with least lead score can be deprioritized.

- They also wanted to avoid unnecessary calls during some periods and wanted the calls to be made to the most important leads. In such case we advised them to contact leads having lead score more than 90%