Summary

- X Education gets a lot of leads, its lead conversion rate is very poor at around 30%.
- The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have higher conversion chance.
- CEO's target for lead conversion rate is around 80%.

Data Cleaning:

- Columns with >40% nulls were dropped.
- Numerical categorical data were imputed with median and continuous columns with ode

EDA:

- Data imbalance checked- only ~40% leads converted.
- Performed univariate and bivariate analysis for categorical and numerical variables.

Data Preparation:

- Created dummy features for categorical variables
- Splitting Train & Test Sets: 70:30 ratio
- Feature Scaling using StandardScaler
- Dropped few columns, they were highly correlated with each other

Model Building:

- Used RFE to reduce variables to get most promising 15 variables.
- Manual Feature selection is done for the remaining variables with dropping variales where p – value > 0.05.
- After iteration Model 5 was stable with (p-values < 0.05). and VIF < 5.

Model Evaluation:

- Confusion matrix was made and cut off point of 0.41 was selected based on accuracy, sensitivity and specificity plot.
- As to solve business problem CEO asked to boost conversion rate to 80%. We have got the sensitivity / recall ~91.68% and accuracy as 81.38% on train data set

Making Predictions on Test Data:

- Making Predictions on Test: Scaling and predicting using final model.
- Evaluation metrics for test gave sensitivity ~92.32% and . accuracy as 81.02%

- Top 3 features are:
  - Tags_Closed by Horizzon
  - Tags_Lost to EINS
  - Tags_Will revert after reading the email

Recommendations:

- Top 3 features are:
  - Tags_Closed by Horizzon
  - Tags_Lost to EINS
  - Tags_Will revert after reading the email. Team should rigourously focus on this category
- Working professionals to be aggressively targeted as they have high conversion rate