Assignment Process Overview

The case study involved enhancing lead conversion rates for X Education through predictive analytics.

Data Cleaning:

Addressed missing values, replaced 'Select' with NaN, and handled columns with high missing values. Imputed missing values based on mode and mean, creating a cleaned dataset.

Data Analysis & Dummy Variable Creation:

Conducted extensive data analysis using visualizations to identify patterns and relationships in the dataset. Generated dummy variables for categorical columns, preparing the dataset for modeling.

Feature Scaling & Selection:

Scaled numerical features and performed feature importance analysis using Random Forest Classifier. Addressed multicollinearity by dropping highly correlated variables to refine the feature set.

Model Building & Evaluation:

Split the data and applied Logistic Regression with Recursive Feature Elimination for feature selection. Evaluated the model's performance using various metrics like confusion matrix, precision-recall trade-offs, and ROC curves. Optimized the model's threshold for prediction to improve performance.

Prediction on Test Set:

Prepared and predicted on the test set, assessing the model's performance with metrics and evaluation methods.

Learnings & Conclusion:

The model achieved around 80% accuracy with balanced precision and recall, indicating decent predictive capability. Key insights were drawn from data analysis, identifying influential features crucial for lead conversion. Threshold optimization enhanced model performance, ensuring better prediction outcomes. The model holds promise for aiding lead conversion predictions and targeted marketing efforts.