

Assignment Summary

The aim of this analysis was to identify strategies for X Education to attract more industry professionals to enroll in their courses. Through basic data analysis, information was gathered on customer website visits, time spent on the site, traffic sources, and conversion rates. The following steps were taken to perform the analysis:

- i. **Data quality checks:** The data was read and examined using commands like shape and info. Duplicate and missing values were identified and cleaned.
- ii. **Data cleaning and treatment:** The data was partially cleaned, with null values replaced and irrelevant variables removed.
- iii. **Exploratory data analysis (EDA):** A quick EDA was conducted, which involved dropping columns with high missing values, removing irrelevant variables, and addressing skewed data. EDA plots were used to identify insights.
- iv. **Dummy variable creation:** Categorical variables were converted to dummy variables, and numeric variables were resealed.
- v. **Train-test split:** The data was split into 70% for training and 30% for testing using the sk-learn package.
- vi. **Re-scaling** numerical variables.
- vii. **Model building:** RFE was used to obtain the top 15 relevant variables. Variables with high VIF and p-values were manually removed.
- viii. **Model evaluation:** A confusion matrix was created, and the optimum cut-off value was identified using the ROC curve. The model achieved around 90% accuracy, sensitivity, and specificity.
- ix. **Prediction:** The model was used to predict outcomes on the test data frame using an optimum cut-off of 0.2, with an accuracy, sensitivity, and specificity of 90%.

The analysis identified the variables that mattered most in attracting potential buyers, ranked in descending order. Using these insights, X Education has a high chance of attracting potential buyers and increasing course enrollments.