# Lead Scoring Case Study Summary:

**Problem Statement:**

X Education, an online education company, aims to increase its lead conversion rate by identifying and prioritizing "Hot Leads" with the highest potential for conversion. They want to implement a lead scoring model to assign scores to leads based on their likelihood of conversion. The objective is to focus their efforts on leads with higher scores and achieve a lead conversion rate of approximately 80%.

**Solution Summary:**

**Step 1:** Data Reading and Understanding:

- Read and inspected the data.

**Step 2:** Data Cleaning:

- Dropped variables with unique values.
- Replaced 'Select' values with null values.
- Dropped columns with more than 35% null values.
- Removed imbalanced and redundant variables.
- Imputed missing values with median values for numerical variables.
- Created new classification variables for categorical variables.
- Handled identical labels with different cases.
- Removed sales team-generated variables.

**Step 3:** Data Transformation:

- Transformed binary variables into '0' and '1'.

**Step 4:** Dummy Variables Creation:

- Created dummy variables for categorical variables.
- Removed repeated and redundant variables.

**Step 5:** Test Train Split:

- Divided the dataset into a 70-30 train-test split.

**Step 6:** Feature Rescaling:

- Applied Min-Max scaling to numerical variables.

- Checked correlations among variables using a heatmap.

- Dropped highly correlated dummy variables.

**Step 7:** Model Building:

- Used Recursive Feature Elimination to select the top 15 important features.

- Examined p-values and dropped insignificant variables.

- Selected 11 significant variables with good VIF values.

- Determined the optimal probability cutoff based on accuracy, sensitivity, and specificity.

- Plotted an ROC curve with an area under the curve of 86%.

- Verified prediction accuracy and evaluated model metrics on the training set.

- Selected a cutoff value of approximately 0.3 for the trade-off between precision and recall.

- Applied the final model to the test set, resulting in an accuracy of 77.52%, sensitivity of 83.01%, and specificity of 74.13%.

**Step 8:** Conclusion:

- The lead score calculated in the test set achieved a conversion rate of 83%, surpassing the target of 80%.

- The model's high sensitivity value is beneficial in identifying the most promising leads.

- Key features contributing significantly to conversion probability include Lead Origin (Lead Add Form), What is your current occupation (Working Professional), and Total Time Spent on the Website.

By implementing the lead scoring model and utilizing the insights gained from the analysis, X Education can effectively prioritize leads and improve their lead conversion rate, ultimately leading to increased success in their online education business.