**Problem Statement:**

X Education, an online course provider, is facing a significant challenge with a low lead conversion rate of approximately 30%. To address this issue and optimize its sales process, the company aims to implement a predictive model that assigns a lead score to each lead, prioritizing those with a higher likelihood of conversion.

**Approach:**

To develop the predictive model, a comprehensive data analysis and modelling approach was employed. The key steps involved:

1. **Data Exploration and Cleaning:**
   * **Exploratory Data Analysis (EDA):** Conducted to understand the distribution of variables, identify potential relationships, and uncover patterns in the data.
   * **Data Cleaning:** Addressed missing values, outliers, and inconsistent data to ensure data quality and model accuracy.
   * **Feature Engineering:** Created new features from existing ones to capture additional insights and improve model performance.
2. **Model Building and Evaluation:**
   * **Model Selection:** Evaluated various classification algorithms, including Logistic Regression with RFE to determine the most suitable features and created the model using GLM for predicting lead conversion.
   * **Model Training and Tuning:** Split the data into 70:30 ratio for training and testing. Trained and fine-tuned the selected model and finally used for predicting the test set lead conversion rate using threshold value optimization.
   * **Model Evaluation:** Assessed the model's performance using metrics such as accuracy, precision, recall, and ROC-AUC curve.
3. **Lead Scoring and Prioritization:**
   * **Calibrated Probabilities:** Calibrated the model's predicted probabilities to ensure accurate lead scoring.
   * **Lead Scoring Mechanism:** Assigned a lead score to each lead based on the predicted probability of conversion.
   * **Lead Prioritization:** Prioritized leads with higher scores for focused sales efforts.

**Challenges and Lessons Learned:**

* **Data Quality and Imputation:** Handling missing values and inconsistent data required careful consideration and appropriate imputation techniques.
* **Feature Engineering:** Creating relevant features from raw data was crucial for improving model performance. Experimentation with different feature engineering approaches was necessary to identify the most effective ones.
* **Imbalanced Dataset:** The dataset exhibited a class imbalance, with a higher proportion of non-converted leads. Addressing this imbalance required techniques and addressed the Select values for few of the features.
* **Model Selection and Tuning:** Choosing the right model and optimizing its hyperparameters was essential for achieving optimal performance.
* **Model Interpretability:** Understanding the model's decision-making process was important for gaining insights and building trust in the predictions.

By successfully addressing these challenges and applying the lessons learned, X Education can significantly improve its lead conversion rate and optimize its sales efforts.