We got an online education company dataset,which was struggling with a low lead conversion rate despite generating a large number of leads. The company wanted to improve the efficiency of their lead conversion process by identifying the most potential leads, or "hot leads", who are more likely to convert into paying customers. The CEO set the target lead conversion rate of around 80%. The company provided us the dataset of past leads, including various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc., along with the target variable 'Converted' which indicates whether a lead was converted or not. The task is to build a logistic regression model to assign a lead score between 0 and 100 to each lead, which can be used by the company to target potential leads more effectively.

Analysis Approach:

1. Data Understanding: we began by exploring and understanding the provided dataset, including checking for missing values, handling categorical variables with 'Select' levels, and analyzing the distribution of variables.

2. Data Cleaning: Then we cleaned the data by addressing missing values, handling categorical variables, and preparing the dataset for model building.

3. Feature Engineering: we Identified and created relevant features that could potentially impact lead conversion, such as lead source, total time spent on website, total visits, last activity, etc.

4. Model Building: Build a logistic regression model using the cleaned and transformed dataset, and evaluated its performance using appropriate evaluation metrics such as accuracy, precision, etc.

5. Model Interpretation: Interpreted the coefficients of the logistic regression model to understand the impact of different features on lead conversion, and identified the most significant factors that influence lead conversion.

6. Lead Scoring: Assigned a lead score to each lead based on the logistic regression model's predicted probabilities, with higher scores indicating higher conversion chances.

7. Model Evaluation: Validated the performance of the lead scoring model using appropriate evaluation metrics, and compared it with the initial lead conversion rate to assess its effectiveness in identifying potential leads.

8. Recommendations: Provided recommendations to the company based on the analysis results, including insights on significant features, lead scoring, and potential improvements in the lead conversion process.

Results in Business Terms: The logistic regression model developed in this analysis can assign lead scores between 0 and 100 to each lead, with higher scores indicating higher conversion chances. This lead scoring system can help X Education identify potential leads more efficiently and focus their efforts on communicating with the most promising leads, which can potentially increase the lead conversion rate. The model's performance can be evaluated using metrics such as accuracy, precision,etc to assess its effectiveness in identifying potential leads. The model's interpretation can provide insights on the significant factors that influence lead conversion, allowing X Education to tailor their marketing strategies and communication approach accordingly. The lead scoring system can be integrated into X

Education's existing lead management process to prioritize leads and allocate resources more effectively, potentially improving the overall conversion rate and increasing revenue.