Summary

The first step we took was to analyze the dataset to determine which variables were relevant and which variables needed to be preprocessed. We found that many of the categorical variables had a level called 'Select' which needed to be handled because it was as good as a null value. We also discovered that some variables, such as 'Lead Source', had a large number of unique categories, which could potentially affect the model's performance. Therefore, we decided to create a new category for all values with a frequency less than 100.

After preprocessing, we used logistic regression to build a predictive model that could assign a lead score to each lead. We trained our model on 70% of the data and tested it on the remaining 30%. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

Our final model achieved an accuracy of 80% and a precision of 73%. This means that our model correctly identified 73% of the potential customers who would convert into paying customers. Our model also identified several key factors that influenced a lead's conversion, such as the Total Time Spent on Website, Total Visits, and the Lead Origin.

In conclusion, lead scoring is a crucial process for companies like X Education, as it can help identify the most potential leads, resulting in higher lead conversion rates. With the help of data analytics and machine learning techniques, we were able to build a predictive model that can accurately assign a lead score to each lead. Our model's performance has helped X Education achieve its target lead conversion rate of 80%.