Summary Report

Project Overview:

The goal of this project was to assist X Education, an online education company, in improving its lead conversion rate by building a predictive model to identify potential leads. The dataset provided contained various attributes related to leads, including lead source, total visits, last activity, and conversion status. Our task was to develop a logistic regression model to assign lead scores and optimize the lead conversion process.

Methodology:

The analysis began with reading and understanding the dataset, which comprises 9240 entries and 37 columns. The dataset contains both numerical and categorical variables, along with missing values in some columns. The missing values were handled by forward filling, and categorical variables were encoded using one-hot encoding. After preprocessing, we split the data into training and testing sets and built a logistic regression model using the training data. The model was trained to predict the probability of lead conversion based on the available features.

Model Building and Evaluation:

Two models were trained and evaluated: Linear Regression and Logistic Regression. In the case of Linear Regression, the Ridge regression variant was utilized due to its ability to handle multicollinearity. The mean squared error (MSE) and R-squared were used as evaluation metrics, resulting in an *MSE of 0.0986* and an *R-squared of 0.5896*.

For Logistic Regression, the coefficients of the model were extracted and analyzed to identify the top three variables contributing most towards lead conversion probability. The top three variables were identified as follows:

1. Total Time Spent on Website
2. Lead Number
3. Tags indicating the lead will revert after reading the email

Recommendations:

Based on our analysis, we recommend that during the period when X Education reaches its sales target for a quarter ahead of schedule and aims to minimize the rate of useless phone calls, here's a suggested strategy they could employ:

a) Improve Lead Scoring
b) Use Automated Emails and Ads
c) Share Helpful Content
d) Be Active on Social Media
e) Analyze Data to Improve
f) Work with Others

With these changes, X Education can keep sales going strong without making too many phone calls when they're ahead of schedule.

Conclusion:

In conclusion, the analysis demonstrates the importance of various factors in predicting lead conversion probability for X Education. By implementing the recommendations outlined in this report, X Education can enhance its marketing and sales strategies, ultimately leading to increased customer acquisition and revenue generation.