Summary Report
Leads Case Study

Introduction:

The goal of this case study was to analyze and improve the lead conversion rate for X Education, a company that offers online courses. The data used for the analysis was a CSV file containing leads data, which was processed and analyzed using Python and various libraries such as Pandas, Seaborn, Matplotlib, Sklearn, and Statsmodel.

Data Analysis and Model Building:

The data was cleaned and preprocessed, including handling missing values, dropping irrelevant columns, creating dummy variables for categorical columns, and addressing outliers. Exploratory data analysis (EDA) was conducted to gain insights into the data and identify potential patterns. A logistic regression model was then built using Recursive Feature Elimination (RFE) to select important features, and Variance Inflation Factor (VIF) to check for multicollinearity. Columns with high p-values and VIF were eliminated to improve the model's accuracy and interpretability.

Model Evaluation and Results:

The final logistic regression model achieved an accuracy of 91%, sensitivity (recall) of around 81%, and specificity of around 97% using a cutoff value of 0.3. The model was also applied to test data, and the results showed an accuracy of 90%, sensitivity of 96%, specificity of around 87%, precision score of 83%, and recall score of 96%. These results indicate that the model has good predictive performance and can effectively identify potential leads with a higher likelihood of converting.

Recommendations:

Based on the analysis, several recommendations were made to improve the lead conversion rate for X Education. These include:

Focusing on leads that have shown higher engagement with the website, such as those with higher total time spent on the website and higher page views per visit.

Targeting leads that have shown specific interests, such as those who have submitted a free demo request or have been in touch with the company's sales team.

Providing personalized communication and follow-up to leads who have shown higher intent to convert, such as those with a higher lead score or have attended webinars or workshops. Continuously monitoring and analyzing data to identify new patterns and opportunities for optimization.

Conclusion:

The case study successfully analyzed the leads data, built a logistic regression model, and provided recommendations to improve the lead conversion rate for X Education. The code and results are available in this repository, and can be used as a reference for similar analyses or as a starting point for further optimization efforts.