SUMMARY REPORT

Introduction

The goal of this analysis was to enhance lead conversion rates by identifying key factors that influence customer conversions. By employing a data-driven approach, we aimed to uncover actionable insights and recommend strategies for improving marketing efficiency and effectiveness.

Data Cleaning

The dataset contained several columns with null values, which required cleaning for accurate analysis. Columns with more than 15% null values were dropped, as they were deemed too sparse to provide meaningful insights. For numerical columns with missing values, the median was used to fill in gaps, ensuring that the central tendency of the data was preserved without being unduly influenced by outliers. For Object datatype columns, the mode was used to fill in missing values, maintaining the most common category and minimizing disruption to the dataset's structure.

Exploratory Data Analysis (EDA)

During the EDA phase, we explored various aspects of the dataset to understand patterns and relationships:

1. Conversion Rates:

- **Overall Conversion:** We found that 38.54% of leads converted, indicating a significant portion of leads did not result in sales.

- **Lead Sources and Activities:** Analysis revealed that certain lead origins and last activities were more effective in converting leads. For example, leads from landing page submissions had higher conversion rates.

2. Impact of Do Not Email/Call Preferences:

- Leads who opted out of emails or calls showed different conversion behaviors. Although the conversion rates via email and calls were similar (38.5%), a substantial number of leads preferred not to be contacted via phone.

3. Website Engagement:

- Leads who spent more time on the website were more likely to convert. Interestingly, leads with 1-10 visits showed higher conversion rates, indicating optimal engagement.

4.Demographic and Behavioral Insights:

- Analysis of variables like time spent on the website, page views per visit, and last activity provided deeper insights into lead behavior and preferences.

Feature Selection and Model Building

After cleaning the data, we dropped non-contributory columns like `Prospect ID` and `Lead Number`. Dummy variables were created for categorical data, enabling us to use these features in our models. Using Recursive Feature Elimination (RFE), we selected the most relevant features for our logistic regression model.

Model Evaluation

The logistic regression model was evaluated using several metrics:

- **Accuracy:** Initial accuracy was promising, but further optimization was needed.
- **ROC Curve and AUC:** The model's AUC score of 0.87 indicated good predictive power.
- **Optimal Cutoff Point:** By analyzing the precision-recall tradeoff, we identified 0.36 as the optimal cutoff for making predictions.

Test Set Predictions

We validated the model on the test set, achieving an accuracy of approximately 79.6%. The model's precision and recall were also calculated, providing a balanced view of its performance.

Lessons Learned

1. Importance of Data Cleaning: Effective data cleaning is crucial for accurate analysis. Handling missing values appropriately ensures that the dataset remains robust and reliable.

2. Feature Selection: Using techniques like RFE helps in building a more efficient model by focusing on the most impactful features.

3. Model Evaluation: Continuous evaluation and optimization of the model are necessary to improve its performance and reliability.

4. Business Insights: Translating technical findings into business insights is essential for actionable recommendations. Understanding customer behavior and preferences can significantly enhance marketing strategies.

Conclusion

The analysis provided actionable insights into lead conversion factors and identified strategies to optimize marketing efforts. Implementing these recommendations and continuously monitoring performance will drive better results and improve lead conversion rates.