Summary Report: Lead Scoring Case Study

The lead scoring case study aimed to optimize lead conversion rates by identifying factors influencing conversion and predicting the likelihood of success for each lead. The project involved building a predictive model using logistic regression, supported by a comprehensive process of data cleaning, exploratory data analysis (EDA), and feature selection.

The process began with **data understanding and preparation**. The dataset, *Leads.csv*, was explored for missing values, inconsistencies, and redundant columns. Variables with more than 3000 missing data were removed, and entries like "Select," representing unfilled responses, were treated as null. Categorical variables were encoded using dummy variables, while numerical features were scaled using **MinaxScaler** to ensure compatibility with the model. These steps resulted in a clean, structured dataset, ready for analysis.

In the **EDA phase**, key features were explored to understand their relationships with the target variable, conversion status. Correlation analysis and visualizations, including heatmaps, boxplots, and distribution plots, revealed significant predictors like Total Time Spent on Website, Lead Origin, and Last Activity. These insights highlighted the importance of engagement metrics in lead conversion, guiding further analysis.

The **modeling phase** focused on logistic regression. A Recursive Feature Elimination (RFE) approach was used to select the most relevant features, reducing multicollinearity and improving interpretability. The data was split into training and testing sets, and the model was trained and evaluated on metrics like accuracy, precision, recall. The model showed robust performance, demonstrating its suitability for the task.

The final phase involved extracting **business insights and recommendations**. Key predictors provided actionable strategies, such as prioritizing leads with high engagement, refining CRM systems to incorporate real-time scoring, and investing in features driving conversions. Monitoring model performance and periodic retraining were recommended to ensure long-term effectiveness.

Key Learnings

- **Data Preparation**: Effective handling of missing data and categorical encoding ensured a robust dataset.
- **Feature Selection**: Identifying the most impactful variables improved model interpretability and performance.
- **Model Utility**: Logistic regression proved a practical, scalable solution for lead scoring, balancing simplicity with effectiveness.
- **Insights and Actionability**: The study emphasized how predictive analytics can drive targeted strategies and improve business outcomes.

In conclusion, this case study demonstrated a structured approach to predictive modelling, offering insights that are not only actionable but also scalable for broader applications.