Predicting Lead Conversion:

A DATA-DRIVEN APPROACH TO SALES PRIORITIZATION

PROBLEM STATEMENT AND OBJECTIVE

Problem Statement:

Predict the likelihood of lead conversion to help the sales team prioritize efforts.

Objective:

Conduct data cleaning and preparation for effective analysis.
Build, tune, and evaluate machine learning models to predict lead conversion.
Provide actionable insights based on model results.

ANALYSIS APPROACH

Data Cleaning:

Handled missing values by dropping columns with >25% missing data and imputing others.
Replaced irrelevant values (e.g., "Select") in categorical columns with NaN.

Feature Engineering:

Dropped high-cardinality and irrelevant columns (e.g., "Prospect ID").
Applied one-hot encoding for categorical variables.

Model Building:

Used Logistic Regression, Random Forest, and SVM.
Tuned hyperparameters using GridSearchCV.

Evaluation:

Assessed models using metrics like accuracy, precision, and recall.

RESULTS IN BUSINESS TERMS

Key Insights:

Random Forest identified the most important features contributing to lead conversion.
Logistic Regression and SVM provided lead scores (probabilities) for prioritization.

Business Recommendations:

Focus on leads with high conversion probabilities for aggressive follow-up.
Minimize efforts on low-probability leads to reduce unnecessary calls

VISUALIZATIONS

Lead Conversion Funnel:

Bar plot showing the number of initial leads vs. converted leads.
Highlights the conversion rate and potential for improvement.

Feature Importance:

Top features contributing to lead conversion (e.g., "Lead Source", "Last Activity").
Helps prioritize efforts on impactful variables.

SUMMARY OF RESULTS

Top Features:

Random Forest identified the top 3 features driving lead conversion.

Categorical Variables:

Focus on key dummy variables (e.g., "Lead Source_Email", "Last Activity_SMS").

Strategies:

Aggressive follow-up for high-probability leads during peak periods.
Reduced efforts on low-probability leads during off-peak periods.