# Analysis Report

## Insights from Feature Importance and Model Comparisons

Feature importance analysis was performed using various models, including Linear Regression, Decision Tree, and Random Forest. Key findings are as follows:  
1. Linear Regression coefficients indicated the relationship and magnitude of features with the target variable.  
2. Decision Tree revealed hierarchical splits of features, showcasing non-linear relationships.  
3. Random Forest provided feature importance scores based on ensemble learning, highlighting the robustness of influential predictors.

Comparisons between models showed that:  
- Random Forest consistently performed better in capturing complex relationships.  
- Decision Tree was easier to interpret but prone to overfitting.  
- Linear Regression struggled with high variance in data.

## Evaluation Metrics for Each Model

Regression Metrics:  
- Linear Regression: High Mean Squared Error (MSE), indicating poor fit for non-linear data.  
- Decision Tree: Moderate MSE, showing overfitting on training data but better generalization than Linear Regression.  
- Random Forest: Lowest MSE, signifying the best predictive performance among regression models.

Classification Metrics:  
- Logistic Regression: Reasonable accuracy, precision, recall, and F1-score, suitable for binary classification tasks.  
- Random Forest (Classification): High precision and recall, indicating balanced predictions for imbalanced datasets.

## Justification for the Most Suitable Model

Based on evaluation metrics and feature importance, Random Forest emerged as the most suitable model due to:  
- Superior performance in both regression and classification tasks.  
- Robustness to overfitting compared to Decision Tree.  
- Enhanced interpretability through feature importance visualization.

While Linear Regression and Logistic Regression are simpler and interpretable, they were less effective with the given dataset's complexity.