Titanic Dataset - EDA & Modeling Summary

This report summarizes the exploratory data analysis (EDA) and machine learning modeling performed on the Titanic dataset. The goal was to explore feature relationships with survival and build basic predictive models (Logistic Regression and Random Forest).

Key EDA Findings: Overall survival rate: ~38.4%. Passenger class strongly influences survival (1st class highest survival). Sex is highly predictive: females survived at much higher rates than males. Children (Age ≤ 12) had higher survival rates. Family size and being alone both impact survival probability. Engineered features: Title (from Name), FamilySize, IsAlone improved prediction.

Model Performance Metrics (Test Set)

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.8436	0.8154	0.7681	0.7910
Random Forest	0.8101	0.7778	0.7101	0.7424

In this analysis, Logistic Regression slightly outperformed Random Forest on the test split. However, Random Forest offers interpretability via feature importance. The engineered features, especially Title and FamilySize, likely contributed to the performance. Further tuning or advanced models may yield higher accuracy.