**Titanic Survival Prediction Analysis**

**Project Overview** This analysis investigates the Titanic dataset to understand and predict factors influencing passenger survival. Using logistic regression, the project aims to create a reliable predictive model for survival outcomes.

**Data Cleaning Process**

1. **Missing Data Imputation**: Missing values in the Age column were filled using median ages based on passenger class.
2. **Column Removal**: Dropped the Cabin column due to excessive missing values.
3. **Data Transformation**: Converted categorical columns (e.g., Sex, Embarked) into binary features.

**Exploratory Data Analysis (EDA)**

* **Gender and Survival**: Visualizations reveal higher survival rates among female passengers.
* **Class and Survival**: First-class passengers had significantly higher survival rates.
* **Age Distribution**: Younger passengers generally had higher survival rates.
* **Family Size**: Traveling with family members positively impacted survival odds, observed through sibling/spouse counts.

**Modeling and Results**

Using logistic regression, a model was developed to predict survival. Key performance metrics included:

* **Accuracy**: 80%
* **Precision, Recall, F1 Score**: Evaluated to assess model reliability.

The model accurately captured survival patterns and confirmed the impact of passenger class, gender, and family size on survival odds.

**Conclusion**

This analysis highlights demographic and socioeconomic factors affecting survival on the Titanic. The logistic regression model performed well, achieving 80% accuracy, and validated that factors such as gender, age, and class significantly influenced survival likelihood. Future steps could include testing additional models or tuning parameters for higher accuracy.