

Task 5 - Exploratory Data Analysis (Titanic Dataset)

Dataset: Titanic train.csv (rows: 891, columns: 12)

Missing values: Top missing columns - {'Cabin': 687, 'Age': 177, 'Embarked': 2}.

Survival rate overall: 38.38%

Observations:

- Females had a higher survival rate than males (see 'sex_survived.png').
- Higher passenger class (1st) had better survival rates compared to lower classes.
- Age distribution shows many passengers were young adults; Age had missing values handled by imputation in example.
- Fare distribution is skewed to the right; outliers present.
- Correlation heatmap shows moderate correlation between Pclass and Fare (negative/positive as expected).

Recommendations: handle missing Age (median or predictive imputation), engineer titles from Name, and one-hot encode categorical variables before modelling.

Key Visualizations:









