Exploratory Data Analysis (EDA) Report - Titanic Dataset

Objective: Extract insights using visual and statistical exploration.

Tools Used: Python (Pandas, Matplotlib, Seaborn)

EDA Steps:

1. Loaded train and test datasets.
2. Used .info(), .describe(), and .value_counts() to understand data.
3. Checked for missing values — 'Age', 'Cabin', and 'Embarked' had nulls.
4. Created visualizations: histograms, boxplots, scatterplots, and pairplots.
5. Explored relationships between features and survival.

Key Findings:

1. Females had a much higher survival rate than males.
2. Higher-class passengers (Pclass 1) were more likely to survive.
3. Younger passengers had better survival rates.
4. Passengers paying higher fares had higher chances of survival.
5. Passengers from 'C' port of embarkation had slightly better survival chances.
6. Large families (SibSp + Parch > 3) had lower survival probabilities.
7. Cabin column had excessive missing values, less reliable for analysis.

Feature	Observation
Sex	Females survived more frequently than males.
Pclass	Higher classes had better survival rates.
Age	Younger passengers more likely to survive.
Fare	Higher fares correlated with survival.
Embarked	Port 'C' had slightly more survivors.

Conclusion: This analysis highlights demographic and socio-economic factors influencing Titanic passenger survival. These insights can support predictive modeling or deeper feature engineering.