Title: Exploratory Data Analysis on Titanic Dataset

Dataset Overview

- Dataset contains information about Titanic passengers including survival status.
- Total entries in training set: 891 rows, 12 columns.
- Contains both numerical and categorical features.

Missing Value Analysis

- Age: Moderate missing values (~20%) → Imputed with **median**
- Embarked: Few missing values → Filled with **most frequent value (mode)**
- Cabin: Very high missing values → Filled with **'Unknown'** label

Key Statistical Insights

- **Age** distributed mostly around young adults and middle aged.
- **Fare** is highly right-skewed with extreme outliers indicating wealth disparity.

Univariate Insights

- Most passengers are from **3rd class**, indicating lower socioeconomic status.
- Majority of passengers paid low fares.
- Males are significantly higher in number than females.

Bivariate Insights

- **Females survived more** frequently than males.
- **First-class passengers** had a significantly higher survival rate.
- **Younger passengers** had better chances of survival.

Multivariate / Correlation Insights

- Positive correlation between **Fare** and Survival.
- Negative correlation between **Pclass** and Survival.
- SibSp and Parch show mutual correlation due to family traveling together.

Conclusion

Survival outcomes show strong dependence on:

- **Gender** (females prioritized)
- **Passenger Class / Fare** (wealthier passengers survived more)
- **Age** (children more likely saved)