Exploratory Data Analysis(EDA) Report – Titanic Dataset

Introduction

Data analysis to find patterns and trends in Titanic dataset.

Dataset Overview

- Tools, Rows, Columns.
- Missing values summary.
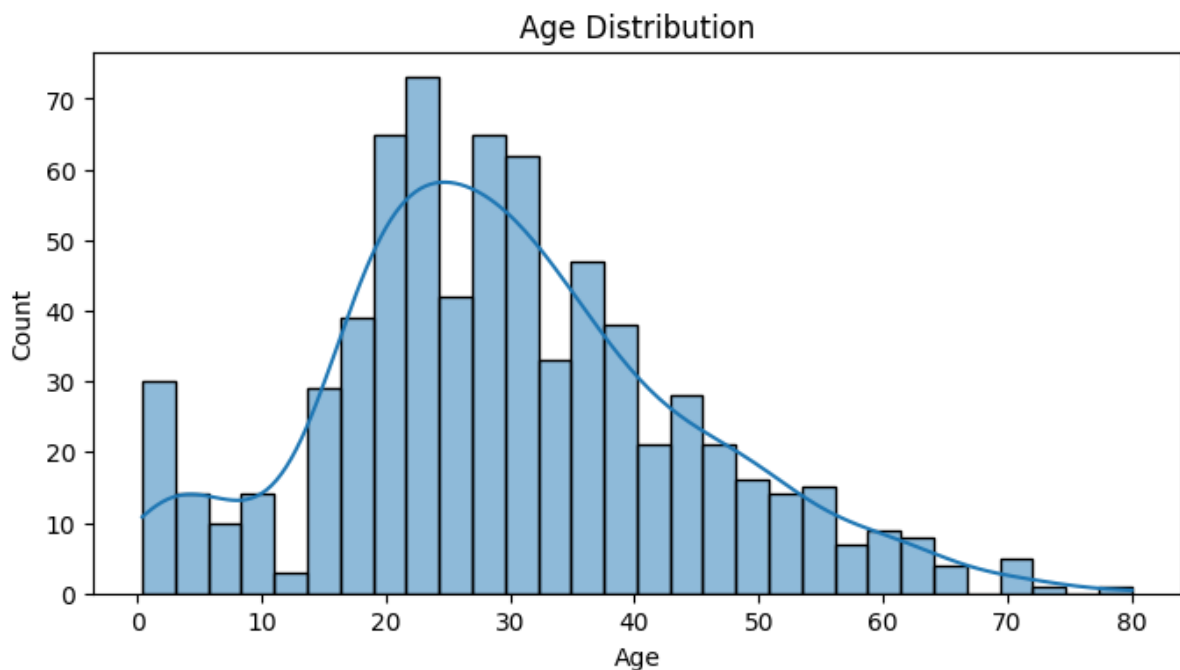- Key columns description. (Age, gender, etc.)

Dataset Summary

- Output from .info() and .describe()

- Example: "Age mean is 29.7 years, fare distribution is right skewed."
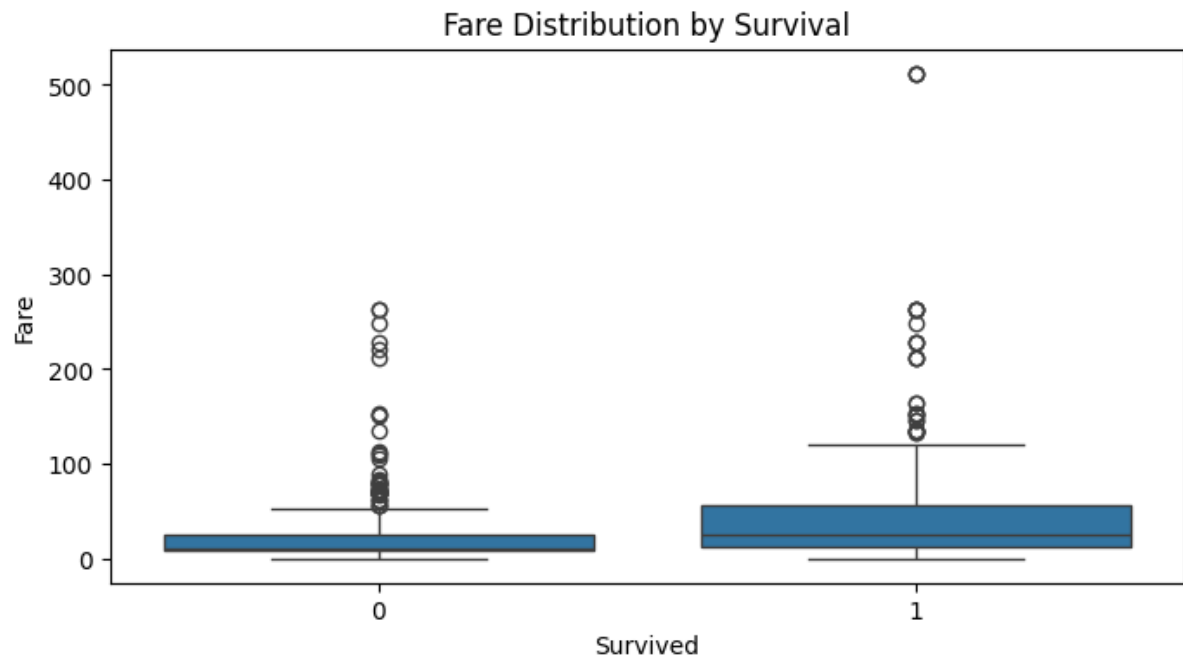
Visualization and Observations

Age Distribution Histogram

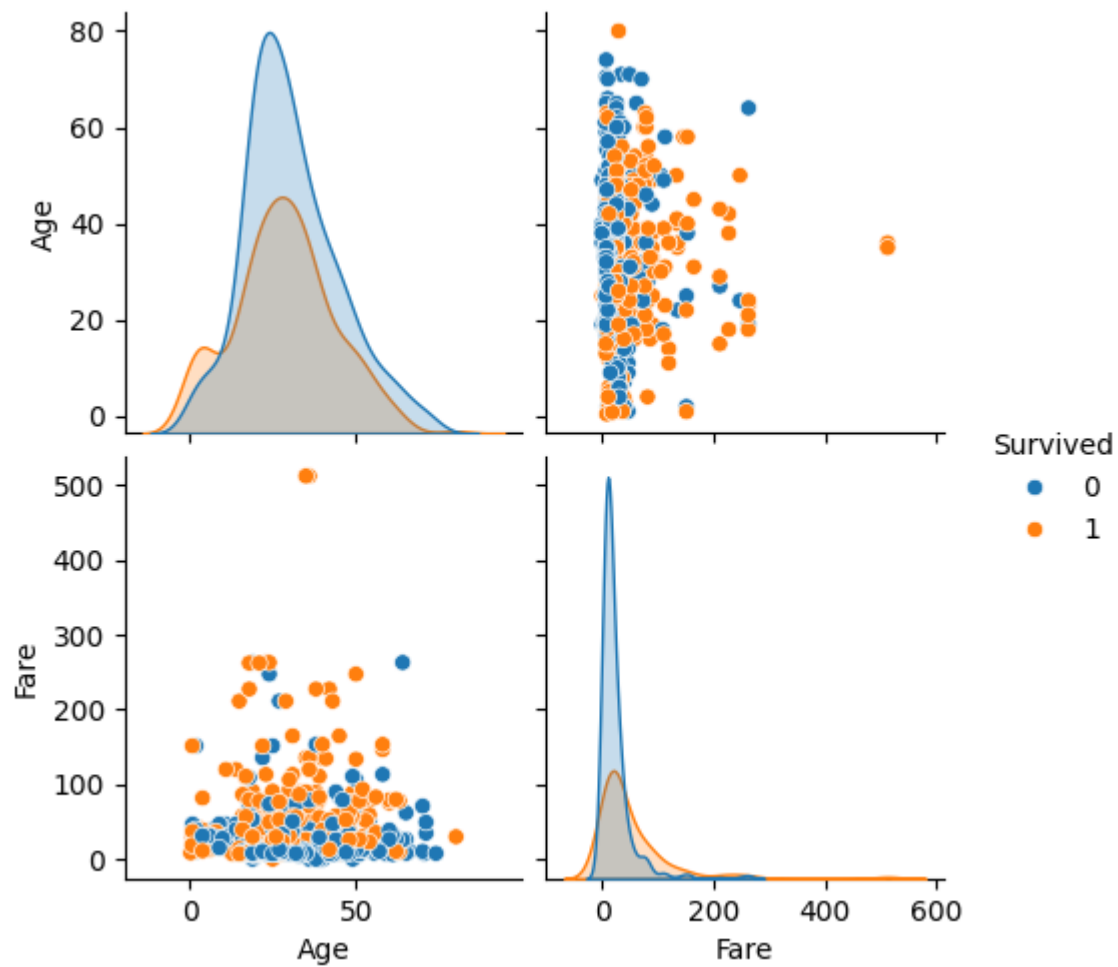*Observation:* Most passengers are between 20-40 years. Some missing Age values.

Fare Boxplot by Survival

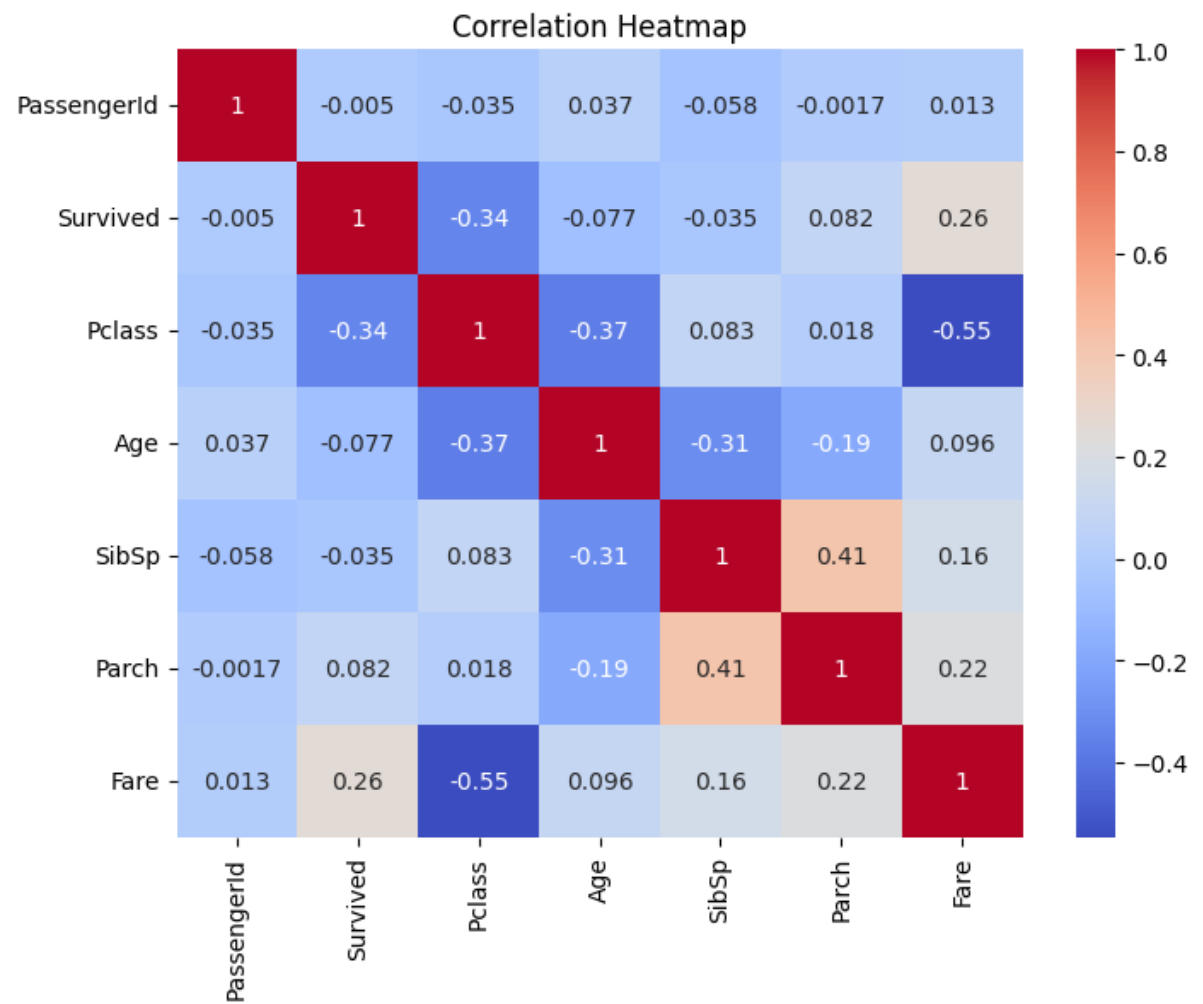*Observation:* Survivors generally paid higher fare compared to non-survivors.

Pairplot (Age, Fare, Survival)

Observation: Age and Fare show some correlation with survival.

Correlation Heatmap

Observation: Pclass and Fare negatively correlated with survival.

Summary of Findings

- Age distribution is slightly right-skewed.
- Higher fare passengers had better survival chances.
- Missing Age values might affect analysis accuracy.

Conclusion

EDA helped understand data better and future analysis direction.