Titanic Data Analysis Report
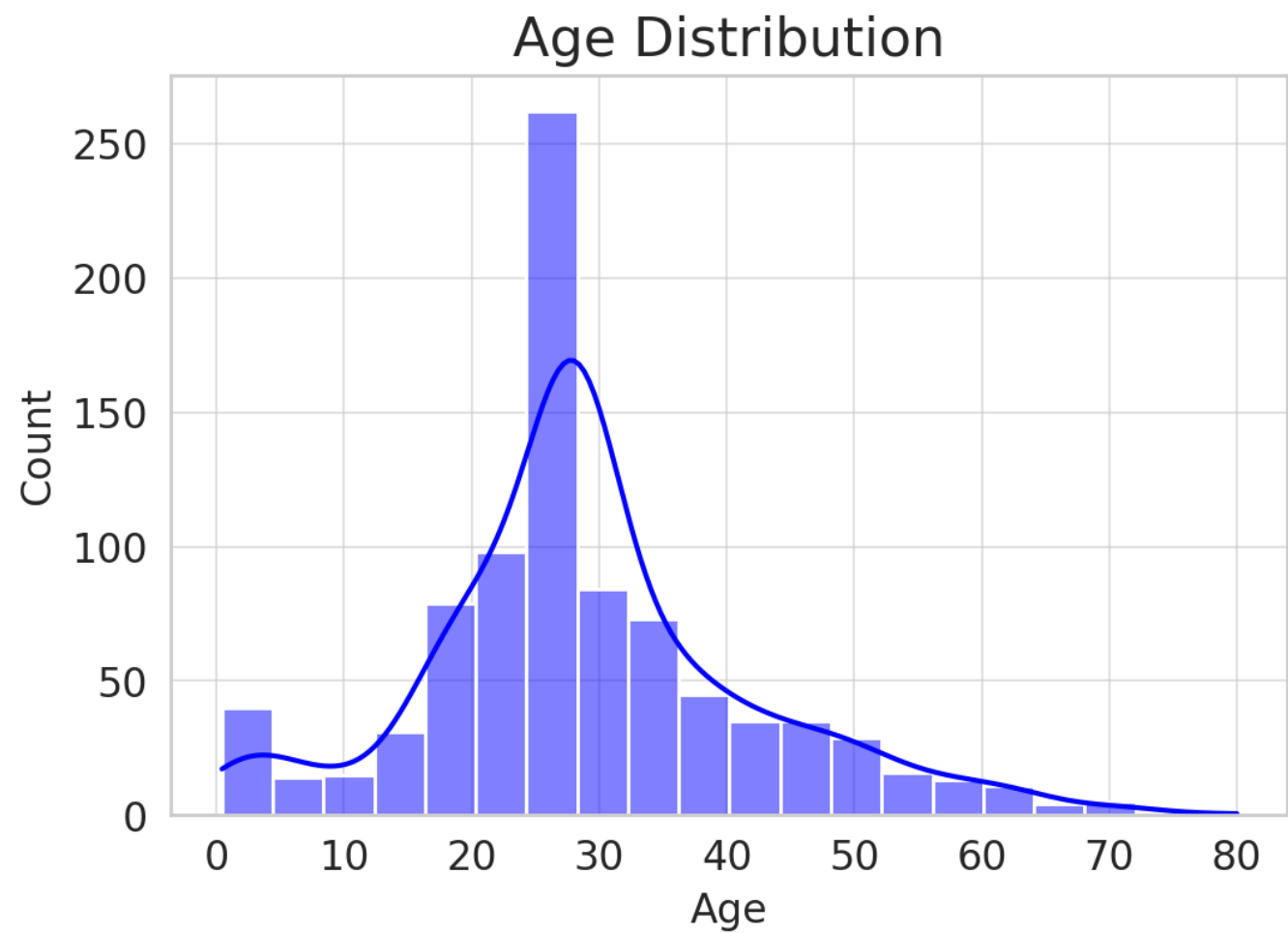
1. Data Cleaning Summary

- Missing values in 'Age' were filled with the median.

- Missing values in 'Embarked' were filled with the mode.

- 'Cabin' column was dropped due to excessive missing values.

- No duplicate entries were found in the dataset.

2. Exploratory Data Analysis (EDA)

- Age distribution shows most passengers were between 20-40 years old.

- Fare distribution is right-skewed, indicating some passengers paid much higher prices.

- Women had a much higher survival rate across all classes.

- First-class passengers had the highest survival rate, third-class the lowest.

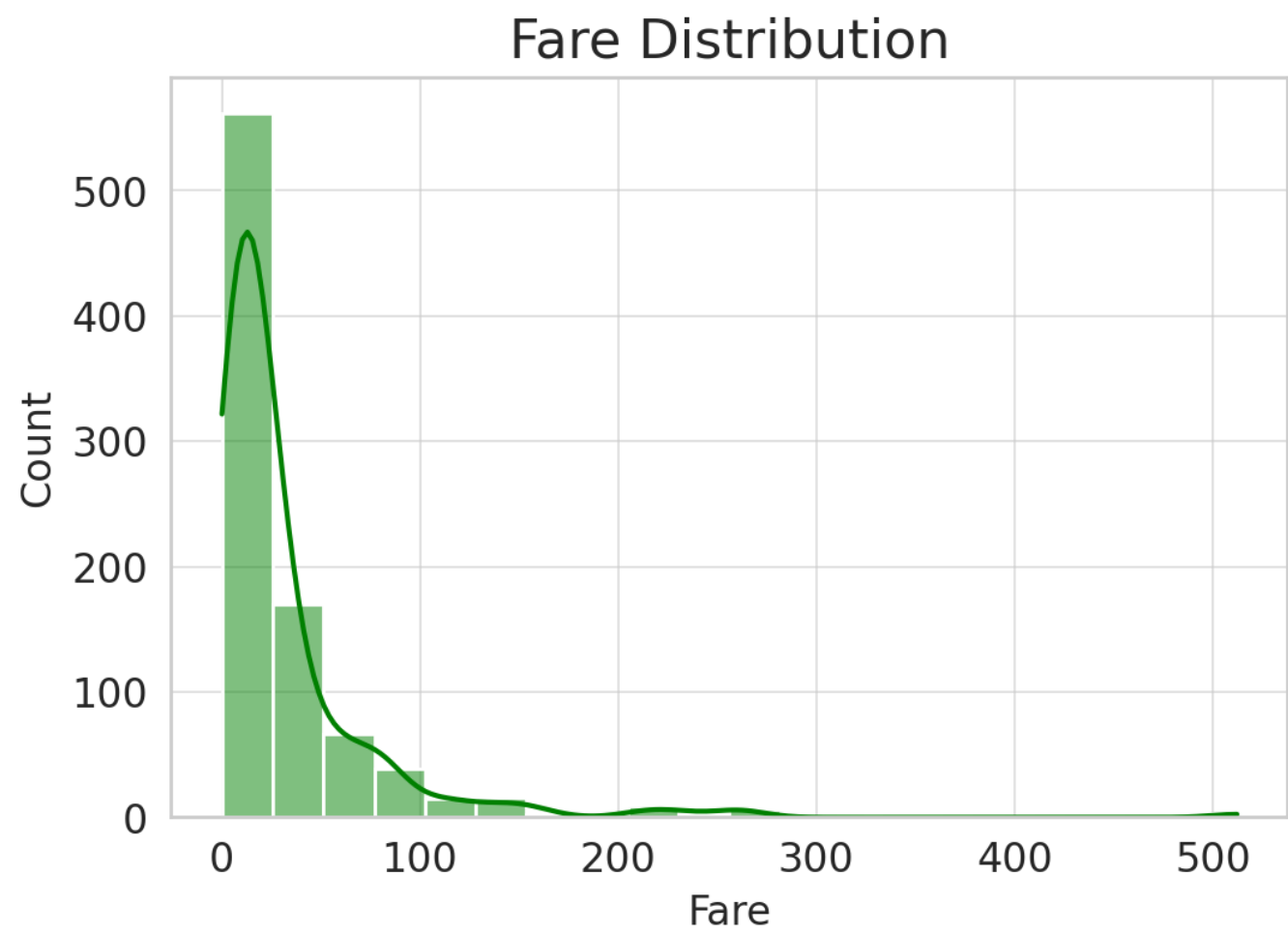- Strong correlation between fare, class, and survival.

Age Distribution

Fare Distribution

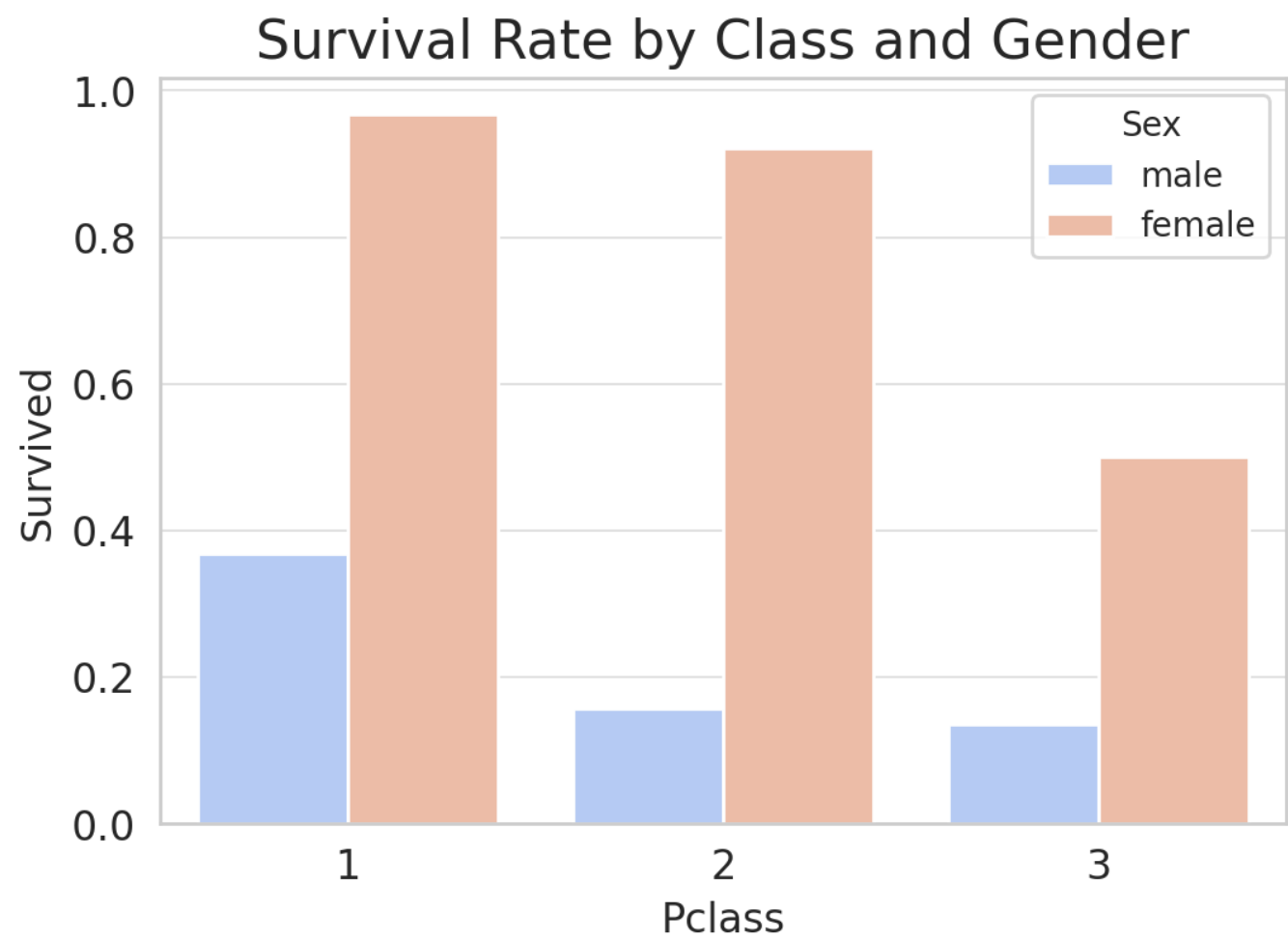**Survival Rate by Class and Gender**

Titanic Data Analysis Report

Correlation Heatmap

Correlation Heatmap

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.00	-0.01	-0.04	0.03	-0.06	-0.00	0.01
Survived	-0.01	1.00	-0.34	-0.06	-0.04	0.08	0.26
Pclass	-0.04	-0.34	1.00	-0.34	0.08	0.02	-0.55
Age	0.03	-0.06	-0.34	1.00	-0.23	-0.17	0.10
SibSp	-0.06	-0.04	0.08	-0.23	1.00	0.41	0.16
Parch	-0.00	0.08	0.02	-0.17	0.41	1.00	0.22
Fare	0.01	0.26	-0.55	0.10	0.16	0.22	1.00