

Titanic Dataset - EDA Summary

Dataset Overview:

- **Rows:** 891
- **Columns:** 9
- **Key Variables:** Survived, Age, Fare, Sex, Pclass, SibSp, Parch, Embarked

Missing Values:

- Age - 20% missing values (handled by median imputation)
- Embarked - 2 missing entries (filled with the mode)

Key Insights:

1. Survival Rate:

- **Overall Survival Rate:** 38%
- **Females** had a much higher survival rate compared to males.

2. Passenger Class:

- Higher-class passengers (Pclass1) had better survival rates than lower classes.

3. Age Distribution:

- Most passengers were aged between 20-40 years.
- Children (Age < 10) had relatively higher survival rates.

4. Fare:

- Passengers who paid higher fares were more likely to survive.

5. Family Influence:

- Small families (1–2 members) had better survival rates than those traveling alone or with large families.

6. Embarked Port:

- Most passengers boarded at 'S' (Southampton).
- Passengers boarding from 'C' (Cherbourg) had the highest survival rate among ports.

7. Correlation Heatmap:

- Strong negative correlation between Pclass and Survived (higher class → higher survival).
- Moderate positive correlation between Fare and Survived.

Visuals Included:

- Histograms
- Boxplots
- Pairplots
- Correlation Heatmap
- Barplots

Conclusion:

- Gender, Passenger Class, Fare, and Family Size were major factors affecting survival.
- Females, First Class passengers, and those with higher fares had better survival chances.
- Traveling with a small family and boarding from Cherbourg also slightly increased survival odds.