

## # Summary of Findings - Titanic EDA

### ## 🔎 Data Cleaning Summary:

- Filled missing Age with median
- Filled missing Embarked with mode
- Dropped Cabin due to large null values
- Created new features:
  - FamilySize = SibSp + Parch + 1
  - IsAlone = 1 if FamilySize == 1 else 0

### ## 📊 Visualization Insights:

- ♦ Survived vs Sex: Females had a much higher survival rate than males
- ♦ Survived vs Pclass: 1st class passengers had highest survival rate
- ♦ Survived vs Age: Children had better survival rate; elderly less so
- ♦ Correlation Heatmap: Fare, Pclass, and Sex were strongly correlated with survival
- ♦ IsAlone: Solo travelers had lower chances of survival

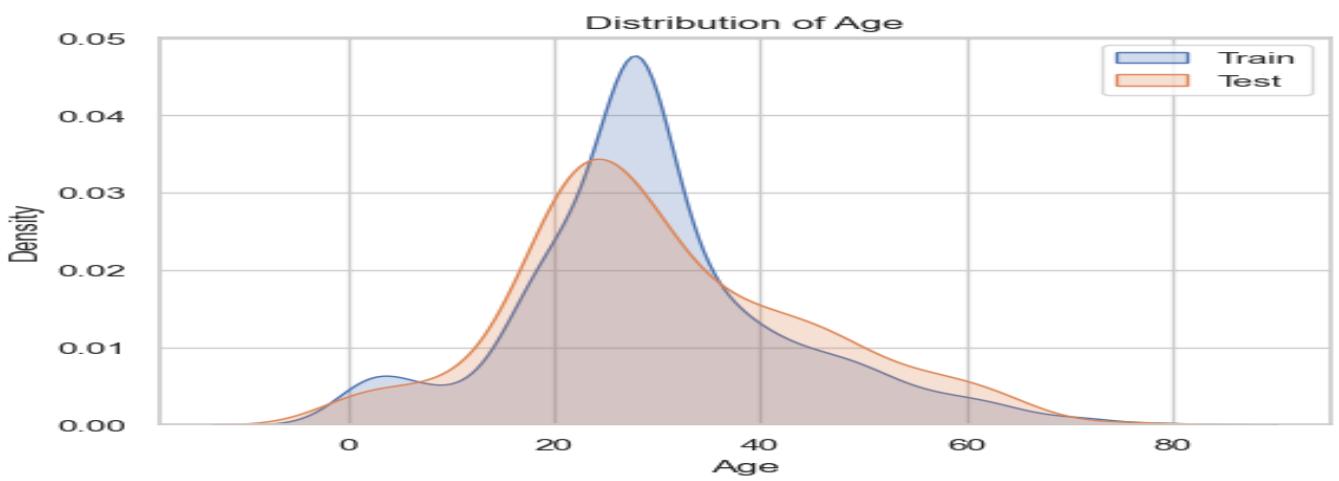
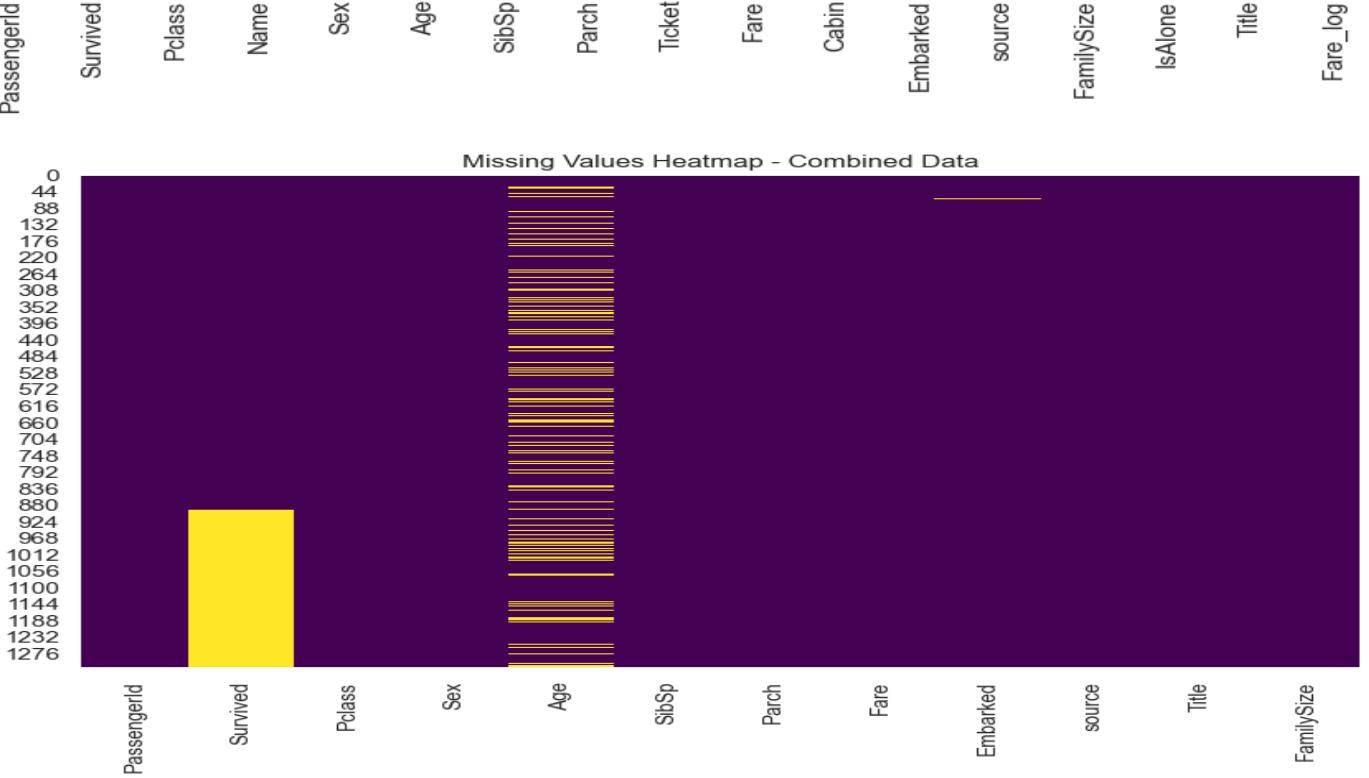
### ## 🚩 VIF Analysis:

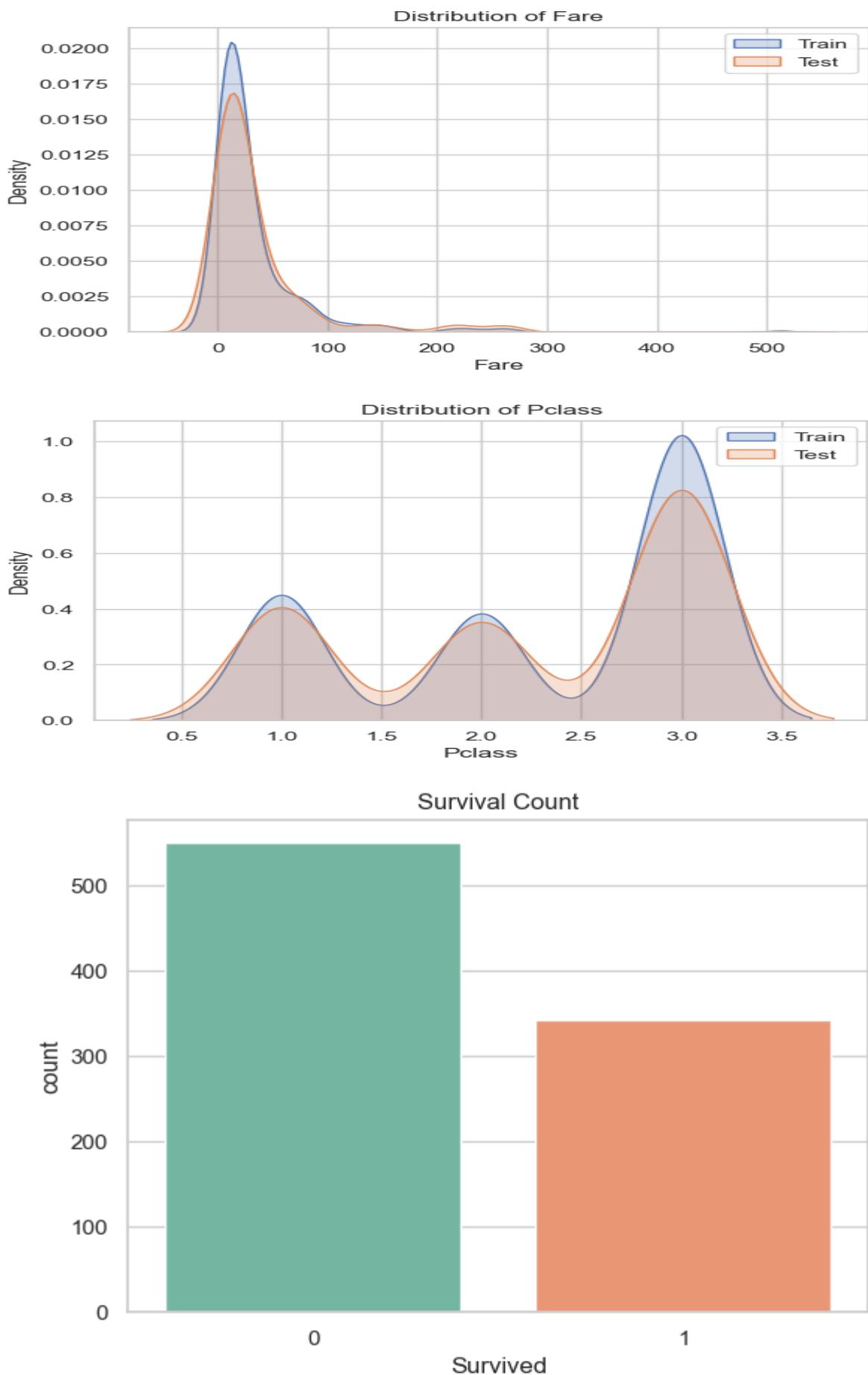
- Checked multicollinearity between Age, Fare, FamilySize
- No VIF > 5 — no significant multicollinearity

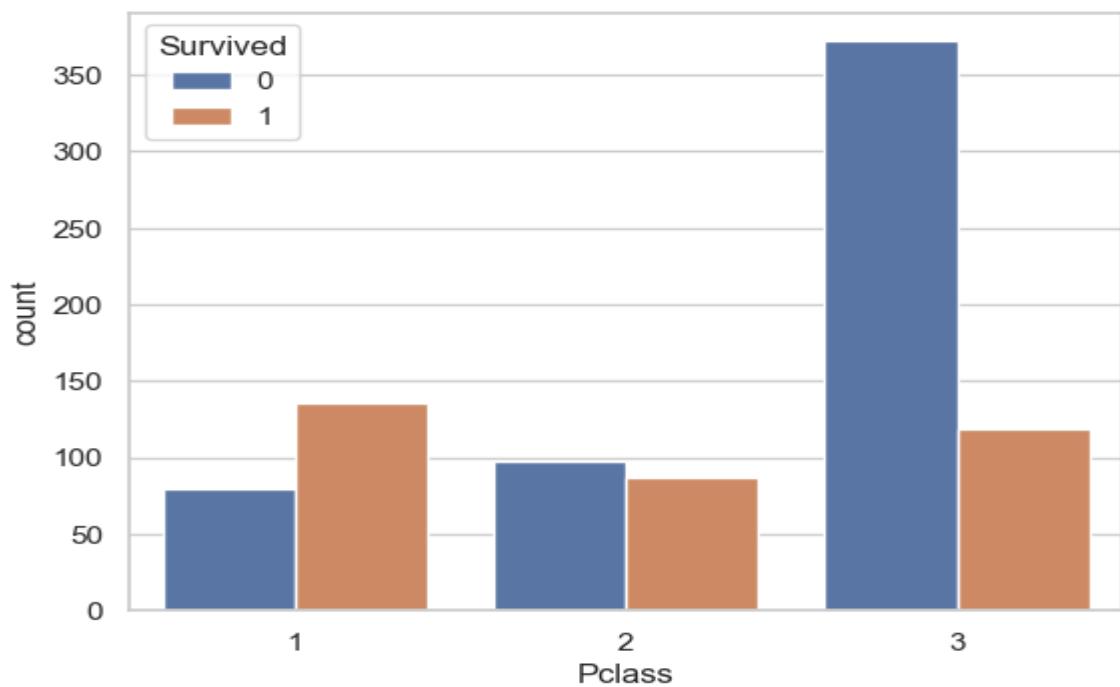
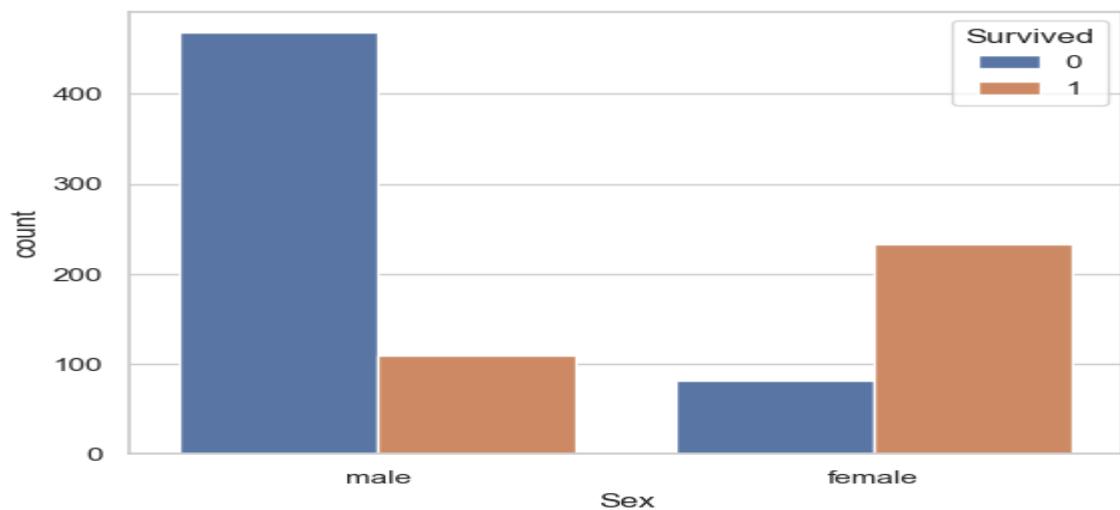
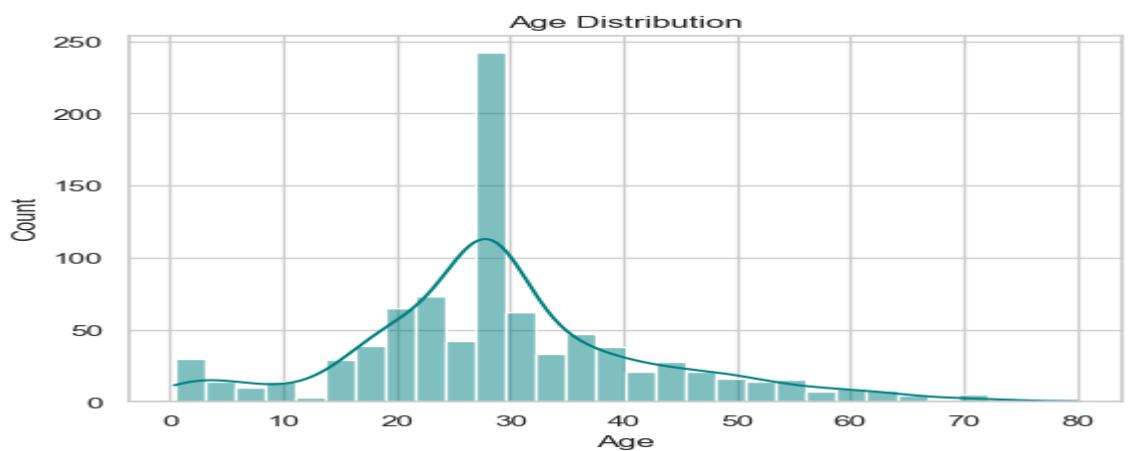
### ## 📦 Summary:

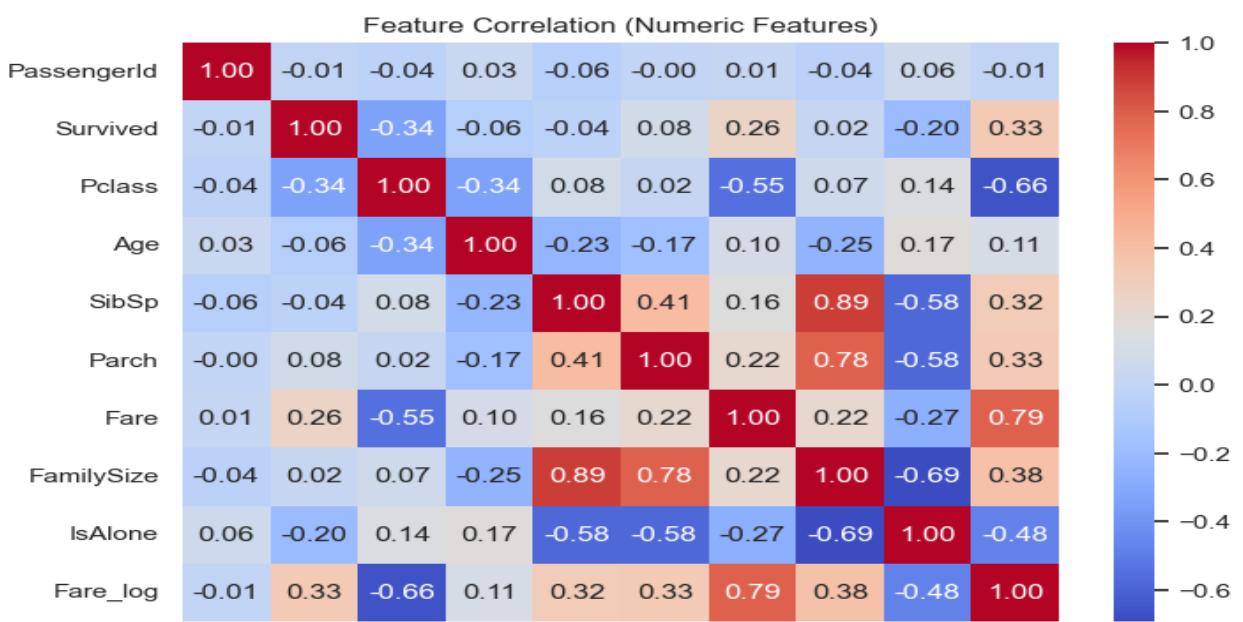
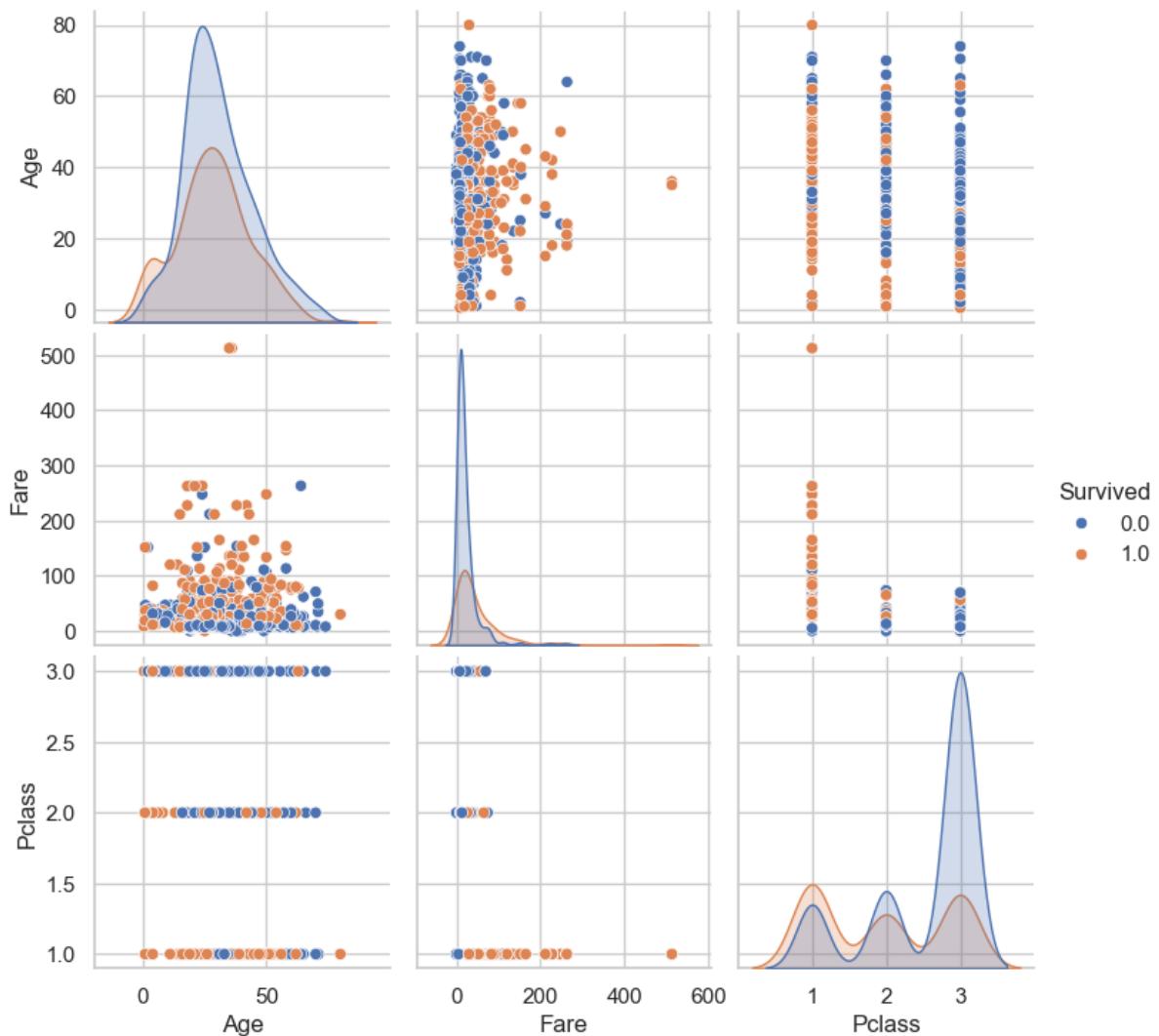
- Cleaned and explored the Titanic dataset
- Engineered meaningful features
- Found key survival patterns useful for model building

Prepared for GitHub and presentation in both notebook and report format.









PassengerId      Survived      Pclass      Age      SibSp      Parch      Fare      FamilySize      IsAlone      Fare\_log

