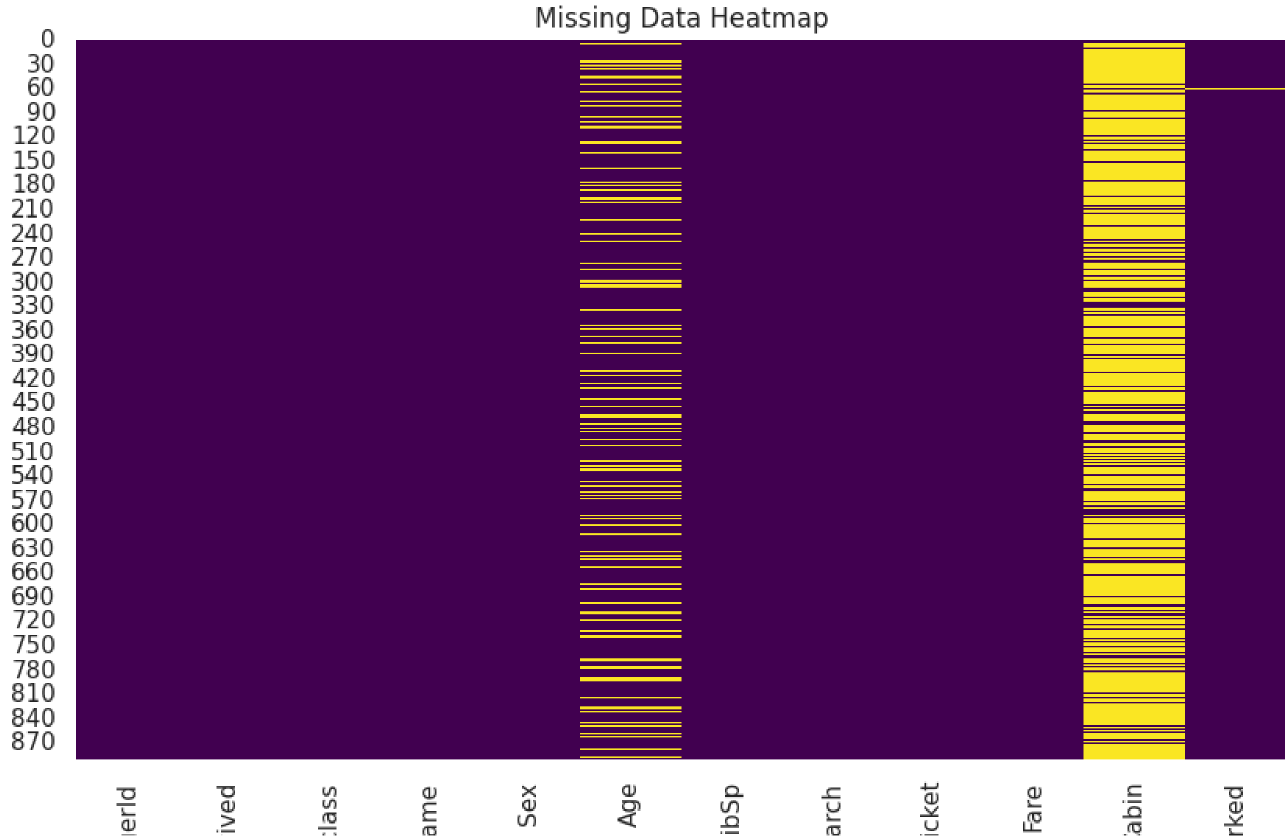
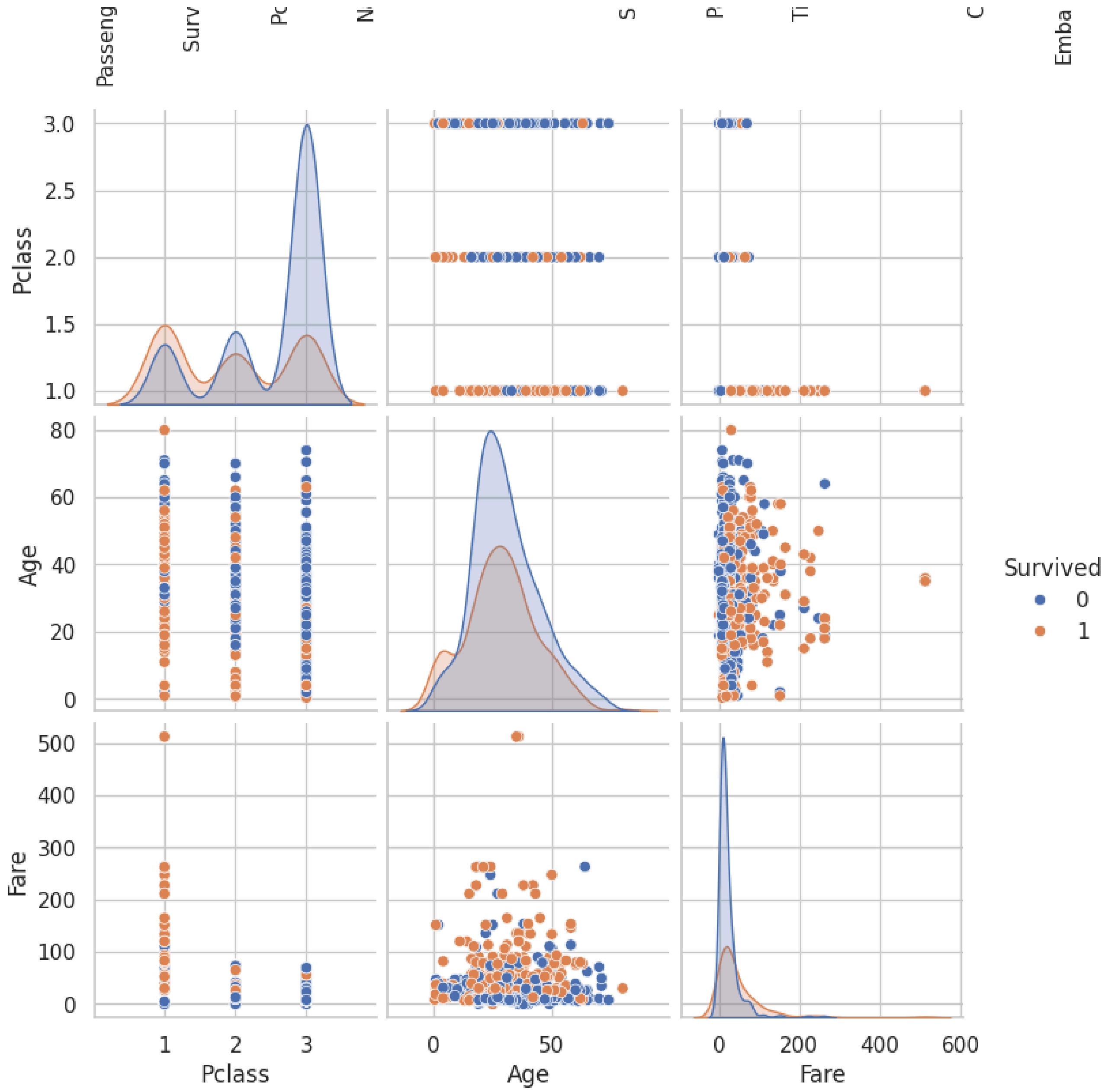
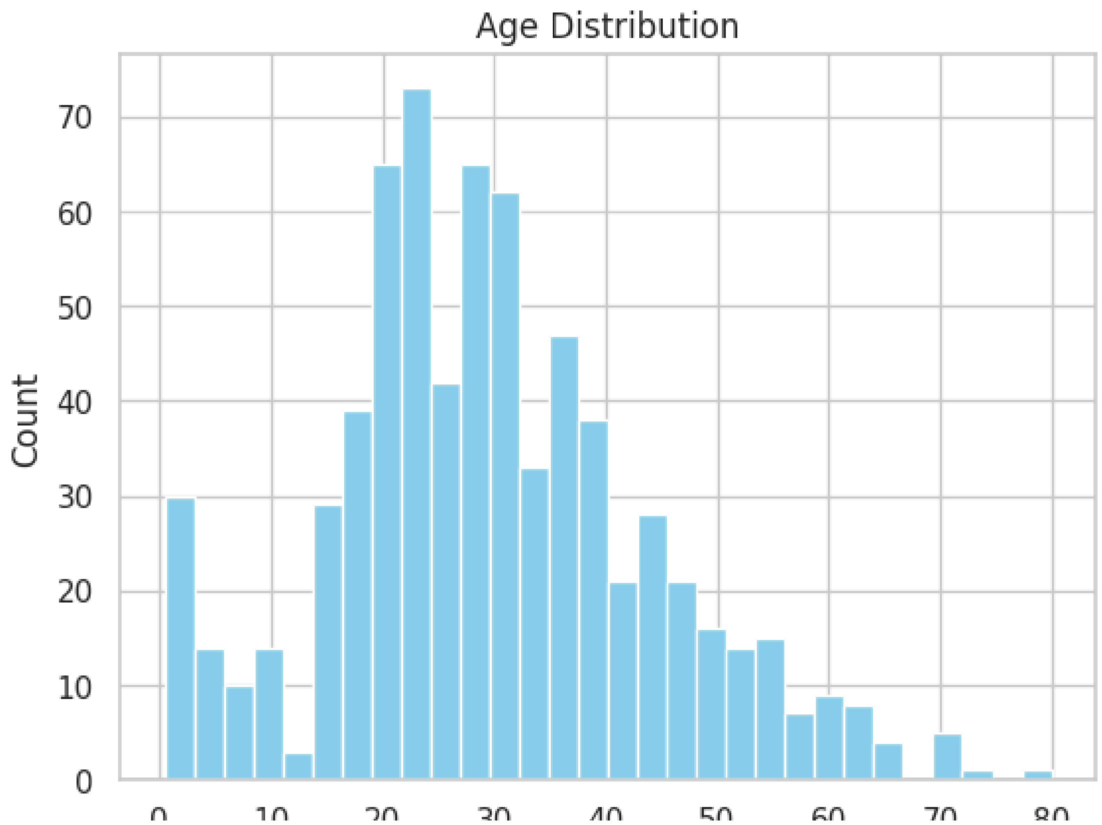
Choose Files train.csv

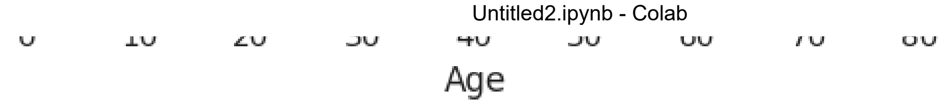
- **train.csv**(text/csv) - 61194 bytes, last modified: 6/30/2025 - 100% done

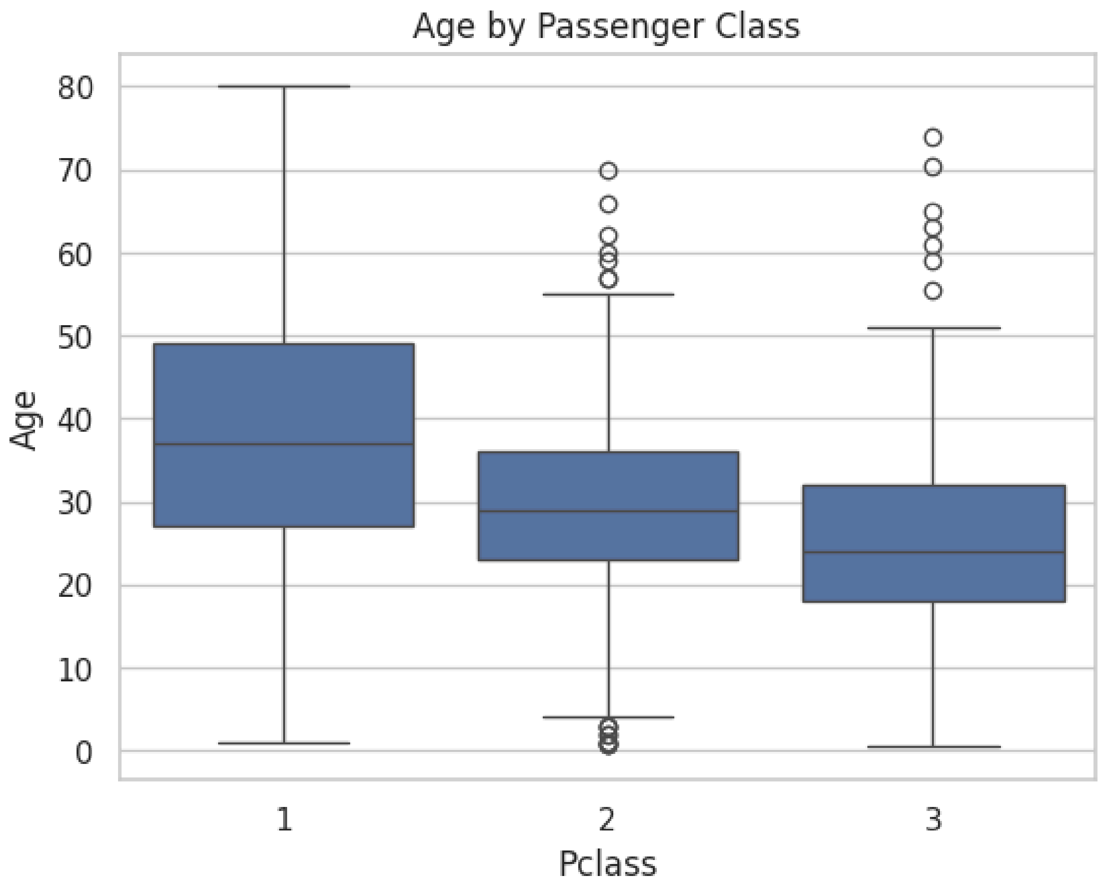
```
Saving train.csv to train (1).csv
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
Sex
male      577
female    314
Name: count, dtype: int64
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64
```

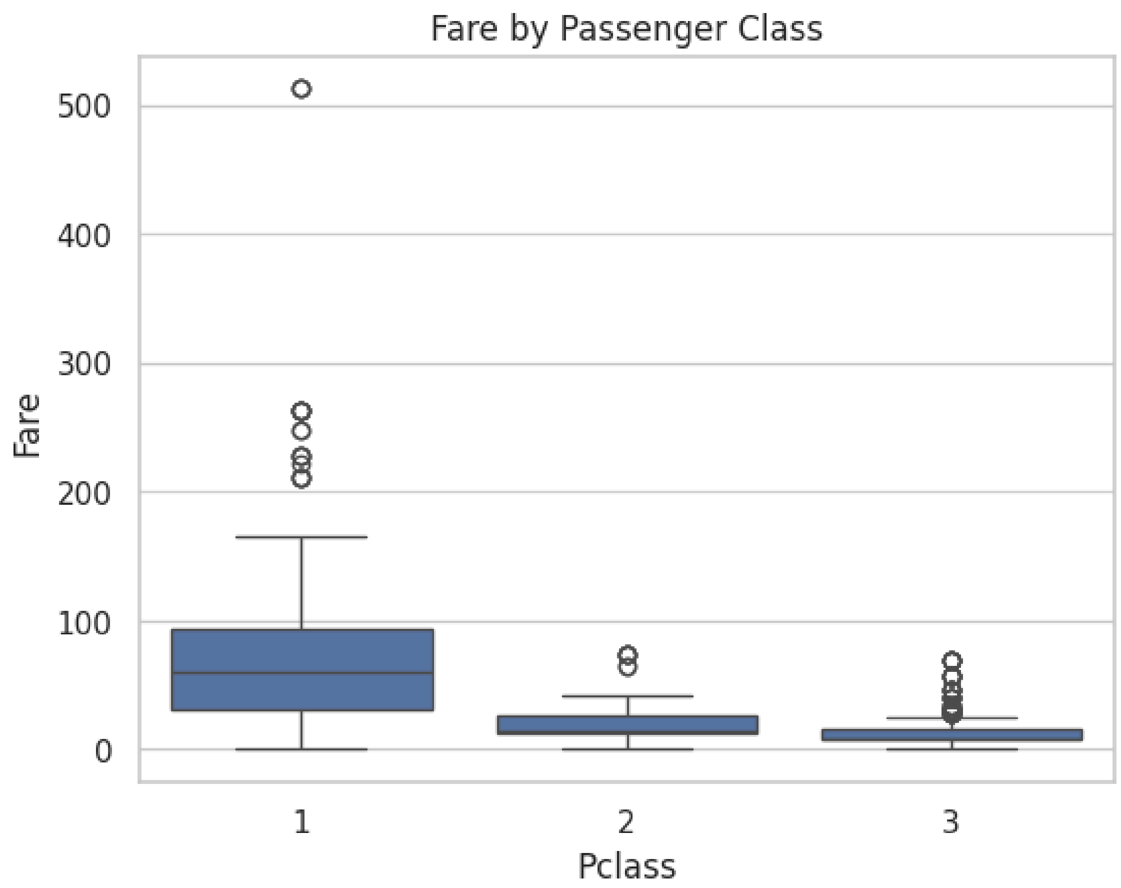


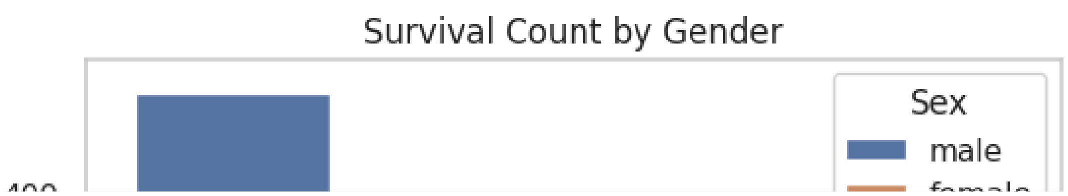

Passeng
Surv
Pc
N
S
P
Ti
C
Emba
Pclass
Age
Fare
Survived
0
1

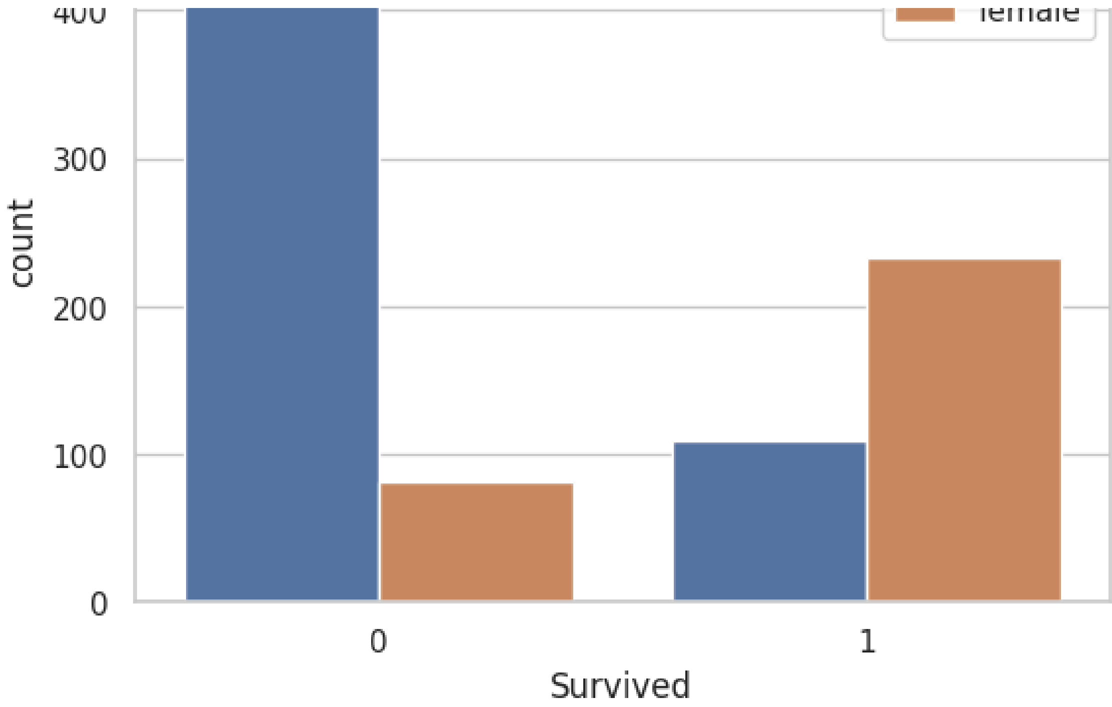

Age Distribution
Count


Age


Age by Passenger Class
80
70
60
50
40
30
20
10
0
Age
1
2
3
Pclass


Fare by Passenger Class
500
400
300
200
100
0
Fare
1
2
3
Pclass


Survival Count by Gender
Sex
male



Summary of Findings

- **Data Quality Issues**: Significant missing data in `Age`, `Cabin`, and some in `Embarked` need cleaning.
- **Survival Correlates**: Gender and Passenger Class have strong relationships with survival outcomes. Females and 1st class passengers survived more.
- **Age Trends**: Younger passengers were mostly in 3rd class. Age has right-skewed distribution, impacting modeling.
- **Fare Patterns**: Fare shows strong skew and varies significantly by class, indicating socioeconomic differences.
- **Next Steps**: Impute or drop missing values, encode categorical variables, and use these insights for building predictive models.