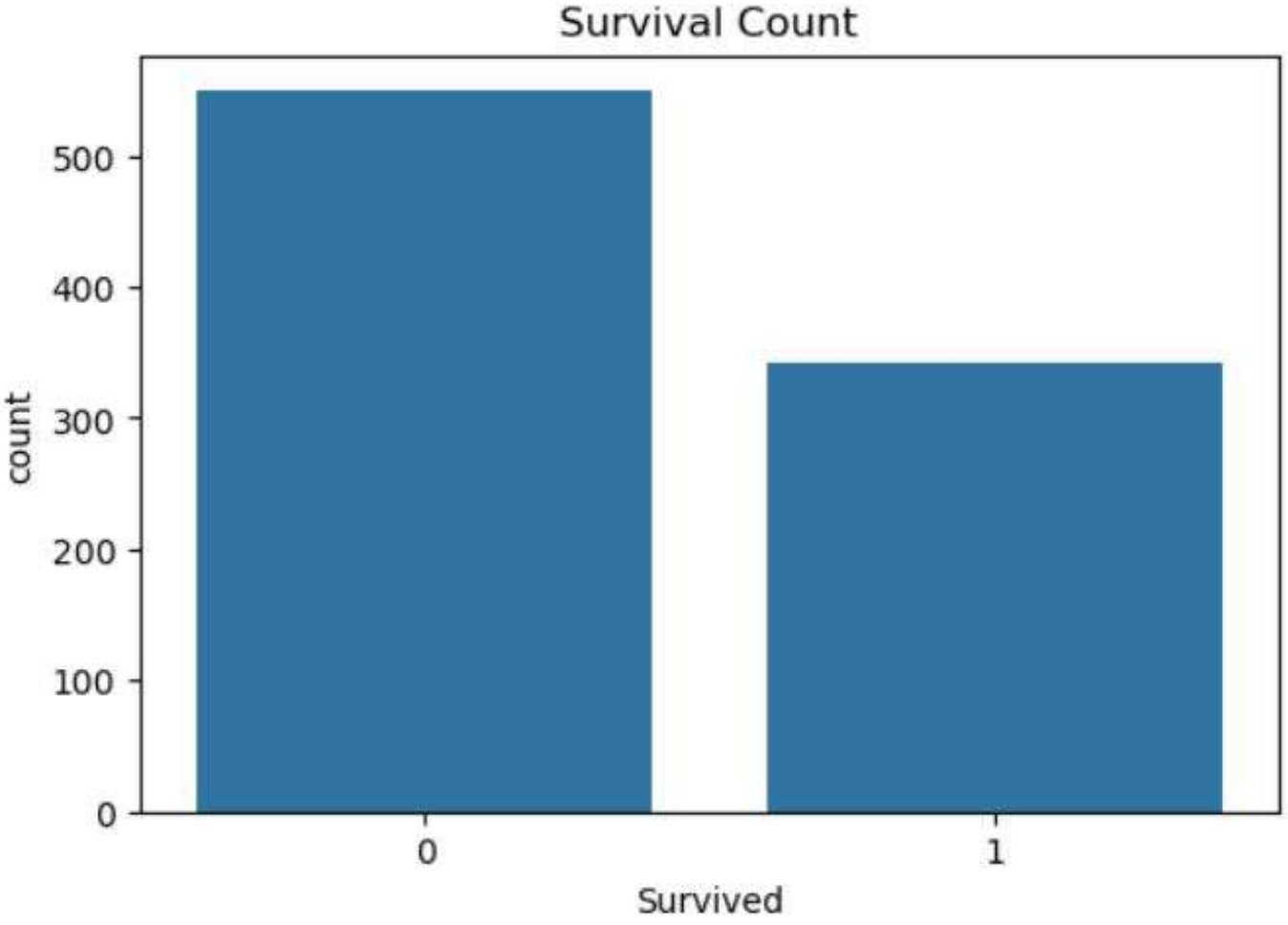
```
Shape of dataset: (891, 12)

Dataset Info:

<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
None
```

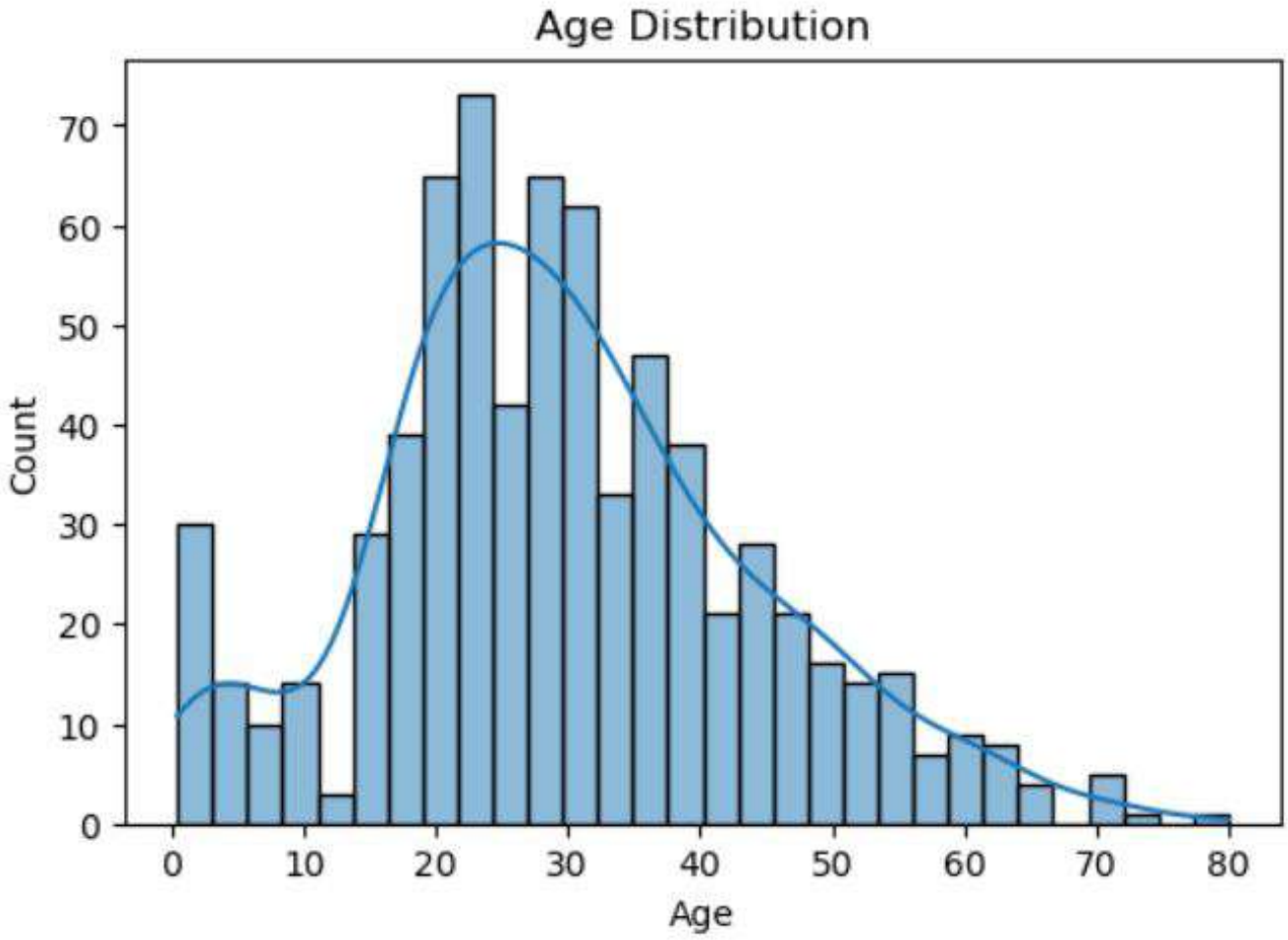
Summary Statistics:

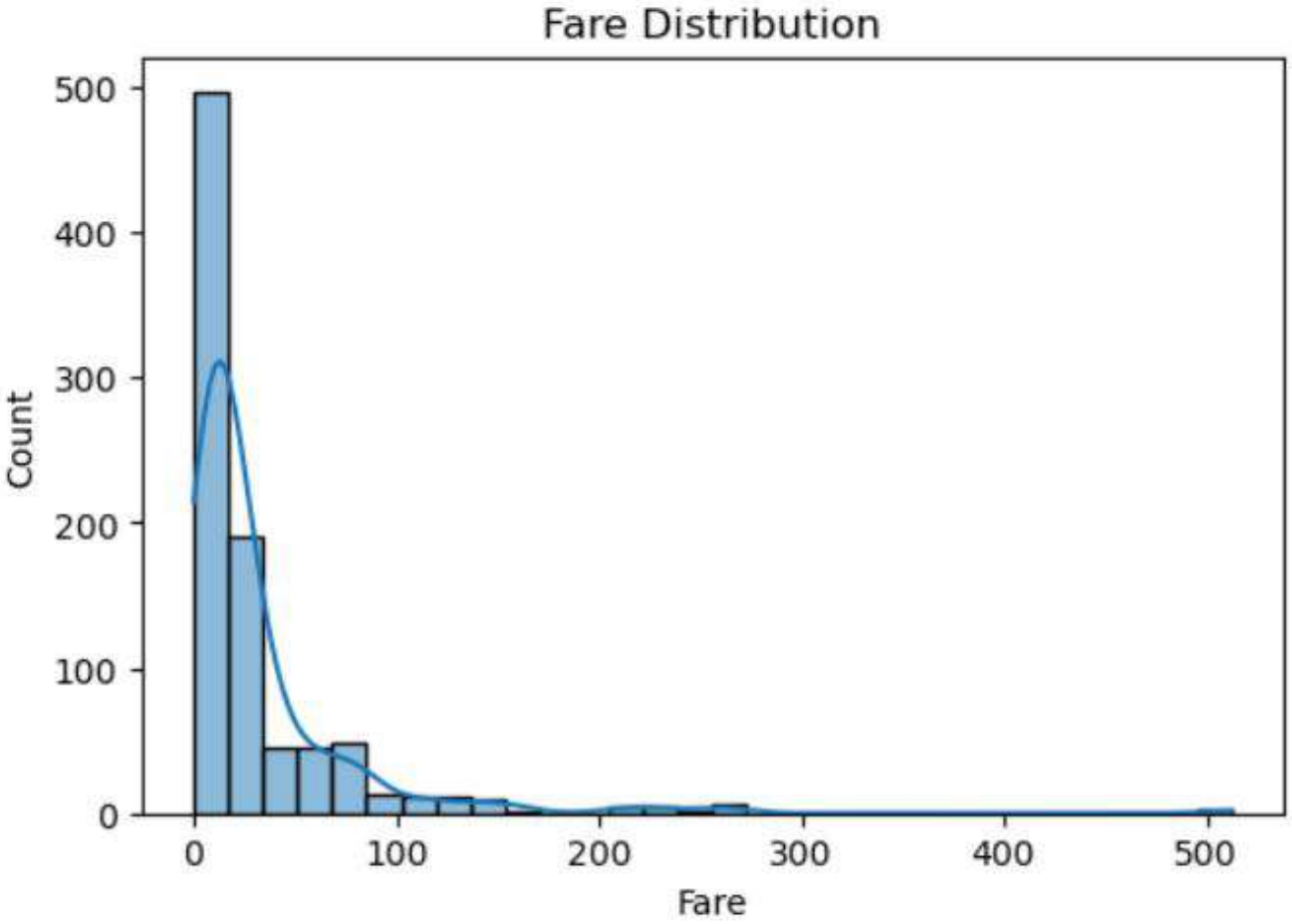
| | PassengerId | Survived | Pclass | Name | Sex | \ |
|---|---|---|---|---|---|---|
| count | 891.000000 | 891.000000 | 891.000000 | 891 | 891 | |
| unique | NaN | NaN | NaN | 891 | 2 | |
| top | NaN | NaN | NaN | Dooley, Mr. Patrick | male | |
| freq | NaN | NaN | NaN | 1 | 577 | |
| mean | 446.000000 | 0.383838 | 2.308642 | NaN | NaN | |
| std | 257.353842 | 0.486592 | 0.836071 | NaN | NaN | |
| min | 1.000000 | 0.000000 | 1.000000 | NaN | NaN | |
| 25% | 223.500000 | 0.000000 | 2.000000 | NaN | NaN | |
| 50% | 446.000000 | 0.000000 | 3.000000 | NaN | NaN | |
| 75% | 668.500000 | 1.000000 | 3.000000 | NaN | NaN | |
| max | 891.000000 | 1.000000 | 3.000000 | NaN | NaN | |

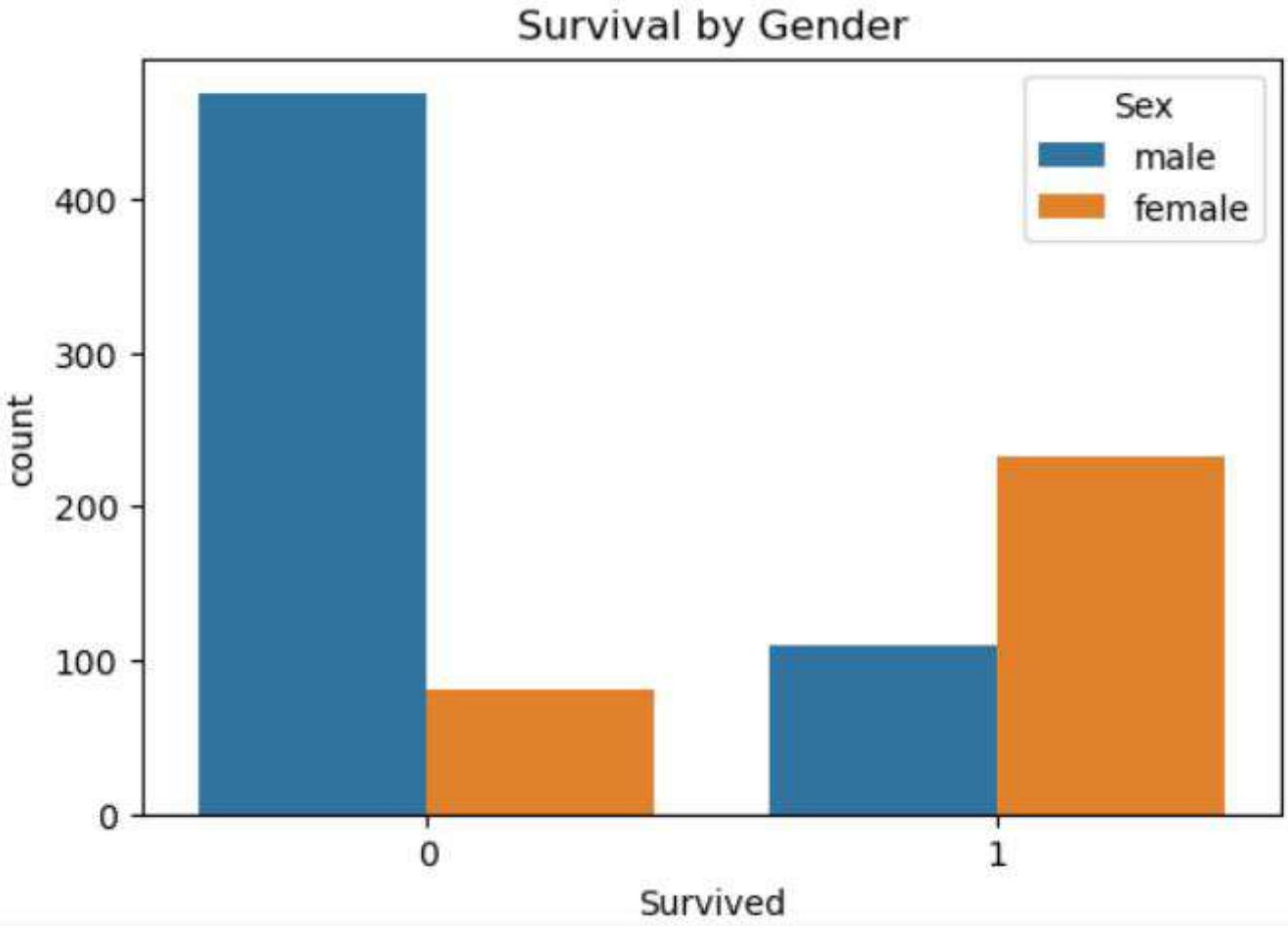
| | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|
| count | 714.000000 | 891.000000 | 891.000000 | 891 | 891.000000 | 204 | 889 |
| unique | NaN | NaN | NaN | 681 | NaN | 147 | 3 |
| top | NaN | NaN | NaN | 347082 | NaN | G6 | S |
| freq | NaN | NaN | NaN | 7 | NaN | 4 | 644 |
| mean | 29.699118 | 0.523008 | 0.381594 | NaN | 32.204208 | NaN | NaN |
| std | 14.526497 | 1.102743 | 0.806057 | NaN | 49.693429 | NaN | NaN |
| min | 0.420000 | 0.000000 | 0.000000 | NaN | 0.000000 | NaN | NaN |
| 25% | 20.125000 | 0.000000 | 0.000000 | NaN | 7.910400 | NaN | NaN |
| 50% | 28.000000 | 0.000000 | 0.000000 | NaN | 14.454200 | NaN | NaN |
| 75% | 38.000000 | 1.000000 | 0.000000 | NaN | 31.000000 | NaN | NaN |
| max | 80.000000 | 8.000000 | 6.000000 | NaN | 512.329200 | NaN | NaN |

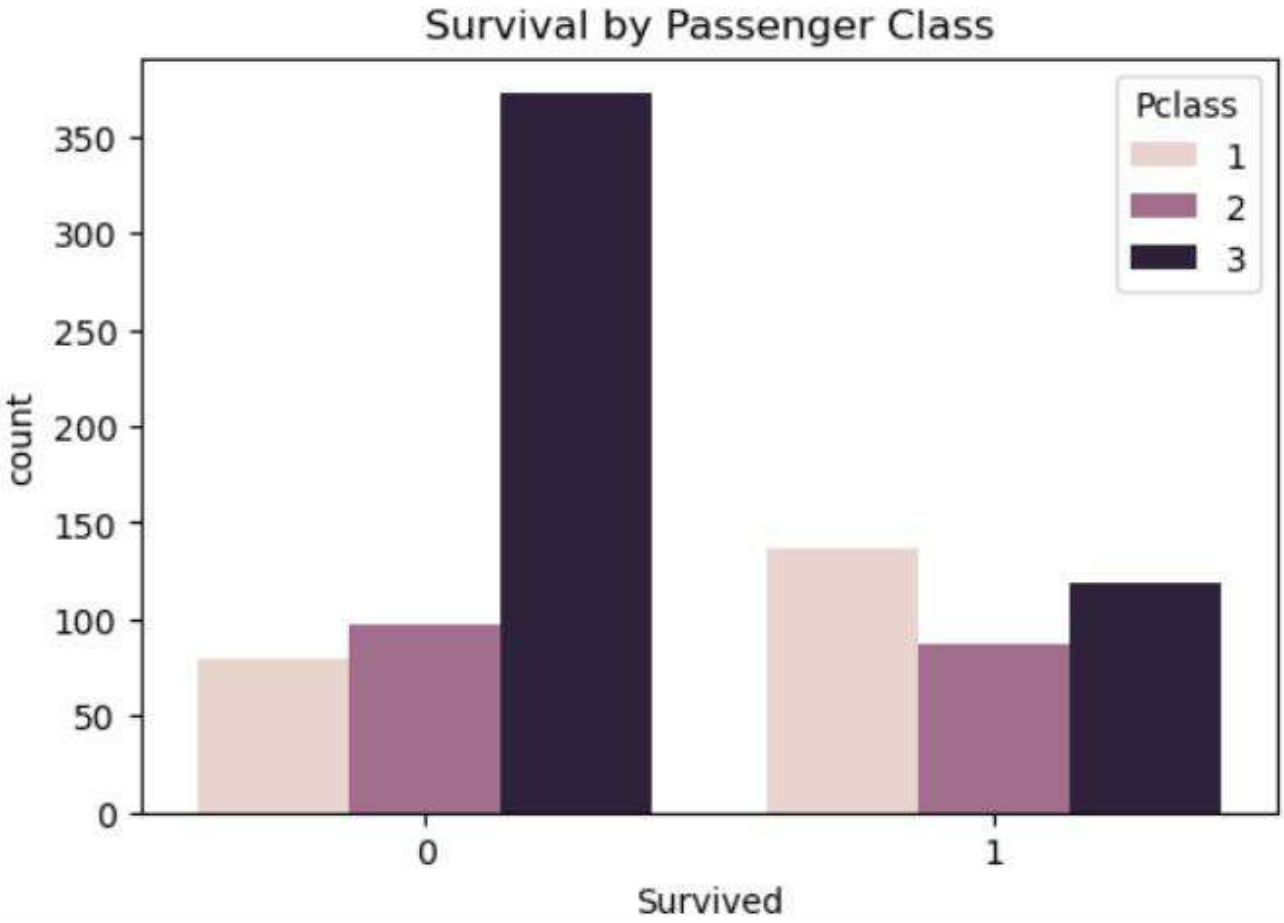
```
Missing Values:

PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64

<Figure size 1000x600 with 0 Axes>
```

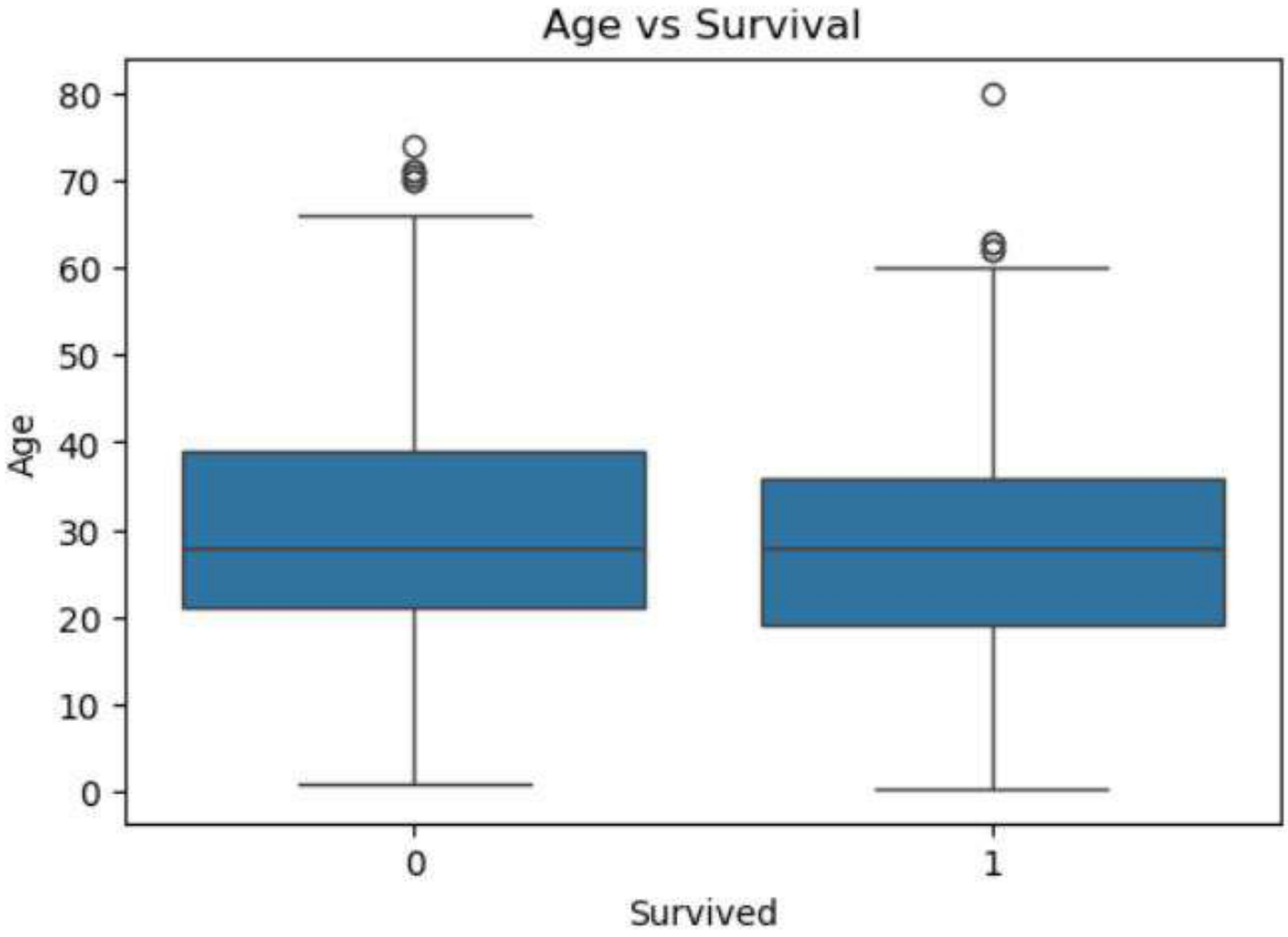

Survival Count
500
400
300
200
100
0
count
0
1
Survived

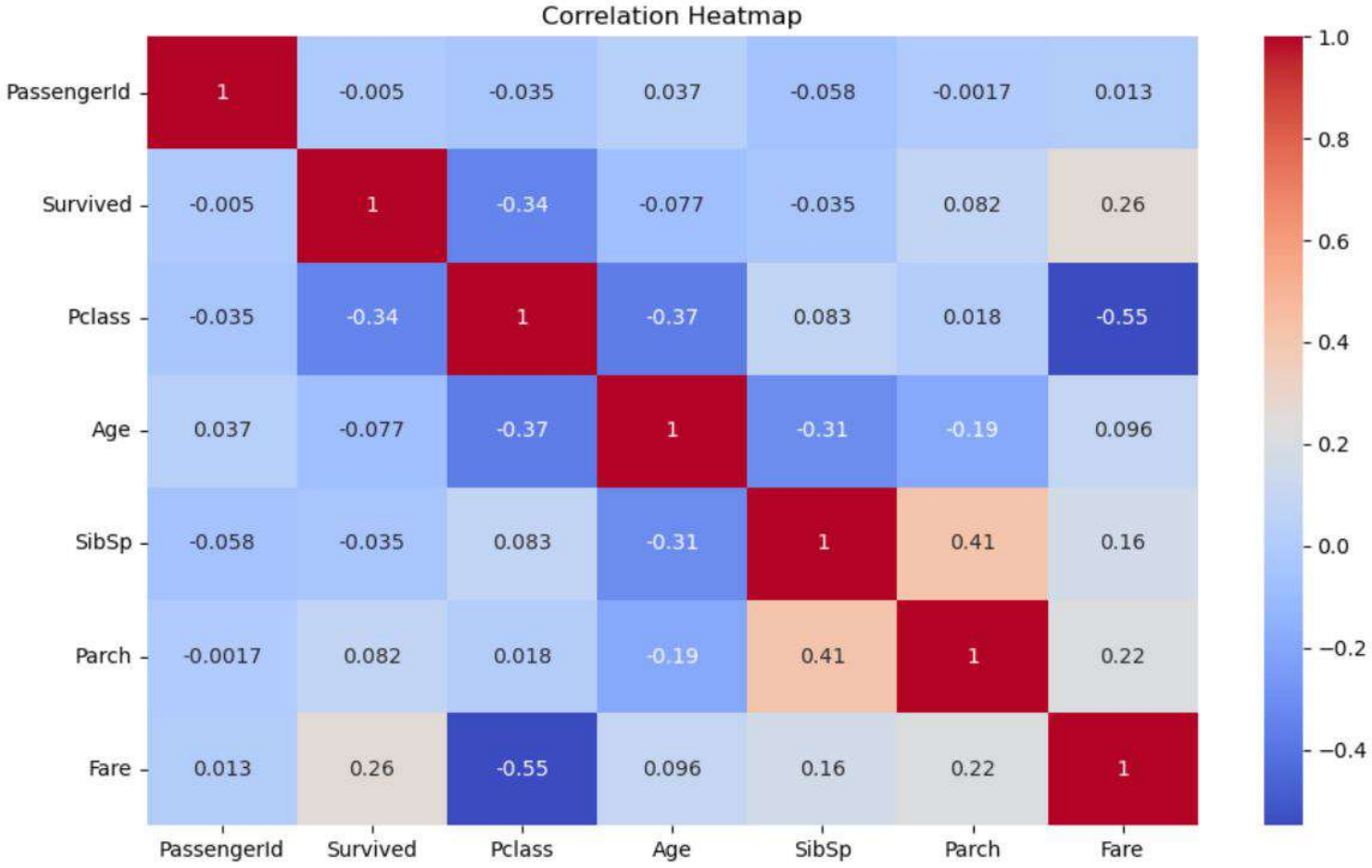

Age Distribution
Count
Age
0
10
20
30
40
50
60
70
80

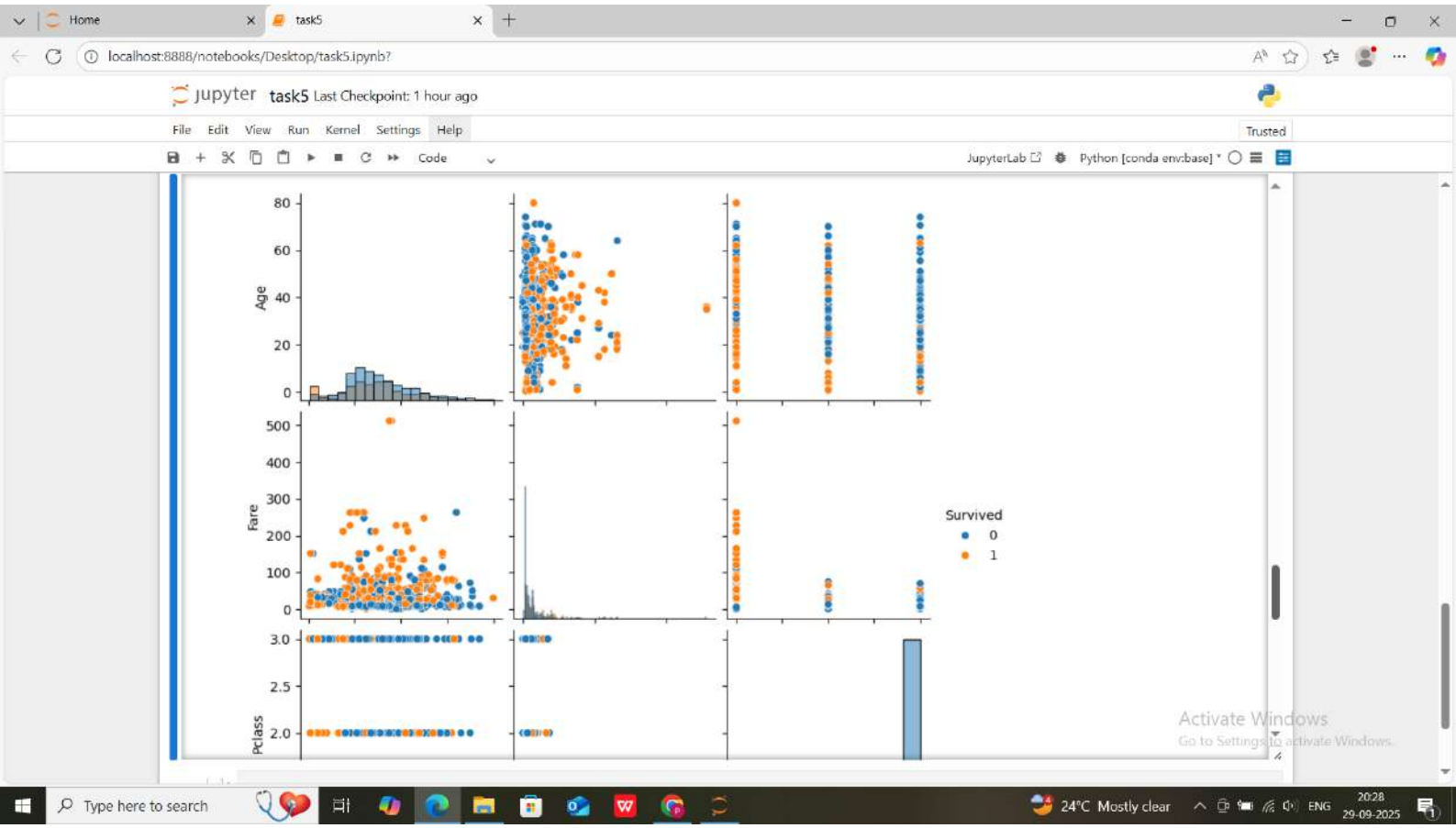

Fare Distribution
Count
Fare
0
100
200
300
400
500

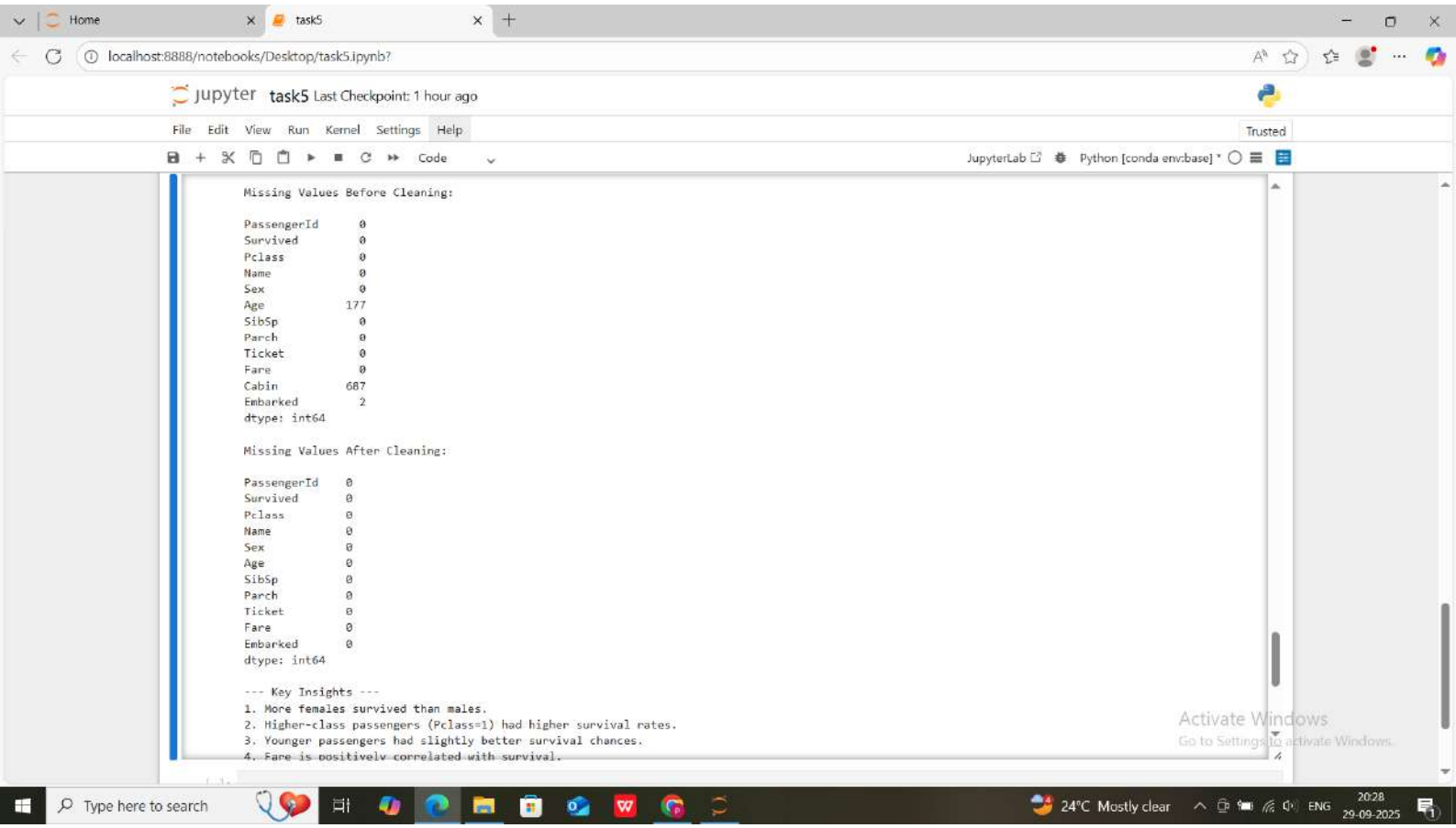

Survival by Gender
Sex
male
female
400
300
200
100
0
count
0
1
Survived

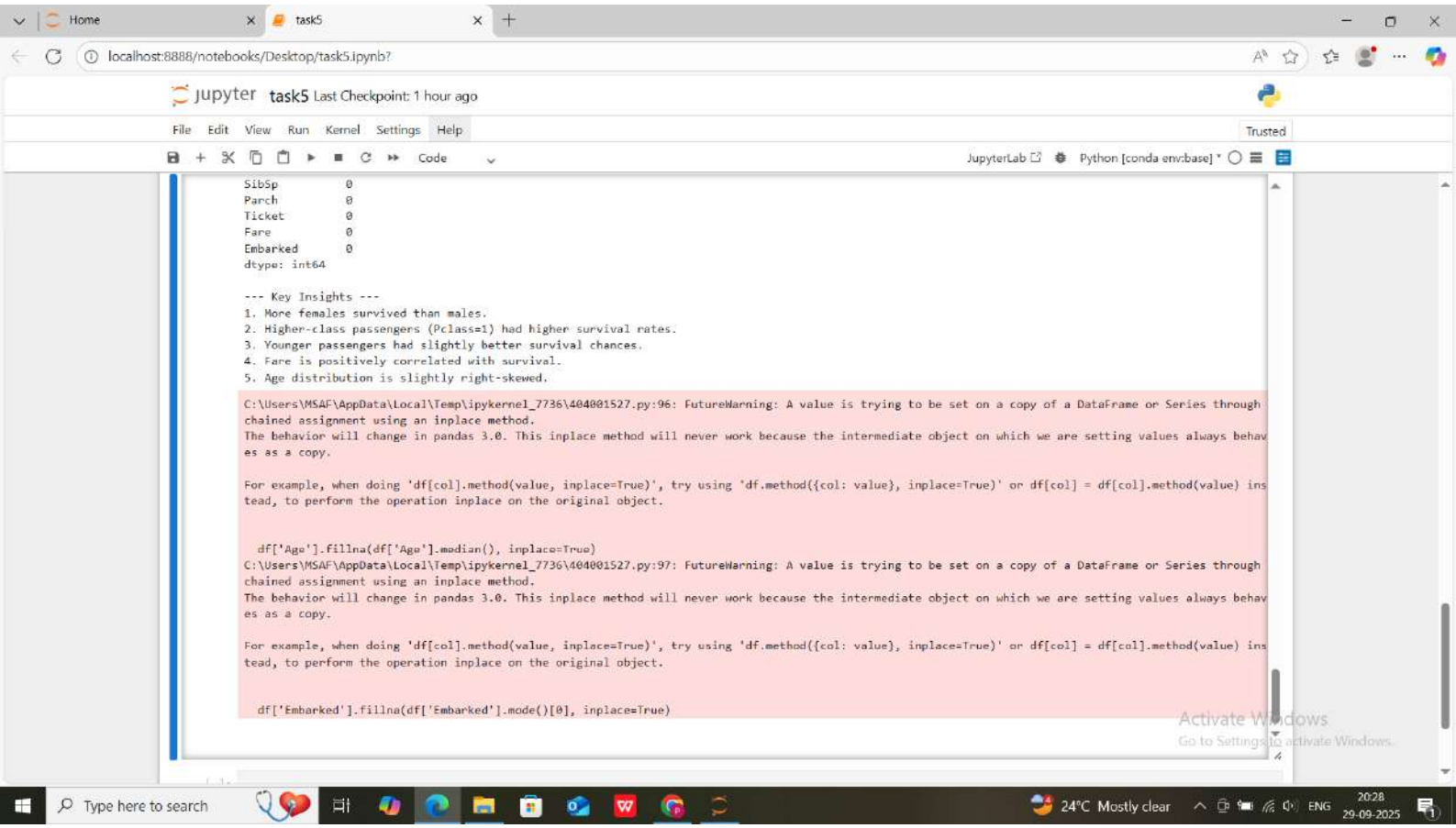

Survival by Passenger Class
Pclass
1
2
3
count
350
300
250
200
150
100
50
0
0
1
Survived


Age vs Survival
Age
80
70
60
50
40
30
20
10
0
0
1
Survived


Correlation Heatmap
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1 -0.005 -0.035 0.037 -0.058 -0.0017 0.013
-0.005 1 -0.34 -0.077 -0.035 0.082 0.26
-0.035 -0.34 1 -0.37 0.083 0.018 -0.55
0.037 -0.077 -0.37 1 -0.31 -0.19 0.096
-0.058 -0.035 0.083 -0.31 1 0.41 0.16
-0.0017 0.082 0.018 -0.19 0.41 1 0.22
0.013 0.26 -0.55 0.096 0.16 0.22 1
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Home
task5
localhost:8888/notebooks/Desktop/task5.ipynb?
jupyter task5 Last Checkpoint: 1 hour ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base]
Age
Fare
Pclass
Survived
0
1
Activate Windows
Go to Settings to activate Windows.
Type here to search
24°C Mostly clear
ENG
20:28
29-09-2025


Home
task5
localhost:8888/notebooks/Desktop/task5.ipynb?
jupyter task5 Last Checkpoint: 1 hour ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base]
Missing Values Before Cleaning:
PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 177
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 2
dtype: int64
Missing Values After Cleaning:
PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0
Embarked 0
dtype: int64
--- Key Insights ---
1. More females survived than males.
2. Higher-class passengers (Pclass=1) had higher survival rates.
3. Younger passengers had slightly better survival chances.
4. Fare is positively correlated with survival.
Activate Windows
Go to Settings to activate Windows.
Type here to search
24°C Mostly clear
ENG
20:28
29-09-2025


Home
task5
localhost:8888/notebooks/Desktop/task5.ipynb?
jupyter task5 Last Checkpoint: 1 hour ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python [conda env:base]
SibSp 0
Parch 0
Ticket 0
Fare 0
Embarked 0
dtype: int64
--- Key Insights ---
1. More females survived than males.
2. Higher-class passengers (Pclass=1) had higher survival rates.
3. Younger passengers had slightly better survival chances.
4. Fare is positively correlated with survival.
5. Age distribution is slightly right-skewed.
C:\Users\MSAF\AppData\Local\Temp\ipykernel_7736\404001527.py:96: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.
df['Age'].fillna(df['Age'].median(), inplace=True)
C:\Users\MSAF\AppData\Local\Temp\ipykernel_7736\404001527.py:97: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.
For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
Activate Windows
Go to Settings to activate Windows.
Type here to search
24°C Mostly clear
ENG
20:28
29-09-2025