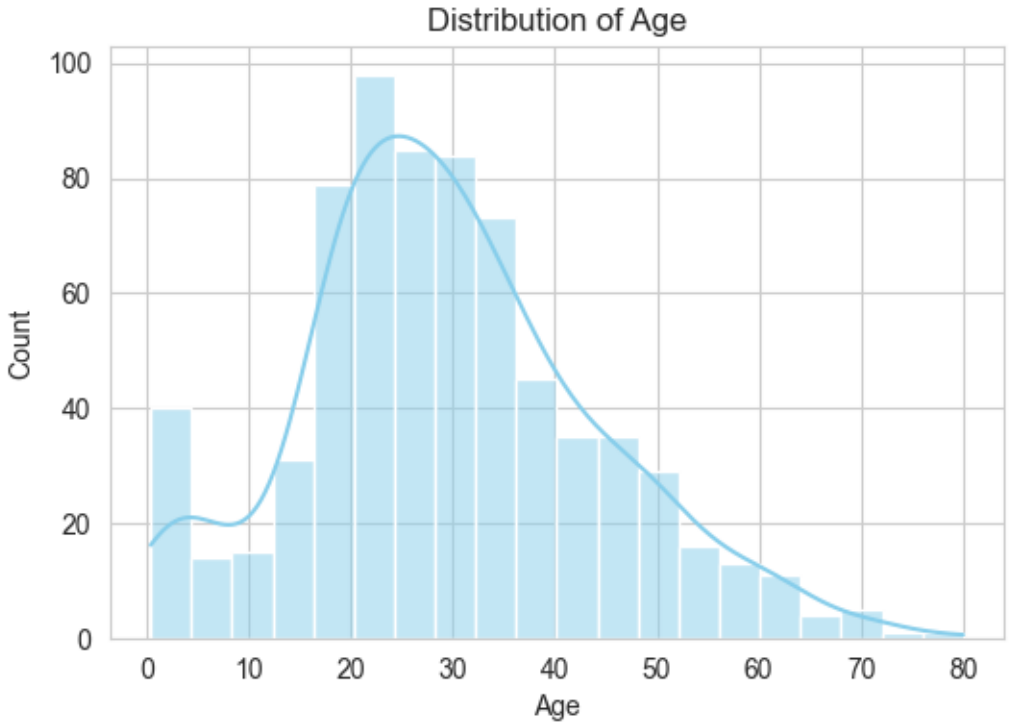
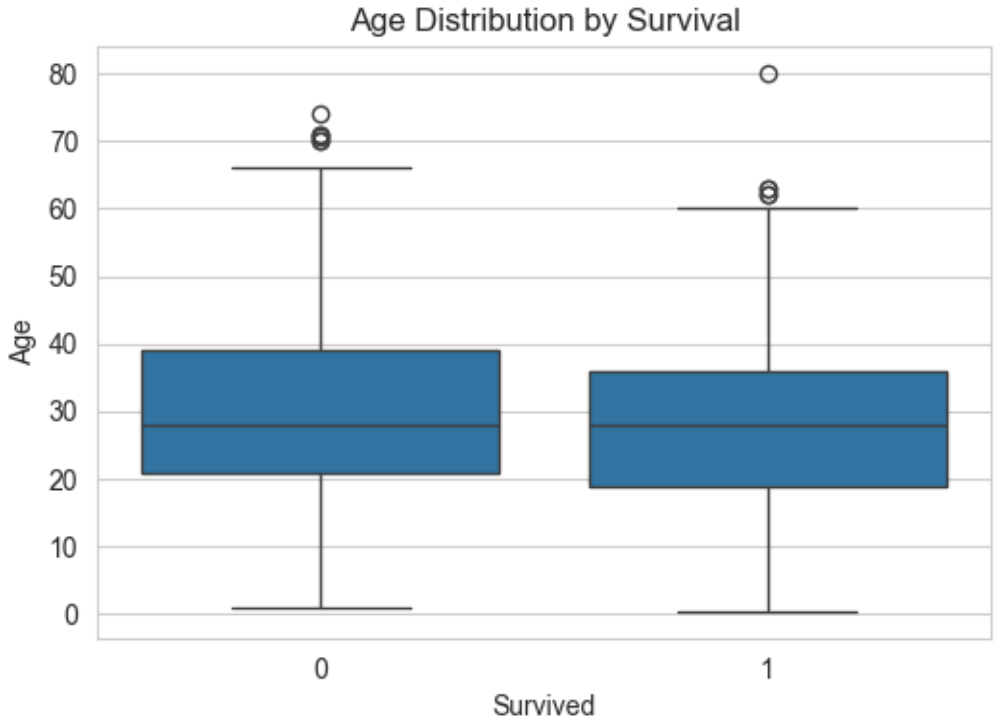
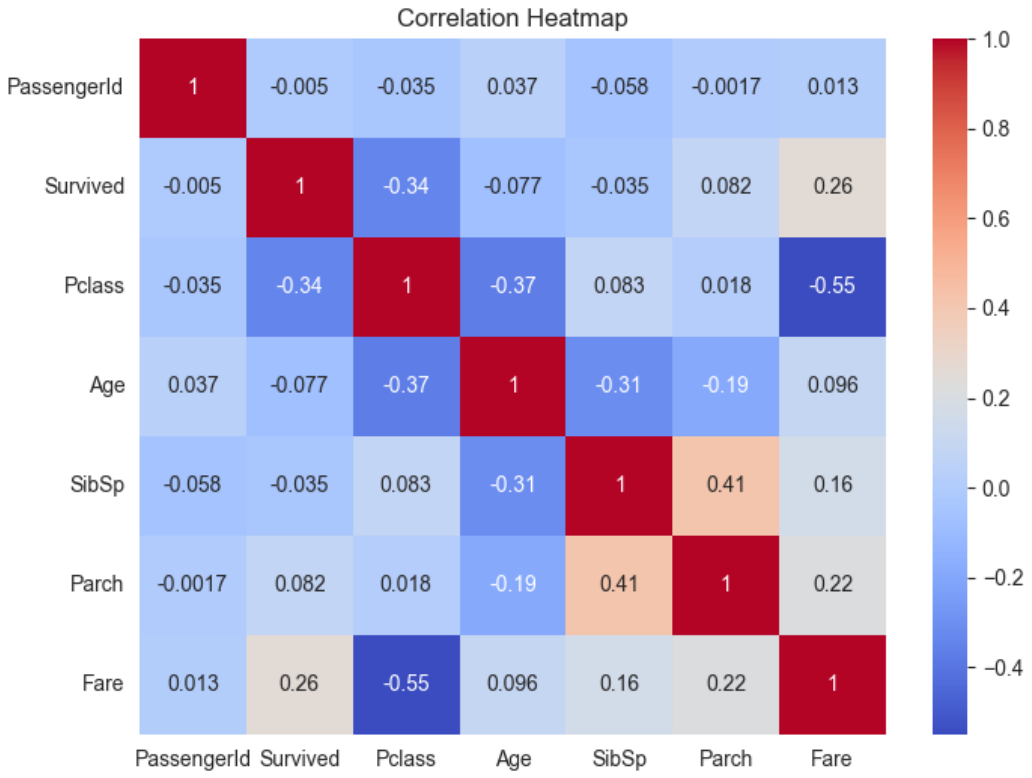
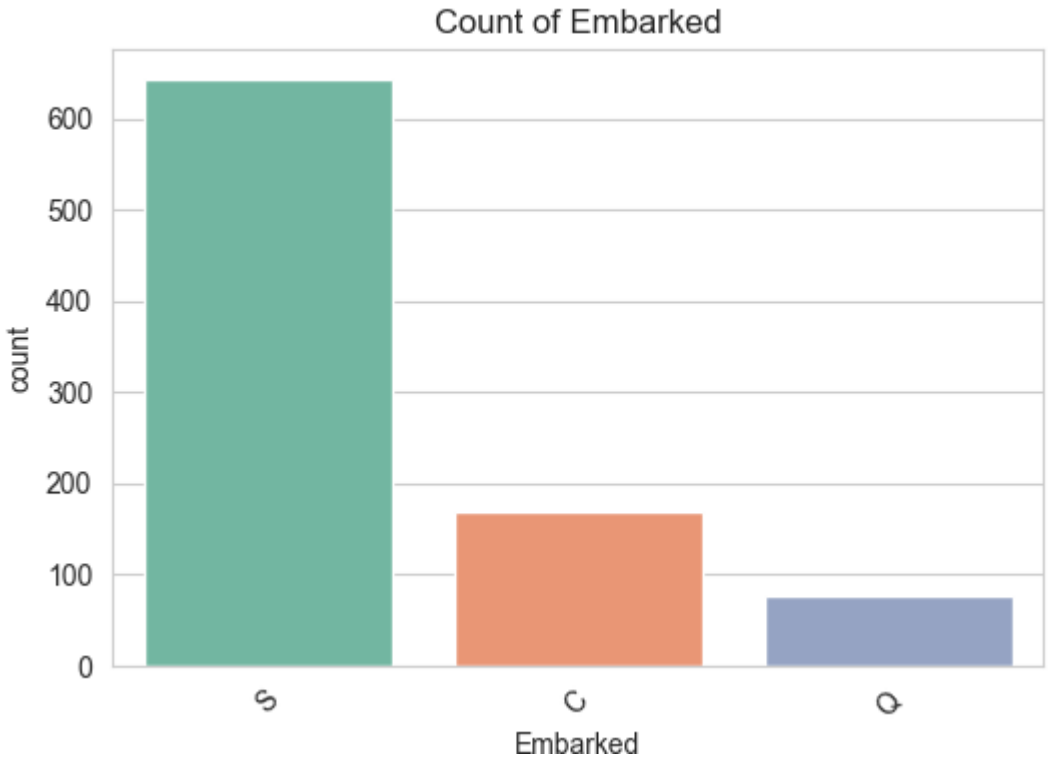
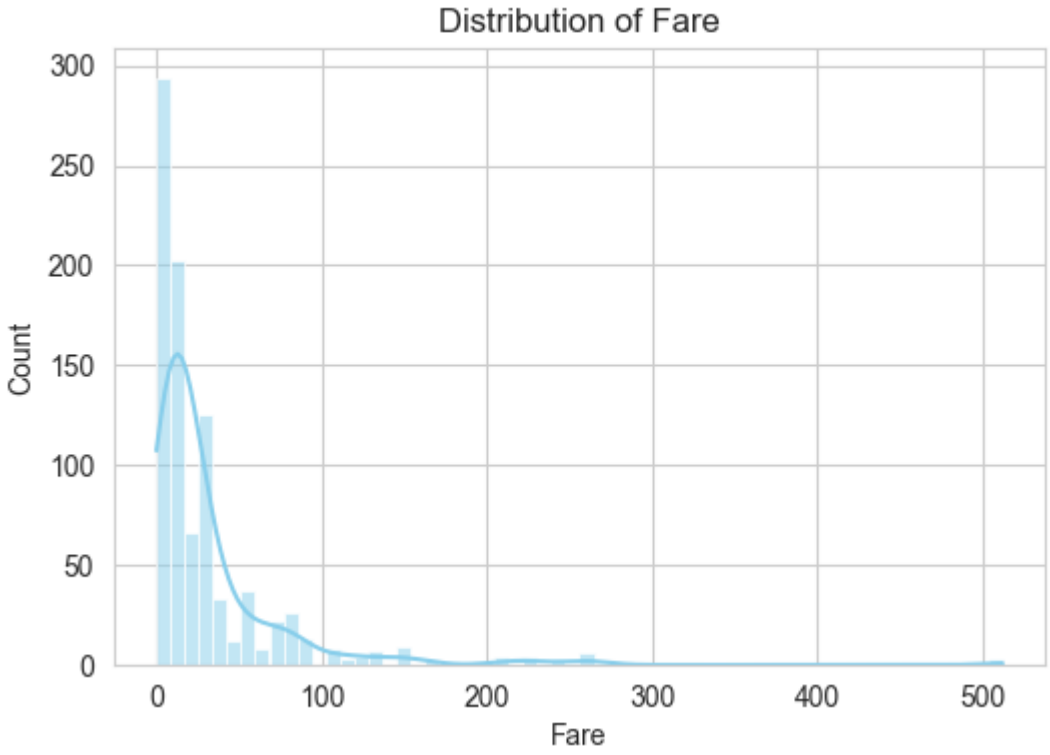
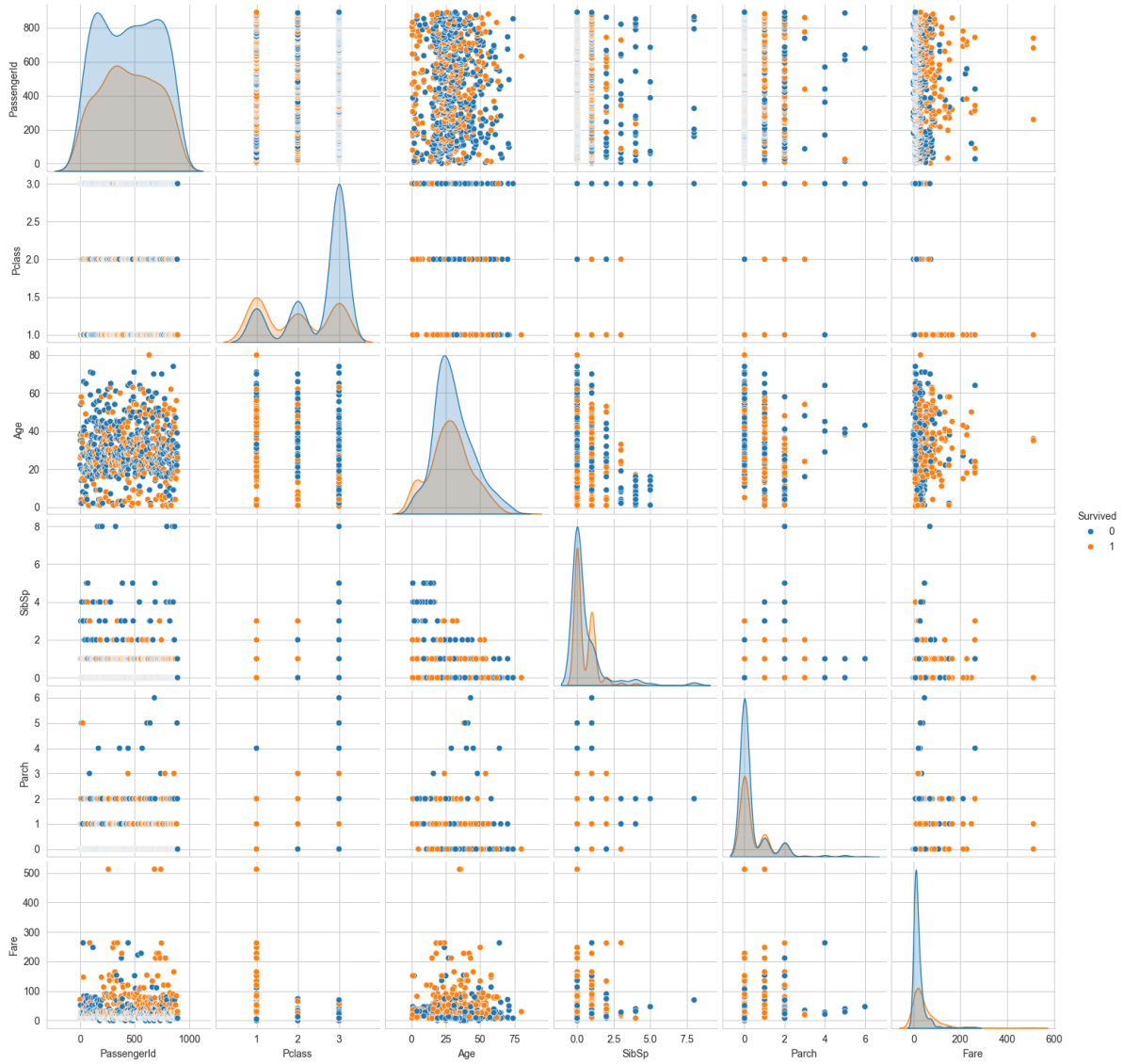

Distribution of Age
Count
Age
0
20
40
60
80
100
0
10
20
30
40
50
60
70
80

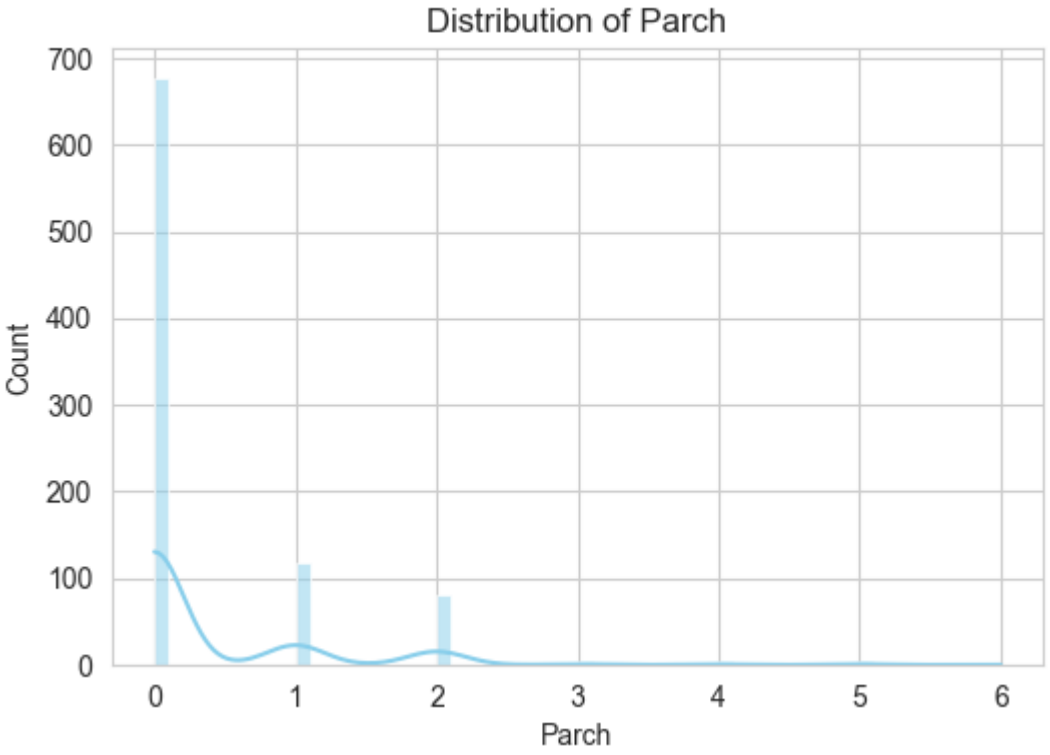

Age Distribution by Survival
Age
Survived
0
10
20
30
40
50
60
70
80
0
1

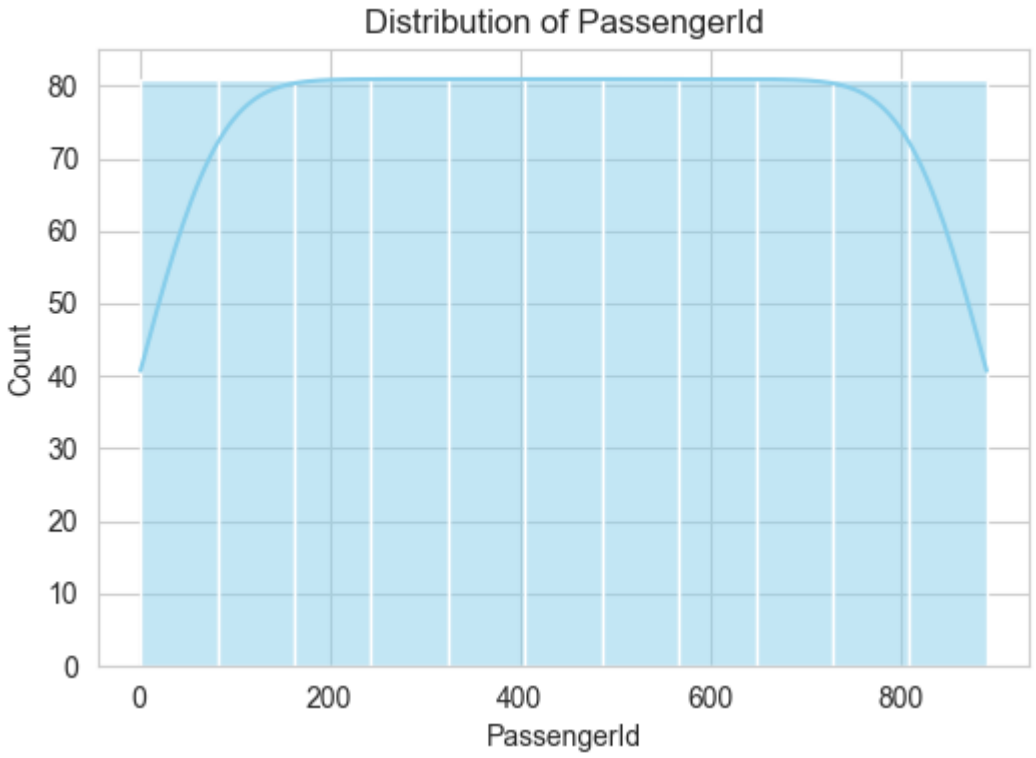

Correlation Heatmap
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1 -0.005 -0.035 0.037 -0.058 -0.0017 0.013
-0.005 1 -0.34 -0.077 -0.035 0.082 0.26
-0.035 -0.34 1 -0.37 0.083 0.018 -0.55
0.037 -0.077 -0.37 1 -0.31 -0.19 0.096
-0.058 -0.035 0.083 -0.31 1 0.41 0.16
-0.0017 0.082 0.018 -0.19 0.41 1 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

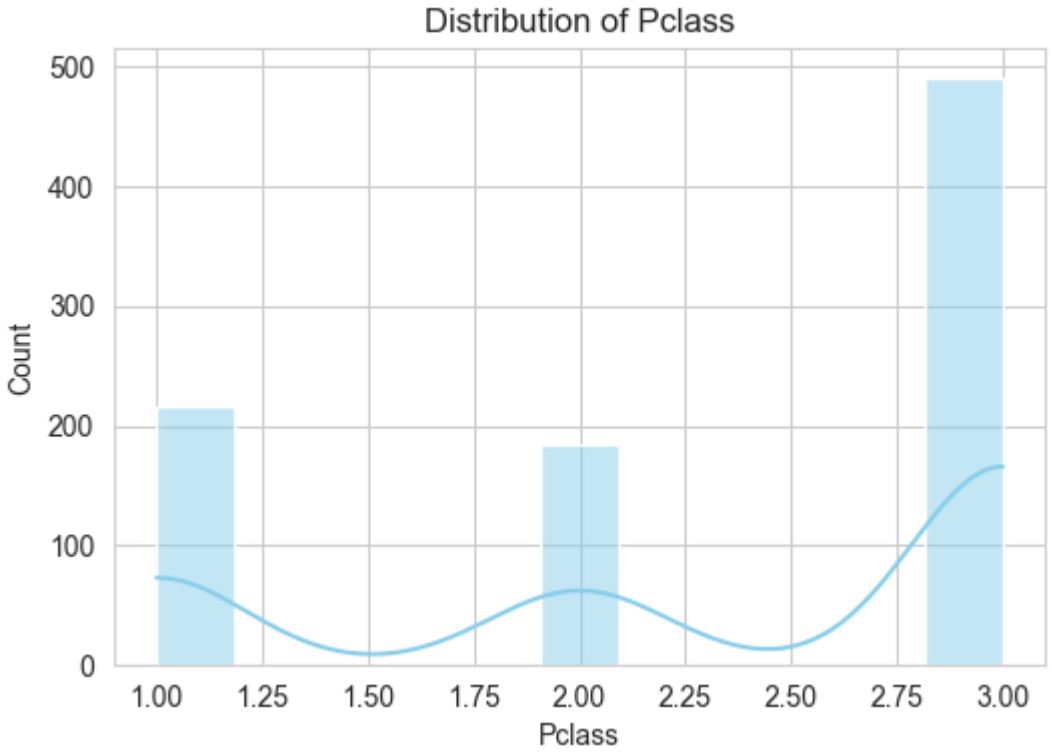

Count of Embarked
count
600
500
400
300
200
100
0
S
C
Q
Embarked

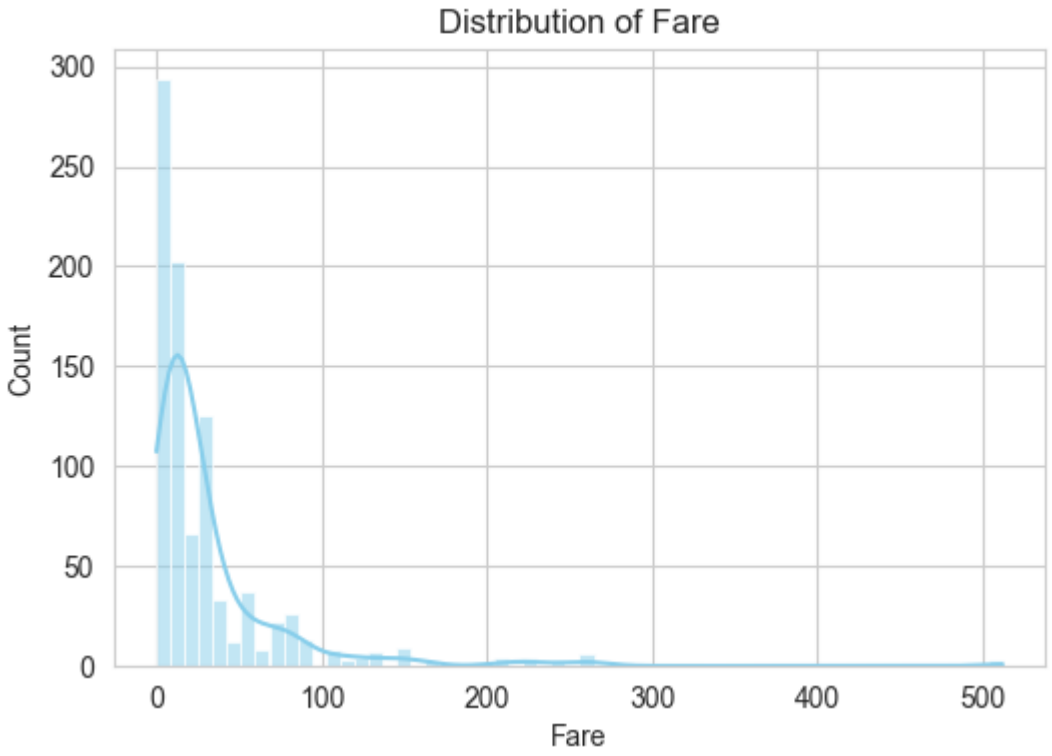

Distribution of Fare
Count
300
250
200
150
100
50
0
0
100
200
300
400
500
Fare

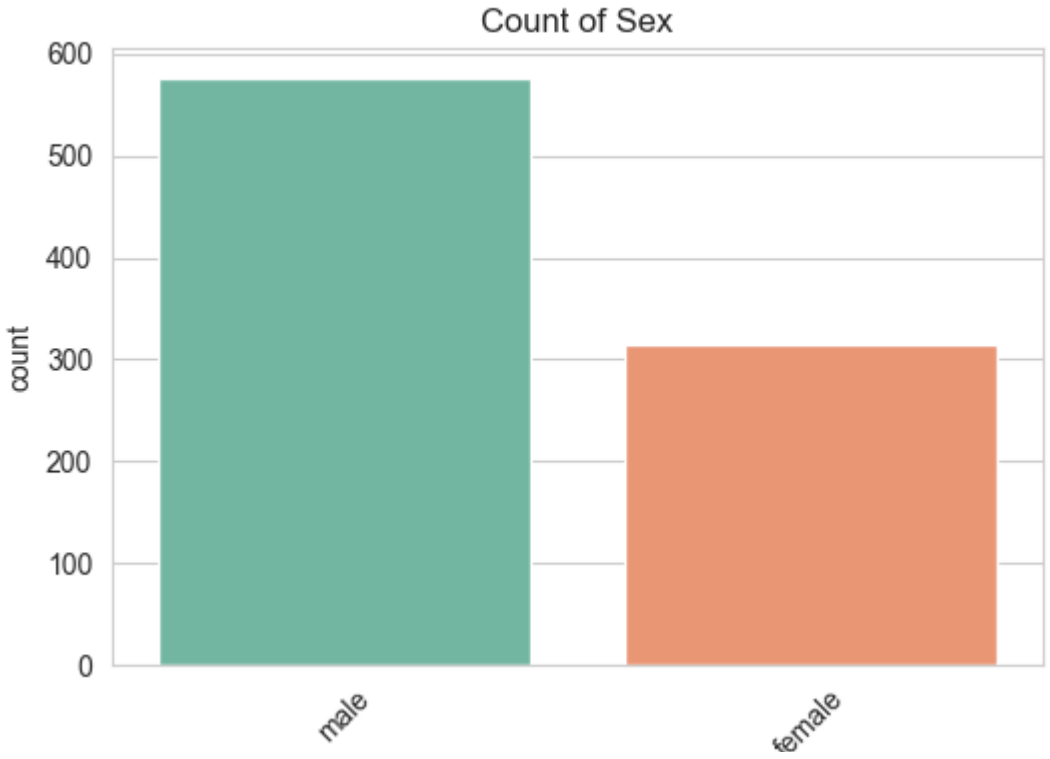

PassengerId
Pclass
Age
SibSp
Parch
Fare
Survived
0
1

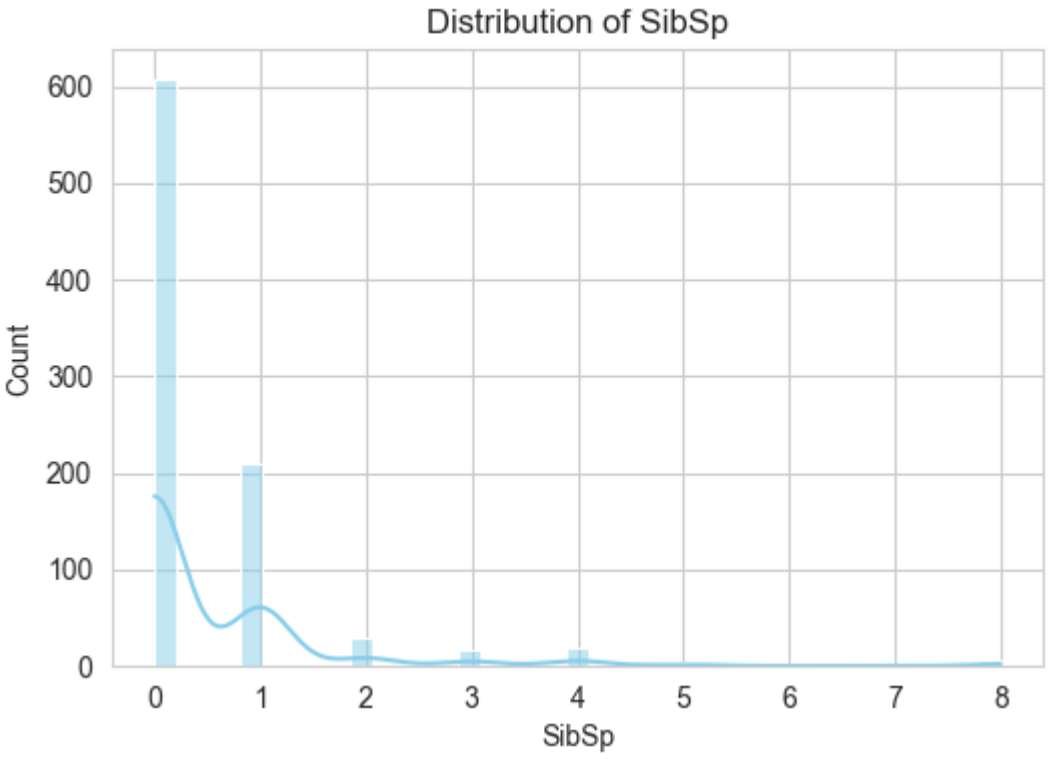

Distribution of Parch
Count
Parch
0
100
200
300
400
500
600
700
0
1
2
3
4
5
6

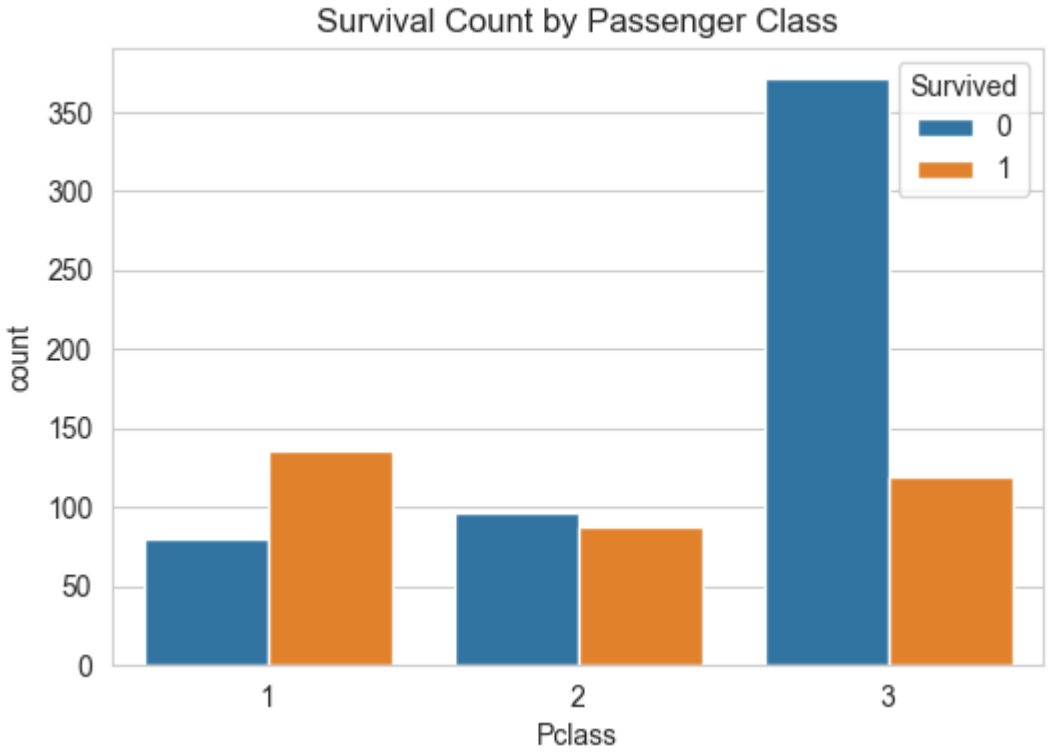

Distribution of PassengerId
Count
PassengerId
0
10
20
30
40
50
60
70
80
0
200
400
600
800

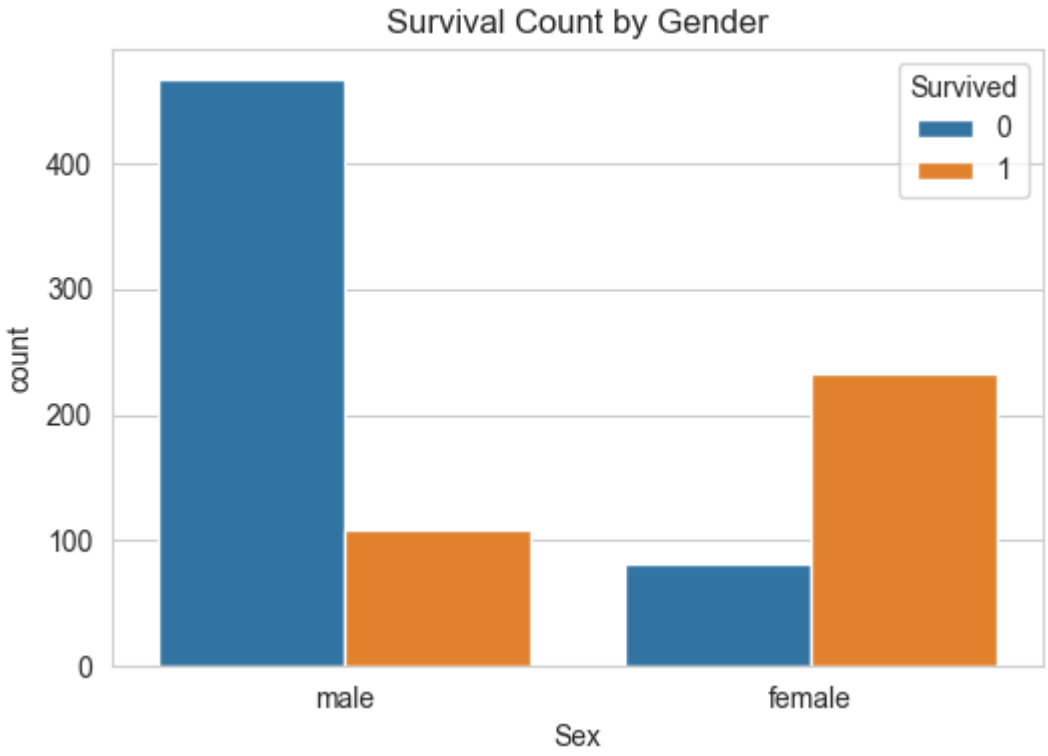

Distribution of Pclass
Count
Pclass
0
100
200
300
400
500
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00

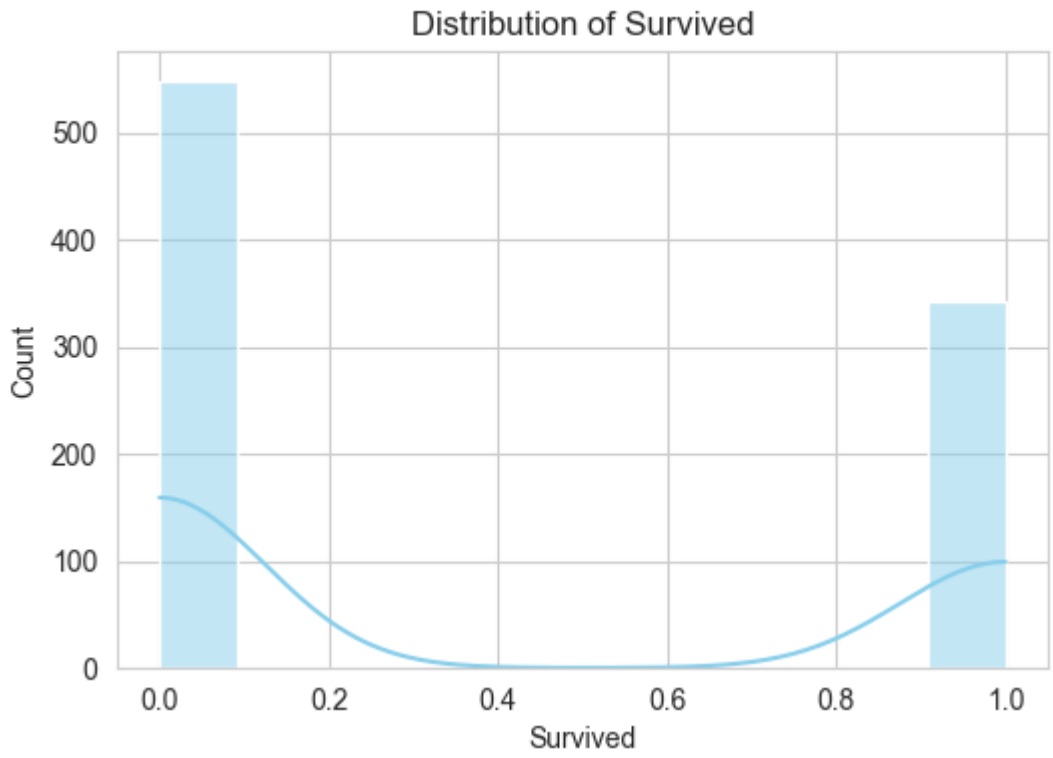

Distribution of Fare
Count
Fare
0
50
100
150
200
250
300
0
100
200
300
400
500

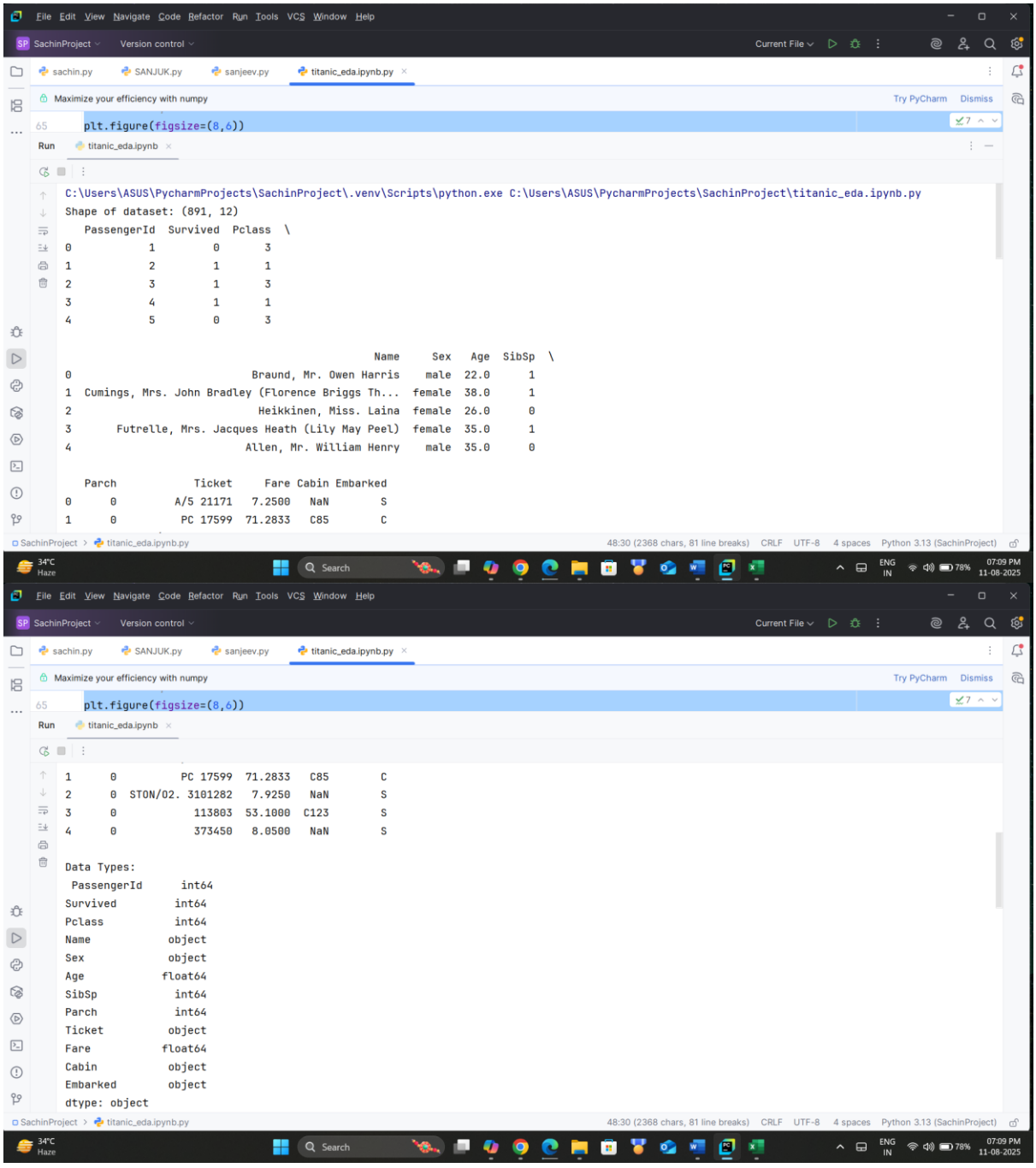

Count of Sex
600
500
400
300
200
100
0
count
male
female


Distribution of SibSp
600
500
400
300
200
100
0
Count
0
1
2
3
4
5
6
7
8
SibSp

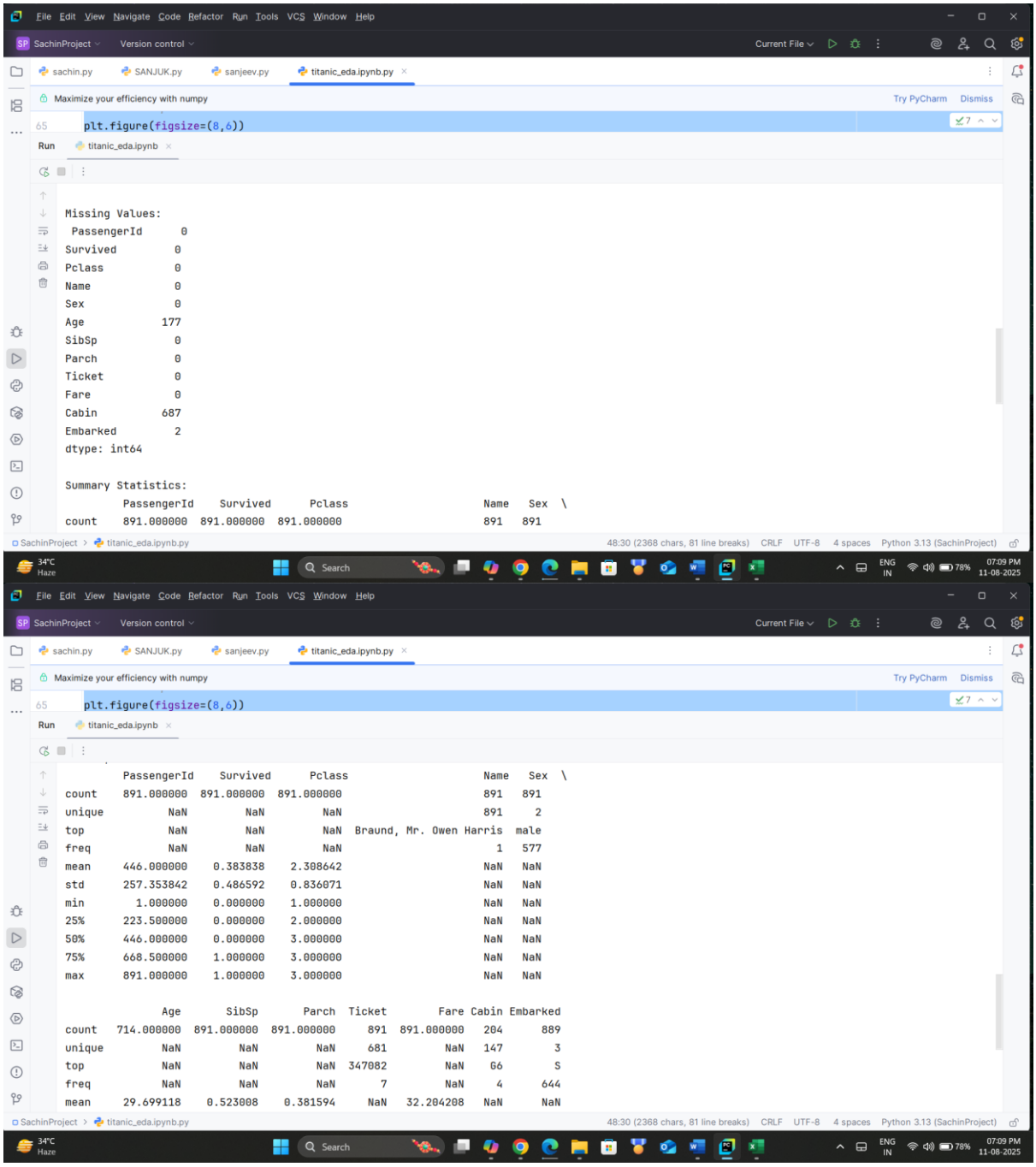

Survival Count by Passenger Class
Survived
0
1
count
350
300
250
200
150
100
50
0
1
2
3
Pclass


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

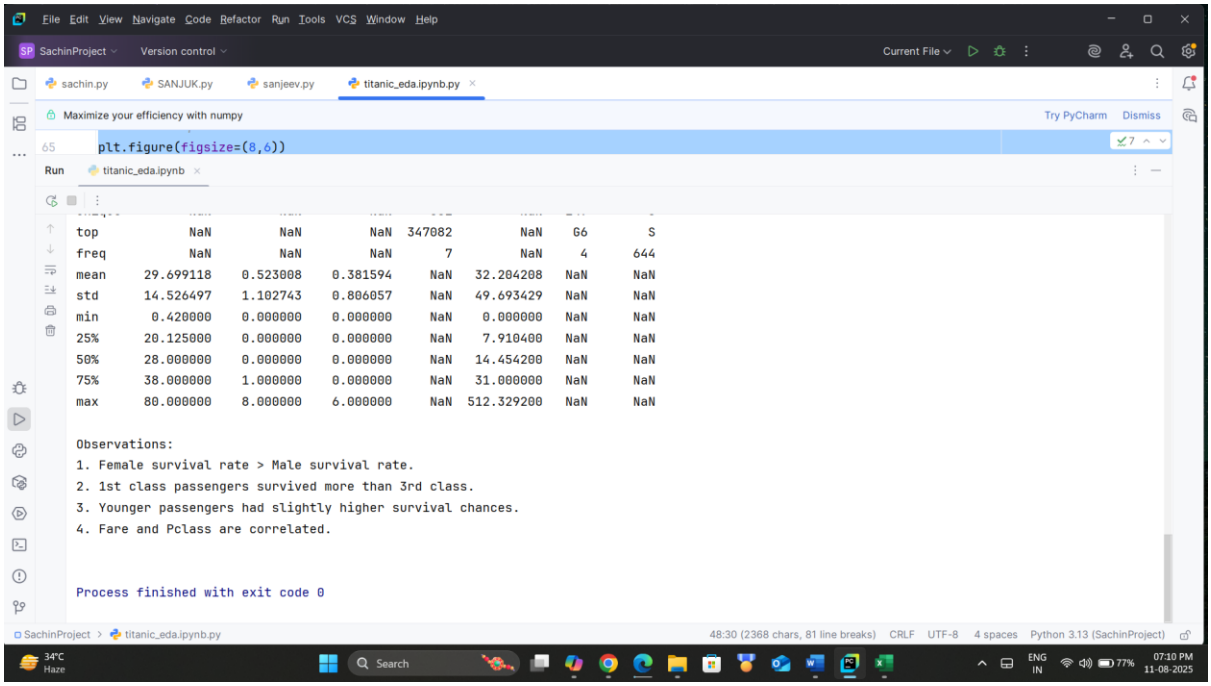

Distribution of Survived
Count
500
400
300
200
100
0
0.0
0.2
0.4
0.6
0.8
1.0
Survived


File Edit View Navigate Code Refactor Run Tools VCS Window Help
SachinProject
Version control
Current File
sachin.py
SANJUK.py
sanjeev.py
titanic_eda.ipynb.py
Maximize your efficiency with numpy
Try PyCharm
Dismiss
65 plt.figure(figsize=(8,6))
Run
titanic_eda.ipynb
C:\Users\ASUS\PycharmProjects\SachinProject\.venv\Scripts\python.exe C:\Users\ASUS\PycharmProjects\SachinProject\titanic_eda.ipynb.py
Shape of dataset: (891, 12)
PassengerId Survived Pclass \
0 1 0 3
1 2 1 1
2 3 1 3
3 4 1 1
4 5 0 3
Name Sex Age SibSp \
0 Braund, Mr. Owen Harris male 22.0 1
1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 1
2 Heikkinen, Miss. Laina female 26.0 0
3 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 1
4 Allen, Mr. William Henry male 35.0 0
Parch Ticket Fare Cabin Embarked
0 0 A/5 21171 7.2500 NaN S
1 0 PC 17599 71.2833 C85 C
SachinProject > titanic_eda.ipynb.py
48:30 (2368 chars, 81 line breaks) CRLF UTF-8 4 spaces Python 3.13 (SachinProject)
34°C Haze
Search
ENG IN
78%
07:09 PM 11-08-2025

File Edit View Navigate Code Refactor Run Tools VCS Window Help
SachinProject
Version control
Current File
sachin.py
SANJUK.py
sanjeev.py
titanic_eda.ipynb.py
Maximize your efficiency with numpy
Try PyCharm
Dismiss
65 plt.figure(figsize=(8,6))
Run
titanic_eda.ipynb
1 0 PC 17599 71.2833 C85 C
2 0 STON/02. 3101282 7.9250 NaN S
3 0 113803 53.1000 C123 S
4 0 373450 8.0500 NaN S
Data Types:
PassengerId int64
Survived int64
Pclass int64
Name object
Sex object
Age float64
SibSp int64
Parch int64
Ticket object
Fare float64
Cabin object
Embarked object
dtype: object
SachinProject > titanic_eda.ipynb.py
48:30 (2368 chars, 81 line breaks) CRLF UTF-8 4 spaces Python 3.13 (SachinProject)
34°C Haze
Search
ENG IN
78%
07:09 PM 11-08-2025


File Edit View Navigate Code Refactor Run Tools VCS Window Help
SachinProject
Version control
Current File
sachin.py
SANJUK.py
sanjeev.py
titanic_eda.ipynb.py
Maximize your efficiency with numpy
Try PyCharm
Dismiss
65 plt.figure(figsize=(8,6))
Run
titanic_eda.ipynb
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
Summary Statistics:
PassengerId Survived Pclass Name Sex \
count 891.000000 891.000000 891.000000 891 891
SachinProject > titanic_eda.ipynb.py
48:30 (2368 chars, 81 line breaks) CRLF UTF-8 4 spaces Python 3.13 (SachinProject)
34°C Haze
Search
ENG IN
78%
07:09 PM 11-08-2025

File Edit View Navigate Code Refactor Run Tools VCS Window Help
SachinProject
Version control
Current File
sachin.py
SANJUK.py
sanjeev.py
titanic_eda.ipynb.py
Maximize your efficiency with numpy
Try PyCharm
Dismiss
65 plt.figure(figsize=(8,6))
Run
titanic_eda.ipynb
PassengerId Survived Pclass Name Sex \
count 891.000000 891.000000 891.000000 891 891
unique NaN NaN NaN 891 2
top NaN NaN NaN Braund, Mr. Owen Harris 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
SachinProject > titanic_eda.ipynb.py
48:30 (2368 chars, 81 line breaks) CRLF UTF-8 4 spaces Python 3.13 (SachinProject)
34°C Haze
Search
ENG IN
78%
07:09 PM 11-08-2025



Observations from EDA

1. Dataset Overview

Shape: 891 rows × 12 columns

Contains both numerical (Age, Fare, SibSp, Parch, Survived) and categorical (Sex, Embarked, Pclass, Cabin, Ticket) variables.

Missing values:

Age → ~177 missing

Cabin → ~687 missing

Embarked → 2 missing

2. Univariate Analysis

Numerical Columns

Age: Most passengers were between 20–40 years old; slightly right-skewed.

Fare: Highly skewed; a few passengers paid very high fares.

SibSp & Parch: Majority had 0 siblings/spouses and 0 parents/children onboard.

Categorical Columns

Sex: ~65% male passengers, ~35% female.

Pclass: Most passengers were in 3rd class (~55%).

Embarked: Majority boarded from Southampton (~72%).

3. Bivariate Analysis

Survival by Gender: Female survival rate was much higher (~74%) compared to males (~19%).

Survival by Passenger Class: 1st class had the highest survival (~63%), 3rd class the lowest (~24%).

Age vs Survival: Survivors were slightly younger on average, but the difference was not very large.

4. Correlation Analysis

Strong negative correlation between Fare and Pclass (~-0.55).

SibSp and Parch had a weak positive relationship with survival.

Age showed very low correlation with survival.

5. Pairplot Insights

First-class passengers paid higher fares and had a much higher survival rate.

Third-class passengers paid lower fares and had the lowest survival rate.

Female passengers in all classes had higher survival chances than males.

6. Summary of Findings

Survival was strongly influenced by gender and passenger class.

Higher fares were generally associated with higher survival rates.

Age had only a minor effect on survival.