Distribution of Age

Age Distribution by Survival

Correlation Heatmap

Count of Embarked

Distribution of Fare

Distribution of Parch

Distribution of PassengerId

Distribution of Pclass

Distribution of Fare

Count of Sex

Distribution of SibSp

Survival Count by Passenger Class

Survival Count by Gender

Distribution of Survived

```python
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
```

```python
65    plt.figure(figsize=(8,6))
```

Run   🐍 titanic_eda.ipynb ×

```
1      0         PC 17599  71.2833   C85        C
2      0  STON/O2. 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
```

```
65    plt.figure(figsize=(8,6))
```

```
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
```

```
65    plt.figure(figsize=(8,6))
```

```
        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
```

The screenshot shows a PyCharm IDE with a titanic_eda.ipynb file. The output console displays:

```
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


Observations:
1. Female survival rate > Male survival rate.
2. 1st class passengers survived more than 3rd class.
3. Younger passengers had slightly higher survival chances.
4. Fare and Pclass are correlated.


Process finished with exit code 0
```

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.