

Titanic Dataset - Data Cleaning and EDA Report

1) File info

Original file: titanic.csv
Original shape: (891, 12)
Duplicates before cleaning: 0
Duplicates after cleaning: 0

2) Cleaning actions

- Filled missing 'Age' with mean = 29.70
- Filled missing 'Embarked' with mode
- Dropped duplicate rows

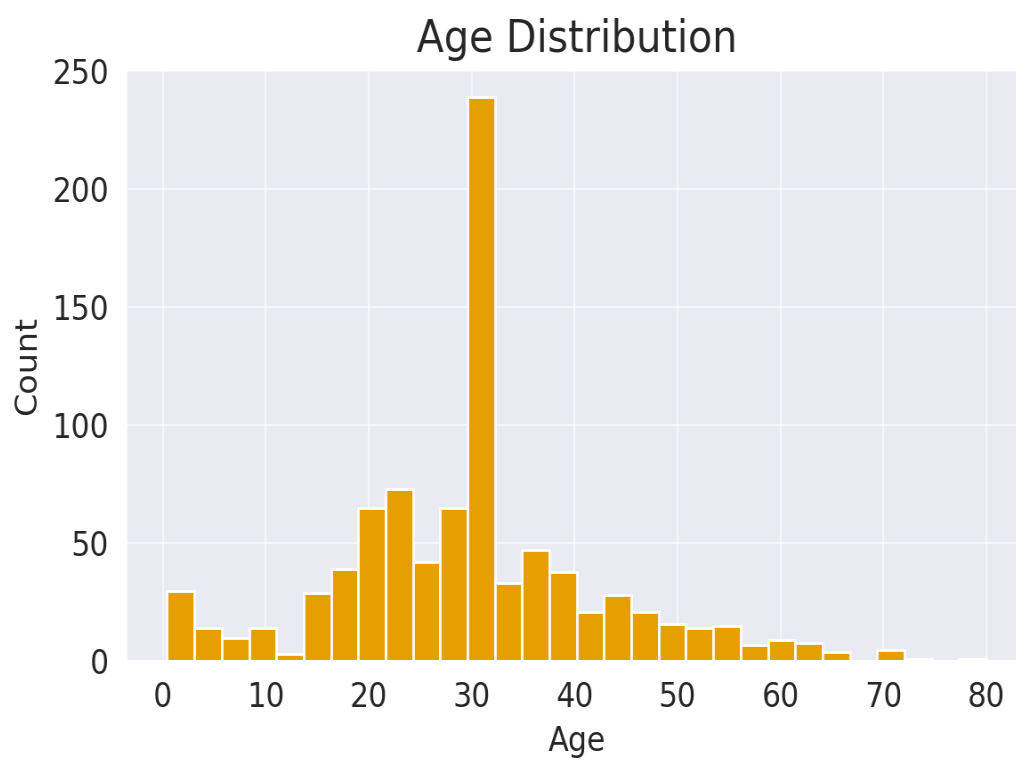
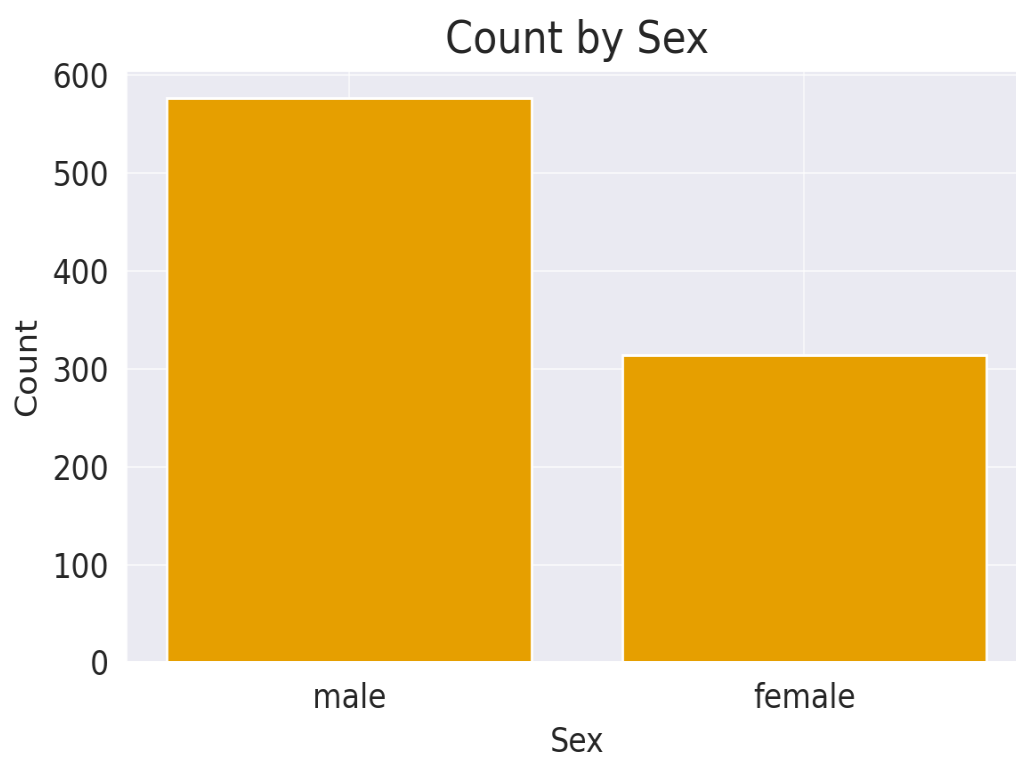
3) Sample of cleaned data (first 10 rows):

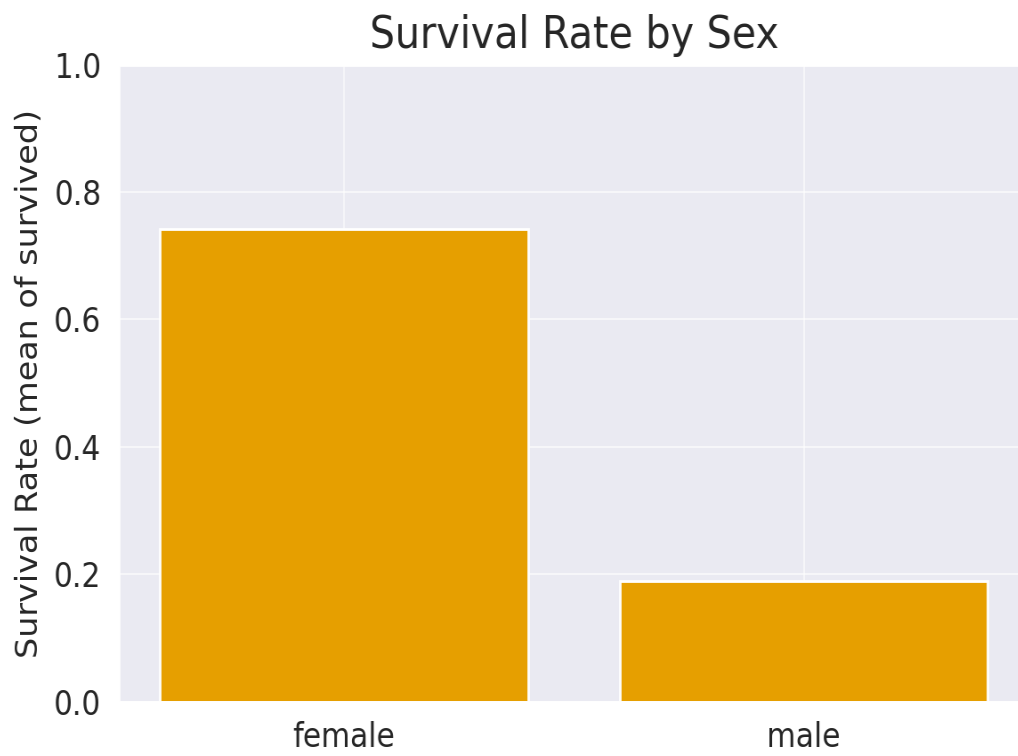
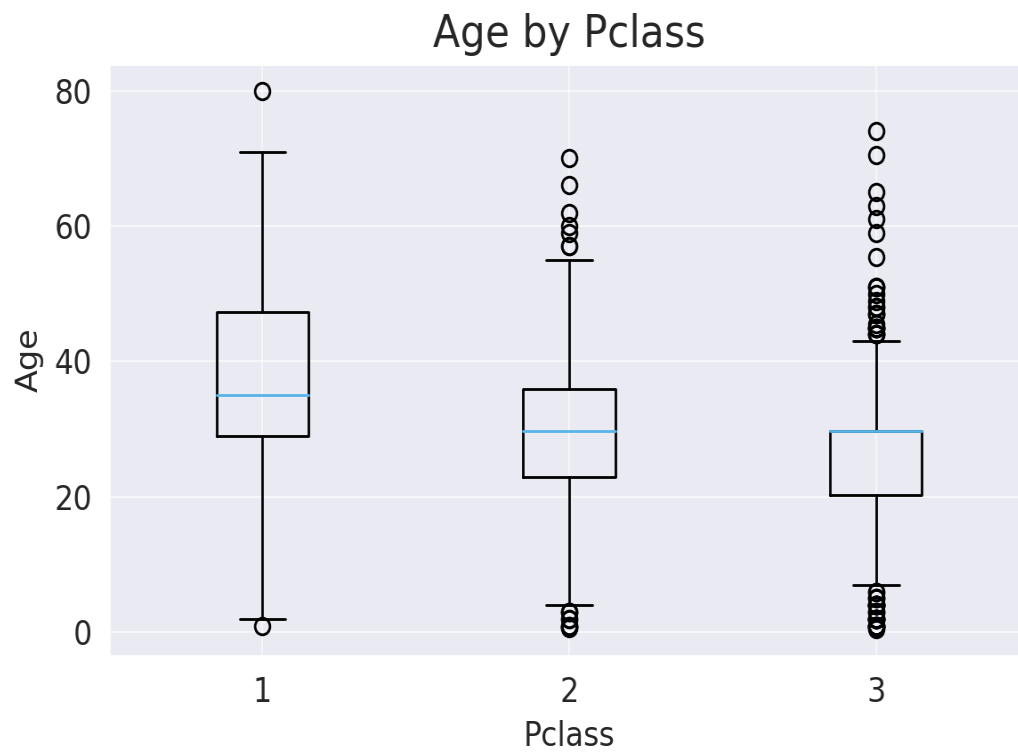
```
PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked 1 0 3 Braund, Mr. Owen Harris male
22.000000 1 0 A/5 21171 7.2500 NaN S 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female 38.000000 1 0 PC
17599 71.2833 C85 C 3 1 3 Heikkinen, Miss. Laina female 26.000000 0 0 STON/O2. 3101282 7.9250 NaN S 4 1 1 Futrelle, Mrs.
Jacques Heath (Lily May Peel) female 35.000000 1 0 113803 53.1000 C123 S 5 0 3 Allen, Mr. William Henry male 35.000000 0 0
373450 8.0500 NaN S 6 0 3 Moran, Mr. James male 29.699118 0 0 330877 8.4583 NaN Q 7 0 1 McCarthy, Mr. Timothy J male
54.000000 0 0 17463 51.8625 E46 S 8 0 3 Palsson, Master. Gosta Leonard male 2.000000 3 1 349909 21.0750 NaN S 9 1 3
Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg) female 27.000000 0 2 347742 11.1333 NaN S 10 1 2 Nasser, Mrs. Nicholas
(Adele Achem) female 14.000000 1 0 237736 30.0708 NaN C
```

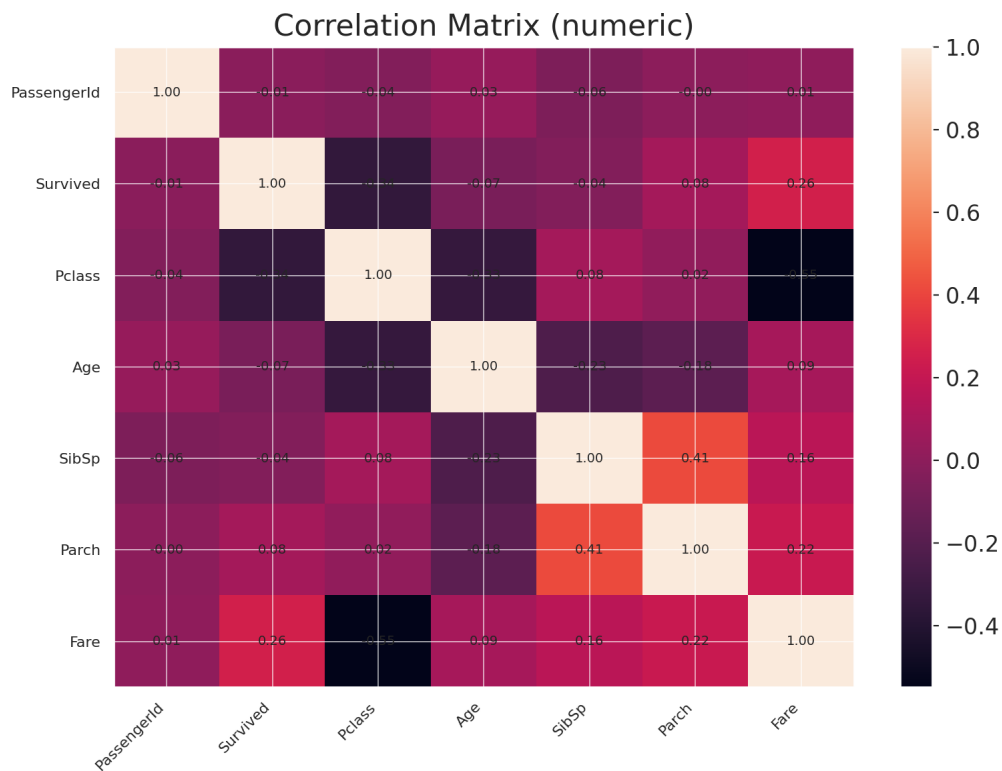
4) Summary statistics:

```
PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked count 891.000000 891.000000
891.000000 891 891 891.000000 891.000000 891.000000 891 891.000000 204 891 unique NaN NaN NaN 891 2 NaN NaN NaN
681 NaN 147 3 top NaN NaN NaN Braund, Mr. Owen Harris male NaN NaN NaN 347082 NaN B96 B98 S freq NaN NaN NaN 1
577 NaN NaN NaN 7 NaN 4 646 mean 446.000000 0.383838 2.308642 NaN NaN 29.699118 0.523008 0.381594 NaN
32.204208 NaN NaN std 257.353842 0.486592 0.836071 NaN NaN 13.002015 1.102743 0.806057 NaN 49.693429 NaN NaN
min 1.000000 0.000000 1.000000 NaN NaN 0.420000 0.000000 0.000000 NaN 0.000000 NaN NaN 25% 223.500000 0.000000
2.000000 NaN NaN 22.000000 0.000000 0.000000 NaN 7.910400 NaN NaN 50% 446.000000 0.000000 3.000000 NaN NaN
29.699118 0.000000 0.000000 NaN 14.454200 NaN NaN 75% 668.500000 1.000000 3.000000 NaN NaN 35.000000 1.000000
0.000000 NaN 31.000000 NaN NaN max 891.000000 1.000000 3.000000 NaN NaN 80.000000 8.000000 6.000000 NaN
512.329200 NaN NaN
```

5) Visualizations







6) Insights (automated):

- female had higher survival rate on average (0.74).
- Pclass with highest survival rate: 1 (0.63).
- Mean age (survived=1): 28.55, (survived=0): 30.42.