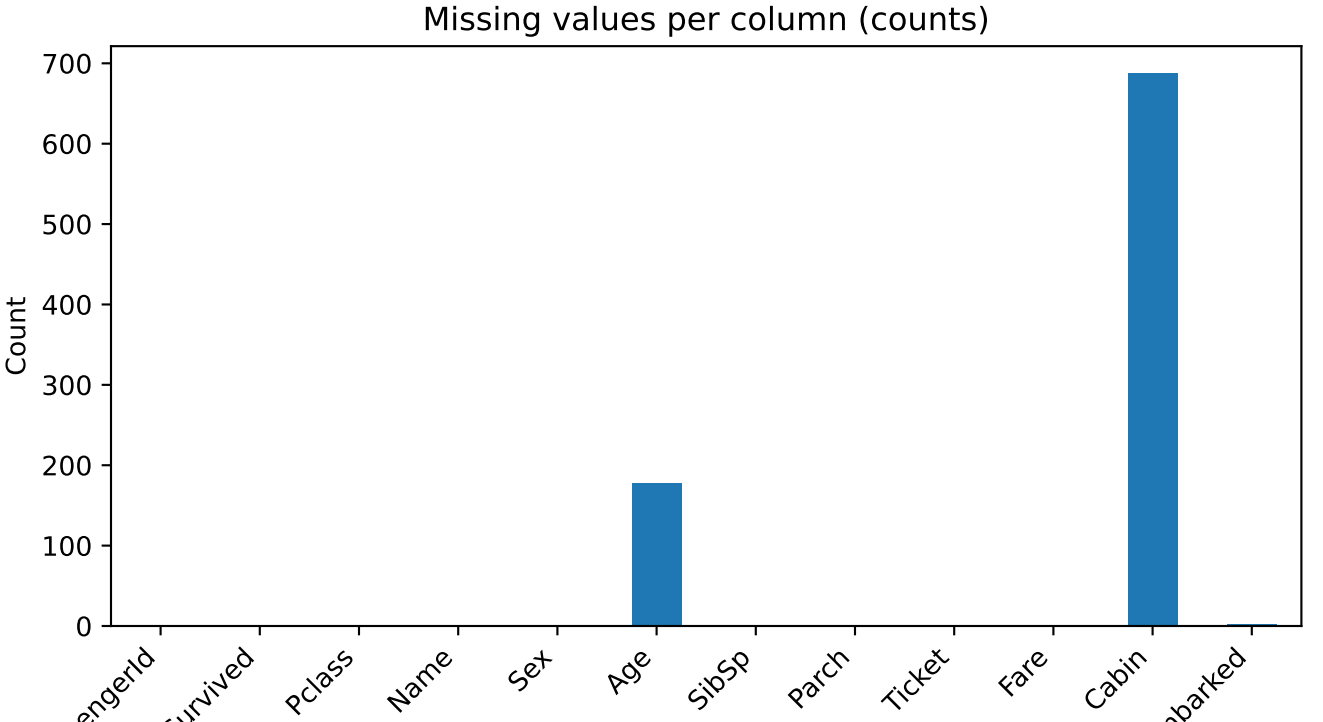
Titanic Dataset — Exploratory Data Analysis (EDA)

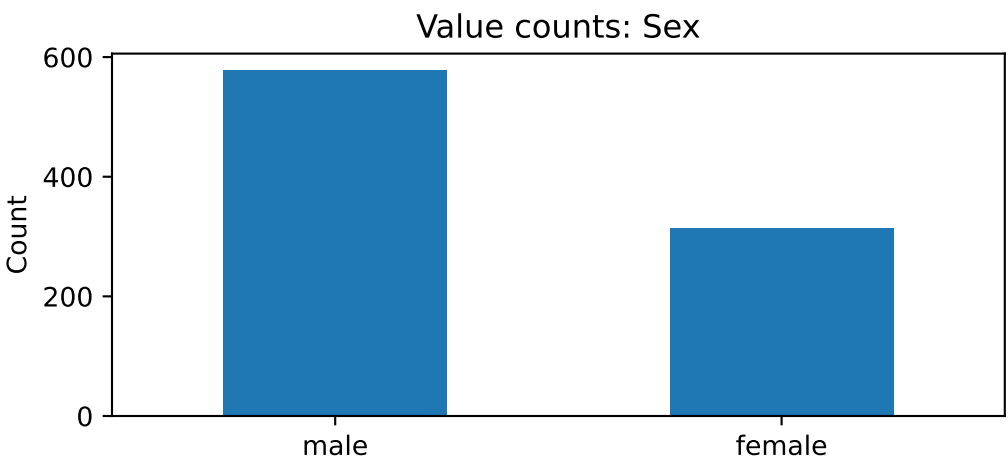
Tools: Python (pandas, matplotlib)

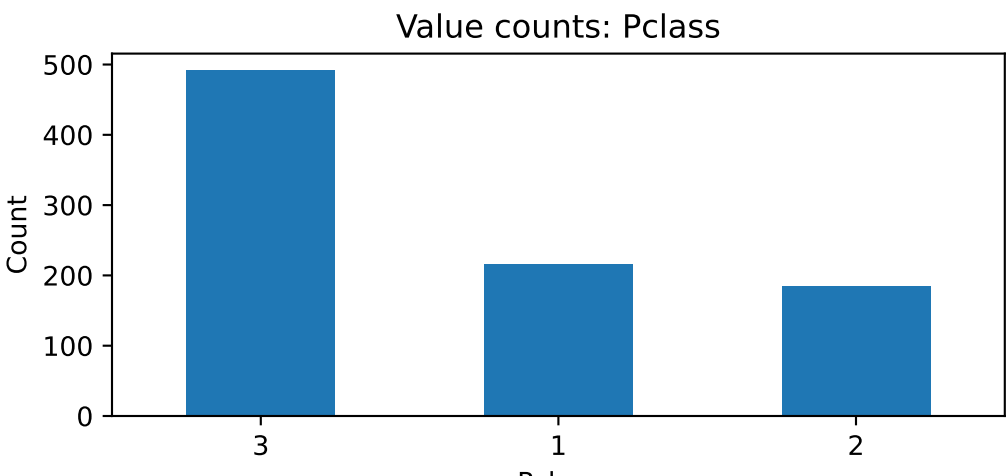
This report contains tables, plots, and observations from the Titanic train.csv dataset.

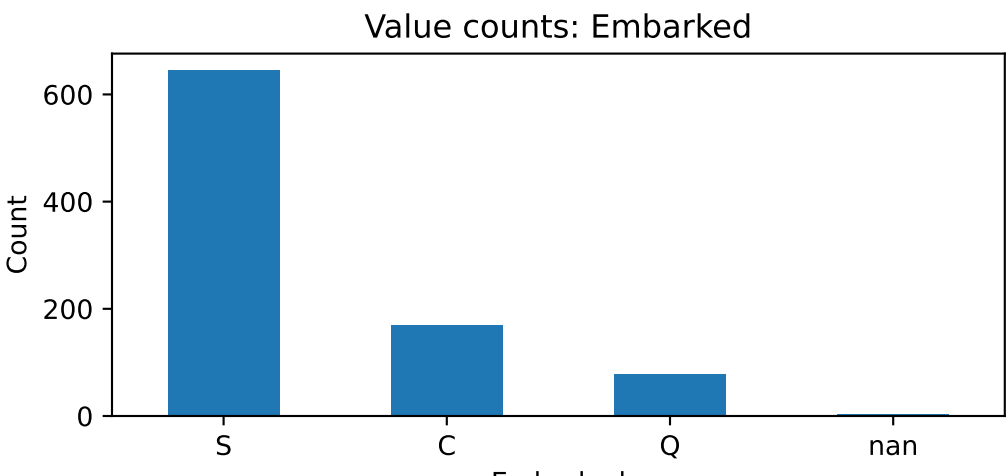
column	dtype	non_null_count
PassengerId	int64	891
Survived	int64	891
Pclass	int64	891
Name	object	891
Sex	object	891
Age	float64	714
SibSp	int64	891
Parch	int64	891
Ticket	object	891
Fare	float64	891
Cabin	object	204
Embarked	object	889

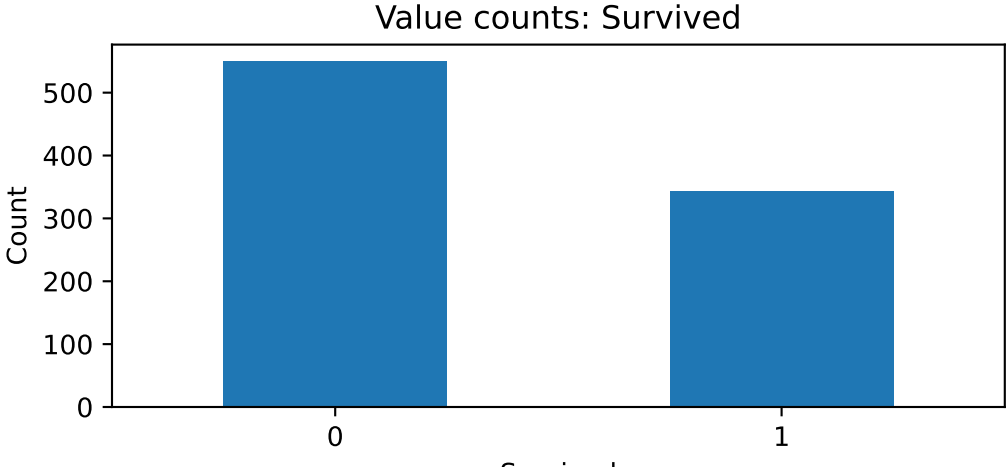
	count	mean	std	min	25%	50%	75%	max
PassengerId	891.0	446.0	257.354	1.0	223.5	446.0	668.5	891.0
Survived	891.0	0.384	0.487	0.0	0.0	0.0	1.0	1.0
Pclass	891.0	2.309	0.836	1.0	2.0	3.0	3.0	3.0
Age	714.0	29.699	14.526	0.42	20.125	28.0	38.0	80.0
SibSp	891.0	0.523	1.103	0.0	0.0	0.0	1.0	8.0
Parch	891.0	0.382	0.806	0.0	0.0	0.0	0.0	6.0
Fare	891.0	32.204	49.693	0.0	7.91	14.454	31.0	512.329

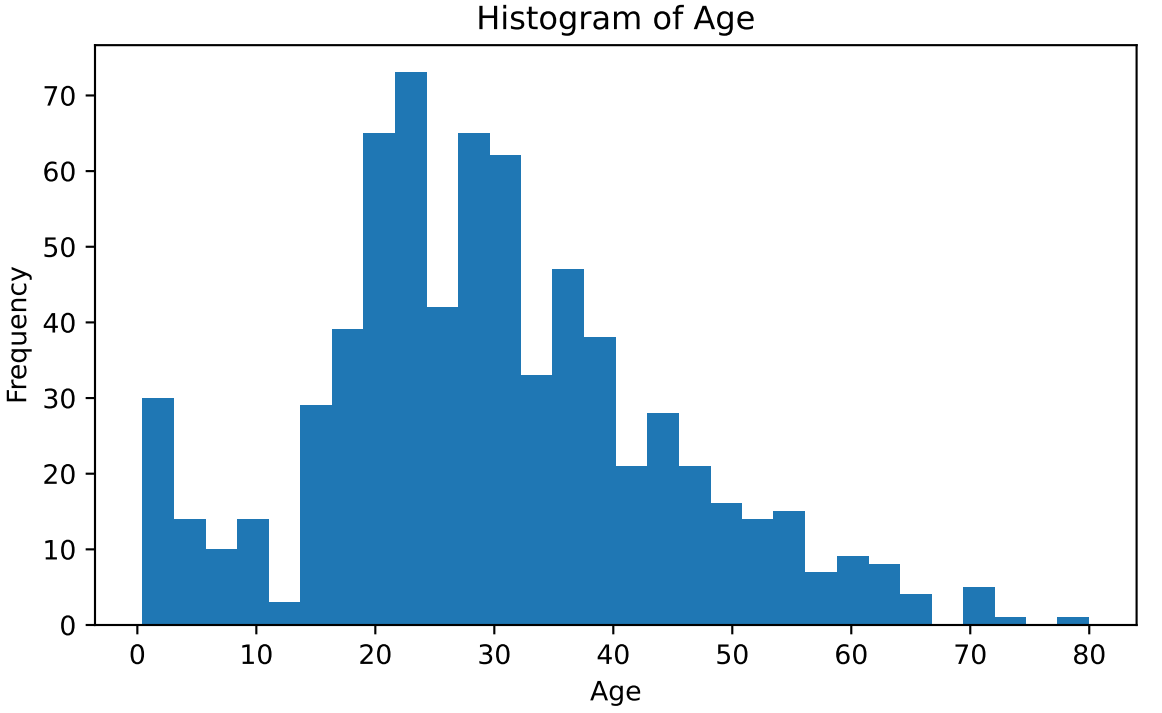

Missing values per column (counts)
Count
0
100
200
300
400
500
600
700
engerId
urvived
Pclass
Name
Sex
Age
SibSp
Parch
Ticket
Fare
Cabin
barked

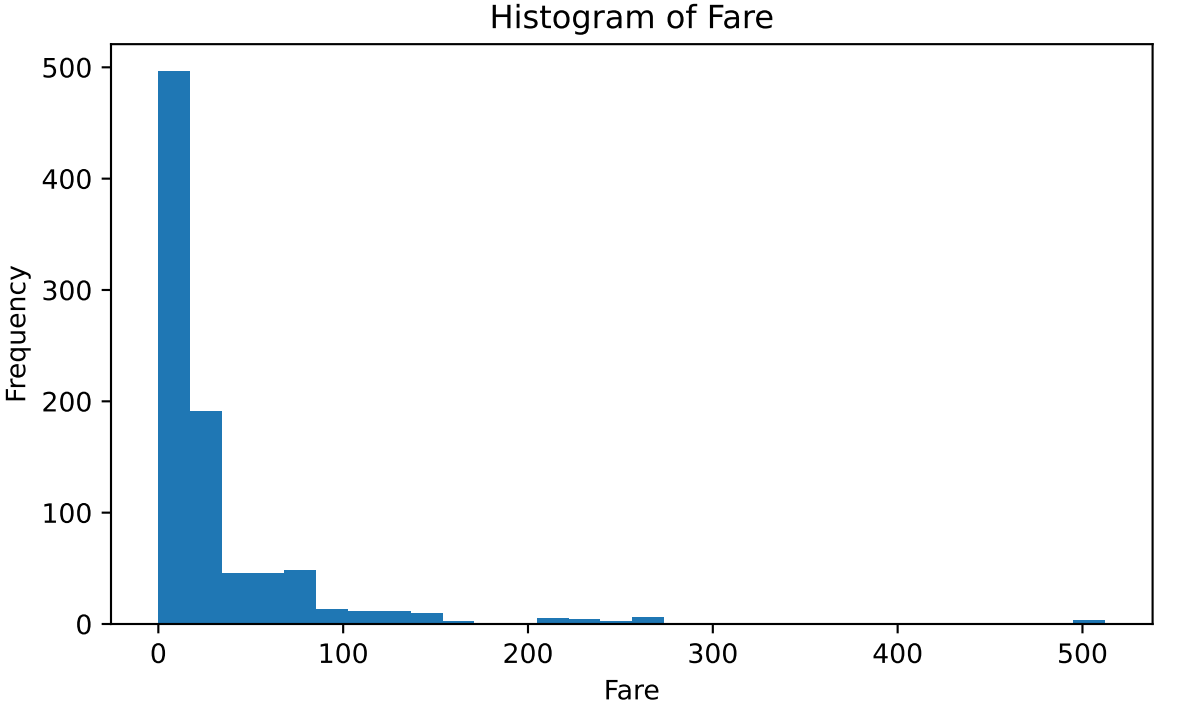

Value counts: Sex
600
400
200
0
Count
male
female

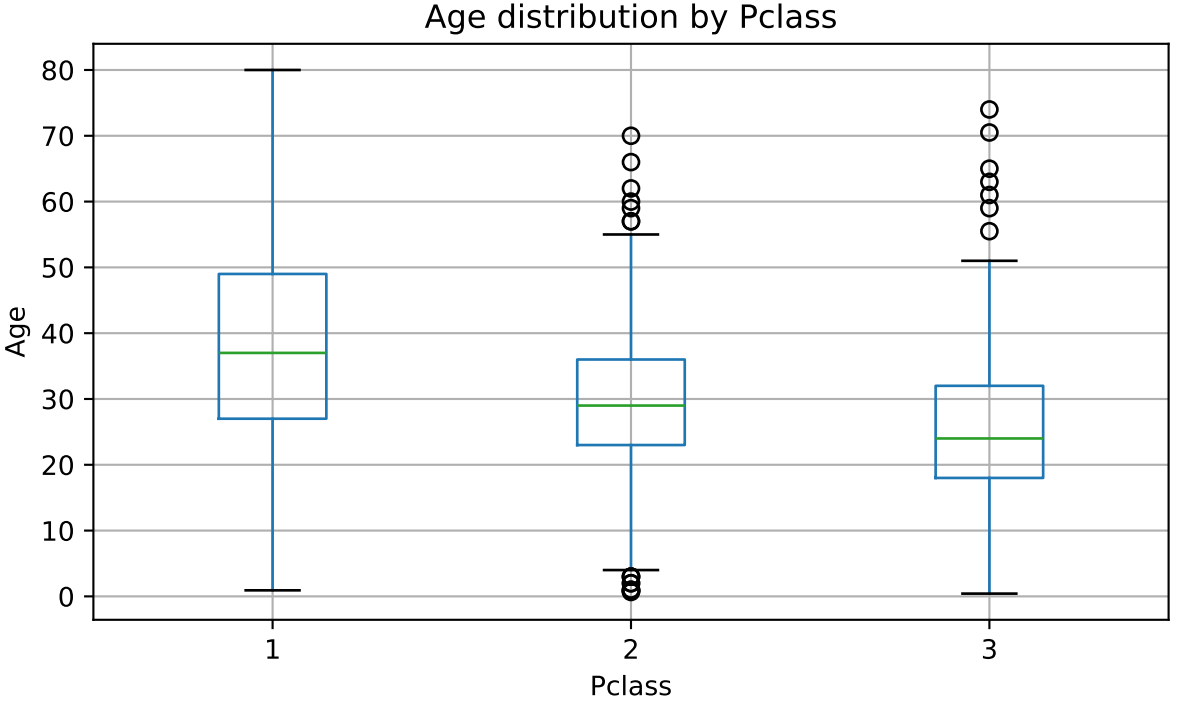

Value counts: Pclass
Count
500
400
300
200
100
0
3
1
2

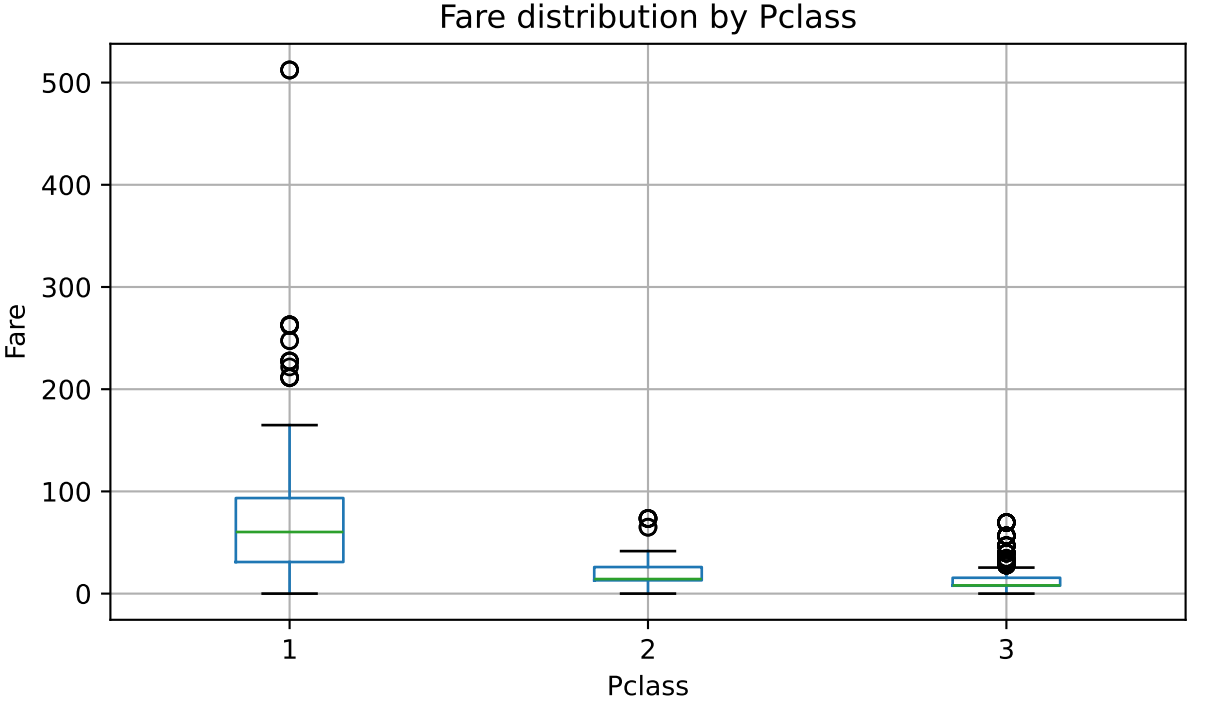

Value counts: Embarked
Count
600
400
200
0
S
C
Q
nan

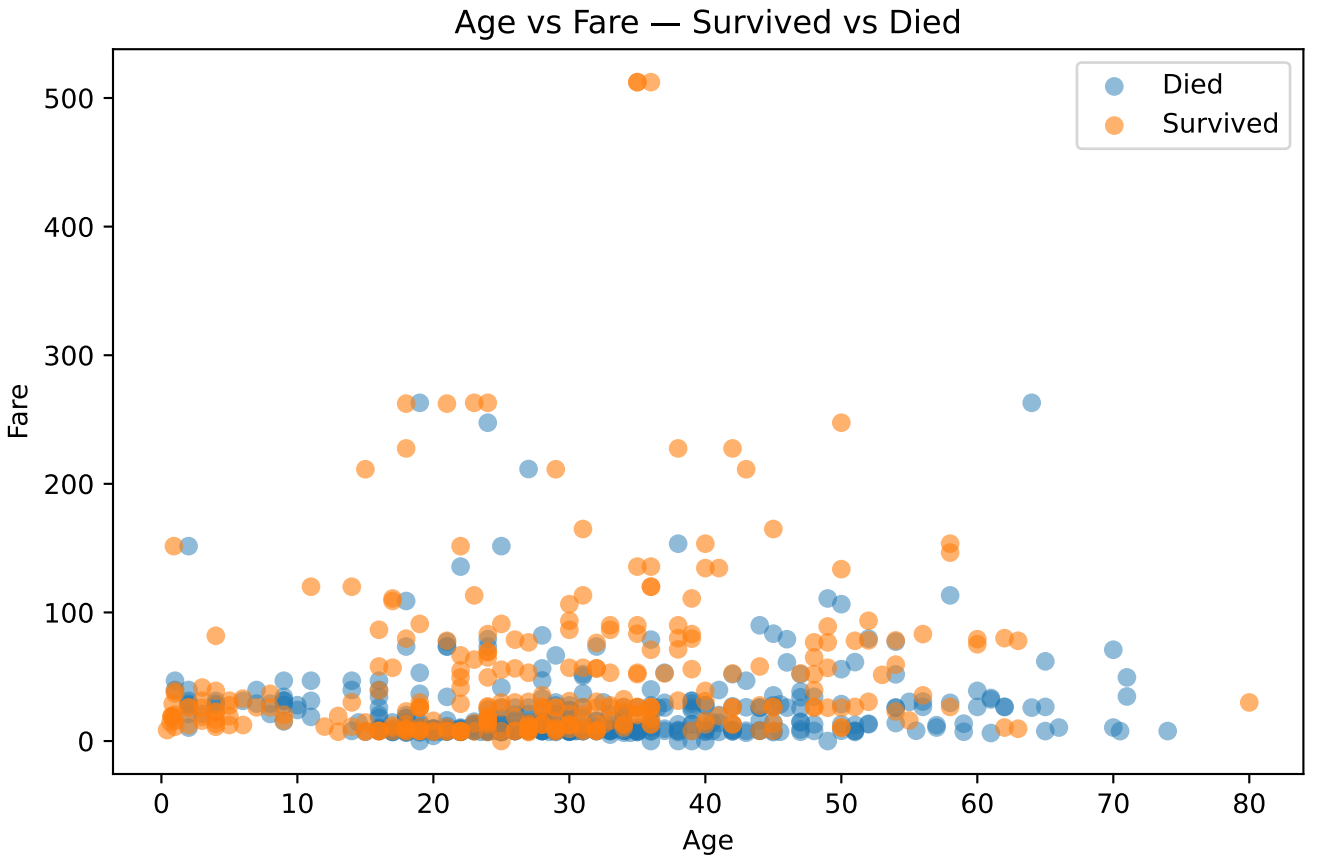

Value counts: Survived
Count
500
400
300
200
100
0
0
1

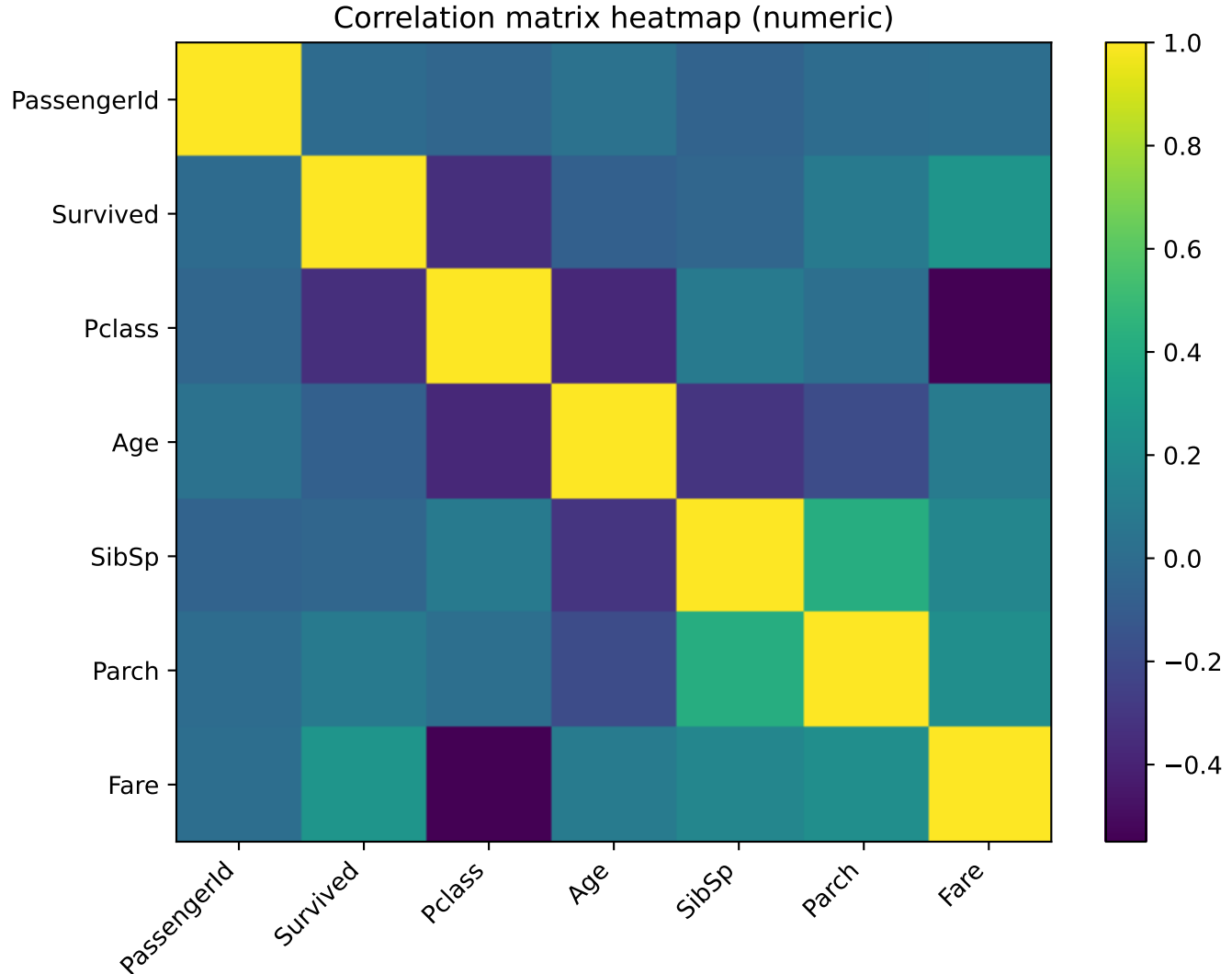

Histogram of Age
Frequency
Age
0
10
20
30
40
50
60
70
80

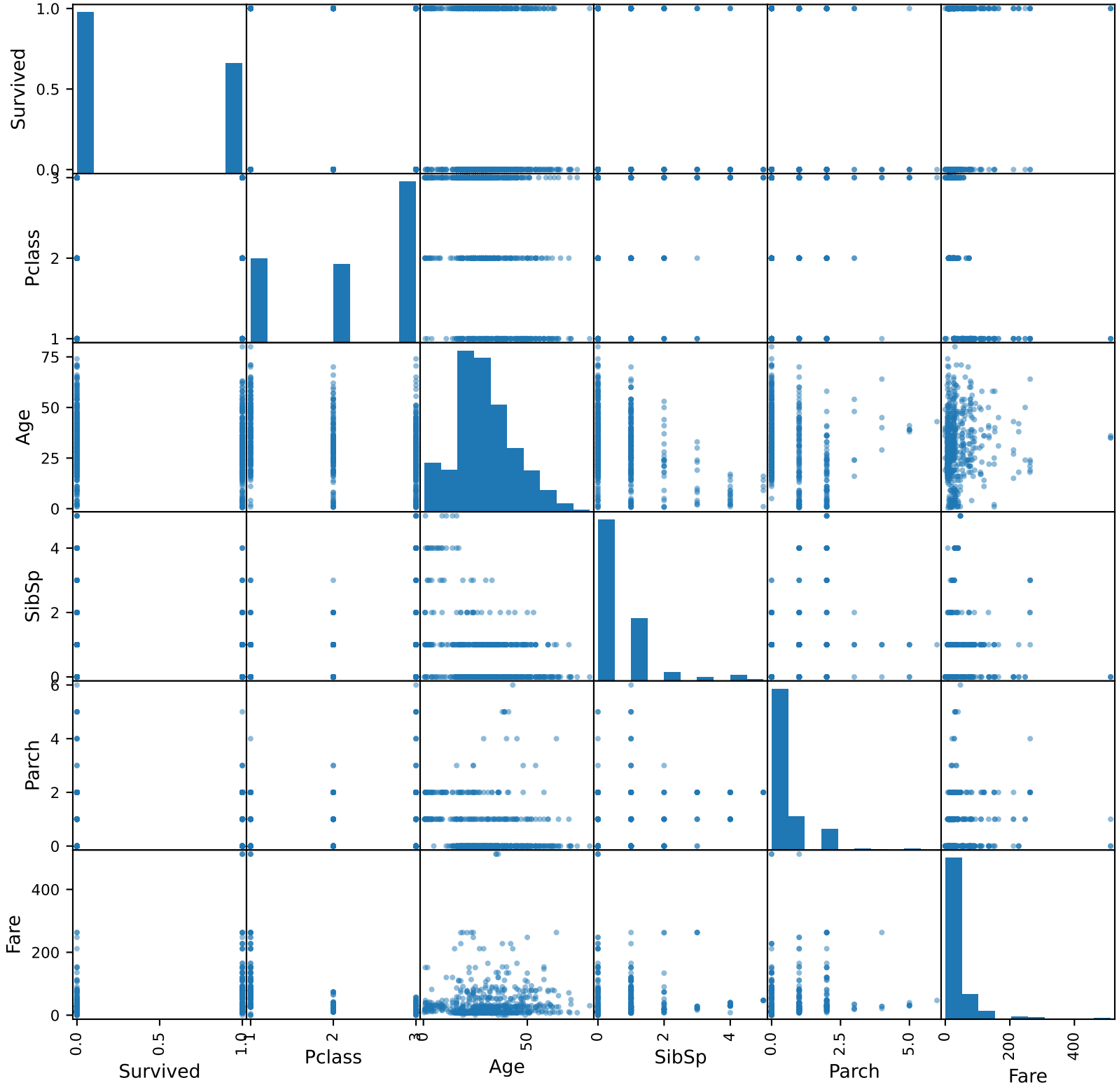

Histogram of Fare
Frequency
Fare
0
100
200
300
400
500
0
100
200
300
400
500

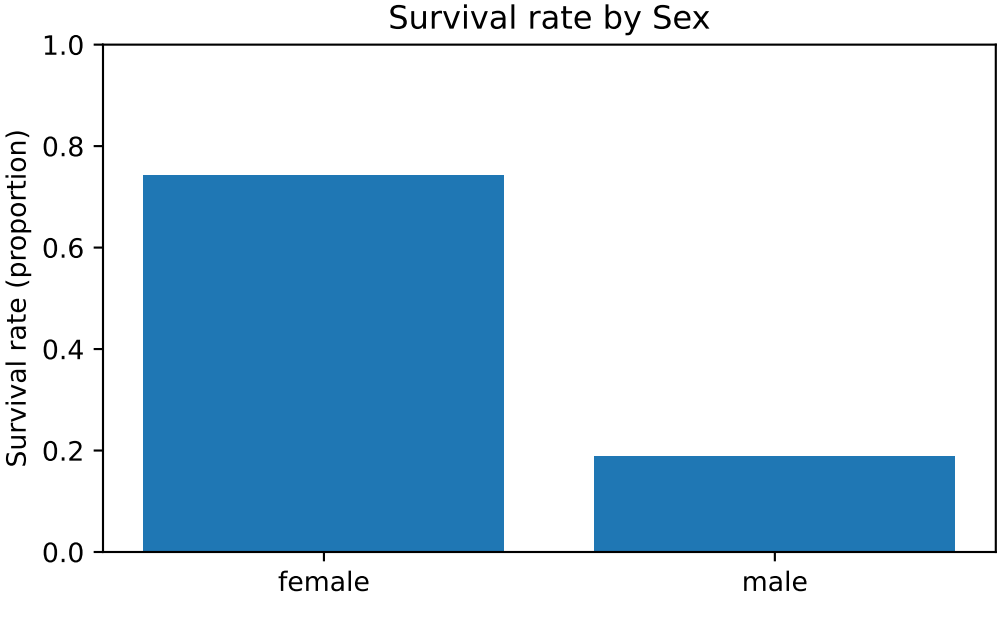

Age distribution by Pclass
Age
Pclass
0
10
20
30
40
50
60
70
80
1
2
3

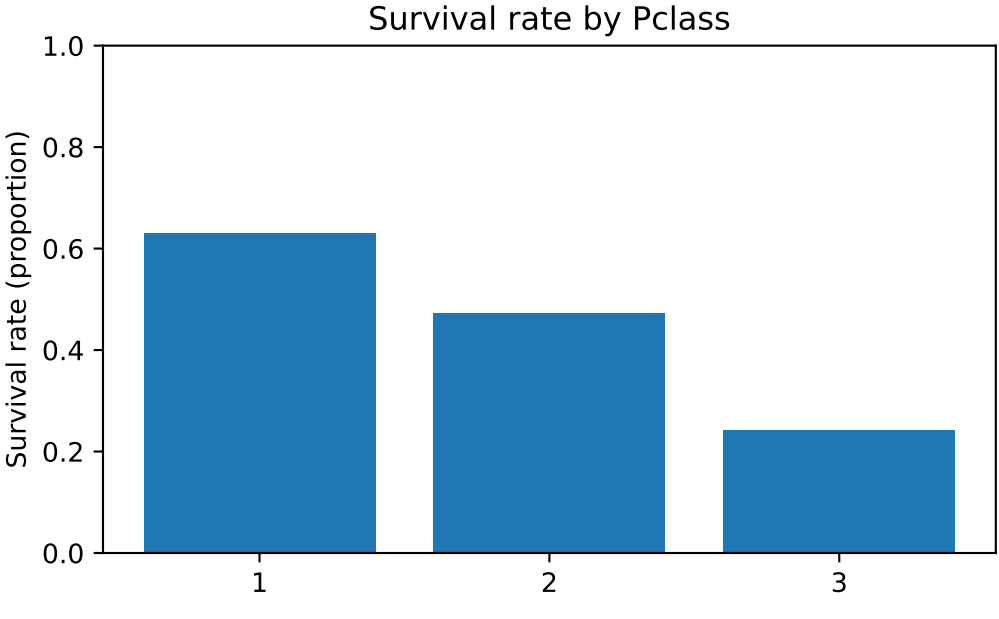

Fare distribution by Pclass
Fare
500
400
300
200
100
0
1
2
3
Pclass

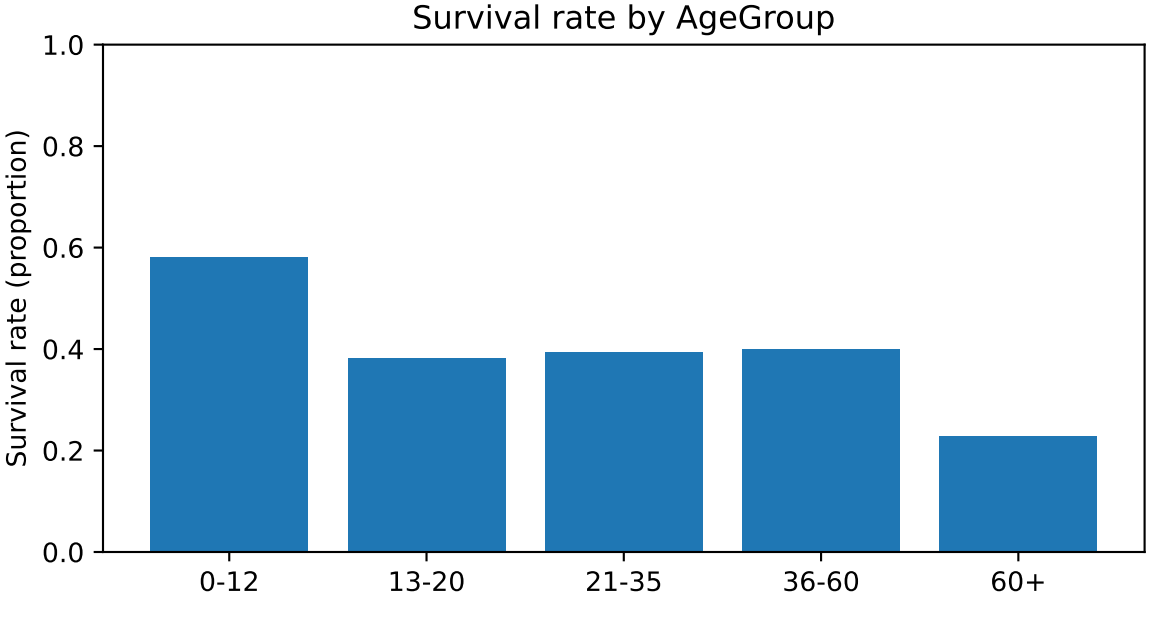

Age vs Fare — Survived vs Died
Died
Survived
Fare
Age
0
100
200
300
400
500
0
10
20
30
40
50
60
70
80


Correlation matrix heatmap (numeric)
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Survived
Pclass
Age
SibSp
Parch
Fare


Survival rate by Sex
Survival rate (proportion)
1.0
0.8
0.6
0.4
0.2
0.0
female
male


Survival rate by Pclass
Survival rate (proportion)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3


Survival rate by AgeGroup
Survival rate (proportion)
1.0
0.8
0.6
0.4
0.2
0.0
0-12
13-20
21-35
36-60
60+

Summary of observations

Observations & Key Findings:

- Missing values: 'Age' and 'Cabin' have substantial missing data; 'Cabin' has the most missin

- Sex & Survival: Females have a significantly higher survival rate than males.

- Pclass & Survival: 1st class passengers had higher survival rates than 2nd and 3rd class.

- Age: Children (0-12) show relatively high survival proportions.

- Fare: Fare distribution is right-skewed; some high-value outliers exist.

- Correlations: 'Pclass' correlates with 'Fare'; associations with 'Survived' are present but mod