Titanic Dataset – Exploratory Data Analysis (EDA)

Univariate, Bivariate & Multivariate Insights

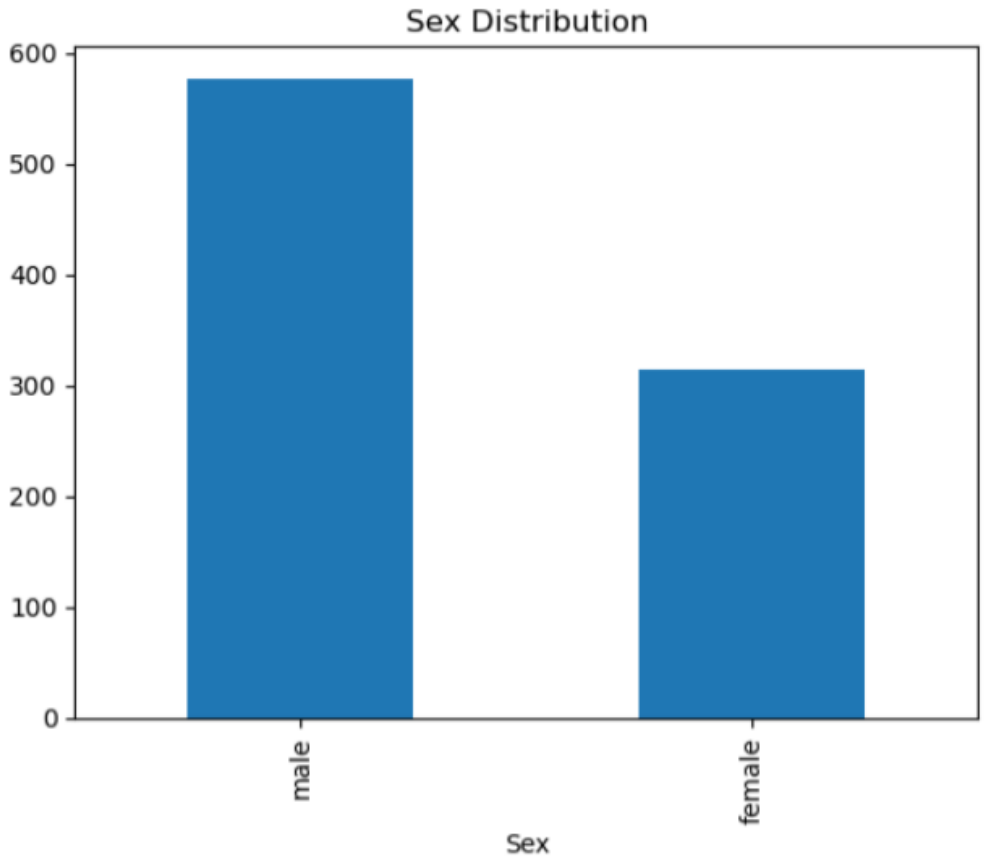
Prepared by Deepika Dohare | Data Analyst |
Tools used (Pandas, Matplotlib, Seaborn)

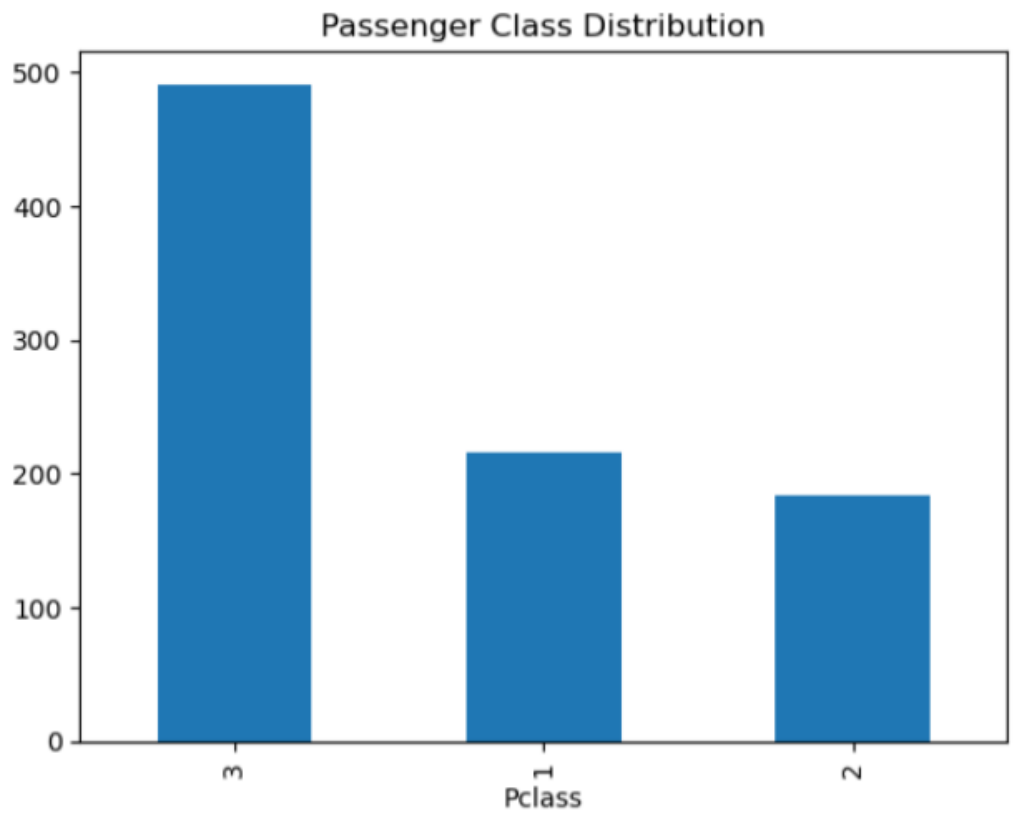
Objective of the Analysis

- Extract insights using visual and statistical exploration
- Gain skills in identifying patterns, trends, and anomalies
- Understand survival patterns among Titanic passengers
- Explore relationships between variables such as class, fare, age, and family size
- Develop intuition for data through univariate, bivariate, and multivariate analysis

Univariate Analysis

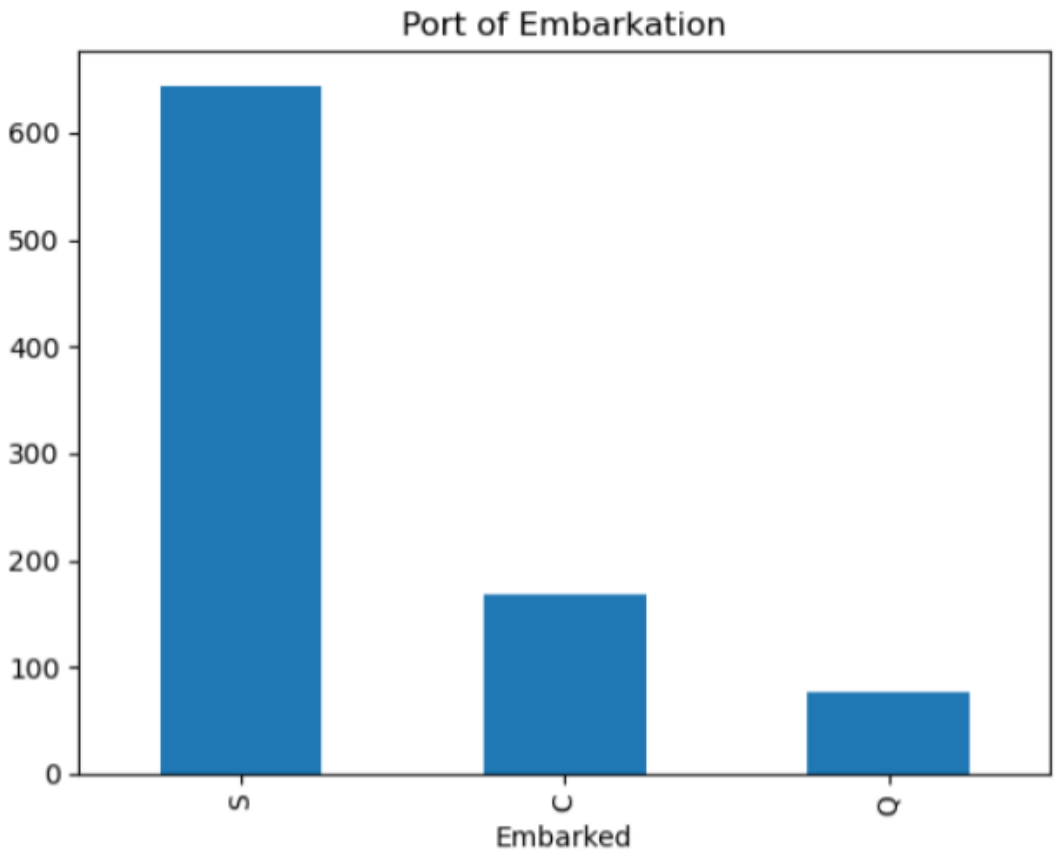
Univariate Analysis – Categorical Features

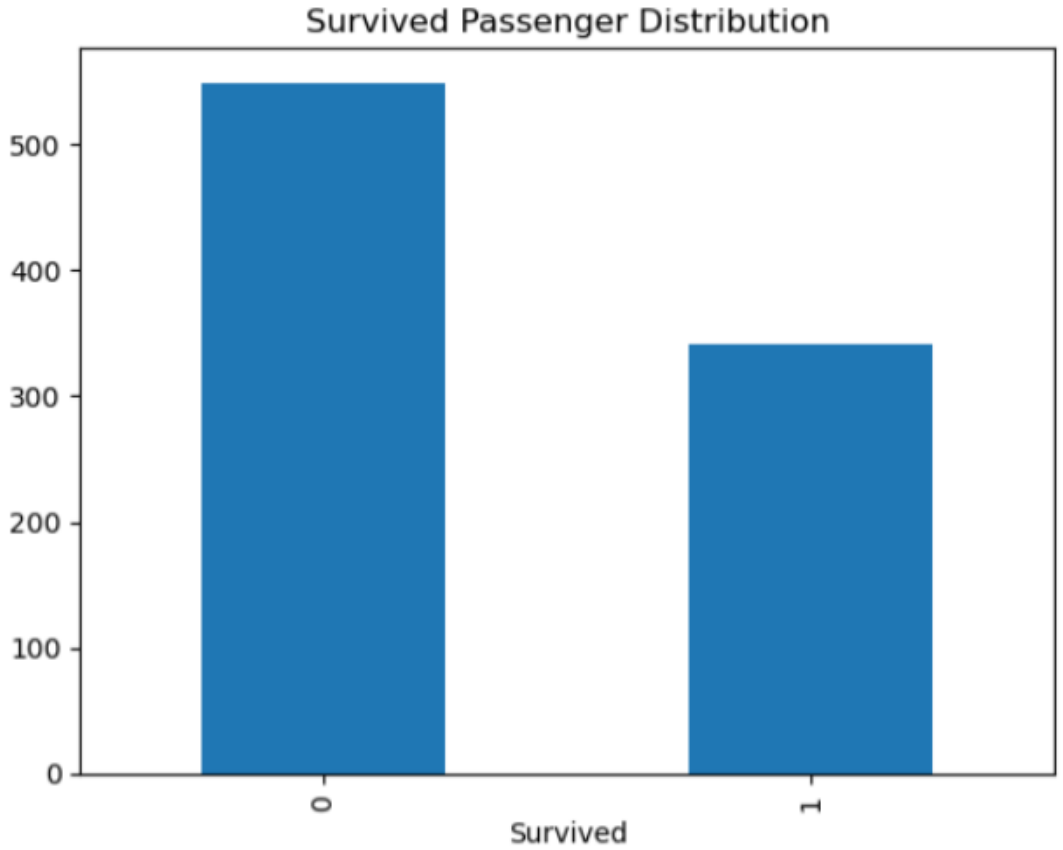




- There were more males than females on board.
- Most passengers belonged to 3rd class, followed by 1st and 2nd.

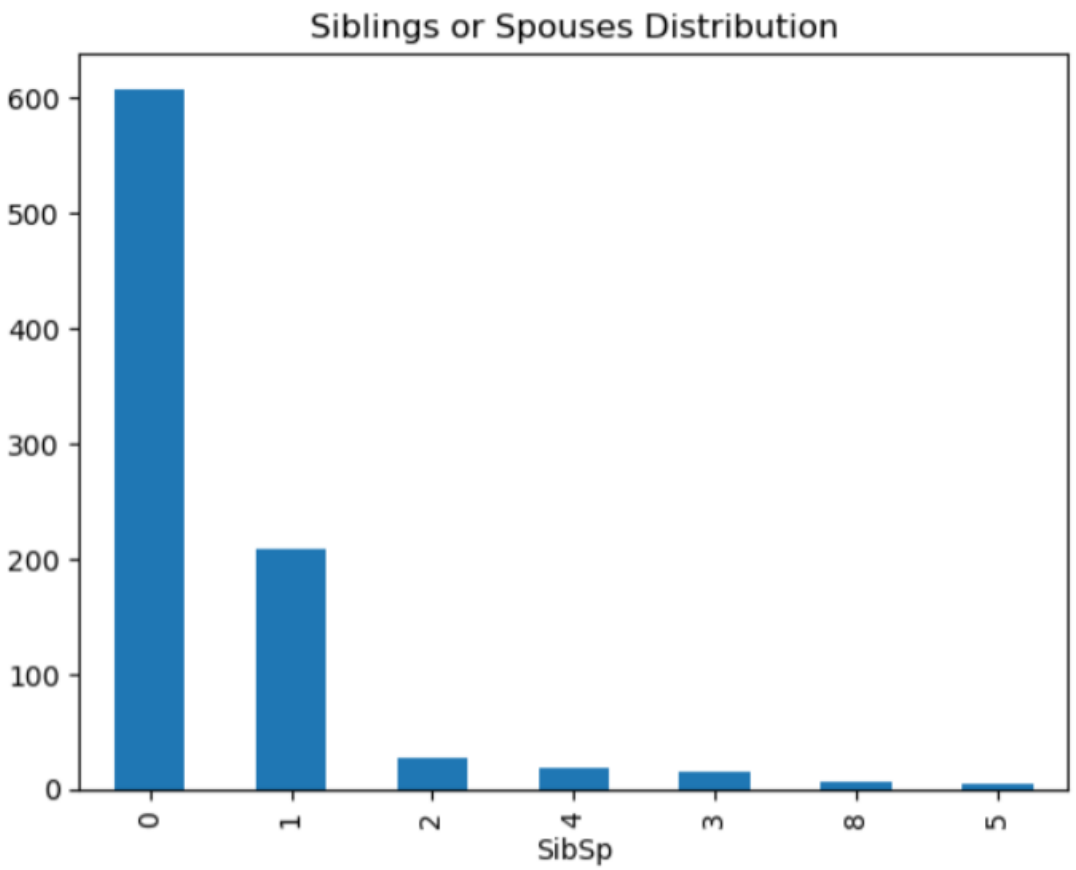
Univariate Analysis – Categorical Features

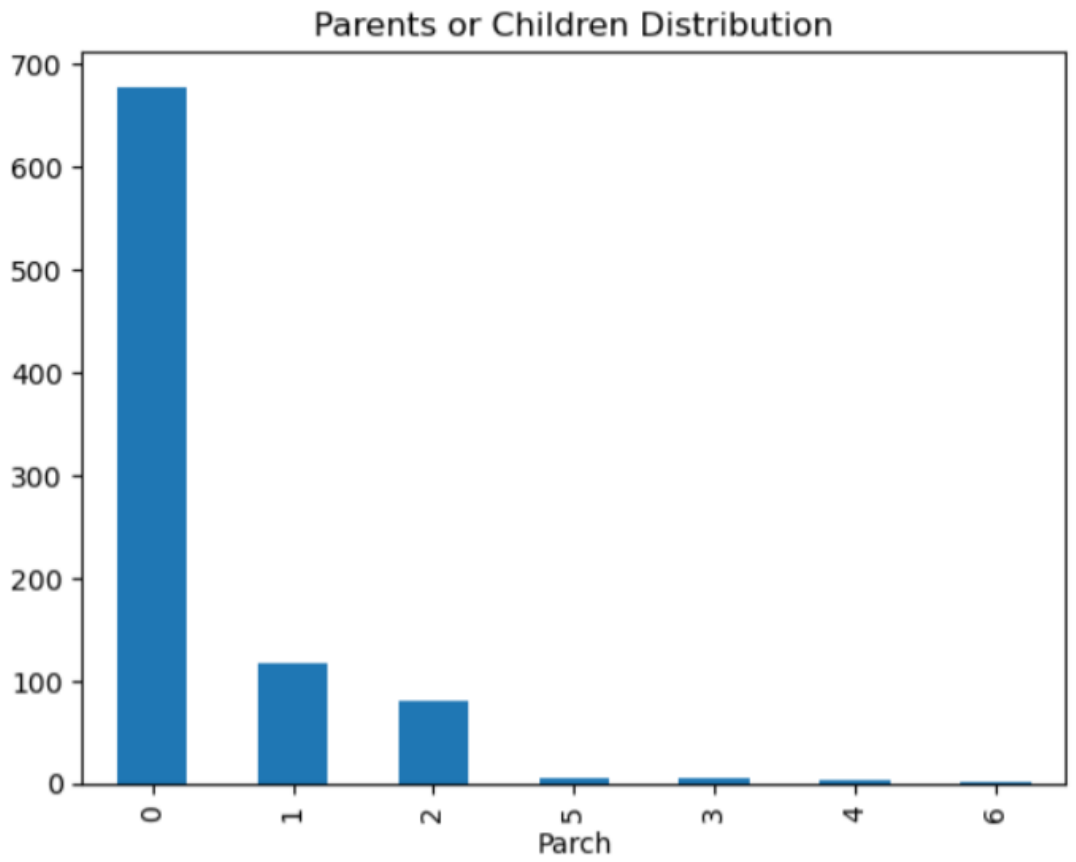




- Most passengers embarked from Southampton (S), followed by Cherbourg (C) and Queenstown (Q).
- About 62% of passengers did not survive, while around 38% survived.

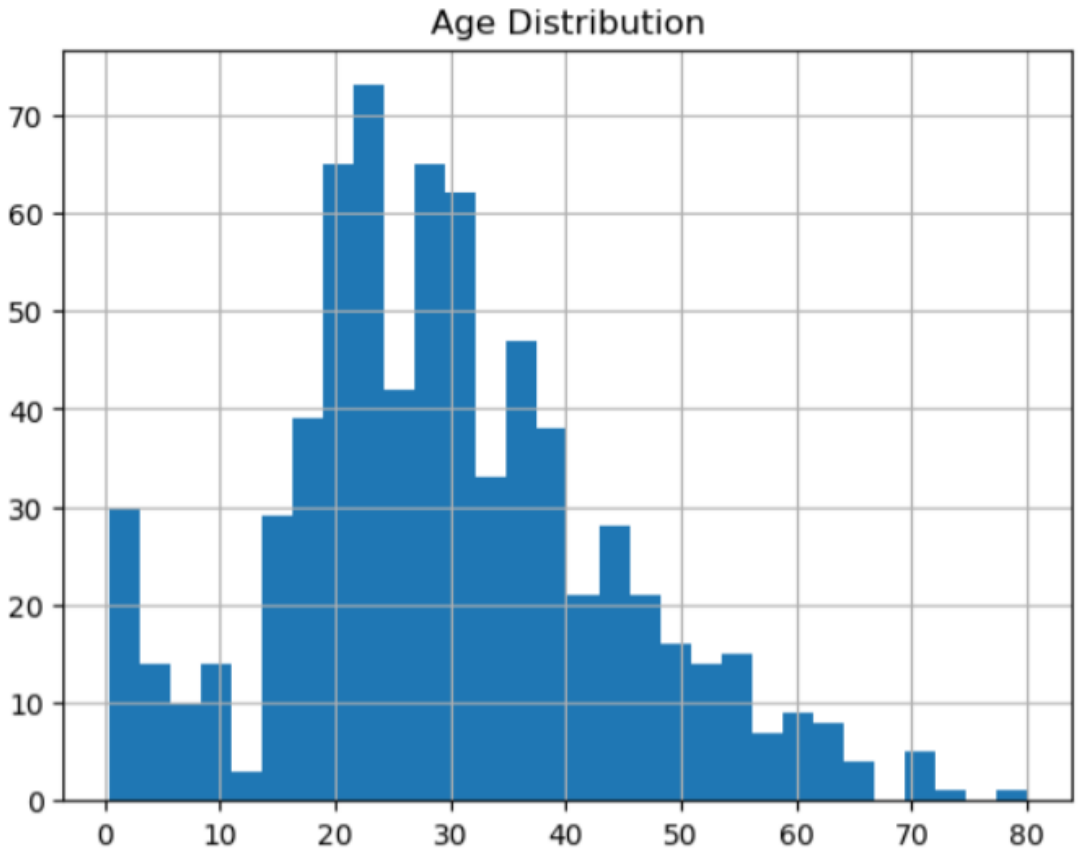
Univariate Analysis – Categorical Features

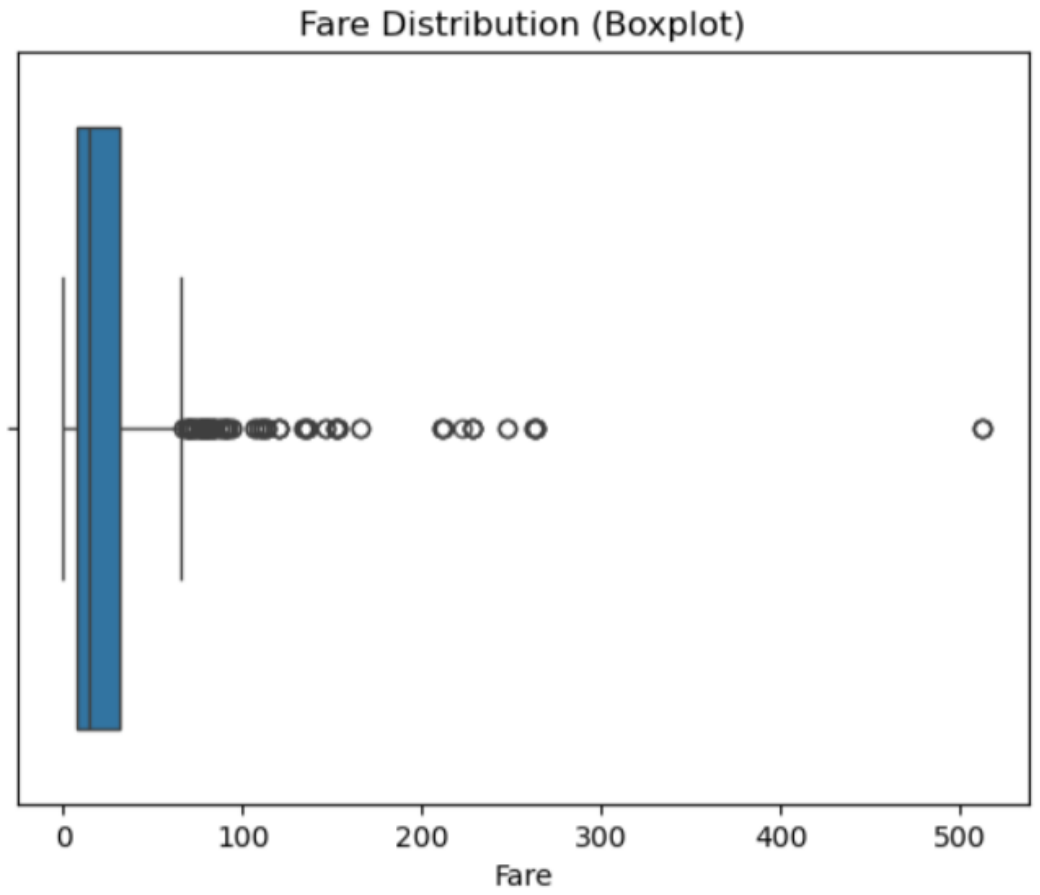




- SibSp (Siblings/Spouses Aboard) - Most passengers had no siblings/spouses aboard.
- Parch (Parents/Children Aboard) - Similar to SibSp, most had no parents/children aboard.

Univariate Analysis – Numerical Features





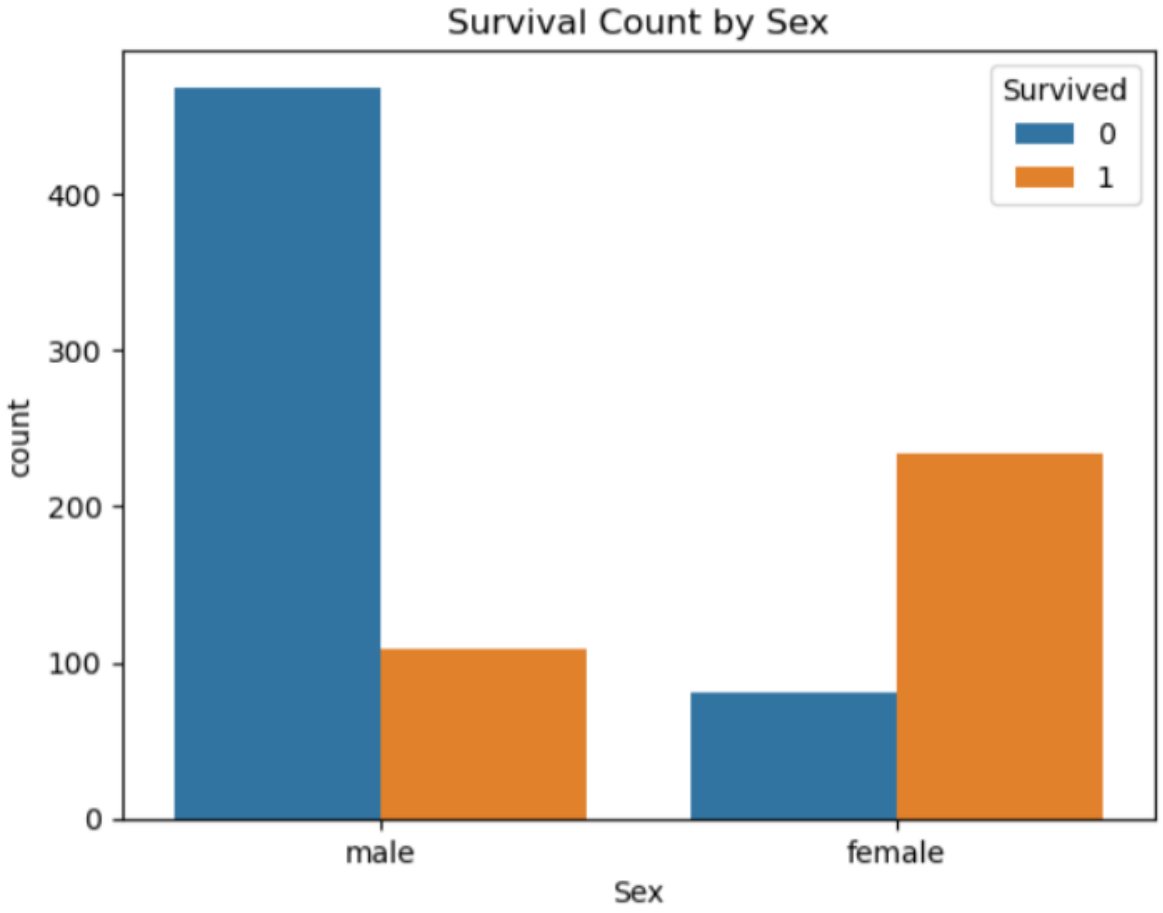
- Most passengers were in the 15–40 years age range.
- Majority of fares were low-priced, most likely for 3rd class passengers. A few outliers paid significantly higher fares, likely 1st class.

Bivariate Analysis

Bivariate Analysis

Survival Rate by Sex

Survived	0	1
Sex		
female	81	233
male	468	109

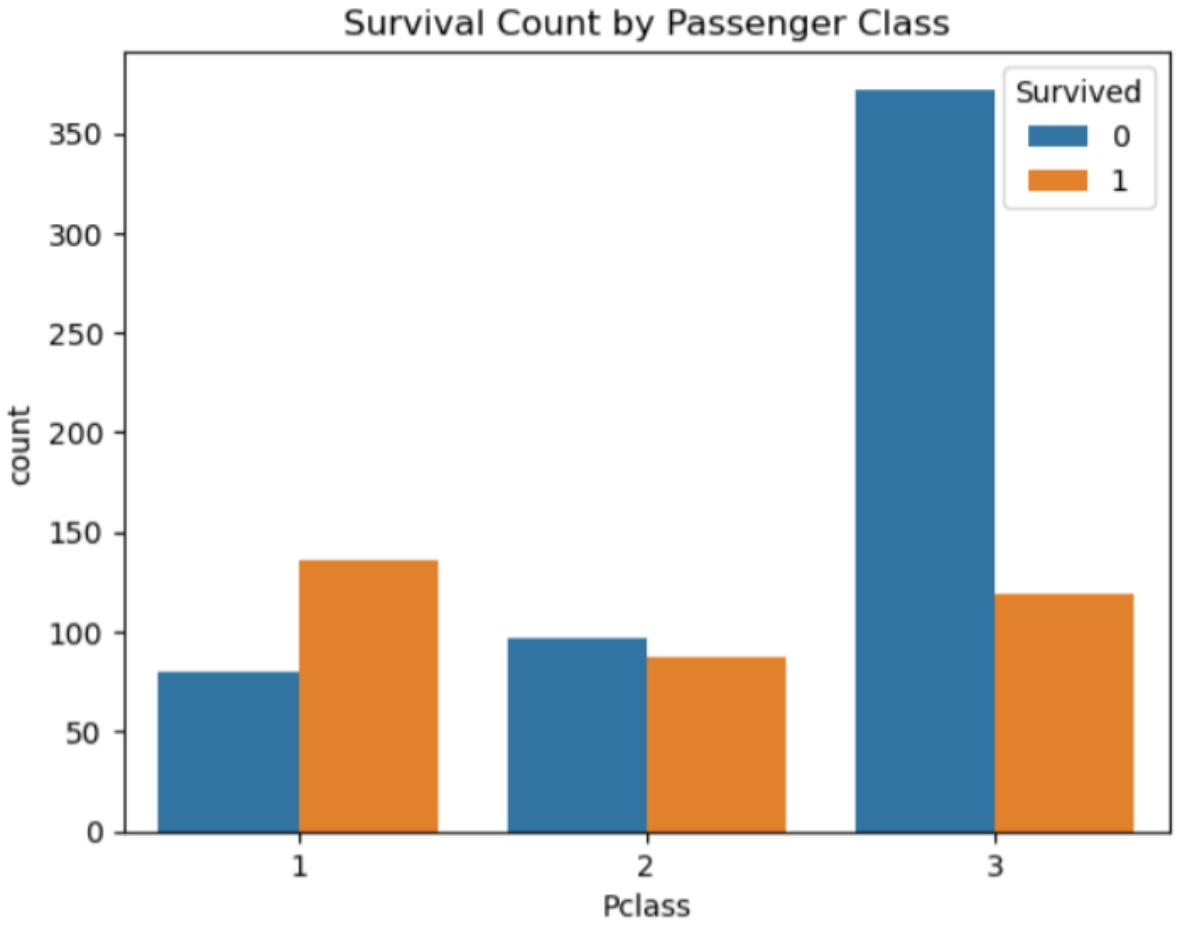


- Females had a much higher survival rate than males.
- Most male passengers did not survive, indicating a possible “women and children first” evacuation policy.

Bivariate Analysis

Survival Rate by Class

Survived	0	1
Pclass		
1	80	136
2	97	87
3	372	119

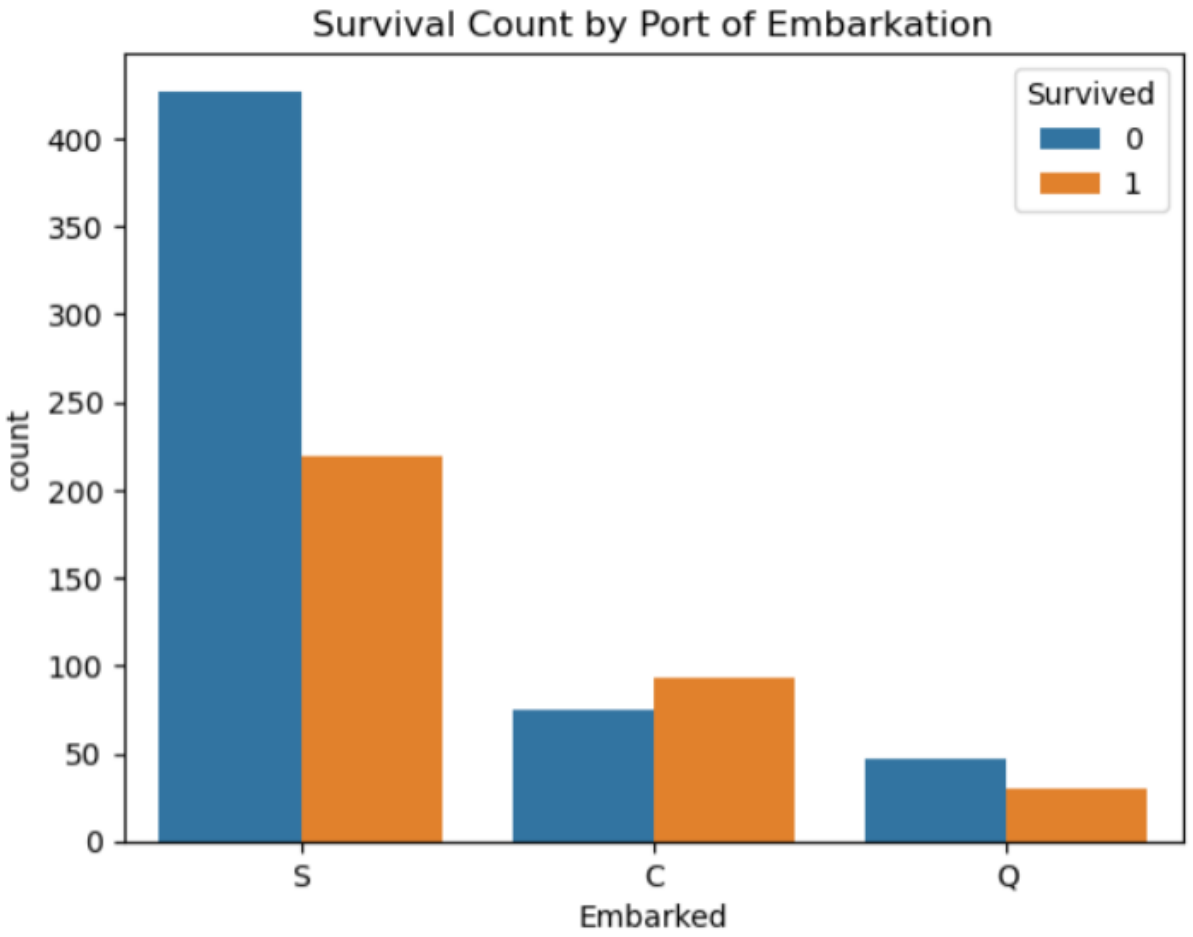


- Passengers in 1st class had the highest survival rate.
- Those in 3rd class were least likely to survive, showing a clear socioeconomic divide in survival.

Bivariate Analysis

Survival vs Embarked

Survived	0	1
Embarked		
C	75	93
Q	47	30
S	427	219

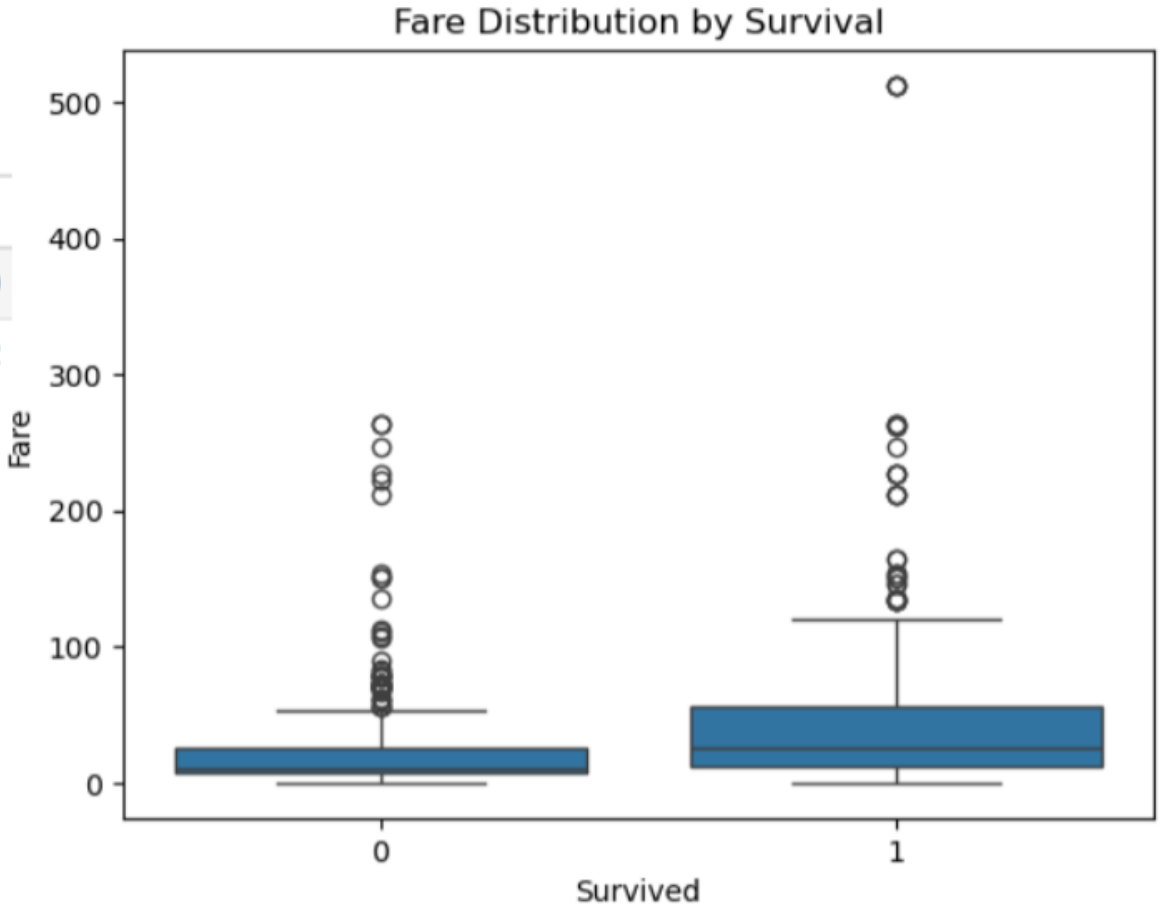


- Passengers who embarked from port ‘C’ (Cherbourg) had a higher survival rate.
- Those from ‘S’ (Southampton) had the lowest.

Bivariate Analysis

Survival vs Fare

Survived	count	mean	std	min	25%	50%	75%	max
0	549.0	22.117887	31.388207	0.0	7.8542	10.5	26.0	263.0000
1	342.0	48.395408	66.596998	0.0	12.4750	26.0	57.0	512.3292

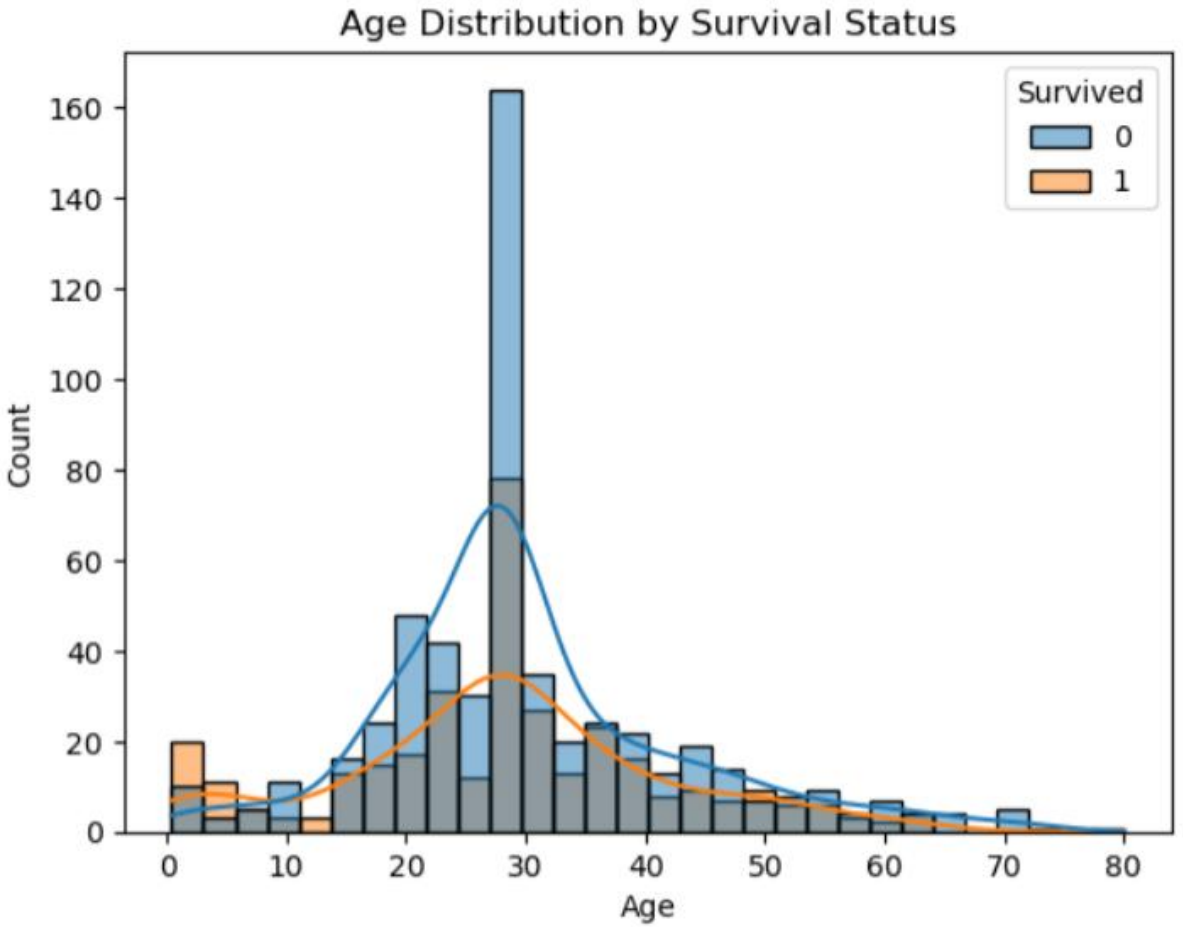


- Survivors generally paid higher fares, likely indicating 1st class travel.
- Fare distribution is skewed, with some very high-paying passengers mostly surviving.

Bivariate Analysis

Survival vs Age

Survived	count	mean	std	min	25%	50%	75%	max
0	549.0	30.028233	12.499986	1.00	23.0	28.0	35.0	74.0
1	342.0	28.291433	13.764425	0.42	21.0	28.0	35.0	80.0

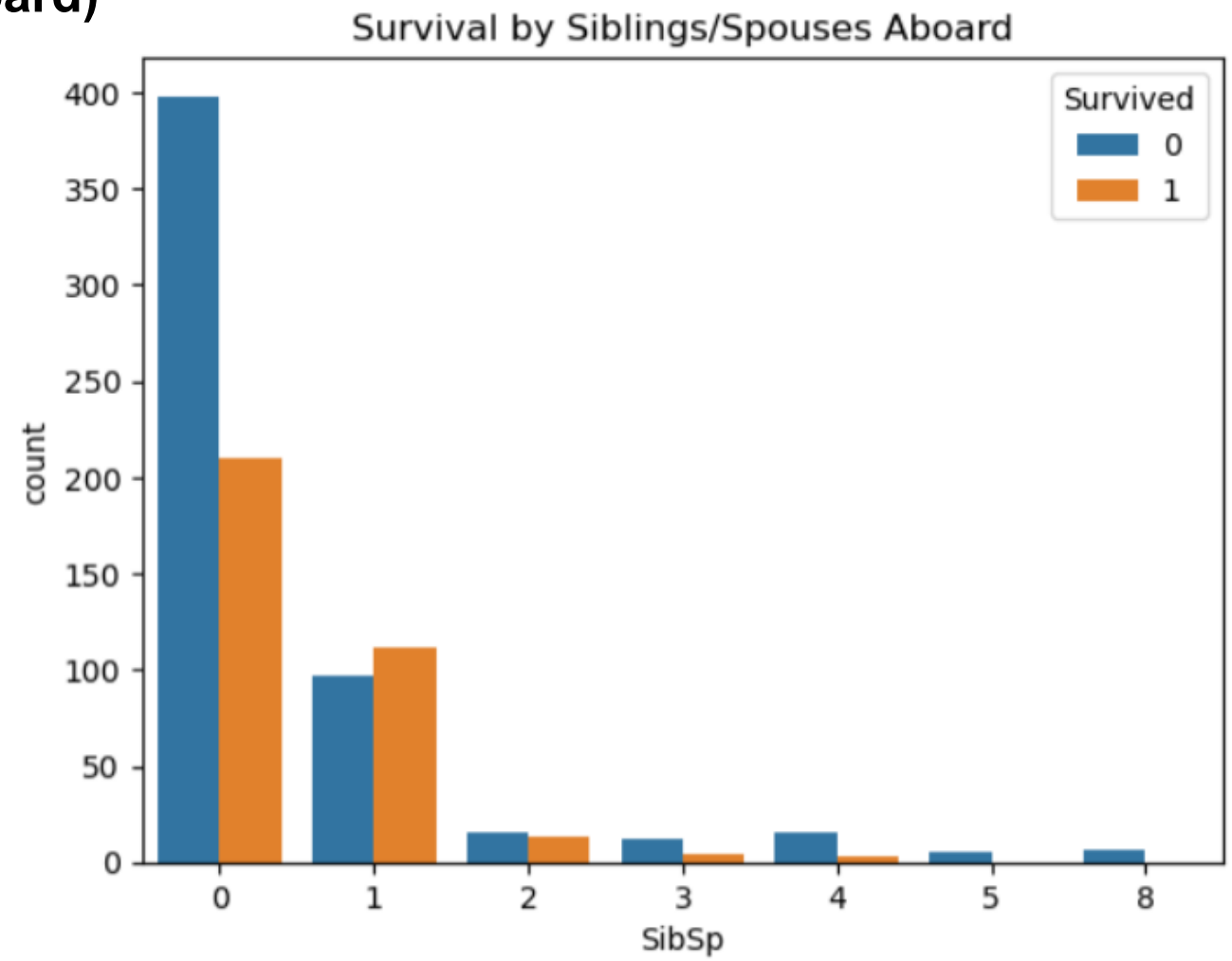


- Children (especially <10 years old) had higher survival rates.
- Elderly passengers had a lower survival chance overall.
- Many adult males in the age group 20–40 did not survive.

Bivariate Analysis

Survival vs SibSp (Siblings/Spouses Aboard)

Survived	0	1
SibSp		
0	398.0	210.0
1	97.0	112.0
2	15.0	13.0
3	12.0	4.0
4	15.0	3.0
5	5.0	NaN
8	7.0	NaN

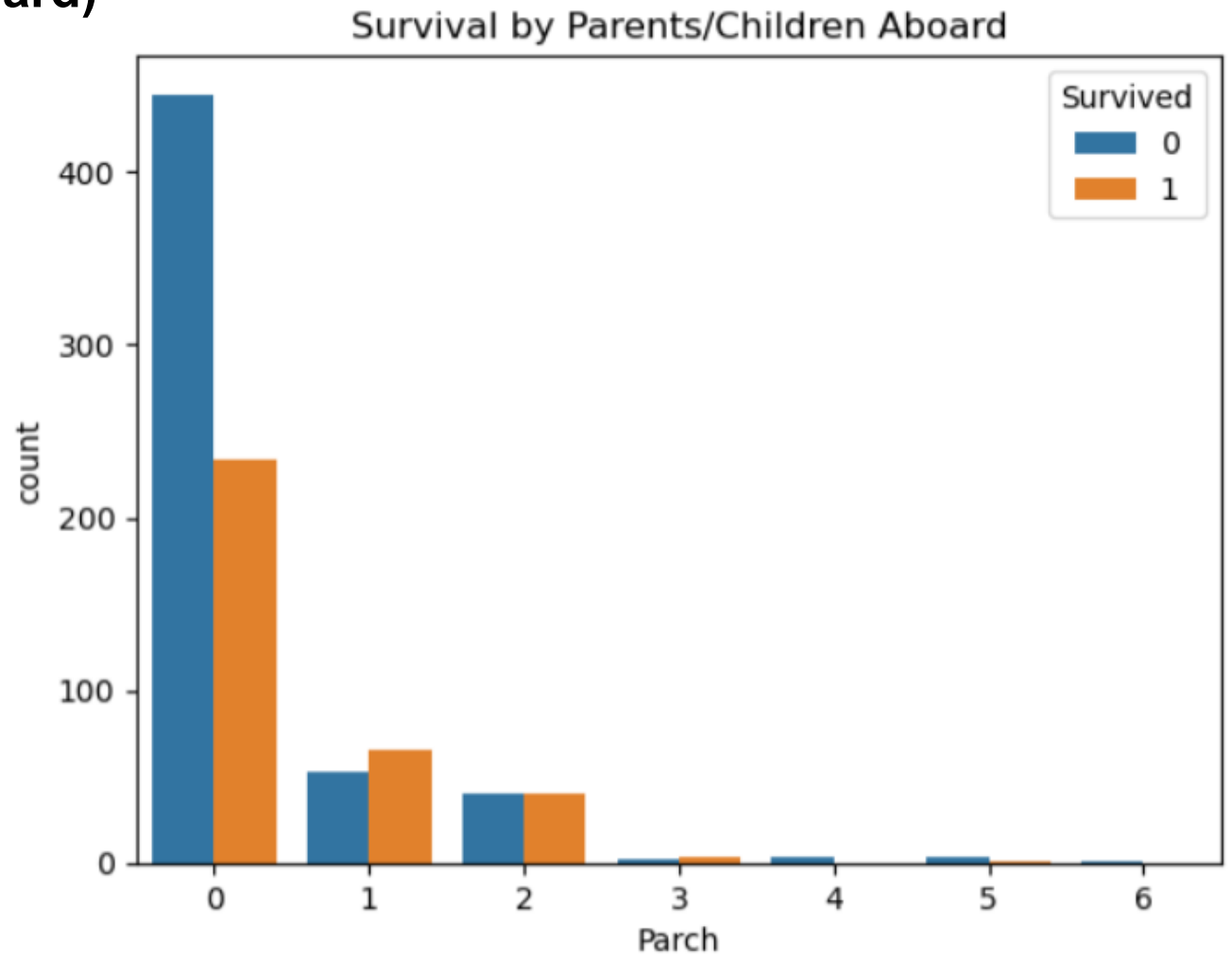


- Passengers with 1–2 family members had better survival chances.
- Those traveling alone or with large families (≥3) had lower survival.

Bivariate Analysis

Survival vs Parch (Parents/Children Aboard)

Survived	0	1
Parch		
0	445.0	233.0
1	53.0	65.0
2	40.0	40.0
3	2.0	3.0
4	4.0	NaN
5	4.0	1.0
6	1.0	NaN



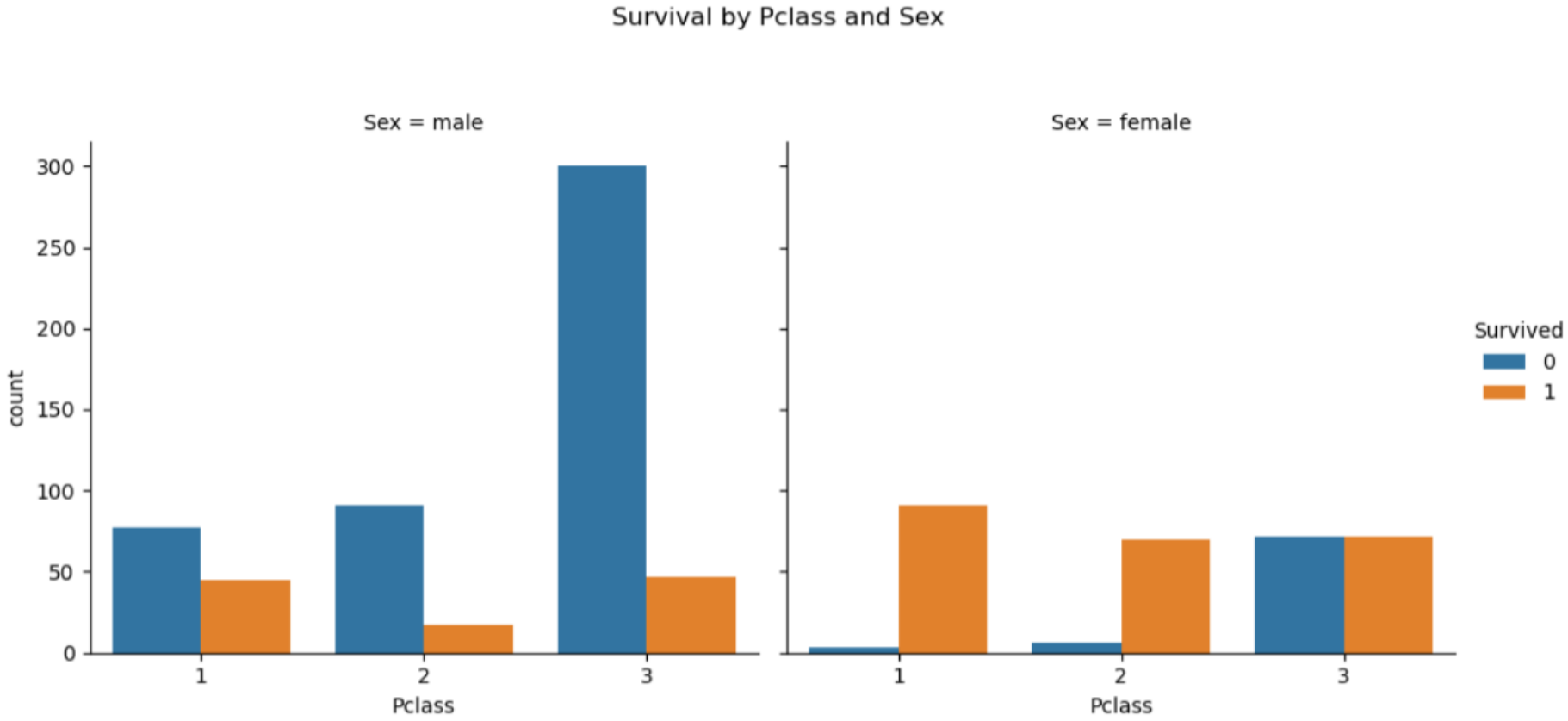
- Similar to SibSp: Having 1–2 parents/children with you increased survival odds.
- Very large family groups had poor survival rates.

Multivariate Analysis and Pairplot()

Multivariate Analysis

Sex + Pclass + Survival

	Survived	0	1
Sex	Pclass		
female	1	3	91
	2	6	70
	3	72	72
male	1	77	45
	2	91	17
	3	300	47

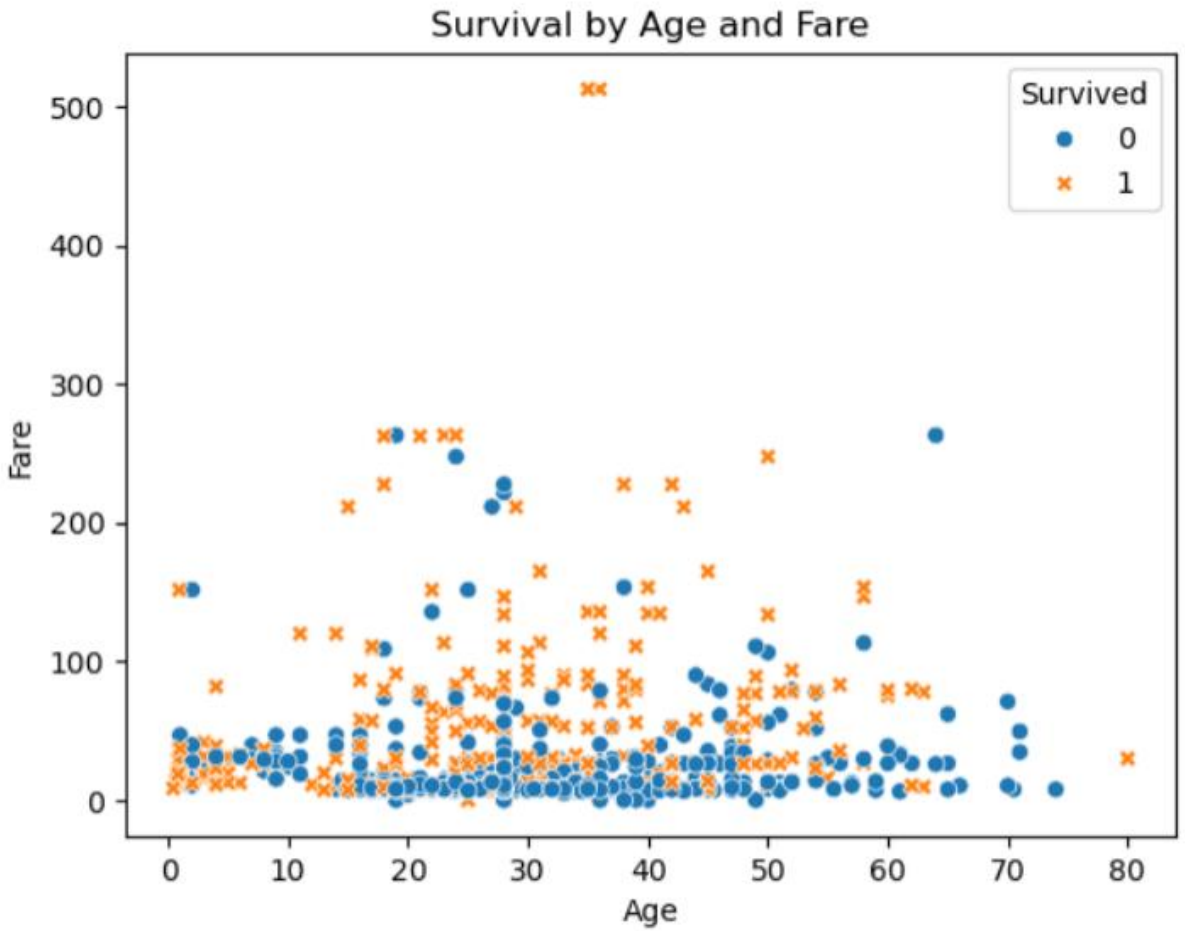


- Females in 1st and 2nd class had very high survival rates (close to 100% in 1st).
- Males in 3rd class had the lowest survival rates — very few survived.
- Survival probability depends heavily on a combination of gender and class.

Multivariate Analysis

Age + Fare + Survival

Survived	Age	Fare
0	30.028233	22.117887
1	28.291433	48.395408

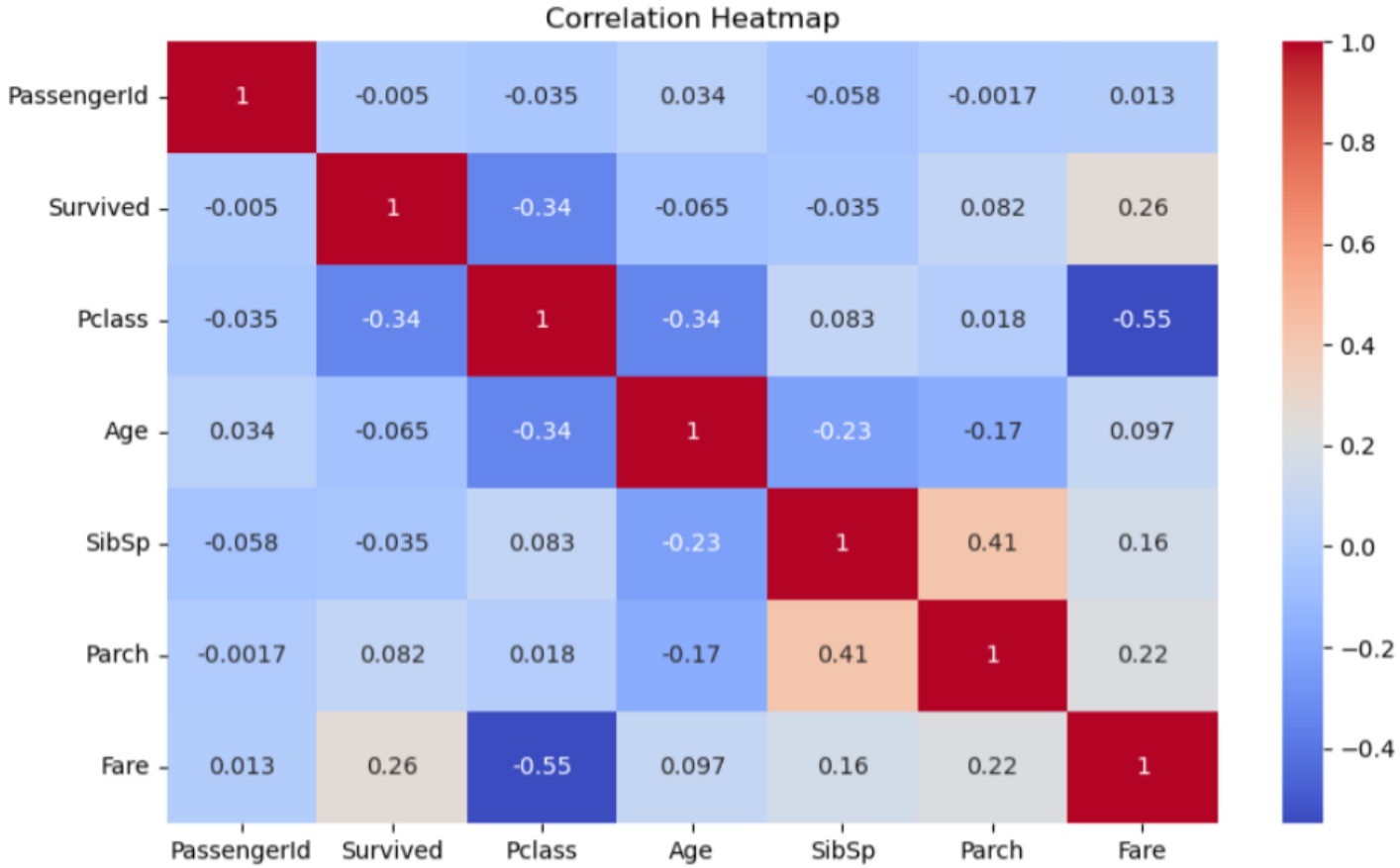


- Survivors on average were younger and paid more.
- Non-survivors were adults and paid lower fares (more common in 3rd class).

Multivariate Analysis

Correlation Table (for numerical features)

	Survived	Age	Fare	Parch	SibSp
Survived	1.000000	-0.064910	0.257307	0.081629	-0.035322
Age	-0.064910	1.000000	0.096688	-0.172482	-0.233296
Fare	0.257307	0.096688	1.000000	0.216225	0.159651
Parch	0.081629	-0.172482	0.216225	1.000000	0.414838
SibSp	-0.035322	-0.233296	0.159651	0.414838	1.000000

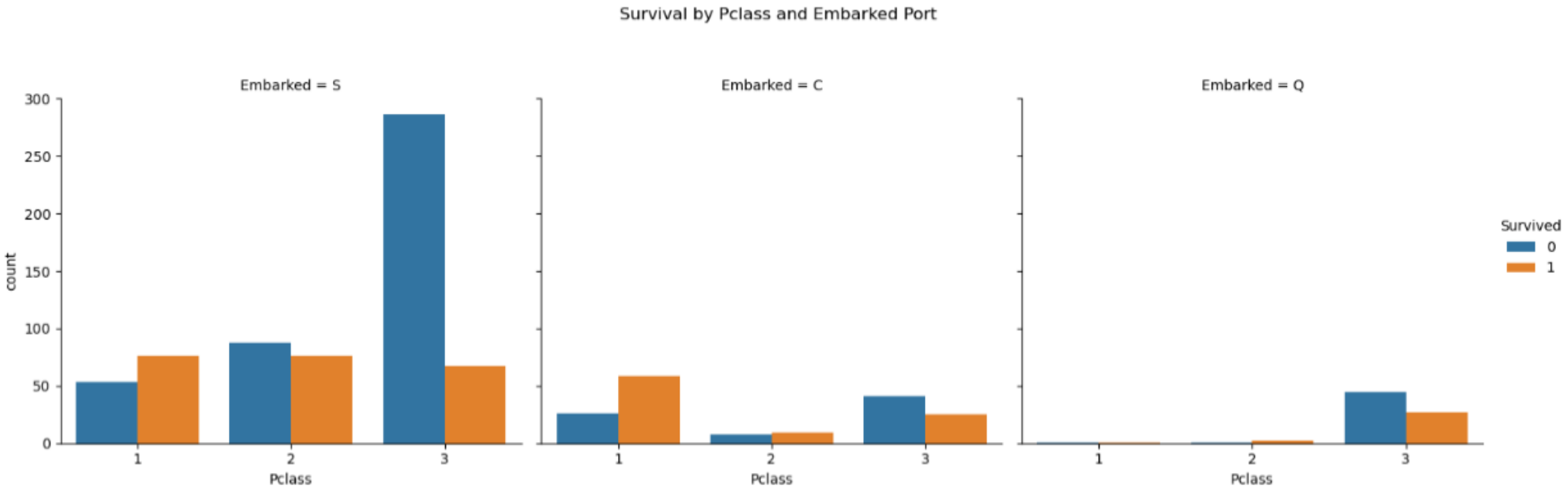


- Fare and Pclass are negatively correlated (higher class = higher fare).
- Survived is moderately correlated with Pclass (higher class = more survival).
- SibSp and Parch are slightly positively correlated, suggesting families traveled together.

Multivariate Analysis

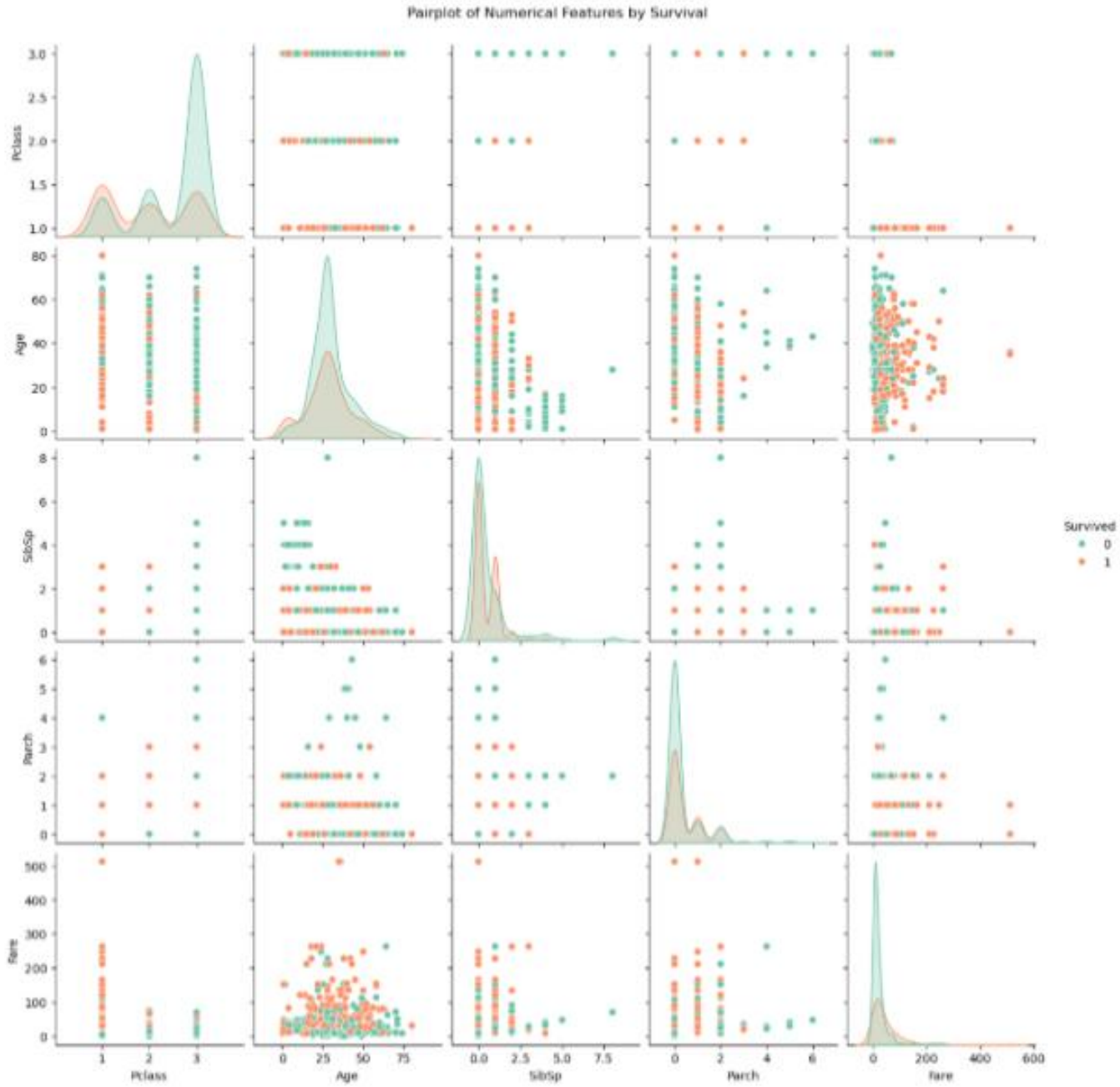
Embarked + Pclass + Survival

	Survived	0	1
Embarked	Pclass		
C	1	26	59
	2	8	9
	3	41	25
Q	1	1	1
	2	1	2
	3	45	27
S	1	53	76
	2	88	76
	3	286	67



- 1st class passengers from Cherbourg had the highest survival rate.
- Most 3rd class passengers from Southampton did not survive.

Pairplot()


Pairplot of Numerical Features by Survival
Pclass
Age
SibSp
Parch
Fare
Survived
0
1

Summary

Summary

- Being female, younger, and in a higher class greatly increased survival chances.
- Survival was most strongly influenced by ticket class (Pclass) and fare paid.
- Family presence (1–2 members) improved outcomes, but large families fared worse.
- Port of embarkation had an influence, likely due to class distribution across ports.
- Strong interplay between socioeconomic status, age, and gender in survival outcomes.

Thank You