In [ ]:
```
1. Dataset Overview
Total Records: 891

Features: 12 (e.g., PassengerId, Survived, Pclass, Age, Fare, Sex, etc.)

Target Variable: Survived (0 = did not survive, 1 = survived)
```

In [ ]:
```
2. Data Summary
.info() showed:

Age has 177 missing values (714 non-null).

Cabin is mostly missing (only 204 non-null).

Embarked has 2 missing values.

.describe() statistics:

Mean Age: ~29.7 years

Fare ranged from 0 to 512, with a median around 14.45

Most passengers were in 3rd class (Pclass = 3)
```

In [ ]:
```
 3. Categorical Value Counts
Sex:

Male: 577

Female: 314

Embarked:

Southampton (S): 644

Cherbourg (C): 168

Queenstown (Q): 77

Cabins & Tickets:

Highly varied, many unique entries

Top repeated ticket: 347082 (7 passengers)
```

In [ ]:
```
4. Visualizations
a. Histograms
Variables like Age, Fare, SibSp, and Parch are right-skewed.

Some fares were very high outliers.

b. Boxplots
Fare had extreme outliers in each passenger class.

Median fare is highest for 1st class, lowest for 3rd class.

c. Pairplot
Visual correlations were explored between variables like Age, Fare, Pclass, ar

Overlap between classes visible, but survivors were more common among younger

d. Correlation Heatmap
Moderate negative correlation between Pclass and Survived (~-0.34)

Positive correlation between Fare and Survived (~0.26)

Weak or no strong linear relationships among many variables

e. Scatter Plot (Age vs. Fare)
Most passengers were clustered in the low-fare, 20–40 age range.

Some high-fare passengers spanned various ages.
```

In [ ]:
```
Key Findings
Survival Trends:

Women and higher-class passengers were more likely to survive.

Survival drops in 3rd class dramatically.

Fare & Class:

1st class passengers paid more and had a higher survival rate.

Outliers in Fare indicate luxury accommodations.

Missing Data:

Consider imputing or removing rows with missing Age, Embarked, or dropping Cab

Feature Engineering Potential:

Split Name for title extraction (e.g., Mr., Mrs., Miss)

Group Fare and Age into bins for categorical modeling
```