In [4]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("train.csv")
df.head()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500

In [5]:
```python
sns.set(style="whitegrid")
df.info()
df.describe()
df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   PassengerId  891 non-null     int64
 1   Survived     891 non-null     int64
 2   Pclass       891 non-null     int64
 3   Name         891 non-null     object
 4   Sex          891 non-null     object
 5   Age          714 non-null     float64
 6   SibSp        891 non-null     int64
 7   Parch        891 non-null     int64
 8   Ticket       891 non-null     object
 9   Fare         891 non-null     float64
 10  Cabin        204 non-null     object
 11  Embarked     889 non-null     object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

Out[5]:
```
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
```
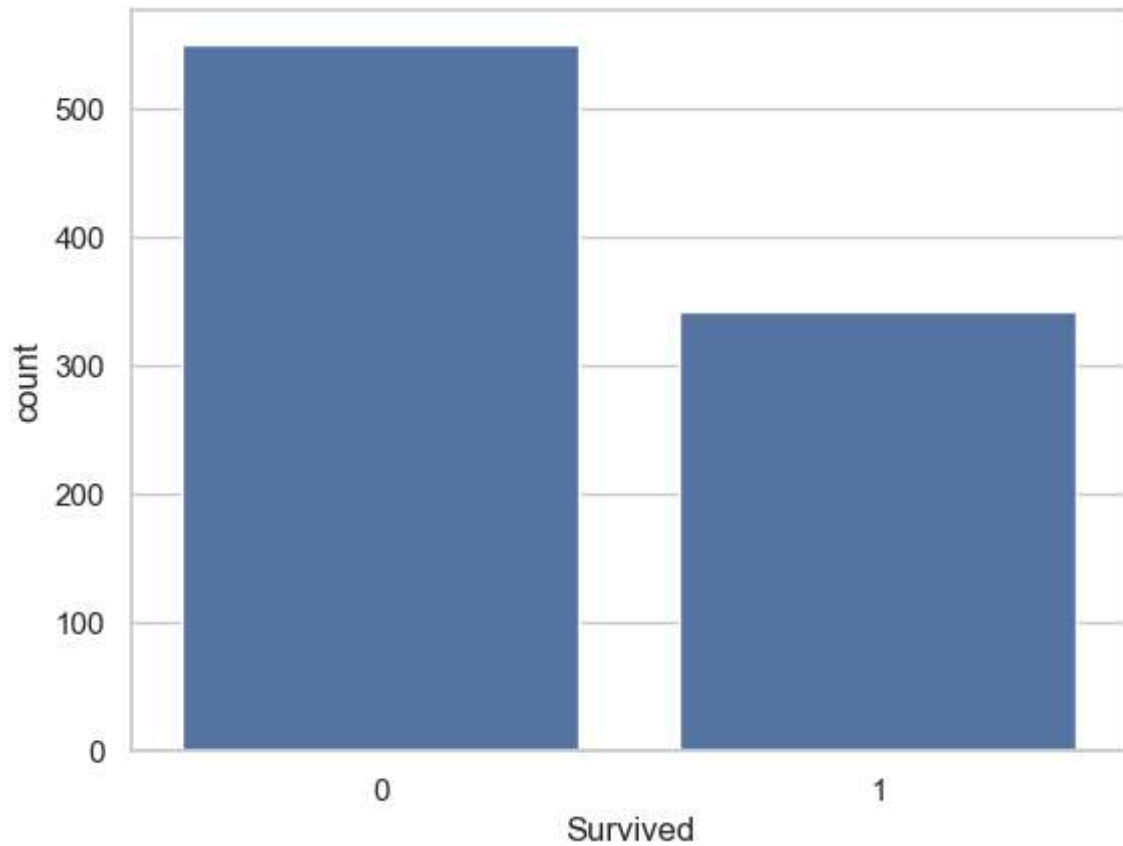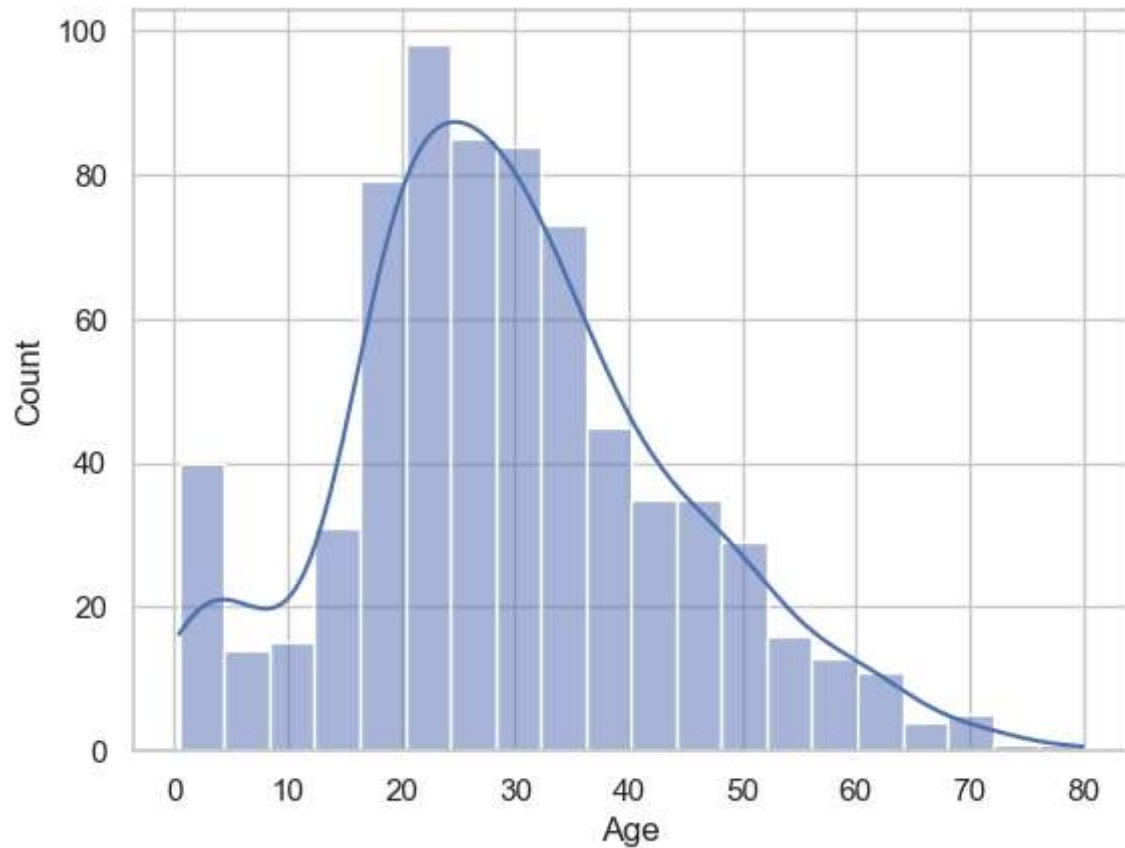
In [7]:
```python
sns.countplot(x="Survived", data=df)
plt.show()
```

The majority of passengers did not survive. Survivors are significantly fewer compared to non-survivors, indicating that the Titanic disaster had a high fatality rate.

```
In [8]:  sns.histplot(df["Age"], kde=True)
         plt.show()
```
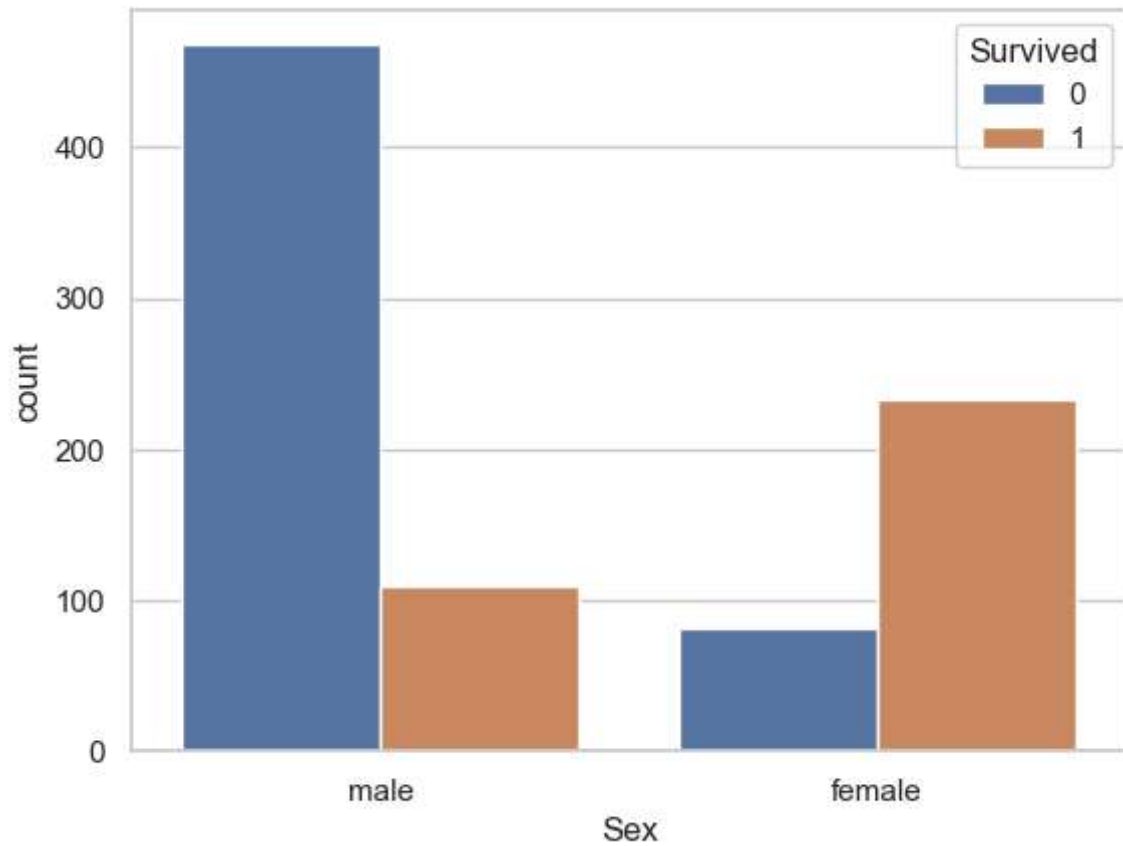
Most passengers were between 20–40 years old. There are fewer passengers in the older age groups, and some missing values in the Age column.

```
In [9]:  sns.histplot(df["Fare"], kde=True)
         plt.show()
```
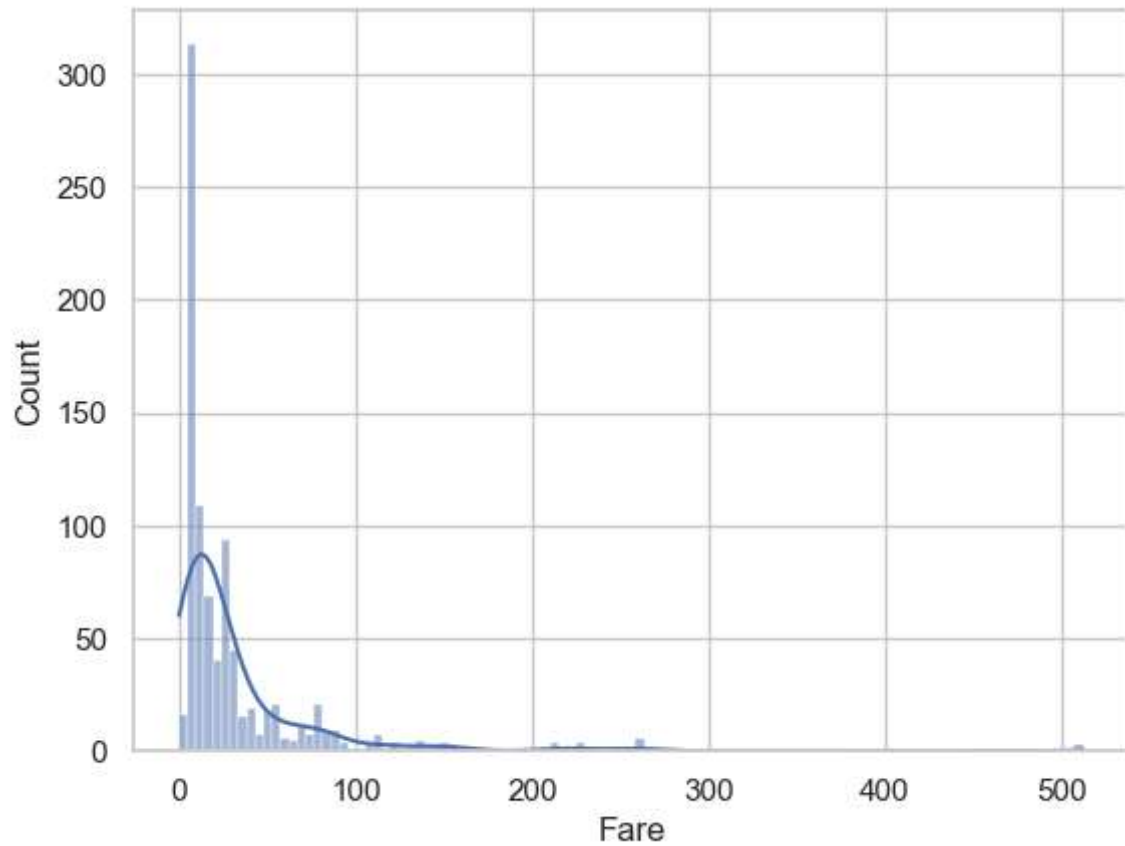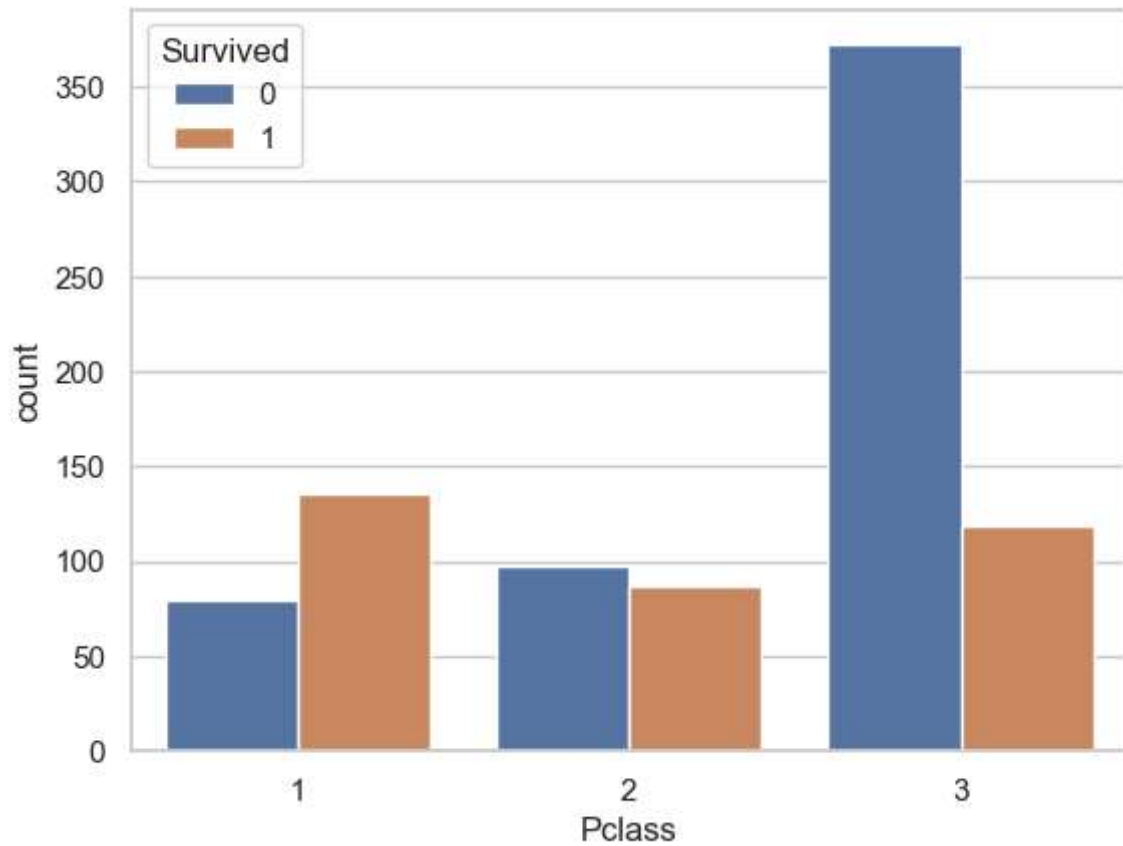
Most fares are concentrated between 0 and 100, with a few very high outliers. Higher fares often indicate first-class tickets.

```
In [10]:   sns.countplot(x="Sex", hue="Survived", data=df)
           plt.show()
```

Females had a much higher survival rate compared to males. This supports the "women and children first" rescue policy.

In [11]:
```
sns.countplot(x="Pclass", hue="Survived", data=df)
plt.show()
```

First-class passengers had the highest survival rate, followed by second class. Third-class passengers had the lowest survival rate.

```
In [12]:  sns.scatterplot(x="Age", y="Fare", hue="Survived", data=df)
          plt.show()
```
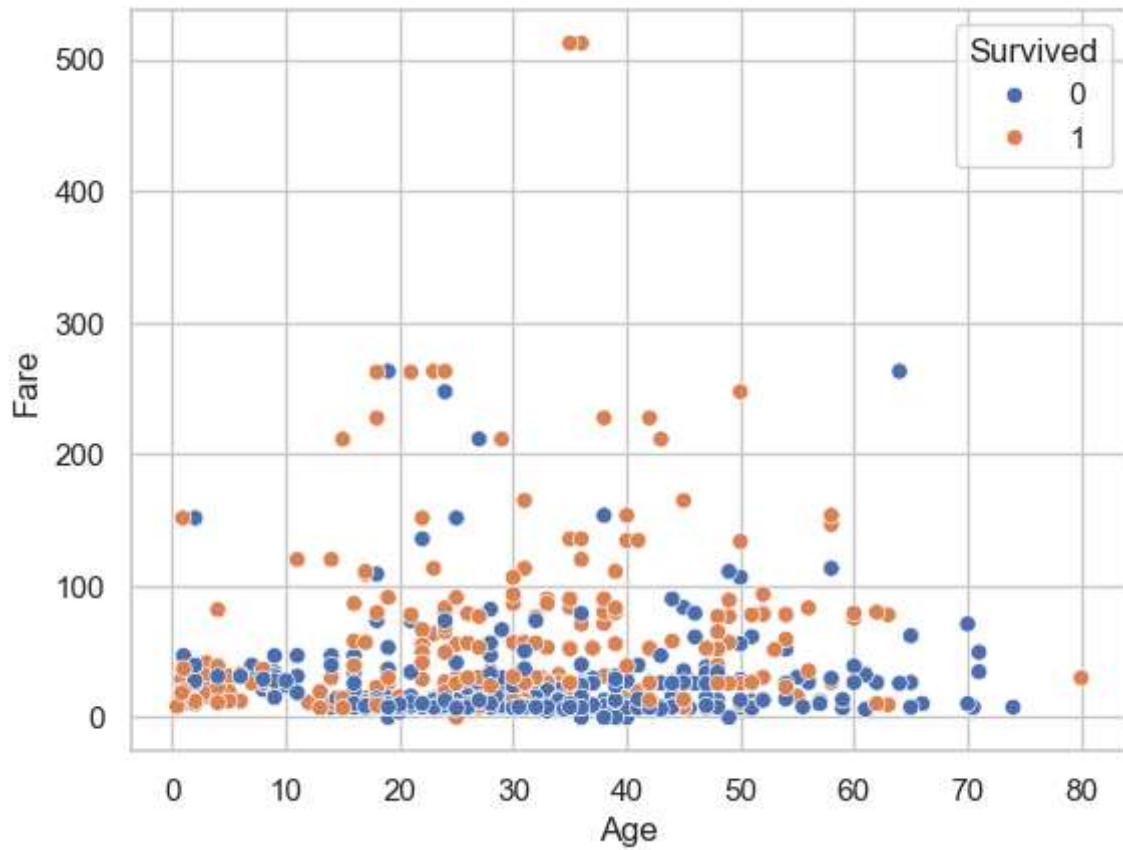
Passengers who paid higher fares (often first-class) had a better chance of survival. Younger passengers also show slightly better survival rates.

In [13]:
```python
sns.heatmap(df.select_dtypes(include=["number"]).corr(), annot=True, cmap="coolwarm
plt.show()
```
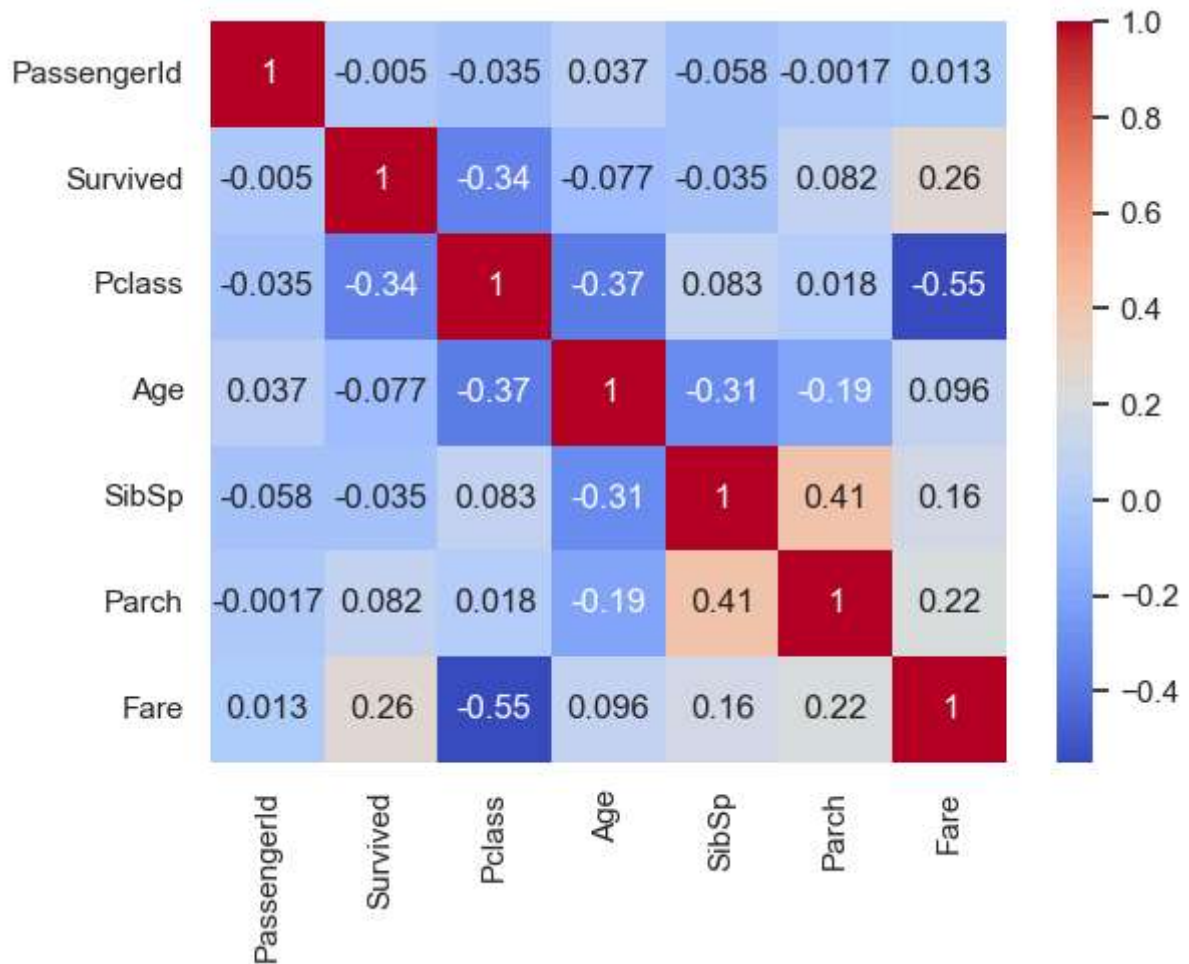
Survival is positively correlated with Fare and negatively correlated with Passenger Class (lower class number means higher class). Age shows a weaker correlation with survival.

Summary

- The majority of passengers did not survive the Titanic disaster.
- Females had a much higher survival rate than males.
- First-class passengers survived more often than those in lower classes.
- Younger passengers had slightly higher chances of survival.
- Higher fares were generally associated with better survival odds.
- The Age column has missing values that might require imputation in further analysis.
- Fare distribution shows a few extreme outliers.
- Correlation analysis confirms the importance of class and fare in survival chances.