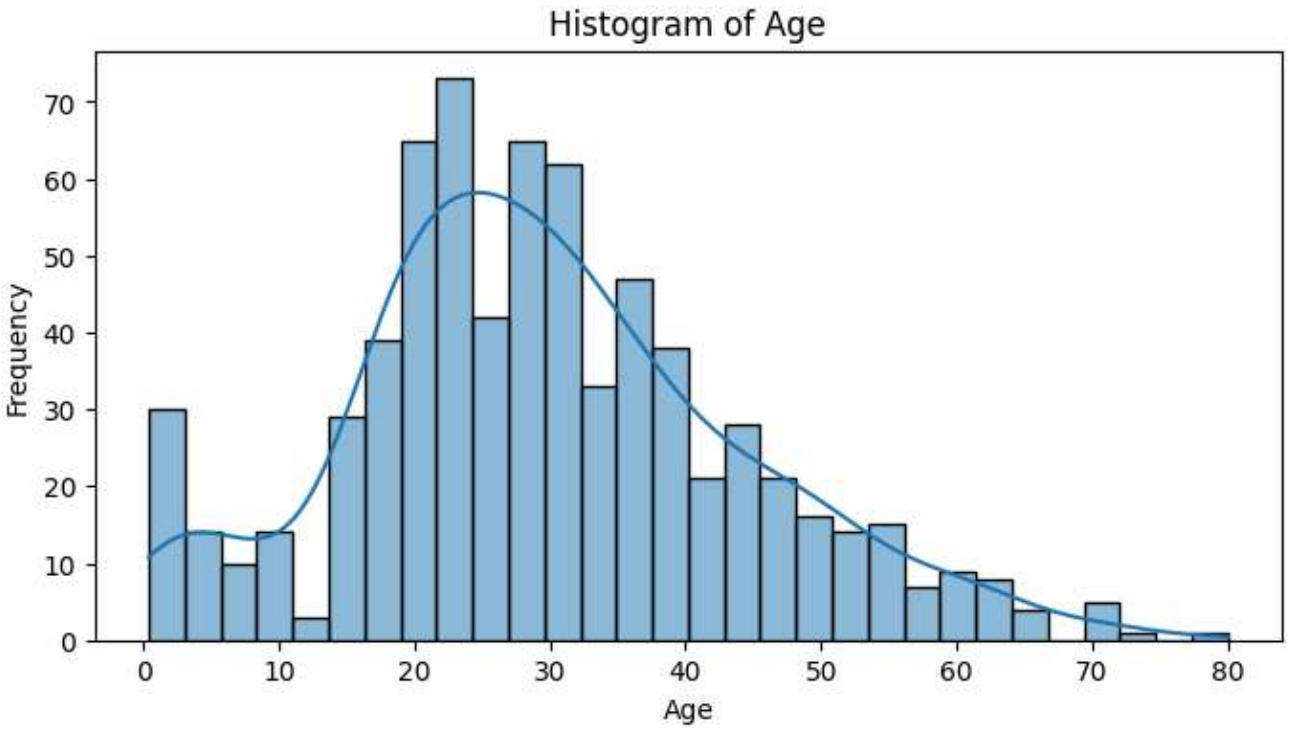
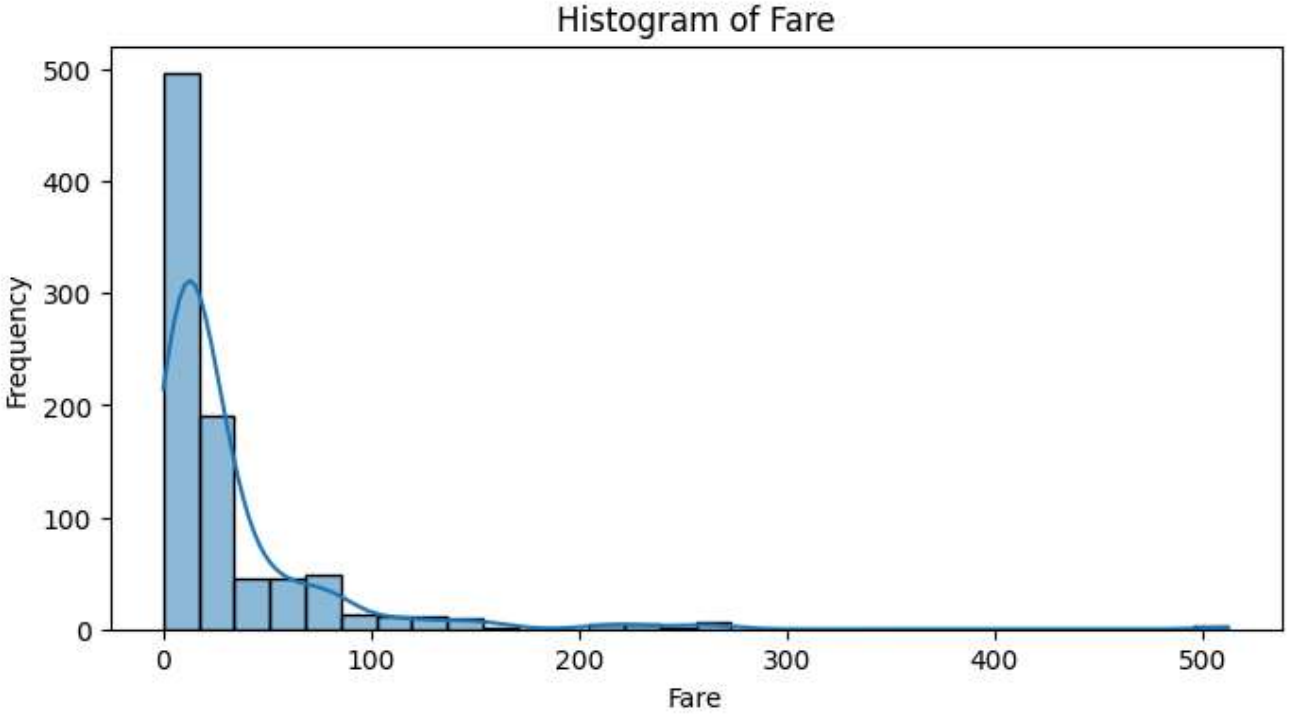
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px


file_path = '/content/Titanic-Dataset (1).csv'
titanic_data = pd.read_csv(file_path)

print("Summary Statistics:")
print(titanic_data.describe(include='all'))
```

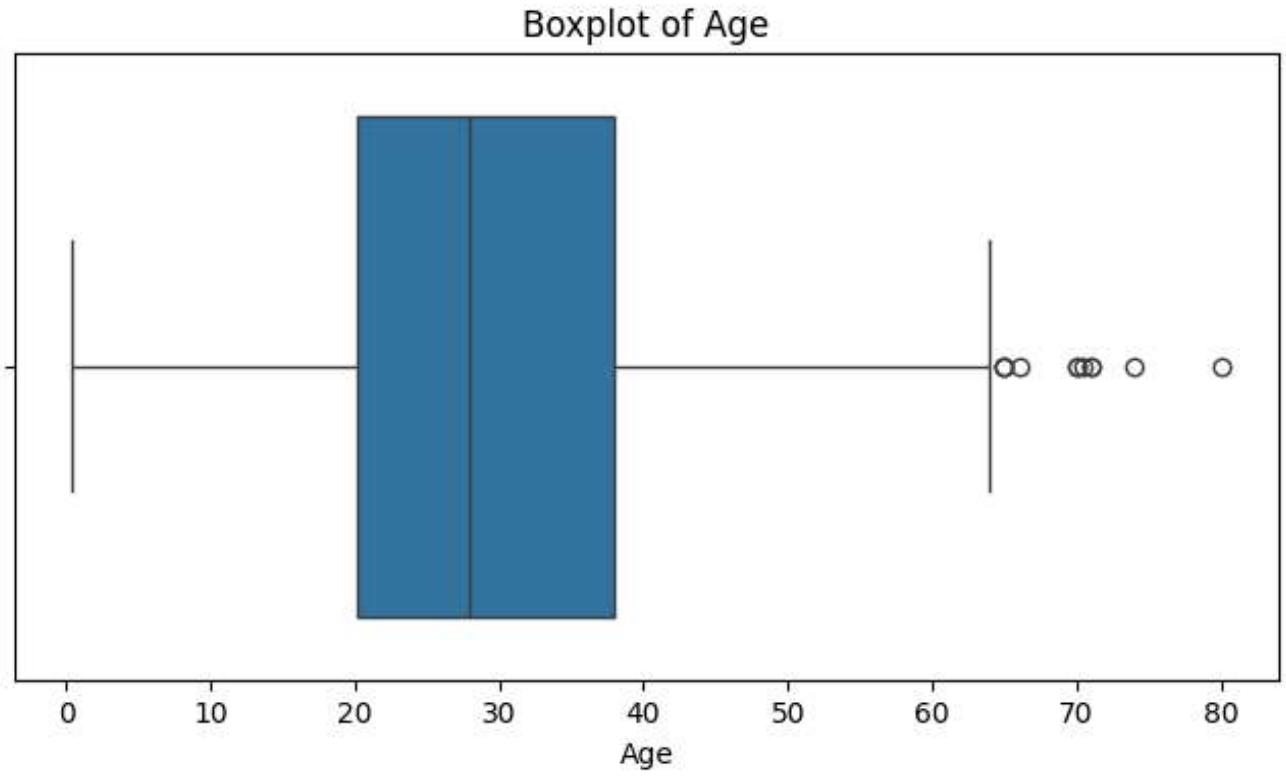
```
Summary Statistics:
        PassengerId    Survived      Pclass                 Name   S
count    891.000000  891.000000  891.000000                  891   8
unique          NaN         NaN         NaN                  891
top             NaN         NaN         NaN  Dooley, Mr. Patrick  ma
freq            NaN         NaN         NaN                    1   5
mean     446.000000    0.383838    2.308642                  NaN   N
std      257.353842    0.486592    0.836071                  NaN   N
min        1.000000    0.000000    1.000000                  NaN   N
25%      223.500000    0.000000    2.000000                  NaN   N
50%      446.000000    0.000000    3.000000                  NaN   N
75%      668.500000    1.000000    3.000000                  NaN   N
max      891.000000    1.000000    3.000000                  NaN   N

                Age       SibSp       Parch  Ticket        Fare Cabi
count   714.000000  891.000000  891.000000     891  891.000000   204
unique         NaN         NaN         NaN     681         NaN   147
top            NaN         NaN         NaN  347082         NaN    G6
freq           NaN         NaN         NaN       7         NaN     4
mean     29.699118    0.523008    0.381594     NaN   32.204208   NaN
std      14.526497    1.102743    0.806057     NaN   49.693429   NaN
min       0.420000    0.000000    0.000000     NaN    0.000000   NaN
25%      20.125000    0.000000    0.000000     NaN    7.910400   NaN
50%      28.000000    0.000000    0.000000     NaN   14.454200   NaN
75%      38.000000    1.000000    0.000000     NaN   31.000000   NaN
max      80.000000    8.000000    6.000000     NaN  512.329200   NaN
```

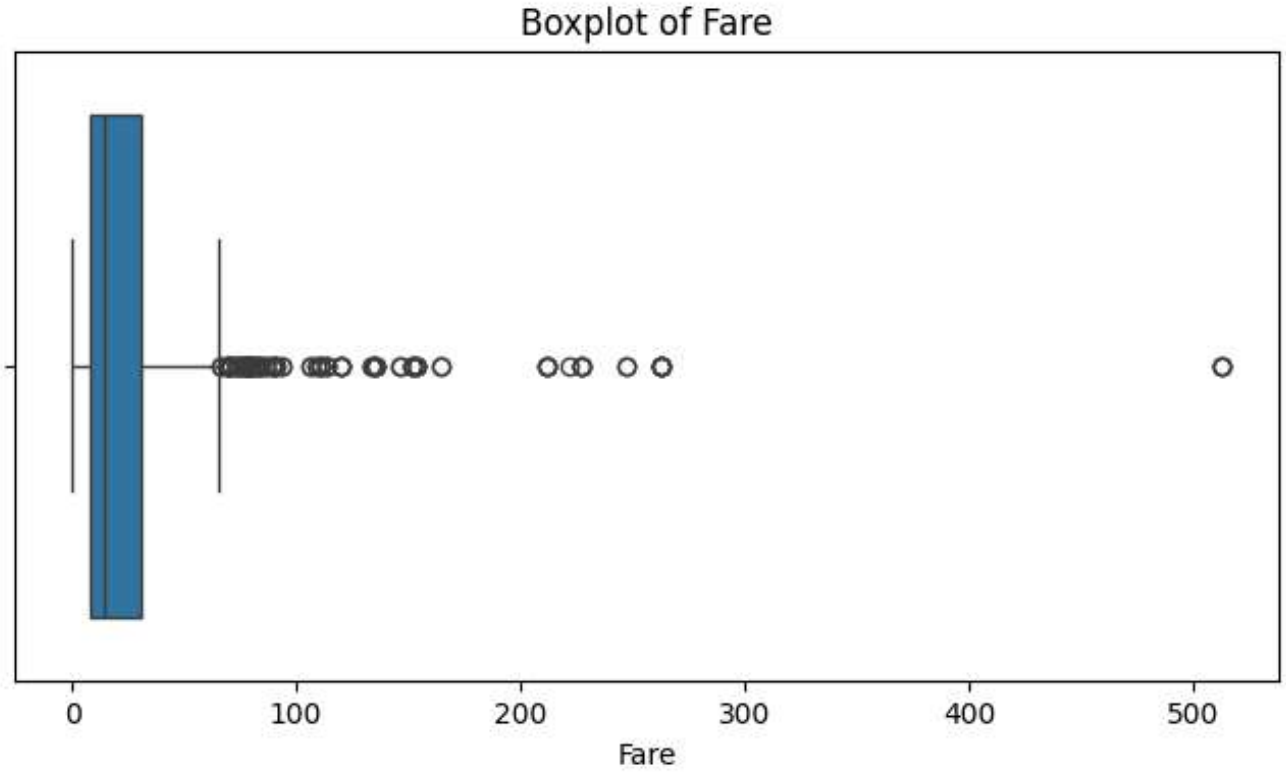
```
    numeric_features = ['Age', 'Fare']
    for feature in numeric_features:
        plt.figure(figsize=(8, 4))
        sns.histplot(titanic_data[feature], kde=True, bins=30)
        plt.title(f'Histogram of {feature}')
        plt.xlabel(feature)
        plt.ylabel('Frequency')
        plt.show()
```


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

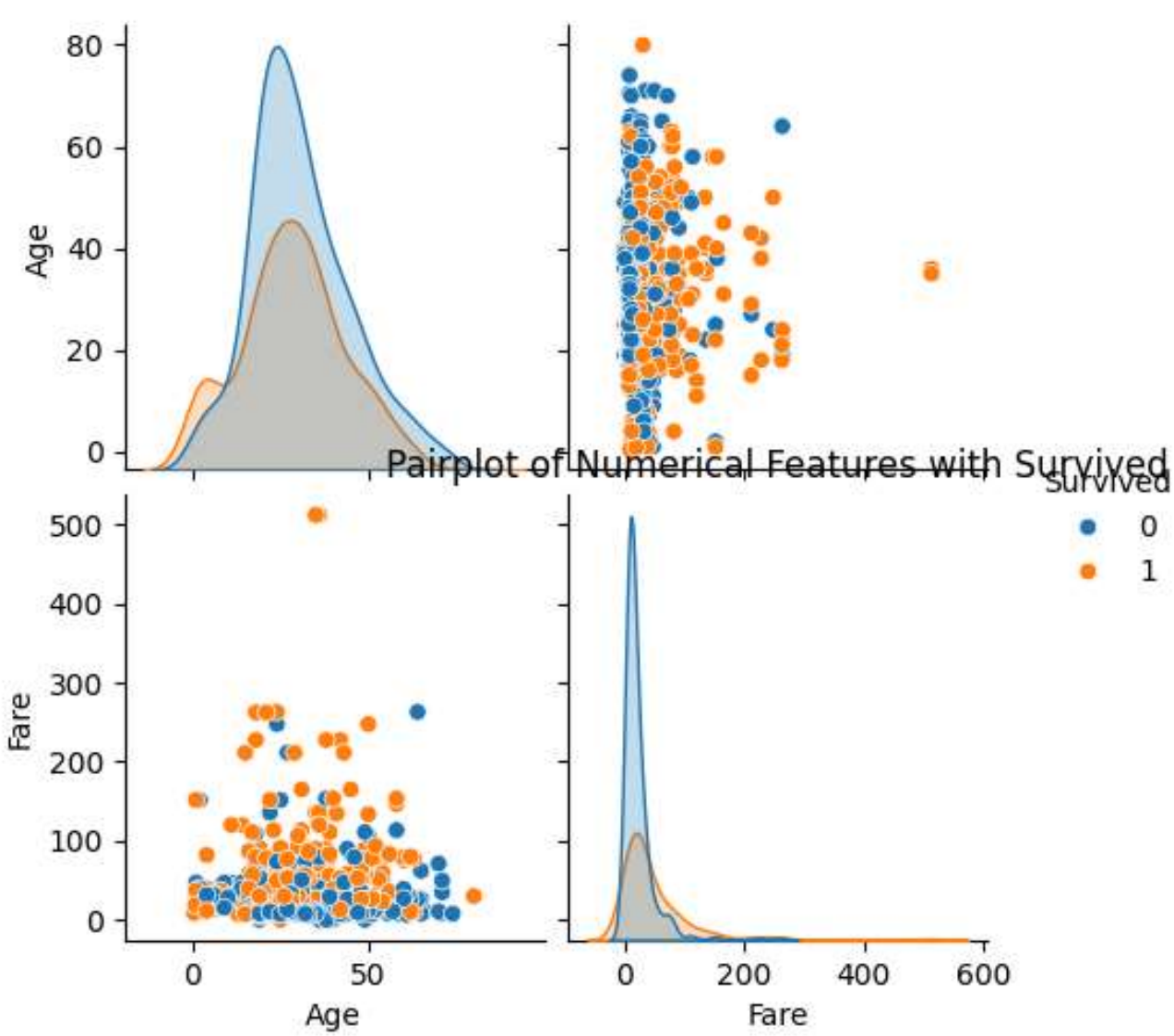

Histogram of Fare
Frequency
Fare
0
100
200
300
400
500

```
for feature in numeric_features:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=titanic_data[feature])
    plt.title(f'Boxplot of {feature}')
    plt.xlabel(feature)
    plt.show()
```


Boxplot of Age
0
10
20
30
40
50
60
70
80
Age

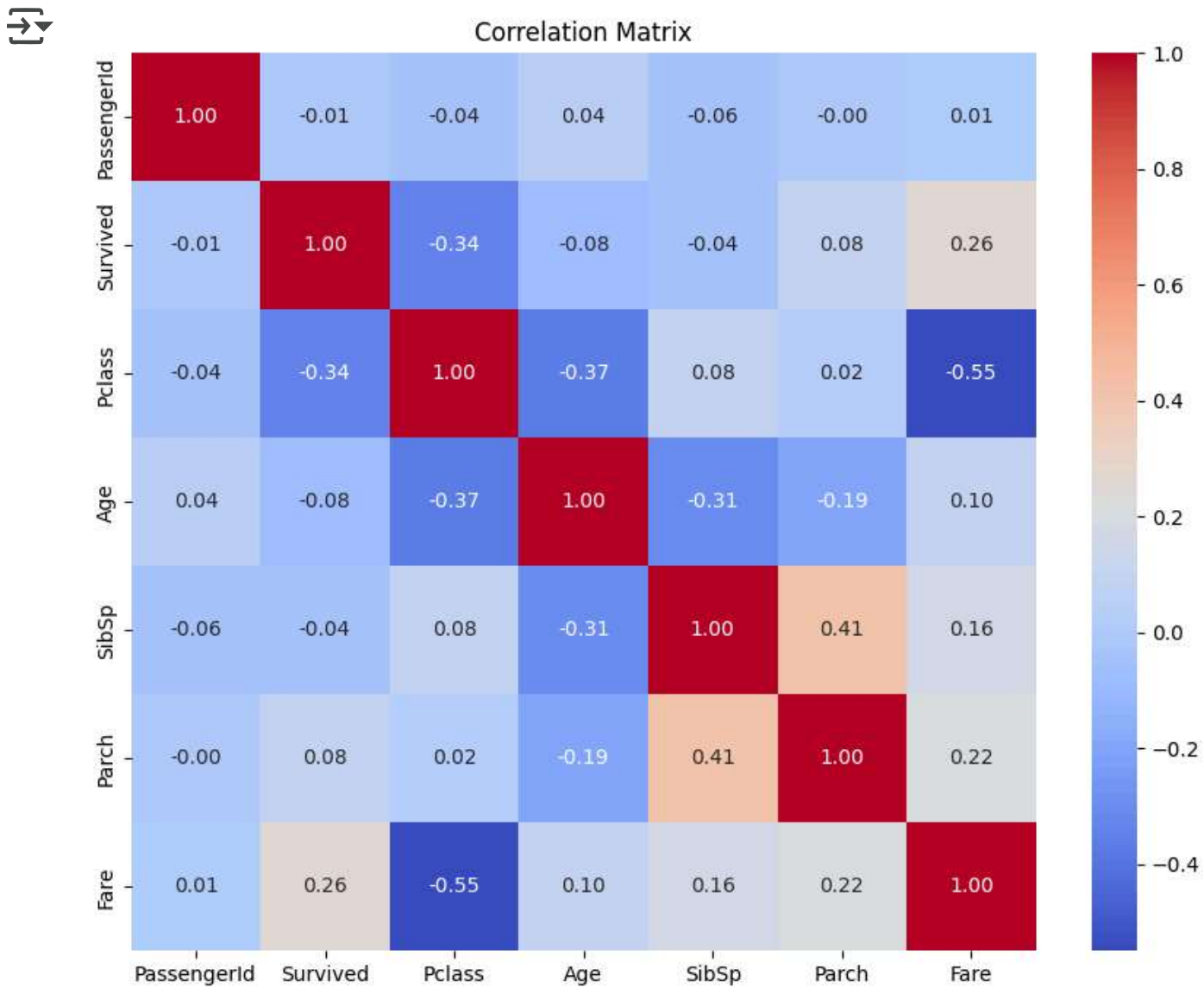

Boxplot of Fare
0
100
200
300
400
500
Fare

```
sns.pairplot(titanic_data, hue='Survived', vars=numeric_features)
plt.title('Pairplot of Numerical Features with Survived')
plt.show()
```

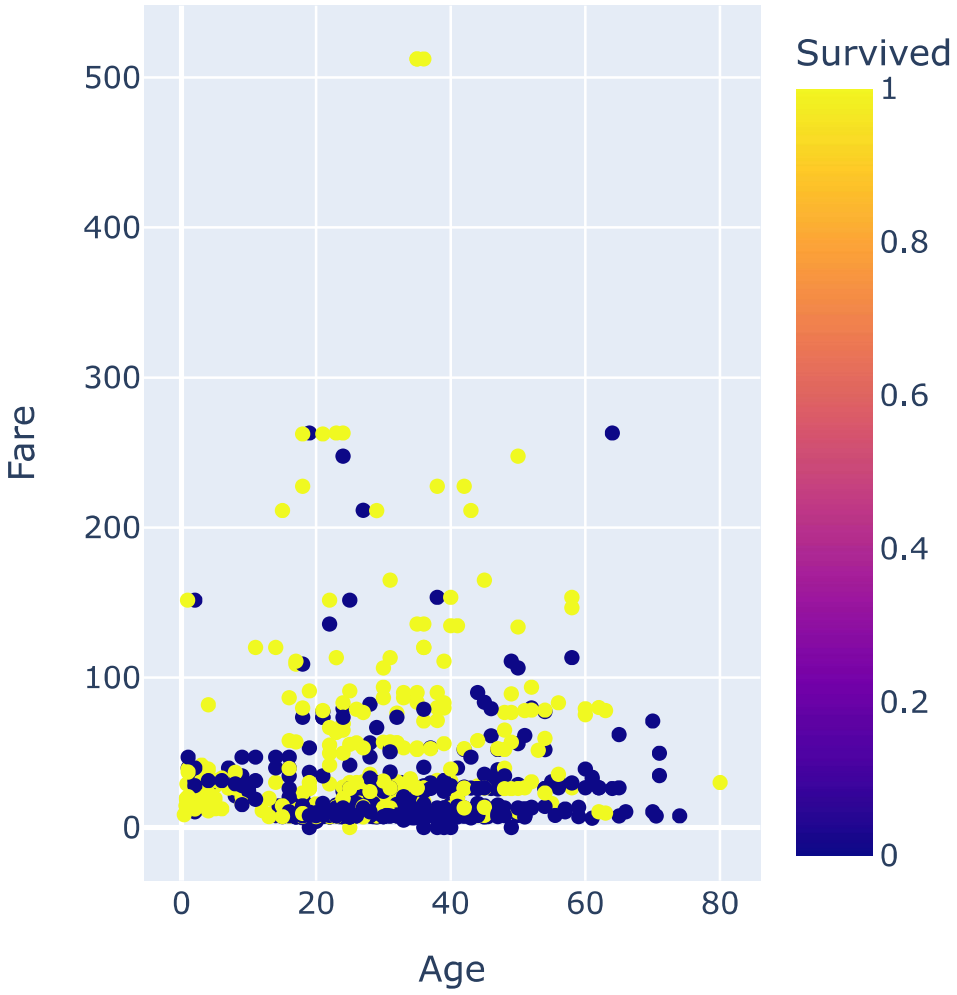



```
numeric_columns = titanic_data.select_dtypes(include=['number'])

correlation_matrix = numeric_columns.corr()

plt.figure(figsize=(10, 8))
```

```
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```



```
fig = px.scatter(titanic_data, x='Age', y='Fare', color='Survived', title
fig.show()
```




```
grouped_stats = titanic_data.groupby('Survived')[['Age', 'Fare']].mean(
print("Mean Age and Fare by Survival:")
print(grouped_stats)
```

```
Mean Age and Fare by Survival:
                Age       Fare
```

```
Survived
0         30.626179  22.117887
1         28.343690  48.395408
```

```
plt.figure(figsize=(8, 4))
sns.countplot(x='Pclass', hue='Survived', data=titanic_data)
plt.title('Survival Counts by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Count')
plt.legend(title='Survived')
plt.show()
```

```
plt.figure(figsize=(8, 4))
sns.countplot(x='Sex', hue='Survived', data=titanic_data)
plt.title('Survival Counts by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.legend(title='Survived')
plt.show()
```