In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Optional: Better plot visuals
sns.set(style="whitegrid")
```

In [2]:
```
df = pd.read_csv("titanic.csv")  # Load Titanic dataset
```

In [3]:
```
df = sns.load_dataset("titanic")
```

In [4]:
```
df.head()           # View first 5 rows
df.tail()           # Last 5 rows
df.shape            # (rows, columns)
df.columns          # Column names
df.info()           # Data types, non-null counts
df.describe()       # Summary statistics (numerical)
df.describe(include='object')  # For categorical columns
```

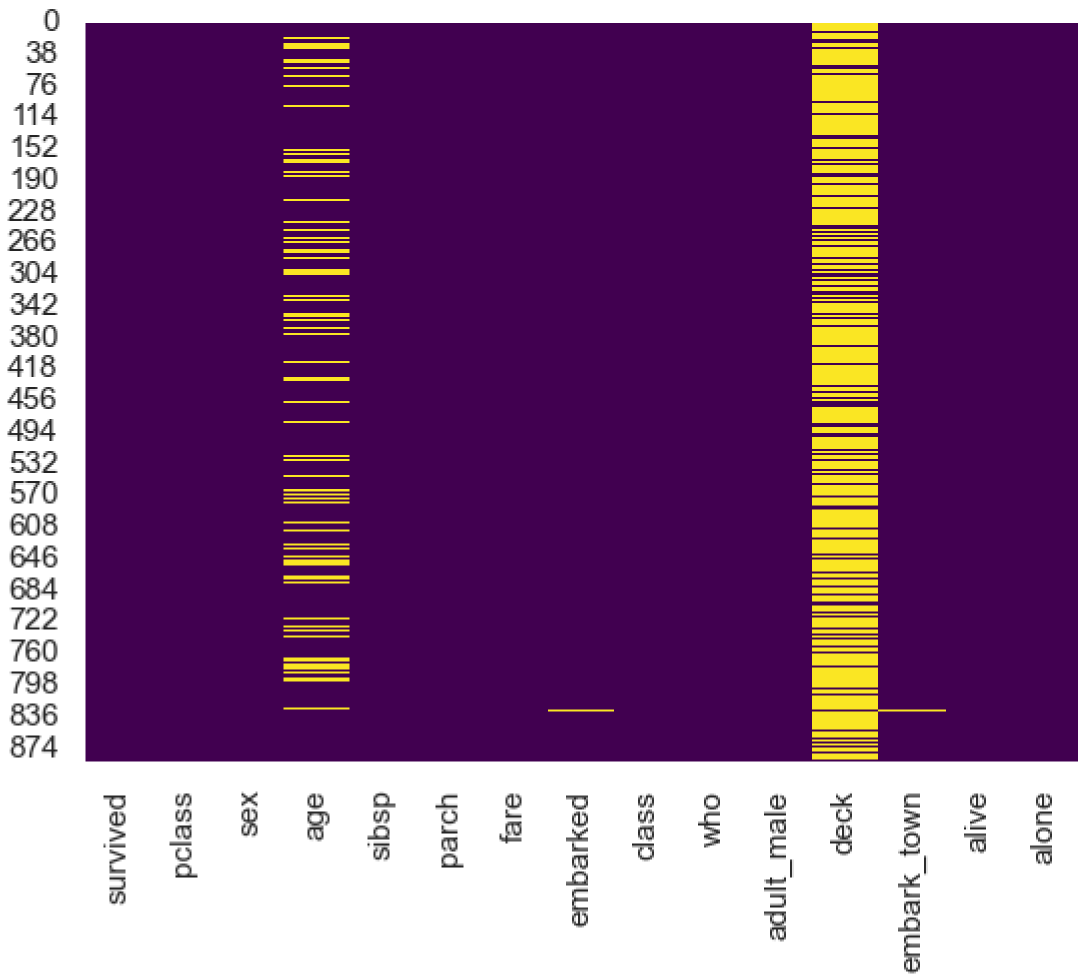
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```

Out[4]:

	sex	embarked	who	embark_town	alive
count	891	889	891	889	891
unique	2	3	3	3	2
top	male	S	man	Southampton	no
freq	577	644	537	644	549

In [5]:
```
df.isnull().sum()   # Missing values per column
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')  # Visualize missing data
```

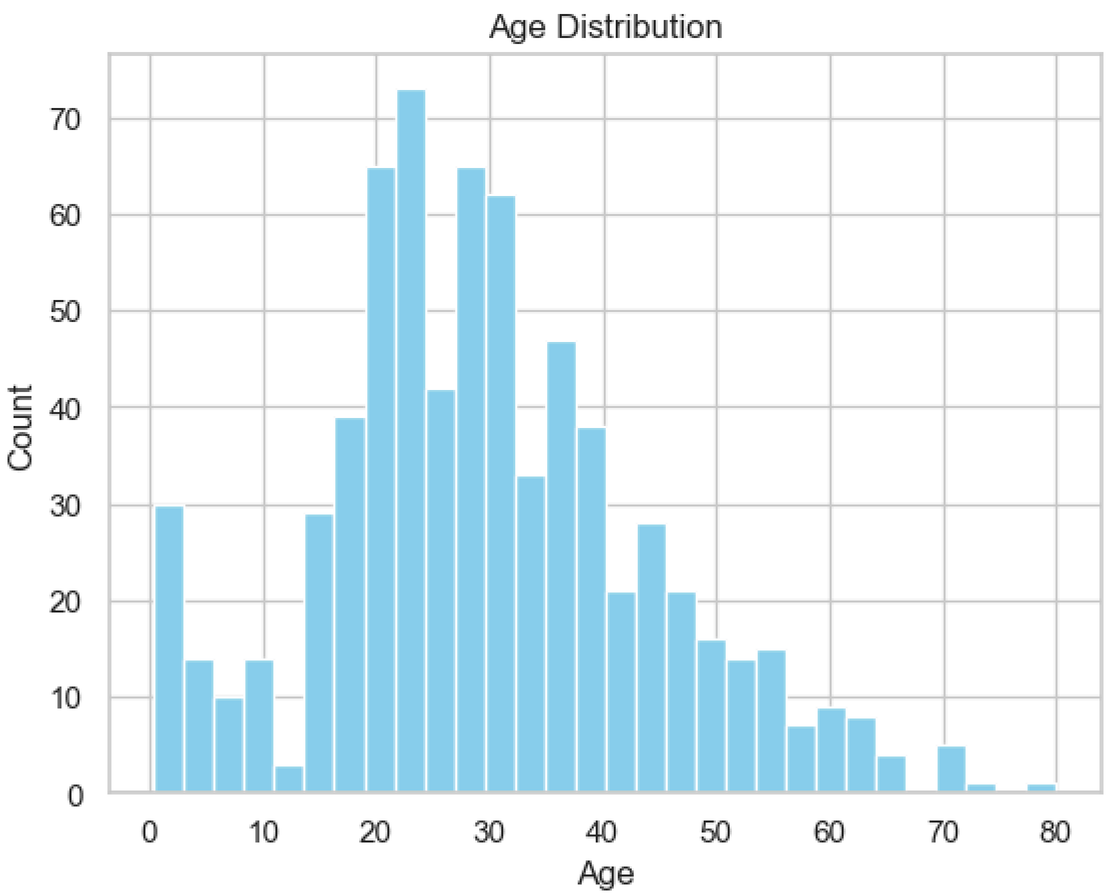
Out[5]: <Axes: >



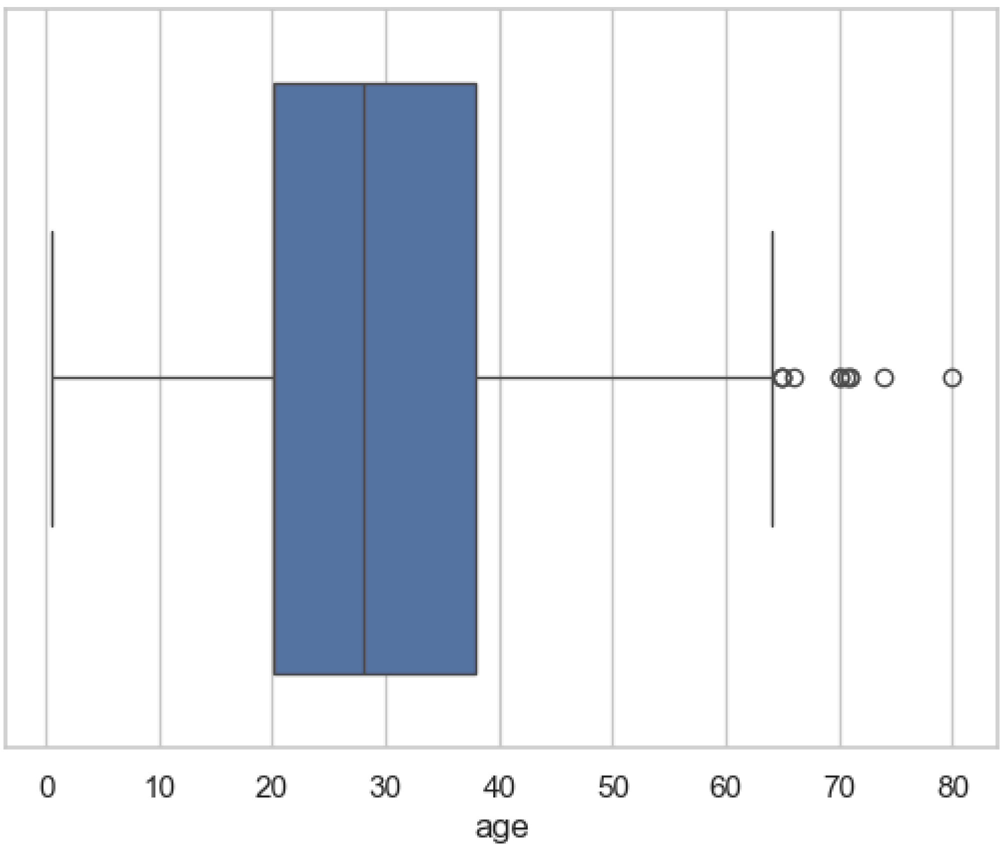
In [6]:

```
df['age'].hist(bins=30, color='skyblue')
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()

sns.boxplot(x='age', data=df)
```

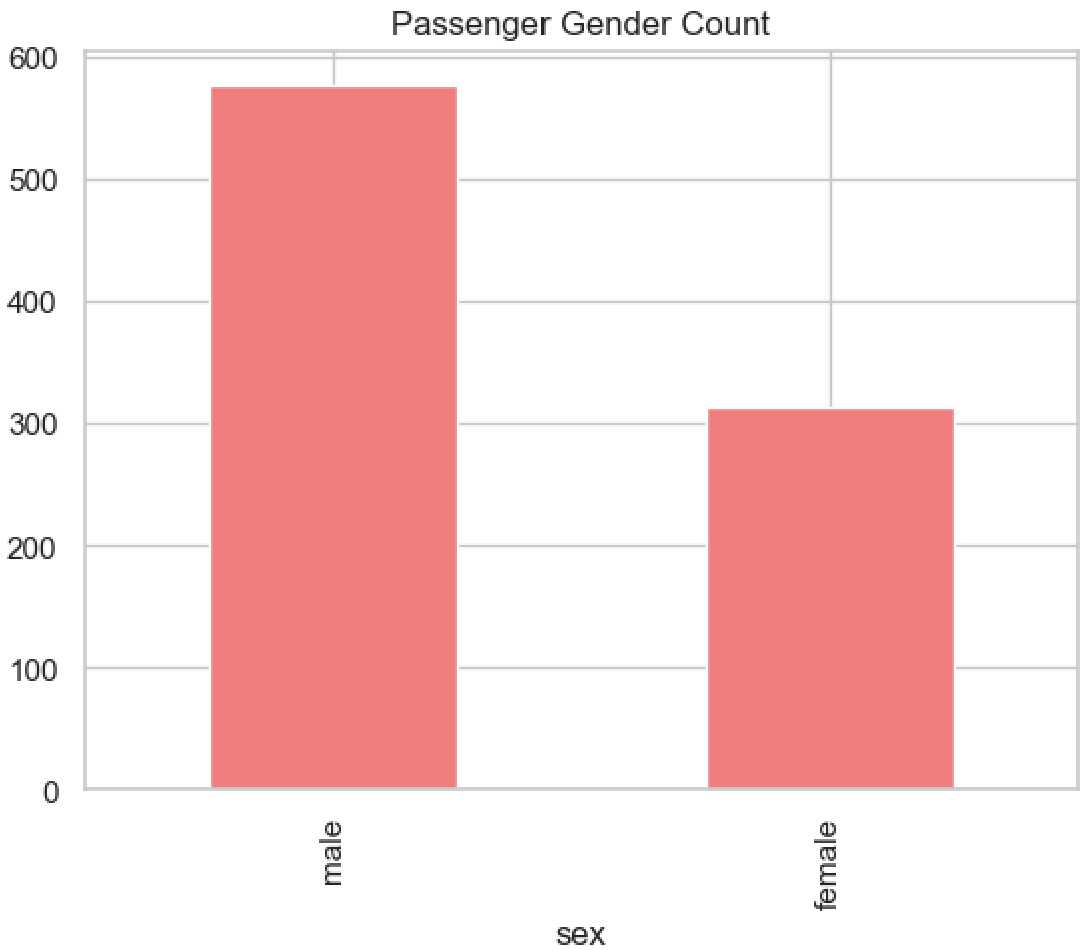


Out[6]: <Axes: xlabel='age'>

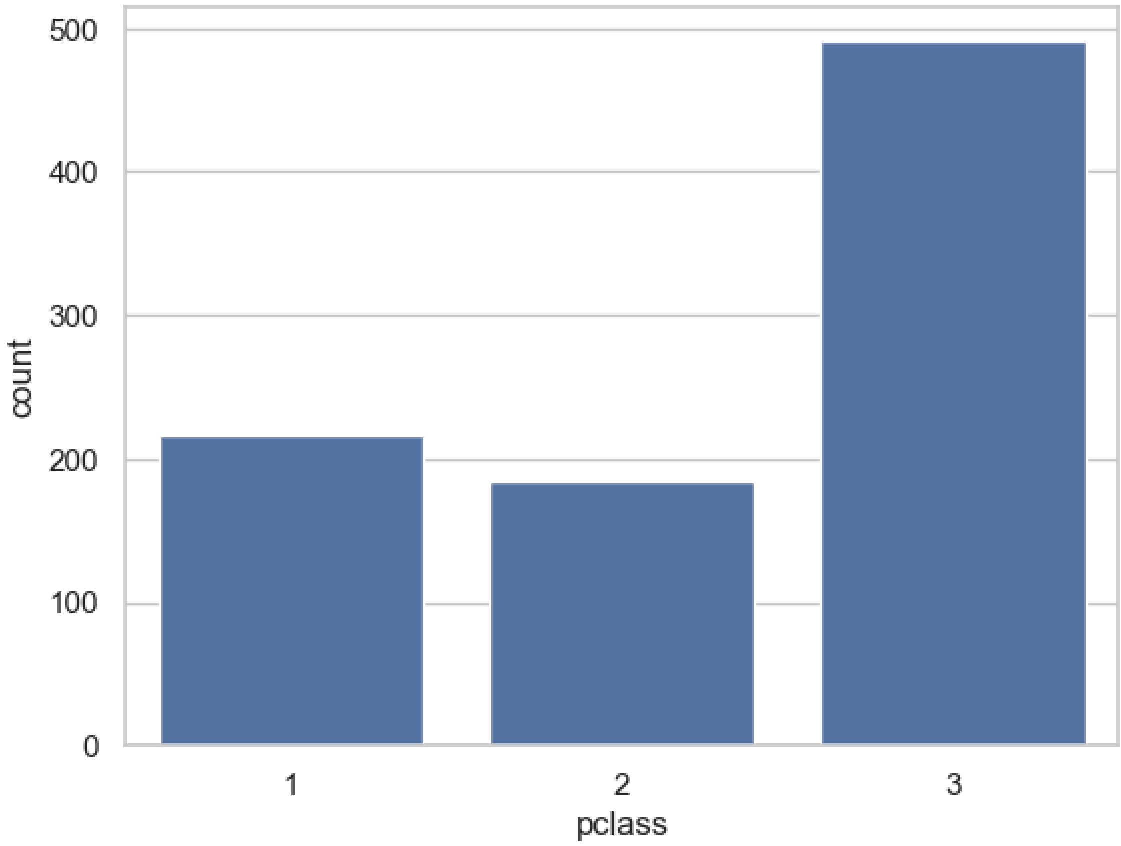


In [7]:
```
df['sex'].value_counts().plot(kind='bar', color='lightcoral')
plt.title("Passenger Gender Count")
plt.show()
```

```
sns.countplot(x='pclass', data=df)
```

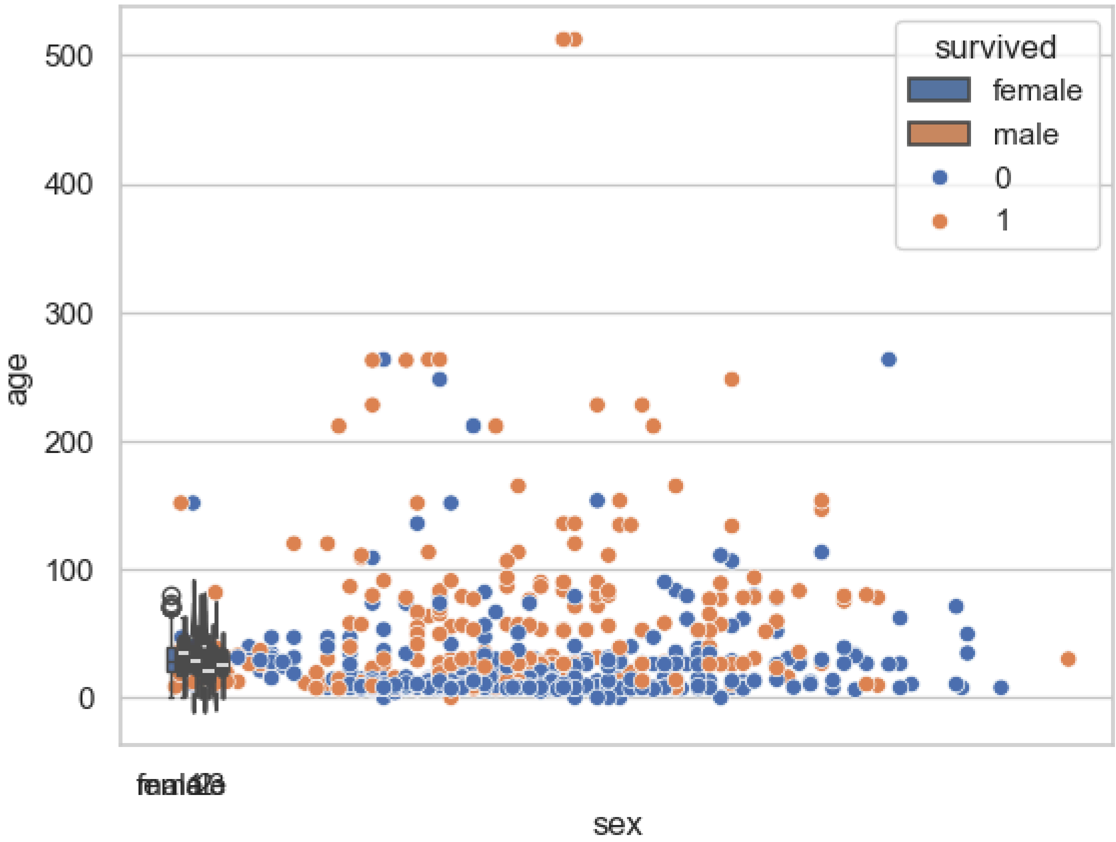


Out[7]: `<Axes: xlabel='pclass', ylabel='count'>`



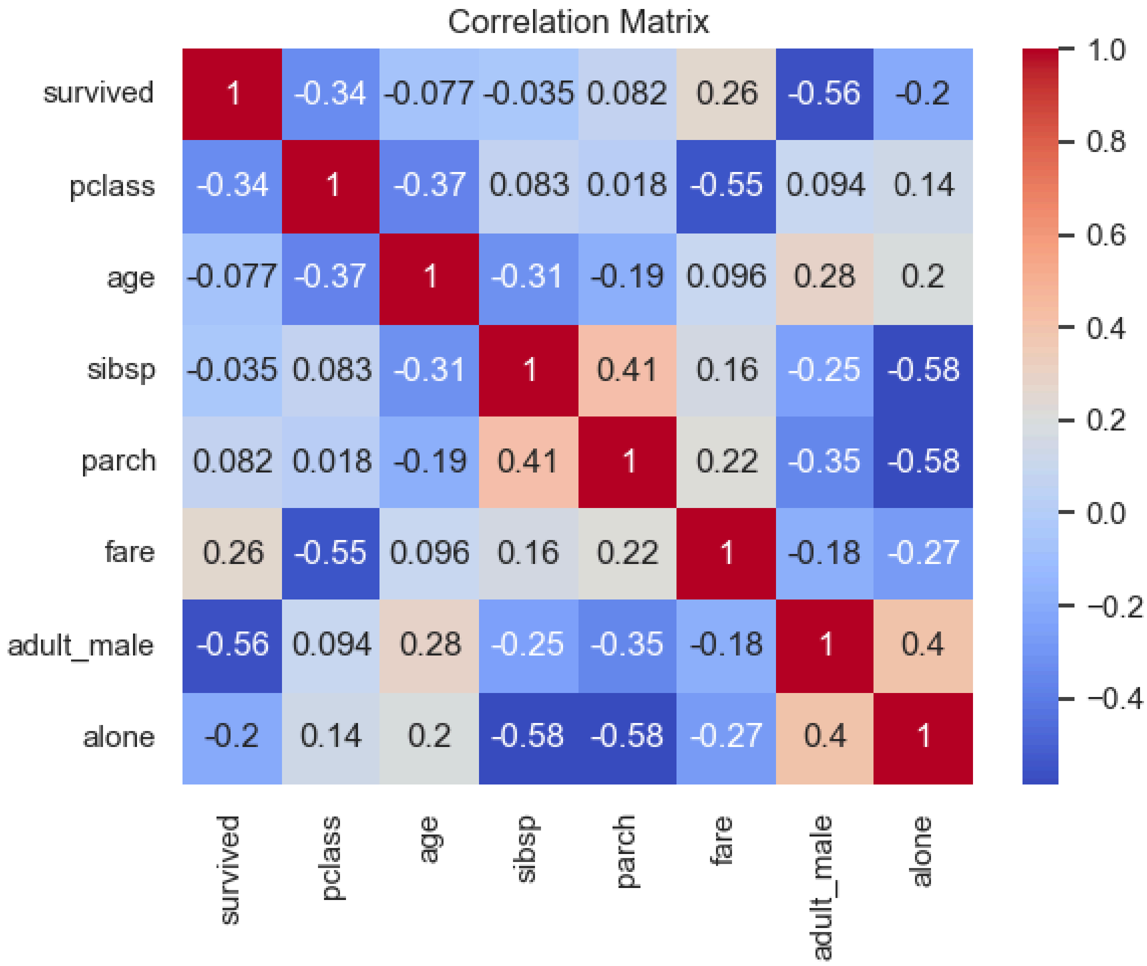
In [8]:
```
sns.boxplot(x='sex', y='age', data=df)
sns.violinplot(x='pclass', y='age', hue='sex', data=df, split=True)
sns.scatterplot(x='age', y='fare', hue='survived', data=df)
```

Out[8]: <Axes: xlabel='sex', ylabel='age'>



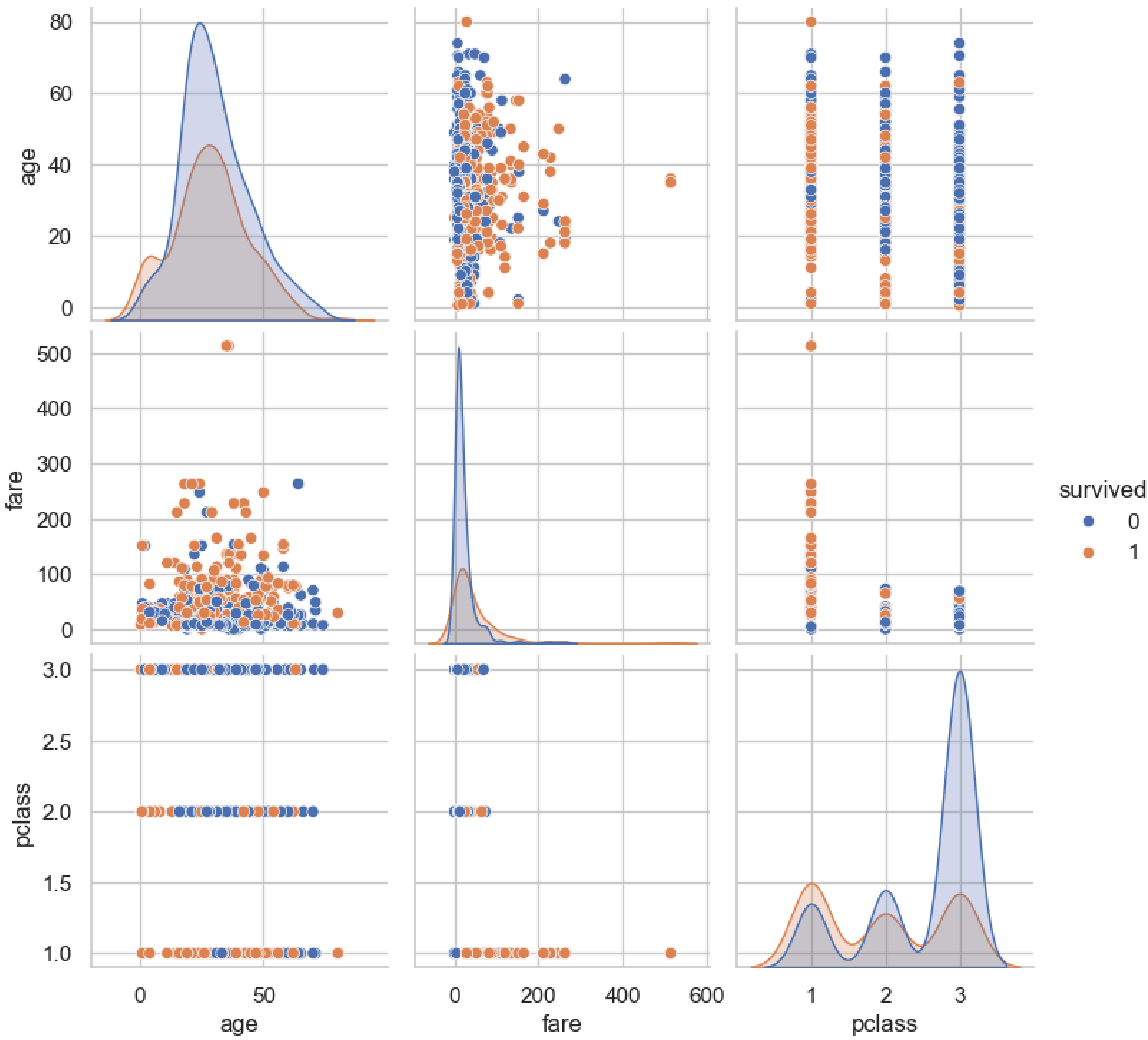
In [9]:
```
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
```

Out[9]: Text(0.5, 1.0, 'Correlation Matrix')



In [10]:
```
sns.pairplot(df[['age', 'fare', 'pclass', 'survived']], hue='survived')
```

Out[10]:
```
<seaborn.axisgrid.PairGrid at 0x1f059f01670>
```



Observation:

- Most passengers are between 20-40 years.
- Survival rate appears higher for females.
- Higher class (1st) passengers had better survival odds.

🔍 Summary of EDA Findings:

- **Missing Data**: Age, cabin, and embarked have missing values.
- **Gender**: Female passengers had a higher survival rate.
- **Class**: First class passengers were more likely to survive.
- **Age**: Younger passengers had a slightly higher chance of survival.
- **Fare**: Passengers who paid more fares tended to survive more.