```
from google.colab import files
uploaded = files.upload()
```

Choose Files tested.csv

- **tested.csv**(text/csv) - 29474 bytes, last modified: 4/15/2025 - 100% done

```
Saving tested.csv to tested (2).csv
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")
```

```
df = pd.read_csv("tested.csv")
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.829
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.000
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.687
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.662
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.287

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.info()
df.describe()
df.isnull().sum()
df.columns
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  418 non-null    int64  
 1   Survived     418 non-null    int64  
 2   Pclass       418 non-null    int64  
 3   Name         418 non-null    object 
 4   Sex          418 non-null    object 
 5   Age          332 non-null    float64
 6   SibSp        418 non-null    int64  
 7   Parch        418 non-null    int64  
 8   Ticket       418 non-null    object 
 9   Fare         417 non-null    float64
 10  Cabin        91 non-null     object 
 11  Embarked     418 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

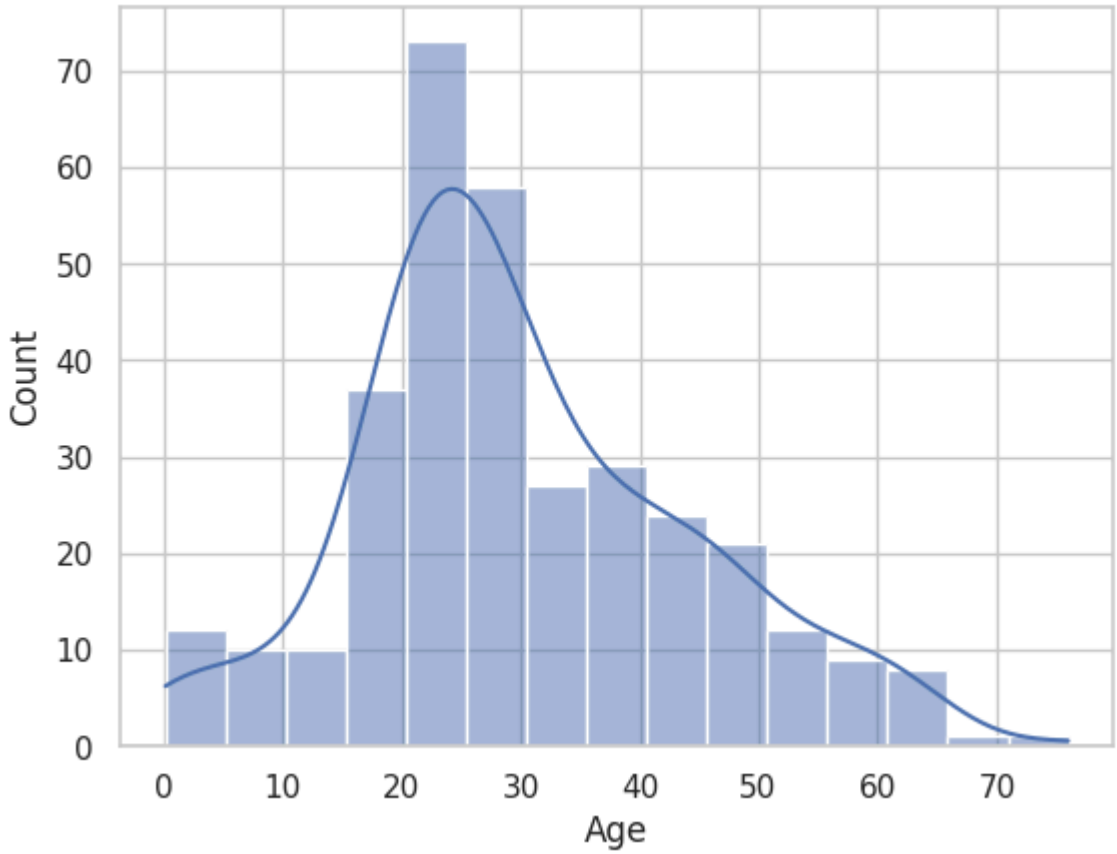
```
df['Sex'].value_counts()
df['Pclass'].value_counts()
df['Embarked'].value_counts()
df['Survived'].value_counts()
```

Survived	count
0	266
1	152

dtype: int64

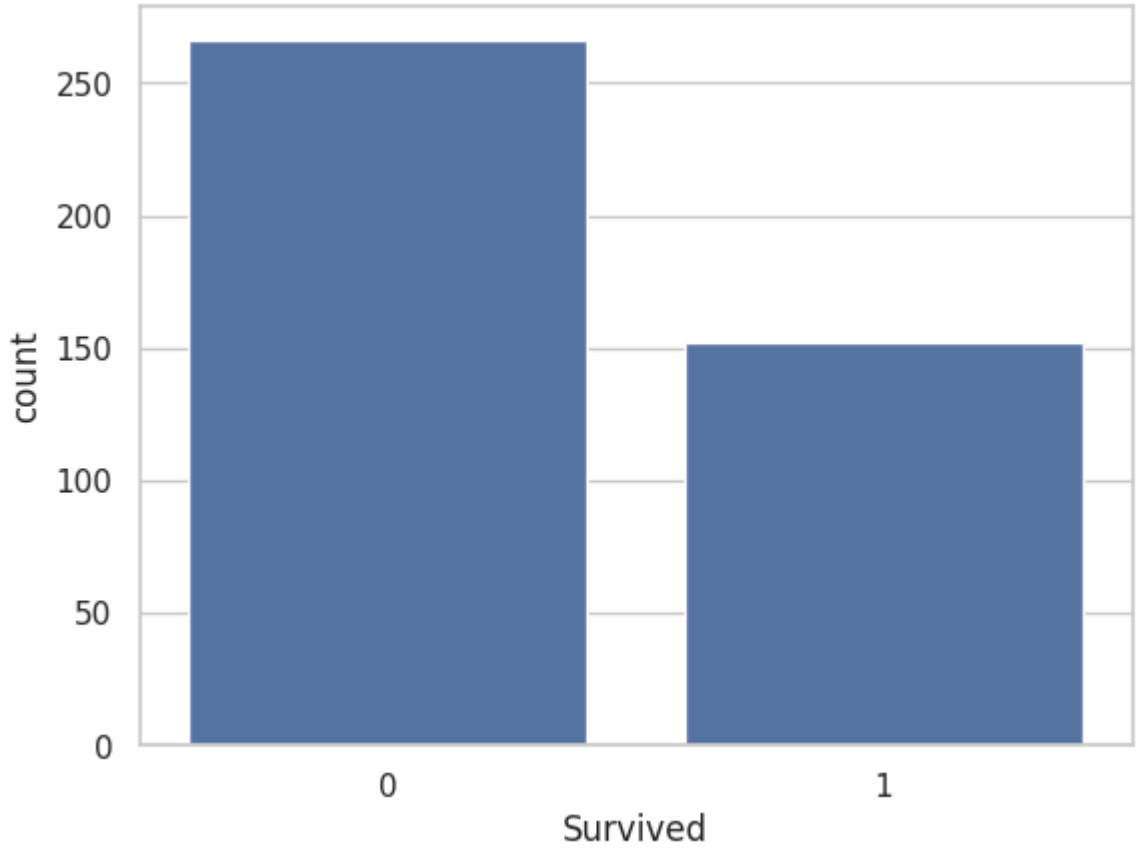
```
sns.histplot(df['Age'], kde=True)
```

```
<Axes: xlabel='Age', ylabel='Count'>
```



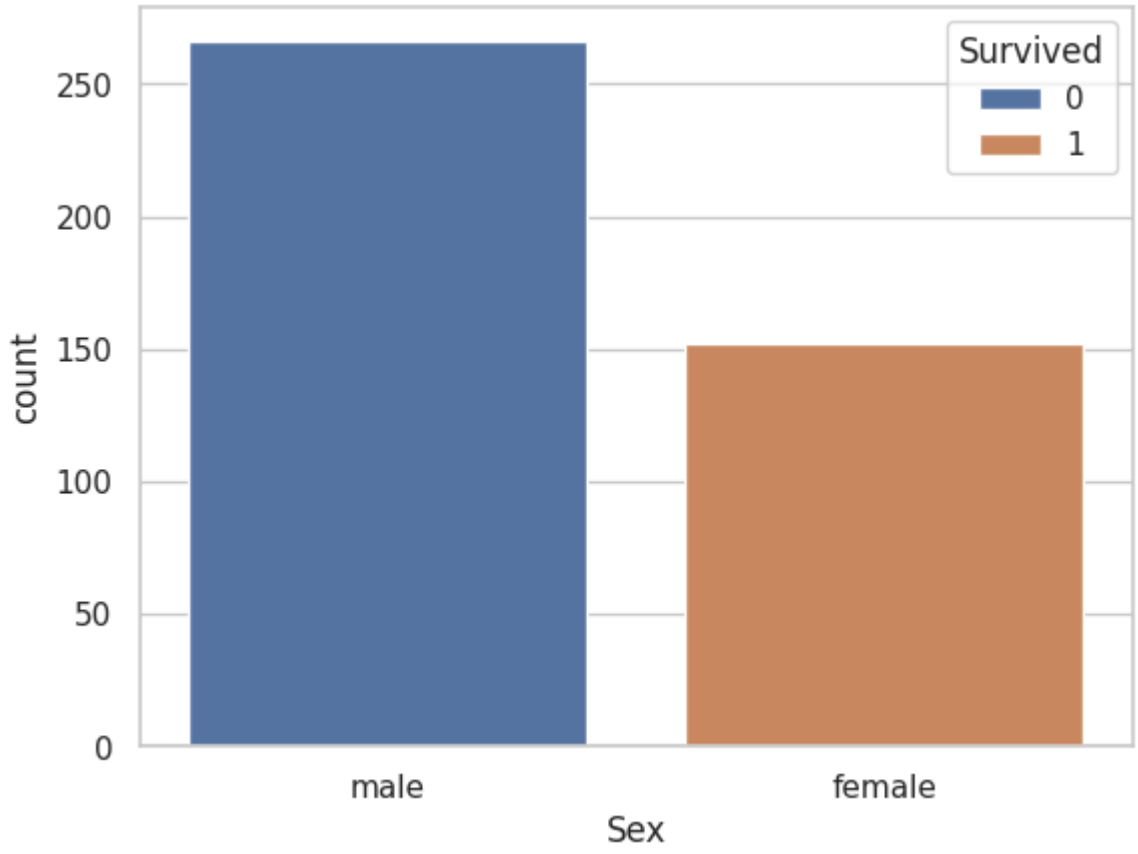

```
sns.countplot(x='Survived', data=df)
```

```
<Axes: xlabel='Survived', ylabel='count'>
```



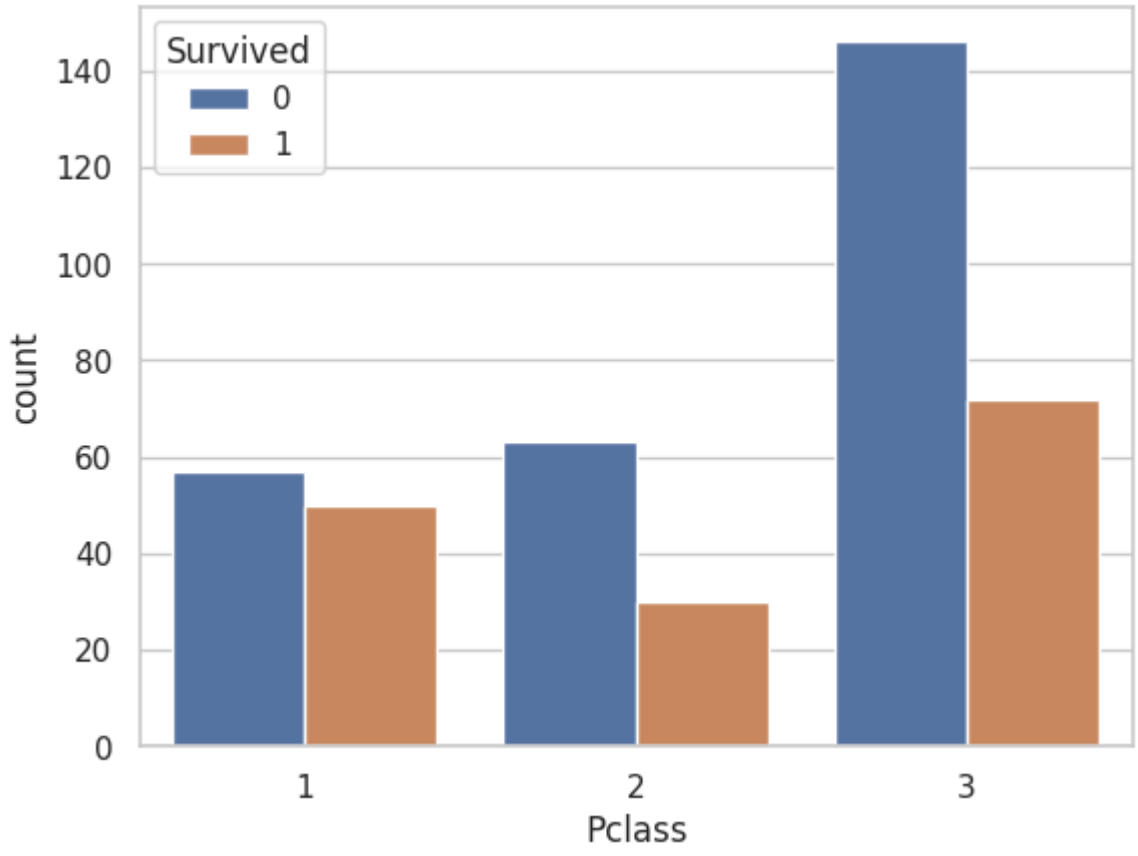
```
sns.countplot(x='Sex', hue='Survived', data=df)
```

```
<Axes: xlabel='Sex', ylabel='count'>
```



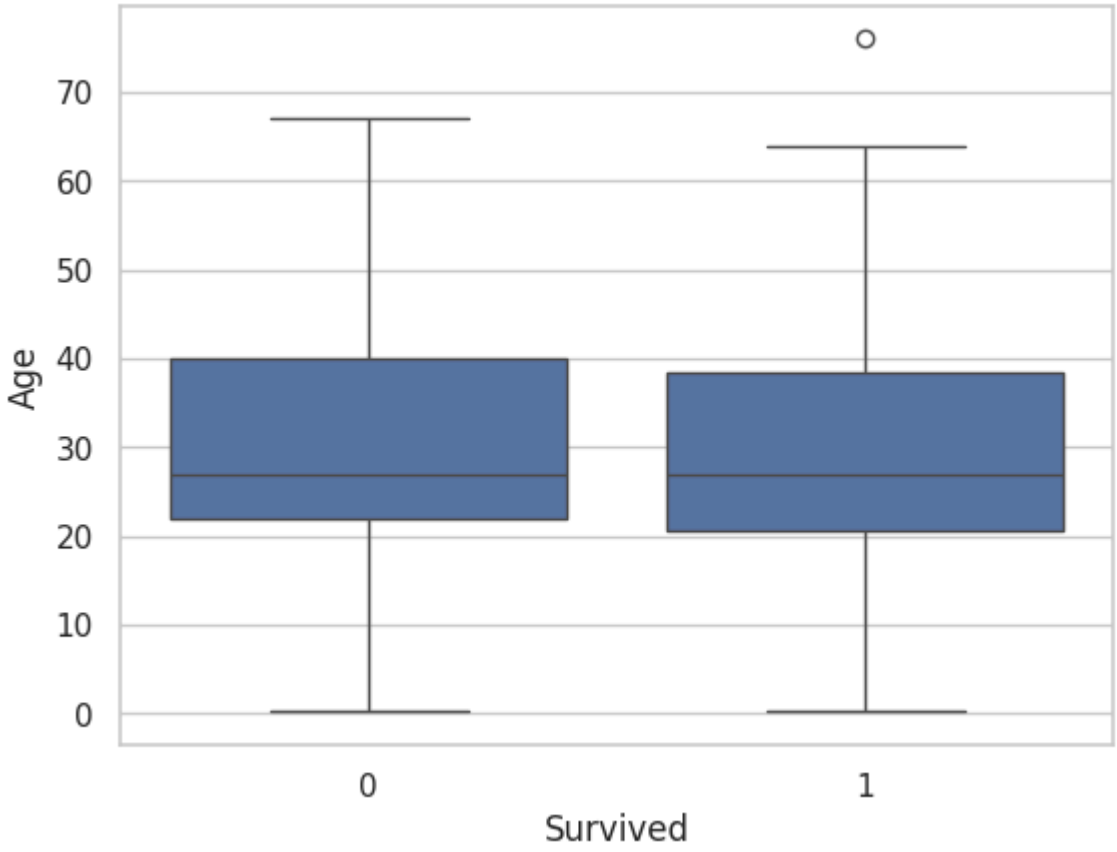

```
sns.countplot(x='Pclass', hue='Survived', data=df)
```

```
<Axes: xlabel='Pclass', ylabel='count'>
```



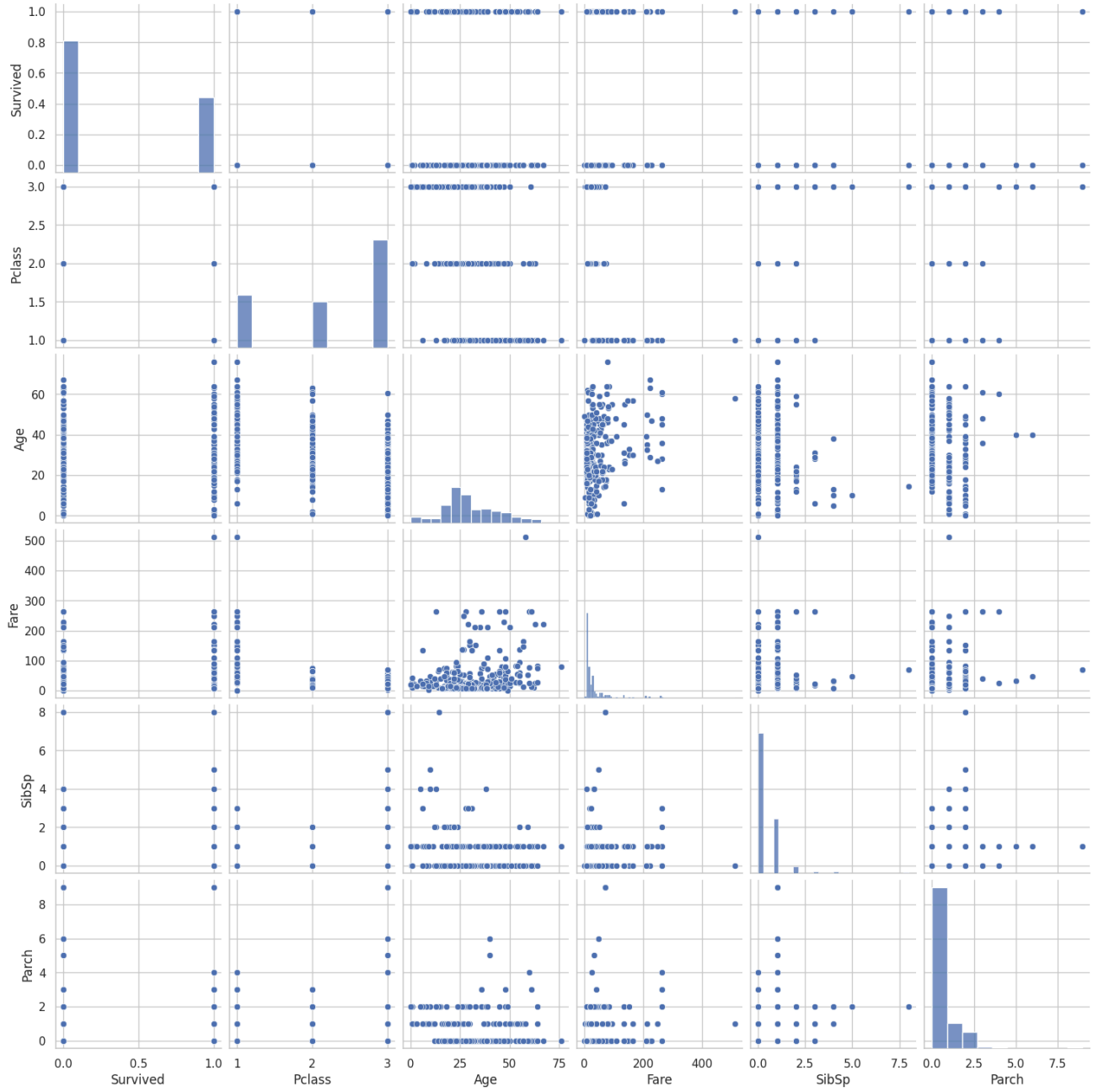
```
sns.boxplot(x='Survived', y='Age', data=df)
```

```
<Axes: xlabel='Survived', ylabel='Age'>
```



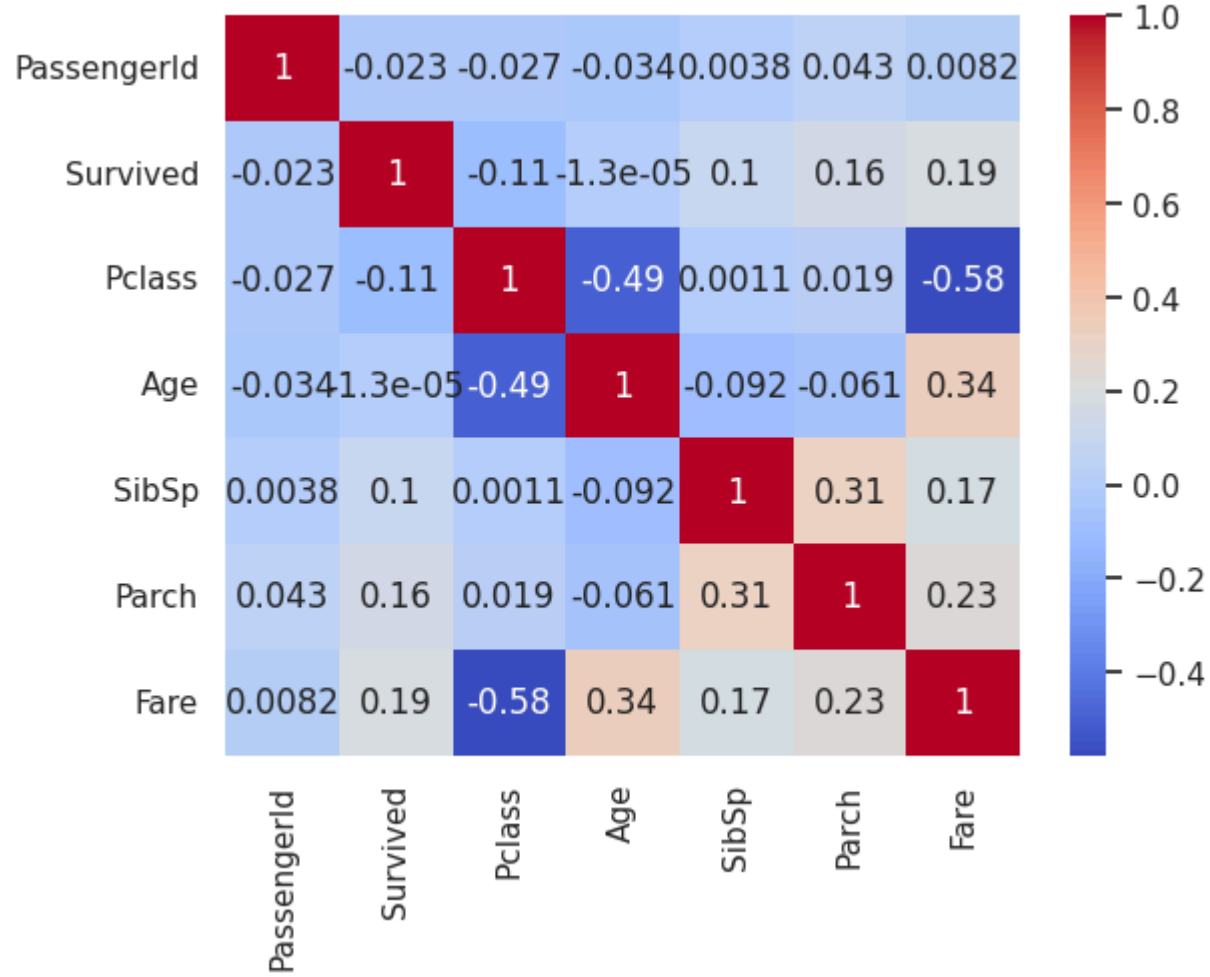

```
sns.pairplot(df[['Survived', 'Pclass', 'Age', 'Fare', 'SibSp', 'Parch']])
```

```
<seaborn.axisgrid.PairGrid at 0x79759720c7d0>
```



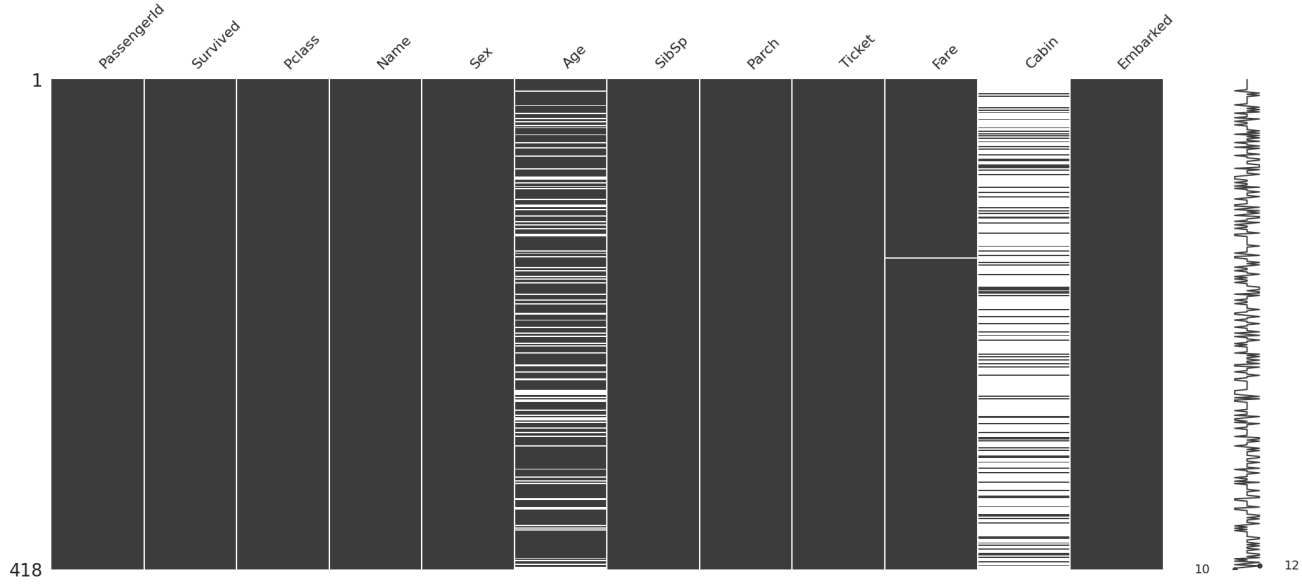

```
numeric_df = df.select_dtypes(include='number')

sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
```

<Axes: >




```
import missingno as msno
msno.matrix(df)
```

<Axes: >




```
df.groupby('Survived')['Age'].mean()
```

Survived	Age
0	30.272732
1	30.272362

```
df.pivot_table(index='Pclass', columns='Sex', values='Survived')
```

Pclass	Sex female	male
1	1.0	0.0
2	1.0	0.0
3	1.0	0.0

```
sns.violinplot(x='Survived', y='Age', hue='Sex', data=df, split=True)
```

```
<Axes: xlabel='Survived', ylabel='Age'>
```