```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
df=pd.read_csv('titanic_dataset.csv')
df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

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

```
df.info()
```

```
<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
```

```
df.shape
```

```
(891, 12)
```

```
df.head(4)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S

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

```
df.nunique()
```

	0
PassengerId	891
Survived	2
Pclass	3
Name	891
Sex	2
Age	88
SibSp	7
Parch	7
Ticket	681
Fare	248
Cabin	147
Embarked	3

dtype: int64

```
df["Survived"].value_counts()
```

Survived	count
0	549
1	342

dtype: int64

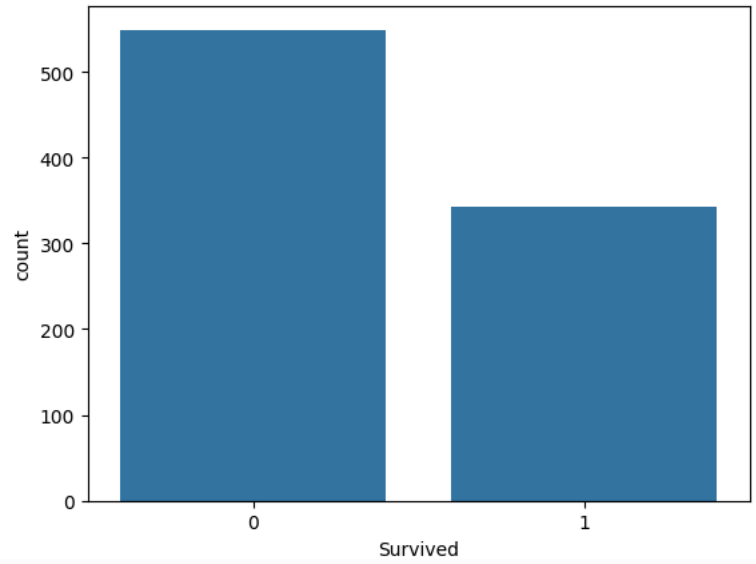
```
per=(df["Survived"].value_counts()/df.shape[0]*100).round(2)
per
```

Survived	count
0	61.62
1	38.38

dtype: float64

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

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




```
df.Pclass.unique()
```

```
array([3, 1, 2])
```

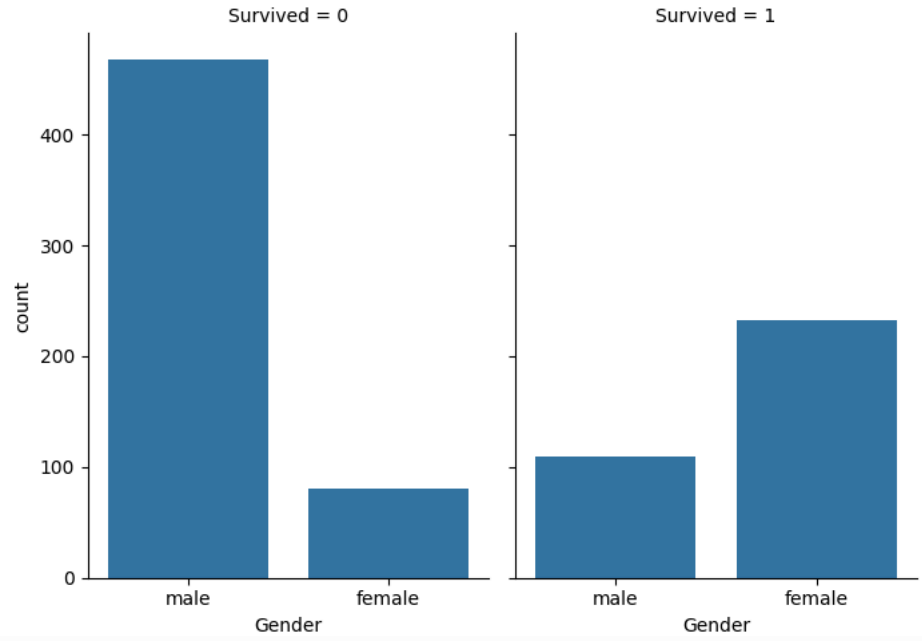
```
df.rename(columns={"Sex" : 'Gender'}, inplace=True)
df
```

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

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

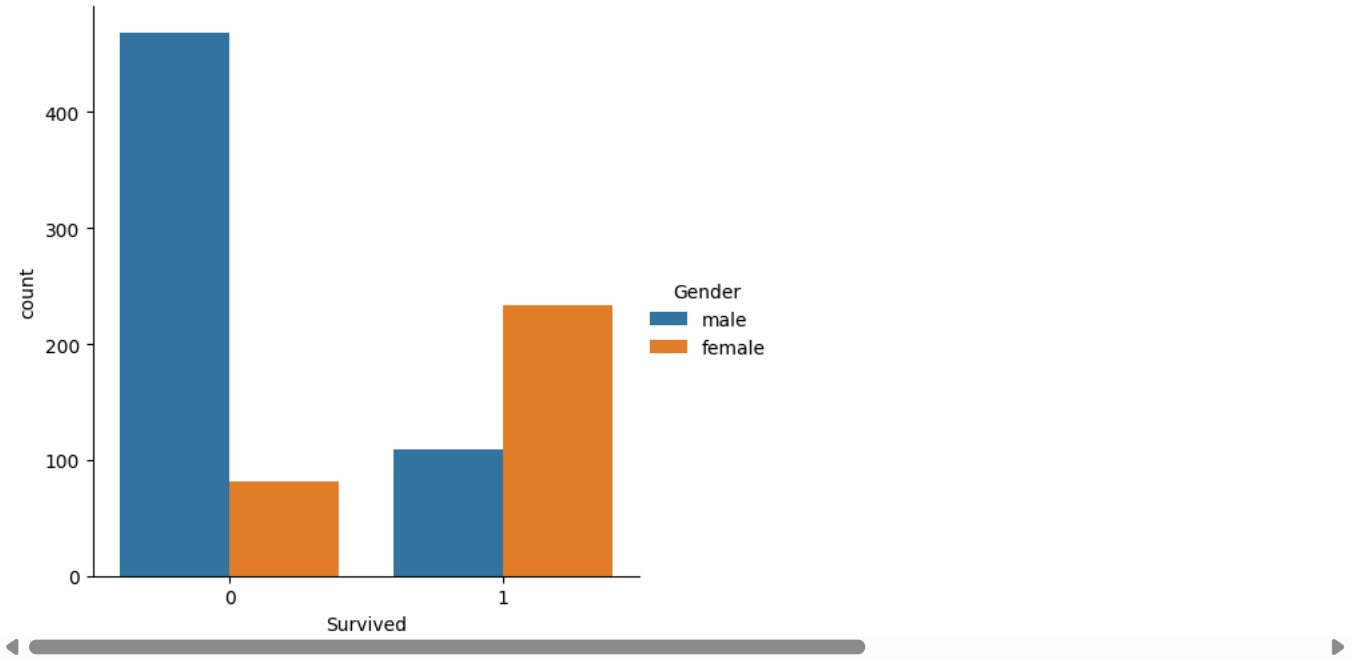
```
sns.catplot(x="Gender",col="Survived",kind="count", data=df,height=5,aspect=.7)
```

```
<seaborn.axisgrid.FacetGrid at 0x7d138d1f5690>
```



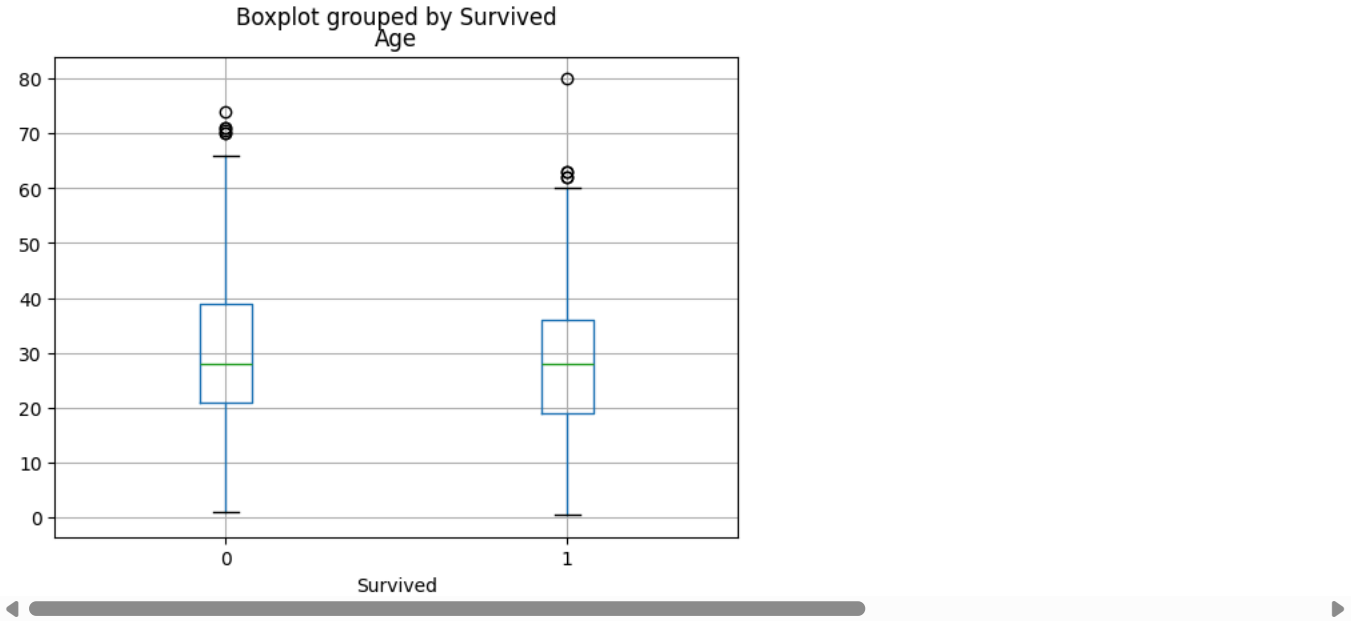

```
sns.catplot(x="Survived",hue="Gender",data=df,kind="count")
```

```
<seaborn.axisgrid.FacetGrid at 0x7d138d2f3f90>
```



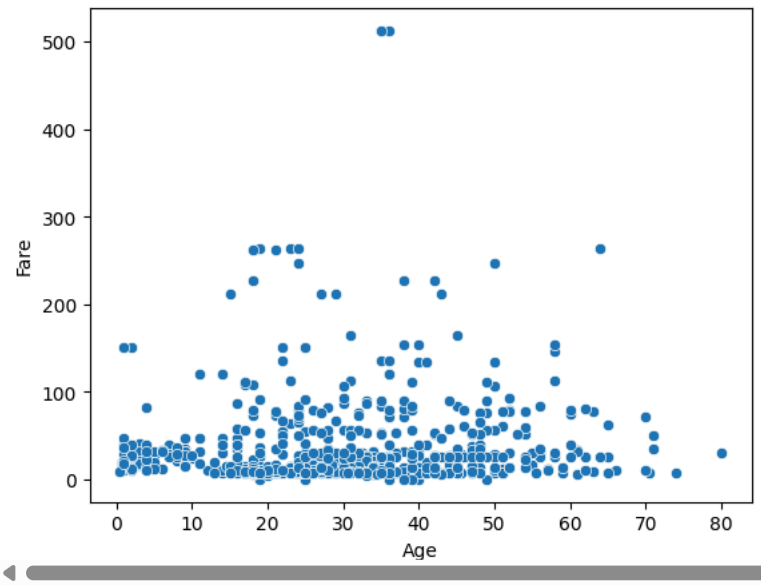

```
df.boxplot(column="Age",by="Survived")
```

```
<Axes: title={'center': 'Age'}, xlabel='Survived'>
```



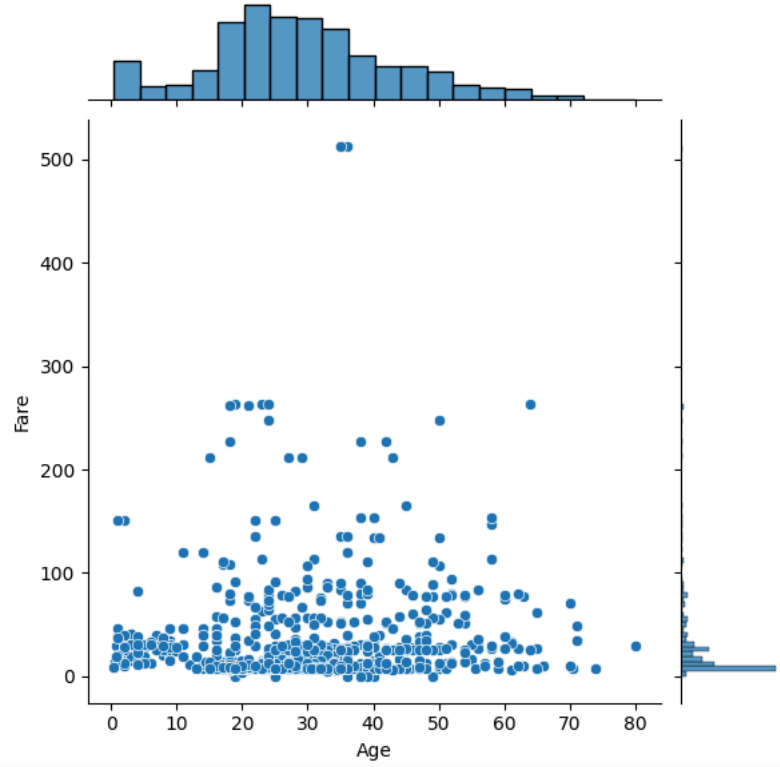

```
sns.scatterplot(x=df["Age"],y=df['Fare'])
```

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



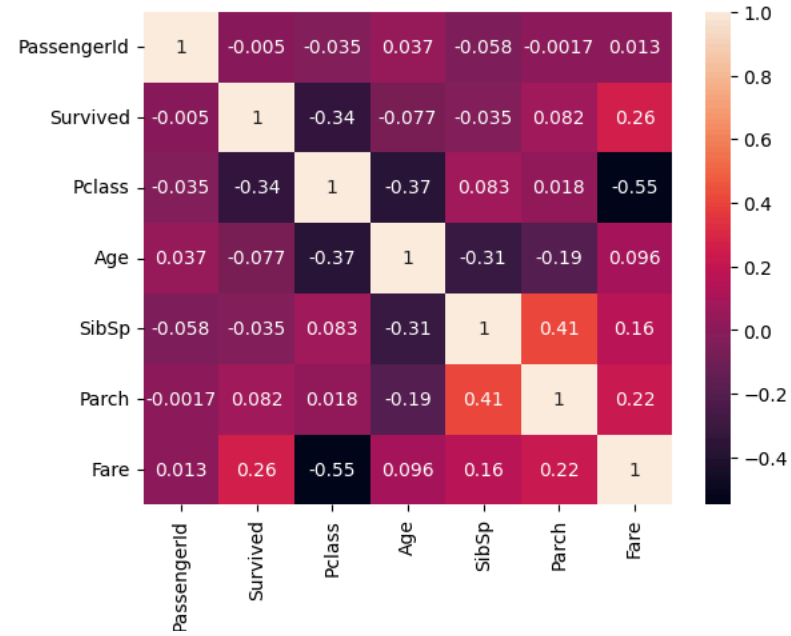

```
sns.jointplot(x="Age",y="Fare",data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7d138d7e3d10>
```

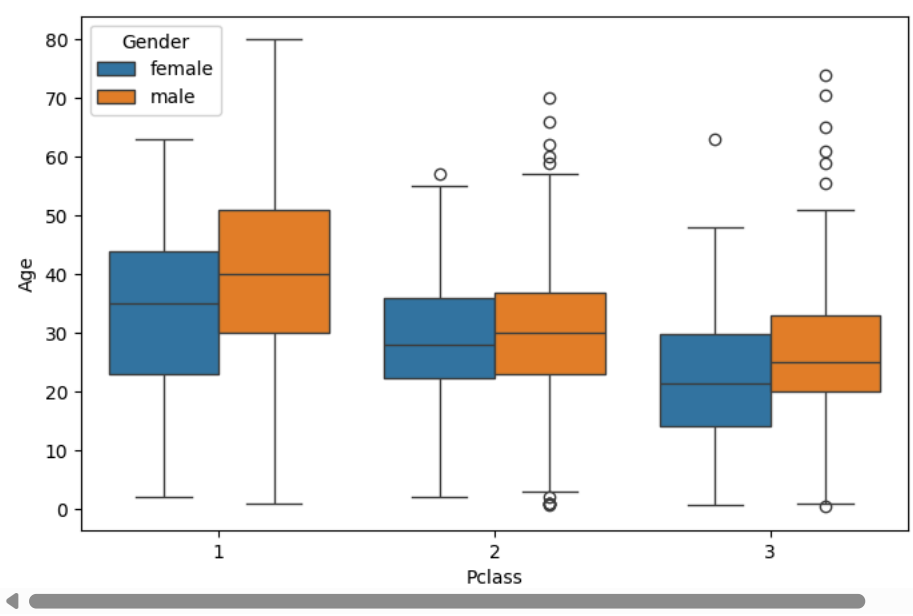



```
corr = df. select_dtypes (include=['number']). corr()
sns. heatmap(corr,annot=True)
```

<Axes: >




```
corr = dt. select_dtypes (include=['number']). corr() sns. heatmap(corr, annot=True)
```




```
sns.pairplot(df)
```

<seaborn.axisgrid.PairGrid at 0x7d138cdd4850>