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

```python
dataset = sns.load_dataset('titanic')
dataset.head()
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alo
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	Fa
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	Fa
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	Ti
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	Fa
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	Ti

```python
dataset.shape
```

```
(891, 15)
```

```python
dataset.info()
```

```
<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
```

```python
dataset.describe()
```

	survived	pclass	age	sibsp	parch	fare

count	891.000000 survived	891.000000 pclass	714.000000 age	891.000000 sibsp	891.000000 parch	891.000000 fare
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [7]:

```python
dataset.isnull().sum()
```

Out[7]:

```
survived         0
pclass           0
sex              0
age            177
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck           688
embark_town      2
alive            0
alone            0
dtype: int64
```

In [8]:

```python
dataset['age']=dataset['age'].fillna(np.mean(dataset['age']))
dataset['deck']=dataset['deck'].fillna('A')
```

In [9]:

```python
dataset.isnull().sum()
```

Out[9]:

```
survived         0
pclass           0
sex              0
age              0
sibsp            0
parch            0
fare             0
embarked         2
class            0
who              0
adult_male       0
deck             0
embark_town      2
alive            0
alone            0
dtype: int64
```

In [10]:

```python
sns.boxplot(x='sex', y='age', data=dataset, hue="survived")
```
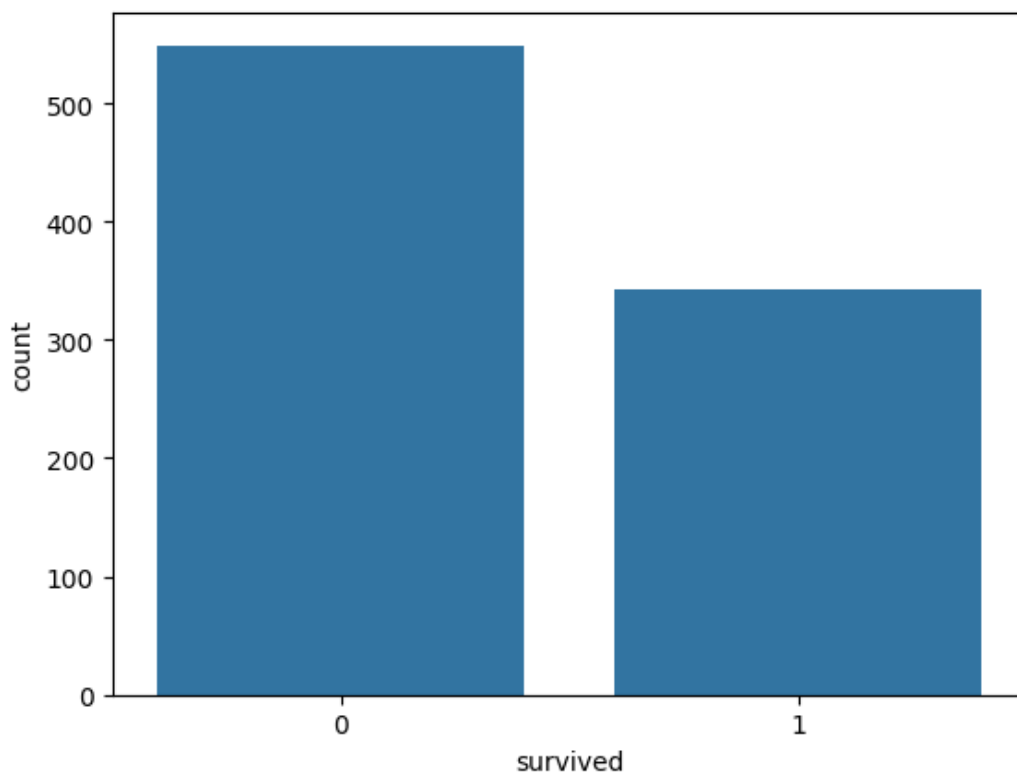
Out[10]:

```
<Axes: xlabel='sex', ylabel='age'>
```

```
sns.countplot(x='survived', data=dataset);
```
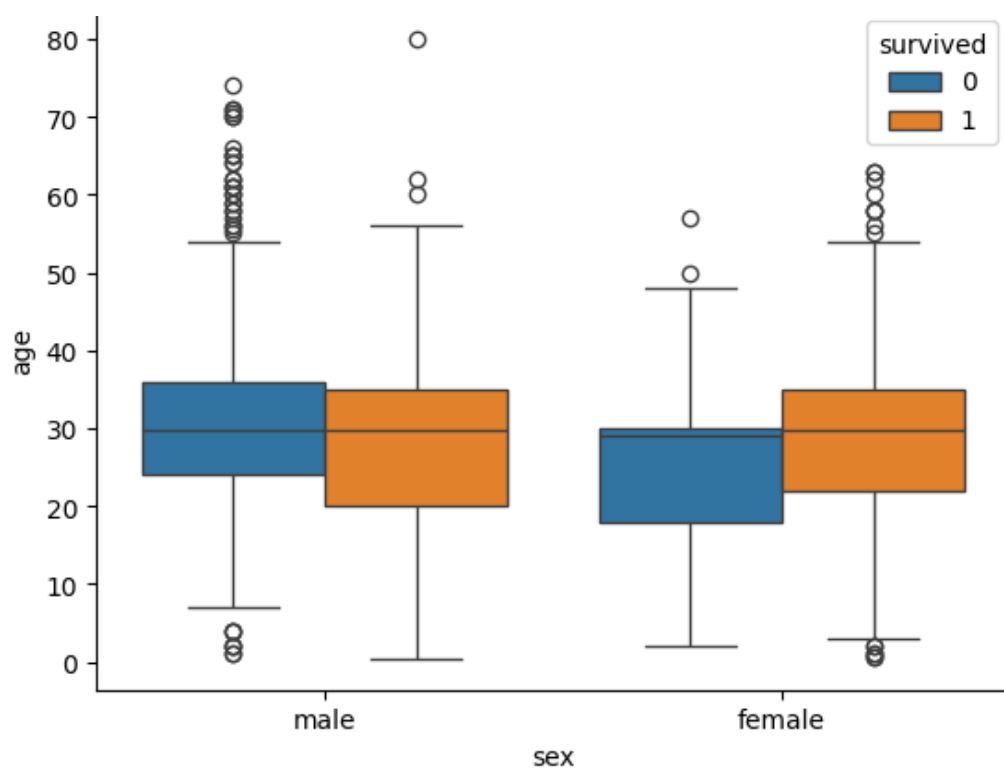
```
dataset.groupby(['survived','sex'])['survived'].count()
```
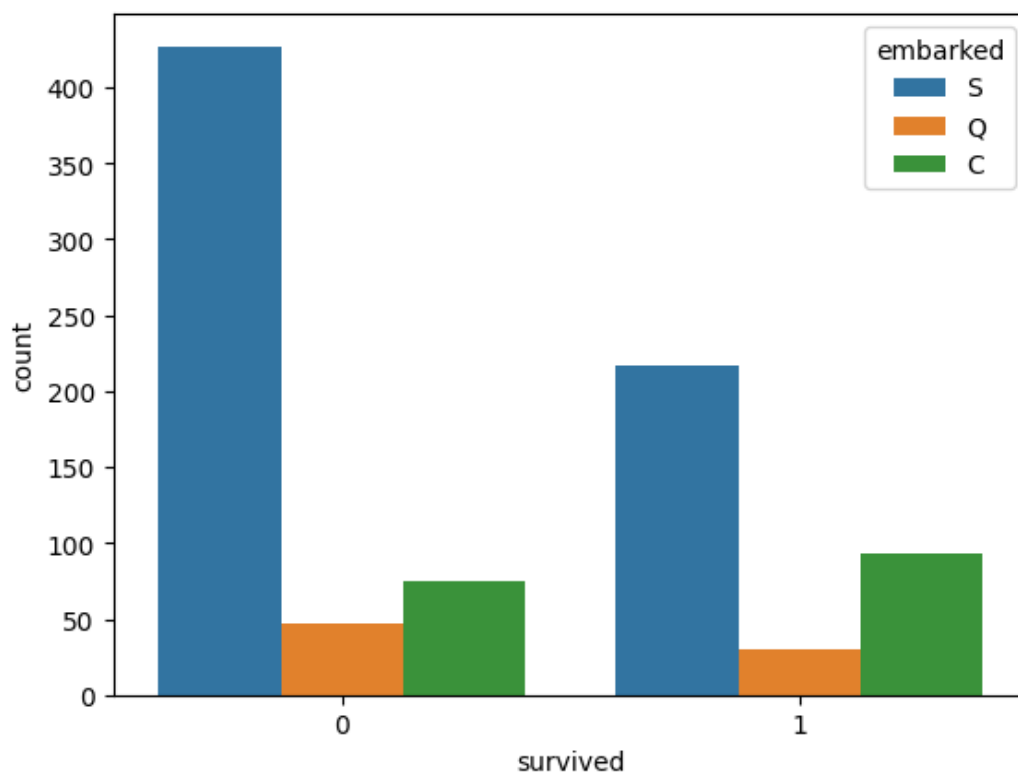
```
survived  sex
0         female      81
          male       468
1         female     233
          male       109
Name: survived, dtype: int64
```

```
sns.countplot(x='survived', data=dataset hue = 'embarked');
```

In [14]:

```
pd.crosstab([dataset.sex, dataset.survived], [dataset.sibsp, dataset.pclass], margins=True)
```

Out[14]:

sibsp		0			1			2			3			4	5	8	All
pclass		1	2	3	1	2	3	1	2	3	1	2	3	3	3	3	
sex	survived																
female	0	1	3	33	2	3	21	0	0	3	0	0	7	4	1	3	81
	1	48	41	48	38	25	17	3	3	4	2	1	1	2	0	0	233
male	0	59	67	235	16	20	35	1	4	7	1	0	4	11	4	4	468
	1	29	9	35	15	7	10	1	1	1	0	0	0	1	0	0	109
All		137	120	351	71	55	83	5	8	15	3	1	12	18	5	7	891

In [15]:

```
pd.crosstab([dataset.sex, dataset.survived], [dataset.parch, dataset.pclass], margins=True)
```
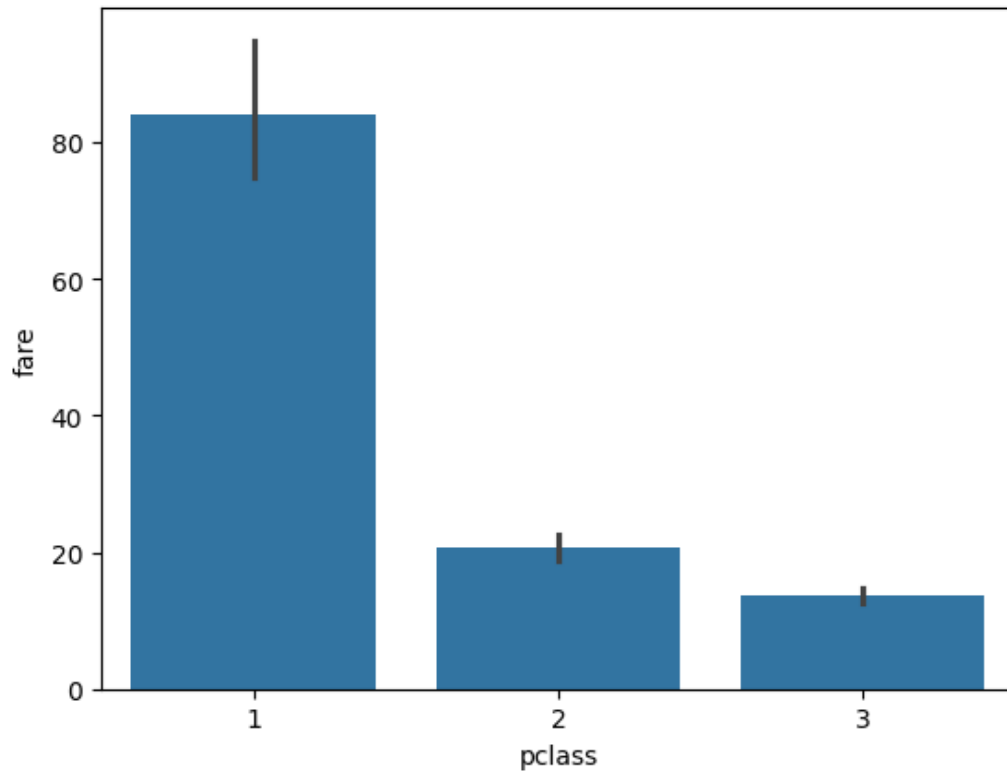
Out[15]:

parch		0			1			2			3		4		5	6	All
pclass		1	2	3	1	2	3	1	2	3	2	3	1	3	3	3	
sex	survived																
female	0	1	5	35	0	1	13	2	0	17	0	1	0	2	3	1	81
	1	63	40	50	17	17	12	11	11	8	2	1	0	0	1	0	233
male	0	63	81	260	10	7	22	3	3	15	0	1	1	1	1	0	468
	1	36	8	36	4	7	8	5	2	3	0	0	0	0	0	0	109
All		163	134	381	31	32	55	21	16	43	2	3	1	3	5	1	891

In [16]:

```
sns.barplot(y = "fare",x = "pclass",data = dataset)
```

Out[16]:

```
<Axes: xlabel='pclass', ylabel='fare'>
```


In [17]:

```
sns.boxplot(y = "fare",x = "pclass",data = dataset[dataset["fare"] < 200])
```
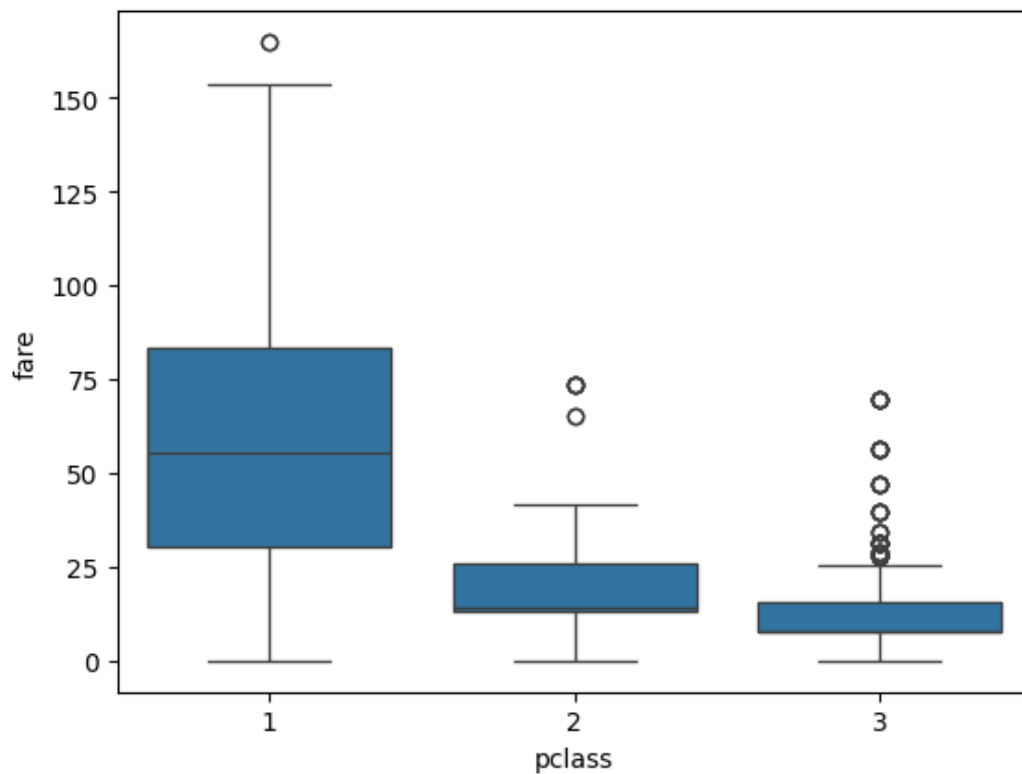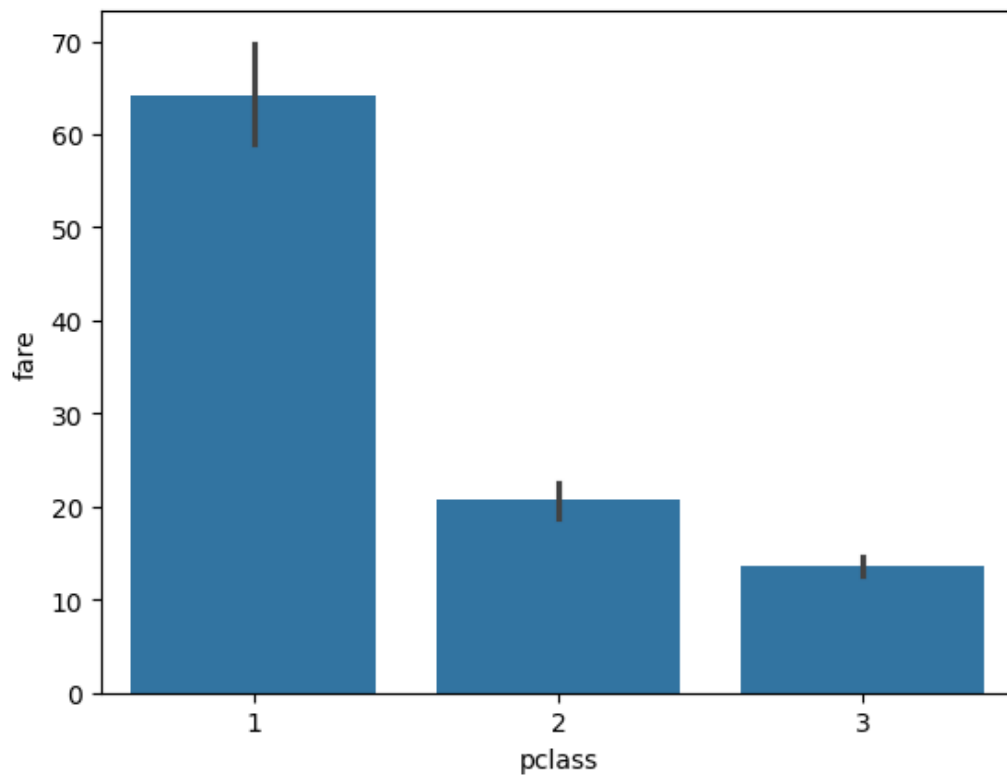
Out[17]:

```
<Axes: xlabel='pclass', ylabel='fare'>
```


In [18]:

```
sns.barplot(y = "fare",x = "pclass",data = dataset[dataset["fare"] < 200])
```
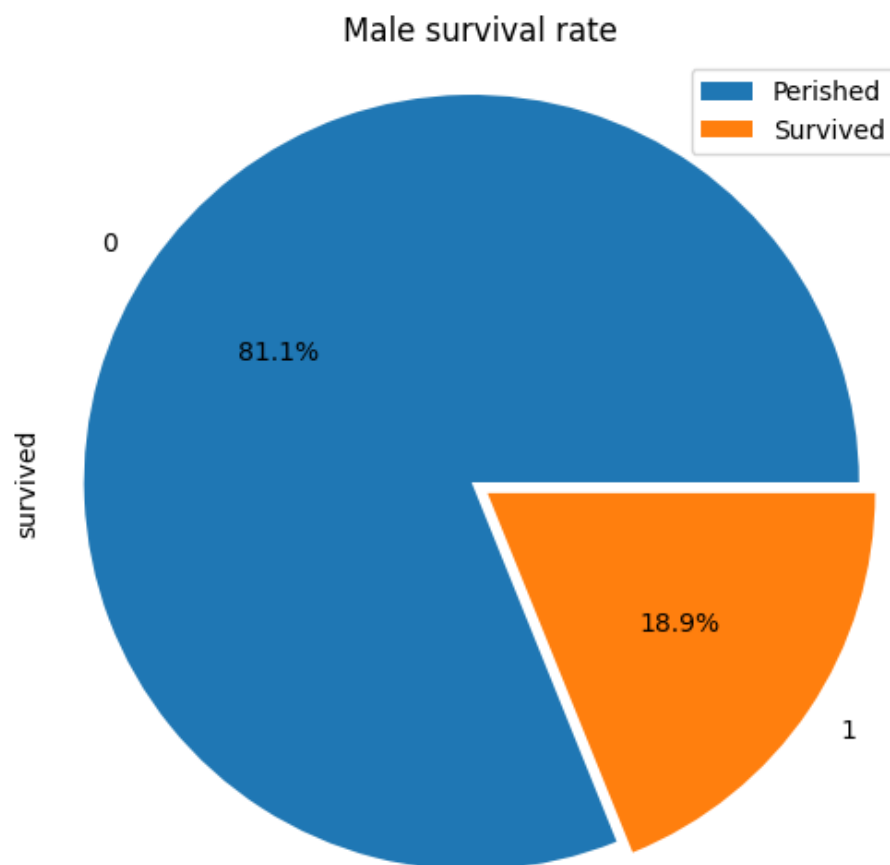
Out[18]:

```
<Axes: xlabel='pclass', ylabel='fare'>
```

```python
dataset[dataset['sex'] == 'male'].survived.groupby(dataset.survived).count().plot(kind='
pie', figsize=(6, 6),explode=[0,0.05],autopct='%1.1f%%')
plt.axis('equal')
plt.legend(["Perished","Survived"])
plt.title("Male survival rate")
plt.show()
```


In [ ]: