

```
In [1]: import pandas as pd;
import numpy as np;
import seaborn as sns;
import matplotlib.pyplot as plt
```

```
In [2]: titanic = pd.read_csv('titanic.csv')
```

```
In [3]: titanic.head()
```

```
Out[3]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cat
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	Ni
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	Ni
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	Ni
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	Ni
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	Ni

```
In [4]: titanic.isna().sum()
```

```
Out[4]: PassengerId      0
Survived      0
Pclass      0
Name      0
Sex      0
Age      86
SibSp      0
Parch      0
Ticket      0
Fare      1
Cabin      327
Embarked      0
dtype: int64
```

```
In [5]: print(titanic['Age'].mean())

titanic['Age'] = titanic['Age'].fillna(titanic['Age'].mean())

30.272590361445783
```

In []:

In [6]: `titanic['Fare'] = titanic['Fare'].fillna(titanic['Fare'].mean())`

In [7]: `titanic.isna().sum()`

Out[7]:

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	0
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	327
Embarked	0
dtype:	int64

```
In [8]: titanic
```

Out[8]:

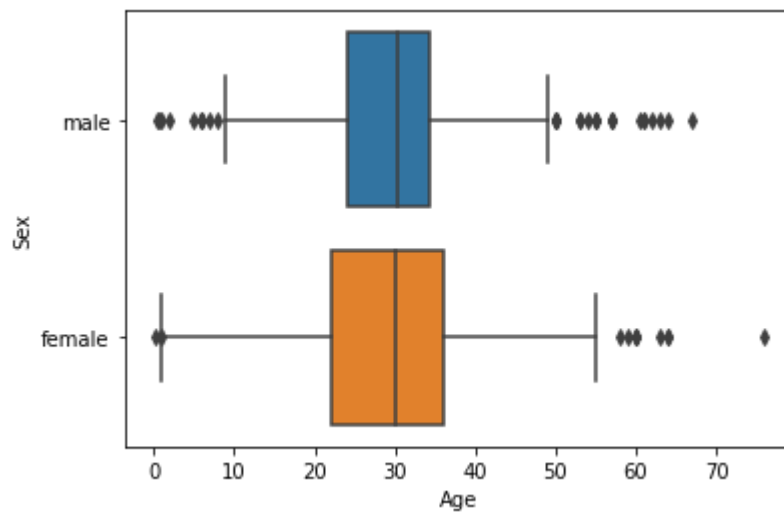
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	892	0	3	Kelly, Mr. James	male	34.50000	0	0	330911
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.00000	1	0	363272
2	894	0	2	Myles, Mr. Thomas Francis	male	62.00000	0	0	240276
3	895	0	3	Wirz, Mr. Albert	male	27.00000	0	0	315154
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.00000	1	1	3101298
...
413	1305	0	3	Spector, Mr. Woolf	male	30.27259	0	0	A.5. 3236
414	1306	1	1	Oliva y Ocana, Dona. Fermina	female	39.00000	0	0	PC 17758
415	1307	0	3	Saether, Mr. Simon Sivertsen	male	38.50000	0	0	SOTON/O.Q. 3101262
416	1308	0	3	Ware, Mr. Frederick	male	30.27259	0	0	359309
417	1309	0	3	Peter, Master. Michael J	male	30.27259	1	1	2668

418 rows × 12 columns



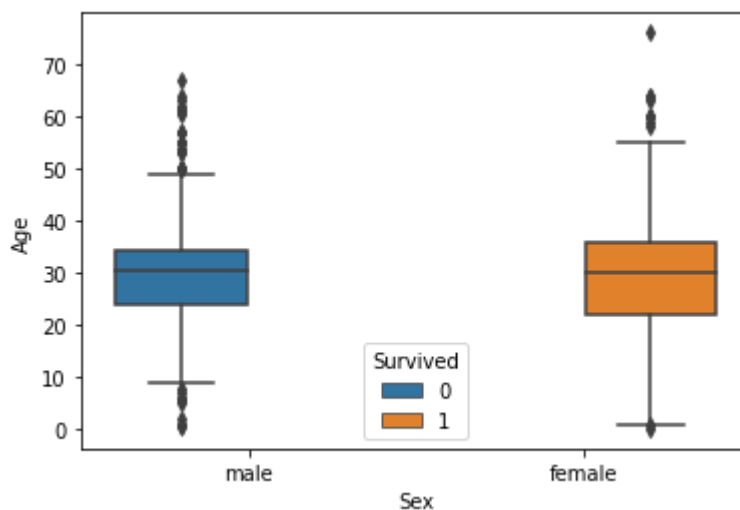
```
In [9]: sns.boxplot(data=titanic,x='Age',y='Sex')
```

```
Out[9]: <AxesSubplot:xlabel='Age', ylabel='Sex'>
```



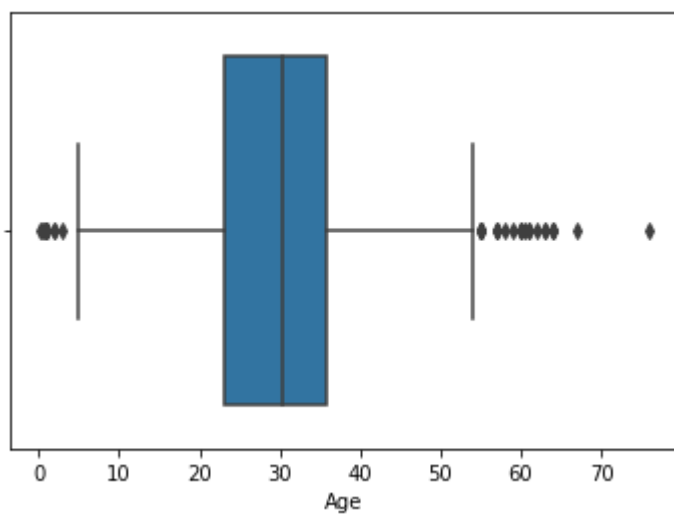
```
In [10]: sns.boxplot(data=titanic,y='Age',x='Sex',hue='Survived')
```

```
Out[10]: <AxesSubplot:xlabel='Sex', ylabel='Age'>
```



```
In [11]: sns.boxplot(data=titanic,x='Age')
```

```
Out[11]: <AxesSubplot:xlabel='Age'>
```



```
In [12]: titanic['Age']
```

```
Out[12]: 0      34.50000
1      47.00000
2      62.00000
3      27.00000
4      22.00000
...
413    30.27259
414    39.00000
415    38.50000
416    30.27259
417    30.27259
Name: Age, Length: 418, dtype: float64
```

```
In [13]: mn = ny.mean(titanic['Age'])
print(mn)
stdd = ny.std(titanic['Age'])
threshold = 0.8
th = -0.8
outlierhandler = []
for i in titanic['Age']:
    z = (i-mn)/stdd
    # print(z)
    if z>0 and z >= threshold :
        outlierhandler.append(0);
    elif z<0 and z<= th:
        outlierhandler.append(0);
    else:
        outlierhandler.append(i);

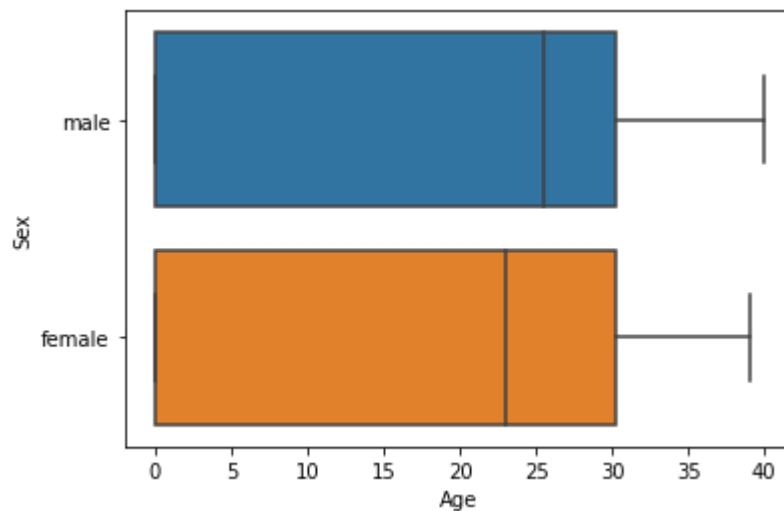
titanic['Age'] = outlierhandler;

30.272590361445783
```

```
In [14]: # titanic['Age']
```

```
In [16]: sns.boxplot(data=titanic,x='Age',y='Sex')
```

```
Out[16]: <AxesSubplot:xlabel='Age', ylabel='Sex'>
```



```
In [ ]:
```