**1) import the library and load the dataset** import seaborn as sns data = sns.load_dataset("titanic")

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data=sns.load_dataset("titanic")
data
```

```
     survived  pclass     sex   age  ...  deck  embark_town  alive
alone
0           0       3    male  22.0  ...   NaN  Southampton     no
False
1           1       1  female  38.0  ...     C    Cherbourg    yes
False
2           1       3  female  26.0  ...   NaN  Southampton    yes
True
3           1       1  female  35.0  ...     C  Southampton    yes
False
4           0       3    male  35.0  ...   NaN  Southampton     no
True
..        ...     ...     ...   ...  ...   ...          ...    ...
...
886         0       2    male  27.0  ...   NaN  Southampton     no
True
887         1       1  female  19.0  ...     B  Southampton    yes
True
888         0       3  female   NaN  ...   NaN  Southampton     no
False
889         1       1    male  26.0  ...     C    Cherbourg    yes
True
890         0       3    male  32.0  ...   NaN   Queenstown     no
True

[891 rows x 15 columns]
```

**2) Deal with Missing values, Outliers and remove unwanted columns**

```
data.isnull().sum()
```

```
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
```

```
data.describe()
```

```
         survived      pclass         age       sibsp       parch
fare
count  891.000000  891.000000  714.000000  891.000000  891.000000
891.000000
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
```

```
data['age'].fillna(data['age'].mean(),inplace = True)
```

```
data['age'].head()
```

```
0    22.0
1    38.0
2    26.0
3    35.0
4    35.0
Name: age, dtype: float64
```

```
data['embarked'].fillna(data['embarked'].mode(),inplace = True)
data['embarked'].head()
```

```
0    S
1    C
2    S
3    S
4    S
Name: embarked, dtype: object
```

```
data['deck'].mode()[0]
```

```
{"type":"string"}
```

```
data['deck'].fillna(data['deck'].mode()[0],inplace = True)
data['deck'].head()
```

```
0    C
1    C
2    C
3    C
4    C
Name: deck, dtype: category
Categories (7, object): ['A', 'B', 'C', 'D', 'E', 'F', 'G']
```

```
data['embark_town'].mode()[0]
```

```
{"type":"string"}
```

```
data['embark_town'].fillna(data['embark_town'].mode()[0],inplace =
True)
```

```
data['embark_town'].head()
```

```
0    Southampton
1      Cherbourg
2    Southampton
3    Southampton
4    Southampton
Name: embark_town, dtype: object
```

**3.Perform exploratory Data analysis**

**1) Perform Descriptive statistics**

**2) Do Data visualisation using seaborn and matplotlib(10-15 visualizations)**

```
import matplotlib.pyplot as plt
import seaborn as sns
data = sns.load_dataset("titanic")
data.head()
```

```
   survived  pclass     sex   age  ...  deck  embark_town  alive
alone
0         0       3    male  22.0  ...   NaN  Southampton     no
False
1         1       1  female  38.0  ...     C    Cherbourg    yes
False
2         1       3  female  26.0  ...   NaN  Southampton    yes
True
3         1       1  female  35.0  ...     C  Southampton    yes
False
4         0       3    male  35.0  ...   NaN  Southampton     no
True

[5 rows x 15 columns]
```

```
data.fare.mean()
```

```
32.2042079685746
```

```
data.fare.mode()
```

```
0    8.05
dtype: float64
```

```
data.fare.median()
```

```
14.4542
```

```
data.age.mean()
```

```
29.69911764705882
```

```
data.age.mode()
```

```
0    24.0
dtype: float64
```

```
data.age.median()
```

```
28.0
```

```
data.sibsp.mean()
```

```
0.5230078563411896
```

```
data.embarked.mode()
```

```
0    S
dtype: object
```

```
data.embark_town.mode()
```

```
0    Southampton
dtype: object
```

```
data.deck.mode()
```

```
0    C
Name: deck, dtype: category
Categories (7, object): ['A', 'B', 'C', 'D', 'E', 'F', 'G']
```

```
data.who.mode()
```

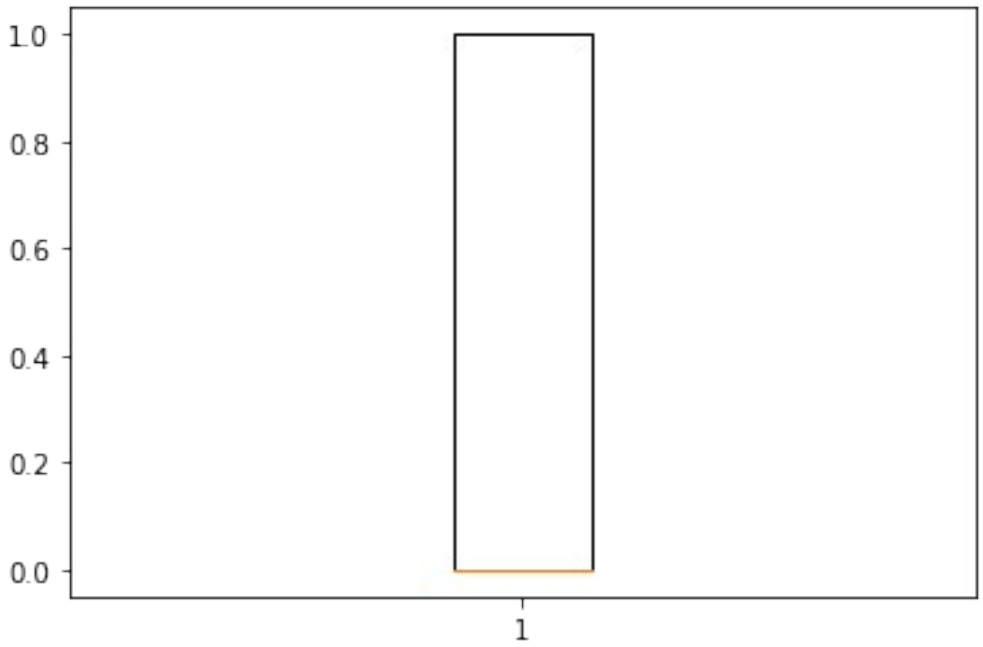
```
0    man
dtype: object
```

```
data.describe()
```

```
         survived      pclass         age       sibsp       parch
fare
count  891.000000  891.000000  714.000000  891.000000  891.000000
891.000000
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
```

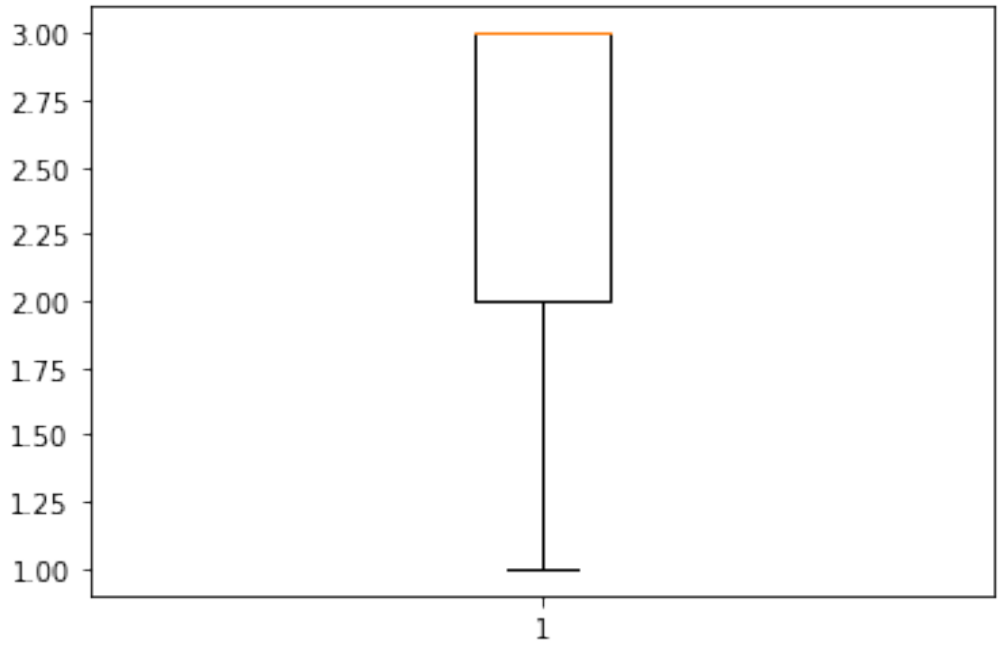
```
plt.boxplot(data.survived)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b83ab50>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b7cc710>,
  <matplotlib.lines.Line2D at 0x7efe5b7d9490>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5ba2acd0>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b7d9650>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5ba2fb90>,
  <matplotlib.lines.Line2D at 0x7efe5b7ccd50>]}
```



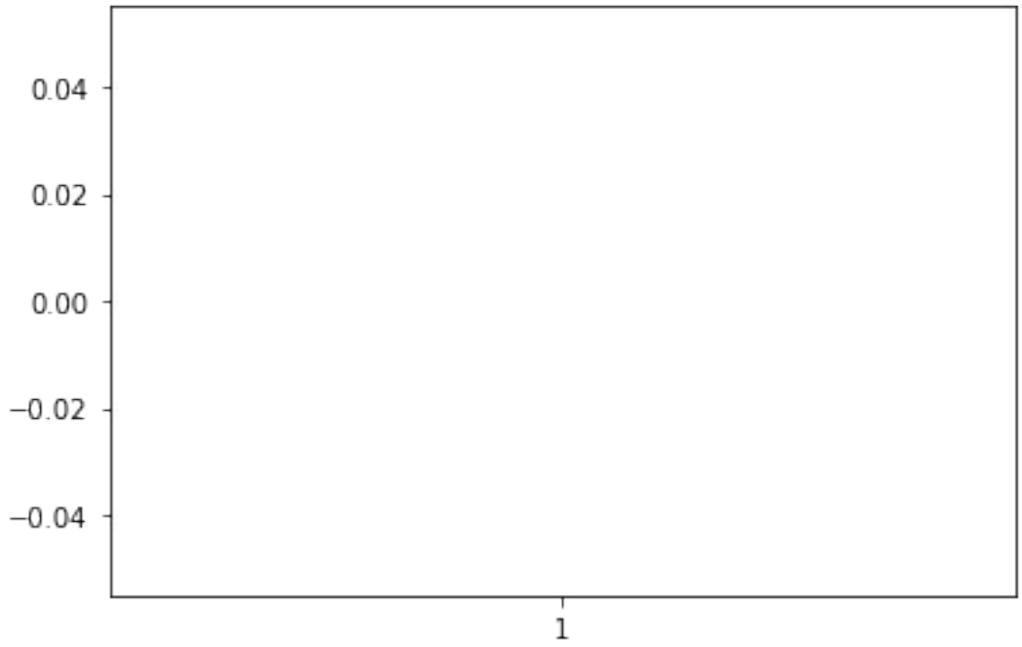
```
plt.boxplot(data.pclass)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b1bda10>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b1c0a50>,
  <matplotlib.lines.Line2D at 0x7efe5b1c0f90>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5b1c7a90>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b1c7550>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5b1bdf90>,
  <matplotlib.lines.Line2D at 0x7efe5b1c0510>]}
```



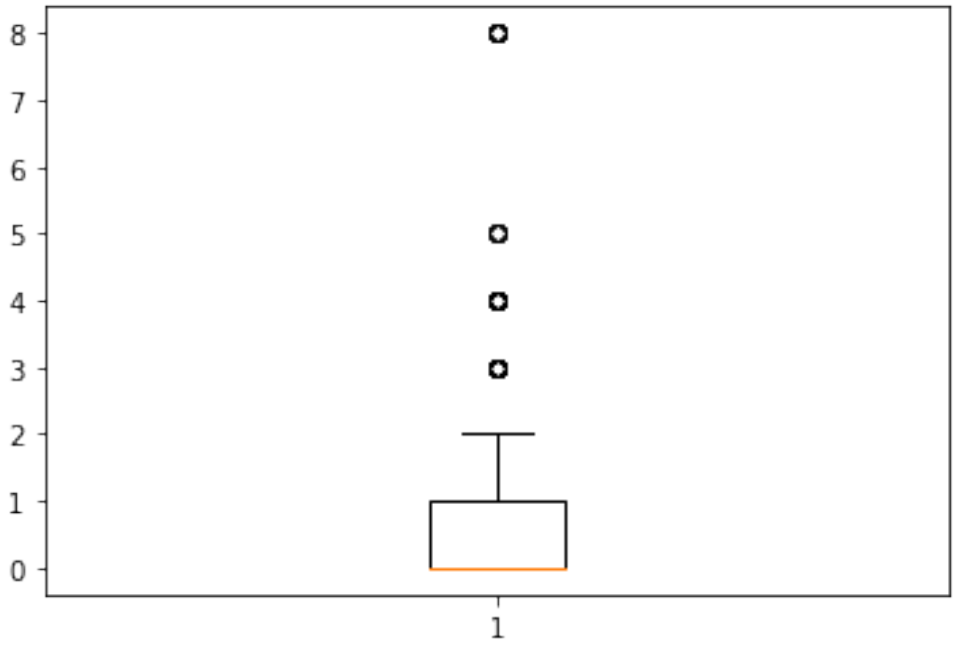

```
plt.boxplot(data.age)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b1b4f50>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b13df90>,
  <matplotlib.lines.Line2D at 0x7efe5b145510>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5b145fd0>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b145a90>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5b13d510>,
  <matplotlib.lines.Line2D at 0x7efe5b13da50>]}
```



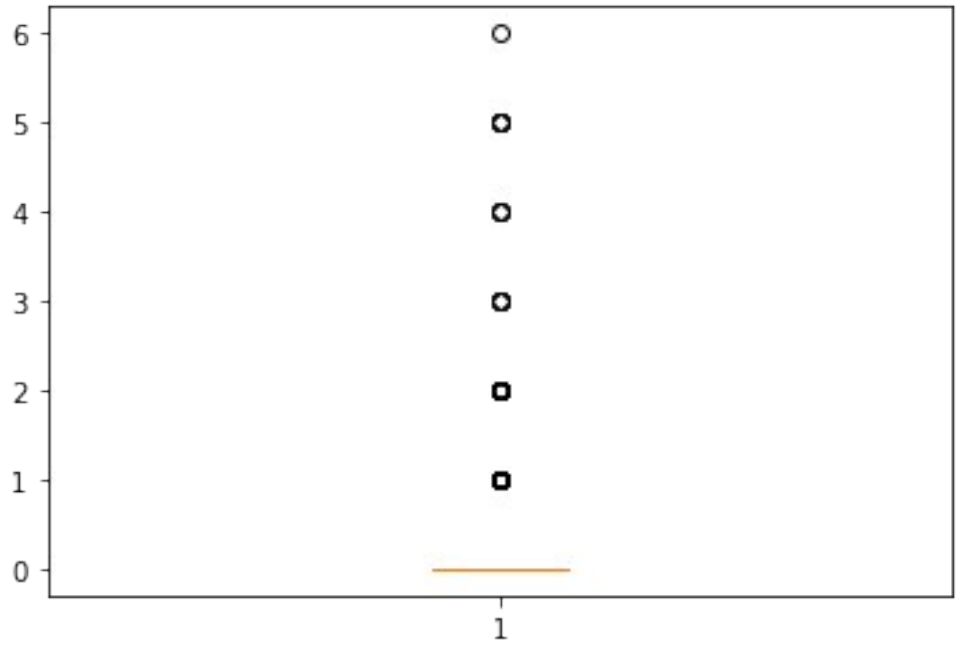

```
plt.boxplot(data.sibsp)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b1294d0>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b130510>,
  <matplotlib.lines.Line2D at 0x7efe5b130a50>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5b137550>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b130fd0>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5b129a50>,
  <matplotlib.lines.Line2D at 0x7efe5b129f90>]}
```



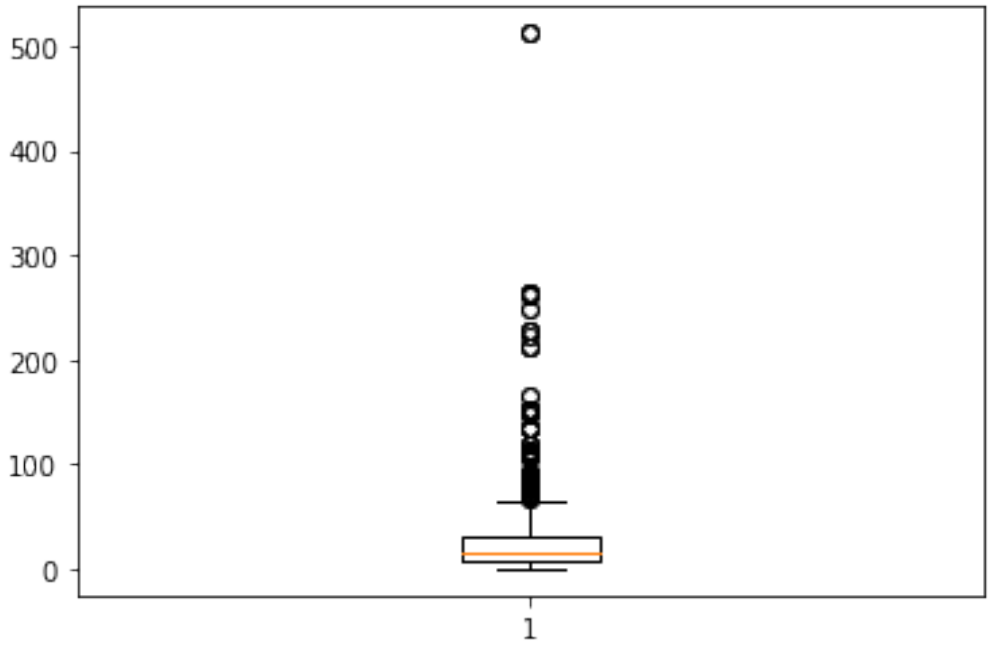
```
plt.boxplot(data.parch)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b0a43d0>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b0a9410>,
  <matplotlib.lines.Line2D at 0x7efe5b0a9950>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5b0b2450>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b0a9ed0>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5b0a4950>,
  <matplotlib.lines.Line2D at 0x7efe5b0a4e90>]}
```



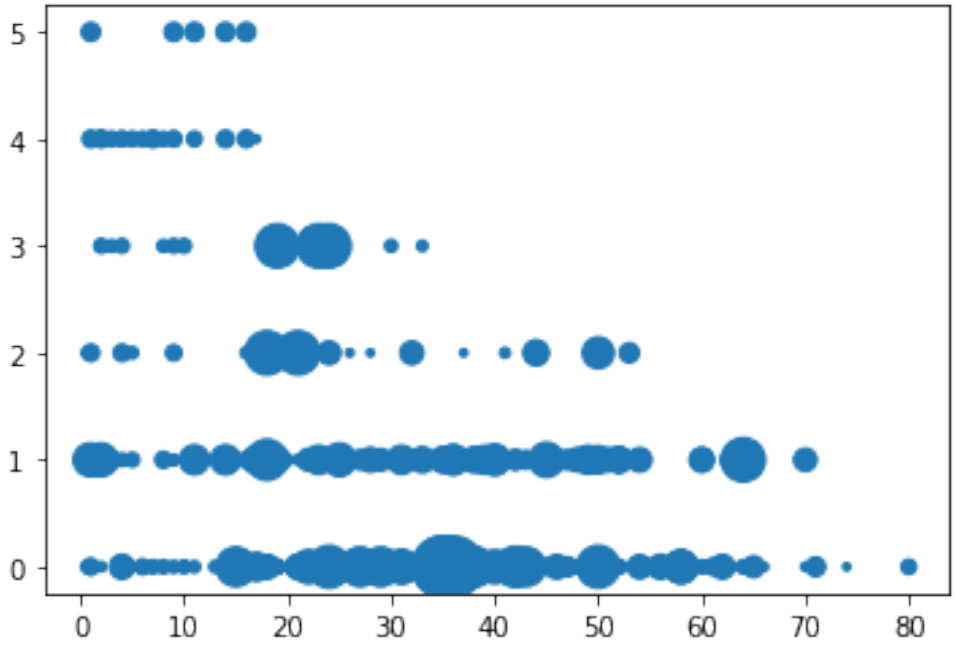

```
plt.boxplot(data.fare)
```

```
{'boxes': [<matplotlib.lines.Line2D at 0x7efe5b00f290>],
 'caps': [<matplotlib.lines.Line2D at 0x7efe5b0192d0>,
  <matplotlib.lines.Line2D at 0x7efe5b019810>],
 'fliers': [<matplotlib.lines.Line2D at 0x7efe5b01f310>],
 'means': [],
 'medians': [<matplotlib.lines.Line2D at 0x7efe5b019d90>],
 'whiskers': [<matplotlib.lines.Line2D at 0x7efe5b00f810>,
  <matplotlib.lines.Line2D at 0x7efe5b00fd50>]}
```



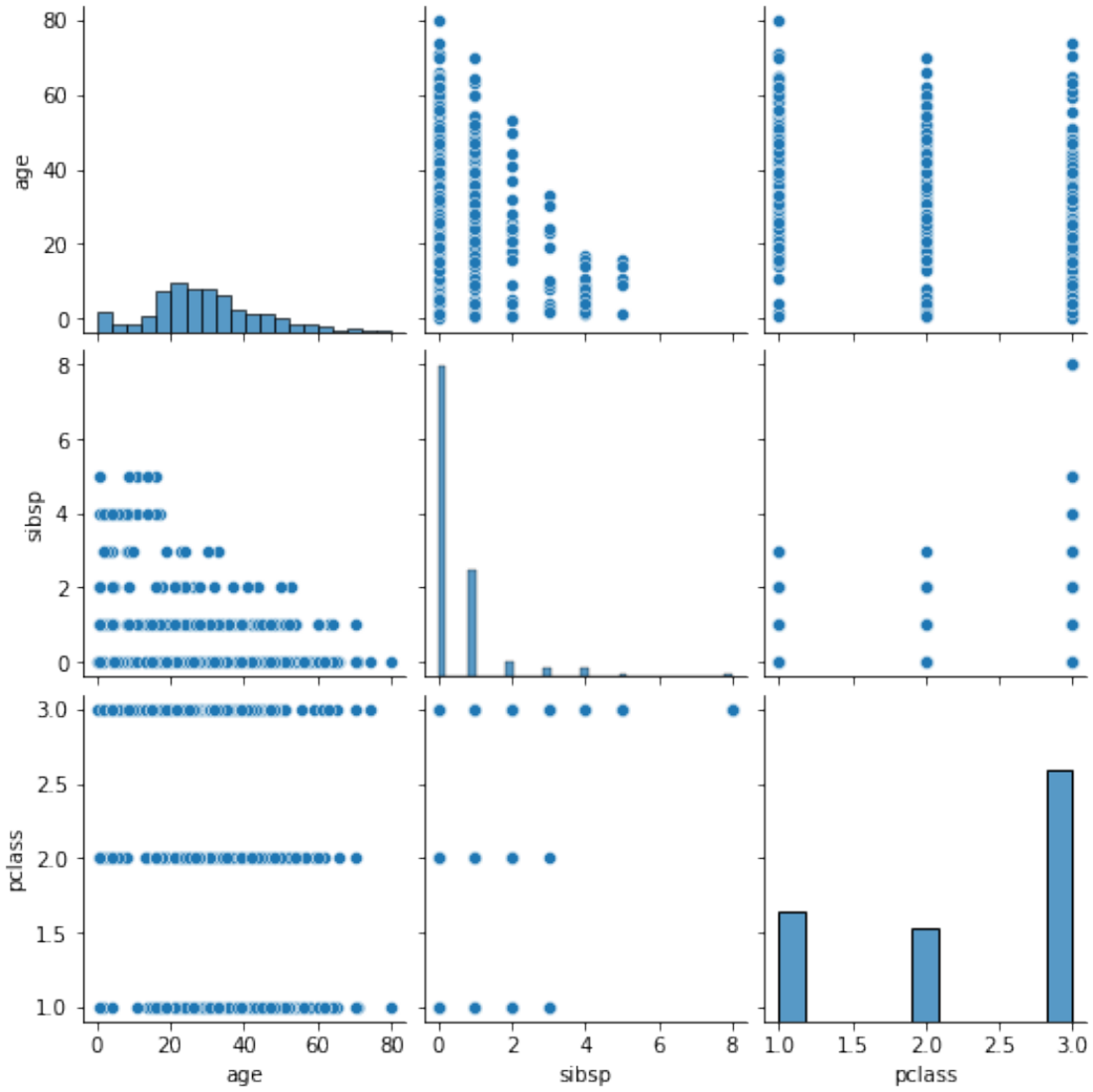

```
plt.scatter(data.age,data.sibsp,data.fare)
```

```
<matplotlib.collections.PathCollection at 0x7efe5afaac10>
```




```
sns.pairplot(data[['age','sibsp','pclass']])
```

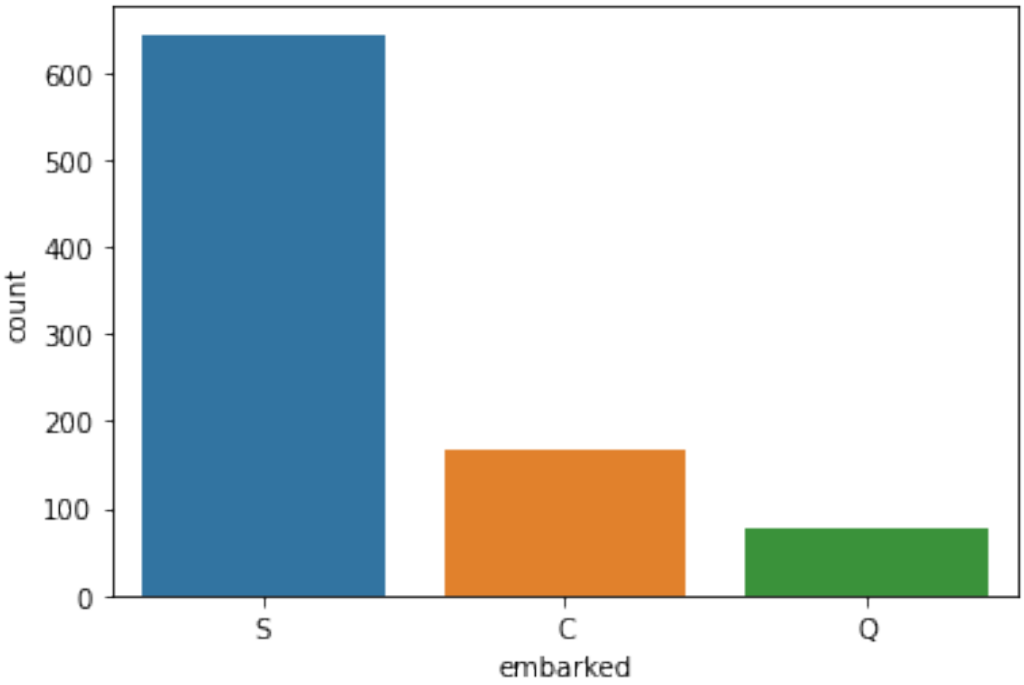
```
<seaborn.axisgrid.PairGrid at 0x7efe5afaafd0>
```




```
sns.countplot(data['embarked'])
```

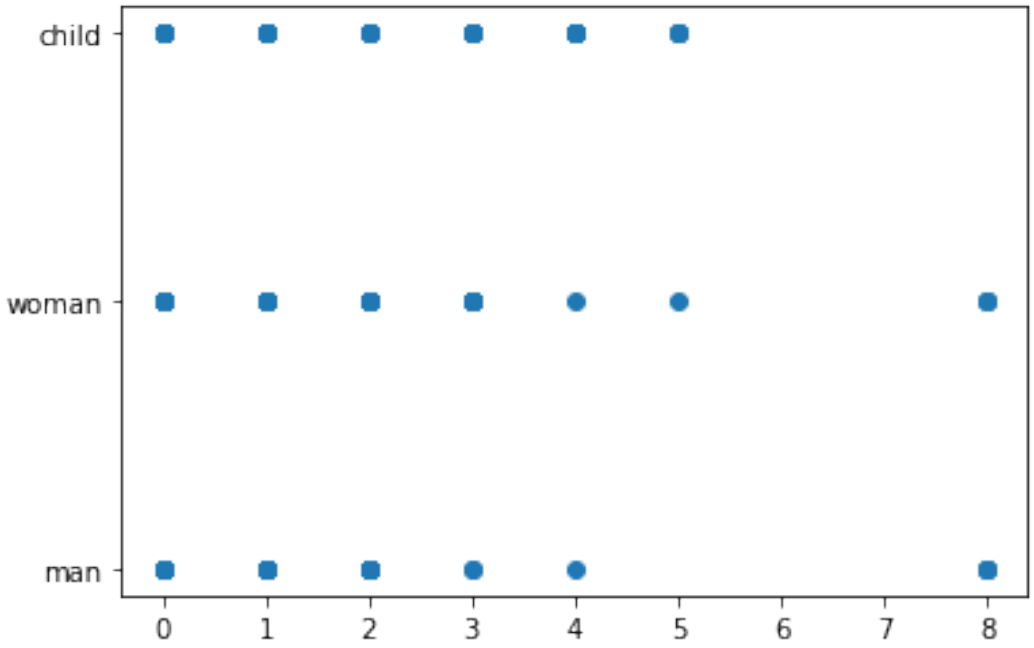
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7efe5b009d50>
```




```
plt.scatter(data.sibsp,data.who)
```

```
<matplotlib.collections.PathCollection at 0x7efe529d7ed0>
```




```
fdt_c = data["embarked"].value_counts()
```

```
fdt_c
```

```
S    644
C    168
Q     77
Name: embarked, dtype: int64
```

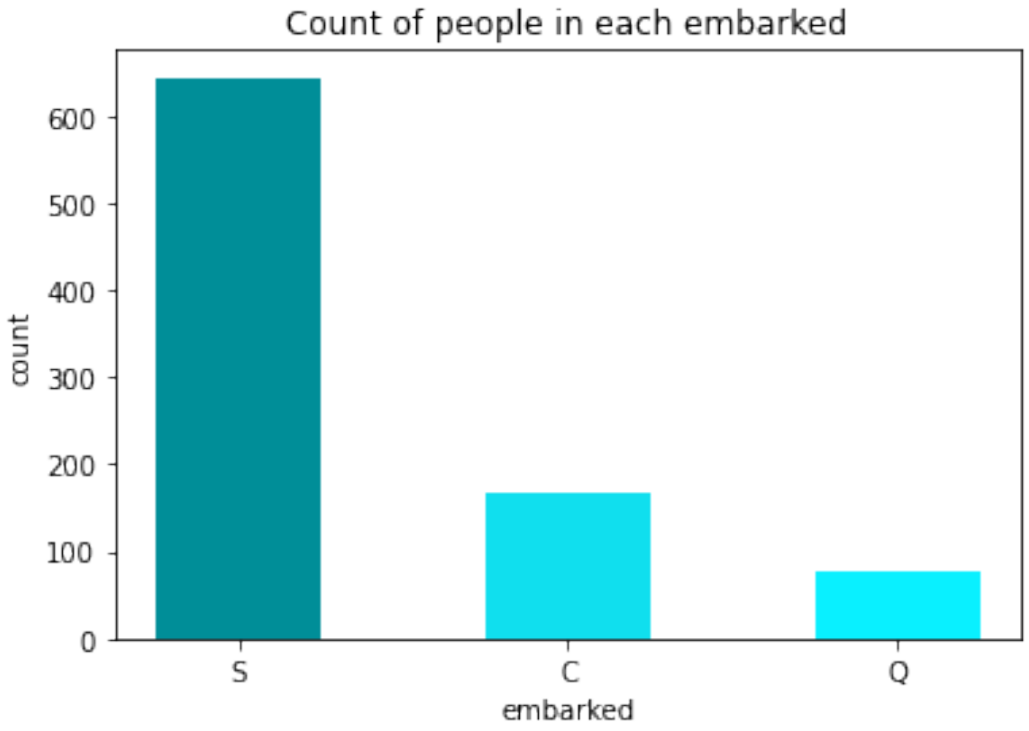
```
fdt_c.index
```

```
Index(['S', 'C', 'Q'], dtype='object')
```

```
fdt_c.values
```

```
array([644, 168,  77])
```

```
plt.bar(fdt_c.index, fdt_c.values, color = ["#008E98", "#10DFEE",
"#09F0FF"], width = 0.5)
plt.xlabel("embarked")
plt.ylabel("count")
plt.title("Count of people in each embarked")
plt.show()
```




```
fdt_c
```

```
S    644
C    168
Q     77
Name: embarked, dtype: int64
```

```
len(fdt_c.sort_values(ascending=False))
```

3

```
fdt_c = fdt_c.sample(frac = 1)
```

```
import matplotlib.pyplot as plt
labels=['embark_town','embarked','sibsp','fare']
sizes = [15, 30, 45, 10]
explode = (0, 0.1, 0, 0)  # only "explode" the 2nd slice (i.e. 'Hogs')

fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=90)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a
circle.

plt.show()
```