assignment-8

August 28, 2023

[53]:
```
import pandas as pd
```

[54]:
```
df = pd.read_excel('/content/titanic3.xls')
```

[55]:
```
#1
df.head()
```

[55]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0000	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON

[56]:
```
#2
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pclass     1309 non-null   int64
 1   survived   1309 non-null   int64
```

```
 2   name        1309 non-null   object
 3   sex         1309 non-null   object
 4   age         1046 non-null   float64
 5   sibsp       1309 non-null   int64
 6   parch       1309 non-null   int64
 7   ticket      1309 non-null   object
 8   fare        1308 non-null   float64
 9   cabin       295 non-null    object
 10  embarked    1307 non-null   object
 11  boat        486 non-null    object
 12  body        121 non-null    float64
 13  home.dest   745 non-null    object
dtypes: float64(3), int64(4), object(7)
memory usage: 143.3+ KB
```

[57]:
```
df.dtypes.isnull()
```

[57]:
```
pclass       False
survived     False
name         False
sex          False
age          False
sibsp        False
parch        False
ticket       False
fare         False
cabin        False
embarked     False
boat         False
body         False
home.dest    False
dtype: bool
```

[58]:
```
#3
df.describe()
```

[58]:

	pclass	survived	age	sibsp	parch
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000
mean	2.294882	0.381971	29.881135	0.498854	0.385027
std	0.837836	0.486055	14.413500	1.041658	0.865560
min	1.000000	0.000000	0.166700	0.000000	0.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000

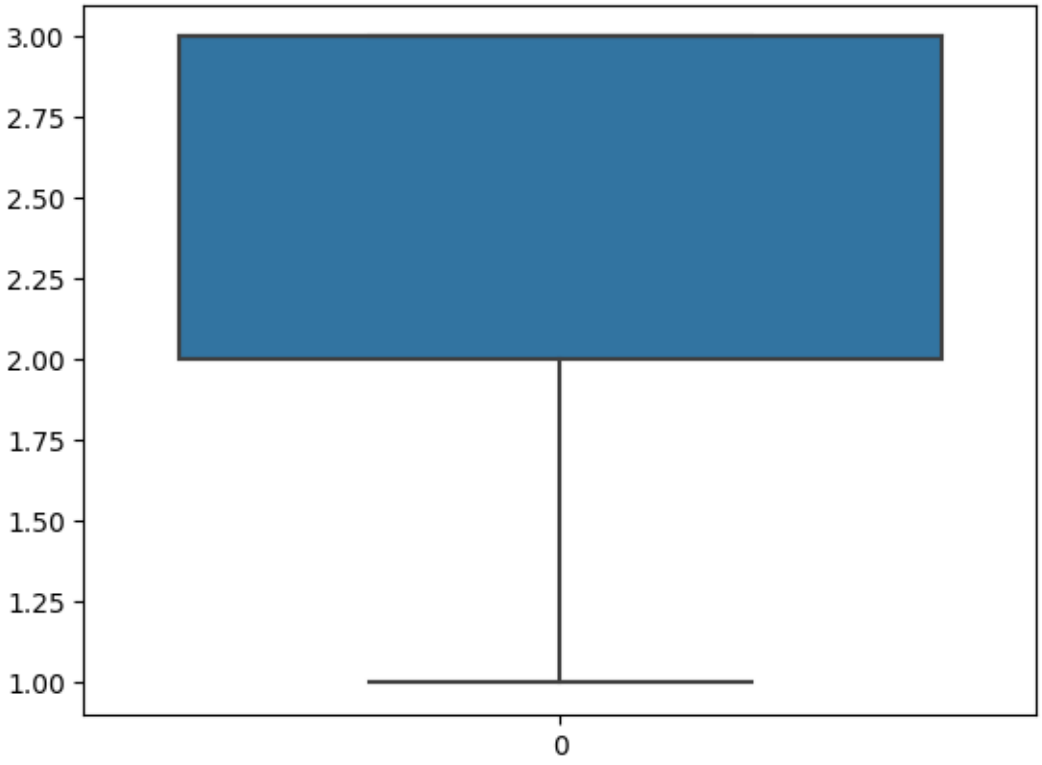
	fare	body

```
count  1308.000000  121.000000
mean     33.295479  160.809917
std      51.758668   97.696922
min       0.000000    1.000000
25%       7.895800   72.000000
50%      14.454200  155.000000
75%      31.275000  256.000000
max     512.329200  328.000000
```

[59]:
```
import numpy as np
import seaborn as sns
```

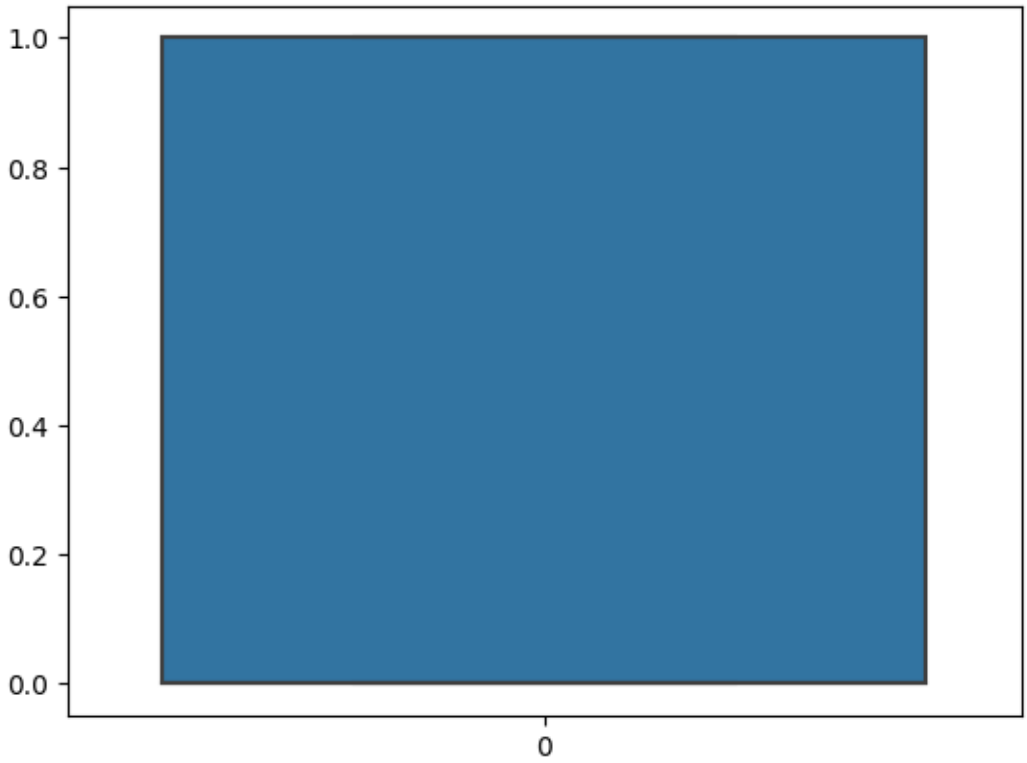
[60]:
```
sns.boxplot(df['pclass']) # there are no outliers in pclass
```

[60]: <Axes: >



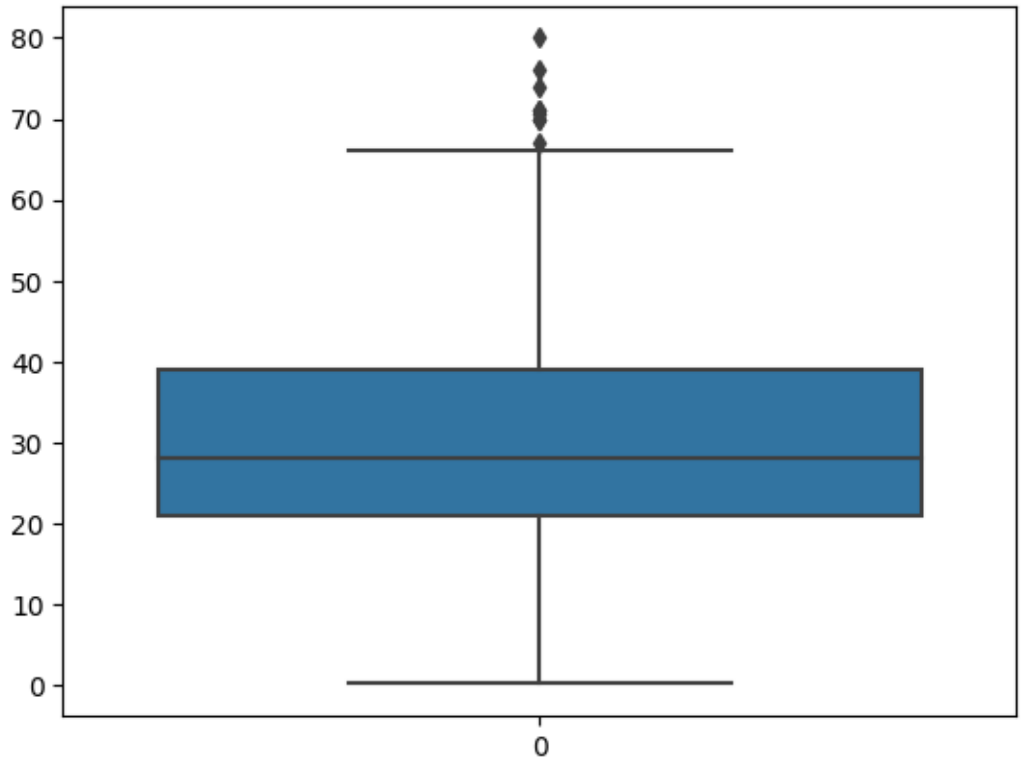
[61]:
```
sns.boxplot(df["survived"]) #No outliers in this class
```

[61]: <Axes: >



[62]:
```
sns.boxplot(df['age']) #Yes there are outliers
```

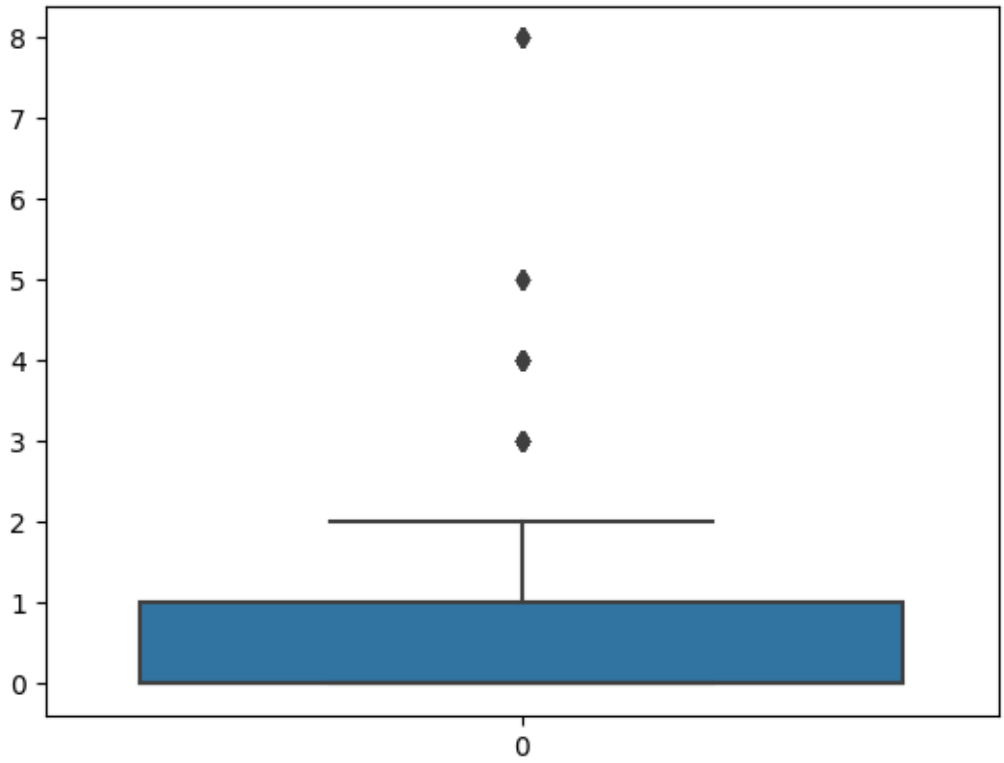
[62]: <Axes: >



[62]:

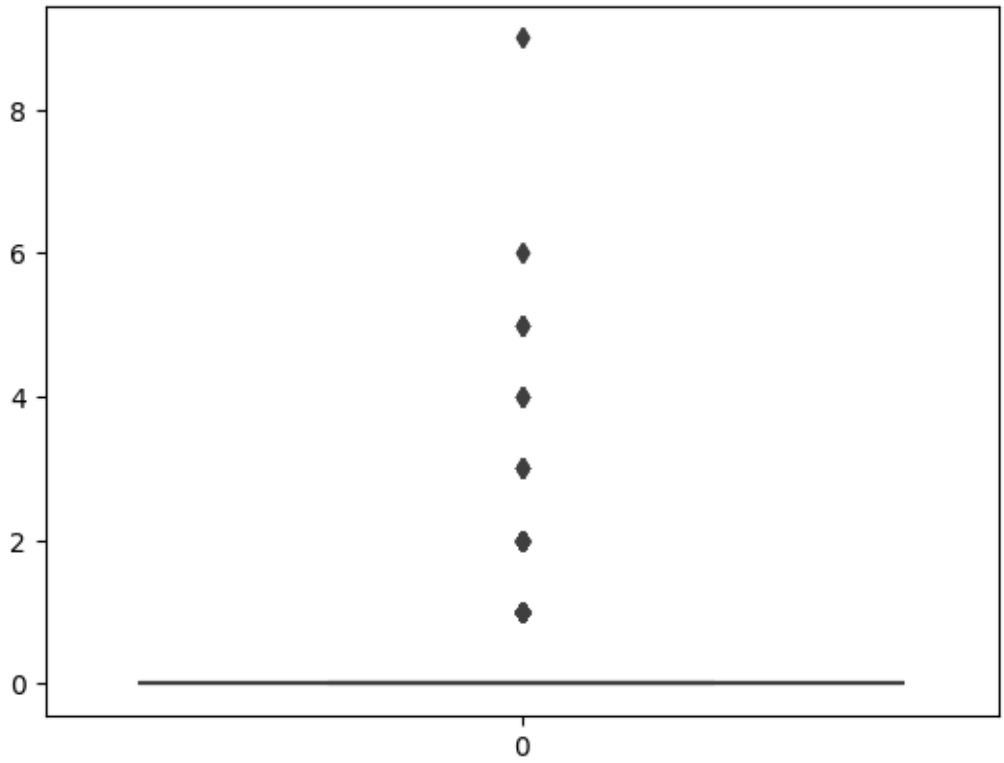
[63]:
```
sns.boxplot(df['sibsp']) #Yes there are outliers
```

[63]: <Axes: >



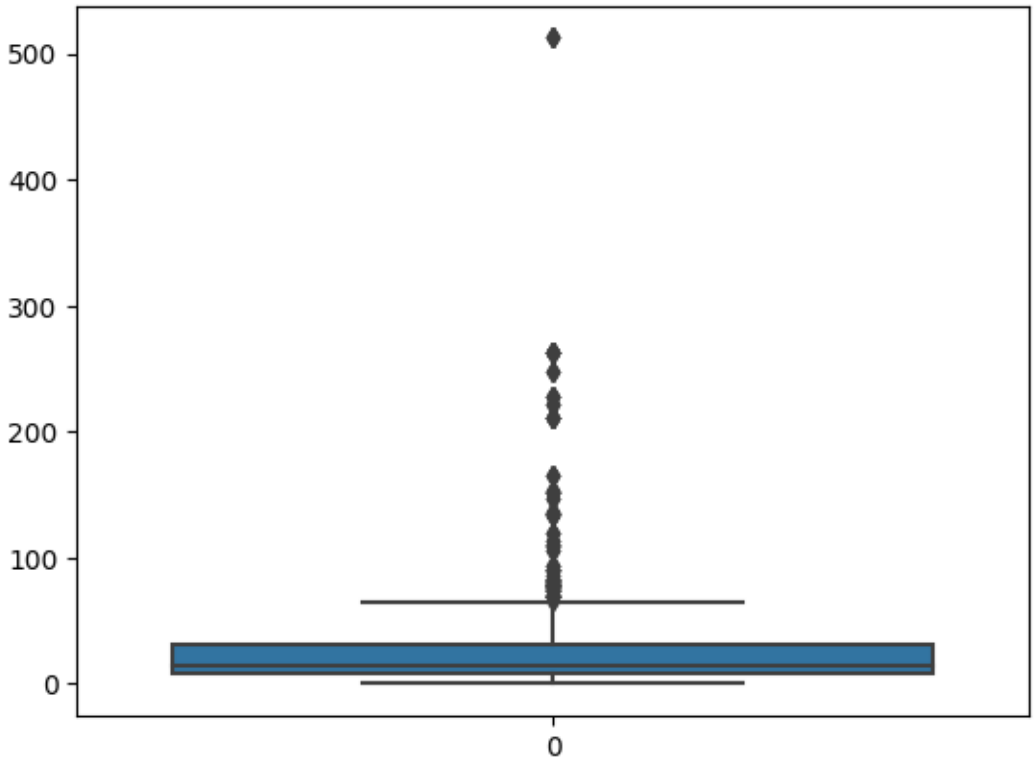
[64]:
```
sns.boxplot(df["parch"]) #yes there are outliers
```

[64]: <Axes: >



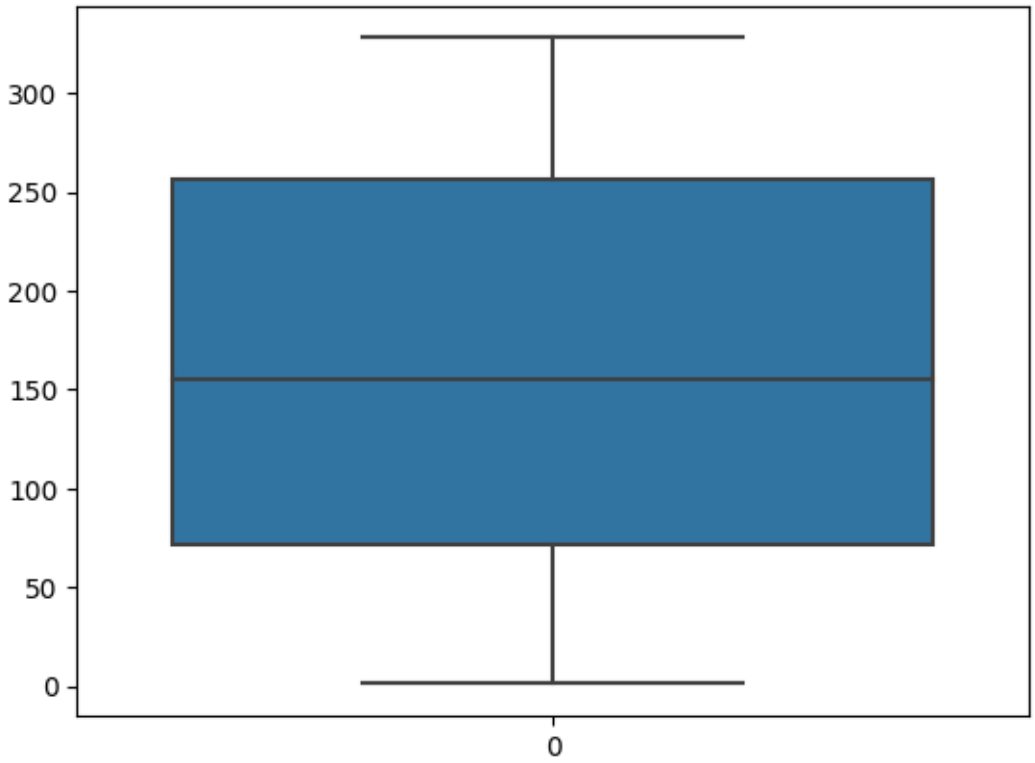
[65]:
```
sns.boxplot(df["fare"]) #Yes there are outliers
```

[65]: <Axes: >



[66]:
```
sns.boxplot(df["body"]) #There are no outliers
```

[66]: <Axes: >



[67]:
```
#4
df.head()
```

[67]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0000	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON

```
4  Montreal, PQ / Chesterville, ON
```

[68]:
```
df['pclass'].groupby(df['fare'])
```

[68]:
```
<pandas.core.groupby.generic.SeriesGroupBy object at 0x7c16abde4e50>
```

[69]:
```
df.groupby(['pclass']).agg({'fare':'mean'})
```

[69]:
```
              fare
pclass
1       87.508992
2       21.179196
3       13.302889
```

[70]:
```
#5
df.columns
```

[70]:
```
Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticket',
       'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

[71]:
```
df['pclass'].value_counts(ascending=True)
```

[71]:
```
2    277
1    323
3    709
Name: pclass, dtype: int64
```

[72]:
```
df['pclass'].value_counts(normalize=True)
```

[72]:
```
3    0.541635
1    0.246753
2    0.211612
Name: pclass, dtype: float64
```

[73]:
```
df['survived'].value_counts()
```

[73]:
```
0    809
1    500
Name: survived, dtype: int64
```

[74]:
```
df['survived'].value_counts(normalize=True)
```

[74]:
```
0    0.618029
1    0.381971
Name: survived, dtype: float64
```

[75]:
```
df['name'].value_counts()
```

[75]:
```
Connolly, Miss. Kate             2
Kelly, Mr. James                 2
Allen, Miss. Elisabeth Walton    1
Ilmakangas, Miss. Ida Livija     1
Ilieff, Mr. Ylio                 1
                                ..
Hart, Miss. Eva Miriam           1
Harris, Mr. Walter               1
Harris, Mr. George               1
Harper, Rev. John                1
Zimmerman, Mr. Leo               1
Name: name, Length: 1307, dtype: int64
```

[76]:
```
df['name'].value_counts(normalize=True)
```

[76]:
```
Connolly, Miss. Kate             0.001528
Kelly, Mr. James                 0.001528
Allen, Miss. Elisabeth Walton    0.000764
Ilmakangas, Miss. Ida Livija     0.000764
Ilieff, Mr. Ylio                 0.000764
                                    …
Hart, Miss. Eva Miriam           0.000764
Harris, Mr. Walter               0.000764
Harris, Mr. George               0.000764
Harper, Rev. John                0.000764
Zimmerman, Mr. Leo               0.000764
Name: name, Length: 1307, dtype: float64
```

[77]:
```
df['sex'].value_counts(ascending=True)
```

[77]:
```
female    466
male      843
Name: sex, dtype: int64
```

[78]:
```
df['sex'].value_counts(normalize=True)
```

[78]:
```
male      0.644003
female    0.355997
Name: sex, dtype: float64
```

[79]:
```
df['age'].value_counts(ascending=True)
```

[79]:
```
26.5     1
66.0     1
67.0     1
```

```
70.5     1
22.5     1
        ..
18.0    39
30.0    40
21.0    41
22.0    43
24.0    47
Name: age, Length: 98, dtype: int64
```

[80]:
```
df['age'].value_counts(normalize=True)
```

[80]:
```
24.0000    0.044933
22.0000    0.041109
21.0000    0.039197
30.0000    0.038241
18.0000    0.037285
              …
0.3333     0.000956
22.5000    0.000956
70.5000    0.000956
0.6667     0.000956
26.5000    0.000956
Name: age, Length: 98, dtype: float64
```

[81]:
```
df['sibsp'].value_counts()
```

[81]:
```
0    891
1    319
2     42
4     22
3     20
8      9
5      6
Name: sibsp, dtype: int64
```

[82]:
```
df['sibsp'].value_counts(normalize=True)
```

[82]:
```
0    0.680672
1    0.243697
2    0.032086
4    0.016807
3    0.015279
8    0.006875
5    0.004584
Name: sibsp, dtype: float64
```

[83]:
```
df.columns
```

[83]:
```
Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticket',
       'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

[84]:
```
df['parch'].value_counts()
```

[84]:
```
0    1002
1     170
2     113
3       8
4       6
5       6
6       2
9       2
Name: parch, dtype: int64
```

[85]:
```
df['parch'].value_counts(normalize=True)
```

[85]:
```
0    0.765470
1    0.129870
2    0.086325
3    0.006112
4    0.004584
5    0.004584
6    0.001528
9    0.001528
Name: parch, dtype: float64
```

[86]:
```
df['ticket'].value_counts()
```

[86]:
```
CA. 2343        11
CA 2144          8
1601             8
S.O.C. 14879     7
3101295          7
                ..
C 7076           1
341826           1
7546             1
3474             1
315082           1
Name: ticket, Length: 939, dtype: int64
```

[87]:
```
df['ticket'].value_counts(normalize=True)
```

[87]:
```
CA. 2343         0.008403
CA 2144          0.006112
1601             0.006112
S.O.C. 14879     0.005348
3101295          0.005348
                   …
C 7076           0.000764
341826           0.000764
7546             0.000764
3474             0.000764
315082           0.000764
Name: ticket, Length: 939, dtype: float64
```

[88]:
```
df['fare'].value_counts(bins=10)
```

[88]:
```
(-0.513, 51.233]      1070
(51.233, 102.466]      154
(102.466, 153.699]      42
(204.932, 256.165]      21
(256.165, 307.398]      13
(153.699, 204.932]       4
(461.096, 512.329]       4
(307.398, 358.63]        0
(358.63, 409.863]        0
(409.863, 461.096]       0
Name: fare, dtype: int64
```

[89]:
```
df['fare'].value_counts(bins=10,normalize=True)
```

[89]:
```
(-0.513, 51.233]      0.817418
(51.233, 102.466]     0.117647
(102.466, 153.699]    0.032086
(204.932, 256.165]    0.016043
(256.165, 307.398]    0.009931
(153.699, 204.932]    0.003056
(461.096, 512.329]    0.003056
(307.398, 358.63]     0.000000
(358.63, 409.863]     0.000000
(409.863, 461.096]    0.000000
Name: fare, dtype: float64
```

[90]:
```
#Answer
df['embarked'].value_counts(ascending=True)
```

[90]:
```
Q    123
C    270
S    914
```

```
Name: embarked, dtype: int64
```

[91]:
```
df['embarked'].value_counts(normalize=True)
```

[91]:
```
S    0.699311
C    0.206580
Q    0.094109
Name: embarked, dtype: float64
```

[92]:
```
#6
df['age'].value_counts(bins=[0,12,19,59])
```

[92]:
```
(19.0, 59.0]      781
(12.0, 19.0]      131
(-0.001, 12.0]     94
Name: age, dtype: int64
```

[93]:
```
df=pd.read_excel('/content/titanic3.xls')
```

[94]:
```
df
```

[94]:

	pclass	survived	name	\
0	1	1	Allen, Miss. Elisabeth Walton	
1	1	1	Allison, Master. Hudson Trevor	
2	1	0	Allison, Miss. Helen Loraine	
3	1	0	Allison, Mr. Hudson Joshua Creighton	
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	
…	…	…	…	
1304	3	0	Zabour, Miss. Hileni	
1305	3	0	Zabour, Miss. Thamine	
1306	3	0	Zakarian, Mr. Mapriededer	
1307	3	0	Zakarian, Mr. Ortin	
1308	3	0	Zimmerman, Mr. Leo	

	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	\
0	female	29.0000	0	0	24160	211.3375	B5	S	2	
1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11	
2	female	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	
3	male	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	
4	female	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	
…	…	…	…	…	…	…	…	…	…	
1304	female	14.5000	1	0	2665	14.4542	NaN	C	NaN	
1305	female	NaN	1	0	2665	14.4542	NaN	C	NaN	
1306	male	26.5000	0	0	2656	7.2250	NaN	C	NaN	
1307	male	27.0000	0	0	2670	7.2250	NaN	C	NaN	
1308	male	29.0000	0	0	315082	7.8750	NaN	S	NaN	

```
       body                        home.dest
0       NaN                    St Louis, MO
1       NaN  Montreal, PQ / Chesterville, ON
2       NaN  Montreal, PQ / Chesterville, ON
3     135.0  Montreal, PQ / Chesterville, ON
4       NaN  Montreal, PQ / Chesterville, ON
…       …                               …
1304  328.0                             NaN
1305    NaN                             NaN
1306  304.0                             NaN
1307    NaN                             NaN
1308    NaN                             NaN

[1309 rows x 14 columns]
```

[95]:
```
df.loc[df['age']<= 12,"AgeRange"]="Child"
```

[96]:
```
df.head()
```

[96]:
```
   pclass  survived                                             name     sex  \
0       1         1                    Allen, Miss. Elisabeth Walton  female
1       1         1                   Allison, Master. Hudson Trevor    male
2       1         0                     Allison, Miss. Helen Loraine  female
3       1         0             Allison, Mr. Hudson Joshua Creighton    male
4       1         0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)  female

       age  sibsp  parch  ticket      fare    cabin embarked boat   body  \
0  29.0000      0      0   24160  211.3375       B5        S    2    NaN
1   0.9167      1      2  113781  151.5500  C22 C26        S   11    NaN
2   2.0000      1      2  113781  151.5500  C22 C26        S  NaN    NaN
3  30.0000      1      2  113781  151.5500  C22 C26        S  NaN  135.0
4  25.0000      1      2  113781  151.5500  C22 C26        S  NaN    NaN

                        home.dest AgeRange
0                    St Louis, MO      NaN
1  Montreal, PQ / Chesterville, ON    Child
2  Montreal, PQ / Chesterville, ON    Child
3  Montreal, PQ / Chesterville, ON      NaN
4  Montreal, PQ / Chesterville, ON      NaN
```

[97]:
```
df.loc[(df['age'] > 12) & (df['age'] <= 19),'AgeRange']="Teen"
```

[97]:
```
```

[98]:
```
df.loc[(df['age'] > 19) & (df['age'] <= 59),'AgeRange']="Adult"
```

[98]:
```
```

[99]:
```
df.loc[df['age']>59,"AgeRange"]="Senior"
```

[100]:
```
df.head()
```

[100]:
```
   pclass  survived                                             name     sex  \
0       1         1                    Allen, Miss. Elisabeth Walton  female
1       1         1                   Allison, Master. Hudson Trevor    male
2       1         0                     Allison, Miss. Helen Loraine  female
3       1         0             Allison, Mr. Hudson Joshua Creighton    male
4       1         0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)  female

       age  sibsp  parch  ticket      fare    cabin embarked boat   body  \
0  29.0000      0      0   24160  211.3375       B5        S    2    NaN
1   0.9167      1      2  113781  151.5500  C22 C26        S   11    NaN
2   2.0000      1      2  113781  151.5500  C22 C26        S  NaN    NaN
3  30.0000      1      2  113781  151.5500  C22 C26        S  NaN  135.0
4  25.0000      1      2  113781  151.5500  C22 C26        S  NaN    NaN

                        home.dest AgeRange
0                     St Louis, MO    Adult
1  Montreal, PQ / Chesterville, ON    Child
2  Montreal, PQ / Chesterville, ON    Child
3  Montreal, PQ / Chesterville, ON    Adult
4  Montreal, PQ / Chesterville, ON    Adult
```

[101]:
```
df.tail()
```

[101]:
```
      pclass  survived                        name     sex   age  sibsp  parch  \
1304       3         0       Zabour, Miss. Hileni  female  14.5      1      0
1305       3         0      Zabour, Miss. Thamine  female   NaN      1      0
1306       3         0  Zakarian, Mr. Mapriededer    male  26.5      0      0
1307       3         0        Zakarian, Mr. Ortin    male  27.0      0      0
1308       3         0         Zimmerman, Mr. Leo    male  29.0      0      0

      ticket     fare cabin embarked boat   body home.dest AgeRange
1304    2665  14.4542   NaN        C  NaN  328.0       NaN     Teen
1305    2665  14.4542   NaN        C  NaN    NaN       NaN      NaN
1306    2656   7.2250   NaN        C  NaN  304.0       NaN    Adult
1307    2670   7.2250   NaN        C  NaN    NaN       NaN    Adult
1308  315082   7.8750   NaN        S  NaN    NaN       NaN    Adult
```

[102]:
```
from pandas import Series,DataFrame
```

[123]:
```
df['sex'].describe()
```

[123]:
```
count     1309
unique       2
```

```
top       male
freq       843
Name: sex, dtype: object
```

[125]:
```
df['sex'].value_counts()
```

[125]:
```
male      843
female    466
Name: sex, dtype: int64
```

[127]:
```
df['survived'].value_counts()
```

[127]:
```
0    809
1    500
Name: survived, dtype: int64
```

[138]:
```
#7
df.pivot_table(index='sex',columns='survived',values='pclass')
```

[138]:
```
survived         0         1
sex
female    2.826772  1.902655
male      2.439883  2.086957
```

[ ]:
```

```

[139]:
```
df.pivot_table(index='AgeRange',columns='pclass',values='survived')
```

[139]:
```
pclass           1         2         3
AgeRange
Adult     0.646552  0.391089  0.230548
Child     0.800000  1.000000  0.400000
Senior    0.384615  0.125000  0.166667
Teen      0.809524  0.407407  0.289157
```

[ ]:
```

```

[137]:
```
df.pivot_table(index='survived',columns='sex',values='pclass')
```

[137]:
```
sex         female      male
survived
0         2.826772  2.439883
1         1.902655  2.086957
```

[146]:
```
#8
import matplotlib.pyplot as plt
%matplotlib inline
```

```
import numpy as np
```

[185]:
```
s=[]
 for i in df['AgeRange'].unique():
  s.append(i)
k=[]
for i in df['AgeRange'].value_counts():
  k.append(i)
s.pop()
s,k
```

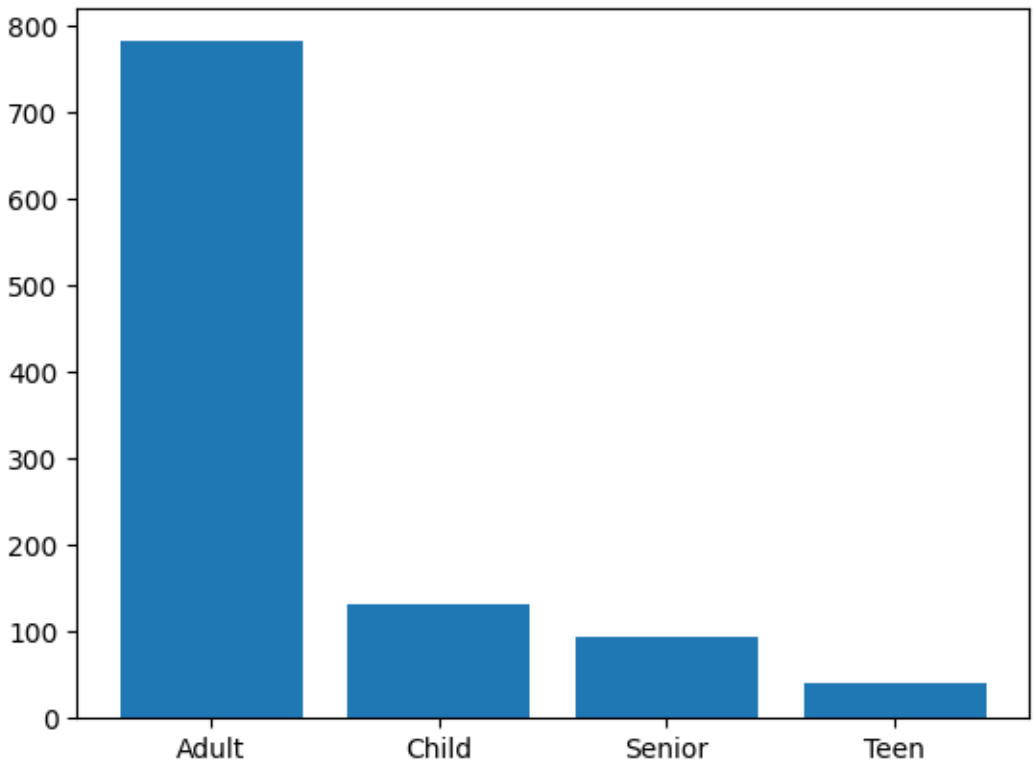
[185]:
```
(['Adult', 'Child', 'Senior', 'Teen'], [781, 131, 94, 40])
```

[188]:
```
ypos=np.arange(len(s))
ypos
```

[188]:
```
array([0, 1, 2, 3])
```

[190]:
```
plt.bar(ypos,k)
plt.xticks(ypos,s)
```

[190]:
```
([<matplotlib.axis.XTick at 0x7c16b5fd4880>,
  <matplotlib.axis.XTick at 0x7c16b5fd49a0>,
  <matplotlib.axis.XTick at 0x7c16abe791b0>,
  <matplotlib.axis.XTick at 0x7c16ab53b0a0>],
 [Text(0, 0, 'Adult'),
  Text(1, 0, 'Child'),
  Text(2, 0, 'Senior'),
  Text(3, 0, 'Teen')])
```



[190]:

[191]:
```
#9
df.head()
```

[191]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	AgeRange
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0000	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO	Adult
1	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON	Child
2	1	0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN		
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1	2	113781	151.5500	C22 C26	S	NaN	135.0		
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1	2	113781	151.5500	C22 C26	S	NaN	NaN		

```
2  Montreal, PQ / Chesterville, ON    Child
3  Montreal, PQ / Chesterville, ON    Adult
4  Montreal, PQ / Chesterville, ON    Adult
```

[209]:

[204]:
```
data=pd.DataFrame(df,columns=['age','fare','survived'])
```

[205]:
```
data
```

[205]:
```
          age      fare  survived
0     29.0000  211.3375         1
1      0.9167  151.5500         1
2      2.0000  151.5500         0
3     30.0000  151.5500         0
4     25.0000  151.5500         0
…        …        …        …
1304  14.5000   14.4542         0
1305      NaN   14.4542         0
1306  26.5000    7.2250         0
1307  27.0000    7.2250         0
1308  29.0000    7.8750         0

[1309 rows x 3 columns]
```

[210]:
```
df['fare'].value_counts()
```

[210]:
```
8.0500     60
13.0000    59
7.7500     55
26.0000    50
7.8958     49
           ..
15.0500     1
9.6875      1
15.5792     1
12.0000     1
7.8750      1
Name: fare, Length: 281, dtype: int64
```

[213]:
```
l=[]
for i in df['fare']:
  l.append(i)
```

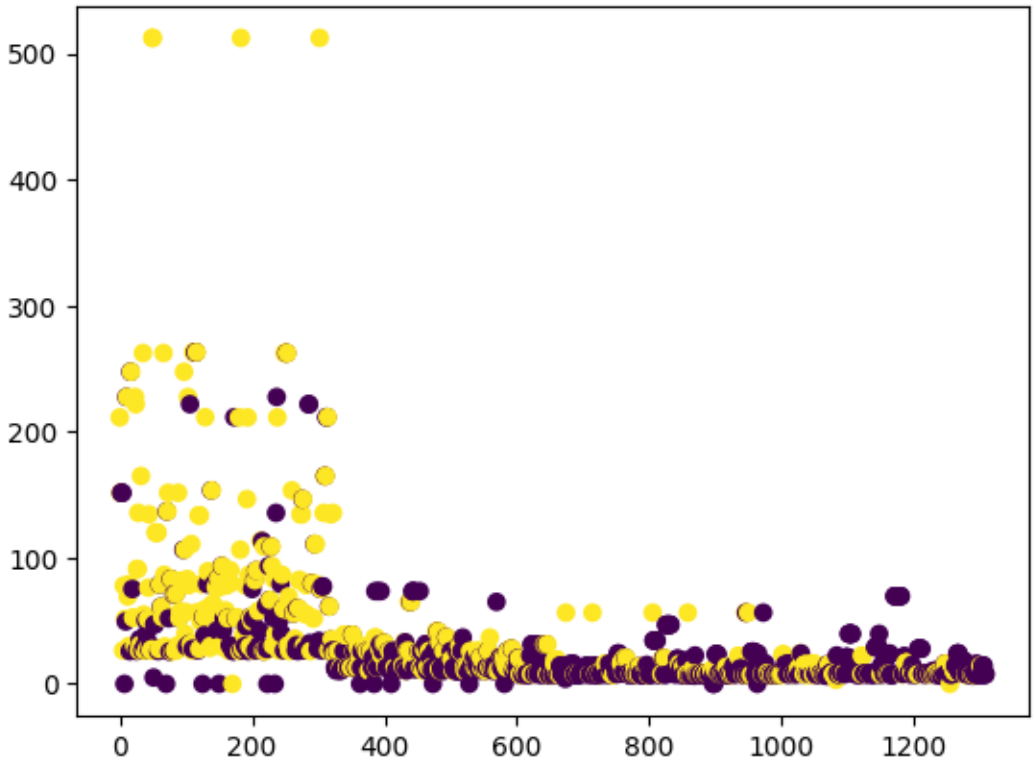
[230]:
```
p=[]
for i in df['age']:
  p.append(i)
```

```
q=[]
for i in df['survived']:
  q.append(i)
```

[231]:
```
ycos=np.arange(len(p))
```

[232]:
```
plt.scatter(ycos,l,c=q)
```

[232]: <matplotlib.collections.PathCollection at 0x7c16aa950310>



[ ]: