28-Ronit_khalate-Assign-9

April 18, 2023

Assignment No 6

Data Visualization II

Use the inbuilt dataset 'titanic' as used in the above problem. Plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not. (Column names : 'sex' and 'age')

Import required libraries

```python
import pandas as pd
import matplotlib.pyplot as plt
```

Load Dataset

```python
df= pd.read_csv("Titanic-Dataset.csv")
df
```

```
[ ]:      PassengerId  Survived  Pclass  \
     0              1         0       3
     1              2         1       1
     2              3         1       3
     3              4         1       1
     4              5         0       3
     ..           ...       ...     ...
     886          887         0       2
     887          888         1       1
     888          889         0       3
     889          890         1       1
     890          891         0       3

                                                      Name     Sex   Age  SibSp  \
     0                              Braund, Mr. Owen Harris    male  22.0      1
     1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
     2                               Heikkinen, Miss. Laina  female  26.0      0
     3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
     4                             Allen, Mr. William Henry    male  35.0      0
     ..                                                 …       …     …      …
     886                                Montvila, Rev. Juozas    male  27.0      0
     887                         Graham, Miss. Margaret Edith  female  19.0      0
     888          Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
```

1

```
889                            Behr, Mr. Karl Howell    male  26.0      0
890                            Dooley, Mr. Patrick      male  32.0      0

      Parch            Ticket     Fare Cabin Embarked
0         0         A/5 21171   7.2500   NaN        S
1         0          PC 17599  71.2833   C85        C
2         0  STON/O2. 3101282   7.9250   NaN        S
3         0            113803  53.1000  C123        S
4         0            373450   8.0500   NaN        S
..      ...               ...      ...   ...      ...
886       0            211536  13.0000   NaN        S
887       0            112053  30.0000   B42        S
888       2        W./C. 6607  23.4500   NaN        S
889       0            111369  30.0000  C148        C
890       0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

Preprocessing data set

```python
required_df = df[["Name","Sex","Age","Survived"]]
required_df
```

```
                                               Name     Sex   Age  Survived
0                            Braund, Mr. Owen Harris    male  22.0         0
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0         1
2                             Heikkinen, Miss. Laina  female  26.0         1
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0         1
4                           Allen, Mr. William Henry    male  35.0         0
..                                              ...     ...   ...       ...
886                            Montvila, Rev. Juozas    male  27.0         0
887                     Graham, Miss. Margaret Edith  female  19.0         1
888          Johnston, Miss. Catherine Helen "Carrie"  female   NaN         0
889                            Behr, Mr. Karl Howell    male  26.0         1
890                            Dooley, Mr. Patrick      male  32.0         0

[891 rows x 4 columns]
```

```python
required_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      891 non-null    object
 1   Sex       891 non-null    object
 2   Age       714 non-null    float64
```

```
 3    Survived   891 non-null     int64
dtypes: float64(1), int64(1), object(2)
memory usage: 28.0+ KB
```

[ ]: `required_df.isnull().sum()`

[ ]:
```
Name         0
Sex          0
Age        177
Survived     0
dtype: int64
```

[ ]:
```
proc_df = required_df.dropna()
proc_df
```

[ ]:

	Name	Sex	Age	Survived
0	Braund, Mr. Owen Harris	male	22.0	0
1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	1
2	Heikkinen, Miss. Laina	female	26.0	1
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4	Allen, Mr. William Henry	male	35.0	0
..	…	…	…	…
885	Rice, Mrs. William (Margaret Norton)	female	39.0	0
886	Montvila, Rev. Juozas	male	27.0	0
887	Graham, Miss. Margaret Edith	female	19.0	1
889	Behr, Mr. Karl Howell	male	26.0	1
890	Dooley, Mr. Patrick	male	32.0	0

```
[714 rows x 4 columns]
```

[ ]: `proc_df.isnull().sum()`

[ ]:
```
Name        0
Sex         0
Age         0
Survived    0
dtype: int64
```

Grouping data and ploting box plot

[ ]:
```
df2 = proc_df.groupby(["Sex","Survived"])
df2

for m , f in df2:
    print(m)
    print(f)
```

```
('female', 0)
```

	Name	Sex	Age	Survived

```
14              Vestrom, Miss. Hulda Amanda Adolfina  female  14.0         0
18  Vander Planke, Mrs. Julius (Emelia Maria Vande…  female  31.0         0
24                      Palsson, Miss. Torborg Danira  female   8.0         0
38                  Vander Planke, Miss. Augusta Maria  female  18.0         0
40      Ahlin, Mrs. Johan (Johanna Persdotter Larsson)  female  40.0         0
..                                                 …      …     …          …
816                     Heininen, Miss. Wendla Maria  female  23.0         0
852                          Boulos, Miss. Nourelain  female   9.0         0
854      Carter, Mrs. Ernest Courtenay (Lilian Hughes)  female  44.0         0
882                      Dahlberg, Miss. Gerda Ulrika  female  22.0         0
885              Rice, Mrs. William (Margaret Norton)  female  39.0         0

[64 rows x 4 columns]
('female', 1)
                                                 Name     Sex   Age  Survived
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0         1
2                            Heikkinen, Miss. Laina  female  26.0         1
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0         1
8   Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0         1
9                   Nasser, Mrs. Nicholas (Adele Achem)  female  14.0         1
..                                                 …      …     …          …
874           Abelson, Mrs. Samuel (Hannah Wizosky)  female  28.0         1
875                  Najib, Miss. Adele Kiamie "Jane"  female  15.0         1
879      Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)  female  56.0         1
880      Shelley, Mrs. William (Imanita Parrish Hall)  female  25.0         1
887                      Graham, Miss. Margaret Edith  female  19.0         1

[197 rows x 4 columns]
('male', 0)
                             Name   Sex   Age  Survived
0          Braund, Mr. Owen Harris  male  22.0         0
4          Allen, Mr. William Henry  male  35.0         0
6            McCarthy, Mr. Timothy J  male  54.0         0
7    Palsson, Master. Gosta Leonard  male   2.0         0
12   Saundercock, Mr. William Henry  male  20.0         0
..                              …     …     …          …
881              Markun, Mr. Johann  male  33.0         0
883   Banfield, Mr. Frederick James  male  28.0         0
884          Sutehall, Mr. Henry Jr  male  25.0         0
886           Montvila, Rev. Juozas  male  27.0         0
890            Dooley, Mr. Patrick  male  32.0         0

[360 rows x 4 columns]
('male', 1)
                             Name   Sex    Age  Survived
21            Beesley, Mr. Lawrence  male  34.00         1
23    Sloper, Mr. William Thompson  male  28.00         1
74                     Bing, Mr. Lee  male  32.00         1
```

```
78        Caldwell, Master. Alden Gates   male    0.83         1
81          Sheerlinck, Mr. Jan Baptist   male   29.00         1
..                                    …    …       …           …
831   Richards, Master. George Sibley    male    0.83         1
838                     Chip, Mr. Chang   male   32.00         1
857              Daly, Mr. Peter Denis    male   51.00         1
869   Johnson, Master. Harold Theodor    male    4.00         1
889              Behr, Mr. Karl Howell    male   26.00         1

[93 rows x 4 columns]
```

```
p=df2.get_group(("male",1))
p
```

```
                               Name   Sex      Age   Survived
21              Beesley, Mr. Lawrence   male   34.00          1
23      Sloper, Mr. William Thompson   male   28.00          1
74                     Bing, Mr. Lee   male   32.00          1
78      Caldwell, Master. Alden Gates   male    0.83          1
81        Sheerlinck, Mr. Jan Baptist   male   29.00          1
..                                 …    …       …            …
831   Richards, Master. George Sibley   male    0.83          1
838                    Chip, Mr. Chang   male   32.00          1
857             Daly, Mr. Peter Denis   male   51.00          1
869   Johnson, Master. Harold Theodor   male    4.00          1
889             Behr, Mr. Karl Howell   male   26.00          1

[93 rows x 4 columns]
```
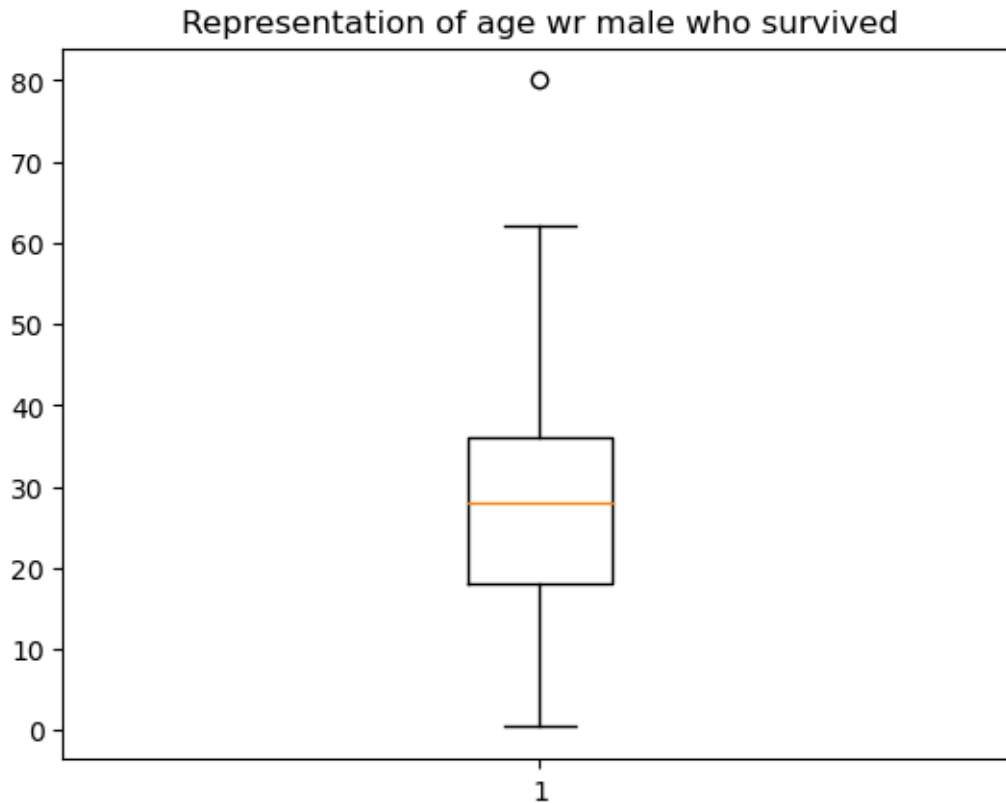
```
plt.title("Representation of age wr male who survived")
plt.boxplot(p["Age"])
print(p["Age"].mean())
```

```
27.276021505376345
```

Representation of age wr male who survived

```
q= df2.get_group(("male",0))
q
```

	Name	Sex	Age	Survived
0	Braund, Mr. Owen Harris	male	22.0	0
4	Allen, Mr. William Henry	male	35.0	0
6	McCarthy, Mr. Timothy J	male	54.0	0
7	Palsson, Master. Gosta Leonard	male	2.0	0
12	Saundercock, Mr. William Henry	male	20.0	0
..	...	...	...	...
881	Markun, Mr. Johann	male	33.0	0
883	Banfield, Mr. Frederick James	male	28.0	0
884	Sutehall, Mr. Henry Jr	male	25.0	0
886	Montvila, Rev. Juozas	male	27.0	0
890	Dooley, Mr. Patrick	male	32.0	0

[360 rows x 4 columns]

```
plt.title("Representation of age wr male who not survived")
plt.boxplot(q["Age"])
print(q["Age"].mean())
```

31.618055555555557

Representation of age wr male who not survived


```
[ ]: r= df2.get_group(("female",1))
     r
```

```
[ ]:                                                     Name     Sex   Age  Survived
     1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0         1
     2                            Heikkinen, Miss. Laina  female  26.0         1
     3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0         1
     8    Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0         1
     9                   Nasser, Mrs. Nicholas (Adele Achem)  female  14.0         1
     ..                                                …       …     …         …
     874         Abelson, Mrs. Samuel (Hannah Wizosky)  female  28.0         1
     875               Najib, Miss. Adele Kiamie "Jane"  female  15.0         1
     879    Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)  female  56.0         1
     880    Shelley, Mrs. William (Imanita Parrish Hall)  female  25.0         1
     887                    Graham, Miss. Margaret Edith  female  19.0         1

     [197 rows x 4 columns]
```
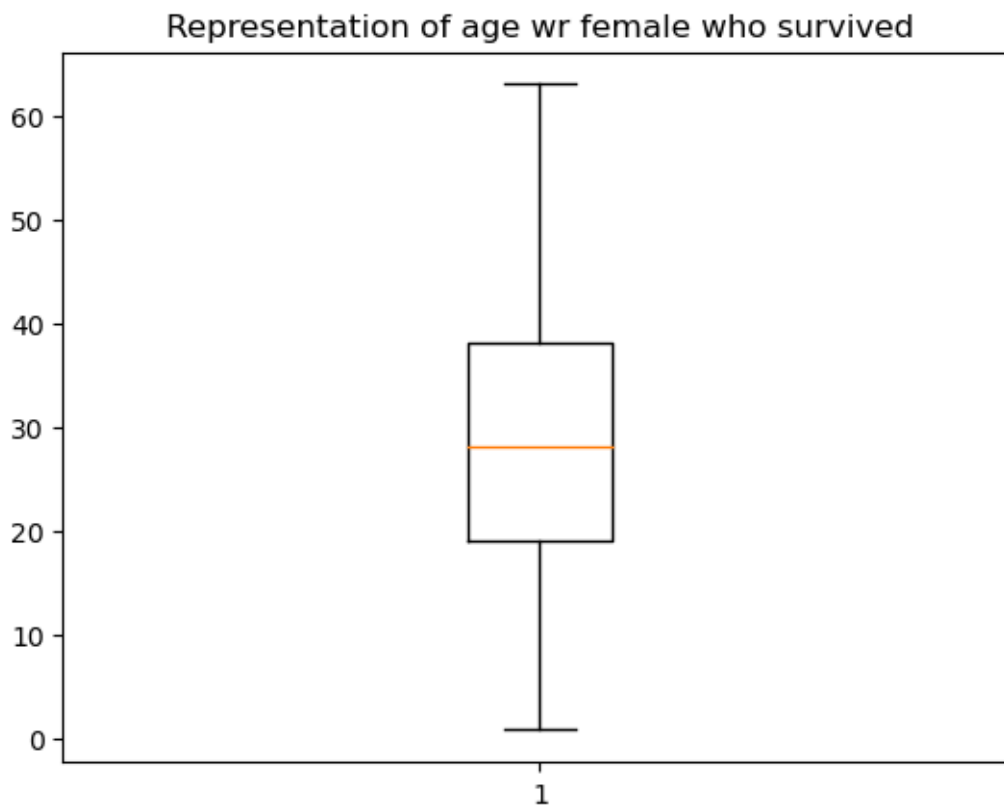
```
plt.title("Representation of age wr female who survived")
plt.boxplot(r["Age"])
print(r["Age"].mean())
```

28.84771573604061

Representation of age wr female who survived

```
s=df2.get_group(("female",0))
s
```

	Name	Sex	Age	Survived
14	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0
18	Vander Planke, Mrs. Julius (Emelia Maria Vande…	female	31.0	0
24	Palsson, Miss. Torborg Danira	female	8.0	0
38	Vander Planke, Miss. Augusta Maria	female	18.0	0
40	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0	0
..	…	…	…	…
816	Heininen, Miss. Wendla Maria	female	23.0	0
852	Boulos, Miss. Nourelain	female	9.0	0
854	Carter, Mrs. Ernest Courtenay (Lilian Hughes)	female	44.0	0
882	Dahlberg, Miss. Gerda Ulrika	female	22.0	0
885	Rice, Mrs. William (Margaret Norton)	female	39.0	0
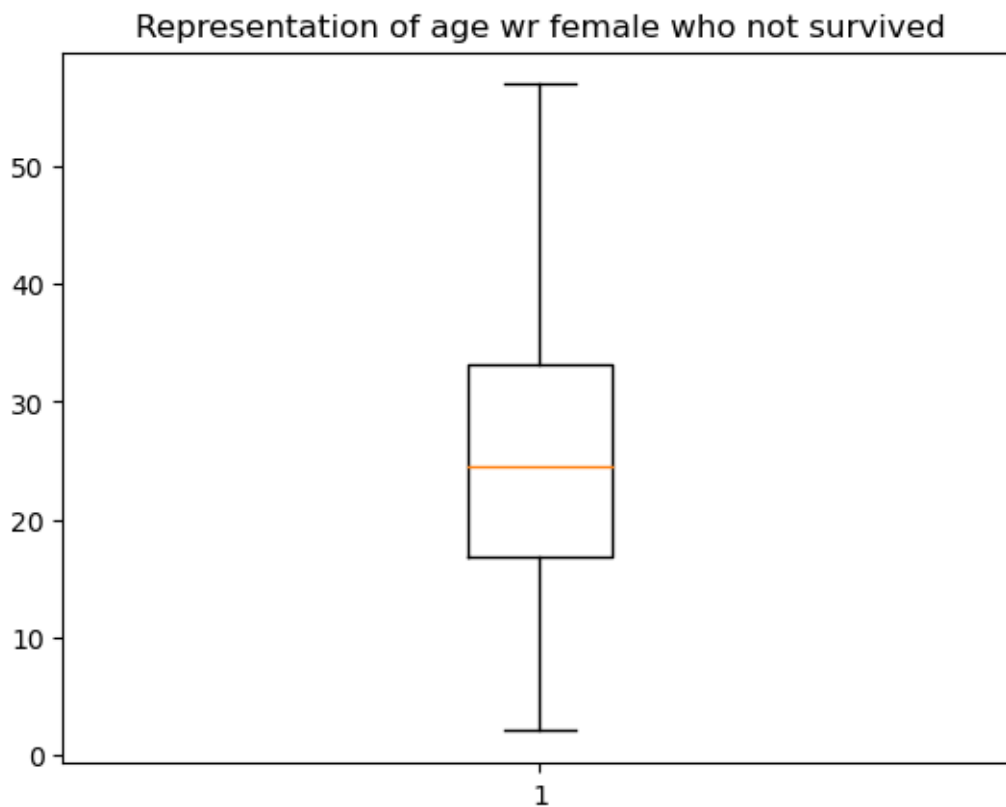
```

```
[64 rows x 4 columns]
```

```
[ ]: plt.title("Representation of age wr female who not survived")
     plt.boxplot(s["Age"])
     print(s["Age"].mean())
```

25.046875

Representation of age wr female who not survived

Observations :-

1.As per above box plot we can say nearabout 50% of males and only 8.96% of females are survived.
2.overall 29.5% survived and 70.50% not survived. 3.avarage age of males and females who survived
is 27.27 and 28.84. 4.avarage age of males and females who not survived is 31.61 and 25.04.