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

In [42]:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

# Load dataset

In [43]:

df**=**pd.read\_csv('titanic.csv')

df

Out[44]:

**me**

|  |  |  |
| --- | --- | --- |
| **PassengerId** | **Survived** | **Pclass Na** |
| **0** 1 | 0 | Brau  3 Mr. Ow Ha |
|  |  | Cumin Mrs. |
| **1** 2 | 1 | 1 Bra |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sex** | **Age** | **SibSp** | **Parch** | **Ticket** | **Fare** |
| male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 |
| female | 38.0 | 1 | 0 | PC 17599 | 71.2833 |

nd, en rris

gs, John dley

(Florence

Briggs Th...

Heikkinen,

Miss. Laina

|  |  |  |  |
| --- | --- | --- | --- |
| **2** | 3 | 1 | 3 |
| **3** | 4 | 1 | 1 |

Futrelle,

Mrs.

female 26.0 0 0

STON/O2. 7.9250

3101282

Jacques Heath (Lily May

Peel)

Allen, Mr. William Henry

|  |  |  |  |
| --- | --- | --- | --- |
| **4**  **...** | 5  ... | 0  ... | 3  ... |
| **886** | 887 | 0 | 2 |
| **887** | 888 | 1 | 1 |

...

Montvila,

Rev. Juozas

Graham,

Miss. Margaret

Edith

Johnston,

Miss.

female 35.0 1 0 113803 53.1000

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| male | 35.0 | 0 | 0 | 373450 | 8.0500 |
| ... | ... | ... | ... | ... | ... |
| male | 27.0 | 0 | 0 | 211536 | 13.0000 |
| female | 19.0 | 0 | 0 | 112053 | 30.0000 |

W./C.

Catherine

|  |  |  |  |
| --- | --- | --- | --- |
| **888** | 889 | 0 | 3 |
| **889** | 890 | 1 | 1 |
| **890** | 891 | 0 | 3 |

Helen "Carrie"

Behr, Mr.

Karl Howell

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| male | 26.0 | 0 | 0 | 111369 | 30.0000 |
| male | 32.0 | 0 | 0 | 370376 | 7.7500 |

Dooley, Mr. Patrick

female NaN 1 2

6607 23.4500

891 rows × 12 columns

# Preprocessing dataset

required\_df**=**df[["Name","Sex","Age","Survived"]] required\_df

Out[45]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 × 4 columns

In [46]:

required\_df.info() *#provide information such as column name ,data type and count of no null*

<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

In [47]:

required\_df.isnull().sum() *#checking null values in dataset*

|  |  |
| --- | --- |
| Out[47]: |  |
| Name | 0 |
| Sex | 0 |
| Age | 177 |
| Survived | 0 |

dtype: int64

Proc\_df**=**required\_df.dropna() *#removing rows which contains null values*

Proc\_df

Out[48]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 × 4 columns

In [49]:

Proc\_df.isnull().sum()

Out[49]:

Name 0

Sex 0

Age 0

Survived 0

dtype: int64

# Grouping data and ploting box plot

df2**=**Proc\_df.groupby(["Sex","Survived"])*# grouping data according to sex and survied*

df2

**for** M,F **in** df2: print(M)

print(F)

|  |  |  |  |
| --- | --- | --- | --- |
| ('female', 0) |  | | |
| Name | Sex | Age | Surviv |
| ed |  |  |  |
| 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)

ed

Name Sex Age Surviv

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)) *#data of male people those are survived*

p

Out[51]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 × 4 columns

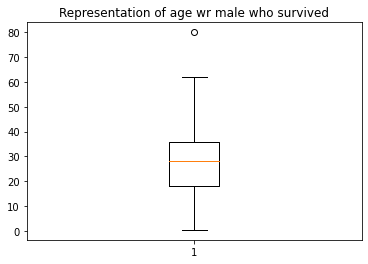
In [52]:

*#boxplot of age wrt male who survived*

plt.title("Representation of age wr male who survived") plt.boxplot(p["Age"])

print(p["Age"].mean())

27.276021505376345



q**=**df2.get\_group(("male",0))*#data of male people those are not survived*

q

Out[53]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 × 4 columns



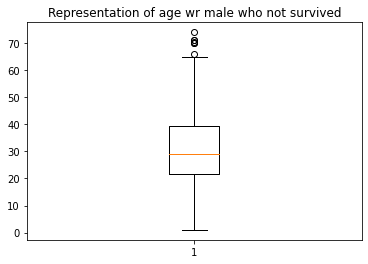
In [54]:

*#boxplot of age wrt male who survived*

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

print(q["Age"].mean())

31.618055555555557



*#data of female people those are survived*

r**=**df2.get\_group(("female",1)) r

Out[55]:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 × 4 columns

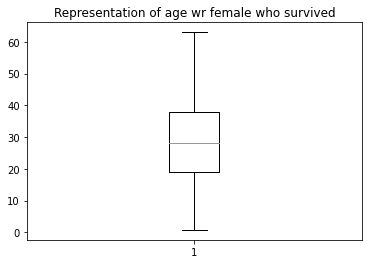
In [56]:

*#boxplot of age wrt female who survived*

plt.title("Representation of age wr female who survived") plt.boxplot(r["Age"])

print(r["Age"].mean())

28.84771573604061



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **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 |

In [58]:

*#boxplot of age wrt female who not survived*

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

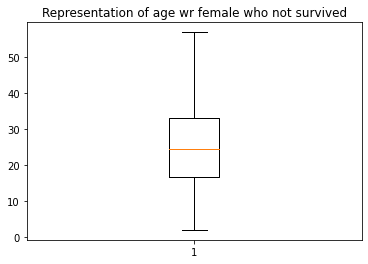
print(s["Age"].mean())



*#data of female people those are not survived*

s**=**df2.get\_group(("female",0)) s

25.046875



# Observations :-

As per above box plot we can say nearabout 50% of males and only 8.96% of females are survived. overall 29.5% survived and 70.50% not survived.

avarage age of males and females who survived is 27.27 and 28.84. avarage age of males and females who not survived is 31.61 and 25.04.