lab1-1

September 8, 2024

[ ]:
```
#LAB 1 : Aim: To calculate and analyze the mean, median, mode, standard␣
 ↪deviation, and quartiles (Q1, Q2, Q3) of a given dataset, and to visually␣
 ↪represent the data distribution using a box plot
import numpy as np ;
import pandas as pd ;
import matplotlib.pyplot as mat ;
# reading the titani file

titenic = pd.read_csv('titanic.csv') ;
print(titenic) ;
```

```
     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
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]
```

[ ]:
```
#find the mean of the age using pandas
mean = titenic['Age'].mean() ;
print(mean) ;
```

```
29.69911764705882
```

[ ]:
```
#find the median of the age
median = titenic['Age'].median() ;
print(median) ;
```

```
28.0
```

[ ]:
```
#find the mode of the age
mode = titenic['Age'].mode()[0] ;
print(mode) ;
```

```
24.0
```

[ ]:
```
#find the minum age of the infilter data freme using the pandas
mimnumage = titenic['Age'].min();
print(mimnumage) ;
```

```
0.42
```

[ ]:
```
#find the stander deviation using the pandas
standerd = titenic['Age'].std() ;
print("standerd deviation " , standerd) ;
```

```
standerd deviation  14.526497332334042
```

[ ]:
```
#find the variace using the standerd deviation
varience = np.sqrt(standerd) ;
print('varience ' , varience) ;
```

```
varience  3.8113642350651875
```

[ ]:
```
#find the q1 and q2 and q3 quertial
Q1 = titenic['Age'].quantile(0.25) ;
Q2 = titenic['Age'].quantile(0.50) ;
Q3 = titenic['Age'].quantile(0.75) ;

print("Q1" , Q1) ;
print("Q2" , Q2) ;
print('Q3' , Q3) ;
```

```
Q1 20.125
Q2 28.0
Q3 38.0
```

[ ]:
```
iqr = Q3 - Q1 ;
print(iqr) ;
```

```
17.875
```

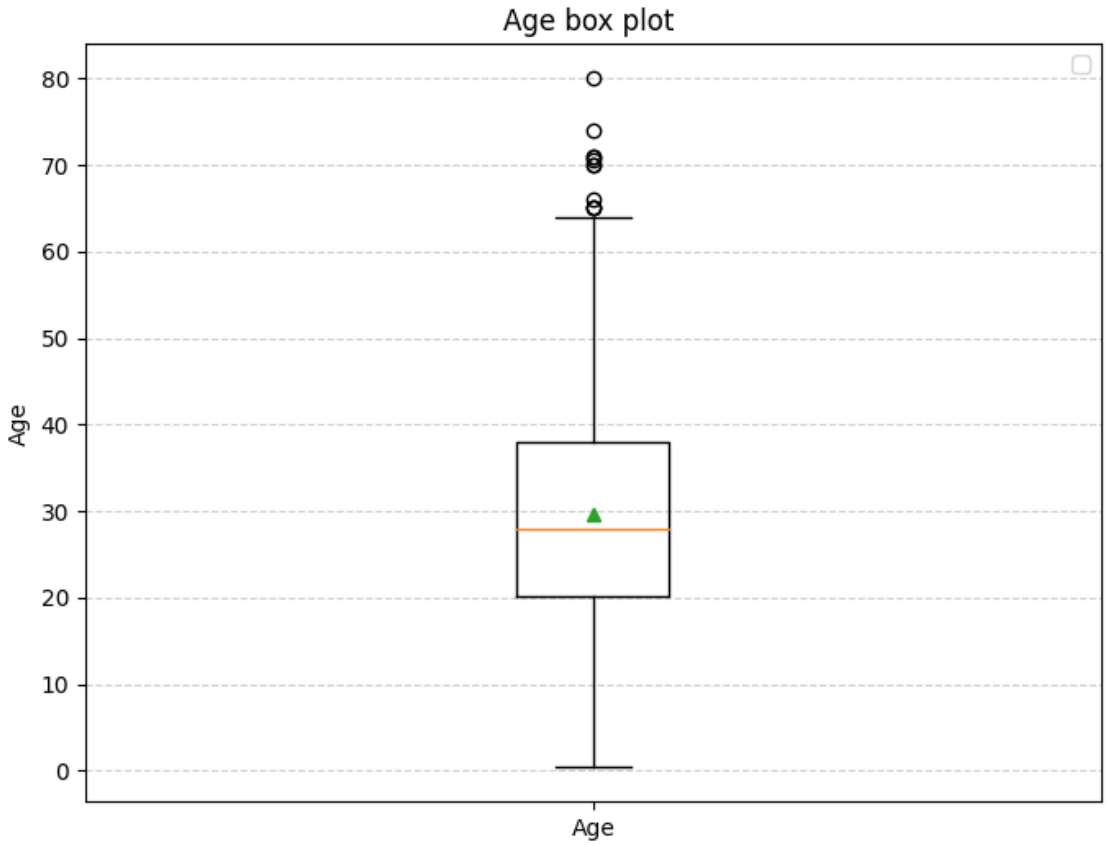
[ ]:
```
#Now finding the upper bound
lowerBound = Q1 - 1.5 * iqr ;
upperbound = Q3 - 1.5 * iqr ;

print(lowerBound) ;
print(upperbound) ;
```

```
-6.6875
11.1875
```

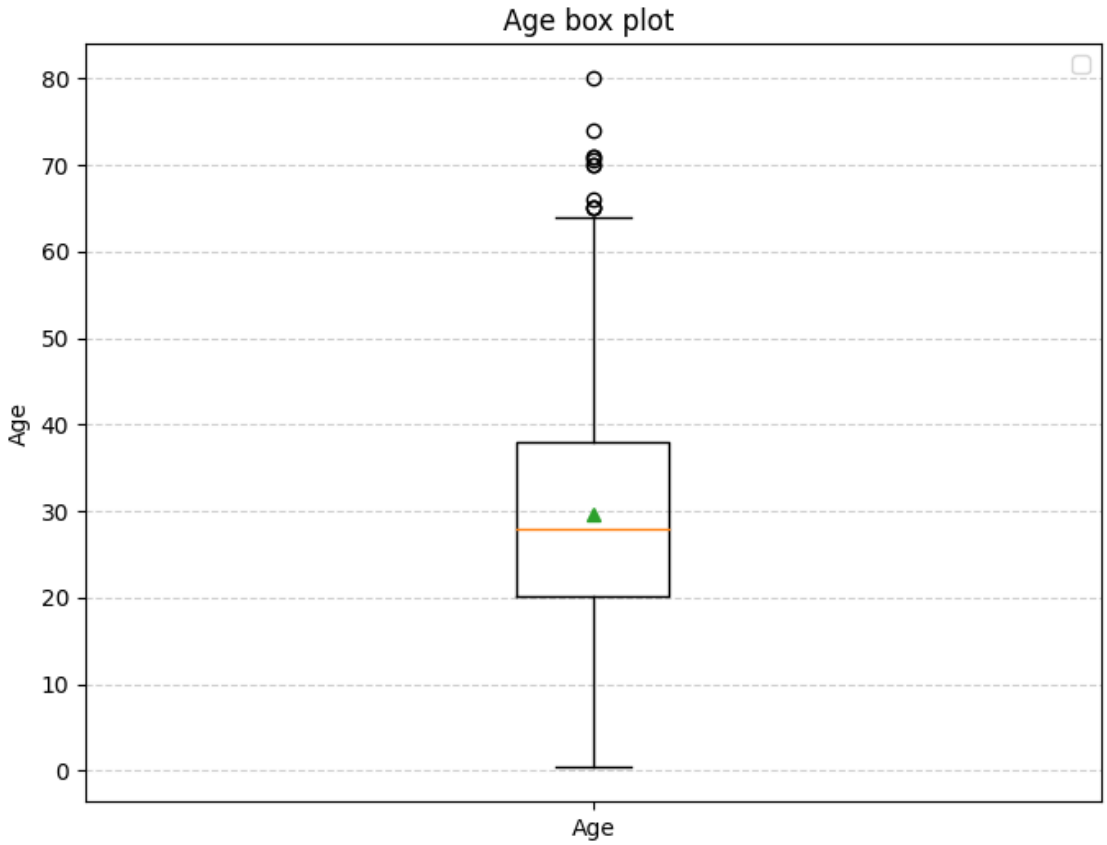
[ ]:
```
#Creating a box plot
#mat is box plot
data = titenic['Age'].dropna() ;
mat.figure(figsize=(8,6)) ;
mat.boxplot(data , showmeans=True) ;
mat.title("Age box plot ") ;
mat.ylabel('Age') ;
mat.xticks([1] , ['Age']) ;
mat.grid(axis='y' , linestyle = '--' , alpha = 0.6 ) ;
outliers = data[(data < lowerBound) | (data > upperbound)] ;
#mat.plot(1 ,  , 'ro' , label = 'Outliers') ;
mat.legend()
mat.show()
```

```
WARNING:matplotlib.legend:No artists with labels found to put in legend.  Note
that artists whose label start with an underscore are ignored when legend() is
called with no argument.
```



[ ]:
```
#Creating a box plot
#mat is box plot
data = titenic['Age'].dropna() ;
mat.figure(figsize=(8,6)) ;
mat.boxplot(data , showmeans=True) ;
mat.title("Age box plot ") ;
mat.ylabel('Age') ;
mat.xticks([1] , ['Age']) ;
mat.grid(axis='y' , linestyle = '--' , alpha = 0.6 ) ;
outliers = data[(data < lowerBound) | (data > upperbound)] ;
#mat.plot(1 , outliers) ;
mat.legend()
mat.show()
```

```
WARNING:matplotlib.legend:No artists with labels found to put in legend.  Note
that artists whose label start with an underscore are ignored when legend() is
called with no argument.
```


Age box plot
80
70
60
50
40
30
20
10
0
Age
Age

[ ]: