NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

# Assignment 3

September 20, 2023

[19]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[20]:
```
dat=pd.read_csv("/content/Titanic-Dataset.csv")
```

[21]:
```
dat.head()
```

[21]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th… | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN | S |

[22]:
```
dat.info()
```

```
<class
'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to
890 Data columns (total 12
columns):
 #   Column      Non-Null Count Dtype
--- ------      -------------- -----
 0  PassengerId 891 non-null   int64
 1   Survived   891 non-null   int64
 2   Pclass     891 non-null   int64
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

```
 3   Name        891 non-null  object
 4   Sex         891 non-null  object
 5   Age         714 non-null  float64
 6   SibSp       891 non-null  int64
 7   Parch       891 non-null  int64
 8   Ticket      891 non-null  object
 9   Fare        891 non-null  float64
 10  Cabin       204 non-null  object
 11  Embarked    889 non-null  object
dtypes: float64(2), int64(5),
object(5) memory usage: 83.7+ KB
```

[23]:
```
dat.describe()
```

[23]:

| | PassengerId | Survived | Pclass | Age | SibSp |
|---|---|---|---|---|---|
| count | 891.000000 | 891.000000 | 891.000000 | 714.000000 | 891.000000 |
| mean | 446.000000 | 0.383838 | 2.308642 | 29.699118 | 0.523008 |
| std | 257.353842 | 0.486592 | 0.836071 | 14.526497 | 1.102743 |
| min | 1.000000 | 0.000000 | 1.000000 | 0.420000 | 0.000000 |
| 25% | 223.500000 | 0.000000 | 2.000000 | 20.125000 | 0.000000 |
| 50% | 446.000000 | 0.000000 | 3.000000 | 28.000000 | 0.000000 |
| 75% | 668.500000 | 1.000000 | 3.000000 | 38.000000 | 1.000000 |
| max | 891.000000 | 1.000000 | 3.000000 | 80.000000 | 8.000000 |

| | Parch | Fare |
|---|---|---|
| count | 891.000000 | 891.000000 |
| mean | 0.381594 | 32.204208 |
| std | 0.806057 | 49.693429 |
| min | 0.000000 | 0.000000 |
| 25% | 0.000000 | 7.910400 |
| 50% | 0.000000 | 14.454200 |
| 75% | 0.000000 | 31.000000 |
| max | 6.000000 | 512.329200 |

[24]:
```
corr=dat.corr
corr
```

[24]:
```
<bound method DataFrame.corr of PassengerId Survived Pclass \
0             1   0     3
1             2   1     1
2             3   1     3
3             4   1     1
4             5   0     3
..          ...  ...   ...
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

```
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
3 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0          1

2                            Heikkinen, Miss. Laina female 26.0    0
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]>
```

[31]:
```
dat.isnull().any()
```

[31]:
```
PassengerId     False
Survived        False
Pclass          False
Name            False
Sex             False
Age              True
SibSp           False
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

```
 Parch        False
Ticket        False
 Fare         False
 Cabin         True
Embarked       True
 dtype: bool
```

[32]:
```
dat.isnull().sum()
```

[32]:
```
 PassengerId   0
Survived       0
 Pclass        0
 Name          0
 Sex           0
 Age         177
 SibSp         0
 Parch         0
 Ticket        0
 Fare          0
 Cabin       687
 Embarked      2
 dtype:
 int64
```

[50]:
```
dat.median()
```

```
<ipython-input-50-899baa360a33>:1: FutureWarning: The default value
of numeric_only in DataFrame.median is deprecated. In a future
version, it will default to False. In addition, specifying
'numeric_only=None' is deprecated. Select only valid columns or
specify the value of numeric_only to silence this warning.
 dat.median()
```

[50]:
```
 PassengerId446.0000
Survived       0.0000
 Pclass        3.0000
 Age          28.0000
 SibSp         0.0000
 Parch         0.0000
 Fare         14.4542
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

```
dtype: float64
```

[58]:
```
m_dat=dat.Age.median()
```

[59]:
```
dat["Age"]=dat["Age"].fillna(m_dat)
```

[68]:
```
dat.drop("Cabin",axis=1,inplace=True)
```

[69]:
```
dat.head()
```

[69]:
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         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

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[75]:
```
dat.isnull().any()
```

[75]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

```
Ticket       False
Fare         False
Embarked      True
dtype: bool
```

[76]:
```
mode=dat["Embarked"].mode()
mode
```

[76]:
```
0    S
Name: Embarked, dtype: object
```

[78]:
```
dat["Embarked"]=dat["Embarked"].fillna(m_dat)
```

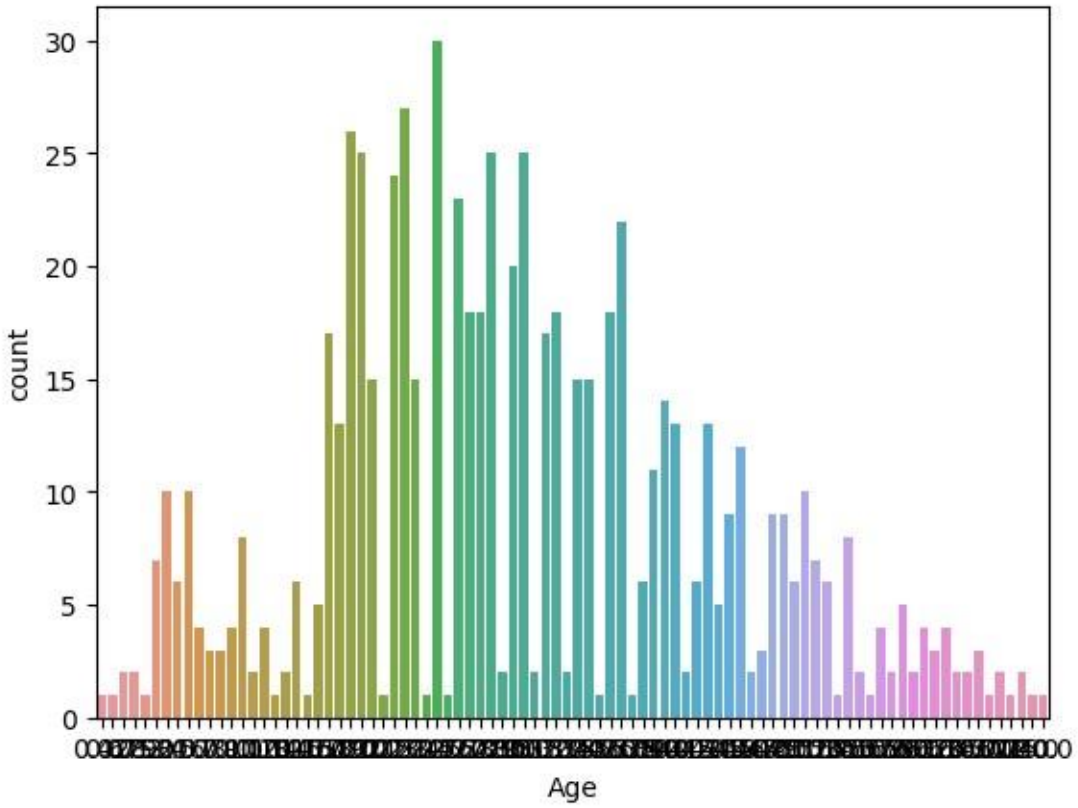
[79]:
```
dat.isnull().any()
```

[79]:
```
PassengerId     False
Survived        False
Pclass          False
Name            False
Sex             False
Age             False
SibSp           False
Parch           False
Ticket          False
Fare            False
Embarked        False
dtype: bool
```

[35]:
```
sns.countplot(x="Age",data=dat)
```

[35]:
```
<Axes: xlabel='Age', ylabel='count'>
```

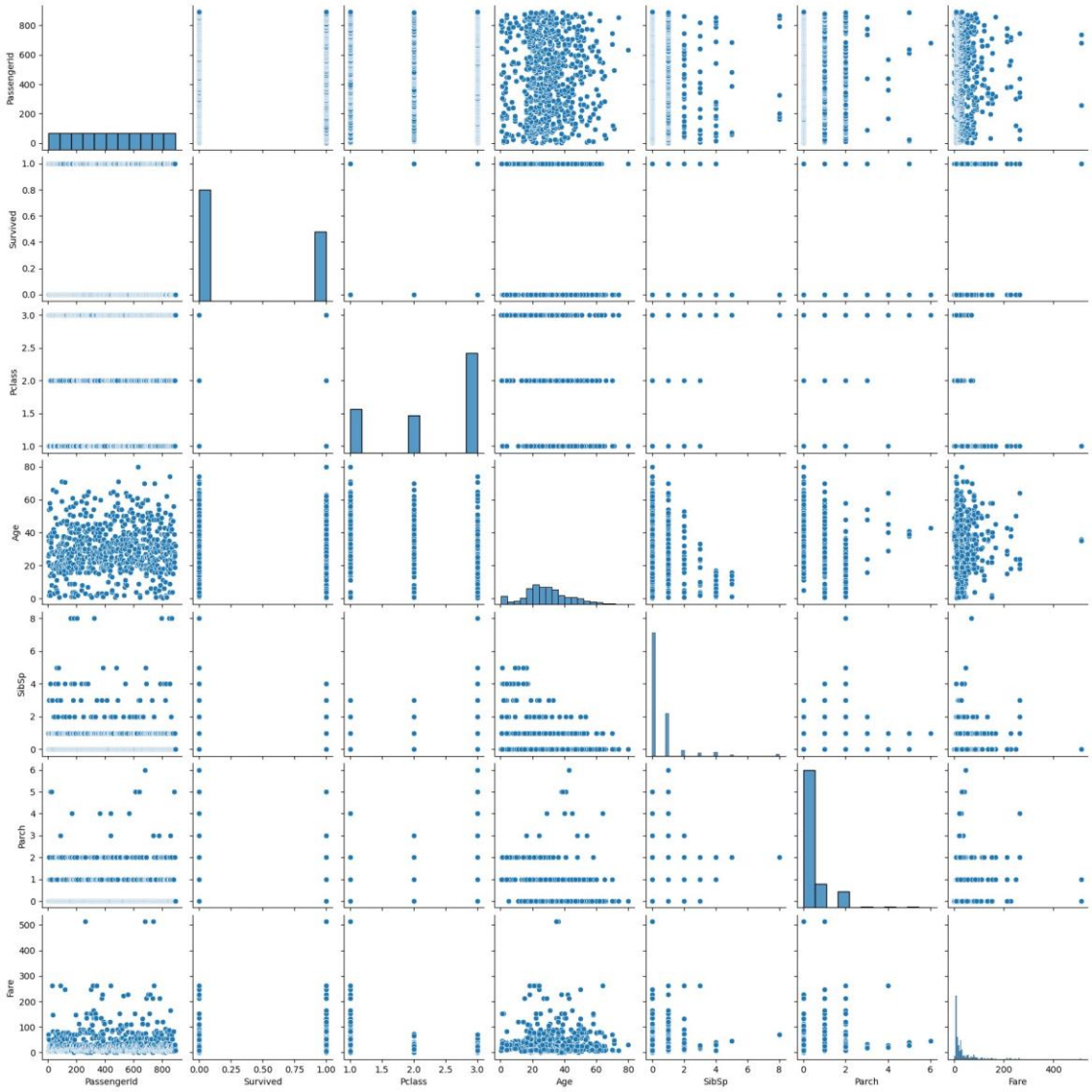
NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE




[37]:
```
sns.pairplot(dat)
```

[37]: <seaborn.axisgrid.PairGrid at 0x79f3f0cb76a0>

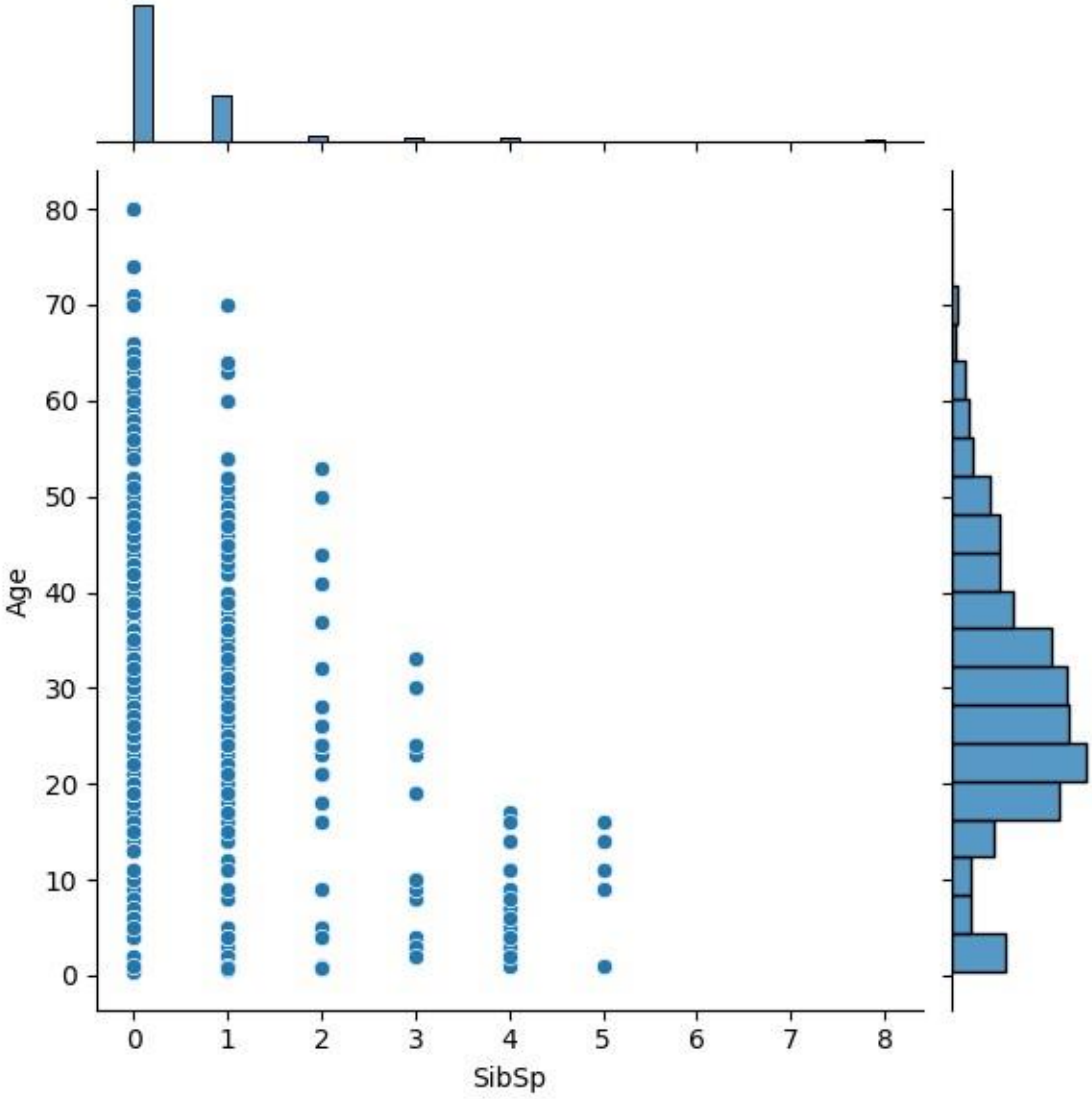
NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE




[39]:
```
sns.jointplot(data=dat,x="SibSp",y="Age")
```

[39]: <seaborn.axisgrid.JointGrid at 0x79f3eead9180>

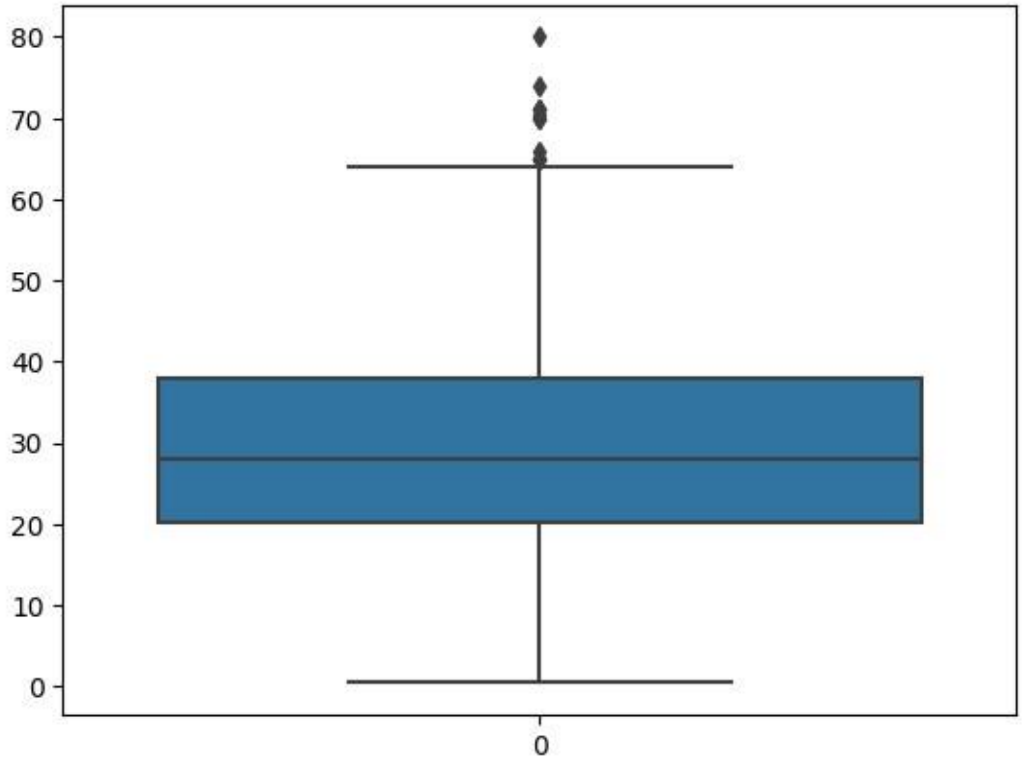
NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE




[40]:
```
sns.boxplot(dat.Age)
```

[40]: <Axes: >

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE




[45]:
```
q1=dat.Age.quantile(0.25)
q3=dat.Age.quantile(0.75)
```

[43]:
```
q1
```

[43]: 38.0

[46]:
```

```
q3

[46]: 38.0

[80]:
```
IQR=q3-q1
```

[82]:
```
upper_limit=q3+1.5*IQR
upper_limit
```

[82]: 64.8125

[86]:
```
lower_limit=q1-1.5*IQR
lower_limit
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

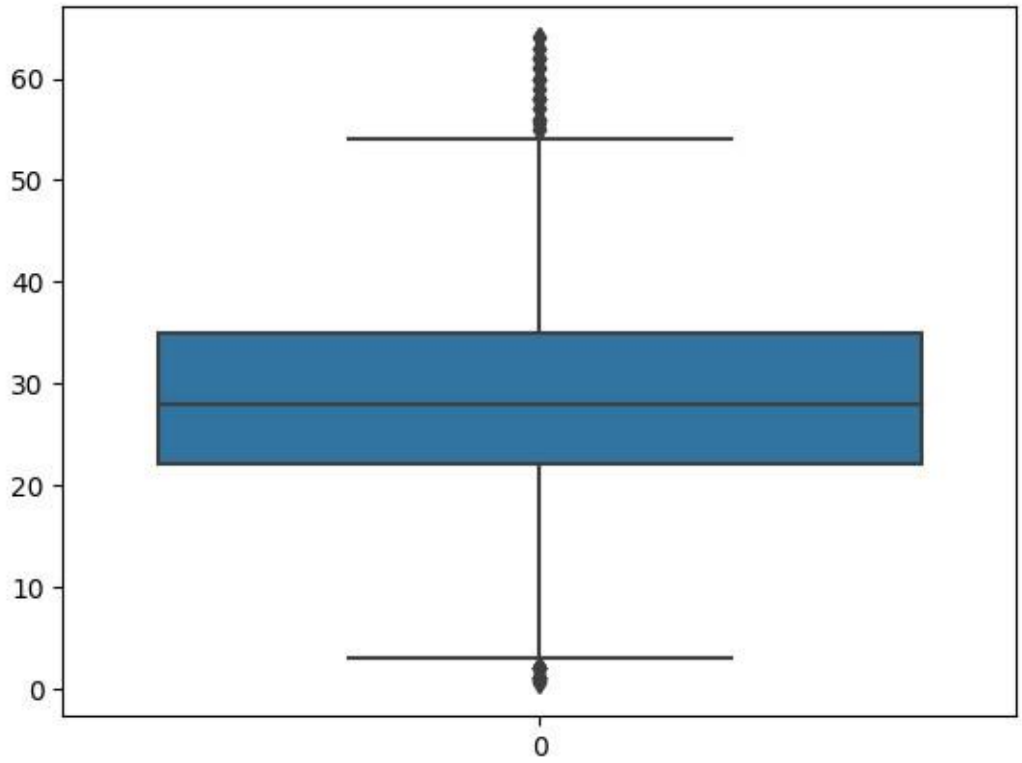
[86]: -6.6875

[90]:
```
d_Age=dat.Age.median()
```

[93]:
```
dat["Age"]=np.where(dat["Age"]>upper_limit,d_Age,dat["Age"])
```

[94]:
```
sns.boxplot(dat.Age)
```

[94]: <Axes:




[95]:
```
dat.describe()
```

[95]:

| | PassengerId | Survived | Pclass | Age | SibSp |
|---|---|---|---|---|---|
| count | 891.000000 | 891.000000 | 891.000000 | 891.000000 | 891.000000 |
| mean | 446.000000 | 0.383838 | 2.308642 | 28.845870 | 0.523008 |
| std | 257.353842 | 0.486592 | 0.836071 | 12.200442 | 1.102743 |
| min | 1.000000 | 0.000000 | 1.000000 | 0.420000 | 0.000000 |
| 25% | 223.500000 | 0.000000 | 2.000000 | 22.000000 | 0.000000 |
| 50% | 446.000000 | 0.000000 | 3.000000 | 28.000000 | 0.000000 |
| 75% | 668.500000 | 1.000000 | 3.000000 | 35.000000 | 1.000000 |

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

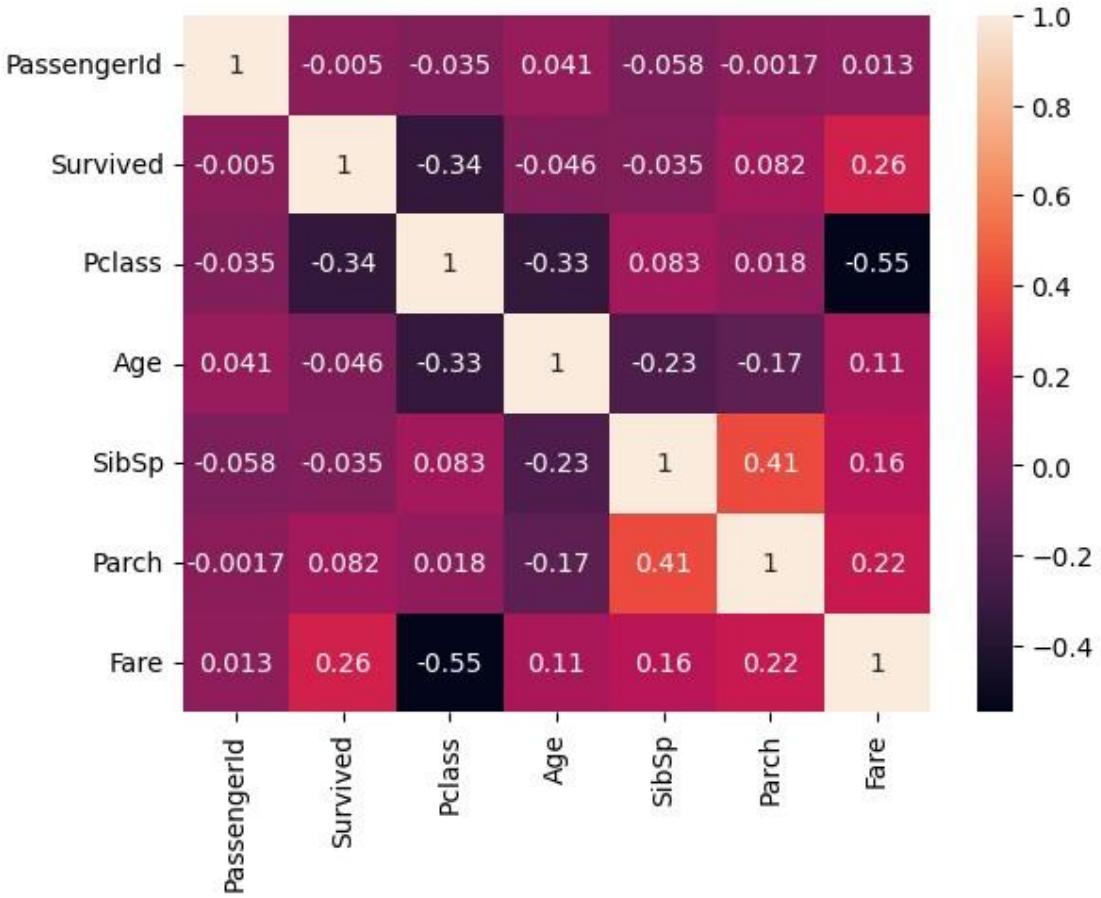
```
max    891.000000   1.000000   3.000000  64.000000   8.000000
```

```
            Parch       Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000
         512.329200
```

[96]:
```
sns.heatmap(dat.corr(),annot=True)
```

```
<ipython-input-96-3a734246ee16>:1: FutureWarning: The default value
of numeric_only in DataFrame.corr is deprecated. In a future
version, it will default to False. Select only valid columns or
specify the value of numeric_only to silence this warning.
sns.heatmap(dat.corr(),annot=True)
```

[96]: <Axes: >



NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

[106]:
```
x=dat.iloc[:,2:11]
x
```

[106]:
```
     Pclass                                                Name     Sex    Age \
0         3 Braund, Mr. Owen Harris     male 22.0
1         1 Cumings, Mrs. John Bradley (Florence Briggs Th… female
          38.0
2         3 Heikkinen, Miss. Laina female 26.0
3         1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0
4         3 Allen, Mr. William Henry    male 35.0
..      …                                                …     …   …
886       2 Montvila, Rev. Juozas male 27.0
887       1 Graham, Miss. Margaret Edith female 19.0
888       3 Johnston, Miss. Catherine Helen "Carrie" female 28.0
889       1 Behr, Mr. Karl Howell male 26.0
890       3 Dooley, Mr. Patrick   male 32.0

            SibSp Parch          Ticket   Fare Embarked
0       1   0     A/5 21171  7.2500     S
1       1   0     PC 17599 71.2833 C
2       0   0 STON/O2. 3101282    7.9250     S
3       1   0     113803 53.1000   S
4       0   0     373450 8.0500    S
      ..    …    …              …      …      …
886     0   0     211536 13.0000   S
887     0   0     112053 30.0000   S
888     1   2     W./C. 6607 23.4500    S
889     0   0     111369 30.0000   C
890     0   0     370376 7.7500    Q
       [891 rows x 9 columns]
```

[107]:
```
y=dat["Survived"]
y
```

[107]:
```
0    0
1    1
2    1
3    1
```

```
4      0
..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

[108]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

[109]:
```
x["Sex"]=le.fit_transform(x["Sex"])
x["Sex"]
```

[109]:
```
0      1
1      0
2      0
3      0
4      1
..
886    1
887    0
888    0
889    1
890    1
Name: Sex, Length: 891, dtype: int64
```

[110]:
```
x["Name"]=le.fit_transform(x["Name"])
x["Name"]
```

[110]:
```
0     108
1     190
2     353
3     272
4      15
```

```
…
886    548
887    303
888    413
889    81
890    220
 Name: Name, Length: 891, dtype: int64
```

[111]:
```
x["Ticket"]=le.fit_transform(x["Ticket"])
x["Ticket"]
```

[111]:
```
0    523
1    596
2    669
3    49
4    472
…
886    101
887    14
888    675
889    8
890    466
      Name: Ticket, Length: 891, dtype: int64
```

[112]:
```
x_Embarked=pd.get_dummies(x["Embarked"],drop_first=True)
x_Embarked
```

[112]:

| | C | Q | S |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 |
| 3 | 0 | 0 | 1 |
| 4 | 0 | 0 | 1 |
| .. | .. | .. | .. |
| 886 | 0 | 0 | 1 |
| 887 | 0 | 0 | 1 |
| 888 | 0 | 0 | 1 |
| 889 | 1 | 0 | 0 |
| 890 | 0 | 1 | 0 |

```
[891 rows x 3 columns]
```

[113]:
```
x_Embarked=pd.get_dummies(x["Embarked"],drop_first=True)
x_Embarked
```

[113]:
```
     C  Q  S
0    0  0  1
1    1  0  0
2    0  0  1
3    0  0  1
4    0  0  1
..  ..  ..  ..
886  0  0  1
887  0  0  1
888  0  0  1
889  1  0  0
890  0  1  0
[891 rows x 3 columns]
```

[114]:
```
x=pd.concat([x,x_Embarked],axis=1)
```

[115]:
```
x.drop("Embarked",axis=1,inplace=True)
```

[117]:
```
x.head()
```

[117]:
```
   Pclass  Name  Sex   Age  SibSp  Parch  Ticket     Fare  C  Q  S
0       3   108    1 22.0      1      0     523   7.2500  0  0  1
1       1   190    0 38.0      1      0     596 71.2833  1  0  0
2       3   353    0 26.0      0      0     669   7.9250  0  0  1  3   1     272
            0 35.0      1      0     49 53.1000  0  0  1
4       3    15    1 35.0      0      0     472   8.0500  0  0  1
```

[119]:
```
from sklearn.model_selection import train_test_split
```

[120]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_sta
te=0)
```

[121]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[121]:
```
((712, 11), (179, 11), (712,), (179,))
```

NAME: PRATYUSH ABHI
REG. NO.: 21BIT0538
CAMPUS: VELLORE

[122]:
```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
```

[127]:
```
xt_train=sc.fit_transform(x_train)
xt_test=sc.fit_transform(x_test)
```

[128]:
```
xt_train
```

[128]:
```
array([[ 0.81925059, -1.32378031, -1.37207547, …, 2.12588331, -
         0.31426968, -1.62827579],
       [-0.38096838, 0.02852784, 0.72882288, …, -0.4703927 , -
        0.31426968, 0.61414657],
       [-0.38096838, 0.25002659, 0.72882288, …, 2.12588331,
        -0.31426968, -1.62827579],
       …,
       [ 0.81925059, 0.630849 , 0.72882288, …, -0.4703927 ,
         3.18198052, -1.62827579],
       [ 0.81925059, 1.73057086, -1.37207547, …, -0.4703927 , -
        0.31426968, 0.61414657],
       [-0.38096838, -1.27326305, 0.72882288, …, -0.4703927 , -
        0.31426968, 0.61414657]])
```

[129]:
```
xt_test
```

[129]:
```
array([[ 0.86022947, 1.61878611, 0.77344314, …, 1.89466187, -
         0.27984505, -1.56278843],
       [ 0.86022947, 1.5600996 , 0.77344314, …, -0.52779866, -
        0.27984505, 0.63988188],
       [ 0.86022947, 0.84021167, 0.77344314, …, -0.52779866,
         3.57340605, -1.56278843],
       …,
       [-1.50871015, 0.45288067, -1.29291987, …, 1.89466187,
        -0.27984505, -1.56278843],
       [ 0.86022947, -1.63244685, 0.77344314, …, -0.52779866, -
        0.27984505, 0.63988188],
       [ 0.86022947, -1.53072356, 0.77344314, …, -0.52779866,
        -0.27984505, 0.63988188]])
```