assignment-3-thridiva

September 20, 2023

0.1 GAJJALA THRIDIVA REDDY

0.2 (MORNING BATCH)

1 1.import the necessary libraries

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

1.1 2.import the dataset

[2]:
```
df=pd.read_csv("/content/Titanic-Dataset.csv")
```

[3]:
```
df
```

[3]:
```
     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]
```

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

[4]:
```
   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 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
```

[5]:
```
df.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
 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
```

[6]:
```
df.shape
```

[6]:
```
(891, 12)
```

[7]:
```
type(df)
```

[7]:
```
pandas.core.frame.DataFrame
```

[8]:
```
df.describe()
```

[8]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

[9]:
```
df.isnull().any()
```

[9]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

[10]:
```
df.isnull().sum()
```

[10]:
```
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
```

INFERENCE: Here we can see that there ar some null values so now we have to handle the null values.We have to use nullvales handling Techniques

[11]:
```
corr=df.corr
corr
```

[11]:
```
<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
..           …       …     …
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]>
```

[12]:
```
df.value_counts()
```

[12]:
```
PassengerId  Survived  Pclass  Name
Sex     Age   SibSp  Parch  Ticket    Fare      Cabin  Embarked
2            1         1       Cumings, Mrs. John Bradley (Florence Briggs
Thayer)   female  38.0  1      0      PC 17599  71.2833   C85    C           1
572          1         1       Appleton, Mrs. Edward Dale (Charlotte Lamson)
female  53.0  2      0      11769     51.4792   C101   S           1
578          1         1       Silvey, Mrs. William Baird (Alice Munger)
female  39.0  1      0      13507     55.9000   E44    S           1
582          1         1       Thayer, Mrs. John Borland (Marian Longstreth
Morris)  female  39.0  1      1      17421     110.8833  C68    C           1
584          0         1       Ross, Mr. John Hugo
male    36.0  0      0      13049     40.1250   A10    C           1
                                                                        ..
```

```
328          1         2       Ball, Mrs. (Ada E Hall)
female  36.0  0      0      28551     13.0000   D      S           1
330          1         1       Hippach, Miss. Jean Gertrude
female  16.0  0      1      111361    57.9792   B18    C           1
332          0         1       Partner, Mr. Austen
male    45.5  0      0      113043    28.5000   C124   S           1
333          0         1       Graham, Mr. George Edward
male    38.0  0      1      PC 17582  153.4625  C91    S           1
890          1         1       Behr, Mr. Karl Howell
male    26.0  0      0      111369    30.0000   C148   C           1
Length: 183, dtype: int64
```

1.2 3.HANDLING THE NULL VALUES

1.delete the null values

2.delete row/column

3.Replace with mean median or mode

[13]:
```
df
```

[13]:
```
     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]
```

[14]:
```
df.isnull().any()
```

[14]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

[15]:
```
df["Age"].mean()
```

[15]:
```
29.69911764705882
```

[16]:
```
df["Age"]=df["Age"].fillna(df["Age"].mean())
```

[17]:
```
df
```

[17]:
```
     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  \
0                                Braund, Mr. Owen Harris    male  22.000000
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.000000
2                                Heikkinen, Miss. Laina  female  26.000000
3          Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.000000
4                              Allen, Mr. William Henry    male  35.000000
..                                                  …     …        …
886                              Montvila, Rev. Juozas    male  27.000000
887                       Graham, Miss. Margaret Edith  female  19.000000
888           Johnston, Miss. Catherine Helen "Carrie"  female  29.699118
889                              Behr, Mr. Karl Howell    male  26.000000
890                                Dooley, Mr. Patrick    male  32.000000

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

[891 rows x 12 columns]
```

[18]:
```
df.isnull().any()
```

[18]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
```

```
dtype: bool
```

[19]:
```
df.drop("Cabin",axis=1,inplace=True)
```

[20]:
```
df
```

[20]:
```
     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  \
0                              Braund, Mr. Owen Harris    male  22.000000
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.000000
2                               Heikkinen, Miss. Laina  female  26.000000
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.000000
4                             Allen, Mr. William Henry    male  35.000000
..                                                 …     …        …
886                             Montvila, Rev. Juozas    male  27.000000
887                      Graham, Miss. Margaret Edith  female  19.000000
888          Johnston, Miss. Catherine Helen "Carrie"  female  29.699118
889                             Behr, Mr. Karl Howell    male  26.000000
890                               Dooley, Mr. Patrick    male  32.000000

     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 11 columns]
```

[21]:
```
df.isnull().any()
```

[21]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Embarked        True
dtype: bool
```

[22]:
```
df.isnull().sum()
```

[22]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64
```

[23]:
```
mode=df["Embarked"].mode()
mode
```

[23]:
```
0    S
Name: Embarked, dtype: object
```

[24]:
```
df.isnull().any()
```

[24]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
```

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

[25]:
```
df["Embarked"]=df["Embarked"].fillna(mode[0])
```

[26]:
```
df.isnull().any()
```

[26]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age            False
SibSp          False
Parch          False
Ticket         False
Fare           False
Embarked       False
dtype: bool
```

INFERENCE: By this we handled all the null Values present in the dataframe

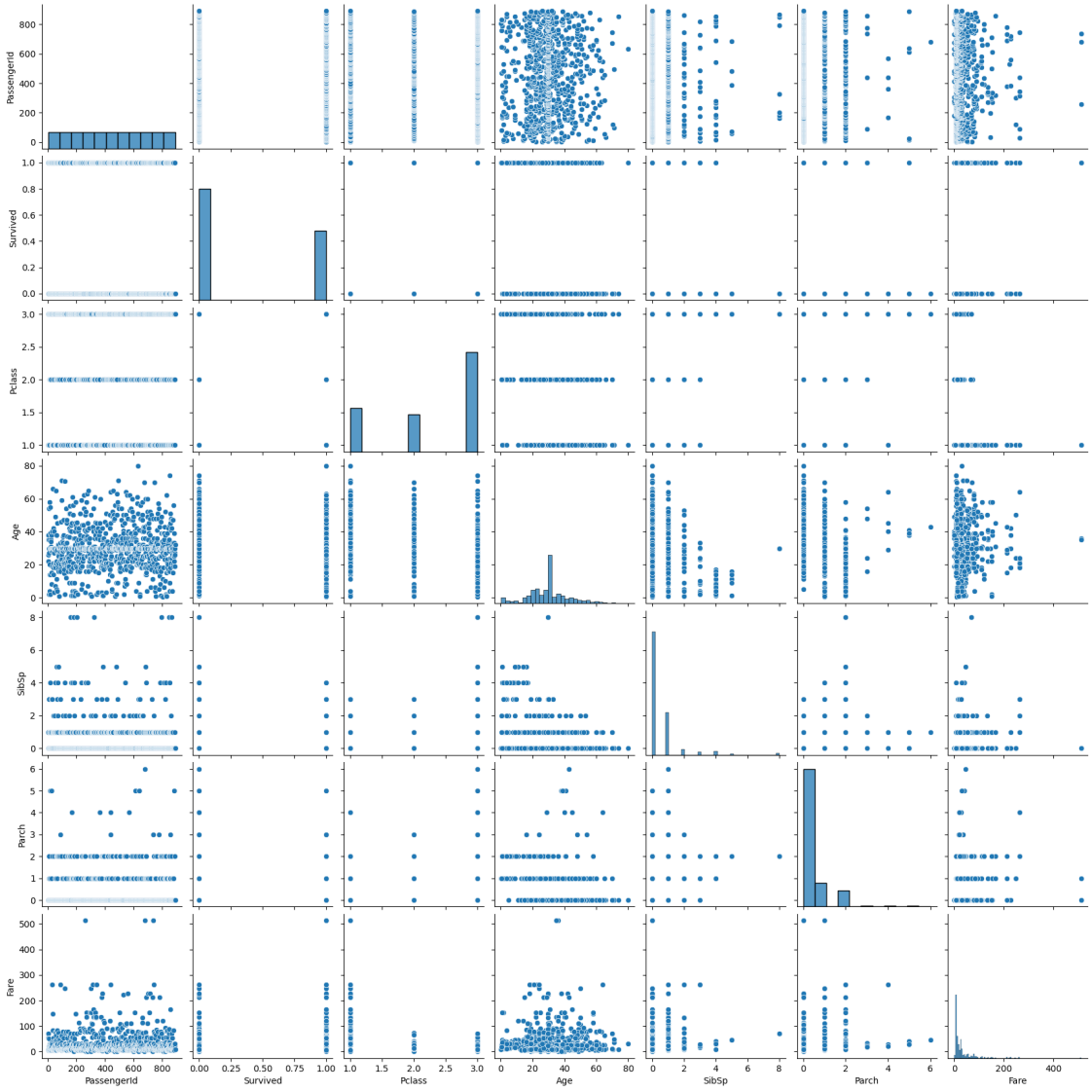
[27]:
```
df.describe()
```

[27]:
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  891.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   13.002015    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   29.699118    0.000000
75%     668.500000    1.000000    3.000000   35.000000    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
```

1.3 4.DATA VISUALIZATION

[28]:
```
plt.figure(figsize=(7,6))
sns.pairplot(df)
plt.show()
```

```
<Figure size 700x600 with 0 Axes>
```

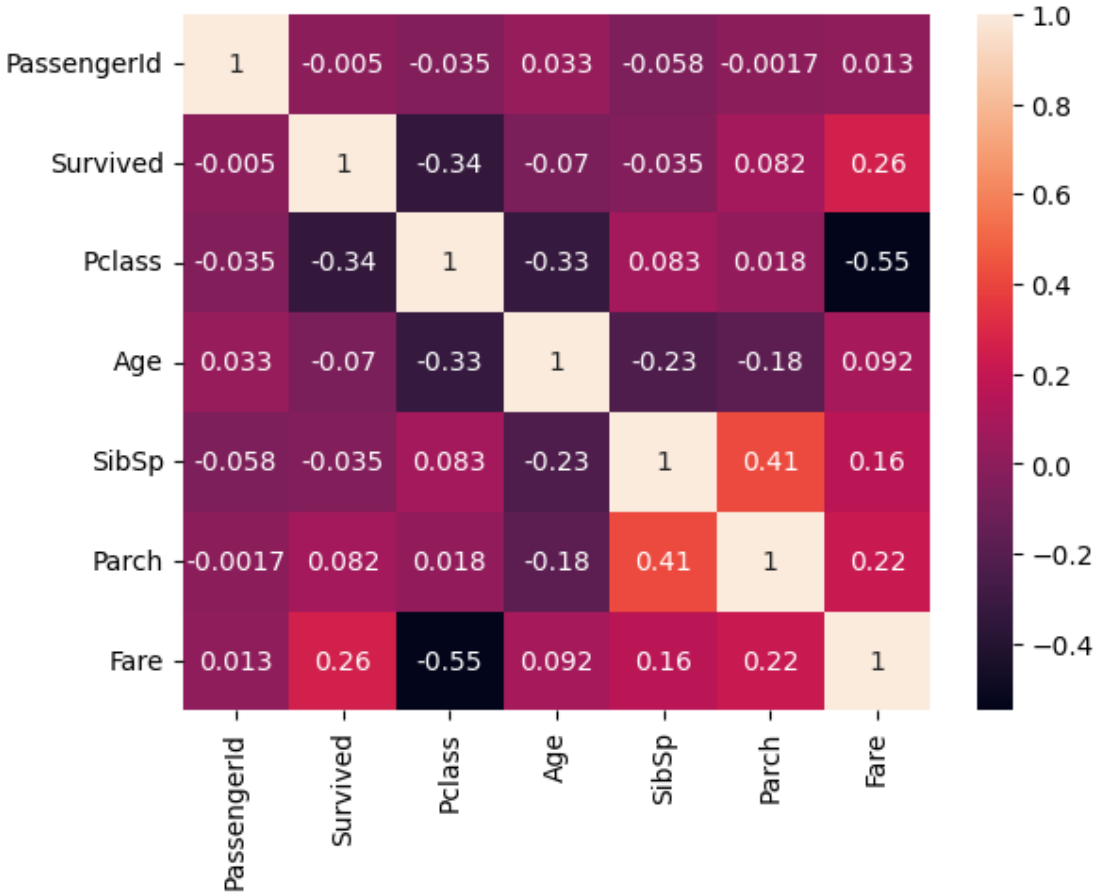


[29]:
```
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-29-8df7bcac526d>: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.
```

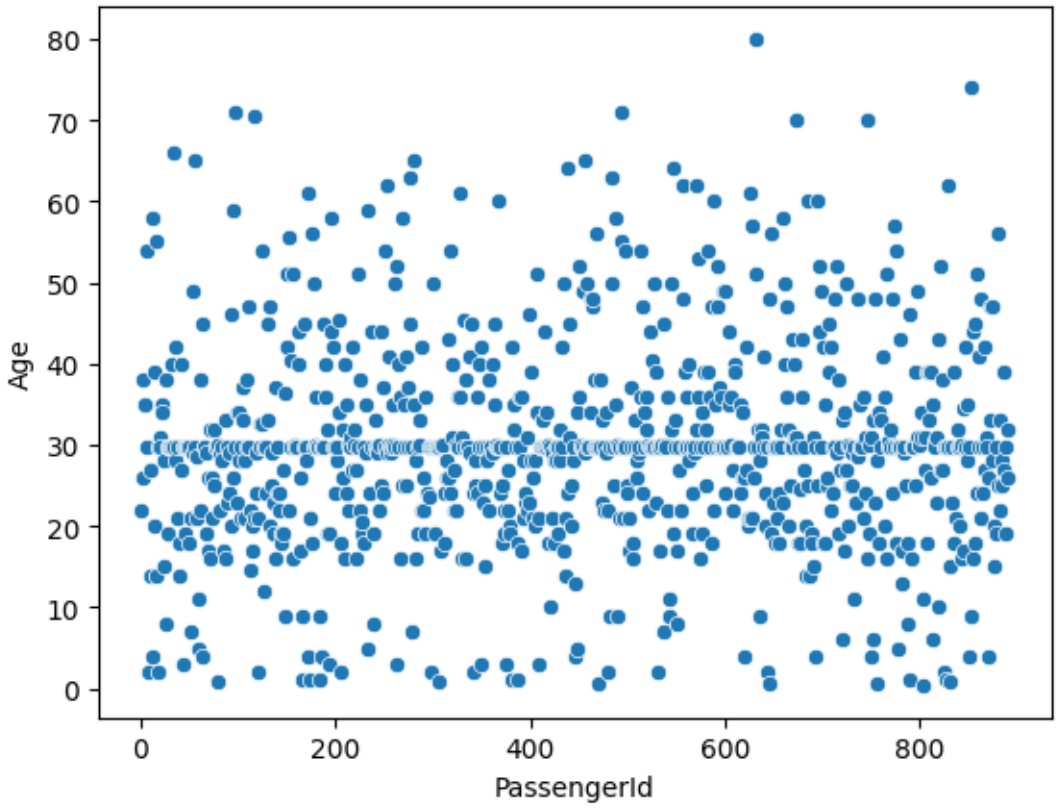
```
sns.heatmap(df.corr(),annot=True)
```

[29]: <Axes: >



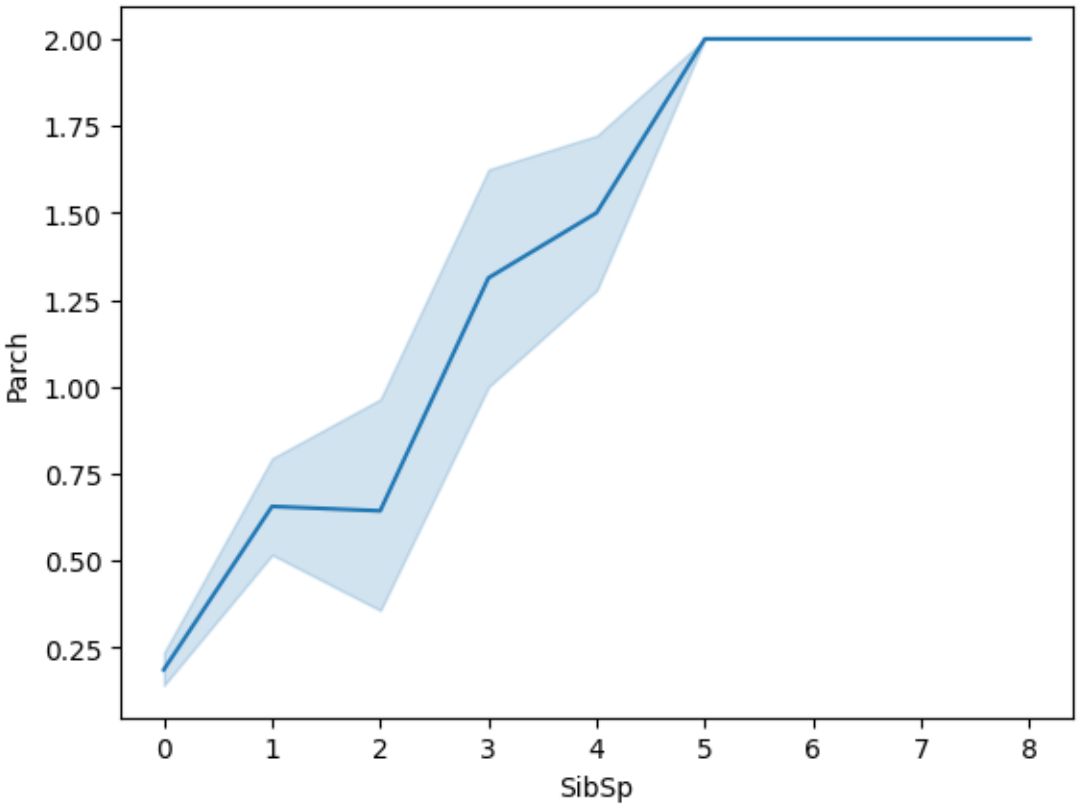
[30]:
```
sns.scatterplot(x="PassengerId",y="Age",data=df)
```

[30]: <Axes: xlabel='PassengerId', ylabel='Age'>



[31]:
```
sns.lineplot(x="SibSp",y="Parch",data=df)
```

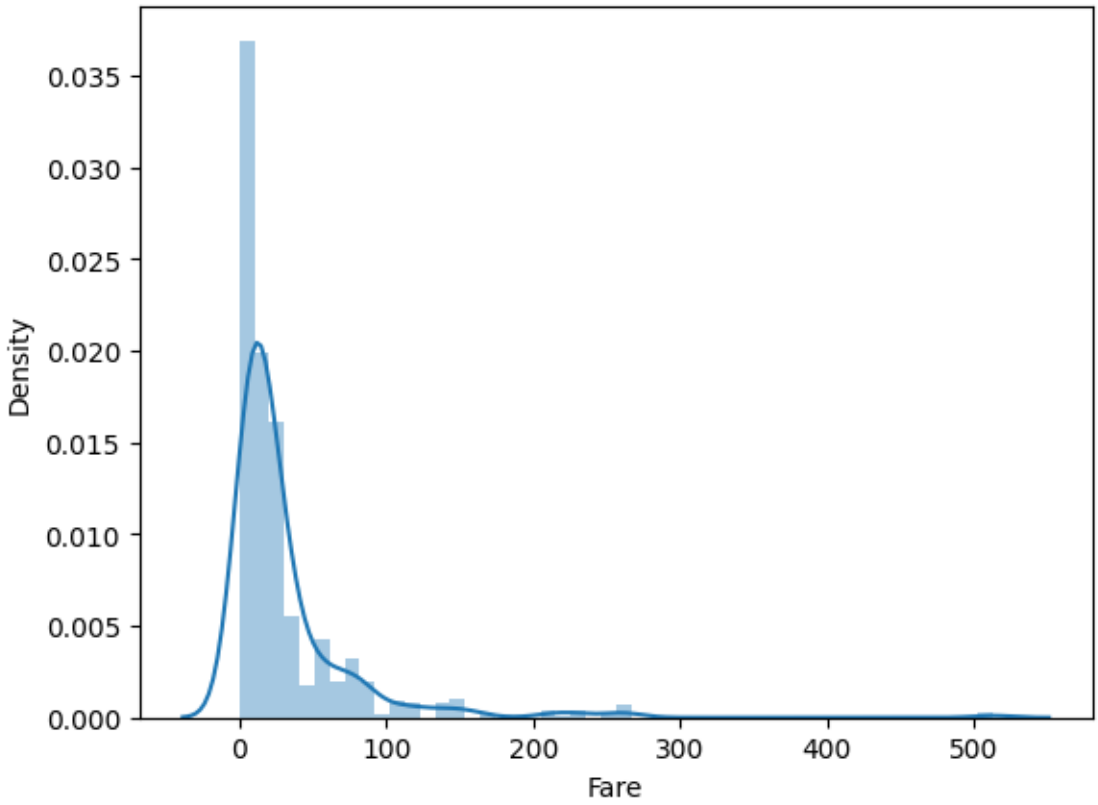
[31]:
```
<Axes: xlabel='SibSp', ylabel='Parch'>
```



[32]:
```
sns.distplot(df["Fare"])
```

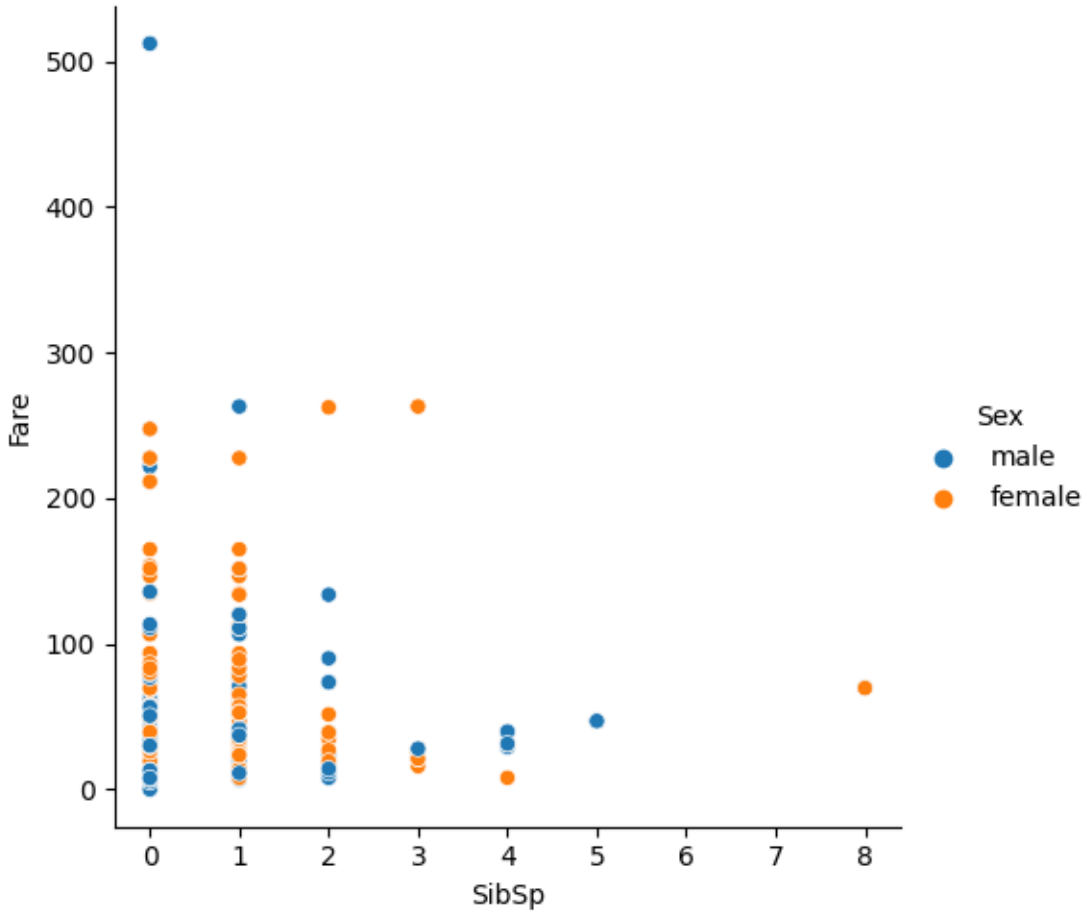
```
<ipython-input-32-5eb648105375>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["Fare"])
```

[32]: `<Axes: xlabel='Fare', ylabel='Density'>`



[33]:
```
sns.relplot(x="SibSp",y="Fare",data=df,hue="Sex")
```

[33]: <seaborn.axisgrid.FacetGrid at 0x7a3b495bf580>



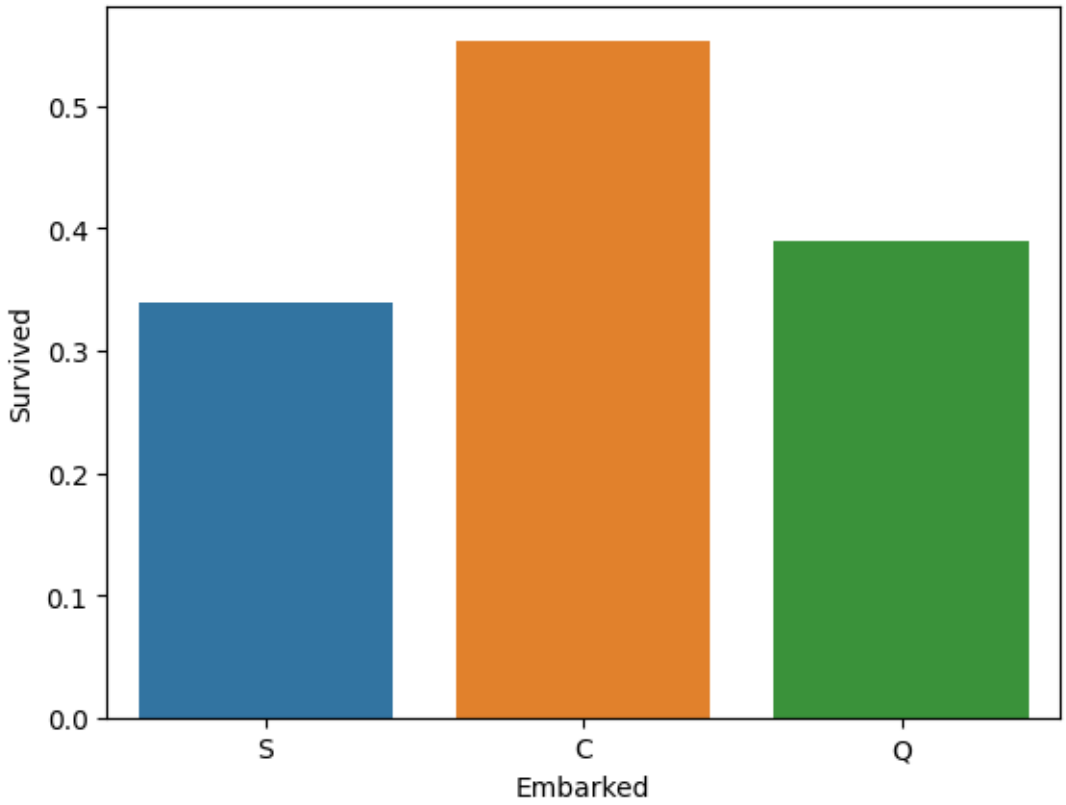
[34]:
```
df["Sex"].value_counts()
```

[34]:
```
male      577
female    314
Name: Sex, dtype: int64
```

[35]:
```
sns.barplot(data=df,x="Embarked",y="Survived",ci=None)
```

```
<ipython-input-35-6e700934b942>:1: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(data=df,x="Embarked",y="Survived",ci=None)
```

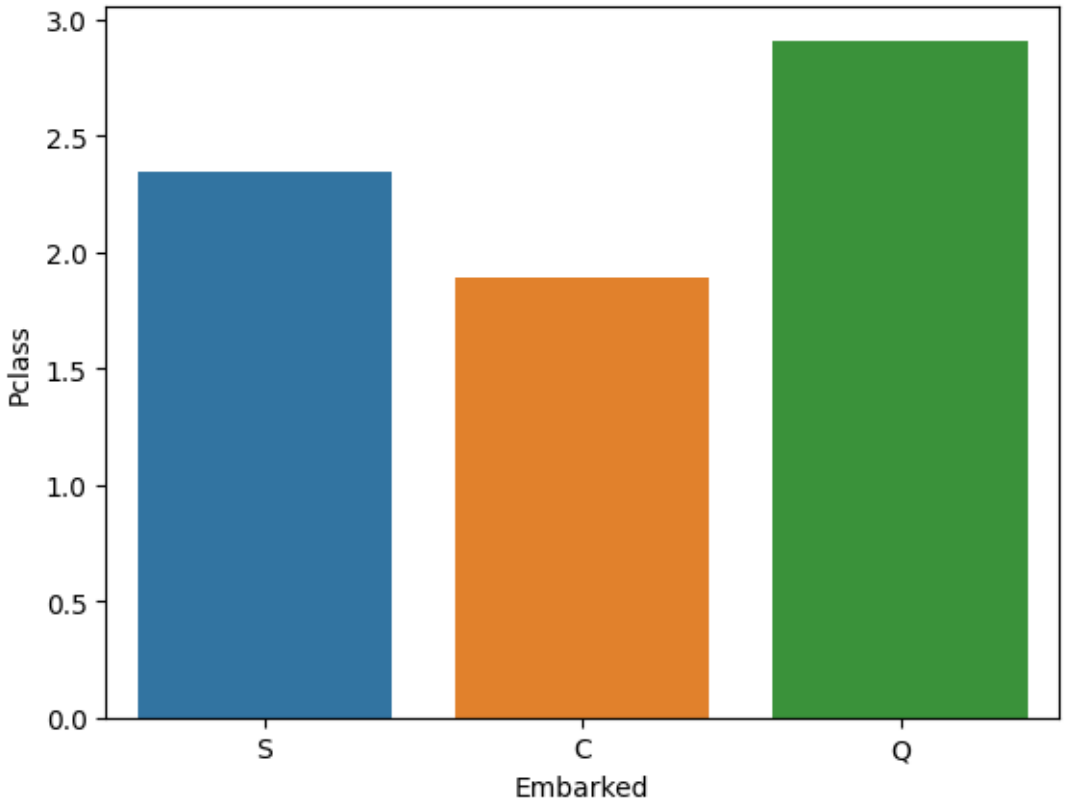
[35]:
```
<Axes: xlabel='Embarked', ylabel='Survived'>
```



[36]:
```
sns.barplot(data=df,x="Embarked",y="Pclass",ci=None)
```

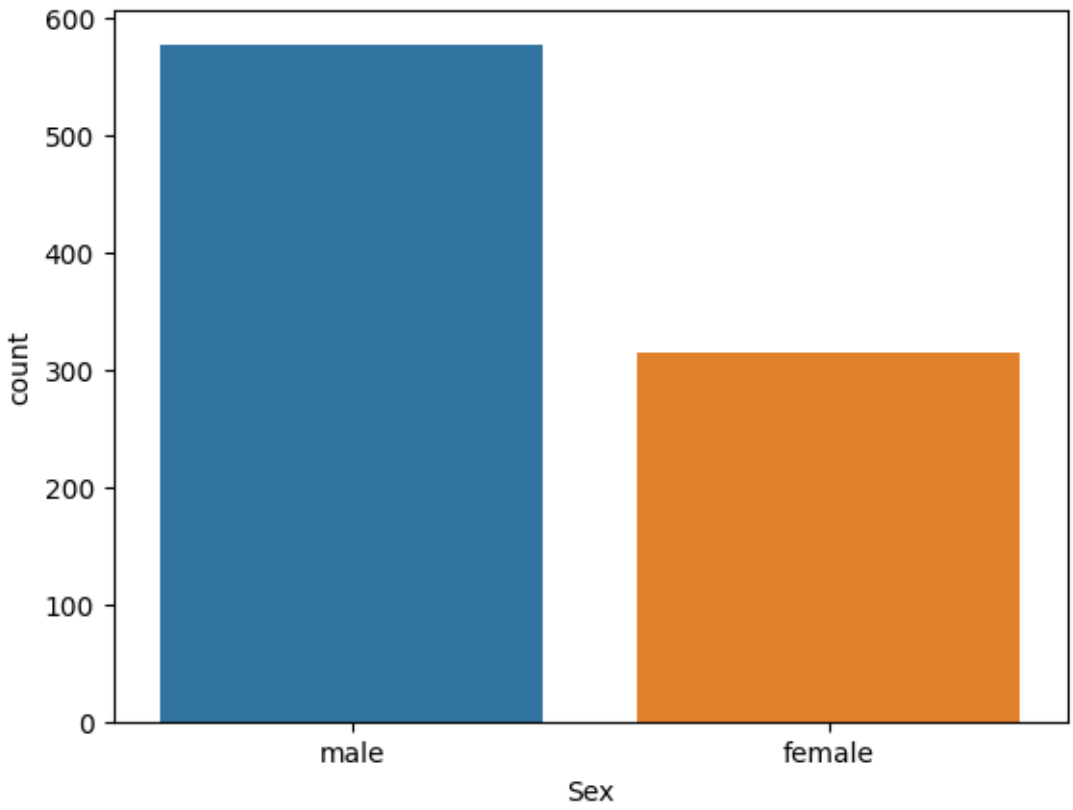
```
<ipython-input-36-49e25b3f0df1>:1: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(data=df,x="Embarked",y="Pclass",ci=None)
```

[36]:
```
<Axes: xlabel='Embarked', ylabel='Pclass'>
```



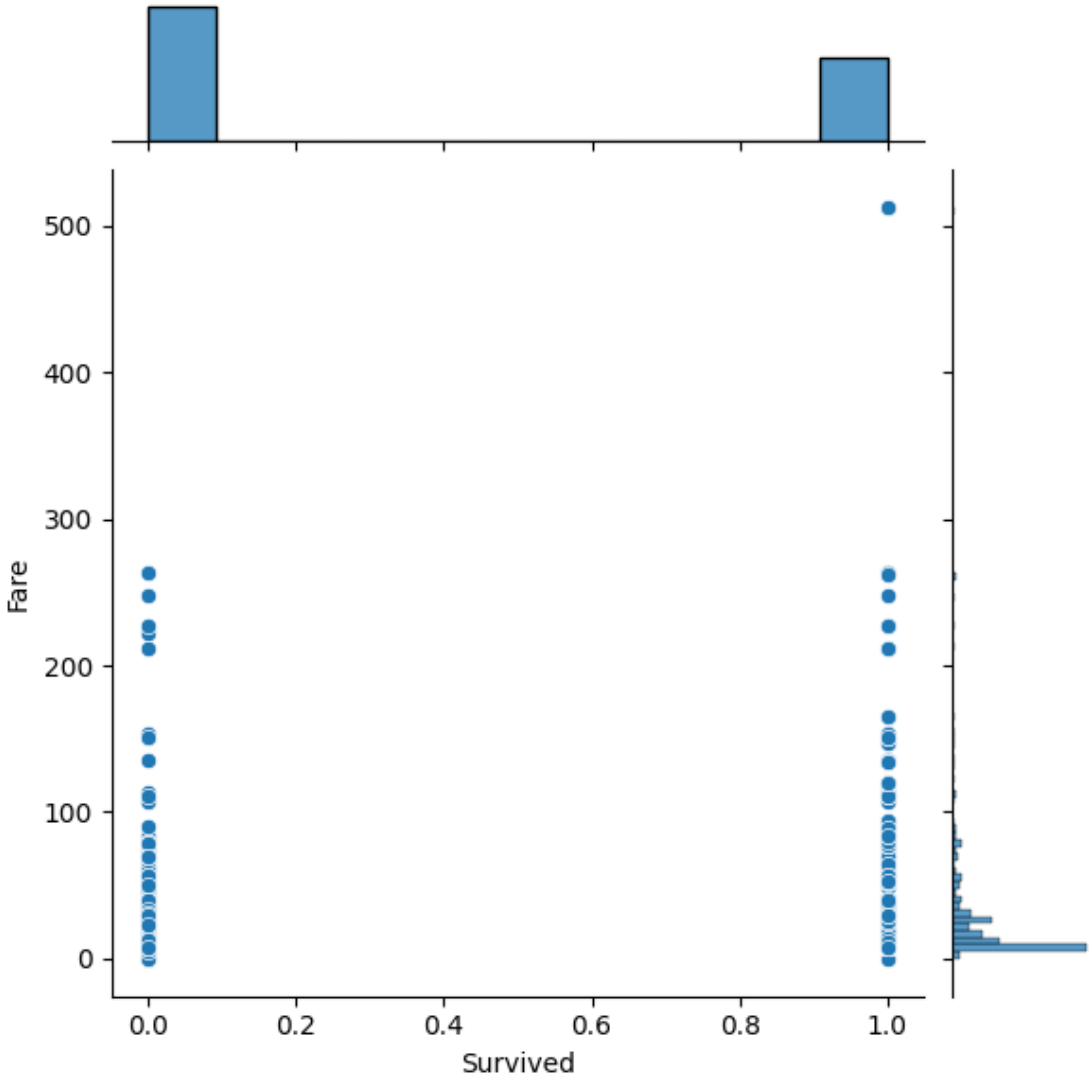
[37]:
```
sns.countplot(x="Sex",data=df)
```

[37]:
```
<Axes: xlabel='Sex', ylabel='count'>
```



[38]:
```
sns.jointplot(x="Survived",y="Fare",data=df)
```

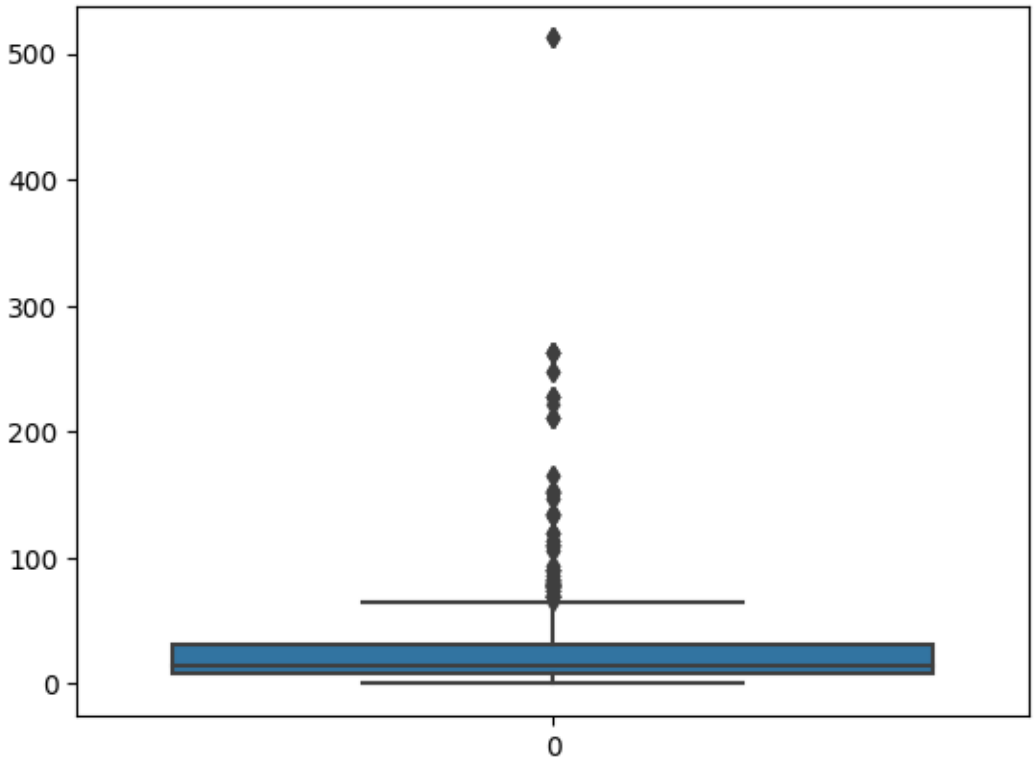
[38]: <seaborn.axisgrid.JointGrid at 0x7a3b49558d90>



1.4 5.Outlier Detection

[39]:
```
sns.boxplot(df.Fare)
```

[39]: <Axes: >



INFERENCE: Here We can see there are some outliers, We have to follow the outliers removing techniques and remove them

1.Inter Quatile Range(IQR Method)

2.Z-Score Method

3.Percentile Method

we can remove the outliers by using the any one of those methods.

Outlier removal by replacement with median

[40]:
```
q1=df.Fare.quantile(0.25)
q3=df.Fare.quantile(0.75)
```

[41]:
```
q1
```

[41]: 7.9104

[42]:
```
q3
```

[42]: 31.0

[43]:
```
IQR=q3-q1
IQR
```

[43]: 23.0896

[44]:
```
upper_limit=q3+1.5*IQR
upper_limit
```

[44]: 65.6344

[45]:
```
fare_outliers=(df["Fare"]>upper_limit).sum()
fare_outliers
```

[45]: 116

[46]:
```
median_Fare=df.Fare.median()
median_Fare
```

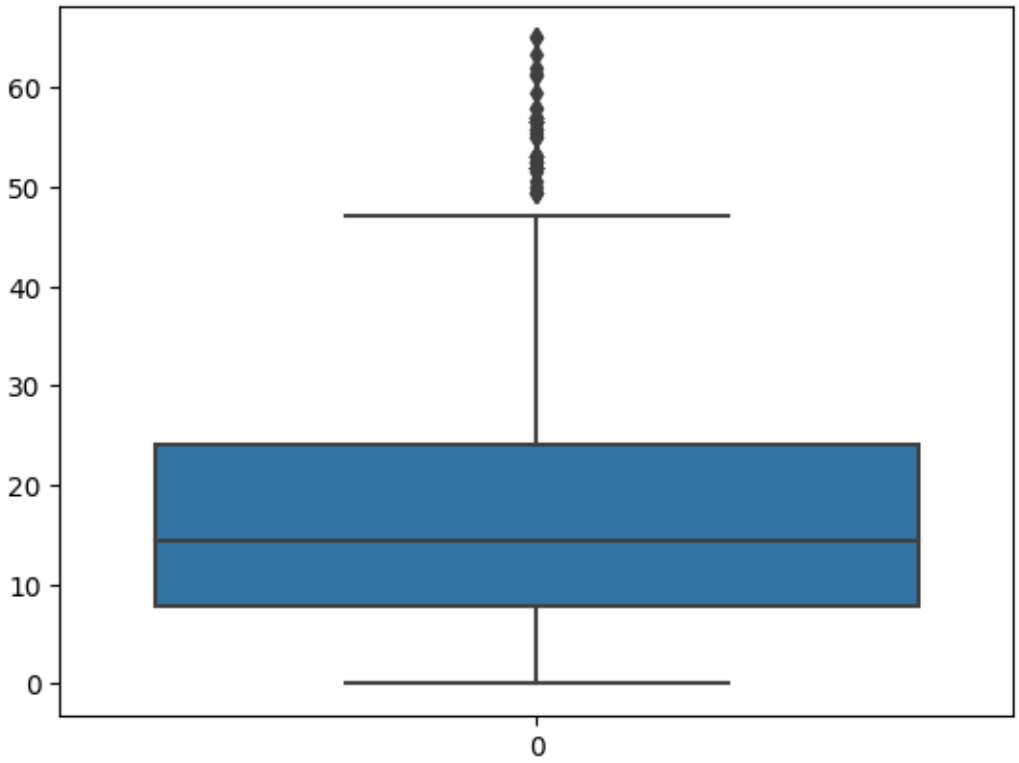
[46]: 14.4542

1.4.1 Outlier Removal

[47]:
```
df["Fare"]=np.where(df["Fare"]>upper_limit,median_Fare,df["Fare"])
```

[48]:
```
sns.boxplot(df.Fare)
```

[48]: <Axes: >



[49]:
```
df.describe()
```

[49]:

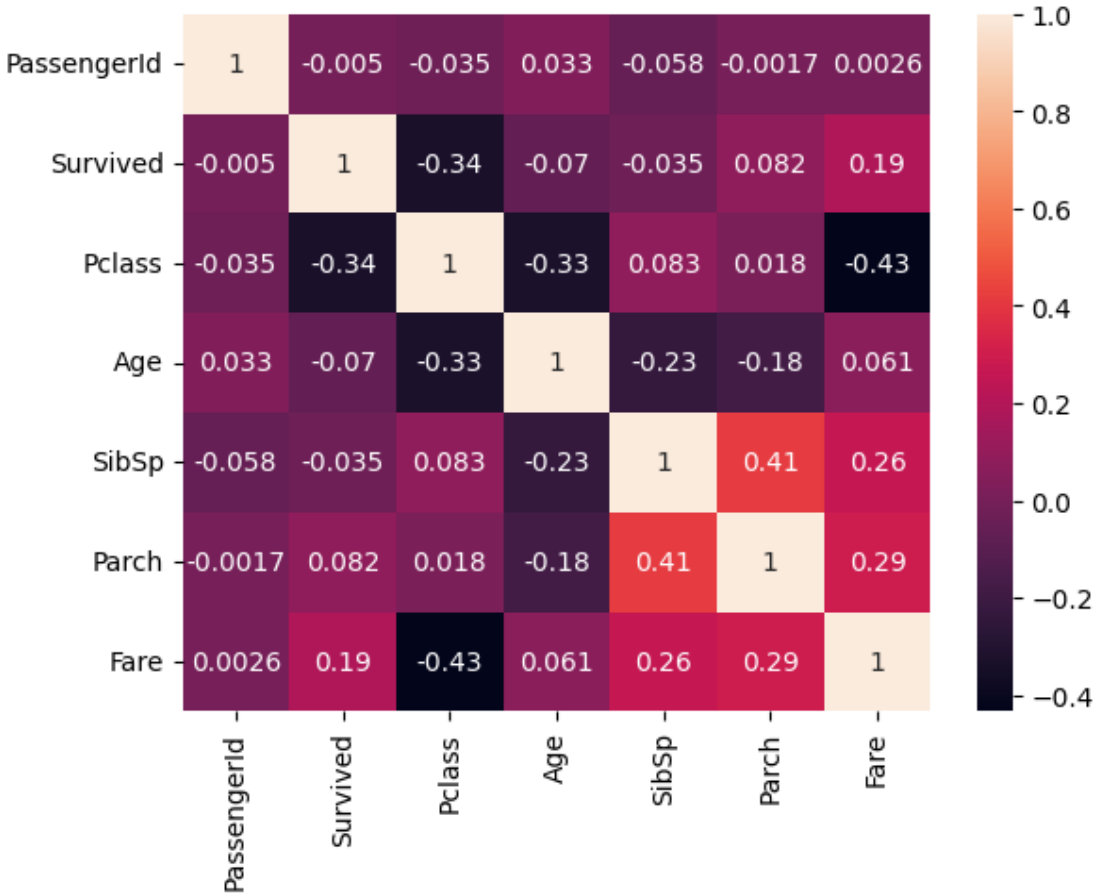
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	17.383622
std	257.353842	0.486592	0.836071	13.002015	1.102743	0.806057	12.713016
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	29.699118	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	24.150000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	65.000000

1.5 6.spliting of dependent and independent variables

[50]:
```
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-50-8df7bcac526d>: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(df.corr(),annot=True)
```

[50]: <Axes: >



[51]:
```
df.head()
```

[51]:

	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  14.4542        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

[52]:
```
df.shape
```

[52]:
```
(891, 11)
```

[53]:
```
x=df.iloc[:,2:10]
x
```

[53]:
```
     Pclass                                               Name     Sex  \
0         3                            Braund, Mr. Owen Harris    male
1         1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female
2         3                             Heikkinen, Miss. Laina  female
3         1       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female
4         3                           Allen, Mr. William Henry    male
..      …                                                  …     …
886       2                              Montvila, Rev. Juozas    male
887       1                       Graham, Miss. Margaret Edith  female
888       3           Johnston, Miss. Catherine Helen "Carrie"  female
889       1                              Behr, Mr. Karl Howell    male
890       3                                Dooley, Mr. Patrick    male

            Age  SibSp  Parch            Ticket     Fare
0     22.000000      1      0         A/5 21171   7.2500
1     38.000000      1      0          PC 17599  14.4542
2     26.000000      0      0  STON/O2. 3101282   7.9250
3     35.000000      1      0            113803  53.1000
4     35.000000      0      0            373450   8.0500
..          …    …    …                …      …
886   27.000000      0      0            211536  13.0000
887   19.000000      0      0            112053  30.0000
888   29.699118      1      2        W./C. 6607  23.4500
889   26.000000      0      0            111369  30.0000
```

```
890  32.000000      0      0            370376   7.7500

[891 rows x 8 columns]
```

[54]:
```
x=df.drop("Survived",axis=1)
type(x)
```

[54]:
```
pandas.core.frame.DataFrame
```

[55]:
```
x.head()
```

[55]:
```
   PassengerId  Pclass                                               Name  \
0            1       3                            Braund, Mr. Owen Harris
1            2       1  Cumings, Mrs. John Bradley (Florence Briggs Th…
2            3       3                             Heikkinen, Miss. Laina
3            4       1       Futrelle, Mrs. Jacques Heath (Lily May Peel)
4            5       3                           Allen, Mr. William Henry

      Sex   Age  SibSp  Parch            Ticket     Fare Embarked
0    male  22.0      1      0         A/5 21171   7.2500        S
1  female  38.0      1      0          PC 17599  14.4542        C
2  female  26.0      0      0  STON/O2. 3101282   7.9250        S
3  female  35.0      1      0            113803  53.1000        S
4    male  35.0      0      0            373450   8.0500        S
```

[56]:
```
y=df["Survived"]
y
```

[56]:
```
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
```

[57]:
```
type(y)
```

[57]:
```
pandas.core.series.Series
```

[58]:
```
type(x)
```

[58]:
```
pandas.core.frame.DataFrame
```

##7.Encoding

[59]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

[60]:
```
x["Sex"]=le.fit_transform(x["Sex"])
x["Sex"]
```

[60]:
```
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
```

[61]:
```
x.head()
```

[61]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	0	38.0	1	0	PC 17599	14.4542	C
2	3	3	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.0	1	0	113803	53.1000	S
4	5	3	Allen, Mr. William Henry	1	35.0	0	0	373450	8.0500	S

[62]:
```
x["Name"]=le.fit_transform(x["Name"])
x["Name"]
```

[62]:
```
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
```

[63]:
```
x["Ticket"]=le.fit_transform(x["Ticket"])
x["Ticket"]
```

[63]:
```
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
```

[64]:
```
x_Embarked=pd.get_dummies(x["Embarked"],drop_first=True)
x_Embarked
```

[64]:
```
     Q  S
0    0  1
1    0  0
2    0  1
3    0  1
4    0  1
..  .. ..
886  0  1
887  0  1
888  0  1
889  0  0
890  1  0

[891 rows x 2 columns]
```

[65]:
```
x=pd.concat([x,x_Embarked],axis=1)
```

[66]:
```
x.drop("Embarked",axis=1,inplace=True)
```

[67]:
```
x.head()
```

[67]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Q	S
0	1	3	108	1	22.0	1	0	523	7.2500	0	1
1	2	1	190	0	38.0	1	0	596	14.4542	0	0
2	3	3	353	0	26.0	0	0	669	7.9250	0	1
3	4	1	272	0	35.0	1	0	49	53.1000	0	1
4	5	3	15	1	35.0	0	0	472	8.0500	0	1

1.6 8.Train test split

[68]:
```
from sklearn.model_selection import train_test_split
```

[69]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

[70]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[70]:
```
((712, 11), (179, 11), (712,), (179,))
```

[71]:
```
x_train.head()
```

[71]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Q	S
140	141	3	99	0	29.699118	0	2	203	15.2458	0	0
439	440	2	447	1	31.000000	0	0	547	10.5000	0	1
817	818	2	504	1	31.000000	1	1	618	37.0042	0	0
378	379	3	85	1	20.000000	0	0	183	4.0125	0	0
491	492	3	871	1	21.000000	0	0	649	7.2500	0	1

[72]:
```
x_test.head()
```

[72]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Q	S
495	496	3	880	1	29.699118	0	0	176	14.4583	0	0
648	649	3	865	1	29.699118	0	0	620	7.5500	0	1
278	279	3	681	1	7.000000	4	1	480	29.1250	1	0
31	32	1	776	0	29.699118	1	0	586	14.4542	0	0
255	256	3	819	0	29.000000	0	2	185	15.2458	0	0

```
255  0  0
```

1.7 9.Feature Scaling

[73]:
```
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
```

[74]:
```
xs_train=sc.fit_transform(x_train)
xs_test=sc.fit_transform(x_test)
```

[75]:
```
xs_train
```

[75]:
```
array([[-1.16343003,  0.81925059, -1.32378031, …, -0.17726299,
        -0.31426968, -1.6398534 ],
       [-0.01263834, -0.38096838,  0.02852784, …, -0.54667438,
        -0.31426968,  0.60981061],
       [ 1.44220868, -0.38096838,  0.25002659, …,  1.51640316,
        -0.31426968, -1.6398534 ],
       …,
       [ 0.71863397,  0.81925059,  0.630849  , …, -0.76203333,
         3.18198052, -1.6398534 ],
       [ 0.44921786,  0.81925059,  1.73057086, …, -0.00958083,
        -0.31426968,  0.60981061],
       [ 0.93031806, -0.38096838, -1.27326305, …,  1.67175552,
        -0.31426968,  0.60981061]])
```

[76]:
```
xs_test
```

[76]:
```
array([[ 0.1591693 ,  0.86022947,  1.61878611, …, -0.19571051,
        -0.27984505, -1.56278843],
       [ 0.78048767,  0.86022947,  1.5600996 , …, -0.76604362,
        -0.27984505,  0.63988188],
       [-0.72204694,  0.86022947,  0.84021167, …,  1.01513799,
         3.57340605, -1.56278843],
       …,
       [-0.97788392, -1.50871015,  0.45288067, …, -0.19604899,
        -0.27984505, -1.56278843],
       [ 1.53175497,  0.86022947, -1.63244685, …, -0.74092958,
        -0.27984505,  0.63988188],
       [-0.34032193,  0.86022947, -1.53072356, …, -0.72476479,
        -0.27984505,  0.63988188]])
```