In [1]:
```python
# Name : Kamal Nayak
# Reg No : 21BCE9722

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

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

In [3]:
```python
df.head()
```

Out[3]:

	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

In [4]:
```python
df.describe()
```

Out[4]:

	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

In [5]:
```python
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
```

In [6]:
```python
df.corr()
```

```
C:\Users\A C\AppData\Local\Temp\ipykernel_14880\1134722465.py: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.
  df.corr()
```

Out[6]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
Age	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067
SibSp	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000

In [7]:
```python
df.corr().Fare.sort_values(ascending=False)
```

```
C:\Users\A C\AppData\Local\Temp\ipykernel_14880\60082530.py: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.
  df.corr().Fare.sort_values(ascending=False)
```

Out[7]:
```
Fare           1.000000
Survived       0.257307
Parch          0.216225
SibSp          0.159651
Age            0.096067
PassengerId    0.012658
Pclass        -0.549500
Name: Fare, dtype: float64
```

In [8]:
```python
df.isnull().any()
```

Out[8]:
```
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
```

In [9]:
```python
df.isnull().sum()
```

Out[9]:
```
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
```

In [10]:
```python
df.Age.nunique()
```

Out[10]:
```
88
```

In [11]:
```python
df.Age.unique()
```

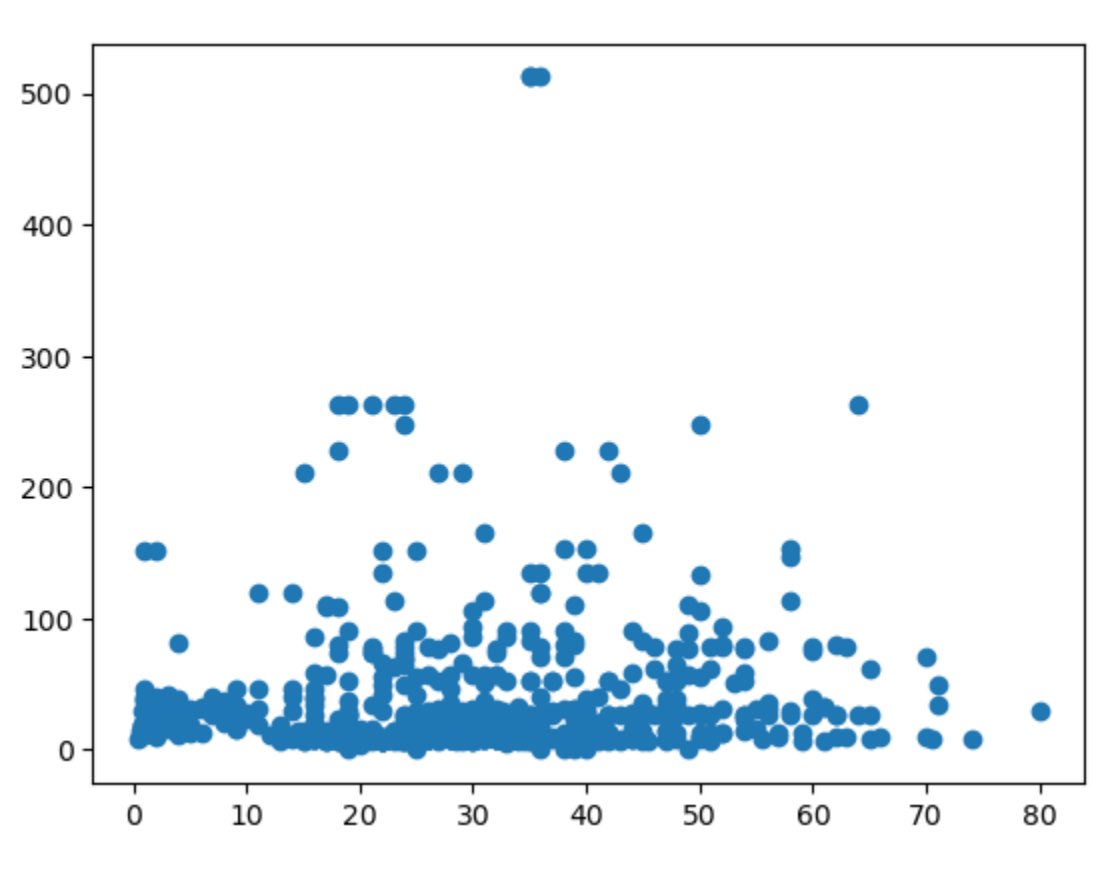
Out[11]:
```
array([22.  , 38.  , 26.  , 35.  ,   nan, 54.  ,  2.  , 27.  , 14.  ,
        4.  , 58.  , 20.  , 39.  , 55.  , 31.  , 34.  , 15.  , 28.  ,
        8.  , 19.  , 40.  , 66.  , 42.  , 21.  , 18.  ,  3.  ,  7.  ,
       49.  , 29.  , 65.  , 28.5 ,  5.  , 11.  , 45.  , 17.  , 32.  ,
       16.  , 25.  ,  0.83, 30.  , 33.  , 23.  , 24.  , 46.  , 59.  ,
       71.  , 37.  , 47.  , 14.5 , 70.5 , 32.5 , 12.  ,  9.  , 36.5 ,
       51.  , 55.5 , 40.5 , 44.  ,  1.  , 61.  , 56.  , 50.  , 36.  ,
       45.5 , 20.5 , 62.  , 41.  , 52.  , 63.  , 23.5 ,  0.92, 43.  ,
       60.  , 10.  , 64.  , 13.  , 48.  ,  0.75, 53.  , 57.  , 80.  ,
       70.  , 24.5 ,  6.  ,  0.67, 30.5 ,  0.42, 34.5 , 74.  ])
```

In [12]:
```python
df.Age.value_counts()
```

Out[12]:
```
24.00    30
22.00    27
18.00    26
19.00    25
28.00    25
         ..
36.50     1
55.50     1
0.92      1
23.50     1
74.00     1
Name: Age, Length: 88, dtype: int64
```

In [13]:
```python
plt.scatter(df["Age"],df["Fare"])
```

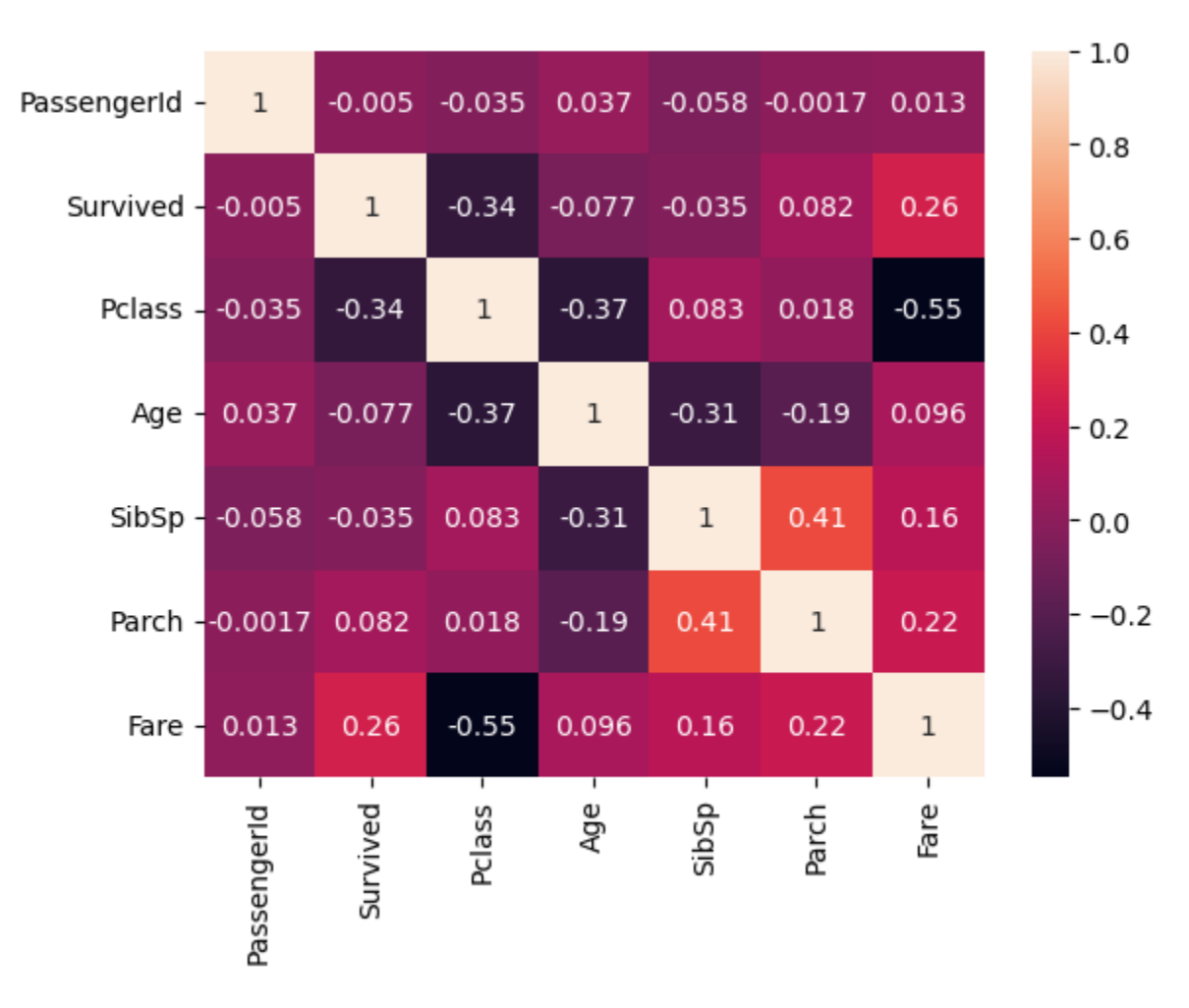
Out[13]:
```
<matplotlib.collections.PathCollection at 0x22c05ad2d40>
```



In [14]:
```python
sns.heatmap(df.corr(),annot=True)
```

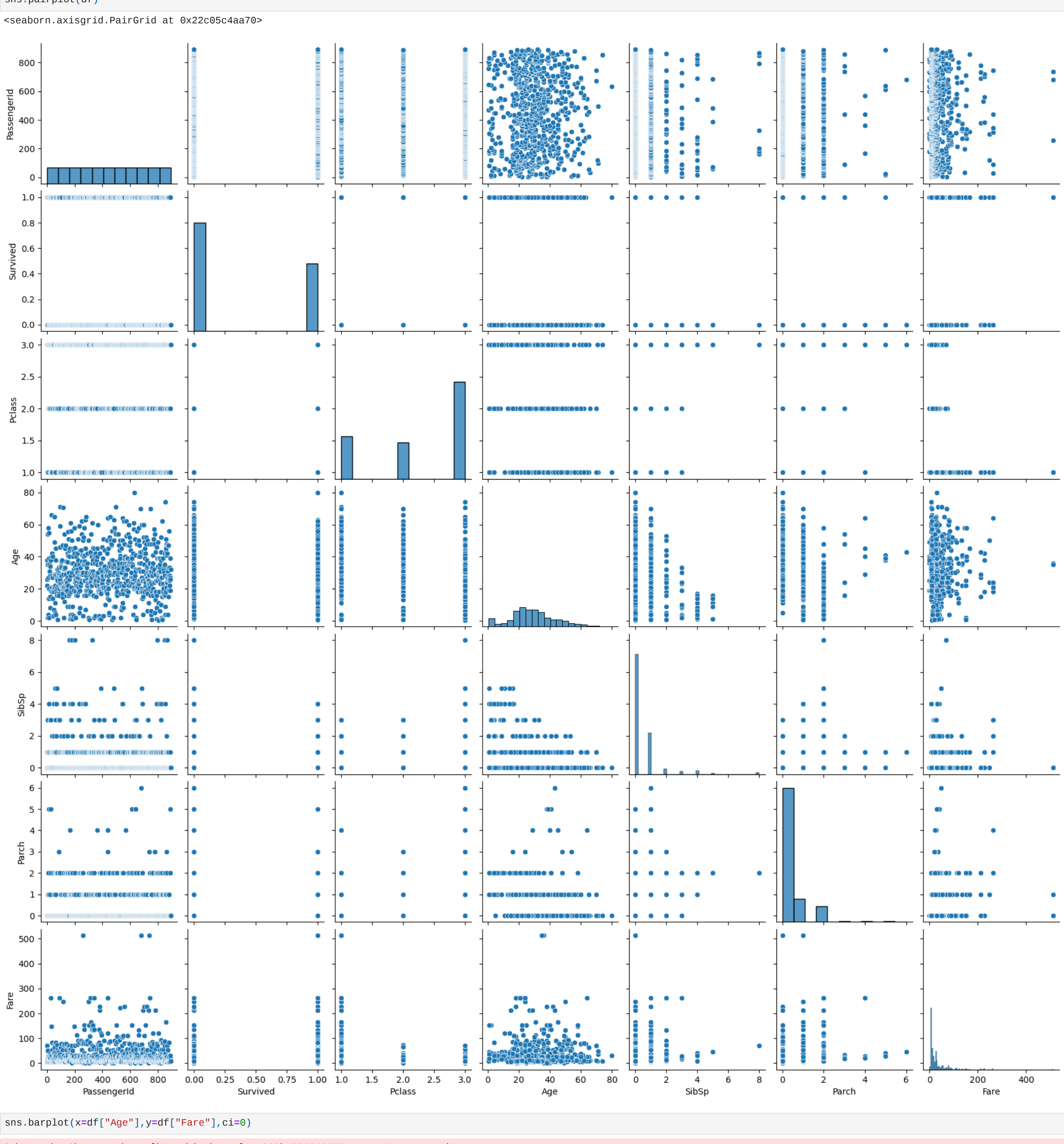
```
C:\Users\A C\AppData\Local\Temp\ipykernel_14880\4277794465.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  sns.heatmap(df.corr(),annot=True)
```

Out[14]:
```
<Axes: >
```



In [15]:
```python
sns.pairplot(df)
```

Out[15]:
```
<seaborn.axisgrid.PairGrid at 0x22c05c4aa70>
```



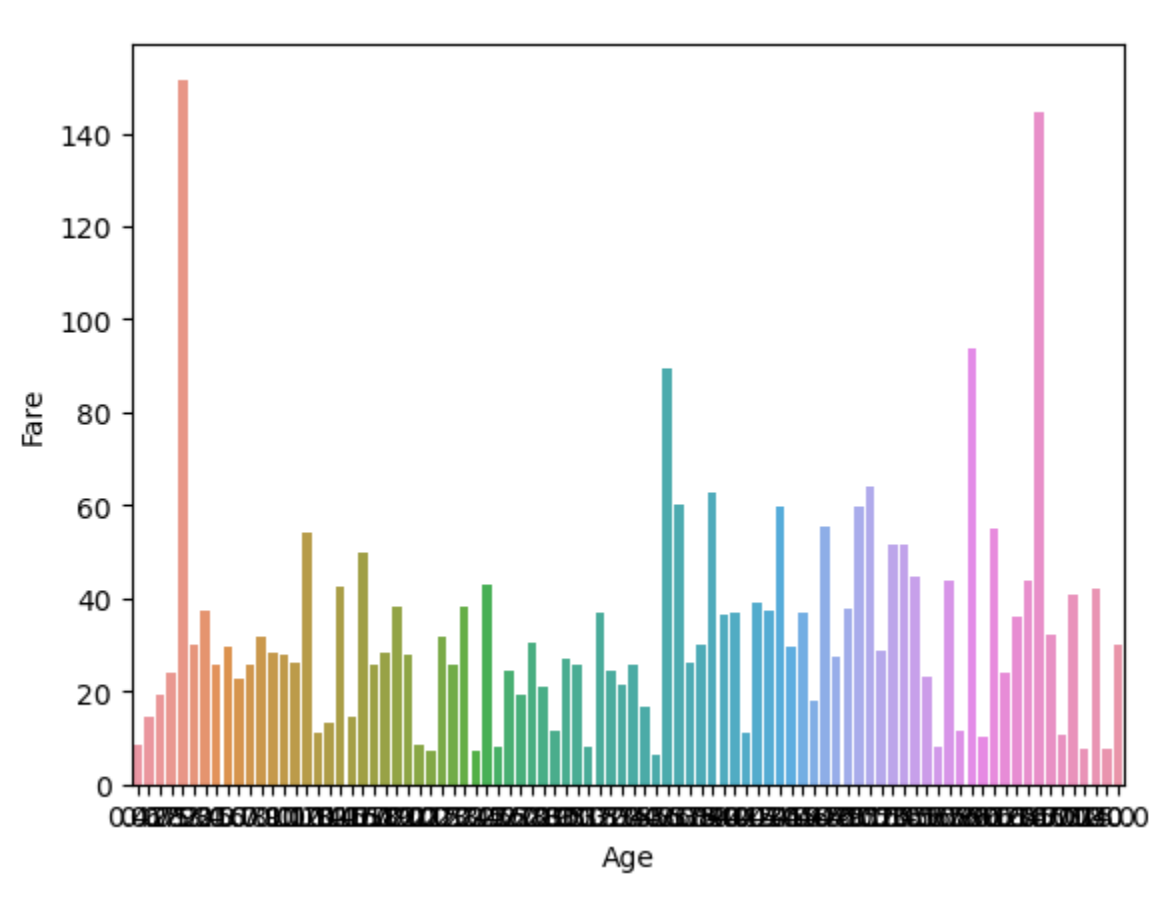
In [16]:
```python
sns.barplot(x=df["Age"],y=df["Fare"],ci=0)
```

```
C:\Users\A C\AppData\Local\Temp\ipykernel_14880\1989656056.py:1: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=('ci', 0)` for the same effect.

  sns.barplot(x=df["Age"],y=df["Fare"],ci=0)
```

Out[16]:
```
<Axes: xlabel='Age', ylabel='Fare'>
```



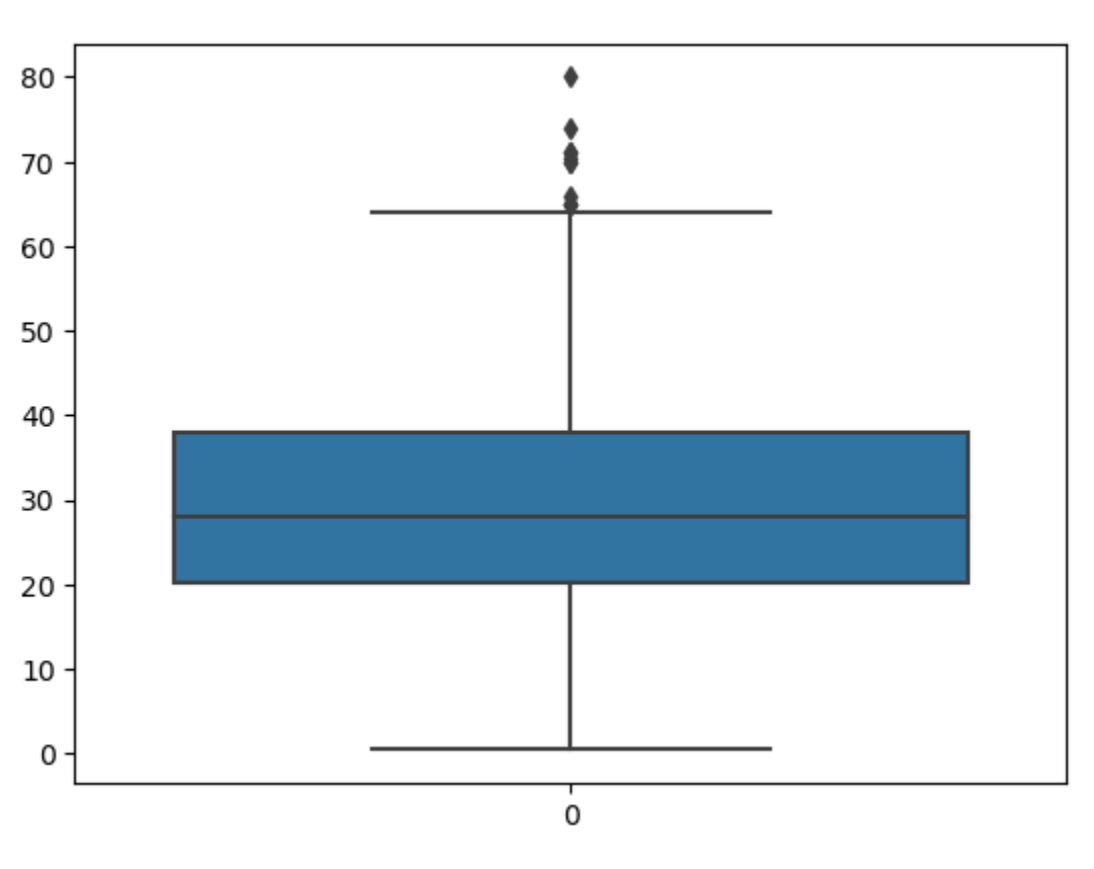
In [17]:
```python
df.head()
```

Out[17]:

	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

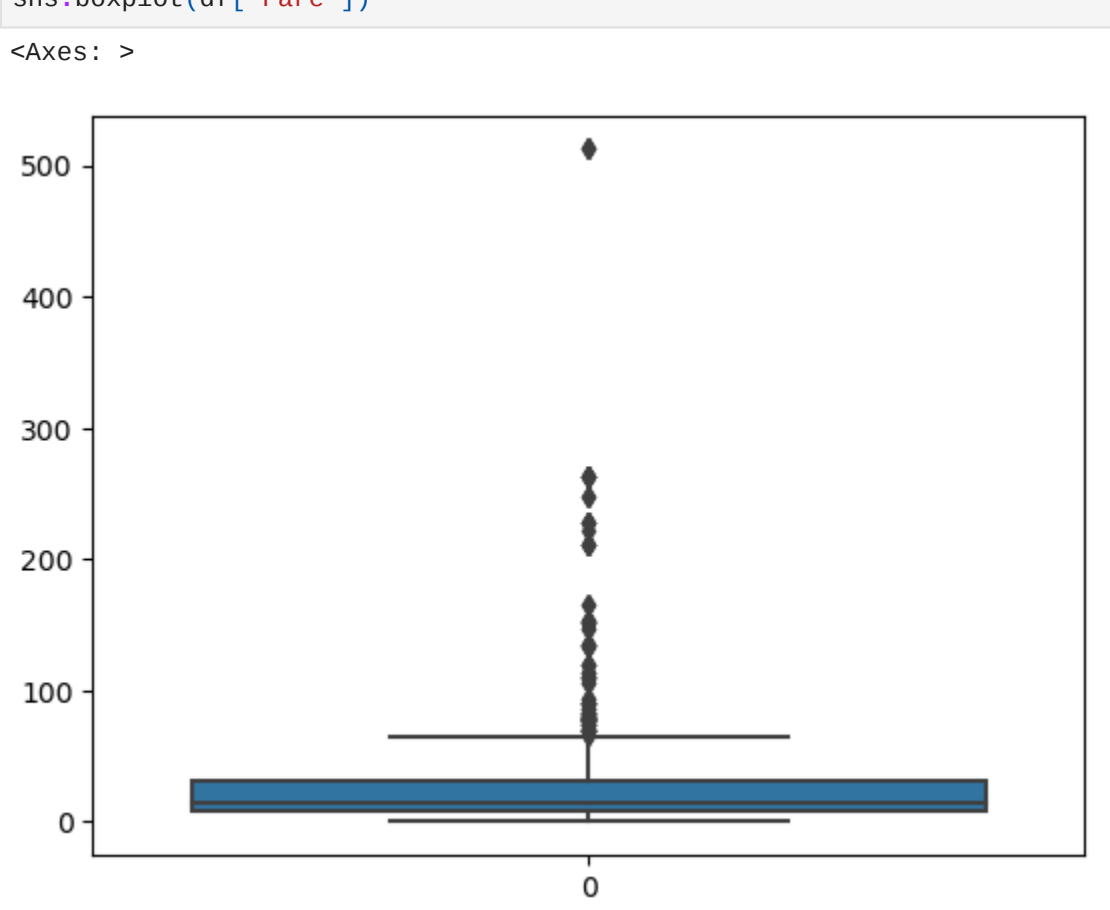
In [18]:
```python
sns.boxplot(df["Age"])
```

Out[18]:
```
<Axes: >
```



In [19]:
```python
sns.boxplot(df["Fare"])
```

Out[19]:
```
<Axes: >
```



In [20]:
```python
df.head()
```

Out[20]:

	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

In [21]:
```python
X=df.drop(columns=["Cabin","Name","Sex","Embarked","Ticket"],axis=1)
X.head()
```

Out[21]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
0	1	0	3	22.0	1	0	7.2500
1	2	1	1	38.0	1	0	71.2833
2	3	1	3	26.0	0	0	7.9250
3	4	1	1	35.0	1	0	53.1000
4	5	0	3	35.0	0	0	8.0500

In [22]:
```python
X.shape
```

Out[22]:
```
(891, 7)
```

In [23]:
```python
type(X)
```

Out[23]:
```
pandas.core.frame.DataFrame
```

In [24]:
```python
y=df["Cabin"]
y.head()
```

Out[24]:
```
0     NaN
1     C85
2     NaN
3    C123
4     NaN
Name: Cabin, dtype: object
```

In [25]:
```python
y=df["Name"]
y.head
```

Out[25]:
```
<bound method NDFrame.head of 0                              Braund, Mr. Owen Harris
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
2                                 Heikkinen, Miss. Laina
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
4                               Allen, Mr. William Henry
                             ...
886                                Montvila, Rev. Juozas
887                         Graham, Miss. Margaret Edith
888             Johnston, Miss. Catherine Helen "Carrie"
889                                Behr, Mr. Karl Howell
890                                  Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object>
```

In [26]:
```python
X.head()
```

Out[26]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
0	1	0	3	22.0	1	0	7.2500
1	2	1	1	38.0	1	0	71.2833
2	3	1	3	26.0	0	0	7.9250
3	4	1	1	35.0	1	0	53.1000
4	5	0	3	35.0	0	0	8.0500

In [28]:
```python
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
X["age"]=le.fit_transform(X["Age"])
X.head()
```

Out[28]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	age
0	1	0	3	22.0	1	0	7.2500	28
1	2	1	1	38.0	1	0	71.2833	51
2	3	1	3	26.0	0	0	7.9250	34
3	4	1	1	35.0	1	0	53.1000	47
4	5	0	3	35.0	0	0	8.0500	47

In [29]:
```python
print(le.classes_)
```

```
[ 0.42  0.67  0.75  0.83  0.92  1.    2.    3.    4.    5.    6.    7.
  8.    9.   10.   11.   12.   13.   14.   14.5  15.   16.   17.   18.
 19.   20.   20.5  21.   22.   23.   23.5  24.   24.5  25.   26.   27.
 28.   28.5  29.   30.   30.5  31.   32.   32.5  33.   34.   34.5  35.
 36.   36.5  37.   38.   39.   40.   40.5  41.   42.   43.   44.   45.
 45.5  46.   47.   48.   49.   50.   51.   52.   53.   54.   55.   55.5
 56.   57.   58.   59.   60.   61.   62.   63.   64.   65.   66.   70.
 70.5  71.   74.   80.     nan]
```

In [30]:
```python
mapping=dict(zip(le.classes_,range(len(le.classes_))))
mapping
```

Out[30]:
```
{0.42: 0,
 0.67: 1,
 0.75: 2,
 0.83: 3,
 0.92: 4,
 1.0: 5,
 2.0: 6,
 3.0: 7,
 4.0: 8,
 5.0: 9,
 6.0: 10,
 7.0: 11,
 8.0: 12,
 9.0: 13,
 10.0: 14,
 11.0: 15,
 12.0: 16,
 13.0: 17,
 14.0: 18,
 14.5: 19,
 15.0: 20,
 16.0: 21,
 17.0: 22,
 18.0: 23,
 19.0: 24,
 20.0: 25,
 20.5: 26,
 21.0: 27,
 22.0: 28,
 23.0: 29,
 23.5: 30,
 24.0: 31,
 24.5: 32,
 25.0: 33,
 26.0: 34,
 27.0: 35,
 28.0: 36,
 28.5: 37,
 29.0: 38,
 30.0: 39,
 30.5: 40,
 31.0: 41,
 32.0: 42,
 32.5: 43,
 33.0: 44,
 34.0: 45,
 34.5: 46,
 35.0: 47,
 36.0: 48,
 36.5: 49,
 37.0: 50,
 38.0: 51,
 39.0: 52,
 40.0: 53,
 40.5: 54,
 41.0: 55,
 42.0: 56,
 43.0: 57,
 44.0: 58,
 45.0: 59,
 45.5: 60,
 46.0: 61,
 47.0: 62,
 48.0: 63,
 49.0: 64,
 50.0: 65,
 51.0: 66,
 52.0: 67,
 53.0: 68,
 54.0: 69,
 55.0: 70,
 55.5: 71,
 56.0: 72,
 57.0: 73,
 58.0: 74,
 59.0: 75,
 60.0: 76,
 61.0: 77,
 62.0: 78,
 63.0: 79,
 64.0: 80,
 65.0: 81,
 66.0: 82,
 70.0: 83,
 70.5: 84,
 71.0: 85,
 74.0: 86,
 80.0: 87,
 nan: 88}
```

In [31]:
```python
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
X_Scaled=ms.fit_transform(X)
X_Scaled=pd.DataFrame(ms.fit_transform(X),columns=X.columns)
X_Scaled.head()
```

Out[31]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	age
0	0.000000	0.0	1.0	0.271174	0.125	0.0	0.014151	0.318182
1	0.001124	1.0	0.0	0.472229	0.125	0.0	0.139136	0.579545
2	0.002247	1.0	1.0	0.321438	0.000	0.0	0.015469	0.386364
3	0.003371	1.0	0.0	0.434531	0.125	0.0	0.103644	0.534091
4	0.004494	0.0	1.0	0.434531	0.000	0.0	0.015713	0.534091

In [32]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(X_Scaled,y,test_size =0.2,random_state =0)
```

In [33]:
```python
print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)
```

```
(712, 8) (179, 8) (712,) (179,)
```

In [ ]: