assignment-3

September 20, 2023

0.1 ASSIGNMENT-3

Shaik Mansoor Vali

1 Steps:

1.import the necessary libraries

2.import the dataset

3.Handling null values

4.outlier detection—surya

5.Seperate Dependent and independent variables

6.Encoding

7.splitting into training and testing set

8.Feature scaling

1.1 1.import the necessary libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
```

1.2 2.import the dataset

[2]:
```
#.csv  .tsv ,json,.excel
dataset=pd.read_csv("Titanic-Dataset.csv")
#dataset=pd.read_csv(r"D:\programs\Python programs\smart␣
 ↪internz\Churn_Modelling.csv")
```

[3]:
```
dataset
```

[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]:
```
dataset.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]:
```
dataset.tail()
```

[5]:
```
     PassengerId  Survived  Pclass                                      Name  \
886          887         0       2                     Montvila, Rev. Juozas
887          888         1       1              Graham, Miss. Margaret Edith
888          889         0       3  Johnston, Miss. Catherine Helen "Carrie"
889          890         1       1                     Behr, Mr. Karl Howell
890          891         0       3                       Dooley, Mr. Patrick

        Sex   Age  SibSp  Parch      Ticket   Fare Cabin Embarked
886    male  27.0      0      0      211536  13.00   NaN        S
887  female  19.0      0      0      112053  30.00   B42        S
888  female   NaN      1      2  W./C. 6607  23.45   NaN        S
889    male  26.0      0      0      111369  30.00  C148        C
890    male  32.0      0      0      370376   7.75   NaN        Q
```

[6]:
```
dataset.shape
```

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

[7]:
```
dataset.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
```

[8]:
```
dataset.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]:
```
corr=dataset.corr()
corr
```

```
C:\Users\Mansoorvali\AppData\Local\Temp\ipykernel_4904\1091080309.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.
  corr=dataset.corr()
```

[9]:

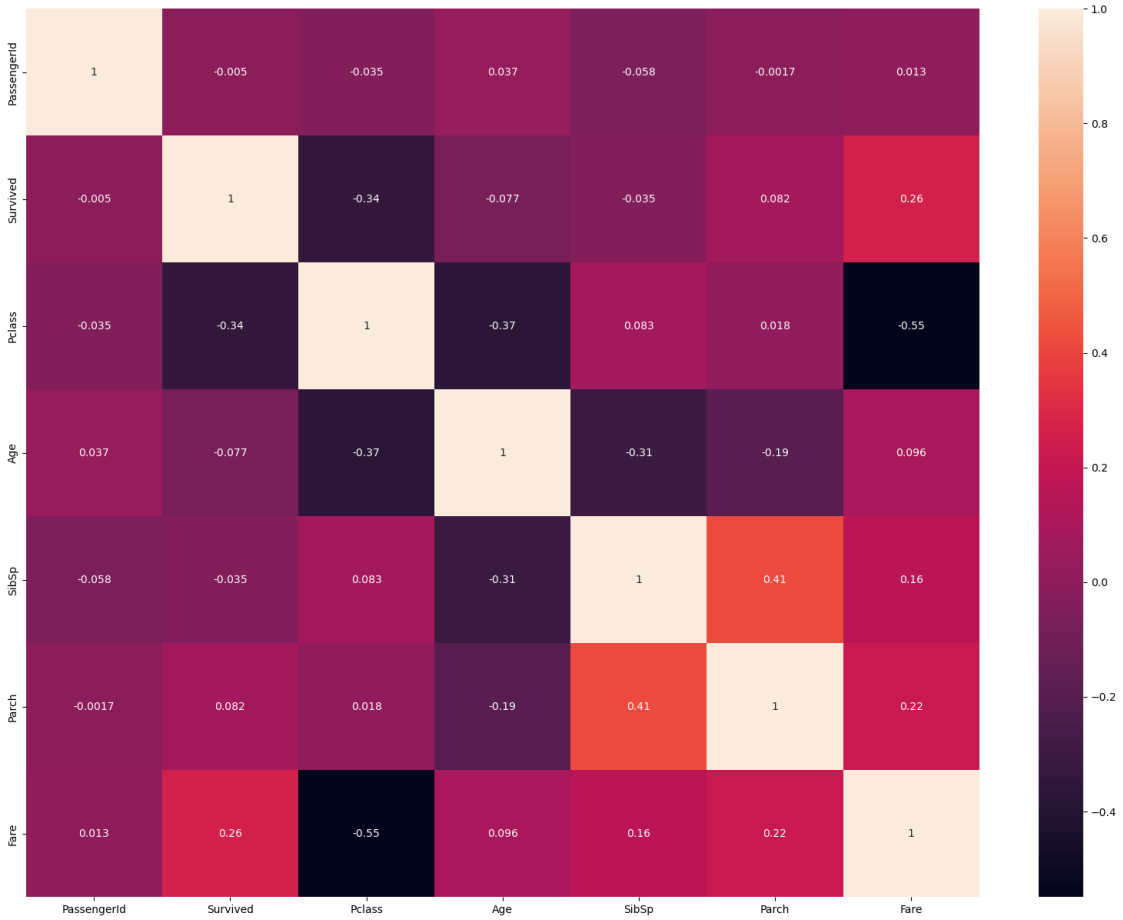
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
Age	0.036847	-0.077221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

```
                Fare
PassengerId  0.012658
Survived     0.257307
Pclass      -0.549500
Age          0.096067
SibSp        0.159651
Parch        0.216225
Fare         1.000000
```

[10]:
```
plt.subplots(figsize=(20,15))
sns.heatmap(corr,annot=True)
```

[10]: <Axes: >



[11]:
```
dataset.PassengerId.value_counts()
```

[11]:
```
1      1
599    1
588    1
589    1
590    1
      ..
301    1
302    1
303    1
304    1
891    1
Name: PassengerId, Length: 891, dtype: int64
```

[12]:
```
dataset.Survived.value_counts()
```

[12]:
```
0    549
1    342
Name: Survived, dtype: int64
```

[13]:
```
dataset.head()
```

[13]:

	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

[14]:
```
dataset.        Pclass.value_counts()
```

[14]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

[ ]:

1.3 3.Handling null values

[15]:
```
dataset.isnull().any()
```

[15]:
```
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
```

[16]:
```
dataset.isnull().sum()
```

[16]:
```
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
```

[17]:
```
dataset["Age"].fillna(dataset["Age"].mean(),inplace=True)
```

[18]:
```
dataset["Cabin"].fillna(dataset["Cabin"].mode()[0],inplace=True)
```

[19]:
```
dataset["Embarked"].fillna(dataset["Embarked"].mode()[0],inplace=True)
```

[20]:
```
dataset.isnull().sum()
```

[20]:
```
PassengerId    0
Survived       0
Pclass         0
```

```
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

[21]:
```
dataset.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	B96 B98	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	B96 B98	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	B96 B98	S

1.3.1 4.outliers

[22]:
```
z_scores = np.abs(stats.zscore(dataset['Age']))
max_threshold=3
outliers = dataset['Age'][z_scores > max_threshold]

# Print and visualize the outliers
print("Outliers detected using Z-Score:")
print(outliers)
```

```
Outliers detected using Z-Score:
96     71.0
116    70.5
493    71.0
```

```
630    80.0
672    70.0
745    70.0
851    74.0
Name: Age, dtype: float64
```

[23]:
```
z_scores = np.abs(stats.zscore(dataset['Fare']))
max_threshold=3
outliers = dataset['Fare'][z_scores > max_threshold]

# Print and visualize the outliers
print("Outliers detected using Z-Score:")
print(outliers)
```

```
Outliers detected using Z-Score:
27     263.0000
88     263.0000
118    247.5208
258    512.3292
299    247.5208
311    262.3750
341    263.0000
377    211.5000
380    227.5250
438    263.0000
527    221.7792
557    227.5250
679    512.3292
689    211.3375
700    227.5250
716    227.5250
730    211.3375
737    512.3292
742    262.3750
779    211.3375
Name: Fare, dtype: float64
```

[24]:
```
column_name = 'Fare'

# Calculate the first quartile (Q1) and third quartile (Q3)
Q1 = dataset[column_name].quantile(0.25)
Q3 = dataset[column_name].quantile(0.75)

# Calculate the IQR
IQR = Q3 - Q1

# Define the lower and upper bounds for outliers
lower_bound = Q1 - 1.5 * IQR
```

```
upper_bound = Q3 + 1.5 * IQR

# Filter rows with values outside the IQR bounds
dataset_cleaned = dataset[(dataset[column_name] > lower_bound) &␣
 ↪(dataset[column_name] <upper_bound)]

# Display the original and cleaned DataFrame sizes
print(f"Original DataFrame size: {dataset.shape}")
print(f"Cleaned DataFrame size: {dataset_cleaned.shape}")
dataset_cleaned
```

```
Original DataFrame size: (891, 12)
Cleaned DataFrame size: (775, 12)
```

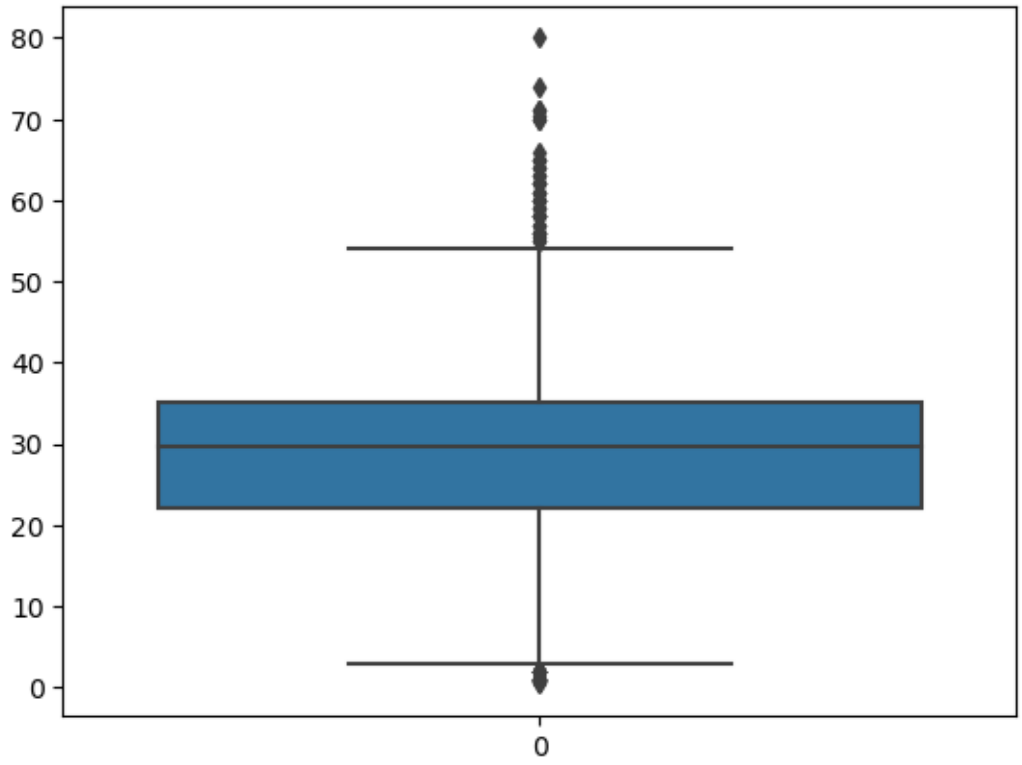
[24]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	B96 B98	S
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9250	B96 B98	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0500	B96 B98	S
5	6	0	3	Moran, Mr. James	male	29.699118	0	0	330877	8.4583	B96 B98	Q
..	…	…	…	…	…	…	…	…	…	…	…	…
886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0				
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0				
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	0				
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0				
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0				

```
886      0            211536  13.0000  B96 B98        S
887      0            112053  30.0000      B42        S
888      2        W./C. 6607  23.4500  B96 B98        S
889      0            111369  30.0000     C148        C
890      0            370376   7.7500  B96 B98        Q

[775 rows x 12 columns]
```

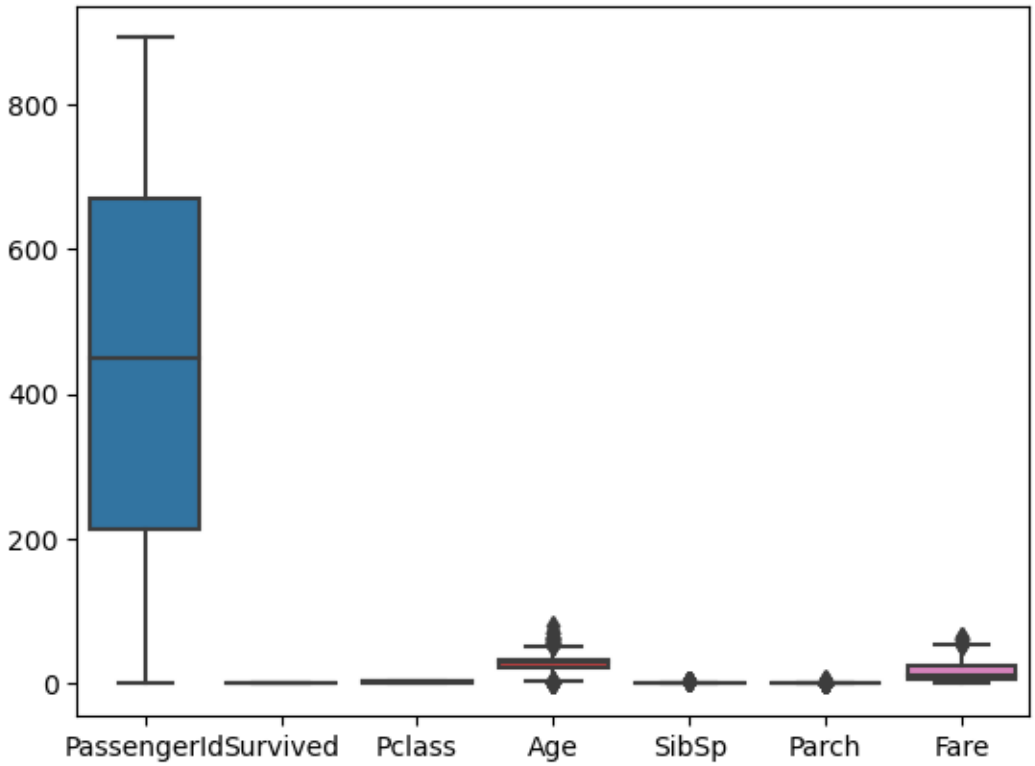
[25]:
```
sns.boxplot(dataset.Age)
```

[25]: <Axes: >



[26]:
```
sns.boxplot(dataset_cleaned)
```

[26]: <Axes: >



[27]:
```
dataset=dataset_cleaned
```

[28]:
```
x=dataset.drop('Survived', axis=1)
y=dataset['Survived']
```

[29]:
```
x.head()
```

[29]:

	PassengerId	Pclass	Name	Sex
0	1	3	Braund, Mr. Owen Harris	male
2	3	3	Heikkinen, Miss. Laina	female
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female
4	5	3	Allen, Mr. William Henry	male
5	6	3	Moran, Mr. James	male

	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	22.000000	1	0	A/5 21171	7.2500	B96 B98	S
2	26.000000	0	0	STON/O2. 3101282	7.9250	B96 B98	S
3	35.000000	1	0	113803	53.1000	C123	S
4	35.000000	0	0	373450	8.0500	B96 B98	S
5	29.699118	0	0	330877	8.4583	B96 B98	Q

[30]:
```
y.head()
```

[30]:
```
0    0
2    1
3    1
4    0
5    0
Name: Survived, dtype: int64
```

1.4 5.Seperate dependent and independent variables

[31]:
```
#datset.iloc[rows,column]
x=dataset.iloc[:,3:13]
y=dataset.iloc[:,13:14]
```

[32]:
```
x.head()
```

[32]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 21171	7.2500	B96 B98	S
2	Heikkinen, Miss. Laina	female	26.000000	0	0	STON/O2. 3101282	7.9250	B96 B98	S
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0	113803	53.1000	C123	S
4	Allen, Mr. William Henry	male	35.000000	0	0	373450	8.0500	B96 B98	S
5	Moran, Mr. James	male	29.699118	0	0	330877	8.4583	B96 B98	Q

[33]:
```
y.head()
```

[33]:
```
Empty DataFrame
Columns: []
Index: [0, 2, 3, 4, 5]
```

[34]:
```
dataset.shape
```

[34]:
```
(775, 12)
```

[35]:
```
x.shape
```

[35]:
```
(775, 9)
```

[36]:
```
y.shape
```

[36]:
```
(775, 0)
```

1.5 6.Encoding

1.5.1 Label encoding on Gender column

[37]:
```
from sklearn.preprocessing import LabelEncoder
```

[38]:
```
le=LabelEncoder()
```

[39]:
```
x["Sex"]=le.fit_transform(x["Sex"])
```

[40]:
```
x["Sex"]
```

[40]:
```
0      1
2      0
3      0
4      1
5      1
      ..
886    1
887    0
888    0
889    1
890    1
Name: Sex, Length: 775, dtype: int32
```

[41]:
```
x["Sex"].value_counts()
```

[41]:
```
1    531
0    244
Name: Sex, dtype: int64
```

[42]:
```
x["Sex"].nunique()
```

[42]:
```
2
```

[43]:
```
x.head()
```

[43]:

	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	Braund, Mr. Owen Harris	1	22.000000	1	0	A/5 21171	7.2500	B96 B98	S
2	Heikkinen, Miss. Laina	0	26.000000	0	0	STON/O2. 3101282	7.9250	B96 B98	S
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.000000	1	0	113803	53.1000	C123	S
4	Allen, Mr. William Henry	1	35.000000	0	0	373450	8.0500	B96 B98	S
5	Moran, Mr. James	1	29.699118	0	0				

```
5            330877   8.4583  B96 B98        Q
```

[44]:
```
x.Sex.value_counts()
```

[44]:
```
1    531
0    244
Name: Sex, dtype: int64
```

1.5.2 One hot encoding on geography column

[45]:
```
x.shape
```

[45]:
```
(775, 9)
```

[46]:
```
sex=pd.get_dummies(x["Sex"],drop_first=True)
```

[47]:
```
sex
```

[47]:
```
     1
0    1
2    0
3    0
4    1
5    1
..  ..
886  1
887  0
888  0
889  1
890  1

[775 rows x 1 columns]
```

[48]:
```
#concat
x=pd.concat([x,sex],axis=1)
```

[49]:
```
x.head()
```

[49]:
```
                                          Name  Sex        Age  SibSp  Parch  \
0                      Braund, Mr. Owen Harris    1  22.000000      1      0
2                       Heikkinen, Miss. Laina    0  26.000000      0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    0  35.000000      1      0
4                     Allen, Mr. William Henry    1  35.000000      0      0
5                             Moran, Mr. James    1  29.699118      0      0

            Ticket     Fare    Cabin Embarked  1
0        A/5 21171   7.2500  B96 B98        S  1
2  STON/O2. 3101282   7.9250  B96 B98        S  0
```

```
3            113803  53.1000     C123        S  0
4            373450   8.0500  B96 B98        S  1
5            330877   8.4583  B96 B98        Q  1
```

[50]:
```
x.drop(["Sex"],axis=1,inplace=True)
```

[51]:
```
x.head(10)
```

[51]:
```
                                                Name        Age  SibSp  \
0                            Braund, Mr. Owen Harris  22.000000      1
2                             Heikkinen, Miss. Laina  26.000000      0
3    Futrelle, Mrs. Jacques Heath (Lily May Peel)  35.000000      1
4                          Allen, Mr. William Henry  35.000000      0
5                                  Moran, Mr. James  29.699118      0
6                           McCarthy, Mr. Timothy J  54.000000      0
7                    Palsson, Master. Gosta Leonard   2.000000      3
8   Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  27.000000      0
9                 Nasser, Mrs. Nicholas (Adele Achem)  14.000000      1
10                    Sandstrom, Miss. Marguerite Rut   4.000000      1

    Parch            Ticket     Fare    Cabin Embarked  1
0       0         A/5 21171   7.2500  B96 B98        S  1
2       0  STON/O2. 3101282   7.9250  B96 B98        S  0
3       0            113803  53.1000     C123        S  0
4       0            373450   8.0500  B96 B98        S  1
5       0            330877   8.4583  B96 B98        Q  1
6       0             17463  51.8625      E46        S  1
7       1            349909  21.0750  B96 B98        S  1
8       2            347742  11.1333  B96 B98        S  0
9       0            237736  30.0708  B96 B98        C  0
10      1          PP 9549  16.7000       G6        S  0
```

[52]:
```
x.shape
```

[52]:
```
(775, 9)
```

1.5.3 7.splitting into training and testing set

[53]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

[54]:
```
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(542, 9)
(233, 9)
```

```
(542, 0)
(233, 0)
```

[55]:
```
a=[1,2,3,4,5,6]
b=[1,0,1,5,6,3]

for i in range(5):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.
 ↪3,random_state=100)
    print("with random state",a_train)
```

```
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
with random state [5, 4, 6, 1]
```

[56]:
```
a=[1,2,3,4,5,6]
b=[1,0,1,5,6,3]

for i in range(5):
    a_train,a_test,b_train,b_test=train_test_split(a,b,test_size=0.3)
    print("without random state",a_train)
```

```
without random state [2, 4, 3, 1]
without random state [4, 6, 5, 2]
without random state [5, 2, 6, 3]
without random state [5, 6, 4, 2]
without random state [3, 6, 1, 2]
```

1.5.4 8.Feature Scaling

[57]:
```
scale = StandardScaler()
x[['Age', 'Fare']] = scale.fit_transform(x[['Age', 'Fare']])
```

[58]:
```
x.head()
```

[58]:

	Name	Age	SibSp	Parch	\
0	Braund, Mr. Owen Harris	-0.556219	1	0	
2	Heikkinen, Miss. Laina	-0.243027	0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0.461654	1	0	
4	Allen, Mr. William Henry	0.461654	0	0	
5	Moran, Mr. James	0.046606	0	0	

	Ticket	Fare	Cabin	Embarked	1
0	A/5 21171	-0.779117	B96 B98	S	1
2	STON/O2. 3101282	-0.729373	B96 B98	S	0
3	113803	2.599828	C123	S	0

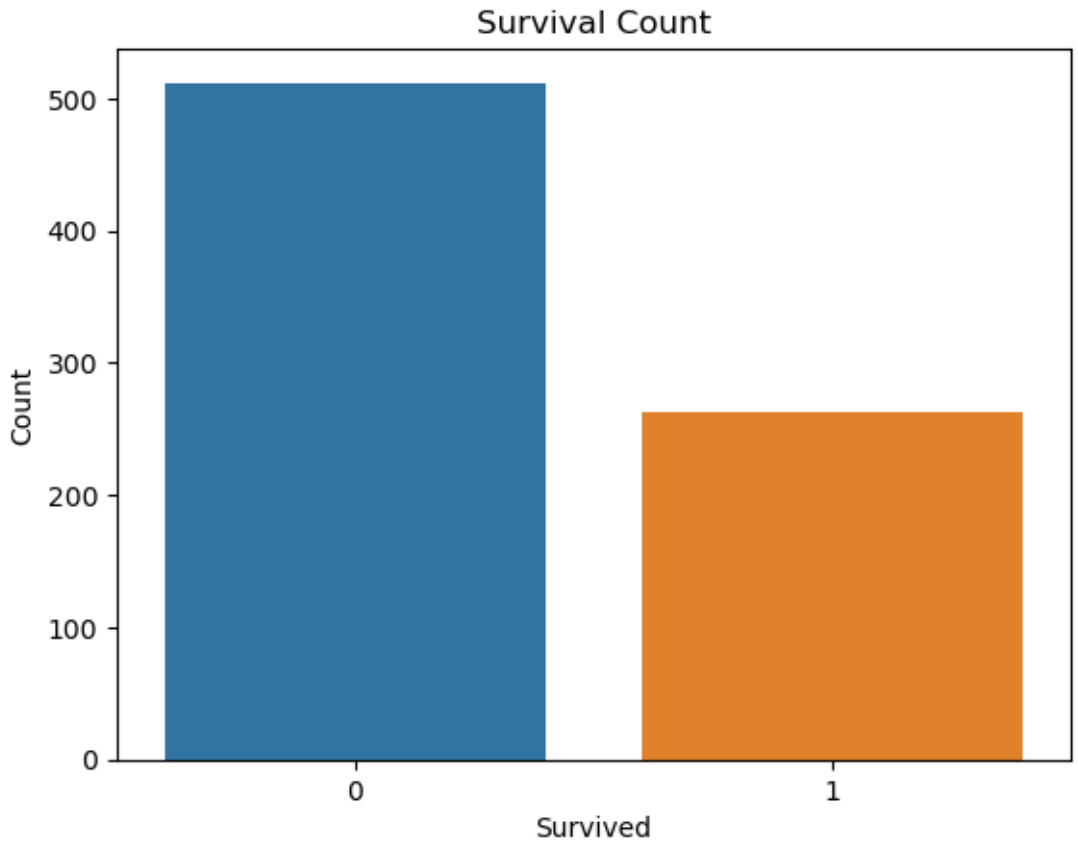
```
4            373450 -0.720161  B96 B98        S  1
5            330877 -0.690071  B96 B98        Q  1
```

[59]:
```
x_train
```

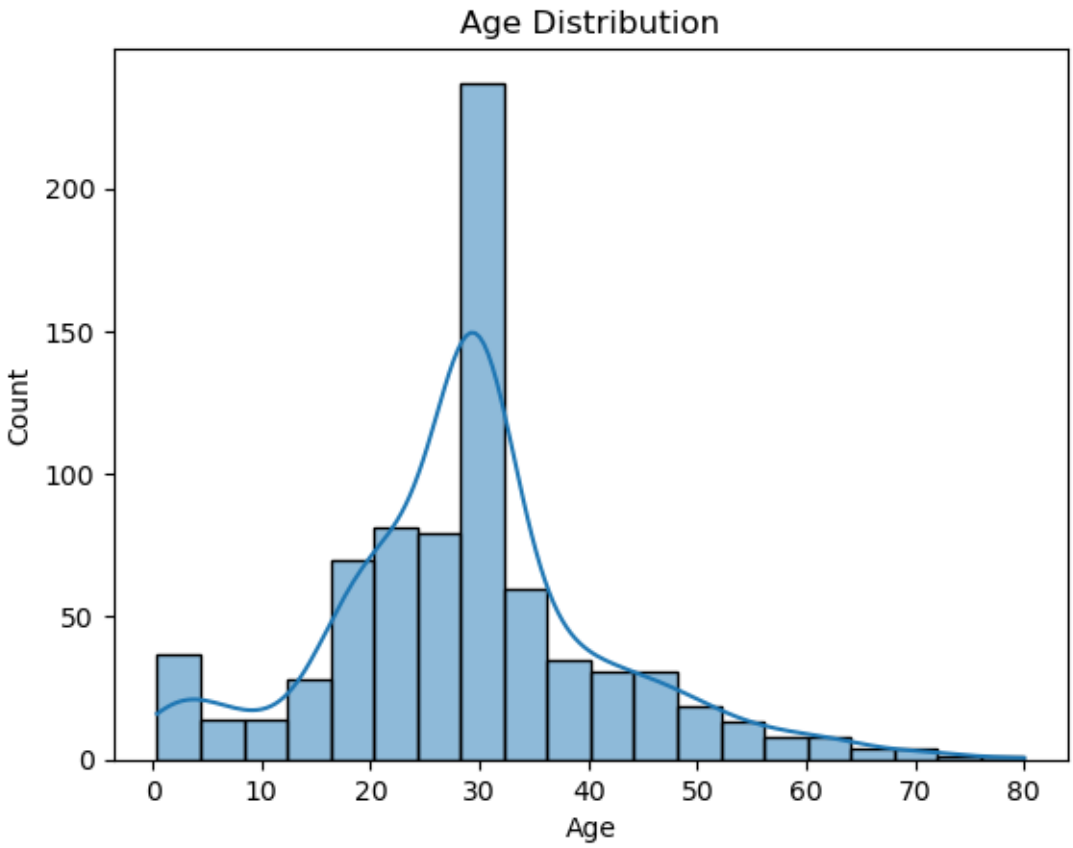
[59]:
```
                                    Name        Age  SibSp  Parch  \
654        Hegarty, Miss. Hanora "Nora"  18.000000      0      0
38   Vander Planke, Miss. Augusta Maria  18.000000      2      0
646                    Cor, Mr. Liudevit  19.000000      0      0
727            Mannion, Miss. Margareth  29.699118      0      0
887        Graham, Miss. Margaret Edith  19.000000      0      0
..                                   …        …    …    …
878                  Laleff, Mr. Kristo  29.699118      0      0
211          Cameron, Miss. Clear Annie  35.000000      0      0
725                Oreskovic, Mr. Luka  20.000000      0      0
643                    Foo, Mr. Choong  29.699118      0      0
790            Keane, Mr. Andrew "Andy"  29.699118      0      0

           Ticket     Fare    Cabin Embarked  1
654        365226   6.7500  B96 B98        Q  0
38         345764  18.0000  B96 B98        S  0
646        349231   7.8958  B96 B98        S  1
727         36866   7.7375  B96 B98        Q  0
887        112053  30.0000      B42        S  0
..            …       …       …      … ..
878        349217   7.8958  B96 B98        S  1
211  F.C.C. 13528  21.0000  B96 B98        S  0
725        315094   8.6625  B96 B98        S  1
643          1601  56.4958  B96 B98        S  1
790         12460   7.7500  B96 B98        Q  1

[542 rows x 9 columns]
```

1.6 DATA VISUALIZATION

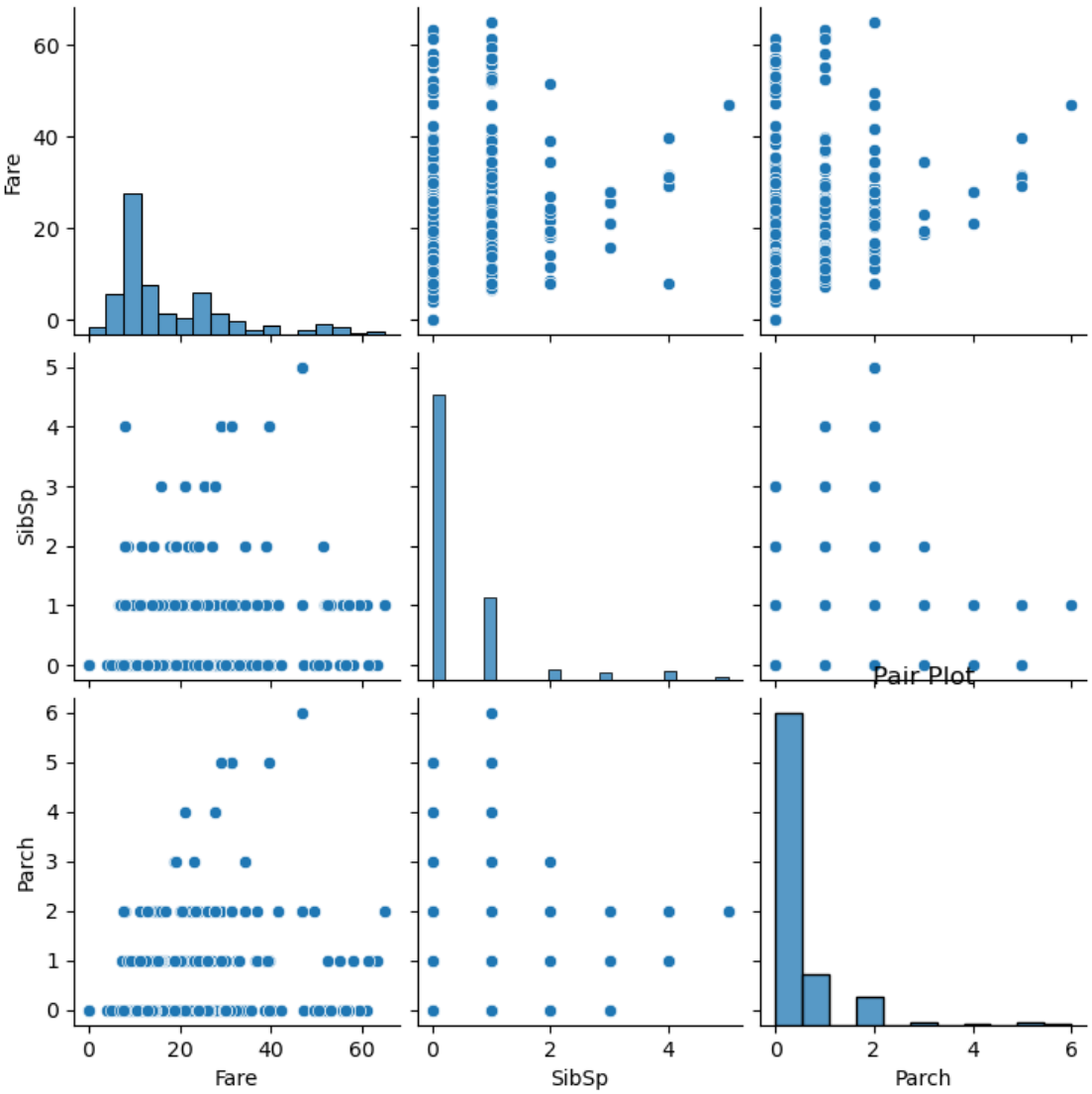
[60]:
```
sns.countplot(data=dataset, x='Survived')
plt.title('Survival Count')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.show()
```



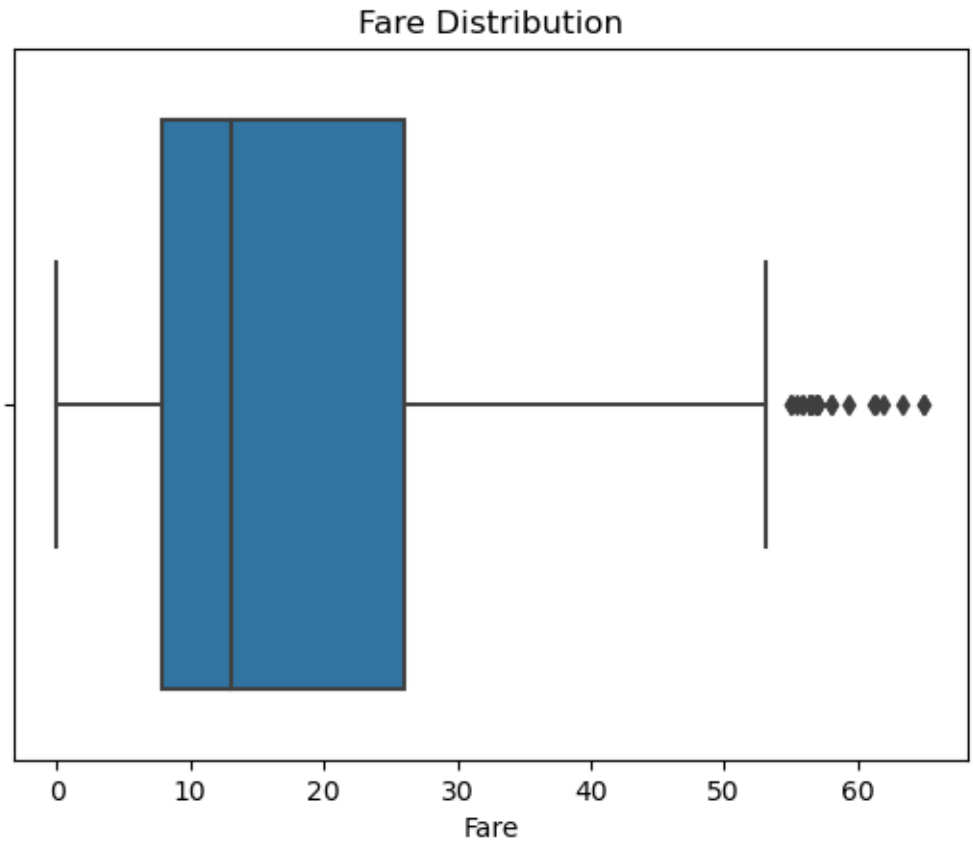
[61]:
```
sns.histplot(data=dataset, x='Age', bins=20, kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```



[62]:
```
sns.pairplot(data=dataset[['Fare', 'SibSp', 'Parch']])
plt.title('Pair Plot')
plt.show()
```

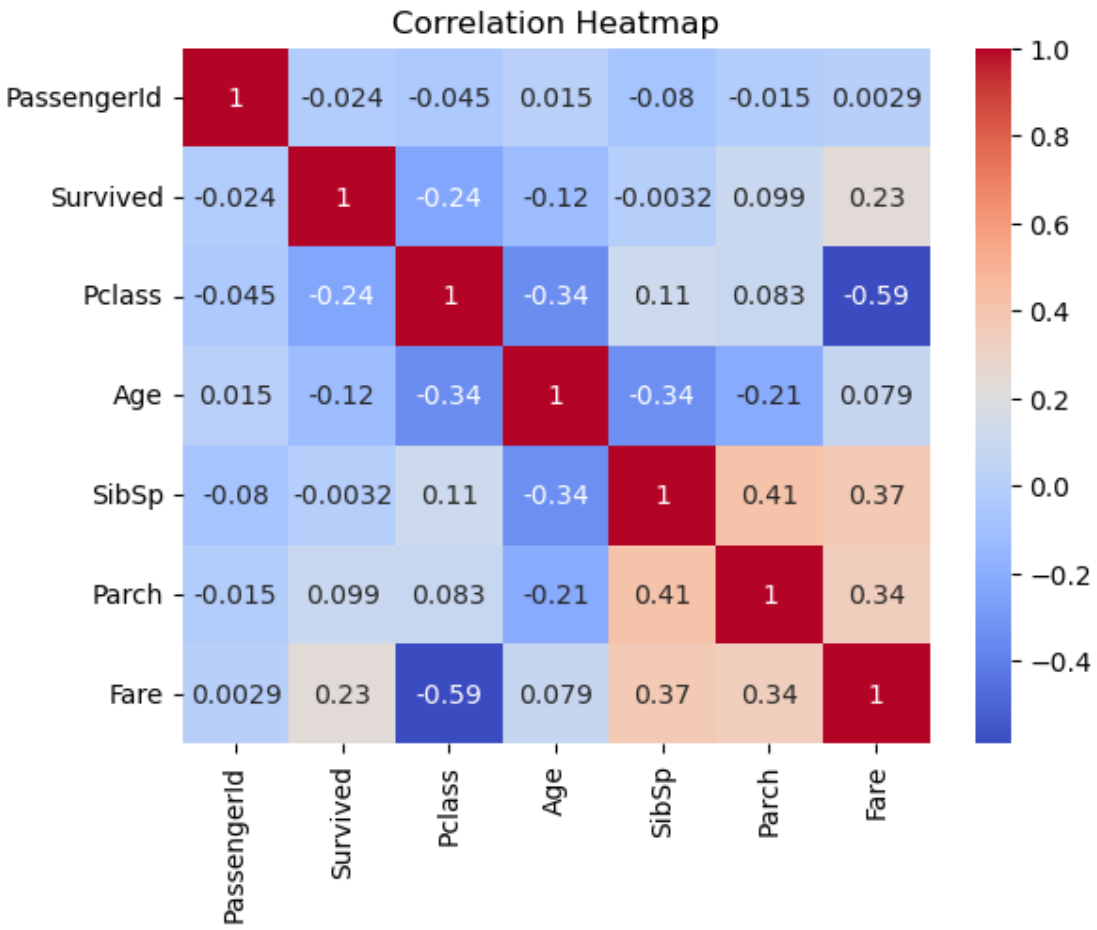


[63]:
```
sns.boxplot(data=dataset, x='Fare')
plt.title('Fare Distribution')
plt.xlabel('Fare')
plt.show()
```



[64]:
```
corr_matrix = dataset.corr()
sns.heatmap(corr_matrix, annot=True,cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

```
C:\Users\Mansoorvali\AppData\Local\Temp\ipykernel_4904\3492499936.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.
  corr_matrix = dataset.corr()
```


Correlation Heatmap
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1
-0.024
-0.045
0.015
-0.08
-0.015
0.0029
-0.024
1
-0.24
-0.12
-0.0032
0.099
0.23
-0.045
-0.24
1
-0.34
0.11
0.083
-0.59
0.015
-0.12
-0.34
1
-0.34
-0.21
0.079
-0.08
-0.0032
0.11
-0.34
1
0.41
0.37
-0.015
0.099
0.083
-0.21
0.41
1
0.34
0.0029
0.23
-0.59
0.079
0.37
0.34
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4