**Assignment No 9**

**Code:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from seaborn import load\_dataset

#print(sns.get\_dataset\_names())

df=pd.read\_csv("tested.csv")

#print(df)

tips=load\_dataset("tips")

# Display basic information

print('Information of Dataset:\n', df.info)

print('Shape of Dataset (row x column): ', df.shape)

print('Columns Name: ', df.columns)

print('Total elements in dataset:', df.size)

print('Datatype of attributes (columns):', df.dtypes)

print('First 5 rows:\n', df.head().T)

print('Last 5 rows:\n',df.tail().T)

print('Any 5 rows:\n',df.sample(5).T)

#---------------------------------------------------------------------------------------

# Display Statistical information

print('Statistical information of Numerical Columns: \n',df.describe())

#---------------------------------------------------------------------------------------

# Display and fill the Null values

print('Total Number of Null Values in Dataset:', df.isna().sum())

df['Age'].fillna(df['Age'].median())

print('Total Number of Null Values in Dataset:', df.isna().sum())

#---------------------------------------------------------------------------------------

#One variable

fig, axes = plt.subplots(1,2)

sns.boxplot(data = df, y ='Age', ax=axes[0])

sns.boxplot(data = df, y ='Fare', ax=axes[1])

plt.show()

# Two variables

fig, axes = plt.subplots(1,3, sharey=True)

sns.boxplot(data = df, x='Sex', y ='Age', hue = 'Sex', ax=axes[0])

sns.boxplot(data = df, x='Pclass', y ='Age', hue = 'Pclass', ax=axes[1])

sns.boxplot(data = df, x='Survived', y ='Age', hue = 'Survived', ax=axes[2])

plt.show()

# Two variables

fig, axes = plt.subplots(1,3, sharey=True)

sns.boxplot(data = df, x='Sex', y ='Fare', hue = 'Sex', ax=axes[0], log\_scale = True)

sns.boxplot(data = df, x='Pclass', y ='Fare', hue = 'Pclass', ax=axes[1], log\_scale = True)

sns.boxplot(data = df, x='Survived', y ='Fare', hue = 'Survived', ax=axes[2], log\_scale = True)

plt.show()

#three variables

fig, axes = plt.subplots(1,2, sharey=True)

sns.boxplot(data = df, x='Sex', y ='Age', hue = 'Survived', ax=axes[0])

sns.boxplot(data = df, x='Pclass', y ='Age', hue = 'Survived', ax=axes[1])

plt.show()

fig, axes = plt.subplots(1,2, sharey=True)

sns.boxplot(data = df, x='Sex', y ='Fare', hue = 'Survived', ax=axes[0], log\_scale = True)

sns.boxplot(data = df, x='Pclass', y ='Fare', hue = 'Survived', ax=axes[1], log\_scale = True)

plt.show()

**Output:**

Information of Dataset:

 <bound method DataFrame.info of      PassengerId  Survived  Pclass  ...      Fare Cabin  Embarked

0            892         0       3  ...    7.8292   NaN         Q

1            893         1       3  ...    7.0000   NaN         S

2            894         0       2  ...    9.6875   NaN         Q

3            895         0       3  ...    8.6625   NaN         S

4            896         1       3  ...   12.2875   NaN         S

..           ...       ...     ...  ...       ...   ...       ...

413         1305         0       3  ...    8.0500   NaN         S

414         1306         1       1  ...  108.9000  C105         C

415         1307         0       3  ...    7.2500   NaN         S

416         1308         0       3  ...    8.0500   NaN         S

417         1309         0       3  ...   22.3583   NaN         C

[418 rows x 12 columns]>

Shape of Dataset (row x column):  (418, 12)

Columns Name:  Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',

       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],

      dtype='object')

Total elements in dataset: 5016

Datatype of attributes (columns): PassengerId      int64

Survived         int64

Pclass           int64

Name            object

Sex             object

Age            float64

SibSp            int64

Parch            int64

Ticket          object

Fare           float64

Cabin           object

Embarked        object

dtype: object

First 5 rows:

                             0  ...                                             4

PassengerId               892  ...                                           896

Survived                    0  ...                                             1

Pclass                      3  ...                                             3

Name         Kelly, Mr. James  ...  Hirvonen, Mrs. Alexander (Helga E Lindqvist)

Sex                      male  ...                                        female

Age                      34.5  ...                                          22.0

SibSp                       0  ...                                             1

Parch                       0  ...                                             1

Ticket                 330911  ...                                       3101298

Fare                   7.8292  ...                                       12.2875

Cabin                     NaN  ...                                           NaN

Embarked                    Q  ...                                             S

[12 rows x 5 columns]

Last 5 rows:

                             413  ...                       417

PassengerId                1305  ...                      1309

Survived                      0  ...                         0

Pclass                        3  ...                         3

Name         Spector, Mr. Woolf  ...  Peter, Master. Michael J

Sex                        male  ...                      male

Age                         NaN  ...                       NaN

SibSp                         0  ...                         1

Parch                         0  ...                         1

Ticket                A.5. 3236  ...                      2668

Fare                       8.05  ...                   22.3583

Cabin                       NaN  ...                       NaN

Embarked                      S  ...                         C

[12 rows x 5 columns]

Any 5 rows:

                           116  ...                 152

PassengerId              1008  ...                1044

Survived                    0  ...                   0

Pclass                      3  ...                   3

Name         Thomas, Mr. John  ...  Storey, Mr. Thomas

Sex                      male  ...                male

Age                       NaN  ...                60.5

SibSp                       0  ...                   0

Parch                       0  ...                   0

Ticket                   2681  ...                3701

Fare                   6.4375  ...                 NaN

Cabin                     NaN  ...                 NaN

Embarked                    C  ...                   S

[12 rows x 5 columns]

Statistical information of Numerical Columns:

        PassengerId    Survived      Pclass  ...       SibSp       Parch        Fare

count   418.000000  418.000000  418.000000  ...  418.000000  418.000000  417.000000

mean   1100.500000    0.363636    2.265550  ...    0.447368    0.392344   35.627188

std     120.810458    0.481622    0.841838  ...    0.896760    0.981429   55.907576

min     892.000000    0.000000    1.000000  ...    0.000000    0.000000    0.000000

25%     996.250000    0.000000    1.000000  ...    0.000000    0.000000    7.895800

50%    1100.500000    0.000000    3.000000  ...    0.000000    0.000000   14.454200

75%    1204.750000    1.000000    3.000000  ...    1.000000    0.000000   31.500000

max    1309.000000    1.000000    3.000000  ...    8.000000    9.000000  512.329200

[8 rows x 7 columns]

Total Number of Null Values in Dataset: PassengerId      0

Survived         0

Pclass           0

Name             0

Sex              0

Age             86

SibSp            0

Parch            0

Ticket           0

Fare             1

Cabin          327

Embarked         0

dtype: int64

Total Number of Null Values in Dataset: PassengerId      0

Survived         0

Pclass           0

Name             0

Sex              0

Age             86

SibSp            0

Parch            0

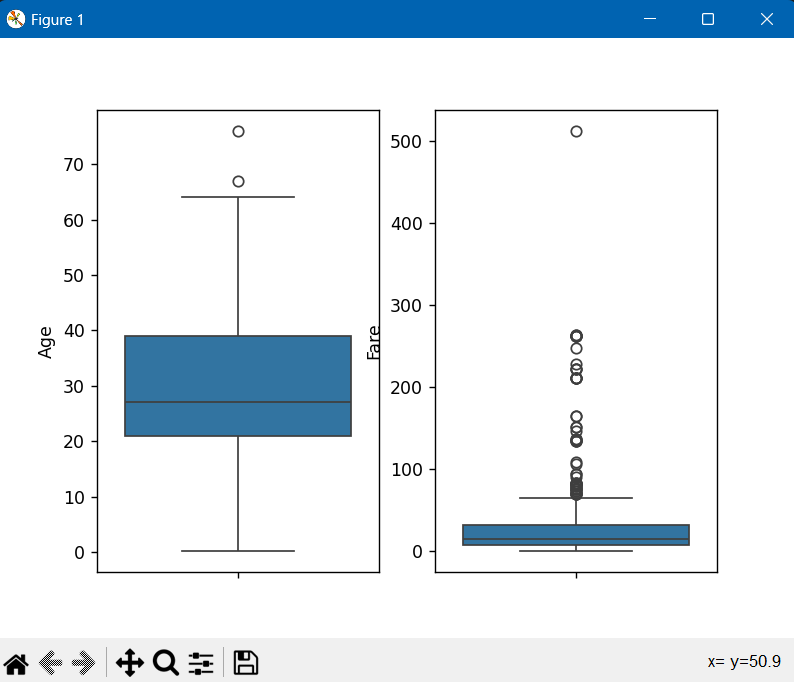
Ticket           0

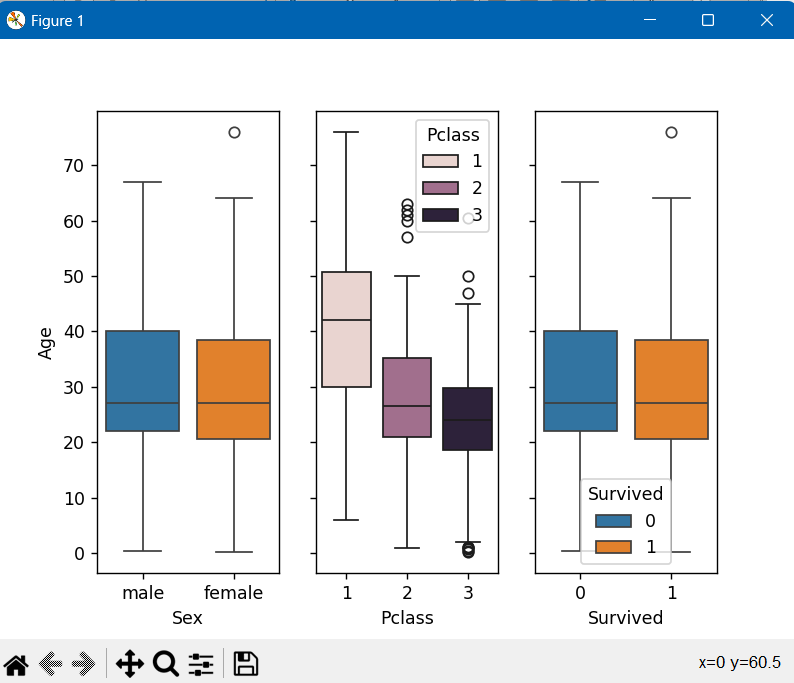
Fare             1

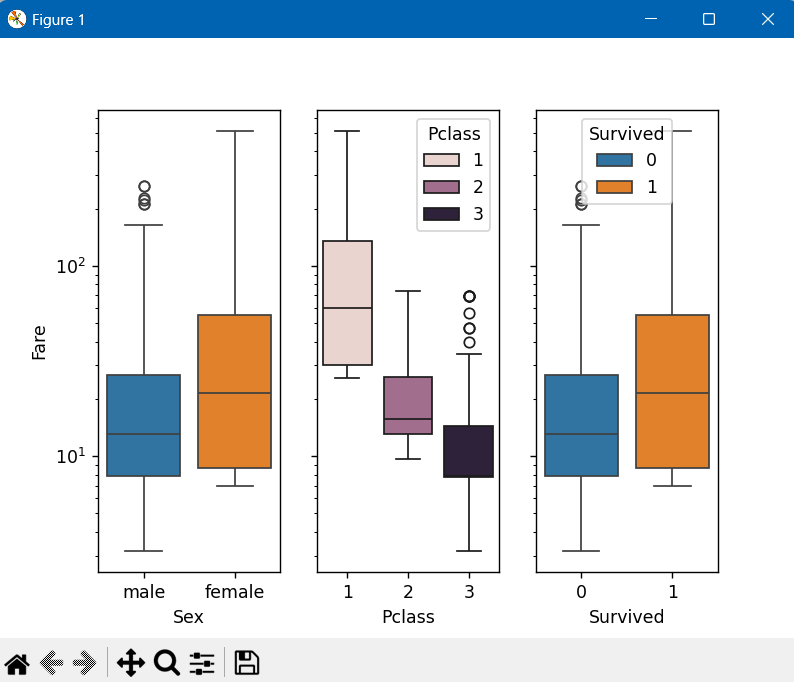
Cabin          327

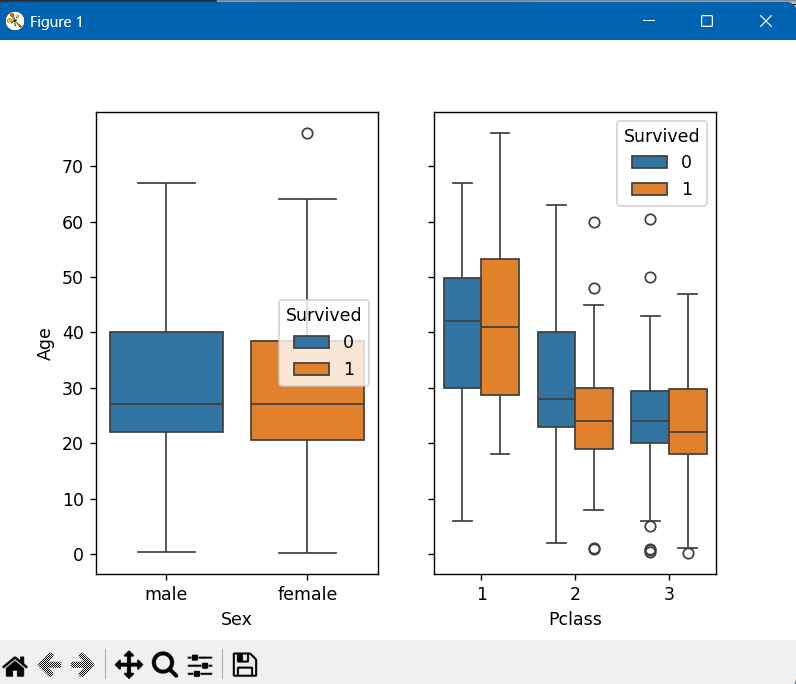
Embarked         0

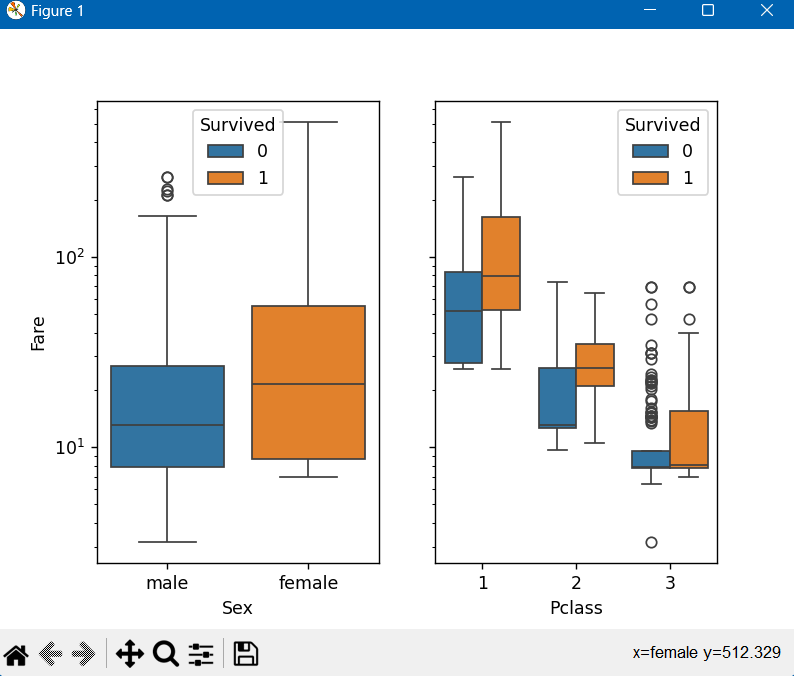
dtype: int64

****

****

****

****

****