Name: Shivam Indrabhan Borse

Roll No: 19

Subject: DATA SCIENCE (LAB)

Branch: AI & DS

Practical No. 01

Problem Statement :

Access an open source dataset
"Titanic".
Apply pre-processing techniques on the raw dataset

Program:

```
import numpy as np
import pandas as pn
import seaborn as sns

df = pn.read_csv ("gender_submission (2).csv")

df.info() df.head(9)
,
df.tail(1)
```

```
In [9]: df.head(9) , df.tail(1)

Out[9]: (   PassengerId  Survived
         0          892         0
         1          893         1
         2          894         0
         3          895         0
         4          896         1
         5          897         0
         6          898         1
         7          899         0
         8          900         1,
             PassengerId  Survived
         417         1309         0)
```

```
df.describe() , df.sample()
```

```
In [10]: df.describe() , df.sample()

Out[10]: (       PassengerId    Survived
          count   418.000000  418.000000
          mean   1100.500000    0.363636
          std     120.810458    0.481622
          min     892.000000    0.000000
          25%     996.250000    0.000000
          50%    1100.500000    0.000000
          75%    1204.750000    1.000000
          max    1309.000000    1.000000,
               PassengerId  Survived
          256         1148         0)
```

df.isnull().sum()

import seaborn as sns import
matplotlib.pyplot as pl

dataset = sns.load_dataset('titanic')

```
dataset
```

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	no	True
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes	True
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	no	False
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes	True
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	no	True

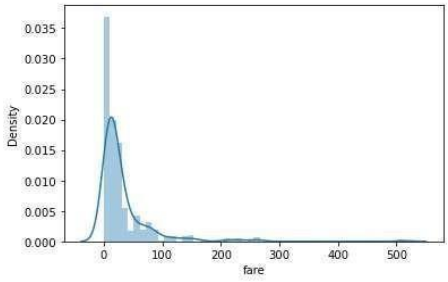
891 rows × 15 columns

sns.distplot(dataset['fare'])

```
sns.distplot(dataset['fare'])
```

```
C:\Users\student\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

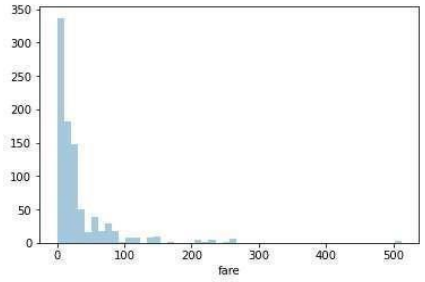
```
<AxesSubplot:xlabel='fare', ylabel='Density'>
```




```
sns.distplot(dataset['fare'], kde = False)
```

```
C:\Users\student\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
<AxesSubplot:xlabel='fare'>
```

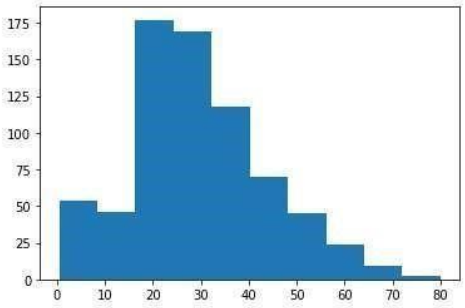


plt.hist(dataset['age'])
plt.show

```
import matplotlib.pyplot as plt
```

```
plt.hist(dataset['age'])
plt.show
```

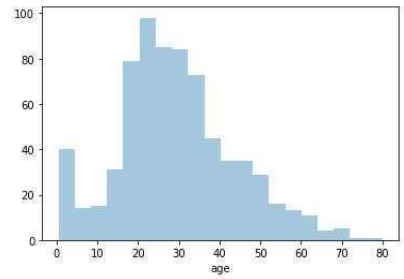
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



sns.distplot(dataset['age'], kde = False)

```
sns.distplot(dataset['age'], kde = False)
```

```
<AxesSubplot:xlabel='age'>
```

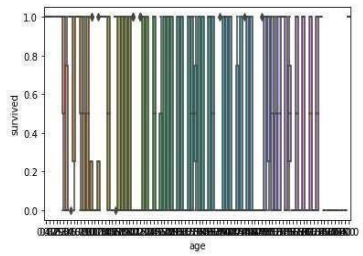


sns.barplot(dataset['fare'],)
sns.boxplot(dataset['age'],dataset['survived'])

```
sns.boxplot(dataset['age'],dataset['survived'])
```

```
C:\Users\student\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as key
word args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without
an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

```
<AxesSubplot:xlabel='age', ylabel='survived'>
```

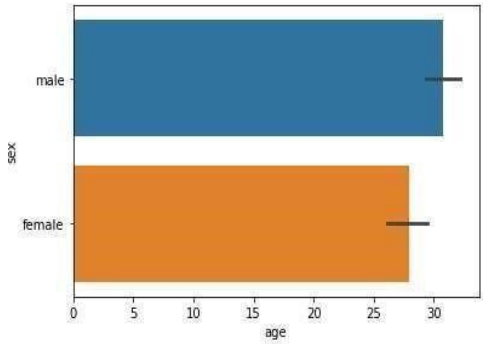


sns.barplot(dataset['age'],dataset['survived'])

sns.barplot(x='age', y='sex', data=dataset)

```
sns.barplot(x='age', y='sex', data=dataset)
```

```
<AxesSubplot:xlabel='age', ylabel='sex'>
```



sns.boxplot(dataset['sex'],dataset['survived'])

```
sns.boxplot(dataset['sex'],dataset['survived'])
```

```
C:\Users\student\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as key
word args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without
an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

```
<AxesSubplot:xlabel='sex', ylabel='survived'>
```