Worksheet 1

Student Name: Dhanuj Tomar UID:20BCS1201

Branch:- CSE Section/Group:701 B

Subject Name:-Machine Learning Lab

Aim/Overview of the practical:-

Implement Exploratory Data Analysis on any data set.

Result/Output/Writing Summary:-

1:-

```
[1] ✓ 0.3s
import matplotlib.pyplot as np
Python

[2] ✓ 0.2s
import pandas as pd
Python

[3] ✓ 0.2s
beat = pd.read_csv('/home/beat_box/Documents/--code/ML/805.csv')
Python

[4] ✓ 0.4s
beat.info()
Python

... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 73 entries, 0 to 72
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   ID        73 non-null     int64
 1   UID       73 non-null     object
 2   Name      73 non-null     object
 3   Section   73 non-null     object
 4   Groups    73 non-null     object
 5   Age       73 non-null     int64
 6   Email Id  73 non-null     object
dtypes: int64(2), object(5)
memory usage: 4.1+ KB
```

import pandas as pd

import :-it is used to import library in python

Pandas :- it is a data analysis library.

As pd :- Another name for referenceing the pandas as pd.

Matplotlib:-it is a library use to plot different type of graps.

Beat:- it is a variable that store the csv file data

pd.read_csv() :- it is a function use to read the file.

Info():-.it provides the essential details about your dataset, such as the number of rows and columns, non-null values, what type of data is in each column, and how much memory your DataFrame is using.

2:-

```
[5]  beat.describe()
     ✓ 0.5s
```

	ID	Age
count	73.000000	73.000000
mean	37.000000	21.739726
std	21.217131	2.198657
min	1.000000	17.000000
25%	19.000000	20.000000
50%	37.000000	22.000000
75%	55.000000	23.000000
max	73.000000	26.000000

```
[6]  beat.head()
     ✓ 0.8s
```

	ID	UID	Name	Section	Groups	Age	Email Id
0	1	21BCS8059	AKASH KUMAR PRASAD	20BCS_MM-805	A	20	21BCS8059@cuchd.in
1	2	21BCS8027	SAURAV KUMAR	20BCS_MM-805	A	21	21BCS8027@cuchd.in
2	3	21BCS8003	DEEPAK KUMAR	20BCS_MM-805	A	22	21BCS8003@cuchd.in
3	4	20BCS4527	Anubhav Tyagi	20BCS_MM-805	A	23	20BCS4527@cuchd.in
4	5	20BCS4802	KUSHAL AGARWAL	20BCS_MM-805	A	20	20BCS4802@cuchd.in

describe():-we can get a summary of the distribution of continuous variables:

head():-It is used to show the title of dataframe

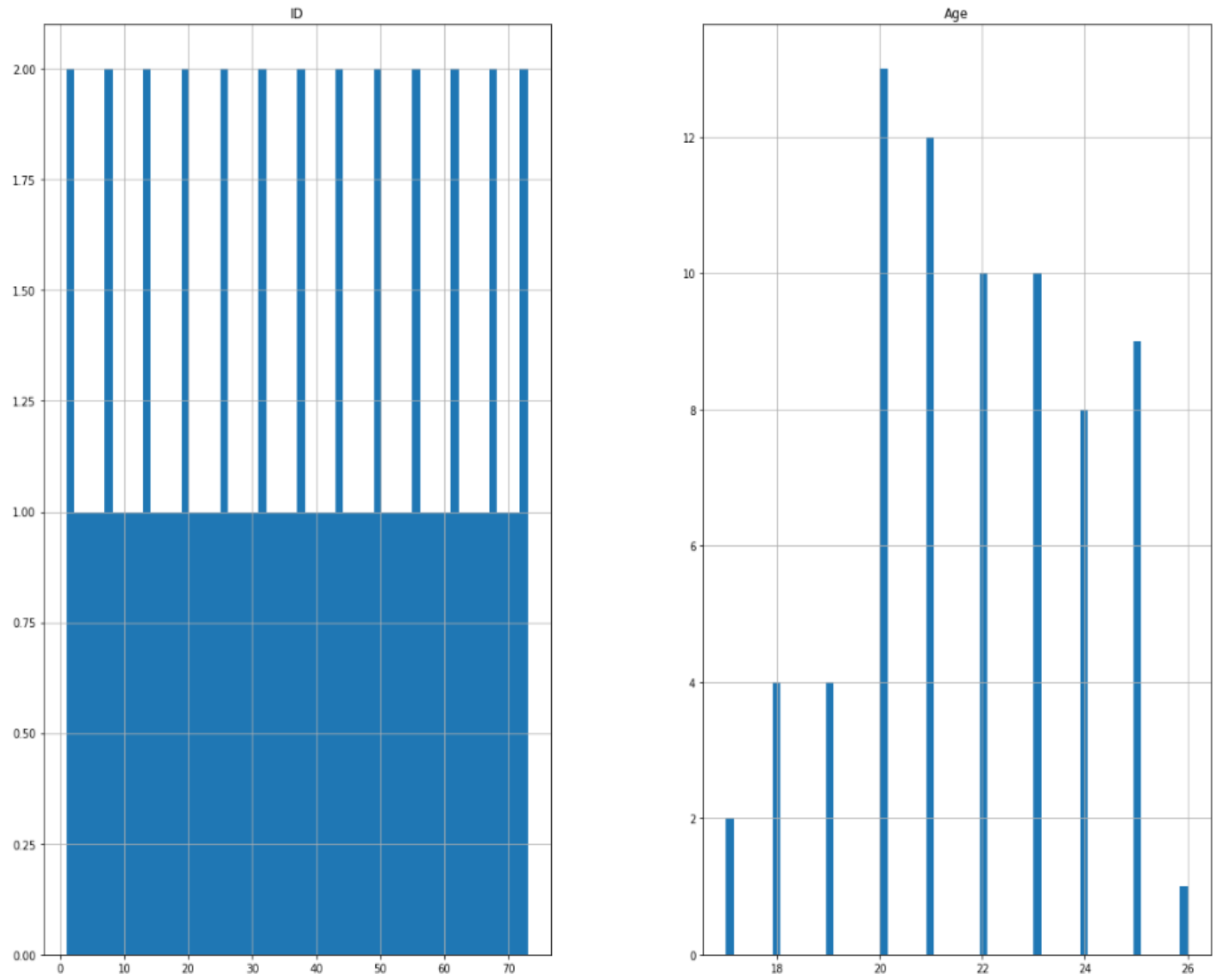
3:-

```
beat.hist(bins=60,figsize=(20,15))
```

[27] ✓ 0.4s

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f9b1ae1d5b0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f9b1ad73250>]],
      dtype=object)
```

h



hist():- it is used to draw the histogram of all the numerical value present in dataset.

4:-

```
[8]  beat.iloc[3]
     ✓ 0.7s
```

```
ID                          4
UID                 20BCS4527
Name              Anubhav Tyagi
Section           20BCS_MM-805
Groups                       A
Age                         23
Email Id    20BCS4527@cuchd.in
Name: 3, dtype: object
```

```
[9]  beat.iloc[0:4]
     ✓ 0.1s
```

	ID	UID	Name	Section	Groups	Age	Email Id
0	1	21BCS8059	AKASH KUMAR PRASAD	20BCS_MM-805	A	20	21BCS8059@cuchd.in
1	2	21BCS8027	SAURAV KUMAR	20BCS_MM-805	A	21	21BCS8027@cuchd.in
2	3	21BCS8003	DEEPAK KUMAR	20BCS_MM-805	A	22	21BCS8003@cuchd.in
3	4	20BCS4527	Anubhav Tyagi	20BCS_MM-805	A	23	20BCS4527@cuchd.in

iloc[]:-it is used to acess data frame with index value (loc - locates by name)

5:-

```
beat[beat.Age==23]
```
[28] ✓ 0.7s

	ID	UID	Name	Section	Groups	Age	Email Id
3	4	20BCS4527	Anubhav Tyagi	20BCS_MM-805	A	23	20BCS4527@cuchd.in
9	10	21BCS8023	IQRAR AHMAD	20BCS_MM-805	A	23	21BCS8023@cuchd.in
15	16	20BCS9355	DEVADEEP KUMAR	20BCS_MM-805	A	23	20BCS9355@cuchd.in
21	22	20BCS8265	Pratyaksh Verma	20BCS_MM-805	A	23	20BCS8265@cuchd.in
27	28	20BCS9621	JUVERIYAH	20BCS_MM-805	A	23	20BCS9621@cuchd.in
34	35	20BCS9500	SANJANA DEVI	20BCS_MM-805	A	23	20BCS9500@cuchd.in
43	44	20BCS3364	PRINCE KUMAR	20BCS_MM-805	B	23	20BCS3364@cuchd.in
55	56	20BCS3240	Aswani Kumar	20BCS_MM-805	B	23	20BCS3240@cuchd.in
61	62	20BCS1725	VISHAL CHAUHAN	20BCS_MM-805	B	23	20BCS1725@cuchd.in
69	70	20BCS7113	KARANDEEP SINGH	20BCS_MM-805	B	23	20BCS7113@cuchd.in

```
beat[(beat.Age==18)&(beat.ID<50)]
```
[11] ✓ 0.3s

	ID	UID	Name	Section	Groups	Age	Email Id
38	39	20BCS5566	DIVIT MITTAL	20BCS_MM-805	B	18	20BCS5566@cuchd.in
47	48	20BCS9970	ANUBHAV	20BCS_MM-805	B	18	20BCS9970@cuchd.in

6:-

```
beat["ID"].unique()
```
[12] ✓ 0.8s

```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
       35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51,
       52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68,
       69, 70, 71, 72, 73])
```

```
print(beat["ID"].mean())
```
[13] ✓ 0.1s

```
37.0
```

```
print(beat["ID"].median())
```
[14] ✓ 0.1s

```
37.0
```

```
print((beat["ID"]=="Age").sum())
```
[15] ✓ 0.4s

```
0
```

7:-

```
[16]  beat["ID"].fillna(994)
      ✓ 0.8s
...   0        1
      1        2
      2        3
      3        4
      4        5
              ..
      68      69
      69      70
      70      71
      71      72
      72      73
      Name: ID, Length: 73, dtype: int64
```

fillna():-It is used to replace null value with some another value

8:-

```
17]   beat.isnull().sum()
      ✓ 0.1s
..    ID          0
      UID         0
      Name        0
      Section     0
      Groups      0
      Age         0
      Email Id    0
      dtype: int64
```

isnull():-it is used to check null value in data frames it return true if null value found and false for not null

9:-

```
beat.groupby("ID")['Age'].mean()
```

[18] ✓ 0.4s

```
ID
1     20.0
2     21.0
3     22.0
4     23.0
5     20.0
      ...
69    22.0
70    23.0
71    24.0
72    25.0
73    26.0
Name: Age, Length: 73, dtype: float64
```

10:-

pivot_table():-Create a spreadsheet-style pivot table as a DataFrame

```
beat.pivot_table(index="ID",columns="Age",values="UID",aggfunc="sum")
```

[19] ✓ 0.1s

Age	17	18	19	20	21	22	23	24	25	26
ID										
1	NaN	NaN	NaN	21BCS8059	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	21BCS8027	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	21BCS8003	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	20BCS4527	NaN	NaN	NaN
5	NaN	NaN	NaN	20BCS4802	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
69	NaN	NaN	NaN	NaN	NaN	20BCS5864	NaN	NaN	NaN	NaN
70	NaN	NaN	NaN	NaN	NaN	NaN	20BCS7113	NaN	NaN	NaN
71	NaN	NaN	NaN	NaN	NaN	NaN	NaN	20BCS5924	NaN	NaN
72	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	20BCS5800	NaN
73	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	20BCS5748

73 rows × 10 columns

Evaluation Grid :

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.	Student Performance (Conduct of experiment) objectives/Outcomes.		12
2.	Viva Voce		10
3.	Submission of Work Sheet (Record)		8
	Total		30