Assignment 1

Charvi Upreti

charvi.upreti2021@vitstudent.ac.in (mailto:charvi.upreti2021@vitstudent.ac.in)

21BCE1440

AI ML Assignment 1	
Task - 1	Create a pandas dataframe (DataFrame name as 'df') with numpy random values (4 features and 4 observation)
Task - 2	Rename the task - 1 'df' dataframe column names to 'Random value 1', 'Random value 2', 'Random value 3' & 'Random value 4'
Task - 3	Find the descriptive statistics of the 'df' dataframe.
Task - 4	Check for the null values in 'df' and find the data type of the columns.
Task - 5	Display the 'Random value 2' & 'Random value 3' columns with location method and index location method.

In [1]:
```
import numpy as np
import pandas as pd
```

Task 1: Create pandas DataFrame named df with numpy random values (4 features and 4 observations)

In [2]:
```
dict1={
    'A':np.random.randint(10,size=4),
    'B':np.random.randint(10,size=4),
    'C':np.random.randint(10,size=4),
    'D':np.random.randint(10,size=4)
}
df=pd.DataFrame(dict1)
```

In [3]:
```
df.shape
```

Out[3]: (4, 4)

In [4]:
```
df
```

Out[4]:

	A	B	C	D
0	7	0	3	3
1	2	5	1	6
2	6	0	2	6
3	5	8	9	0

Task 2: Rename the task - 1 'df' dataframe column names to 'Random value 1', ' Random value 2', 'Random value 3' & 'Random value 4'

In [5]:
```
df.columns=['Random value 1', 'Random value 2', 'Random value 3','Random value 4
```

In [6]:
```
1 df
```

Out[6]:

	Random value 1	Random value 2	Random value 3	Random value 4
0	7	0	3	3
1	2	5	1	6
2	6	0	2	6
3	5	8	9	0

Task 3: Find the descriptive statistics of 'df' dataframe.

In [7]:
```
1 print(df.describe(include='all'))
2 ## or
3 print("\nOR\n")
4 print(df.describe())
```

```
       Random value 1  Random value 2  Random value 3  Random value 4
count        4.000000        4.000000        4.000000        4.000000
mean         5.000000        3.250000        3.750000        3.750000
std          2.160247        3.947573        3.593976        2.872281
min          2.000000        0.000000        1.000000        0.000000
25%          4.250000        0.000000        1.750000        2.250000
50%          5.500000        2.500000        2.500000        4.500000
75%          6.250000        5.750000        4.500000        6.000000
max          7.000000        8.000000        9.000000        6.000000

OR

       Random value 1  Random value 2  Random value 3  Random value 4
count        4.000000        4.000000        4.000000        4.000000
mean         5.000000        3.250000        3.750000        3.750000
std          2.160247        3.947573        3.593976        2.872281
min          2.000000        0.000000        1.000000        0.000000
25%          4.250000        0.000000        1.750000        2.250000
50%          5.500000        2.500000        2.500000        4.500000
75%          6.250000        5.750000        4.500000        6.000000
max          7.000000        8.000000        9.000000        6.000000
```

Task 4: Check for null values in 'df' and find the data types of the columns.

In [8]:
```
1 df.isnull()
```

Out[8]:

	Random value 1	Random value 2	Random value 3	Random value 4
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False

In [9]:
```
df.dtypes
```

Out[9]:
```
Random value 1    int32
Random value 2    int32
Random value 3    int32
Random value 4    int32
dtype: object
```

In [10]:
```
### deliberately putting null value to check
df.loc[0, 'Random value 1'] = np.nan
print(df)
df.isnull()
```

```
   Random value 1  Random value 2  Random value 3  Random value 4
0             NaN               0               3               3
1             2.0               5               1               6
2             6.0               0               2               6
3             5.0               8               9               0
```

Out[10]:

	Random value 1	Random value 2	Random value 3	Random value 4
0	True	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False

Task 5: Display the 'Random value 2' and 'Random value 3' columns with location method and index location method.

In [11]:
```
print(df.loc[:,['Random value 2','Random value 3']])
```

```
   Random value 2  Random value 3
0               0               3
1               5               1
2               0               2
3               8               9
```

In [12]:
```
print(df.iloc[:,[1,2]])
```

```
   Random value 2  Random value 3
0               0               3
1               5               1
2               0               2
3               8               9
```

Other questions given in class:-

Task - 1 Create a pandas dataframe (DataFrame name as 'df') (10 observation and 5 features)

Task- 2 Check the info of 'df'

Task 3- Check the descriptive statistics of 'df'

Task 4- check the 4th index observation with 'loc' slicing operator.

Task 5 - Check the null values in your 'df'

1: Create a pandas dataframe (DataFrame name as 'df') (10 observations and 5 features)

Note:-

Features = Columns

Observation = rows

In [13]:
```
dict1={
    'A':np.random.randint(10,size=10),
    'B':np.random.randint(10,size=10),
    'C':np.random.randint(10,size=10),
    'D':np.random.randint(10,size=10),
    'E':np.random.randint(10,size=10)
}
df=pd.DataFrame(dict1)
```

In [14]:
```
df
```

Out[14]:

	A	B	C	D	E
0	8	6	2	7	6
1	7	2	7	8	9
2	3	2	9	0	7
3	3	9	5	4	3
4	3	7	8	1	9
5	4	0	3	8	7
6	4	7	4	3	0
7	8	0	7	7	1
8	9	1	4	3	9
9	4	3	7	3	3

2: Checking info of 'df'

In [15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   A       10 non-null     int32
 1   B       10 non-null     int32
 2   C       10 non-null     int32
 3   D       10 non-null     int32
 4   E       10 non-null     int32
dtypes: int32(5)
memory usage: 328.0 bytes
```

3: Check the descriptive statistics of 'df'

In [16]:
```
1 df.describe()
```

Out[16]:

	A	B	C	D	E
count	10.000000	10.000000	10.000000	10.00000	10.000000
mean	5.300000	3.700000	5.600000	4.40000	5.400000
std	2.406011	3.267687	2.319004	2.91357	3.405877
min	3.000000	0.000000	2.000000	0.00000	0.000000
25%	3.250000	1.250000	4.000000	3.00000	3.000000
50%	4.000000	2.500000	6.000000	3.50000	6.500000
75%	7.750000	6.750000	7.000000	7.00000	8.500000
max	9.000000	9.000000	9.000000	8.00000	9.000000

4: Check the 4th index observation with 'loc' slicing operator

In [17]:
```
1 df.loc[4]
```

Out[17]:
```
A    3
B    7
C    8
D    1
E    9
Name: 4, dtype: int32
```

5: Check all the null values in your 'df'

In [18]:
```
1 df.isnull().any()
```

Out[18]:
```
A    False
B    False
C    False
D    False
E    False
dtype: bool
```