

21Bec10044
Shreya Gupta
August 29, 2023

ASSIGNMENT - 1

Task-1: Create a pandas dataframe (Dataframe name as 'df' with numpy random values (4 features and 4 observation)

Importing the dependencies:

```
[2]: import numpy as np
import pandas as pd
```

```
[7]: #Given: 4 features and 4 observations
#random values
r_data = np.random.rand(4, 4)

#creating dataframe df:
df = pd.DataFrame(r_data, columns=['a', 'b', 'c', 'd'])

print(df)
```

	a	b	c	d
0	0.304242	0.524756	0.431945	0.291229
1	0.611853	0.139494	0.292145	0.366362
2	0.456070	0.785176	0.199674	0.514234
3	0.592415	0.046450	0.607545	0.170524

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

```
[8]: #Renaming columns:
new_columns = { 'a': 'Random value 1',
                'b': 'Random value 2',
                'c': 'Random value 3',
                'd': 'Random value 4' }

df = df.rename(columns=new_columns)

print(df)
```

	Random value 1	Random value 2	Random value 3	Random value 4
0	0.304242	0.524756	0.431945	0.291229
1	0.611853	0.139494	0.292145	0.366362

2	0.456070	0.785176	0.199674	0.514234
3	0.592415	0.046450	0.607545	0.170524

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

```
[9]: #Describe function:
df.describe()
```

```
[9]:
```

	Random value 1	Random value 2	Random value 3	Random value 4
count	4.000000	4.000000	4.000000	4.000000
mean	0.491145	0.373969	0.382827	0.335587
std	0.142582	0.343548	0.177650	0.143846
min	0.304242	0.046450	0.199674	0.170524
25%	0.418113	0.116233	0.269027	0.261053
50%	0.524242	0.332125	0.362045	0.328795
75%	0.597274	0.589861	0.475845	0.403330
max	0.611853	0.785176	0.607545	0.514234

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

```
[12]: null_values = df.isnull()
print("Null values: \n", null_values)

df.info()
```

Null values:

	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

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 4 entries, 0 to 3

Data columns (total 4 columns):

#	Column	Non-Null Count	Dtype
0	Random value 1	4 non-null	float64
1	Random value 2	4 non-null	float64
2	Random value 3	4 non-null	float64
3	Random value 4	4 non-null	float64

dtypes: float64(4)

memory usage: 256.0 bytes

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

```
[13]: columns_loc = df.loc[:, ['Random value 2', 'Random value 3']]
print("Using location method:\n", columns_loc)
```

```
# Display using index-based location (.iloc)
columns_iloc = df.iloc[:, [1, 2]]
print("\nUsing index location method:\n", columns_iloc)
```

Using location method:

	Random value 2	Random value 3
0	0.524756	0.431945
1	0.139494	0.292145
2	0.785176	0.199674
3	0.046450	0.607545

Using index location method:

	Random value 2	Random value 3
0	0.524756	0.431945
1	0.139494	0.292145
2	0.785176	0.199674
3	0.046450	0.607545