Assignment 2

Smart Internz AI

Reg No : 20BCB0134

Name : Ritika Marlapalli

Email: ritika.marlapalli2020@vitstudent.ac.in

College: VIT Vellore

Campus : Vellore

Create a pandas dataframe (DataFrame name as 'df') with numpy random values (4 features and 4 observation)

```python
import numpy as np
import pandas as pd
# Set the random seed for reproducibility

np.random.seed(25)

# Create random values
data = np.random.rand(4, 4)

# Create the DataFrame
df = pd.DataFrame(data, columns=['Feature1', 'Feature2', 'Feature3', 'Feature4'])
print(df)
```

```
   Feature1  Feature2  Feature3  Feature4
0  0.870124  0.582277  0.278839  0.185911
1  0.411100  0.117376  0.684969  0.437611
2  0.556229  0.367080  0.402366  0.113041
3  0.447031  0.585445  0.161985  0.520719
```

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

```python
# Rename the column names
df.rename(columns={'Feature1': 'Random value 1', 'Feature2': 'Random value 2','Feature3': 'Random value 3', 'Feature4': 'Random value 4'},inplace=True)

print(df)
```

```
   Random value 1  Random value 2  Random value 3  Random value 4
0        0.870124        0.582277        0.278839        0.185911
1        0.411100        0.117376        0.684969        0.437611
2        0.556229        0.367080        0.402366        0.113041
3        0.447031        0.585445        0.161985        0.520719
```

Find the descriptive statistics of the 'df' dataframe.

```
# Display descriptive statistics
statistics = df.describe()

# Print the statistics
print(statistics)
```

```
       Random value 1  Random value 2  Random value 3  Random value 4
count        4.000000        4.000000        4.000000        4.000000
mean         0.571121        0.413044        0.382040        0.314320
std          0.208670        0.222032        0.224539        0.195621
min          0.411100        0.117376        0.161985        0.113041
25%          0.438048        0.304654        0.249625        0.167694
50%          0.501630        0.474679        0.340602        0.311761
75%          0.634703        0.583069        0.473016        0.458388
max          0.870124        0.585445        0.684969        0.520719
```

Check for the null values in 'df' and find the data type of the columns.

```
# Check for null values
null_values = df.isnull().sum()
# Print null values
print("Null Values:\n", null_values)
# Find data types of columns
data_types = df.dtypes
# Print data types
print("\nData Types:\n", data_types)
```

```
Null Values:
 Random value 1    0
Random value 2    0
Random value 3    0
Random value 4    0
dtype: int64

Data Types:
 Random value 1    float64
Random value 2    float64
Random value 3    float64
Random value 4    float64
dtype: object
```

Display the 'Random value 2' & 'Random value 3' columns with location method and index location method.

```
# Display columns using location method
location_method_cols = df.loc[:, ['Random value 2', 'Random value 3']]
print("Columns using location method:\n", location_method_cols)
# Display columns using index location method
index_location_method_cols = df.iloc[:, [1, 2]]
print("\nColumns using index location method:\n", index_location_method_cols)
```

```
Columns using location method:
    Random value 2  Random value 3
0        0.582277        0.278839
1        0.117376        0.684969
2        0.367080        0.402366
3        0.585445        0.161985

Columns using index location method:
    Random value 2  Random value 3
0        0.582277        0.278839
1        0.117376        0.684969
2        0.367080        0.402366
3        0.585445        0.161985
```