Name - Angad Singh. Registration No. - 21BPS1420

Assignment 1:

```python
import pandas as pd
import numpy as np

size = 10
null_prob = 0.4

# Task 1: Create a pandas DataFrame with 10 observations and 5 features
print("Task 1 - DataFrame Creation:\n")
data = {
    'Feature1': np.random.choice([np.random.rand(), np.nan], size=size, p=[1 - null_prob, null_prob]),
    'Feature2': np.random.choice([np.random.rand(), np.nan], size=size, p=[1 - null_prob, null_prob]),
    'Feature3': np.random.choice([np.random.rand(), np.nan], size=size, p=[1 - null_prob, null_prob]),
    'Feature4': np.random.choice([np.random.rand(), np.nan], size=size, p=[1 - null_prob, null_prob]),
    'Feature5': np.random.choice([np.random.rand(), np.nan], size=size, p=[1 - null_prob, null_prob])
}
df = pd.DataFrame(data)
df
```

Task 1 - DataFrame Creation:

	Feature1	Feature2	Feature3	Feature4	Feature5
0	NaN	0.063519	0.593685	0.908126	NaN
1	NaN	0.063519	0.593685	0.908126	0.546318
2	0.39432	0.063519	0.593685	NaN	0.546318
3	0.39432	NaN	NaN	NaN	0.546318
4	0.39432	0.063519	0.593685	NaN	NaN
5	NaN	0.063519	0.593685	0.908126	0.546318
6	0.39432	0.063519	NaN	0.908126	NaN
7	0.39432	0.063519	0.593685	0.908126	NaN
8	0.39432	0.063519	NaN	0.908126	0.546318
9	0.39432	0.063519	NaN	NaN	0.546318

```python
# Task 2: Check the info of 'df'
print("Task 2 - DataFrame Info:\n")
df.info()
```

Task 2 - DataFrame Info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Feature1  7 non-null      float64
 1   Feature2  9 non-null      float64
 2   Feature3  6 non-null      float64
 3   Feature4  6 non-null      float64
 4   Feature5  6 non-null      float64
dtypes: float64(5)
memory usage: 528.0 bytes
```

```python
# Task 3: Check the descriptive statistics of 'df'
print("Task 3 - Descriptive Statistics:\n")
print(df.describe())
```

Task 3 - Descriptive Statistics:

	Feature1	Feature2	Feature3	Feature4	Feature5
count	7.000000e+00	9.000000	6.000000	6.000000	6.000000
mean	3.943203e-01	0.063519	0.593685	0.908126	0.546318
std	5.995890e-17	0.000000	0.000000	0.000000	0.000000
min	3.943203e-01	0.063519	0.593685	0.908126	0.546318
25%	3.943203e-01	0.063519	0.593685	0.908126	0.546318
50%	3.943203e-01	0.063519	0.593685	0.908126	0.546318
75%	3.943203e-01	0.063519	0.593685	0.908126	0.546318
max	3.943203e-01	0.063519	0.593685	0.908126	0.546318

```python
# Task 4: Check the 4th index observation with 'loc' slicing operator.
print("Task 4 - 4th Index Observation:\n")
```

```
observation_4 = df.loc[4]
print(observation_4)
```

    Task 4 - 4th Index Observation:

    Feature1   0.394320
    Feature2   0.063519
    Feature3   0.593685
    Feature4   NaN
    Feature5   NaN
    Name: 4, dtype: float64

```
# Task 5: Check the null values in your 'df'
print("Task 5 - Null Values Check:\n")
print(df.isnull().sum())
```

    Task 5 - Null Values Check:

    Feature1   3
    Feature2   1
    Feature3   4
    Feature4   4
    Feature5   4
    dtype: int64

✓ 0s   completed at 10:38 PM   ● ✕