Hemanth Venkata Addepalli

hemanth.addepalli2021@vitstudent.ac.in

+91 8919950467

6pm-8pm

TASK1:Creating DataFrame

Creating DataFrame
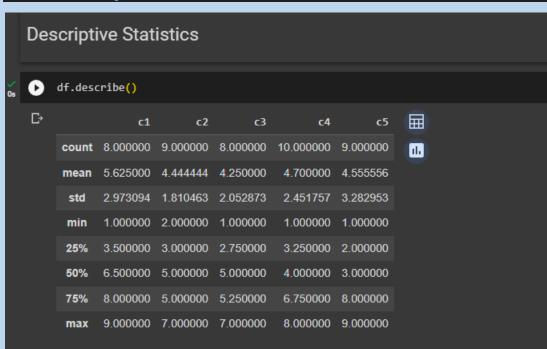
```python
[65] import pandas as pd
     import numpy as np
     df=pd.DataFrame([[7,3,2,8,np.nan],[1,2,np.nan,4,5],[4,5,np.nan,7,8],[6,3,1,4,2],[9,7,5,4,3],[2,5,7,6,9],[8,5,3,2,2],[np.nan,7,5,3,1],[8,np.nan,6,1,2],[np.nan,3,5,8,9]])
```

```python
[66] df.shape

     (10, 5)
```

```python
[67] df.columns=['c1','c2','c3','c4','c5']
     df.index=['obs1','obs2','obs3','obs4','obs5','obs6','obs7','obs8','obs9','obs10']
```

```python
[68] df
```

	c1	c2	c3	c4	c5
obs1	7.0	3.0	2.0	8	NaN
obs2	1.0	2.0	NaN	4	5.0
obs3	4.0	5.0	NaN	7	8.0
obs4	6.0	3.0	1.0	4	2.0
obs5	9.0	7.0	5.0	4	3.0
obs6	2.0	5.0	7.0	6	9.0
obs7	8.0	5.0	3.0	2	2.0
obs8	NaN	7.0	5.0	3	1.0
obs9	8.0	NaN	6.0	1	2.0
obs10	NaN	3.0	5.0	8	9.0

TASK 2:Checking the Info of dataframe

Checking the Info of dataframe

```python
[69] df.info()

     <class 'pandas.core.frame.DataFrame'>
     Index: 10 entries, obs1 to obs10
     Data columns (total 5 columns):
      #   Column  Non-Null Count  Dtype
     ---  ------  --------------  -----
      0   c1      8 non-null      float64
      1   c2      9 non-null      float64
      2   c3      8 non-null      float64
      3   c4      10 non-null     int64
      4   c5      9 non-null      float64
     dtypes: float64(4), int64(1)
     memory usage: 480.0+ bytes
```

Descriptive Statistics

```
df.describe()
```

	c1	c2	c3	c4	c5
count	8.000000	9.000000	8.000000	10.000000	9.000000
mean	5.625000	4.444444	4.250000	4.700000	4.555556
std	2.973094	1.810463	2.052873	2.451757	3.282953
min	1.000000	2.000000	1.000000	1.000000	1.000000
25%	3.500000	3.000000	2.750000	3.250000	2.000000
50%	6.500000	5.000000	5.000000	4.000000	3.000000
75%	8.000000	5.000000	5.250000	6.750000	8.000000
max	9.000000	7.000000	7.000000	8.000000	9.000000

4th Index

```
df.loc['obs4']
```

```
c1    6.0
c2    3.0
c3    1.0
c4    4.0
c5    2.0
Name: obs4, dtype: float64
```

Checking Null Values

```
df.isnull()
```

	c1	c2	c3	c4	c5
obs1	False	False	False	False	True
obs2	False	False	True	False	False
obs3	False	False	True	False	False
obs4	False	False	False	False	False
obs5	False	False	False	False	False
obs6	False	False	False	False	False
obs7	False	False	False	False	False
obs8	True	False	False	False	False
obs9	False	True	False	False	False
obs10	True	False	False	False	False

```
[73] df.isnull().any()
```

```
c1     True
c2     True
c3     True
c4     False
c5     True
dtype: bool
```

```
[74] df.isnull().sum()
```

```
c1    2
c2    1
c3    2
c4    0
c5    1
dtype: int64
```