AI&ML ASSIGNMENT-1

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

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

In [75]:
```
a=np.random.randint(1,25,10)
b=np.random.randint(25,50,10)
c=np.random.randint(50,75,10)
d=np.random.randint(75,100,10)
e=np.random.randint(100,125,10)
```

In [76]:
```
print("1st numpy feature with random values --> ", a)
print("2nd numpy feature with random values --> ",b)
print("3rd numpy feature with random values --> ",c)
print("4th numpy feature with random values --> ",d)
print("5th numpy feature with random values --> ",e)
```

```
1st numpy feature with random values -->  [16  3  8 22 12 10 21  8  2  7]
2nd numpy feature with random values -->  [40 29 37 48 39 27 42 28 46 36]
3rd numpy feature with random values -->  [71 57 67 61 50 68 56 74 51 53]
4th numpy feature with random values -->  [84 78 99 78 79 76 94 85 77 89]
5th numpy feature with random values -->  [115 107 101 120 106 122 105 121
101 104]
```

In [37]:
```
type(a)
```

Out[37]:
```
numpy.ndarray
```

In [38]:
```
x={"a":a,"b":b,"c":c,"d":d,"e":e}
x
```

Out[38]:
```
{'a': array([18, 22, 17, 16, 16, 12,  4, 21,  3, 24]),
 'b': array([42, 30, 37, 36, 41, 41, 26, 40, 49, 26]),
 'c': array([65, 74, 52, 68, 56, 54, 73, 72, 69, 66]),
 'd': array([81, 98, 97, 85, 78, 77, 77, 89, 96, 95]),
 'e': array([118, 108, 108, 119, 111, 109, 103, 110, 113, 110])}
```

In [39]:
```
df=pd.DataFrame(x)
```

In [40]:
```
df
```

Out[40]:

	a	b	c	d	e
0	18	42	65	81	118
1	22	30	74	98	108
2	17	37	52	97	108
3	16	36	68	85	119
4	16	41	56	78	111
5	12	41	54	77	109
6	4	26	73	77	103
7	21	40	72	89	110
8	3	49	69	96	113
9	24	26	66	95	110

In [63]:
```
df.head()
```

Out[63]:

	Random value 1	Random value 2	Random value 3	Random value 4	Random value 5
0	18	42	65	81	118
1	22	30	74	98	108
2	17	37	52	97	108
3	16	36	68	85	119
4	16	41	56	78	111

In [49]:
```
df.shape
```

Out[49]:
```
(10, 5)
```

Renaming the column names to 'Random value 1','Random value 2','Random value 3','Random value 4','Random value 5'

In [41]:
```
df.columns
```

Out[41]:
```
Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

In [42]:
```
df.columns=['Random value 1','Random value 2','Random value 3','Random valu
```

In [43]:
```
df.columns
```

Out[43]:
```
Index(['Random value 1', 'Random value 2', 'Random value 3', 'Random value
4',
       'Random value 5'],
      dtype='object')
```

In [44]:
```
df
```

Out[44]:

	Random value 1	Random value 2	Random value 3	Random value 4	Random value 5
0	18	42	65	81	118
1	22	30	74	98	108
2	17	37	52	97	108
3	16	36	68	85	119
4	16	41	56	78	111
5	12	41	54	77	109
6	4	26	73	77	103
7	21	40	72	89	110
8	3	49	69	96	113
9	24	26	66	95	110

Finding the descriptive statistics of our dataframe df

In [45]:
```
df.describe()
```

Out[45]:

	Random value 1	Random value 2	Random value 3	Random value 4	Random value 5
count	10.000000	10.000000	10.000000	10.000000	10.000000
mean	15.300000	36.800000	64.900000	87.300000	110.900000
std	7.103207	7.465476	8.102812	8.756585	4.771443
min	3.000000	26.000000	52.000000	77.000000	103.000000
25%	13.000000	31.500000	58.250000	78.750000	108.250000
50%	16.500000	38.500000	67.000000	87.000000	110.000000
75%	20.250000	41.000000	71.250000	95.750000	112.500000
max	24.000000	49.000000	74.000000	98.000000	119.000000

checking for null values and also finding the datatypes of columns

In [46]:
```
df.isnull().any()
```

Out[46]:
```
Random value 1    False
Random value 2    False
Random value 3    False
Random value 4    False
Random value 5    False
dtype: bool
```

In [47]:
```
df.isnull().sum()  #no null values in the dataframe
```

Out[47]:
```
Random value 1    0
Random value 2    0
Random value 3    0
Random value 4    0
Random value 5    0
dtype: int64
```

In [48]:
```
df.info() #here dtype shows the datatype of columns of our dataframe
```

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

In [52]:
```
df.dtypes
```

Out[52]:
```
Random value 1    int32
Random value 2    int32
Random value 3    int32
Random value 4    int32
Random value 5    int32
dtype: object
```

check the 4th index observation with 'loc' slicing operator

In [50]:
```
df.loc[4]
```

Out[50]:
```
Random value 1     16
Random value 2     41
Random value 3     56
Random value 4     78
Random value 5    111
Name: 4, dtype: int32
```

display random value2 and random value 3 columns with location method and index method

In [62]:
```
index_2=df.columns.get_loc("Random value 2")
index_3=df.columns.get_loc("Random value 3")
print(index_2,index_3,end=" ")
```

```
1 2
```

In [68]:
```
df.iloc[:,index_2:index_3+1] #iloc method
```

Out[68]:

	Random value 2	Random value 3
0	42	65
1	30	74
2	37	52
3	36	68
4	41	56
5	41	54
6	26	73
7	40	72
8	49	69
9	26	66

In [74]:
```
df[['Random value 2','Random value 3']] #normal method
```

Out[74]:

	Random value 2	Random value 3
0	42	65
1	30	74
2	37	52
3	36	68
4	41	56
5	41	54
6	26	73
7	40	72
8	49	69
9	26	66

In [85]:
```
df.loc[:, df.columns[index_2:index_3+1]] #loc method
```

Out[85]:

	Random value 2	Random value 3
0	42	65
1	30	74
2	37	52
3	36	68
4	41	56
5	41	54
6	26	73
7	40	72
8	49	69
9	26	66