Day 14

Series

Editor Window

```
import pandas as pd
s=pd.read_csv('random.csv',header=None,squeeze=True)
type(s)
st=pd.read_csv('Startups.csv',squeeze=True,usecols=['R&D Spend'])

st.index
st.values
st.size
st.ndim
st.is_unique
st.add_prefix("nielit")
st.sum()
st.max()
st.min()
st.mean()
st.idxmax()
st.head(10)
st.tail(10)
st.describe()
st1=pd.read_csv('D:\pyth\drinks.csv')
a=[12,22,27,30,33,36,35,34,32,27,20,11]
b=['jan','feb','mar','april','may','june','july','aug','sept','oct'
  ,'nov','dec']
s1=pd.Series(data=a,index=b)
s1.sort_values(ascending=False)
st.sort_values()
```

s1.sort_values(inplace=True)

s1=pd.Series(data=a,index=b)

s1.sort_index()

s1[4:]

s1[5:11]

s1[[2,10]]

o/p window

type(s)

Out[5]: pandas.core.series.Series

st=pd.read_csv('Startups.csv',squeeze=True)

st=pd.read_csv('Startups.csv',squeeze=True,usecols=['Rd Sepnd'])

st.index

Out[10]: RangeIndex(start=0, stop=50, step=1)

st.values

Out[11]:

array([165349.2 , 162597.7 , 153441.51, 144372.41, 142107.34, 131876.9 ,
134615.46, 130298.13, 120542.52, 123334.88, 101913.08, 100671.96,
93863.75, 91992.39, 119943.24, 114523.61, 78013.11, 94657.16,
91749.16, 86419.7 , 76253.86, 78389.47, 73994.56, 67532.53,
77044.01, 64664.71, 75328.87, 72107.6 , 66051.52, 65605.48,
61994.48, 61136.38, 63408.86, 55493.95, 46426.07, 46014.02,
28663.76, 44069.95, 20229.59, 38558.51, 28754.33, 27892.92,

23640.93, 15505.73, 22177.74, 1000.23, 1315.46, 0. ,

542.05, 0.])

st.size

Out[12]: 50

st.ndim

Out[13]: 1

st.is_unique

Out[14]: False

st.add_prefix("nielit")

Out[15]:

nielit0 165349.20

nielit1 162597.70

nielit2 153441.51

nielit3 144372.41

nielit4 142107.34

nielit5 131876.90

nielit6 134615.46

nielit7 130298.13

nielit8 120542.52

nielit9 123334.88

nielit10 101913.08

nielit11 100671.96

nielit12 93863.75

nielit13 91992.39

nielit14 119943.24

nielit15 114523.61

nielit16 78013.11

nielit17 94657.16

nielit18 91749.16

nielit19 86419.70

nielit20 76253.86

nielit21 78389.47

nielit22 73994.56

nielit23 67532.53

nielit24 77044.01

nielit25 64664.71

nielit26 75328.87

nielit27 72107.60

nielit28 66051.52

nielit29 65605.48

nielit30 61994.48

nielit31 61136.38

nielit32 63408.86

nielit33 55493.95

nielit34 46426.07

nielit35 46014.02

nielit36 28663.76

nielit37 44069.95

nielit38 20229.59

nielit39 38558.51

nielit40 28754.33

nielit41 27892.92

nielit42 23640.93

nielit43 15505.73

nielit44 22177.74

nielit45 1000.23

nielit46 1315.46

nielit47 0.00

nielit48 542.05

nielit49 0.00

Name: R&D Spend, dtype: float64

st.sum()

Out[16]: 3686080.7800000003

st.max()

Out[17]: 165349.2

st.min()

Out[18]: 0.0

st.mean()

Out[19]: 73721.61559999999

st.idxmax()

Out[20]: 0

st.head(10)

Out[21]:

0 165349.20

1 162597.70

2 153441.51

3 144372.41

4 142107.34

5 131876.90

6 134615.46

7 130298.13

8 120542.52

9 123334.88

Name: R&D Spend, dtype: float64

st.tail(10)

Out[22]:

40 28754.33

41 27892.92

42 23640.93

43 15505.73

44 22177.74

45 1000.23

46 1315.46

47 0.00

48 542.05

49 0.00

Name: R&D Spend, dtype: float64

st.describe()

Out[23]:

count 50.000000

mean 73721.615600

std 45902.256482

min 0.000000

25% 39936.370000

50% 73051.080000

75% 101602.800000

max 165349.200000

Name: R&D Spend, dtype: float64

st1=pd.read_csv('drinks.csv')

out

Traceback (most recent call last):

File "C:\Users\nielit\AppData\Local\Temp/ipykernel_3020/754039757.py", line 1, in <module>

st1=pd.read_csv('drinks.csv')

st1=pd.read_csv('D:\pyth\drinks.csv')

a=[12,22,27,30,33,36,35,34,32,27,20,11]

b=['jan','feb','mar','april','may','june','july','aug','sept','oct'

,'nov','dec']

s1=pd.Series(a,b)

s1=pd.Series(a)

s1=pd.Series(a,b)

s1=pd.Series(data=a,index=b)

s1=pd.Series(data=b,index=a)

s1=pd.Series(data=a,index=b)

s1.sort_values()

Out[34]:

dec 11

jan 12

nov 20

feb 22

mar 27

oct 27

april 30

sept 32

may 33

aug 34

july 35

june 36

dtype: int64

s1.sort_values(ascending=False)

Out[35]:

june 36

july 35

aug 34

may 33

sept 32

april 30

mar 27

oct 27

feb 22

nov 20

jan 12

dec 11

dtype: int64

st.sort_values()

Out[36]:

49 0.00

47 0.00

48 542.05

45 1000.23

46 1315.46

43 15505.73

38 20229.59

44 22177.74

42 23640.93

41 27892.92

36 28663.76

40 28754.33

39 38558.51

37 44069.95

35 46014.02

34 46426.07

33 55493.95

31 61136.38

30 61994.48

32 63408.86

25 64664.71

29 65605.48

28 66051.52

23 67532.53

27 72107.60

22 73994.56

26 75328.87

20 76253.86

24 77044.01

16 78013.11

21 78389.47

19 86419.70

18 91749.16

13 91992.39

12 93863.75

17 94657.16

11 100671.96

10 101913.08

15 114523.61

14 119943.24

8 120542.52

9 123334.88

7 130298.13

5 131876.90

6 134615.46

4 142107.34

3 144372.41

2 153441.51

1 162597.70

0 165349.20

Name: R&D Spend, dtype: float64

s1

Out[37]:

jan 12

feb 22

mar 27

april 30

may 33

june 36

july 35

aug 34

sept 32

oct 27

nov 20

dec 11

dtype: int64

s1.sort_values(inplace=True)

s1=pd.Series(data=a,index=b)

s1.sort_index()

Out[40]:

april 30

aug 34

dec 11

feb 22

jan 12

july 35

june 36

mar 27

may 33

nov 20

oct 27

sept 32

dtype: int64

s1[4:]

Out[41]:

may 33

june 36

july 35

aug 34

sept 32

oct 27

nov 20

dec 11

dtype: int64

s1[5:11]

Out[42]:

june 36

july 35

aug 34

sept 32

oct 27

nov 20

dtype: int64

s1[[2,10]]

Out[43]:

mar 27

nov 20

dtype: int64

Data Frame

Editor Window

```
import pandas as pd
import numpy as np
a=np.random.randn(5,5)
df=pd.DataFrame(a,columns=['col1','col2','col3',
                           'col4','col5'])
df1=pd.read_csv('Salary_Data.csv',names=['exp','salary'])
df2=pd.read_csv('sales_data.csv')
df2.index
df2.values
df2.shape
df2.size
df2.columns
df2.dtypes
df2.info()
df2.describe()
df3=df2.head(20)
df4=df2['MONTH_ID']
type(df4)
df5=df2[['QUANTITYORDERED','CITY']]
df2['vehicle-id']=np.arange(2823)
df2['QUANTITYORDERED']=df2['QUANTITYORDERED']+10
```

o/p window

```
import pandas as pd
import numpy as np
a=np.random.randn(5,5)

df=pd.DataFrame(a)

df=pd.DataFrame(a,columns=['col1','col2','col3',
                           'col4','col5'])



df1=pd.read_csv('Salary_Data.csv')

df1=pd.read_csv('Salary_Data.csv',names=['exp','salary'])

df2=pd.read_csv('sales_data.csv')

df2.index
Out[54]: RangeIndex(start=0, stop=2823, step=1)

df2.values
Out[55]:
array([[30, 95.7, 2, ..., 'USA', nan, 'Small'],
       [34, 81.35, 5, ..., 'France', 'EMEA', 'Small'],
       [41, 94.74, 2, ..., 'France', 'EMEA', 'Medium'],
       ...,
       [43, 100.0, 4, ..., 'Spain', 'EMEA', 'Medium'],
```

```
       [34, 62.24, 1, ..., 'France', 'EMEA', 'Small'],
       [47, 65.52, 9, ..., 'USA', nan, 'Medium']], dtype=object)
```

df2.shape

```
Out[56]: (2823, 13)
```

df2.size

```
Out[57]: 36699
```

df2.columns

```
Out[58]:
Index(['QUANTITYORDERED', 'PRICEEACH', 'ORDERLINENUMBER', 'SALES', 'STATUS',
       'MONTH_ID', 'YEAR_ID', 'PRODUCTLINE', 'CUSTOMERNAME', 'CITY', 'COUNTRY',
       'TERRITORY', 'DEALSIZE'],
      dtype='object')
```

df2.dtypes

```
Out[59]:
QUANTITYORDERED      int64
PRICEEACH          float64
ORDERLINENUMBER      int64
SALES              float64
STATUS              object
MONTH_ID             int64
YEAR_ID              int64
PRODUCTLINE         object
CUSTOMERNAME        object
CITY                object
COUNTRY             object
```

TERRITORY object

DEALSIZE object

dtype: object

df2.info()

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

RangeIndex: 2823 entries, 0 to 2822

Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	QUANTITYORDERED	2823 non-null	int64
1	PRICEEACH	2823 non-null	float64
2	ORDERLINENUMBER	2823 non-null	int64
3	SALES	2823 non-null	float64
4	STATUS	2823 non-null	object
5	MONTH_ID	2823 non-null	int64
6	YEAR_ID	2823 non-null	int64
7	PRODUCTLINE	2823 non-null	object
8	CUSTOMERNAME	2823 non-null	object
9	CITY	2823 non-null	object
10	COUNTRY	2823 non-null	object
11	TERRITORY	1749 non-null	object
12	DEALSIZE	2823 non-null	object

dtypes: float64(2), int64(4), object(7)

memory usage: 286.8+ KB

df2.describe()

Out[61]:

	QUANTITYORDERED	PRICEEACH	...	MONTH_ID	YEAR_ID

count	2823.000000	2823.000000	...	2823.000000	2823.00000
mean	35.092809	83.658544	...	7.092455	2003.81509
std	9.741443	20.174277	...	3.656633	0.69967
min	6.000000	26.880000	...	1.000000	2003.00000
25%	27.000000	68.860000	...	4.000000	2003.00000
50%	35.000000	95.700000	...	8.000000	2004.00000
75%	43.000000	100.000000	...	11.000000	2004.00000
max	97.000000	100.000000	...	12.000000	2005.00000

[8 rows x 6 columns]

df2.head(20)

Out[62]:

	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	...	COUNTRY	TERRITORY	DEALSIZE
0	30	95.70	2	...	USA	NaN	Small
1	34	81.35	5	...	France	EMEA	Small
2	41	94.74	2	...	France	EMEA	Medium
3	45	83.26	6	...	USA	NaN	Medium
4	49	100.00	14	...	USA	NaN	Medium
5	36	96.66	1	...	USA	NaN	Medium
6	29	86.13	9	...	France	EMEA	Small
7	48	100.00	1	...	Norway	EMEA	Medium
8	22	98.57	2	...	USA	NaN	Small
9	41	100.00	14	...	France	EMEA	Medium
10	37	100.00	1	...	Australia	APAC	Medium
11	23	100.00	7	...	USA	NaN	Small
12	28	100.00	2	...	USA	NaN	Medium
13	34	100.00	2	...	USA	NaN	Medium
14	45	92.83	1	...	France	EMEA	Medium

15 36 100.00 6 ... USA NaN Medium

16 23 100.00 9 ... Finland EMEA Small

17 41 100.00 5 ... Norway EMEA Medium

18 46 94.74 1 ... USA NaN Medium

19 42 100.00 1 ... USA NaN Medium

[20 rows x 13 columns]

df3=df2.head(20)

df2['MONTH_ID']

Out[64]:

0 2

1 5

2 7

3 8

4 10

..

2818 12

2819 1

2820 3

2821 3

2822 5

Name: MONTH_ID, Length: 2823, dtype: int64

df4=df2['MONTH_ID']

type(df4)

Out[67]: pandas.core.series.Series

```
df5=df2[['QUANTITYORDERED','CITY']]
```

```
df['vehicle-id']=1
```

```
df2['vehicle-id']=1
```

```
df2['vehicle-id']=np.arange(2823)
```

```
df2['QUANTITYORDERED']=df2['QUANTITYORDERED']+10
```