PYTHON PANDAS TUTORIAL-PART-1

In part we are going to learn about

1.Pandas Dataframe

2.Pandas Series

3.Pandas Basic Operation

In [1]:

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

In [52]:

```
np.arange(0,20)
```

Out[52]:

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19])
```

In [53]:

```
np.arange(0,20).reshape(5,4)
```

Out[53]:

```
array([[ 0,  1,  2,  3],
       [ 4,  5,  6,  7],
       [ 8,  9, 10, 11],
       [12, 13, 14, 15],
       [16, 17, 18, 19]])
```

In [54]:

```
## Creat DAtaframe
df=pd.DataFrame(data=np.arange(0,20).reshape(5,4),index=["Row1","Row2","Row3","Row4","Ro
             ["columns1","columns2","columns3","columns4"])
```

In [55]:

```
df.head()
```

Out[55]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [5]:

```
df.tail()
```

Out[5]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [6]:

```
type(df)
```

Out[6]:

```
pandas.core.frame.DataFrame
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 5 entries, Row1 to Row5
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   columns1  5 non-null      int32
 1   columns2  5 non-null      int32
 2   columns3  5 non-null      int32
 3   columns4  5 non-null      int32
dtypes: int32(4)
memory usage: 120.0+ bytes
```

In [8]:

```
df.describe()
```

Out[8]:

	columns1	columns2	columns3	columns4
count	5.000000	5.000000	5.000000	5.000000
mean	8.000000	9.000000	10.000000	11.000000
std	6.324555	6.324555	6.324555	6.324555
min	0.000000	1.000000	2.000000	3.000000
25%	4.000000	5.000000	6.000000	7.000000
50%	8.000000	9.000000	10.000000	11.000000
75%	12.000000	13.000000	14.000000	15.000000
max	16.000000	17.000000	18.000000	19.000000

In [9]:

```
## Indexing
## Directby using columname,rowindex[loc],rowindexnumber[.iloc]
df.head()
```

Out[9]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [10]:

```
df[['columns1','columns2','columns3']]
```

Out[10]:

	columns1	columns2	columns3
Row1	0	1	2
Row2	4	5	6
Row3	8	9	10
Row4	12	13	14
Row5	16	17	18

In [11]:

```
##Columnsname
df['columns1']
```

Out[11]:

```
Row1     0
Row2     4
Row3     8
Row4    12
Row5    16
Name: columns1, dtype: int32
```

In [12]:

```
df[['columns1','columns2']]
```

Out[12]:

	columns1	columns2
Row1	0	1
Row2	4	5
Row3	8	9
Row4	12	13
Row5	16	17

In [13]:

```
type(df[['columns1','columns2','columns3']])
```

Out[13]:

```
pandas.core.frame.DataFrame
```

In [14]:

```
type(df['columns1'])
```

Out[14]:

```
pandas.core.series.Series
```

In [15]:

```
df.loc['Row3']
```

Out[15]:

```
columns1     8
columns2     9
columns3    10
columns4    11
Name: Row3, dtype: int32
```

In [25]:

```
type(df.loc['Row3'])
```

Out[25]:

```
pandas.core.series.Series
```

In [17]:

```
df.loc[['Row1','Row2']]
```

Out[17]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7

In [18]:

```
df.head()
```

Out[18]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [19]:

```
df.iloc[2:4,0:2]
```

Out[19]:

	columns1	columns2
Row3	8	9
Row4	12	13

In [20]:

```
df.iloc[1:3,1:3]
```

Out[20]:

	columns2	columns3
Row2	5	6
Row3	9	10

In [21]:

```
df.iloc[2:5,1:4]
```

Out[21]:

	columns2	columns3	columns4
Row3	9	10	11
Row4	13	14	15
Row5	17	18	19

In [22]:

```
df.iloc[0:5,:4]
```

Out[22]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [26]:

```
df.iloc[2:5,1:4]
```

Out[26]:

	columns2	columns3	columns4
Row3	9	10	11
Row4	13	14	15
Row5	17	18	19

In [27]:

```
df.iloc[2:,1:]
```

Out[27]:

	columns2	columns3	columns4
Row3	9	10	11
Row4	13	14	15
Row5	17	18	19

In [31]:

```
df.iloc[:,1:]
```

Out[31]:

	columns2	columns3	columns4
Row1	1	2	3
Row2	5	6	7
Row3	9	10	11
Row4	13	14	15
Row5	17	18	19

In [33]:

```
## Covert dataframe into array
df.iloc[:,1:].values
```

Out[33]:

```
array([[ 1,  2,  3],
       [ 5,  6,  7],
       [ 9, 10, 11],
       [13, 14, 15],
       [17, 18, 19]])
```

In [35]:

```
#### Basic Operation
df.isnull().sum()
```

Out[35]:

```
columns1    0
columns2    0
columns3    0
columns4    0
dtype: int64
```

In [38]:

```
df=pd.DataFrame(data=[[1,np.nan,2],[1,2,3]],index=["Row1","Row2"],columns=["columns1","c
```

In [39]:

```
df
```

Out[39]:

	columns1	columns2	columns3
Row1	1	NaN	2
Row2	1	2.0	3

In [40]:

```
df.isnull().sum()
```

Out[40]:

```
columns1    0
columns2    1
columns3    0
dtype: int64
```

In [41]:

```
df.isnull()
```

Out[41]:

	columns1	columns2	columns3
Row1	False	True	False
Row2	False	False	False

In [42]:

```
df.isnull().sum()==0
```

Out[42]:

```
columns1     True
columns2    False
columns3     True
dtype: bool
```

In [47]:

```
df['columns3'].value_counts()
```

Out[47]:

```
2    1
3    1
Name: columns3, dtype: int64
```

In [57]:

```
df
```

Out[57]:

	columns1	columns2	columns3	columns4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [56]:

```
df['columns2'].unique()
```

Out[56]:

```
array([ 1,  5,  9, 13, 17])
```

In [58]:

```
df>2
```

Out[58]:

	columns1	columns2	columns3	columns4
Row1	False	False	False	True
Row2	True	True	True	True
Row3	True	True	True	True
Row4	True	True	True	True
Row5	True	True	True	True

In [60]:

```
df[df['columns2']>2]
```

Out[60]:

	columns1	columns2	columns3	columns4
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [ ]:

```

```