Pandas Tutorial

Pandas is an open source, BSD-licensed library providing high-performance,easy-to-use data sturtures and data analysis tools for the python programming language.

Agenda

What is Data Frame? - it is combination of row and column or greater than one column or row
What is Data series?
Different operation in Pandas

In [1]:

```
## first Step is to import pandas
import pandas as pd
import numpy as np
```

In [6]:

```
## playing with Dataframe
df=pd.DataFrame(np.arange(0,20).reshape(5,4),index=['Row1','Row2','Row3','Row4','Row5'],col
```

In [10]:

```
df.head()
```

Out[10]:

	Column1	Column2	Column3	Column4
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 [11]:

```
df.to_csv('Test1.csv')
```

In [15]:

```
## Accessing the elements
# 1. .loc 2. iloc
df.loc['Row1']
```

Out[15]:

```
Column1    0
Column2    1
Column3    2
Column4    3
Name: Row1, dtype: int32
```

In [16]:

```
type(df.loc['Row1'])        # check the type
```

Out[16]:

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

In [18]:

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

Out[18]:

	Column1	Column2
Row1	0	1
Row2	4	5
Row3	8	9
Row4	12	13

In [19]:

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

Out[19]:

	Column1	Column2	Column3	Column4
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 [21]:

```
## Take the elements from the column2
df.iloc[:,1:]
```

Out[21]:

	Column2	Column3	Column4
Row1	1	2	3
Row2	5	6	7
Row3	9	10	11
Row4	13	14	15
Row5	17	18	19

In [22]:

```
type(df.iloc[:,1:])
```

Out[22]:

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

In [25]:

```
type(df.iloc[:,1:2])
```

Out[25]:

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

In [27]:

```
# convert DataFrame into array
df.iloc[:,:].values
```

Out[27]:

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

In [29]:

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

Out[29]:

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

In [30]:

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

Out[30]:

```
(5, 3)
```

In [31]:

```
## how to check null values
df.isnull().sum()
```

Out[31]:

```
Column1    0
Column2    0
Column3    0
Column4    0
dtype: int64
```

In [32]:

```
df
```

Out[32]:

	Column1	Column2	Column3	Column4
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 [35]:

```
df['Column1'].value_counts()
```

Out[35]:

```
0     1
4     1
8     1
12    1
16    1
Name: Column1, dtype: int64
```

In [36]:

```
df['Column1'].unique
```

Out[36]:

```
<bound method Series.unique of Row1     0
Row2     4
Row3     8
Row4    12
Row5    16
Name: Column1, dtype: int32>
```

In [ ]:

```

```