Pandas Series

April 13, 2020

1 Pandas

Pandas is a Python library for data analysis. It offers a number of data exploration, cleaning and transformation operations that are critical in working with data in Python.

pandas build upon numpy and scipy providing easy-to-use data structures and data manipulation functions with integrated indexing.

The main data structures pandas provides are Series and DataFrames. After a brief introduction to these two data structures and data ingestion, the key features of pandas are:

- Generating descriptive statistics on data
- Data cleaning using built in pandas functions
- Frequent data operations for subsetting, filtering, insertion, deletion and aggregation of data
- Merging multiple datasets using dataframes
- Working with timestamps and time-series data

[3]:
```python
import pandas as pd
```

[4]:
```python
x=pd.Series([100,200,300,400,500],['tom','bob','eric','john','nancy'])
```

[5]:
```python
print(x)
```

```
tom      100
bob      200
eric     300
john     400
nancy    500
dtype: int64
```

[6]:
```python
x
```

[6]:
```
tom      100
bob      200
eric     300
john     400
nancy    500
dtype: int64
```

```
[7]: x.index
```

```
[7]: Index(['tom', 'bob', 'eric', 'john', 'nancy'], dtype='object')
```

```
[8]: x.loc[['john','nancy']]
```

```
[8]: john     400
     nancy    500
     dtype: int64
```

```
[9]: x[[4,3,1]]
```

```
[9]: nancy    500
     john     400
     bob      200
     dtype: int64
```

```
[10]: x.iloc[2]
```

```
[10]: 300
```

```
[11]: x*2
```

```
[11]: tom       200
      bob       400
      eric      600
      john      800
      nancy    1000
      dtype: int64
```

```
[12]: x+2
```

```
[12]: tom      102
      bob      202
      eric     302
      john     402
      nancy    502
      dtype: int64
```

```
[13]: x[['nancy','eric']]**2
```

```
[13]: nancy    250000
      eric      90000
      dtype: int64
```

2 different data types

[14]:
```
import numpy as np
```

[15]:
```
labels=['a','b','c']
arr=np.array([10,20,30])
my_list=[10,20,30]
d={'a':10,'b':20,'c':30}
```

[16]:
```
#data, index
pd.Series(data=my_list)
```

[16]:
```
0    10
1    20
2    30
dtype: int64
```

[17]:
```
pd.Series(data=my_list,index=labels)
```

[17]:
```
a    10
b    20
c    30
dtype: int64
```

[18]:
```
pd.Series(my_list,labels)
```

[18]:
```
a    10
b    20
c    30
dtype: int64
```

[19]:
```
pd.Series(labels,my_list)
```

[19]:
```
10    a
20    b
30    c
dtype: object
```

[20]:
```
pd.Series(arr)
```

[20]:
```
0    10
1    20
2    30
dtype: int32
```

[21]:
```
pd.Series(arr,labels)
```

```
[21]: a    10
      b    20
      c    30
      dtype: int32
```

```
[22]: pd.Series(d)
```

```
[22]: a    10
      b    20
      c    30
      dtype: int64
```

```
[23]: # Using an Index
```

```
[24]: ser1=pd.Series([1,2,3,4],['USA','Germany','USSR','Japan'])
```

```
[25]: print(ser1)
```

```
USA        1
Germany    2
USSR       3
Japan      4
dtype: int64
```

```
[26]: ser2=pd.Series([1,2,5,4],['USA','Germany','Italy','Japan'])
```

```
[27]: print(ser2)
```

```
USA        1
Germany    2
Italy      5
Japan      4
dtype: int64
```

```
[28]: ser1+ser2
```

```
[28]: Germany    4.0
      Italy      NaN
      Japan      8.0
      USA        2.0
      USSR       NaN
      dtype: float64
```

```
[33]: ser3=pd.Series(index=[1,2,5,4],data=['Germany','USA','Italy','Japan'])
```

```
[35]: ser3
```

```
[35]: 1    Germany
      2        USA
      5      Italy
      4      Japan
      dtype: object
```

```
[ ]:
```