Pandas - DataFrame

- A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns.

Features of DataFrame:-

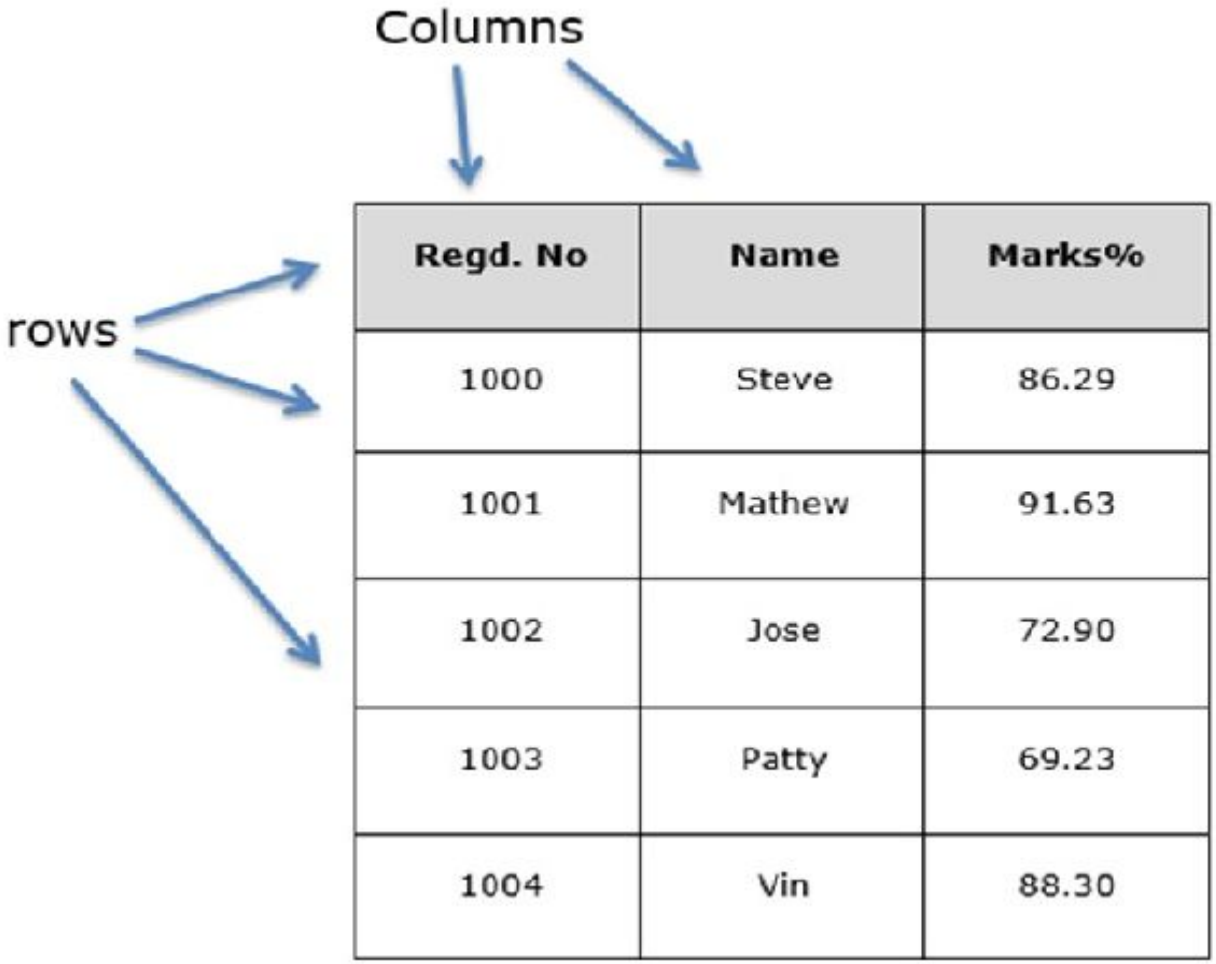
- Columns can be of different types
- Size – Mutable
- Labeled axes (rows and columns)
- Can Perform Arithmetic operations on rows and columns

Structure

Pandas

INTERNSHIPSTUDIO

- Let us assume that we are creating a data frame with student's data.



Regd. No	Name	Marks%
1000	Steve	86.29
1001	Mathew	91.63
1002	Jose	72.90
1003	Patty	69.23
1004	Vin	88.30

DataFrame- constructor

- A pandas DataFrame can be created using the following constructor –

pandas.DataFrame(data, index, columns, dtype, copy)

The parameters of the constructor are as follows –

- Data- data takes various forms like ndarray, series, map, lists, dict, constants and also another DataFrame.
- Index- For the row labels, the Index to be used for the resulting frame is Optional
- Columns- For column labels, the optional default syntax is np.arange(n). This is only true if no index is passed.
- Dtype- Data type of each column.
- Copy- his command (or whatever it is) is used for copying of data, if the default is False.

Create DataFrame

A pandas DataFrame can be created using various inputs like :–

- Lists
- dict
- Series
- Numpy ndarrays
- Another DataFrame

Create an Empty DataFrame

#import the pandas library and aliasing as pd

```
import pandas as pd
df = pd.DataFrame()
print df
```

Its output is as follows −

- Empty DataFrame
- Columns: []
- Index: []

Create a DataFrame from Lists

- Example 1

```
import pandas as pd
data = [1,2,3,4,5]
df = pd.DataFrame(data)
print df
```

Its **output** is as follows –

```
   0
0  1
1  2
2  3
3  4
4  5
```

Example 2

```
import pandas as pd
data = [['Alex',10],['Bob',12],['Clarke',13]]
df=pd.DataFrame(data,columns=['Name','Age'])
print df
```

Its **output** is as follows:−

Name Age
0 Alex 10
1 Bob 12
2 Clarke 13

Example 3

```
import pandas as pd
data = [['Alex',10],['Bob',12],['Clarke',13]]
df=pd.DataFrame(data,columns=['Name','Age'],dtype=float)
print df
```

Its **output** is as follows –

```
Name Age
 0 Alex 10.0
 1 Bob 12.0
 2 Clarke 13.0
```

Note – Observe, the **dtype** parameter changes the type of Age column to floating point.

Create a DataFrame from Dict- ndarrays / Lists

- All the **ndarrays** must be of same length. If index is passed, then the length of the index should equal to the length of the arrays.
- If no index is passed, then by default, index will be range(n), where **n** is the array length.

Example 1

```
import pandas as pd
data={'Name':['Tom','Jack','Steve','Ricky'],'Age':[28
,34,29,42]}
df = pd.DataFrame(data)
print df
```

Its **output** is as follows –

	Age	Name
0	28	Tom
1	34	Jack
2	29	Steve
3	42	Ricky

Note – Observe the values 0,1,2,3. They are the default index assigned to each using the function range(n).

Example 2

- Let us now create an indexed DataFrame using arrays.

```
import pandas as pd
 data = {'Name':['Tom', 'Jack', 'Steve',
'Ricky'],'Age':[28,34,29,42]}
df=pd.DataFrame(data,index=['rank1','rank2',
'rank3','rank4'])
print df
```

Its **output** is as follows –

```
        Age    Name
rank1    28     Tom
rank2    34    Jack
rank3    29   Steve
rank4    42   Ricky
```

Note – Observe, the **index** parameter assigns an index to each row.

Create a DataFrame from List of Dicts

- List of Dictionaries can be passed as input data to create a DataFrame. The dictionary keys are by default taken as column names.

Example 1

- The following example shows how to create a DataFrame by passing a list of dictionaries.

```
import pandas as pd
data = [{'a': 1, 'b': 2},{'a': 5, 'b': 10, 'c': 20}]
df = pd.DataFrame(data)
print df
```

Its **output** is as follows –

```
   a   b    c
0  1   2  NaN
1  5  10  20.0
```

Note – Observe, NaN (Not a Number) is appended in missing areas.

Example 2

- The following example shows how to create a DataFrame by passing a list of dictionaries and the row indices.

```
import pandas as pd
data = [{'a': 1, 'b': 2},{'a': 5, 'b': 10, 'c': 20}]
df = pd.DataFrame(data, index=['first', 'second'])
print df
```

Its **output** is as follows –

```
        a   b     c
first   1   2   NaN
second  5   10  20.0
```

Create a DataFrame from Dict of Series

- Dictionary of Series can be passed to form a DataFrame. The resultant index is the union of all the series indexes passed.

Example-

```
import pandas as pd
 d = {'one' : pd.Series([1, 2, 3], index=['a', 'b', 'c']), 'two' :
    pd.Series([1, 2, 3, 4], index=['a', 'b', 'c', 'd'])}
df = pd.DataFrame(d)
print df
```

Its **output** is as follows –

```
      one    two
a     1.0    1
b     2.0    2
c     3.0    3
d     NaN    4
```

Note – Observe, for the series one, there is no label **'d'** passed, but in the result, for the **d** label, NaN is appended with NaN.

Pandas

INTERNSHIPSTUDIO

- How to create an Empty DataFrame?
- How to create a DataFrame from Lists ?
- How to create a DataFrame from Dict of Series?
- How to create a DataFrame from List of Dicts?
- How to create a DataFrame from Dict of ndarrays / Lists ?