In [1]:
```
import pandas as pd
```

In [8]:
```
l = [1,2,3]
s = pd.Series(l)
s
```

Out[8]:
```
0    1
1    2
2    3
dtype: int64
```

In [10]:
```
s1 =pd.Series([1,2,3,4],index =[1,2,3,4])
s1
```

Out[10]:
```
1    1
2    2
3    3
4    4
dtype: int64
```

In [6]:
```
s1["c"]
```

Out[6]:
```
3
```

In [11]:
```
s1[2]
```

Out[11]:
```
2
```

In [9]:
```
print(s1[0:3])  # 0,1,2
print("====================")
print(s1["a":"d"])

```

```
a    1
b    2
c    3
dtype: int64
====================
a    1
b    2
c    3
d    4
dtype: int64
```

In [14]:
```
d = {"maths":88,"Science":78,"English":92}
pd.Series(d)
```

Out[14]:
```
maths      88
Science    78
English    92
dtype: int64
```

DataFrame

In [18]:
```
import numpy as np
a = np.array([[1,2,3,4],[4,5,6,7]])
pd.DataFrame(a,columns=[1,2,3,4], index= ['a','b'])
```

Out[18]:

	1	2	3	4
a	1	2	3	4
b	4	5	6	7

In [27]:
```
d = {"name":["Sri","lalitha","mits","college"],"marks":[78,90,67,56]}
df = pd.DataFrame(d)
df
```

. . .

In [26]:
```
df[0:2]
```

Out[26]:

	name	marks
0	Sri	78
1	lalitha	90

Dataset

- large amount of data will be stored in a file

Reading a dataset

- pd.read_file extented

formats

- csv====> camma seperated values
  - pd.read_csv
- txt ===> text file
  - pd.read_table
- tsv ===> tab seperated values
  - pd.read_tsv

- excel files
  - pd.read_excel
- html files
  - pd.read_html

In [44]:
```
df = pd.read_csv("marks.csv",header = None,names=["Student name","m1","m2","
df
```

Out[44]:

	Student name	m1	m2	m3
0	Sri	67	78.0	99.0
1	Lalitha	78	78.0	NaN
2	Mits	88	71.0	56.0
3	College	91	56.0	78.0
4	APSSDC	88	91.0	56.2
5	Python	90	NaN	65.0
6	Programming	81	76.0	45.0

In [43]:
```
df.m1
```

Out[43]:
```
0    67
1    78
2    88
3    91
4    88
5    90
6    81
Name: m1, dtype: int64
```

In [45]:
```
df.Student name
```

```
  File "<ipython-input-45-a62af85847d0>", line 1
    df.Student name
                   ^
SyntaxError: invalid syntax
```

In [47]:
```
df["Student name"]
df["m2"]
```

Out[47]:
```
0    78.0
1    78.0
2    71.0
3    56.0
4    91.0
5     NaN
6    76.0
Name: m2, dtype: float64
```

In [50]:
```
df[["Student name","m2"]]
```

...

In [51]:
```
df.index
```

Out[51]:
```
RangeIndex(start=0, stop=7, step=1)
```

In [52]:
```
df.columns
```

Out[52]:
```
Index(['Student name', 'm1', 'm2', 'm3'], dtype='object')
```

In [53]:
```
df.values
```

Out[53]:
```
array([['Sri', 67, 78.0, 99.0],
       ['Lalitha', 78, 78.0, nan],
       ['Mits', 88, 71.0, 56.0],
       ['College', 91, 56.0, 78.0],
       ['APSSDC', 88, 91.0, 56.2],
       ['Python', 90, nan, 65.0],
       ['Programming', 81, 76.0, 45.0]], dtype=object)
```

In [54]:
```
df.dtypes
```

Out[54]:
```
Student name     object
m1                int64
m2              float64
m3              float64
dtype: object
```

In [56]:
```
df.shape
```

Out[56]:
```
(7, 4)
```

Accessing of data from dataset

- index based accessing (iloc)
- location based accessing (loc)

In [57]:
```
df
```

Out[57]:

	Student name	m1	m2	m3
0	Sri	67	78.0	99.0
1	Lalitha	78	78.0	NaN
2	Mits	88	71.0	56.0
3	College	91	56.0	78.0
4	APSSDC	88	91.0	56.2
5	Python	90	NaN	65.0
6	Programming	81	76.0	45.0

In [58]:
```
df.iloc[2:5]
#2,3,4
```

Out[58]:

	Student name	m1	m2	m3
2	Mits	88	71.0	56.0
3	College	91	56.0	78.0
4	APSSDC	88	91.0	56.2

In [61]:
```
df.iloc[2:5:2,1:5:2]
# col = 1,2,3,4
```

Out[61]:

	m1	m3
2	88	56.0
4	88	56.2

In [62]:
```
df.loc[1:5,"Student name":"m2"]
```

Out[62]:

	Student name	m1	m2
1	Lalitha	78	78.0
2	Mits	88	71.0
3	College	91	56.0
4	APSSDC	88	91.0
5	Python	90	NaN

In [65]:
```
df.set_index("Student name", inplace = True)
```

In [66]:
```
1 df
```

Out[66]:

Student name	m1	m2	m3
Sri	67	78.0	99.0
Lalitha	78	78.0	NaN
Mits	88	71.0	56.0
College	91	56.0	78.0
APSSDC	88	91.0	56.2
Python	90	NaN	65.0
Programming	81	76.0	45.0

In [69]:
```
1 df.loc["Lalitha":"Python","m1":"m2"]
```

Out[69]:

Student name	m1	m2
Lalitha	78	78.0
Mits	88	71.0
College	91	56.0
APSSDC	88	91.0
Python	90	NaN

In [70]:
```
1 df = pd.read_csv("marks.csv",header = None,names=["Student name","m1","m2","
2 df
```

Out[70]:

	Student name	m1	m2	m3
0	Sri	67	78.0	99.0
1	Lalitha	78	78.0	NaN
2	Mits	88	71.0	56.0
3	College	91	56.0	78.0
4	APSSDC	88	91.0	56.2
5	Python	90	NaN	65.0
6	Programming	81	76.0	45.0

In [71]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7 entries, 0 to 6
Data columns (total 4 columns):
Student name    7 non-null object
m1              7 non-null int64
m2              6 non-null float64
m3              6 non-null float64
dtypes: float64(2), int64(1), object(1)
memory usage: 236.0+ bytes
```

In [79]:
```
1 df.describe()
2
```

. . .

In [80]:
```
1 df.describe(include = "object")
```

Out[80]:

	Student name
count	7
unique	7
top	Python
freq	1

In [81]:
```
1 df.describe(include="all")
```

Out[81]:

	Student name	m1	m2	m3
count	7	7.000000	6.000000	6.000000
unique	7	NaN	NaN	NaN
top	Python	NaN	NaN	NaN
freq	1	NaN	NaN	NaN
mean	NaN	83.285714	75.000000	66.533333
std	NaN	8.635475	11.419282	19.335632
min	NaN	67.000000	56.000000	45.000000
25%	NaN	79.500000	72.250000	56.050000
50%	NaN	88.000000	77.000000	60.600000
75%	NaN	89.000000	78.000000	74.750000
max	NaN	91.000000	91.000000	99.000000

In [84]:
```
df["total"] = df["m1"] + df["m2"]+ df["m3"]
df
```

Out[84]:

	Student name	m1	m2	m3	total
0	Sri	67	78.0	99.0	244.0
1	Lalitha	78	78.0	NaN	NaN
2	Mits	88	71.0	56.0	215.0
3	College	91	56.0	78.0	225.0
4	APSSDC	88	91.0	56.2	235.2
5	Python	90	NaN	65.0	NaN
6	Programming	81	76.0	45.0	202.0

In [ ]:
```
1
```