In [1]:

```python
import pandas
```

In [2]:

```python
data = {"Name" : ['Joe' , 'Nat' , "harry" , 'Sam' , 'Monica'] ,
        "Age" : [20 , 21 , 19 , 21 , 22] ,
        "marks" : [85.10 , 77.80 , 91.54 , 88.78 , 60.55] ,
        "Grade" : ['A' , "B" , "A" , "A" , "B"] ,
        "Hobby" : ['Swmming' , 'Reading' , "Music" , "Painting" , 'Dancing']}
```

In [3]:

```python
data
```

Out[3]:

```
{'Name': ['Joe', 'Nat', 'harry', 'Sam', 'Monica'],
 'Age': [20, 21, 19, 21, 22],
 'marks': [85.1, 77.8, 91.54, 88.78, 60.55],
 'Grade': ['A', 'B', 'A', 'A', 'B'],
 'Hobby': ['Swmming', 'Reading', 'Music', 'Painting', 'Dancing']}
```

In [4]:

```python
# create  dataframe
info =  pandas.DataFrame(data)
```

In [6]:

```python
info
```

Out[6]:

	Name	Age	marks	Grade	Hobby
0	Joe	20	85.10	A	Swmming
1	Nat	21	77.80	B	Reading
2	harry	19	91.54	A	Music
3	Sam	21	88.78	A	Painting
4	Monica	22	60.55	B	Dancing

attributes on dataframe

- index
- size
- shape
- dtypes
- columns
- values

In [7]:

```python
info
```

Out[7]:

	Name	Age	marks	Grade	Hobby
0	Joe	20	85.10	A	Swmming
1	Nat	21	77.80	B	Reading
2	harry	19	91.54	A	Music
3	Sam	21	88.78	A	Painting
4	Monica	22	60.55	B	Dancing

In [8]:
```
info.columns
```
Out[8]:
```
Index(['Name', 'Age', 'marks', 'Grade', 'Hobby'], dtype='object')
```
In [10]:
```
list(info.index)
```
Out[10]:
```
[0, 1, 2, 3, 4]
```
In [11]:
```
info.shape
```
Out[11]:
```
(5, 5)
```
In [12]:
```
info.size
```
Out[12]:
```
25
```
In [13]:
```
info.dtypes
```
Out[13]:
```
Name      object
Age        int64
marks    float64
Grade     object
Hobby     object
dtype: object
```
In [15]:
```
info.values
```
Out[15]:
```
array([['Joe', 20, 85.1, 'A', 'Swmming'],
       ['Nat', 21, 77.8, 'B', 'Reading'],
       ['harry', 19, 91.54, 'A', 'Music'],
       ['Sam', 21, 88.78, 'A', 'Painting'],
       ['Monica', 22, 60.55, 'B', 'Dancing']], dtype=object)
```

functions on dataframe

- info

- isnull()

```
# information about  dataframe
info.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    5 non-null      object
 1   Age     5 non-null      int64
 2   marks   5 non-null      float64
 3   Grade   5 non-null      object
 4   Hobby   5 non-null      object
dtypes: float64(1), int64(1), object(3)
memory usage: 332.0+ bytes
```

In [21]:

```
# check missing values
info.isnull().sum()
```

Out[21]:

```
Name     0
Age      0
marks    0
Grade    0
Hobby    0
dtype: int64
```

In [23]:

```
# change  index  labels
info.index = ['S1' ,  'S2' ,  'S3 ' , 'S4' , 'S5']
```

In [24]:

```
info.index
```

Out[24]:

```
Index(['S1', 'S2', 'S3 ', 'S4', 'S5'], dtype='object')
```

In [25]:

```
info
```

Out[25]:

	Name	Age	marks	Grade	Hobby
S1	Joe	20	85.10	A	Swmming
S2	Nat	21	77.80	B	Reading
S3	harry	19	91.54	A	Music
S4	Sam	21	88.78	A	Painting
S5	Monica	22	60.55	B	Dancing

In [28]:

```
# rename  column name  marks  -- > Marks
# to store permenantly  use  inplace  =  True
```

```
info.rename(columns = {"marks" : "Marks"} , inplace = True)
```

In [29]:
```
info
```

Out[29]:

	Name	Age	Marks	Grade	Hobby
S1	Joe	20	85.10	A	Swmming
S2	Nat	21	77.80	B	Reading
S3	harry	19	91.54	A	Music
S4	Sam	21	88.78	A	Painting
S5	Monica	22	60.55	B	Dancing

In [33]:
```
info.columns
```

Out[33]:
```
Index(['Name', 'Age', 'Marks', 'Grade', 'Hobby'], dtype='object')
```

In [30]:
```
# create  dataframe  using list
l1  =  ['Joe', 20, 85.1, 'A', 'Swmming']
l2  =  ['Nat', 21, 77.8, 'B', 'Reading']
l3  =  ['harry', 19, 91.54, 'A', 'Music']
l4  =  ['Sam', 21, 88.78, 'A', 'Painting']
l5  =  ['Monica', 22, 60.55, 'B', 'Dancing']
```

In [31]:
```
info1  =  pandas.DataFrame(data  =  [l1  ,  l2  , l3  , l4 ,l5])
```

In [32]:
```
info1
```

Out[32]:

	0	1	2	3	4
0	Joe	20	85.10	A	Swmming
1	Nat	21	77.80	B	Reading
2	harry	19	91.54	A	Music
3	Sam	21	88.78	A	Painting
4	Monica	22	60.55	B	Dancing

In [34]:
```
# after creation  of dataframe
info1.index  =  ['S1' ,  'S2' ,  'S3 ' , 'S4' , 'S5']
info1.columns  =  ['Name', 'Age', 'Marks', 'Grade', 'Hobby']
```

In [35]:
```
info1
```

Out[35]:

	Name	Age	Marks	Grade	Hobby
S1	Joe	20	85.10	A	Swmming
S2	Nat	21	77.80	B	Reading
S3	harry	19	91.54	A	Music
S4	Sam	21	88.78	A	Painting
S5	Monica	22	60.55	B	Dancing

In [37]:

```python
info.columns
```

Out[37]:

```
Index(['Name', 'Age', 'Marks', 'Grade', 'Hobby'], dtype='object')
```

In [38]:

```python
# assign index and  column names
info1  =   pandas.DataFrame(data  =  [l1  ,  l2  , l3  , l4 ,l5] ,
                            index  =  ['S1' ,  'S2' ,  'S3 ' , 'S4'  , 'S5' ] ,
                            columns  = ['Name', 'Age', 'Marks', 'Grade', 'Hobby']  )
```

In [39]:

```python
info1
```

Out[39]:

	Name	Age	Marks	Grade	Hobby
S1	Joe	20	85.10	A	Swmming
S2	Nat	21	77.80	B	Reading
S3	harry	19	91.54	A	Music
S4	Sam	21	88.78	A	Painting
S5	Monica	22	60.55	B	Dancing

In [40]:

```python
info1.index
```

Out[40]:

```
Index(['S1', 'S2', 'S3 ', 'S4', 'S5'], dtype='object')
```

In [41]:

```python
info1.columns
```

Out[41]:

```
Index(['Name', 'Age', 'Marks', 'Grade', 'Hobby'], dtype='object')
```

In [42]:

```python
print(info1.shape ,  info1.size  )
```

```
(5, 5) 25
```

In [43]:

```python
print(info1.values)
```

```
[['Joe' 20 85.1 'A' 'Swmming']
 ['Nat' 21 77.8 'B' 'Reading']
```

```
 ['harry' 19 91.54 'A' 'Music']
 ['Sam' 21 88.78 'A' 'Painting']
 ['Monica' 22 60.55 'B' 'Dancing']]
```

In [44]:

```python
import pandas as  pd
```

In [45]:

```python
info3  =  pd.DataFrame(data)
```

In [46]:

```python
info3
```

Out[46]:

	Name	Age	marks	Grade	Hobby
0	Joe	20	85.10	A	Swmming
1	Nat	21	77.80	B	Reading
2	harry	19	91.54	A	Music
3	Sam	21	88.78	A	Painting
4	Monica	22	60.55	B	Dancing

In [47]:

```python
# import dataframe
from pandas import  DataFrame
```

In [48]:

```python
info4  =  DataFrame(data)
```

In [49]:

```python
info4
```

Out[49]:

	Name	Age	marks	Grade	Hobby
0	Joe	20	85.10	A	Swmming
1	Nat	21	77.80	B	Reading
2	harry	19	91.54	A	Music
3	Sam	21	88.78	A	Painting
4	Monica	22	60.55	B	Dancing

Data Subsetting -Extracting data

In [51]:

```python
a  =  info['Name']
a
```

Out[51]:

```
S1      Joe
S2      Nat
S3     harry
```

```
S4      Sam
S5    Monica
Name: Name, dtype: object
```

```
type(info)
```

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

```
type(a)
```

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

```
a.shape
```

```
(5,)
```

```
info.Name
```

```
S1      Joe
S2      Nat
S3    harry
S4      Sam
S5    Monica
Name: Name, dtype: object
```

```
# data from  name  , age
info[['Name' ,  'Age']]
```

	Name	Age
S1	Joe	20
S2	Nat	21
S3	harry	19
S4	Sam	21
S5	Monica	22