PANDAS

pasdas are use full to represent you data into effective way

1. data freame
2. series
3. panal

   data fram is a part of pandas its useful to represent you data in row and column(2D array)

   series is a part of pandas its useful to represent you data in one 1D array

   panal is a part of pandas. panal are useful to represent you data in 3D array

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
# import DATA From Dic.
dic1={
    "name":['rishiraj','kaushal','abhishak'],
    "City":['Ghaziabad','delhi','Bareilly'],
    "marks":[90,80,85]
}
```

In [3]:

```
df=pd.DataFrame(dic1)
df
```

Out[3]:

	name	City	marks
0	rishiraj	Ghaziabad	90
1	kaushal	delhi	80
2	abhishak	Bareilly	85

In [4]:

```
df.to_csv('raj.csv') # convert data in csv formate
```

In [5]:

```
df.to_csv('raj1.csv',index=False) # convert data in csv formate without index values
```

In [6]:

```
df.head(2) # show some number of data from top
```

Out[6]:

	name	City	marks
0	rishiraj	Ghaziabad	90
1	kaushal	delhi	80

In [7]:

```
df.tail(2) # show some number of data from Bottom
```

Out[7]:

	name	City	marks
1	kaushal	delhi	80
2	abhishak	Bareilly	85

In [8]:

```
df1=pd.read_excel('D:/DATA SCIENCE/test.xlsx', 'Sheet1') # data from excel
df1
```

Out[8]:

	ID	NAME	LAST NAME	FEES	COURSE
0	200	RISHIRAJ	SINGH	18000	DS, ML
1	201	RAHUL	KUMAR	18500	B.Sc
2	202	AKANKSHA	RANI	16000	PYRHON,ML
3	203	MONU	THAKUR	15000	JAVA
4	204	AKBAR	KHAN	12000	M.COM
5	205	KAUSHAL	KUMAR	15500	DIPLOMA
6	206	ABHISHAK	SINGH	20500	IIT
7	207	RAJ KISHOR	KUMAR	17000	B-TECH
8	208	RAKESH	ENG.	19500	DIPLOMA
9	209	UMESH	BHAI	17500	B-TECH

In [9]:

```
df1.describe() # calculate every thing of numerical column
```

Out[9]:

	ID	FEES
count	10.00000	10.000000
mean	204.50000	16950.000000
std	3.02765	2454.587542
min	200.00000	12000.000000
25%	202.25000	15625.000000
50%	204.50000	17250.000000
75%	206.75000	18375.000000
max	209.00000	20500.000000

In [10]:

```
df1[df1['ID']>204] # filter
```

Out[10]:

	ID	NAME	LAST NAME	FEES	COURSE
5	205	KAUSHAL	KUMAR	15500	DIPLOMA
6	206	ABHISHAK	SINGH	20500	IIT
7	207	RAJ KISHOR	KUMAR	17000	B-TECH
8	208	RAKESH	ENG.	19500	DIPLOMA
9	209	UMESH	BHAI	17500	B-TECH

In [11]:

```
df2=pd.read_csv('D:/DATA SCIENCE/raj1.csv') # read csv data
df2
```

Out[11]:

	name	City	marks
0	rishiraj	Ghaziabad	90
1	kaushal	delhi	80
2	abhishak	Bareilly	85
3	monu	Loni	92
4	Rahul	tillamor	76
5	Ram	Ghaziabad	82
6	mohan	Delhi	83

In [12]:

```
f=df2[['name','City']] # Select a particuler column
f
```

Out[12]:

	name	City
0	rishiraj	Ghaziabad
1	kaushal	delhi
2	abhishak	Bareilly
3	monu	Loni
4	Rahul	tillamor
5	Ram	Ghaziabad
6	mohan	Delhi

In [13]:

```
df2.shape
```

Out[13]:

```
(7, 3)
```

In [14]:

```
 df2.size
```

Out[14]:

```
21
```

In [15]:

```
print(np.sum(df2))
```

```
name             rishirajkaushalabhishakmonuRahulRammohan
City     GhaziabaddelhiBareillyLonitillamorGhaziabadDelhi
marks                                                 588
dtype: object
```

In [16]:

```
print(np.sum(df1))
```

```
ID                                                   2045
NAME         RISHIRAJRAHULAKANKSHAMONUAKBARKAUSHALABHISHAKR...
LAST NAME      SINGHKUMARRANITHAKURKHANKUMARSINGHKUMARENG.BHAI
FEES                                                169500
COURSE       DS, MLB.ScPYRHON,MLJAVAM.COMDIPLOMAIITB-TECHDI...
dtype: object
```

In [17]:

```
df1.sum(axis=0)
```

Out[17]:

```
ID                                                       2045
NAME          RISHIRAJRAHULAKANKSHAMONUAKBARKAUSHALABHISHAKR...
LAST NAME       SINGHKUMARRANITHAKURKHANKUMARSINGHKUMARENG.BHAI
FEES                                                   169500
COURSE        DS, MLB.ScPYRHON,MLJAVAM.COMDIPLOMAIITB-TECHDI...
dtype: object
```

In [18]:

```
df1.sum(axis=1) # sum of all int value - column wish
```

Out[18]:

```
0    18200
1    18701
2    16202
3    15203
4    12204
5    15705
6    20706
7    17207
8    19708
9    17709
dtype: int64
```

In [19]:

```
np.ndim(df1)
```

Out[19]:

```
2
```

In [20]:

```
df1
```

Out[20]:

	ID	NAME	LAST NAME	FEES	COURSE
0	200	RISHIRAJ	SINGH	18000	DS, ML
1	201	RAHUL	KUMAR	18500	B.Sc
2	202	AKANKSHA	RANI	16000	PYRHON,ML
3	203	MONU	THAKUR	15000	JAVA
4	204	AKBAR	KHAN	12000	M.COM
5	205	KAUSHAL	KUMAR	15500	DIPLOMA
6	206	ABHISHAK	SINGH	20500	IIT
7	207	RAJ KISHOR	KUMAR	17000	B-TECH
8	208	RAKESH	ENG.	19500	DIPLOMA
9	209	UMESH	BHAI	17500	B-TECH

In [21]:

```
SUM=df1['FEES'].sum() # sum of all student fees
SUM
```

Out[21]:

```
169500
```

In [22]:

```
SUM1=df1['ID'].sum()
SUM1
```

Out[22]:

```
2045
```

In [23]:

```
MAX=df1['FEES'].max() # maximum fee
MAX
```

Out[23]:

```
20500
```

In [24]:

```
MIN=df1['FEES'].min() # minmum fee
MIN
```

Out[24]:

```
12000
```

In [25]:

```
SELECT_DATA=df1[df1["ID"]>204]
SELECT_DATA
```

Out[25]:

	ID	NAME	LAST NAME	FEES	COURSE
5	205	KAUSHAL	KUMAR	15500	DIPLOMA
6	206	ABHISHAK	SINGH	20500	IIT
7	207	RAJ KISHOR	KUMAR	17000	B-TECH
8	208	RAKESH	ENG.	19500	DIPLOMA
9	209	UMESH	BHAI	17500	B-TECH

In [26]:

```
df1["DISCOUNT(%)"]=[10,10.5,9.2,3.8,5,6,5.5,6.2,25,14] # add Column
df1
```

Out[26]:

	ID	NAME	LAST NAME	FEES	COURSE	DISCOUNT(%)
0	200	RISHIRAJ	SINGH	18000	DS, ML	10.0
1	201	RAHUL	KUMAR	18500	B.Sc	10.5
2	202	AKANKSHA	RANI	16000	PYRHON,ML	9.2
3	203	MONU	THAKUR	15000	JAVA	3.8
4	204	AKBAR	KHAN	12000	M.COM	5.0
5	205	KAUSHAL	KUMAR	15500	DIPLOMA	6.0
6	206	ABHISHAK	SINGH	20500	IIT	5.5
7	207	RAJ KISHOR	KUMAR	17000	B-TECH	6.2
8	208	RAKESH	ENG.	19500	DIPLOMA	25.0
9	209	UMESH	BHAI	17500	B-TECH	14.0

In [27]:

```
 df1=pd.read_excel('D:/DATA SCIENCE/test.xlsx', 'Sheet1') # data from excel
df1
```

Out[27]:

	ID	NAME	LAST NAME	FEES	COURSE
0	200	RISHIRAJ	SINGH	18000	DS, ML
1	201	RAHUL	KUMAR	18500	B.Sc
2	202	AKANKSHA	RANI	16000	PYRHON,ML
3	203	MONU	THAKUR	15000	JAVA
4	204	AKBAR	KHAN	12000	M.COM
5	205	KAUSHAL	KUMAR	15500	DIPLOMA
6	206	ABHISHAK	SINGH	20500	IIT
7	207	RAJ KISHOR	KUMAR	17000	B-TECH
8	208	RAKESH	ENG.	19500	DIPLOMA
9	209	UMESH	BHAI	17500	B-TECH

In [28]:

```
df1.pop('LAST NAME') # delete Column
df1
```

Out[28]:

	ID	NAME	FEES	COURSE
0	200	RISHIRAJ	18000	DS, ML
1	201	RAHUL	18500	B.Sc
2	202	AKANKSHA	16000	PYRHON,ML
3	203	MONU	15000	JAVA
4	204	AKBAR	12000	M.COM
5	205	KAUSHAL	15500	DIPLOMA
6	206	ABHISHAK	20500	IIT
7	207	RAJ KISHOR	17000	B-TECH
8	208	RAKESH	19500	DIPLOMA
9	209	UMESH	17500	B-TECH

In [29]:

```
df1.drop(6) # Delete row using Index
```

Out[29]:

	ID	NAME	FEES	COURSE
0	200	RISHIRAJ	18000	DS, ML
1	201	RAHUL	18500	B.Sc
2	202	AKANKSHA	16000	PYRHON,ML
3	203	MONU	15000	JAVA
4	204	AKBAR	12000	M.COM
5	205	KAUSHAL	15500	DIPLOMA
7	207	RAJ KISHOR	17000	B-TECH
8	208	RAKESH	19500	DIPLOMA
9	209	UMESH	17500	B-TECH

In [30]:

```
df1[(df1['ID']>204)&(df1['ID']<208)] # pick Data using LIMITE
```

Out[30]:

	ID	NAME	FEES	COURSE
5	205	KAUSHAL	15500	DIPLOMA
6	206	ABHISHAK	20500	IIT
7	207	RAJ KISHOR	17000	B-TECH

In [31]:

```
df1[df1['ID']==205]
```

Out[31]:

	ID	NAME	FEES	COURSE
5	205	KAUSHAL	15500	DIPLOMA

In [32]:

```
df1[df1['FEES']==15000]
```

Out[32]:

	ID	NAME	FEES	COURSE
3	203	MONU	15000	JAVA

In [33]:

```
df1.T # exachange row in column
```

Out[33]:

	0	1	2	3	4	5	6	7	
ID	200	201	202	203	204	205	206	207	
NAME	RISHIRAJ	RAHUL	AKANKSHA	MONU	AKBAR	KAUSHAL	ABHISHAK	RAJ KISHOR	RA
FEES	18000	18500	16000	15000	12000	15500	20500	17000	
COURSE	DS, ML	B.Sc	PYRHON,ML	JAVA	M.COM	DIPLOMA	IIT	B-TECH	DIP

In [34]:

```
df1.head(6)
```

Out[34]:

	ID	NAME	FEES	COURSE
0	200	RISHIRAJ	18000	DS, ML
1	201	RAHUL	18500	B.Sc
2	202	AKANKSHA	16000	PYRHON,ML
3	203	MONU	15000	JAVA
4	204	AKBAR	12000	M.COM
5	205	KAUSHAL	15500	DIPLOMA

In [35]:

```
df1.tail(4)
```

Out[35]:

	ID	NAME	FEES	COURSE
6	206	ABHISHAK	20500	IIT
7	207	RAJ KISHOR	17000	B-TECH
8	208	RAKESH	19500	DIPLOMA
9	209	UMESH	17500	B-TECH

In [ ]:

```

```