

## PANDAS UNDERSTANDING

pd.Series()

```
import pandas as pd  
h=('AA','09-05-2020',100,50.2)  
s=pd.Series(h)  
print(s)
```

o/p:

```
0      AA  
1  09-05-2020  
2      100  
3     50.2  
dtype: object
```

or,

```
import pandas as pd  
r=['sachin','virat',10,18]  
s=pd.Series(r)  
print(s)
```

o/p:

```
0    sachin  
1    virat  
2         10  
3         18  
dtype: object  
or,
```

```
import pandas as pd

d={'barsman1':'sachin','jarsy no. of
batsman1':10,'batsman2':'virat','jarsy no batsman2':18}

s=pd.Series(d)

print(s)
```

o/p:

```
barsman1          sachin
jarsy no. of batsman1    10
batsman2          virat
jarsy no batsman2      18
dtype: object
```

or,

```
import pandas as pd

f=['sachin',10,'virat',18]

s=pd.Series(f,index=['batsman1','jarsy no of
batsman1','batsman2','jarsy no of batsman 2'])

print(s)
```

o/p:

```
batsman1          sachin
jarsy no of batsman1    10
batsman2          virat
jarsy no of batsman 2    18
dtype: object
```

or,

```

import pandas as pd

f=['sachin',10,'virat',18]

s=pd.Series(f,index=['batsman1','jarsy no of
batsman1','batsman2','jarsy no of batsman 2'])

print(s)

print('\n')

print('jarsy no of batsman1-',s['jarsy no of batsman1'])

print('\n')

print('name of batsman2-',s[2])

print('\n')

print(s[['batsman1','jarsy no of batsman1','batsman2']])

```

o/p:

```

batsman1          sachin
jarsy no of batsman1      10
batsman2          virat
jarsy no of batsman 2     18
dtype: object

```

jarsy no of batsman1- 10

name of batsman2- virat

```

batsman1          sachin
jarsy no of batsman1      10
batsman2          virat
dtype: object

```

pd.DataFrame()

```

import pandas as pd

```

```
dict={'name':['shubha','sujit','babu'],'date of birth':['13-06-1994','12-06-1991','11-06-1994'],
      'shares':[50,30,60],'price':[16.5,17.65,19.8]}
h=pd.DataFrame(dict)
print(h)
```

o/p:

	name	date of birth	shares	price
0	shubha	13-06-1994	50	16.50
1	sujit	12-06-1991	30	17.65
2	babu	11-06-1994	60	19.80

Or,

```
import pandas as pd

player_dict={'opener':['RS','SD','LR'],'base
batsman':['VK','SI','MP'],'allrounder':['HP','RJ','VS'],'spinner':['YC
','KJ','KP']}

h=pd.DataFrame(player_dict)

print(h)

h['pacer']=['MS','JB','VNK']

print('\n')

print(h)

print('\n')

h.index=['one','two','three']

print(h)
```

o/p:

	opener	base	batsman	allrounder	spinner
0	RS		VK	HP	YC
1	SD		SI	RJ	KJ
2	LR		MP	VS	KP

	opener	base	batsman	allrounder	spinner	pacer
0	RS		VK	HP	YC	MS
1	SD		SI	RJ	KJ	JB
2	LR		MP	VS	KP	VNK

	opener	base	batsman	allrounder	spinner	pacer
one	RS		VK	HP	YC	MS
two	SD		SI	RJ	KJ	JB
three	LR		MP	VS	KP	VNK

or,

```
import pandas as pd
```

```
player_dict={'opener':pd.Series(['RS','SD','LR'],index=['one','two','three'])
```

```
, 'base
batsman':pd.Series(['VK','SI','MP'],index=['one','two','three'])
```

```
, 'allrounder':pd.Series(['HP','RJ','VS'],index=['one','two','three']),
```

```
'spinner':pd.Series(['YC','KJ','KP'],index=['one','two','three'])}
```

```
h=pd.DataFrame(player_dict)
```

```
print(h)
```

```
h['pacer']=['MS','JB','VNK']
```

```
print('\n')
```

```
print(h)
```

o/p:

	opener	base	batsman	allrounder	spinner
one	RS		VK	HP	YC
two	SD		SI	RJ	KJ
three	LR		MP	VS	KP

	opener	base	batsman	allrounder	spinner	pacer
one	RS		VK	HP	YC	MS
two	SD		SI	RJ	KJ	JB
three	LR		MP	VS	KP	VNK

or,

```
import pandas as pd

player_dict={'opener':['RS','SD','LR'],'base
batsman':['VK','SI','MP'],'allrounder':['HP','RJ','VS'],'spinner':['YC
','KJ','KP']}

h=pd.DataFrame(player_dict)

print(h)

h['pacer']=['MS','JB','VNK']

print('\n')

print(h)

print('\n')

h=h.set_index(['base batsman'])

print(h)
```

o/p:

	opener	base	batsman	allrounder	spinner
0	RS		VK	HP	YC
1	SD		SI	RJ	KJ
2	LR		MP	VS	KP

	opener	base	batsman	allrounder	spinner	pacers
0	RS		VK	HP	YC	MS
1	SD		SI	RJ	KJ	JB
2	LR		MP	VS	KP	VNK

	base	batsman	opener	allrounder	spinner	pacers
VK			RS	HP	YC	MS
SI			SD	RJ	KJ	JB
MP			LR	VS	KP	VNK

#computing different row and column

import pandas as pd

```
dict={'name':['shubha','sujit','babu'],'date of
birth':['13.06','02.11','05.06'],'share':[30,50,100],'price':[10,15,16
],
```

```
    'car':['BMW','TVS','APACHI']}]
```

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

```
df=df.set_index('name')
```

```
print(df)
```

```
print('\n')
```

```
print(df.loc['shubha'])
```

```
print('\n')
```

```
print(df.loc['shubha','date of birth'])
```

```
print('\n')
```

or,

	date of birth	share	price	car
name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

```

date of birth    13.06
share            30
price           10
car             BMW
Name: shubha, dtype: object

```

13.06

Or,

```

import pandas as pd

dict={'name':['shubha','sujit','babu'],'date of
birth':['13.06','02.11','05.06'],'share':[30,50,100],'price':[10,15,16
],
      'car':['BMW','TVS','APACHI']}

df=pd.DataFrame(dict)

df=df.set_index('name')

print(df)

print('\n')

print(df['share'])

print('\n')

print(df.loc['shubha'])

print('\n')

print(df.loc['shubha','date of birth'])

print('\n')

```



o/p:

	date of birth	share	price	car
name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

```
name
shubha    30
sujit     50
babu     100
Name: share, dtype: int64
```

```
date of birth    13.06
share           30
price           10
car             BMW
Name: shubha, dtype: object
```

13.06

Or,

```
import pandas as pd

dict={'name':['shubha','sujit','babu'],'date of
birth':['13.06','02.11','05.06'],'share':[30,50,100],'price':[10,15,16
],'car':['BMW','TVS','APACHI']}

df=pd.DataFrame(dict)

df=df.set_index('name')

print(df)

print('\n')

print(df.loc[:,'share'])
```

o/p:

	date of birth	share	price	car
--	---------------	-------	-------	-----

name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

```

name
shubha    30
sujit     50
babu     100
Name: share, dtype: int64

```

Or,

```

import pandas as pd

dict={'name':['shubha','sujit','babu'],'date of
birth':['13.06','02.11','05.06'],'share':[30,50,100],'price':[10,15,16],'car'
:['BMW','TVS','APACHI']}

df=pd.DataFrame(dict)

df=df.set_index('name')

print(df)

print('\n')

print(df.loc[:,'share'])

del df['car']

print('\n')

print(df)

```

o/p:

	date of birth	share	price	car
name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

```

name
shubha    30
sujit     50
babu     100
Name: share, dtype: int64

```

	date of birth	share	price
name			
shubha	13.06	30	10
sujit	02.11	50	15
babu	05.06	100	16

or,

```
import pandas as pd

dict={'name':['shubha','sujit','babu'],'date of
birth':['13.06','02.11','05.06'],'share':[30,50,100],'price':[10,15,16],'car'
:['BMW','TVS','APACHI']}
```

	date of birth	share	price	car
name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

```
df=pd.DataFrame(dict)

df=df.set_index('name')

print(df)

print('\n')

print(df.loc[:, 'share'])

del df['car']

print('\n')

print(df)

df=df.drop('share',axis=1)

print('\n')

print(df)

df=df.drop('shubha',axis=0)

print('\n')

print(df)
```

o/p:

	date of birth	share	price	car
name				
shubha	13.06	30	10	BMW
sujit	02.11	50	15	TVS
babu	05.06	100	16	APACHI

name	
shubha	30
sujit	50
babu	100

Name: share, dtype: int64

	date of birth	share	price
name			
shubha	13.06	30	10
sujit	02.11	50	15
babu	05.06	100	16

	date of birth	price
name		
shubha	13.06	10
sujit	02.11	15
babu	05.06	16

	date of birth	price
name		
sujit	02.11	15
babu	05.06	16

read.file:

```
import pandas as pd
```

```
import os
```

```
casts = pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv', index_col=None)
```

```
casts.head()
```

o/p:

title	year	name	type	character	n
-------	------	------	------	-----------	---

0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$hutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	NaN
4	Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	NaN

Or,

```
import pandas as pd

import os

animalss=pd.read_csv(r'C:\Users\Shubhamay\Documents\animals.csv',index_col=None)

print(animalss.head())
```

o/p:

	animal	item	cost
0	cat	2	900
1	cow	4	1000
2	goat	3	600
3	dog	1	500
4	tiger	4	1500

Or,

```
import pandas as pd

import os

title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)

print(title.tail())
```

o/p:

	title	year
49995	Rebel	1970
49996	Suzanne	1996
49997	Bomba	2013
49998	Aao Jao Ghar Tumhara	1984
49999	Mrs. Munck	1995

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
pd.set_option('max_rows',8,'max_columns',8)
print(title)
```

o/p:

	title	year
0	The Rising Son	1990
1	The Thousand Plane Raid	1969
2	Crucea de piatra	1993
3	Country	2000
...	...	...
49996	Suzanne	1996
49997	Bomba	2013
49998	Aao Jao Ghar Tumhara	1984
49999	Mrs. Munck	1995

[50000 rows x 2 columns]

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
len(title)
```

o/p:

50000

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
print(title.head(7))
```

o/p:

	title	year
0	The Rising Son	1990
1	The Thousand Plane Raid	1969
2	Crucea de piatra	1993
3	Country	2000
4	Gaiking II	2011
5	Medusa (IV)	2015
6	The Fresh Air Will Do You Good	2008

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
print(title.tail(7))
```

o/p:

	title	year
49993	Corruption.Gov	2010
49994	Lille Fridolf blir morfar	1957
49995	Rebel	1970
49996	Suzanne	1996
49997	Bomba	2013
49998	Aao Jao Ghar Tumhara	1984
49999	Mrs. Munck	1995

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
title.iloc[0,0]
```

o/p:

'The Rising Son'

Or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=movie['title']
print(t.head())
```

o/p:

0	The Rising Son
1	The Thousand Plane Raid
2	Crucea de piatra
3	Country
4	Gaiking II

Name: title, dtype: object

Or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
print(movie.head())
print('\n')
print(movie.iloc[0])
print('\n')
print(movie.iloc[:,0].head())
```

o/p:

	title	year
0	The Rising Son	1990
1	The Thousand Plane Raid	1969
2	Crucea de piatra	1993
3	Country	2000
4	Gaiking II	2011

```
title    The Rising Son
year      1990
Name: 0, dtype: object
```

```
0          The Rising Son
1  The Thousand Plane Raid
2      Crucea de piatra
3          Country
4      Gaiking II
Name: title, dtype: object
```

Or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
after_85=movie[movie['year']>1985]
print(after_85.head())
print('\n')
print(len(after_85))
```

o/p:

	title	year
0	The Rising Son	1990
1	The Thousand Plane Raid	1969
2	Crucea de piatra	1993
3	Country	2000
4	Gaiking II	2011



	title	year
0	The Rising Son	1990
2	Crucea de piatra	1993
3	Country	2000
4	Gaiking II	2011
5	Medusa (IV)	2015

29814

or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
m=movie
movie90=m[(m['year']>1990)& (m['year']<2000)]
print(movie90.head())
print('\n')
print("no of 90's movie:",len(movie90))
```

o/p:

	title	year
2	Crucea de piatra	1993
12	Poka Makorer Ghar Bosoti	1996
19	Maa Durga Shakti	1999
24	Conflict of Interest	1993
32	Der Brocken	1992

no of 90's movie: 4288

sort by value and sort by index:

or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
m=movie
macbeth=m[m['title']=='Macbeth']
print(macbeth.head())
print('\n')
print("no of macbeth",len(macbeth))
```

o/p:

title	year
-------	------

```
4226    Macbeth    1913
9322    Macbeth    2006
11722   Macbeth    2013
17166   Macbeth    1997
25847   Macbeth    1998
```

no of macbeth 5

or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
m=movie
macbeth=m[m['title']=='Macbeth'].sort_index()
print(macbeth.head())
print('\n')
print("no of macbeth",len(macbeth))
```

o/p:

```
      title  year
4226  Macbeth  1913
9322  Macbeth  2006
11722  Macbeth  2013
17166  Macbeth  1997
25847  Macbeth  1998
```

no of macbeth 5

or,

```
import pandas as pd
import os
movie=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
m=movie
macbeth=m[m['title']=='Macbeth'].sort_values('year')
print(macbeth.head())
print('\n')
print("no of macbeth",len(macbeth))
```

o/p:

```
      title  year
4226  Macbeth  1913
17166  Macbeth  1997
25847  Macbeth  1998
9322   Macbeth  2006
11722  Macbeth  2013
```

no of macbeth 5

Or,

```
import pandas as pd
```

```
import os
```

```
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```
print(cast.iloc[3:5])
```

o/p:

	title	year	name	type	character	n
3	Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	NaN
4	Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	NaN

Or,

```
import pandas as pd
```

```
import os
```

```
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```
c=cast
```

```
print(c['n'].isnull().head())
```

or,

o/p:

```
0    False
1    False
2    False
3     True
4     True
Name: n, dtype: bool
```

Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast

print(c['n'].notnull().head())
```

o/p:

```
0     True
1     True
2     True
3    False
4    False
Name: n, dtype: bool
```

Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```

c=cast

del c['name']

c_fill=c[c['n'].isnull()].fillna('NA')

print(c_fill.head())

```

o/p:

	title	year	type	character	n
3	Secret in Their Eyes	2015	actor	2002 Dodger Fan	NA
4	Steve Jobs	2015	actor	1988 Opera House Patron	NA
5	Straight Outta Compton	2015	actor	Club Patron	NA
6	Straight Outta Compton	2015	actor	Dopeman	NA
7	For Thy Love 2	2009	actor	Thug 1	NA

Or,

```

import pandas as pd

import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

print(c.head(5))

print('\n')

print(c.ffill().head(5))

```

o/p:

	title	year	name	type	character	n
0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$shutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$shutter	actor	2002 Dodger Fan	NaN
4	Steve Jobs	2015	\$shutter	actor	1988 Opera House Patron	NaN

	title	year	name	type	character	n
0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$hutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	10.0
4	Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	10.0

Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

print('\n')

print(c.bfill().head(5))
```

o/p:

	title	year	name	type	character	n
0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$hutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	9.0
4	Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	9.0

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=title
print(t[t['title']=='Maa Durga Shakti'])
```

o/p:

	title	year
19	Maa Durga Shakti	1999

Or,

```
import pandas as pd
```

```
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=title
print(t[t['title']=='Maa'])
```

o/p:

```
      title  year
38880   Maa  1968
```

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=title
print(t[t['title'].str.startswith('Maa')].head(3))
```

o/p:

```
      title  year
19   Maa Durga Shakti  1999
334      Maarek hob  2004
3046   Maa Aur Mamta  1970
```

Or,

```
import pandas as pd
import os
titless=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=titless
t['title'].value_counts().head(5)
```

o/p;

```
The Circle      6
Macbeth         5
Othello         5
Boy             4
Strings         4
Name: title, dtype: int64
```

Or,

```
import pandas as pd
import os
titless=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=None)
t=titless
t['year'].value_counts().head(5) #so the no. of concurrence occurrence
```

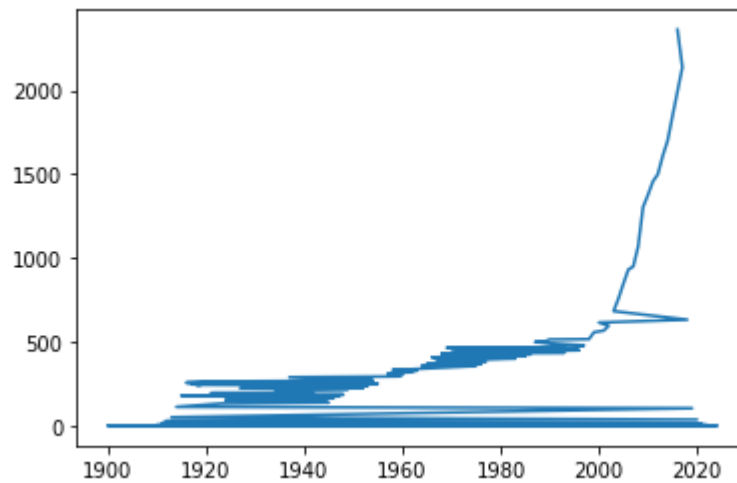
o/p:

```
2016      2363
2017      2138
2015      1849
2014      1701
2013      1609
Name: year, dtype: int64
```

Or,

```
import pandas as pd
from matplotlib import pyplot as plt
import os
titless=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=N
one)
t=titless
p=t['year'].value_counts()
p.plot()
plt.show()
```

o/p:

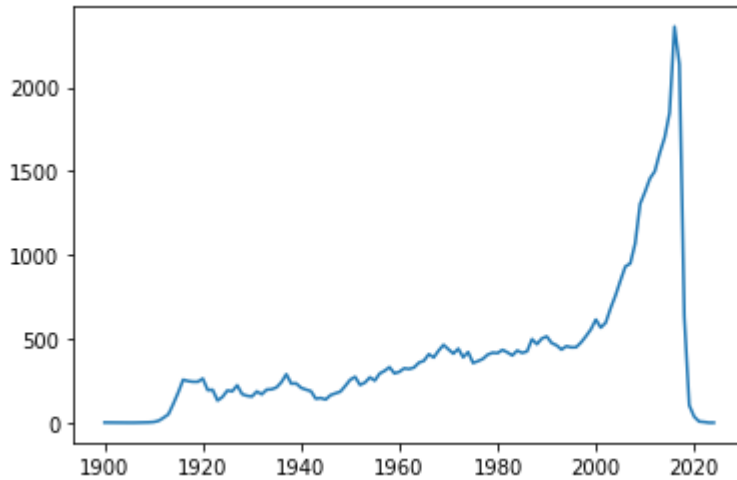


Or,

```
import pandas as pd
from matplotlib import pyplot as plt
import os
titless=pd.read_csv(r'C:\Users\Shubhamay\Documents\titles.csv',index_col=N
one)
t=titless
p=t['year'].value_counts()
p.sort_index().plot()
plt.show()
```

o/p:





Or,

```
import pandas as pd

import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

print(c['year'].value_counts().head(3))
```

o/p:

```
2016      3841
2015      3439
2014      3309
Name: year, dtype: int64
```

Or,

```
import pandas as pd

import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

print(c.groupby(['year']).size().head(3))
```

o/p:

```
year
1912      3
1913     10
1914     61
dtype: int64
```

or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

cg=c[c['year']==2016]

cg['year'].value_counts()
```

o/p:

```
2016      3841
Name: year, dtype: int64
```

Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

cg=c[c['year']==2016]

len(cg)
```

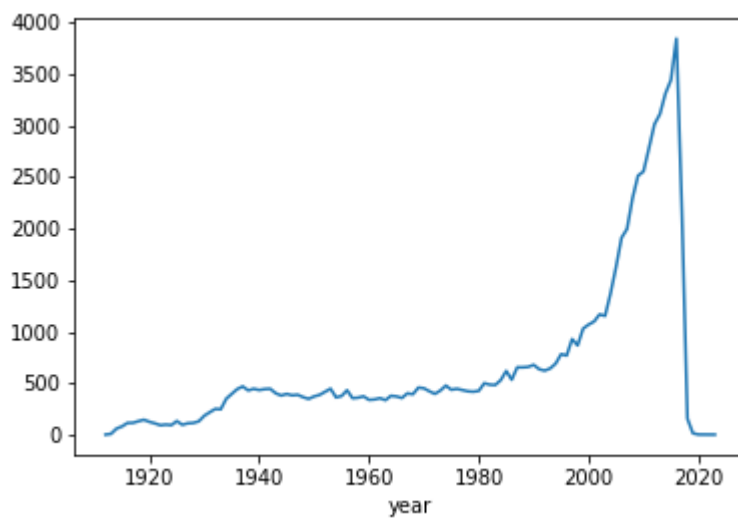
o/p:

3841

Or,

```
import pandas as pd
import os
from matplotlib import pyplot as plt
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast
ck=c.groupby(['year']).size()
ck.plot()
plt.show()
```

o/p;



Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

cf=c[c['name']=='Aaron Abrams']

print(cf.groupby(['year']).size().head())
```

o/p:

```
year
2003    2
2004    2
2005    2
2006    1
2007    2
dtype: int64
```

or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

cf=c[c['name']=='Aaron Abrams']

print(cf.groupby(['year','title']).size())
```

o/p:

```
year  title
2003  The In-Laws
```

	The Visual Bible: The Gospel of John	1
2004	Resident Evil: Apocalypse	1
	Siblings	1
2005	Cinderella Man	1
	Sabah	1
2006	Zoom	1
2007	Firehouse Dog	1
	Young People Fucking	1
2008	Flash of Genius	1
2009	Amelia	1
	At Home by Myself... with You	1
2011	388 Arletta Avenue	1
	Jesus Henry Christ	2
	Take This Waltz	1
	The Chicago 8	1
2013	It Was You Charlie	1
2015	Closet Monster	1
	Regression	1
2017	#FromJennifer	1
	The Go-Getters	1
2018	Code 8	1

dtype: int64

or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast
print(c.groupby(['year']).n.max().head())
```

o/p:

```
year
1912      6.0
1913     14.0
1914     39.0
1915     14.0
1916     35.0
Name: n, dtype: float64
```

Or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast
print(c.groupby(['year']).n.min().head())
```

o/p:

```
year
1912      6.0
```

```
1913    1.0
1914    1.0
1915    1.0
1916    1.0
Name: n, dtype: float64
```

Or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast
print(c.groupby(['year']).n.mean().head())
```

o/p;

```
year
1912    6.000000
1913    4.142857
1914    7.085106
1915    4.236111
1916    5.037736
Name: n, dtype: float64
```

Or,

```
import pandas as pd

import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

decade=(c['year']//10)*10 #1981//10=198,198*10=1980

print(c.groupby(decade).size().head())
```

o/p:

```
year
1910    669
1920   1121
1930   3448
1940   3997
1950   3892
dtype: int64
```

or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast

decade=(c['year']//10)*10 #1981//10=198,198*10=1980

print(c.groupby(decade).n.size().head())
```

o/p:

```
year
1910      669
1920     1121
1930     3448
1940     3997
1950     3892
Name: n, dtype: int64
```

Or,

```
import pandas as pd
import os

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast #1981//10=198,198*10=1980

print(c.groupby((c['year']//10)*10).size().head(4))
```

o/p:

```
year
1910      669
```

```
1920    1121
1930    3448
1940    3997
dtype: int64
```

or,

```
import pandas as pd
```

```
import os
```

```
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```
c=cast #1981//10=198,198*10=1980
```

```
print(c.groupby(['type',(c['year']//10)*10]).size())
```

o/p:

```
type      year
actor    1910      384
         1920      710
         1930     2628
         1940     3014
         1950     2877
         1960     2775
         1970     3044
         1980     3565
         1990     5108
         2000    10368
         2010    15523
         2020         4
actress   1910      285
         1920      411
         1930      820
         1940      983
         1950     1015
         1960      968
         1970     1299
         1980     1989
         1990     2544
         2000     5831
         2010     8853
         2020         3
dtype: int64
```

or,



```

import pandas as pd

import numpy as np

import os

import seaborn as sns

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast #1981//10=198,198*10=1980

c_decade=c.groupby(['type',c['year']//10*10]).size()

df=c_decade.unstack()#change row into column

del df[2020]

print(df)

```

o/p:

year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
type											
actor	384	710	2628	3014	2877	2775	3044	3565	5108	10368	15523
actress	285	411	820	983	1015	968	1299	1989	2544	5831	8853

or,

```

import pandas as pd

import os

from matplotlib import pyplot as plt

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast #1981//10=198,198*10=1980

c_decade=c.groupby(['type',c['year']//10*10]).size()

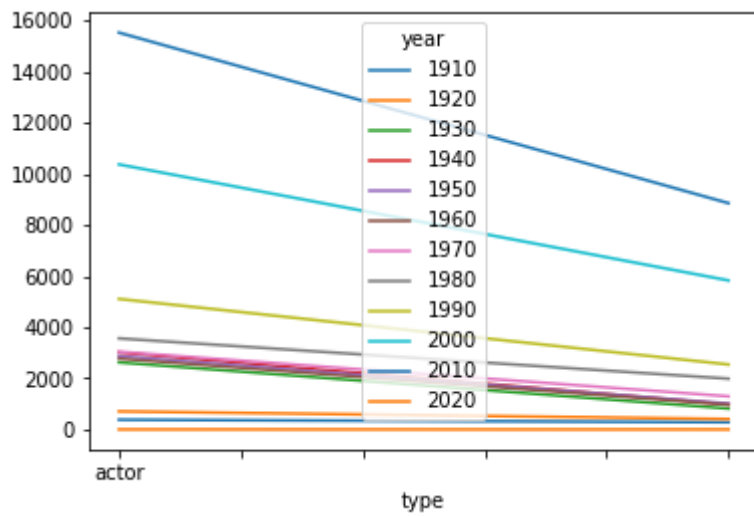
df=c_decade.unstack()

df.plot()

plt.show()

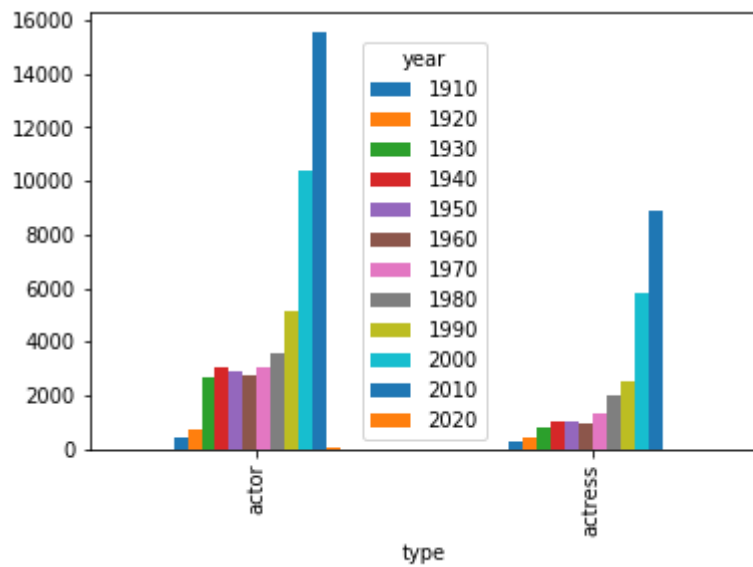
```

o/p,



Or,

```
import pandas as pd
import os
from matplotlib import pyplot as plt
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast #1981//10=198,198*10=1980
c_decade=c.groupby(['type',c['year']//10*10]).size()
df=c_decade.unstack()
df.plot(kind='bar')
plt.show()
```



Or,

```
import pandas as pd
import os
from matplotlib import pyplot as plt

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c=cast #1981//10=198,198*10=1980

c_decade=c.groupby(['type',c['year']//10*10]).size()

df=c_decade.unstack(0)

print(df)
```

o/p:

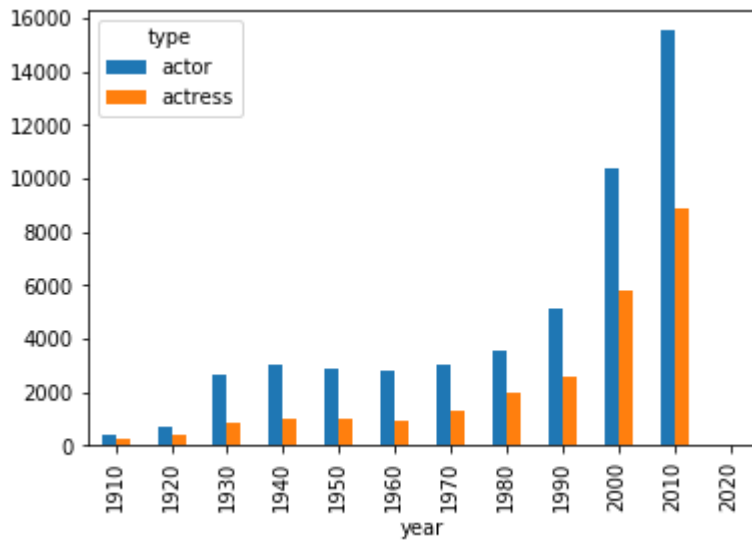
type	actor	actress
year		
1910	384	285
1920	710	411
1930	2628	820

1940	3014	983
1950	2877	1015
1960	2775	968
1970	3044	1299
1980	3565	1989
1990	5108	2544
2000	10368	5831
2010	15523	8853
2020	4	3

Or,

```
import pandas as pd
import os
from matplotlib import pyplot as plt
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c=cast #1981//10=198,198*10=1980
c_decade=c.groupby(['type',c['year']//10*10]).size()
df=c_decade.unstack(0)
df.plot(kind='bar')
plt.show()
```

o/p:



Or,

```
import pandas as pd

import os

release=pd.read_csv(r'C:\Users\Shubhamay\Documents\release_dates.csv',index_c
ol=None)

print(release.head())
```

o/p:

	title	year	country	date
0	#73, Shaanthi Nivaasa	2007	India	2007-06-15
1	#Beings	2015	Romania	2015-01-29
2	#Declimax	2018	Netherlands	2018-01-21
3	#Ewankosau saranghaeyo	2015	Philippines	2015-01-21
4	#Horror	2015	USA	2015-11-20

Or,

```
import pandas as pd

import os
```

```

release=pd.read_csv(r'C:\Users\Shubhamay\Documents\release_dates.csv',index_c
ol=None)

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c_amelia=cast[cast['title']=='Amelia']

print('cast amelia:', '\n', c_amelia.head(7))

print('\n')

r_amelia=release[release['title']=='Amelia']

print('release amelia:', '\n', r_amelia.head(7))

print('\n')

df=c_amelia.merge(r_amelia)

del df['title']

print(df)

```

op:

```

cast amelia:
      title  year      name  type  character    n
5767  Amelia  2009  Aaron Abrams  actor  Slim Gordon  8.0
23319  Amelia  2009  Jeremy Akerman  actor      Sheriff 19.0

```

```

release amelia:
      title  year  country      date
20543  Amelia  1966   Mexico  1966-03-10
20544  Amelia  2009   Canada  2009-10-23
20545  Amelia  2009     USA   2009-10-23
20546  Amelia  2009  Australia  2009-11-12
20547  Amelia  2009  Singapore  2009-11-12
20548  Amelia  2009   Ireland  2009-11-13
20549  Amelia  2009   Mexico  2009-11-13

```

```

      year      name  type  character    n  country      date
0   2009  Aaron Abrams  actor  Slim Gordon  8.0   Canada  2009-10-23
1   2009  Aaron Abrams  actor  Slim Gordon  8.0     USA  2009-10-23
2   2009  Aaron Abrams  actor  Slim Gordon  8.0  Australia  2009-11-12
3   2009  Aaron Abrams  actor  Slim Gordon  8.0  Singapore  2009-11-12
4   2009  Aaron Abrams  actor  Slim Gordon  8.0   Ireland  2009-11-13
5   2009  Aaron Abrams  actor  Slim Gordon  8.0   Mexico  2009-11-13
6   2009  Aaron Abrams  actor  Slim Gordon  8.0     UK  2009-11-13
7   2009  Aaron Abrams  actor  Slim Gordon  8.0   Spain  2009-11-20
8   2009  Aaron Abrams  actor  Slim Gordon  8.0  Philippines  2009-12-02

```

9	2009	Aaron Abrams	actor	Slim Gordon	8.0	Italy	2009-12-23	
10	2009	Aaron Abrams	actor	Slim Gordon	8.0	Sweden	2010-01-01	
11	2009	Aaron Abrams	actor	Slim Gordon	8.0	Turkey	2010-01-08	
12	2009	Aaron Abrams	actor	Slim Gordon	8.0	Belarus	2010-01-14	
13	2009	Aaron Abrams	actor	Slim Gordon	8.0	Kazakhstan	2010-01-14	
14	2009	Aaron Abrams	actor	Slim Gordon	8.0	Russia	2010-01-14	
15	2009	Aaron Abrams	actor	Slim Gordon	8.0	Greece	2010-01-21	
16	2009	Aaron Abrams	actor	Slim Gordon	8.0	Poland	2010-02-12	
17	2009	Aaron Abrams	actor	Slim Gordon	8.0	Belgium	2010-03-03	
18	2009	Aaron Abrams	actor	Slim Gordon	8.0	India	2010-03-12	
19	2009	Aaron Abrams	actor	Slim Gordon	8.0	Netherlands	2010-03-25	
20	2009	Aaron Abrams	actor	Slim Gordon	8.0	Brazil	2010-03-26	
21	2009	Aaron Abrams	actor	Slim Gordon	8.0	France	2010-04-14	
22	2009	Aaron Abrams	actor	Slim Gordon	8.0	Germany	2010-06-17	
23	2009	Aaron Abrams	actor	Slim Gordon	8.0	Japan	2010-11-27	
24	2009	Jeremy Akerman	actor	Sheriff	19.0	Canada	2009-10-23	
25	2009	Jeremy Akerman	actor	Sheriff	19.0	USA	2009-10-23	
26	2009	Jeremy Akerman	actor	Sheriff	19.0	Australia	2009-11-12	
27	2009	Jeremy Akerman	actor	Sheriff	19.0	Singapore	2009-11-12	
28	2009	Jeremy Akerman	actor	Sheriff	19.0	Ireland	2009-11-13	
29	2009	Jeremy Akerman	actor	Sheriff	19.0	Mexico	2009-11-13	
30	2009	Jeremy Akerman	actor	Sheriff	19.0	UK	2009-11-13	
31	2009	Jeremy Akerman	actor	Sheriff	19.0	Spain	2009-11-20	
32	2009	Jeremy Akerman	actor	Sheriff	19.0	Philippines	2009-12-02	
33	2009	Jeremy Akerman	actor	Sheriff	19.0	Italy	2009-12-23	
34	2009	Jeremy Akerman	actor	Sheriff	19.0	Sweden	2010-01-01	
35	2009	Jeremy Akerman	actor	Sheriff	19.0	Turkey	2010-01-08	
36	2009	Jeremy Akerman	actor	Sheriff	19.0	Belarus	2010-01-14	
37	2009	Jeremy Akerman	actor	Sheriff	19.0	Kazakhstan	2010-01-14	
38	2009	Jeremy Akerman	actor	Sheriff	19.0	Russia	2010-01-14	
39	2009	Jeremy Akerman	actor	Sheriff	19.0	Greece	2010-01-21	
40	2009	Jeremy Akerman	actor	Sheriff	19.0	Poland	2010-02-12	
41	2009	Jeremy Akerman	actor	Sheriff	19.0	Belgium	2010-03-03	
42	2009	Jeremy Akerman	actor	Sheriff	19.0	India	2010-03-12	
43	2009	Jeremy Akerman	actor	Sheriff	19.0	Netherlands	2010-03-25	
44	2009	Jeremy Akerman	actor	Sheriff	19.0	Brazil	2010-03-26	
45	2009	Jeremy Akerman	actor	Sheriff	19.0	France	2010-04-14	
46	2009	Jeremy Akerman	actor	Sheriff	19.0	Germany	2010-06-17	
47	2009	Jeremy Akerman	actor	Sheriff	19.0	Japan	2010-11-27	
4	Amelia	2009	Aaron Abrams	actor	Slim Gordon	8.0	Ireland	2009-11-13

Or,

```
import pandas as pd
```

```
import os
```

```
release=pd.read_csv(r'C:\Users\Shubhamay\Documents\release_dates.csv',index_col=None)
```

```
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```

c_amelia=cast[cast['title']=='Amelia']

print('cast amelia:', '\n', c_amelia.head(2))

print('\n')

print(c_amelia.columns)

print('\n')

r_amelia=release[release['title']=='Amelia']

print('release amelia:', '\n', r_amelia.head(2))

print('\n')

print(r_amelia.columns)

print('\n')

print(pd.merge(c_amelia, r_amelia).head(2))

```

o/p:

cast amelia:

	title	year	name	type	character	n
5767	Amelia	2009	Aaron Abrams	actor	Slim Gordon	8.0
23319	Amelia	2009	Jeremy Akerman	actor	Sheriff	19.0

```
Index(['title', 'year', 'name', 'type', 'character', 'n'], dtype='object')
```

release amelia:

	title	year	country	date
20543	Amelia	1966	Mexico	1966-03-10
20544	Amelia	2009	Canada	2009-10-23

```
Index(['title', 'year', 'country', 'date'], dtype='object')
```

	title	year	name	type	character	n	country	date
0	Amelia	2009	Aaron Abrams	actor	Slim Gordon	8.0	Canada	2009-10-23
1	Amelia	2009	Aaron Abrams	actor	Slim Gordon	8.0	USA	2009-10-23

Or,

```
import pandas as pd
```



```

import os
release=pd.read_csv(r'C:\Users\Shubhamay\Documents\release_dates.csv',index_col=None)
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
c_amelia=cast[cast['title']=='Amelia']
print('cast amelia:', '\n', c_amelia.head(2))
print('\n')
print(c_amelia.columns)
print('\n')
r_amelia=release[release['title']=='Amelia']
print('release amelia:', '\n', r_amelia.head(2))
print('\n')
print(r_amelia.columns)
print('\n')
print(pd.merge(r_amelia, c_amelia).head(2))

```

o/p:

```

cast amelia:
      title  year      name  type  character    n
5767  Amelia  2009  Aaron Abrams  actor  Slim Gordon   8.0
23319  Amelia  2009  Jeremy Akerman  actor      Sheriff  19.0

```

```
Index(['title', 'year', 'name', 'type', 'character', 'n'], dtype='object')
```

```

release amelia:
      title  year country      date
20543  Amelia  1966  Mexico  1966-03-10
20544  Amelia  2009  Canada  2009-10-23

```

```
Index(['title', 'year', 'country', 'date'], dtype='object')
```

```

      title  year country      date      name  type  character    n
0  Amelia  2009  Canada  2009-10-23  Aaron Abrams  actor  Slim Gordon   8.0
1  Amelia  2009  Canada  2009-10-23  Jeremy Akerman  actor      Sheriff  19.0

```

Or,

```

import pandas as pd

import os

release=pd.read_csv(r'C:\Users\Shubhamay\Documents\release_dates.csv',index_col=None)

cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)

c_amelia=cast[cast['title']=='Amelia']

```

```

#print('cast amelia:', '\n', c_amelia.head(2))

#print('\n')

#print(c_amelia.columns)

#print('\n')

r_amelia=release[release['title']=='Amelia']

#print('release amelia:', '\n', r_amelia.head(2))

#print('\n')

#print(r_amelia.columns)

#print('\n')

df=pd.merge(r_amelia,c_amelia)

#print(df)

print(c_amelia.shape)

print(r_amelia.shape)

print(df.shape)

```

o/p:

```

(2, 6)
(25, 4)
(48, 8)

```

or:

```

df1=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'temp':[32,34,35]})

print(df1)

print('\n')

df2=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'humidity':[17,19,18],}
)

print(df2)

```

```
print('\n')

print(pd.merge(df1,df2))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35

	city	humidity
0	delhi	17
1	kanpur	19
2	kolkata	18

	city	temp	humidity
0	delhi	32	17
1	kanpur	34	19
2	kolkata	35	18

Or,

```
df1=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'temp':[32,34,35]})

print(df1)

print('\n')

df2=pd.DataFrame({'city':['delhi','sanfransisco','kolkata'],'humidity':[17,19,18],
'rain':[180,200,250]})

print(df2)

print('\n')

print(pd.merge(df1,df2))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35

	city	humidity	rain
0	delhi	17	180
1	sanfransisco	19	200
2	kolkata	18	250

	city	temp	humidity	rain
0	delhi	32	17	180
1	kolkata	35	18	250

Or,

```
df1=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'temp':[32,34,35]})
print(df1)
print('\n')
df2=pd.DataFrame({'city':['delhi','sanfransisco','kolkata'],'humidity':[17,19,18],
'rain':[180,200,250]})
print(df2)
print('\n')
print(pd.merge(df1,df2,how='outer'))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35

	city	humidity	rain
0	delhi	17	180
1	sanfransisco	19	200
2	kolkata	18	250

	city	temp	humidity	rain
0	delhi	32.0	17.0	180.0
1	kanpur	34.0	NaN	NaN

2	kolkata	35.0	18.0	250.0
3	sanfransisco	NaN	19.0	200.0

Or,

```
df1=pd.DataFrame({'city':['delhi','kanpur','kolkata','mumbai','hydrabad'],'temp':[32,34,35,20,22]})
```

```
print(df1)
```

```
print('\n')
```

```
df2=pd.DataFrame({'city':['delhi','london','sanfransisco','kolkata'],'humidity':[17,19,18,23],'rain':[180,200,250,260]})
```

```
print(df2)
```

```
print('\n')
```

```
print(pd.merge(df1,df2,how='outer'))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35
3	mumbai	20
4	hydrabad	22

	city	humidity	rain
0	delhi	17	180
1	london	19	200
2	sanfransisco	18	250
3	kolkata	23	260

	city	temp	humidity	rain
0	delhi	32.0	17.0	180.0
1	kanpur	34.0	NaN	NaN
2	kolkata	35.0	23.0	260.0
3	mumbai	20.0	NaN	NaN
4	hydrabad	22.0	NaN	NaN
5	london	NaN	19.0	200.0
6	sanfransisco	NaN	18.0	250.0

Or,

```
df1=pd.DataFrame({'city':['delhi','kanpur','kolkata','mumbai','hydrabad'],'temp':[32,34,35,20,22]})
```

```
print(df1)
```

```
print('\n')
```

```
df2=pd.DataFrame({'city':['delhi','london','sanfransisco','kolkata'],'humidity':[17,19,18,23],'rain':[180,200,250,260]})
```

```
print(df2)
```

```
print('\n')
```

```
print(pd.merge(df1,df2,how='left'))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35
3	mumbai	20
4	hydrabad	22

	city	humidity	rain
0	delhi	17	180
1	london	19	200
2	sanfransisco	18	250
3	kolkata	23	260

	city	temp	humidity	rain
0	delhi	32	17.0	180.0
1	kanpur	34	NaN	NaN
2	kolkata	35	23.0	260.0
3	mumbai	20	NaN	NaN
4	hydrabad	22	NaN	NaN

Or,

```

df1=pd.DataFrame({'city':['delhi','kanpur','kolkata','mumbai','hydrabad'],'temp':[32,34,35,20,22]})

print(df1)

print('\n')

df2=pd.DataFrame({'city':['delhi','london','sanfransisco','kolkata'],'humidity':[17,19,18,23],'rain':[180,200,250,260]})

print(df2)

print('\n')

print(pd.merge(df1,df2,how='right'))

```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35
3	mumbai	20
4	hydrabad	22

	city	humidity	rain
0	delhi	17	180
1	london	19	200
2	sanfransisco	18	250
3	kolkata	23	260

	city	temp	humidity	rain
0	delhi	32.0	17	180
1	kolkata	35.0	23	260
2	london	NaN	19	200
3	sanfransisco	NaN	18	250

Or,

```

df1=pd.DataFrame({'city':['delhi','kanpur','kolkata','mumbai'],'temp':[32,34,35,20]})

print(df1)

```

```

print('\n')

df2=pd.DataFrame({'city':['delhi','london','sanfransisco'],'humidity':[17,19,18], 'rain':[180,200,250]})

print(df2)

print('\n')

print(pd.merge(df1,df2,how='outer',indicator=True))

```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35
3	mumbai	20

	city	humidity	rain
0	delhi	17	180
1	london	19	200
2	sanfransisco	18	250

	city	temp	humidity	rain	_merge
0	delhi	32.0	17.0	180.0	both
1	kanpur	34.0	NaN	NaN	left_only
2	kolkata	35.0	NaN	NaN	left_only
3	mumbai	20.0	NaN	NaN	left_only
4	london	NaN	19.0	200.0	right_only
5	sanfransisco	NaN	18.0	250.0	right_only

Or,

```

df1=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'humidity':[32,34,35], 'rain':[150,160,190]})

print(df1)

print('\n')

df2=pd.DataFrame({'city':['delhi','sanfransisco','kolkata'],'humidity':[17,19,18], 'rain':[180,200,250]})

```



```

print(df2)

print('\n')

print(pd.merge(df1,df2,on=['city']))

```

o/p:

	city	humidity	rain
0	delhi	32	150
1	kanpur	34	160
2	kolkata	35	190

	city	humidity	rain
0	delhi	17	180
1	sanfransisco	19	200
2	kolkata	18	250

	city	humidity_x	rain_x	humidity_y	rain_y
0	delhi	32	150	17	180
1	kolkata	35	190	18	250

Or,

```

df1=pd.DataFrame({'city':['delhi','kanpur','kolkata'],'humidity':[32,34,35],
rain':[150,160,190]})

```

```

print(df1)

```

```

print('\n')

```

```

df2=pd.DataFrame({'city':['delhi','sanfransisco','kolkata'],'humidity':[17,19
,18], 'rain':[180,200,250]})

```

```

print(df2)

```

```

print('\n')

```

```

print(pd.merge(df1,df2,on=['city'],suffixes=['_left','_right']))

```

o/p:

	city	humidity	rain
0	delhi	32	150
1	kanpur	34	160
2	kolkata	35	190

	city	humidity	rain
0	delhi	17	180
1	sanfransisco	19	200
2	kolkata	18	250

	city	humidity_left	rain_left	humidity_right	rain_right
0	delhi	32	150	17	180
1	kolkata	35	190	18	250

```
import pandas as pd
df1=pd.DataFrame({'city':['delhi','kanpur','kolkata','mumbai'],'temp':[32,
34,35,20]})
print(df1)
print('\n')
df2=pd.DataFrame({'city':['delhi','london','sanfransisco'],'humidity':[17,
19,18],'rain':[180,200,250]})
print(df2)
print('\n')
print(pd.merge(df1,df2))
```

o/p:

	city	temp
0	delhi	32
1	kanpur	34
2	kolkata	35
3	mumbai	20

	city	humidity	rain
0	delhi	17	180
1	london	19	200
2	sanfransisco	18	250

	city	temp	humidity	rain
0	delhi	32	17	180

Indexing:

```
import pandas as pd
import os
```

```
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col='title')
print(cast.head())
```

o/p:

	year	name	type	character	n
title					
Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
Suuri illusioni	1985	Homo \$	actor	Guests	22.0
Battle of the Sexes	2017	\$hutter	actor	Bobby Riggs Fan	10.0
Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	NaN
Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	NaN

Or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col='title')
print(cast.head())
print(cast.loc['Macbeth'])
```

o/p:

	year	name	type	character	n
title					
Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
Suuri illusioni	1985	Homo \$	actor	Guests	22.0
Battle of the Sexes	2017	\$hutter	actor	Bobby Riggs Fan	10.0
Secret in Their Eyes	2015	\$hutter	actor	2002 Dodger Fan	NaN
Steve Jobs	2015	\$hutter	actor	1988 Opera House Patron	NaN

	year	name	type	character	n
title					
Macbeth	2015	Darren Adamson	actor	Soldier	NaN
Macbeth	1916	Spottiswoode Aitken	actor	Duncan	4.0
Macbeth	1948	Robert Alan	actor	Third Murderer	NaN
Macbeth	2016	John Albasing	actor	Doctor	NaN
Macbeth	1948	William Alland	actor	Second Murderer	18.0
Macbeth	1997	Stevie Allen	actor	Murderer	21.0
Macbeth	2014	Moyo Akand?	actress	Witch	NaN
Macbeth	1916	Mary Alden	actress	Lady Macduff	6.0

Or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col='name')
print(cast.head())
print(cast.loc['Macbeth'])
```

o/p:

	title	year	type	character	n
name					
Buffy #1	Closet Monster	2015	actor	Buffy 4	31.0
Homo \$	Suuri illusioni	1985	actor	Guests	22.0
\$hutter	Battle of the Sexes	2017	actor	Bobby Riggs Fan	10.0
\$hutter	Secret in Their Eyes	2015	actor	2002 Dodger Fan	NaN
\$hutter	Steve Jobs	2015	actor	1988 Opera House Patron	NaN

**KeyError** Traceback (most recent call last)

~\Anaconda3\lib\site-packages\pandas\core\indexes\base.py in get\_loc(self, key, method, tolerance)

```

2645         try:
-> 2646             return self._engine.get_loc(key)
2647         except KeyError:

```

pandas\\_libs\index.pyx in pandas.\_libs.index.IndexEngine.get\_loc()

pandas\\_libs\index.pyx in pandas.\_libs.index.IndexEngine.get\_loc()

pandas\\_libs\index.pyx in  
pandas.\_libs.index.IndexEngine.\_get\_loc\_duplicates()

pandas\\_libs\index.pyx in  
pandas.\_libs.index.IndexEngine.\_maybe\_get\_bool\_indexer()

**KeyError:** 'Macbeth'

Or,

```

import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col='name'
)
print(cast.head())
print(cast[cast['title']=='Macbeth'])

```

o/p:

	title	year	type	character	n
name					
Buffy #1	Closet Monster	2015	actor	Buffy 4	31.0
Homo \$	Suuri illusioni	1985	actor	Guests	22.0
\$hutter	Battle of the Sexes	2017	actor	Bobby Riggs Fan	10.0
\$hutter	Secret in Their Eyes	2015	actor	2002 Dodger Fan	NaN
\$hutter	Steve Jobs	2015	actor	1988 Opera House Patron	NaN
	title	year	type	character	n
name					

Darren Adamson	Macbeth	2015	actor	Soldier	NaN
Spottiswoode Aitken	Macbeth	1916	actor	Duncan	4.0
Robert Alan	Macbeth	1948	actor	Third Murderer	NaN
John Albasiny	Macbeth	2016	actor	Doctor	NaN
William Alland	Macbeth	1948	actor	Second Murderer	18.0
Stevie Allen	Macbeth	1997	actor	Murderer	21.0
Moyo Akand?	Macbeth	2014	actress	Witch	NaN
Mary Alden	Macbeth	1916	actress	Lady Macduff	6.0

```
%timeit
```

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
print(cast.head())
%timeit cast[cast['title']=='Macbeth']
```

o/p:

	title	year	name	type	character	n
0	Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
1	Suuri illusioni	1985	Homo \$	actor	Guests	22.0
2	Battle of the Sexes	2017	\$shutter	actor	Bobby Riggs Fan	10.0
3	Secret in Their Eyes	2015	\$shutter	actor	2002 Dodger Fan	NaN
4	Steve Jobs	2015	\$shutter	actor	1988 Opera House Patron	NaN

8.89 ms ± 23.6 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)

Or,

```
import pandas as pd
import os
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col='title')
print(cast.head())
%timeit cast.loc['Macbeth']
```

o/p:

	year	name	type	character	n
title					
Closet Monster	2015	Buffy #1	actor	Buffy 4	31.0
Suuri illusioni	1985	Homo \$	actor	Guests	22.0
Battle of the Sexes	2017	\$shutter	actor	Bobby Riggs Fan	10.0
Secret in Their Eyes	2015	\$shutter	actor	2002 Dodger Fan	NaN
Steve Jobs	2015	\$shutter	actor	1988 Opera House Patron	NaN

3.37 ms ± 80.9 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)

Or,

```
import pandas as pd
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
cm=cast.set_index(['title']).sort_index()
print(cm.loc['Macbeth'])
```

o/p:

	year	name	type	character	n
title					
Macbeth	2015	Darren Adamson	actor	Soldier	NaN
Macbeth	1948	Robert Alan	actor	Third Murderer	NaN
Macbeth	1948	William Alland	actor	Second Murderer	18.0
Macbeth	1916	Spottiswoode Aitken	actor	Duncan	4.0
Macbeth	1916	Mary Alden	actress	Lady Macduff	6.0
Macbeth	2014	Moyo Akand?	actress	Witch	NaN
Macbeth	1997	Stevie Allen	actor	Murderer	21.0
Macbeth	2016	John Albasiny	actor	Doctor	NaN

Or,

```
import pandas as pd
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
cm=cast.set_index(['title','n']).sort_index()
print(cm.loc['Macbeth'])
```

o/p:

	year	name	type	character	n
4.0	1916	Spottiswoode Aitken	actor	Duncan	
6.0	1916	Mary Alden	actress	Lady Macduff	
18.0	1948	William Alland	actor	Second Murderer	
21.0	1997	Stevie Allen	actor	Murderer	
NaN	2015	Darren Adamson	actor	Soldier	
NaN	1948	Robert Alan	actor	Third Murderer	
NaN	2016	John Albasiny	actor	Doctor	
NaN	2014	Moyo Akand?	actress	Witch	

Or,

```
import pandas as pd
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
cm=cast.set_index(['title','n']).sort_index()
print(cm.loc['Macbeth'].loc[4:18])#macbeth with n number 4-18
```

o/p:

	year	name	type	character	n
4.0	1916	Spottiswoode Aitken	actor	Duncan	
6.0	1916	Mary Alden	actress	Lady Macduff	
18.0	1948	William Alland	actor	Second Murderer	

Or,

```
import pandas as pd
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
```

```
cm=cast.set_index(['title','n']).sort_index()
print(cm.loc['Macbeth'].loc[4])#macbeth with n number 4
```

o/p:

```
year                1916
name      Spottiswoode Aitken
type                actor
character          Duncan
Name: 4.0, dtype: object
```

Or,

```
import pandas as pd
cast=pd.read_csv(r'C:\Users\Shubhamay\Documents\cast.csv',index_col=None)
cm=cast.set_index(['title','n']).sort_index()
print(cm.reset_index('n').head(2))
```

o/p:

	n	year	name	type	character
title					
#1 Serial Killer	17.0	2013	Michael Alton	actor	Detective Roberts
#DigitalLivesMatter	NaN	2016	Rashan Ali	actress	News Reporter

List in pandas:

```
import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
title[0:5]
```

o/p:

```
[OrderedDict([('title', 'The Rising Son'), ('year', '1990')]),
 OrderedDict([('title', 'The Thousand Plane Raid'), ('year', '1969')]),
 OrderedDict([('title', 'Crucea de piatra'), ('year', '1993')]),
 OrderedDict([('title', 'Country'), ('year', '2000')]),
 OrderedDict([('title', 'Gaiking II'), ('year', '2011')])]
```

Or,

```
import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
title[-5:]# last 5 row
```

o/p:

```
[OrderedDict([('title', 'Rebel'), ('year', '1970')]),
 OrderedDict([('title', 'Suzanne'), ('year', '1996')]),
```

```
OrderedDict([('title', 'Bomba'), ('year', '2013')]),
OrderedDict([('title', 'Aao Jao Ghar Tumhara'), ('year', '1984')]),
OrderedDict([('title', 'Mrs. Munck'), ('year', '1995')])]
```

Or,

```
import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
for k, v in title[5].items():
    print(k,':', v)
```

o/p:

```
title : Medusa (IV)
year : 2015
```

or,

```
import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
year=[a for a in title]
year[:5]
```

o/p:

```
[OrderedDict([('title', 'The Rising Son'), ('year', '1990')]),
OrderedDict([('title', 'The Thousand Plane Raid'), ('year', '1969')]),
OrderedDict([('title', 'Crucea de piatra'), ('year', '1993')]),
OrderedDict([('title', 'Country'), ('year', '2000')]),
OrderedDict([('title', 'Gaiking II'), ('year', '2011')])]
```

Or,

```
import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
year85=[a for a in title if a['year']=='1985']
year85[:5]
```

o/p:

```
[OrderedDict([('title', 'Insaaf Main Karoonga'), ('year', '1985')]),
OrderedDict([('title', 'Vivre pour survivre'), ('year', '1985')]),
OrderedDict([('title', 'Water'), ('year', '1985')]),
OrderedDict([('title', 'Doea tanda mata'), ('year', '1985')]),
OrderedDict([('title', 'Koritsia gia tsibima'), ('year', '1985')])]
```



```

import pandas as pd

import os

title=pd.read_csv(r'c:\Users\Shubhamay\Documents\titles.csv',index_col=None)

movies90=title[int(title['year'])//10*10]

movies90.head(5)

```

o/p:

**TypeError:** cannot convert the series to <class 'int'>

Or,

```

import pandas as pd

import csv

title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))

year90=[a for a in title if int(a['year'])>int('1989') and
int(a['year'])<int('2000')]

year90[:5]

```

o/p:

```

[OrderedDict([('title', 'The Rising Son'), ('year', '1990')]),
 OrderedDict([('title', 'Crucea de piatra'), ('year', '1993')]),
 OrderedDict([('title', 'Poka Makorer Ghar Bosoti'), ('year', '1996')]),
 OrderedDict([('title', 'Maa Durga Shakti'), ('year', '1999')]),
 OrderedDict([('title', 'Conflict of Interest'), ('year', '1993')])]

```

Or,

```

import pandas as pd
import csv
title=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
macbeth=[m for m in title if m['title']=='Macbeth']

```

Macbeth

o/p:

```
[OrderedDict([('title', 'Macbeth'), ('year', '1913')]),
 OrderedDict([('title', 'Macbeth'), ('year', '2006')]),
 OrderedDict([('title', 'Macbeth'), ('year', '2013')]),
 OrderedDict([('title', 'Macbeth'), ('year', '1997')]),
 OrderedDict([('title', 'Macbeth'), ('year', '1998')])]
```

Or,

```
import pandas as pd
```

```
import csv
```

```
cast=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
```

```
cast[3:5]
```

o/p:

```
[OrderedDict([('title', 'Secret in Their Eyes'),
              ('year', '2015'),
              ('name', '$hutter'),
              ('type', 'actor'),
              ('character', '2002 Dodger Fan'),
              ('n', '')]),
 OrderedDict([('title', 'Steve Jobs'),
              ('year', '2015'),
              ('name', '$hutter'),
              ('type', 'actor'),
              ('character', '1988 Opera House Patron'),
              ('n', '')])]
```

Replace

```
import pandas as pd
row_data={'city':['kolkata','delhi','mumbai','alahabad','hydrabad','bengalore'],
          'rank':['1st','4th','5th','2nd','3rd','6th'],
          'population':['2crore','3crore','5 crore','4 crore','3crore','1.5crore'],
          'area':['250hec','500hec','750hec','600hec','450hec','300hec']}
town=pd.DataFrame(row_data,index=['A','B','C','D','E','F'])
print(town)
print('\n')
town['city'].replace('kolkata','ranchi',inplace=True)
print(town)
```

o/p:

	city	rank	population	area
A	kolkata	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	alahabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

	city	rank	population	area
A	ranchi	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	alahabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

Or,

```
import pandas as pd
row_data={'city':['kolkata','delhi','mumbai','alahabad','hydrabad','bengalore'],
'rank':['1st','4th','5th','2nd','3rd','6th'],
'population':['2crore','3crore','5 crore','4 crore','3crore','1.5crore'],
'area':['250hec','500hec','750hec','600hec','450hec','300hec']}
town=pd.DataFrame(row_data,index=['A','B','C','D','E','F'])
print(town)
print('\n')
town['city'].replace(['kolkata','alahabad'],['ranchi','aurangabad'],inplace=True)
print(town)
```

o/p:

	city	rank	population	area
A	kolkata	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	alahabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

	city	rank	population	area
A	ranchi	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	aurangabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

Or,

```

import pandas as pd
row_data={'city':['kolkata','delhi','mumbai','alahabad','hydrabad','bengalore'],
'rank':['1st','4th','5th','2nd','3rd','6th'],
'population':['2crore','3crore','5 crore','4 crore','3crore','1.5crore'],
'area':['250hec','500hec','750hec','600hec','450hec','300hec']}
town=pd.DataFrame(row_data,index=['A','B','C','D','E','F'])
print(town)
print('\n')
town.at['A','city']='ranchi'# inspite of 'at' we can also use loc

print(town)

```

or,

	city	rank	population	area
A	kolkata	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	alahabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

	city	rank	population	area
A	ranchi	1st	2crore	250hec
B	delhi	4th	3crore	500hec
C	mumbai	5th	5 crore	750hec
D	alahabad	2nd	4 crore	600hec
E	hydrabad	3rd	3crore	450hec
F	bengalore	6th	1.5crore	300hec

Or,

```

import pandas as pd
import csv
cast=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
cast0 = [c['n'] for c in cast ]
cast0[3:5]

```

o/p;

```
['', '']
```

Or,

```

import pandas as pd
import csv
cast=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
cast0 = [c['n'].replace('','0') for c in cast ]
cast0[3:5]

```

o/p:

```
['0', '0']
```

Or,

```
import pandas as pd
import csv
cast=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
cast0 = [{**c, 'n':c['n'].replace(' ', '0')} for c in cast ]
cast0[3:5]
```

o/p:

```
[{'title': 'Secret in Their Eyes',
  'year': '2015',
  'name': '$hutter',
  'type': 'actor',
  'character': '2002 Dodger Fan',
  'n': '0'},
 {'title': 'Steve Jobs',
  'year': '2015',
  'name': '$hutter',
  'type': 'actor',
  'character': '1988 Opera House Patron',
  'n': '0'}]
```

Or,

```
import pandas as pd
import os
title=pd.read_csv(r'c:\Users\Shubhamay\Documents\titles.csv',index_col=None)
movies90=title[title['title'].str.startswith ('Maa')]
print(movies90.head(5))
```

o/p:

	title	year
19	Maa Durga Shakti	1999
334	Maarek hob	2004
3046	Maa Aur Mamta	1970
6304	Maampazhakkaalam	2004
7470	Maa Vaibhav Laxmi	1989

Or,

```
import pandas as pd
import csv
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
maa=[m for m in cinema if m['title'].startswith ('Maa')]
maa[:5]
```

o/p:

```
[OrderedDict([('title', 'Maa Durga Shakti'), ('year', '1999')]),
 OrderedDict([('title', 'Maarek hob'), ('year', '2004')]),
```

```
OrderedDict([('title', 'Maa Aur Mamta'), ('year', '1970')]),
OrderedDict([('title', 'Maampazhakkaalam'), ('year', '2004')]),
OrderedDict([('title', 'Maa Vaibhav Laxmi'), ('year', '1989')])]
```

Or,

```
import pandas as pd
import csv
from collections import Counter
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
by_year=Counter(t['year'] for t in cinema)
by_year.elements
```

o/p:

```
<bound method Counter.elements of Counter({'2016': 2363, '2017': 2138,
'2015': 1849, '2014': 1701, '2013': 1609, '2012': 1500, '2011': 1457,
'2010': 1377, '2009': 1305, '2008': 1070, '2007': 950, '2006': 932,
'2005': 848, '2004': 760, '2003': 684, '2018': 633, '2000': 616, '2002':
597, '2001': 568, '1999': 558, '1998': 517, '1990': 515, '1989': 503,
'1987': 499, '1997': 480, '1991': 476, '1988': 469, '1969': 465, '1992':
462, '1994': 457, '1996': 451, '1995': 450, '1972': 442, '1970': 438,
'1993': 437, '1981': 435, '1984': 431, '1968': 431, '1986': 426, '1974':
422, '1982': 420, '1979': 418, '1985': 417, '1980': 416, '1971': 413,
'1966': 409, '1978': 407, '1983': 402, '1973': 391, '1967': 390, '1977':
381, '1965': 370, '1976': 368, '1964': 359, '1975': 357, '1958': 332,
'1963': 330, '1961': 325, '1962': 321, '1957': 310, '1960': 304, '1959':
295, '1956': 294, '1937': 290, '1951': 274, '1954': 271, '1920': 264,
'1916': 257, '1950': 257, '1955': 252, '1917': 249, '1918': 245, '1919':
245, '1936': 242, '1953': 240, '1939': 235, '1938': 234, '1952': 226,
'1927': 224, '1949': 220, '1935': 211, '1940': 209, '1934': 200, '1941':
198, '1933': 198, '1922': 197, '1921': 195, '1925': 193, '1942': 189,
'1926': 188, '1931': 187, '1948': 186, '1915': 181, '1947': 176, '1928':
171, '1932': 170, '1946': 165, '1929': 161, '1930': 157, '1924': 154,
'1944': 148, '1943': 144, '1945': 139, '1923': 133, '1914': 113, '2019':
105, '1913': 51, '2020': 37, '1912': 30, '1911': 12, '2021': 8, '1910': 5,
'2022': 5, '1909': 3, '1900': 2, '2024': 1, '1906': 1, '1905': 1, '2023':
1})>
```

Or,

```
import pandas as pd
import csv
from collections import Counter
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
by_year=Counter(t['year'] for t in cinema)
by_year.most_common(3)
```

o/p:

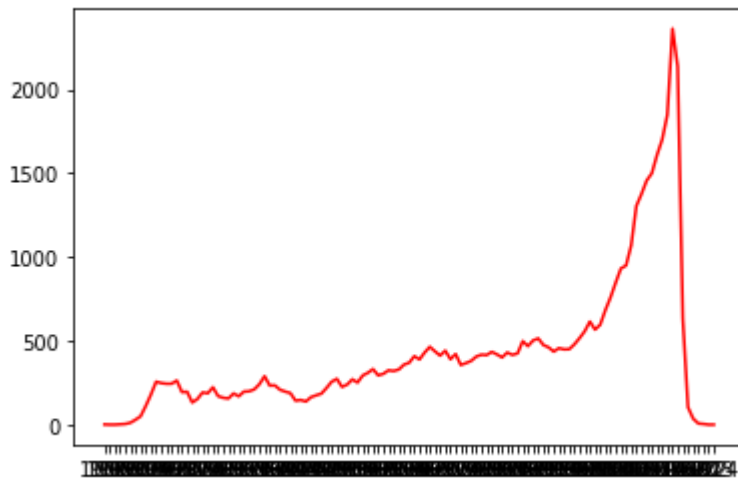
```
[('2016', 2363), ('2017', 2138), ('2015', 1849)]
```

Or,

```
import pandas as pd
import csv
from collections import Counter
from matplotlib import pyplot as plt
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
by_year=Counter(c['year'] for c in cinema)
data=by_year.most_common(len(cinema))
data=sorted(data)
x=[c[0] for c in data]
y=[c[1] for c in data]
plt.plot(x,y,'-r')
plt.show()
```

o/p:

plt.show()



Or,

```
name='shubhamaychand'
d={}
for c in name:
    if c not in d:
        d[c]=1
    else:
        d[c]+=1
print(d)
```

o/p:

```
{'s': 1, 'h': 1, 'u': 1, 'b': 1, 'a': 1, 'm': 1, 'y': 1, 'c': 1, 'n': 1,
'd': 1}
```

Or,

```

name='shubhamaychand'
d={}
for c in name:
    if c not in d:
        d[c]=1
    else:
        d[c]+=1
print(d) # thats what we do with defaultdict

```

o/p:

```

{'s': 1, 'h': 3, 'u': 1, 'b': 1, 'a': 3, 'm': 1, 'y': 1, 'c': 1, 'n': 1,
'd': 1}

```

Or,

```

from collections import defaultdict
name='shubhamaychand'
d=defaultdict(int)
for c in name:
    d[c]+=1
print(d)

```

o/p:

```

defaultdict(<class 'int'>, {'s': 1, 'h': 3, 'u': 1, 'b': 1, 'a': 3, 'm':
1, 'y': 1, 'c': 1, 'n': 1, 'd': 1})

```

or,

```

from collections import defaultdict
name='shubhamaychand'
d=defaultdict(int)
for c in name:
    d[c]+=1
print(d)
d['z']

```

o/p  
:

```

defaultdict(<class 'int'>, {'s': 1, 'h': 3, 'u': 1, 'b': 1, 'a': 3, 'm': 1,
'y': 1, 'c': 1, 'n': 1, 'd': 1})

```

Out[14]:

0

Or,

```

import pandas as pd
import csv
from collections import defaultdict
from matplotlib import pyplot as plt

```



```
d=defaultdict(list)
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
for c in cinema:
    d[c['year']].append(c['title'])
print(d)
```

o/p:

```
defaultdict(<class 'list'>, {'1990': ['The Rising Son', 'Arrive Alive',
'The Neverending Story II: The Next Chapter', 'Torn Apart', 'Thieves of
Fortune', 'Nyaya Anyaya', 'Dryan', 'Pernikahan berdarah', 'Huang jia du
chuan', 'Tutajosok', 'Jack el vigilante', 'Love Your Mama', 'Biao jie, ni
hao ye!', 'Chaparrito pero cumplidor', 'Jiao Yulu', 'Michael and Madonna',
.....
.....
.....
'The Eternal Law', 'Hamlet'], '2024': ['Hyperion: The Wrath of Kronos'],
'1900': ['Family Troubles (II)', "Pierrot's Problem, or How to Make a Fat
Wife Out of Two Lean Ones"], '2022': ['Silver from Water', 'The Cretaceous
Period', 'Oracle: The Last Judgment', 'Spider in the Web', 'Rosie'],
'1906': ['The Joe Gans-Battling Nelson Fight'], '1905': ['Valsons'],
'2023': ['Gnome Alone']})
```

Or,

```
import pandas as pd
import csv
from collections import defaultdict
from matplotlib import pyplot as plt
d=defaultdict(list)
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))
for c in cinema:
    d[c['year']].append(c['title'])
xx=[]
yy=[]
for k,v in d.items():
    xx.append(k)
    yy.append(v)
print(xx)
print(yy)
```

o/p:

```
['1990', '1969', '1993', '2000', '2011', '2015', '2008', '1970', '2016',
'2017', '2014', '1996', '1943', '1973', '2002', '1976', '2012', '1965',
'1999', '1917', '2009', '1989', '1916', '2001', '2013', '1992', '1998',
'1958', '2003', '1918', '1947', '1967', '1997', '2010', '1929', '2005',
'1986', '2006', '1988', '1963', '1935', '1932', '1950', '1985', '1922',
'1981', '1979', '1934', '1977', '1957', '1926', '2004', '2007', '1960',
'2018', '1974', '1953', '1975', '1913', '1928', '1931', '1964', '2019',
'1966', '1956', '1995', '1991', '1920', '1980', '1984', '1938', '1968',
'1982', '1930', '1971', '1972', '1983', '1961', '1924', '1919', '1994',
'1987', '1912', '1923', '1951', '1936', '1937', '1952', '1944', '1948',
'1925', '1959', '1954', '1915', '1941', '1962', '1945', '1933', '1978',
'1955', '1939', '2020', '1927', '1940', '1949', '1921', '1946', '1914',
```

```
'1942', '1909', '1911', '2021', '1910', '2024', '1900', '2022', '1906',  
'1905', '2023']
```

```
[['The Rising Son', 'Arrive Alive', 'The Neverending Story II: The Next  
Chapter', 'Torn Apart', 'Thieves of Fortune', 'Nyaya Anyaya', 'Dryan',  
'Pernikahan berdarah', 'Huang jia du chuan',
```

```
....
```

```
....
```

```
...
```

```
..
```

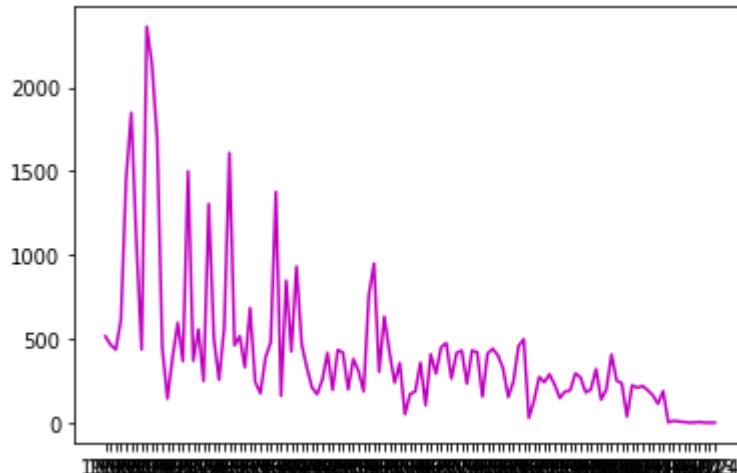
```
leepytime Gal', 'Jiban Sangini', 'Obliging Young Lady', 'Man from  
Cheyenne'], ['O Nono Mandamento', 'La Chicanera', 'Salome Mad'], ['Den  
store flyver', 'Pinocchio', 'The Squatter and the Clown', 'A Tale of the  
Australian Bush', 'Attack on the Gold Escort', 'Den farlige leg', 'Barnaby  
Rudge', 'Way Outback', 'Oborona Sevastopolya', 'The Cup Winner', 'Modieus  
geklede vrouwen tijdens de wedstrijden', 'In the Nick of Time'], ['Edge of  
Time', 'EX', 'Two Wheels', 'Burnt Offerings', 'Hd 164595', 'Prime Time',  
'The Witch of Novemeber', 'Night Breeds'], ['Paz e Amor', 'Lucrezia Borgia  
(II)', 'Jane Eyre', 'The Eternal Law', 'Hamlet'], ['Hyperion: The Wrath of  
Kronos'], ['Family Troubles (II)', "Pierrot's Problem, or How to Make a  
Fat Wife Out of Two Lean Ones"], ['Silver from Water', 'The Cretaceous  
Period', 'Oracle: The Last Judgment', 'Spider in the Web', 'Rosie'], ['The  
Joe Gans-Battling Nelson Fight'], ['Valsons'], ['Gnome Alone']]
```

```
or,
```

```
import pandas as pd  
import csv  
from collections import defaultdict  
from matplotlib import pyplot as plt  
d=defaultdict(list)  
cinema=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\titles.csv')))  
for c in cinema:  
    d[c['year']].append(c['title'])  
xx=[]  
yy=[]  
for k,v in d.items():  
    xx.append(k)  
    yy.append(len(v))  
print(yy)  
plt.plot(sorted(xx),yy,'-m')  
plt.show()
```

```
o/p:
```

```
[515, 465, 437, 616, 1457, 1849, 1070, 438, 2363, 2138, 1701, 451, 144,  
391, 597, 368, 1500, 370, 558, 249, 1305, 503, 257, 568, 1609, 462, 517,  
332, 684, 245, 176, 390, 480, 1377, 161, 848, 426, 932, 469, 330, 211,  
170, 257, 417, 197, 435, 418, 200, 381, 310, 188, 760, 950, 304, 633, 422,  
240, 357, 51, 171, 187, 359, 105, 409, 294, 450, 476, 264, 416, 431, 234,  
431, 420, 157, 413, 442, 402, 325, 154, 245, 457, 499, 30, 133, 274, 242,  
290, 226, 148, 186, 193, 295, 271, 181, 198, 321, 139, 198, 407, 252, 235,  
37, 224, 209, 220, 195, 165, 113, 189, 3, 12, 8, 5, 1, 2, 5, 1, 1, 1]
```



o/p:

```
import pandas as pds
import csv
from collections import defaultdict
casts=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
cf=[c for c in casts  if c['name']=='Aaron Abrams']
cf[:3]
```

or,

```
[OrderedDict([('title', '#FromJennifer'),
              ('year', '2017'),
              ('name', 'Aaron Abrams'),
              ('type', 'actor'),
              ('character', 'Ralph Sinclair'),
              ('n', '')]),
 OrderedDict([('title', '388 Arletta Avenue'),
              ('year', '2011'),
              ('name', 'Aaron Abrams'),
              ('type', 'actor'),
              ('character', 'Alex'),
              ('n', '4')]),
 OrderedDict([('title', 'Amelia'),
              ('year', '2009'),
              ('name', 'Aaron Abrams'),
              ('type', 'actor'),
              ('character', 'Slim Gordon'),
              ('n', '8')])]
```

Or,

```
import pandas as pds
import csv
from collections import defaultdict
casts=list(csv.DictReader(open(r'c:\Users\Shubhamay\Documents\cast.csv')))
cf=[c for c in casts  if c['name']=='Aaron Abrams']
```

```
cf[:3]
dcf=defaultdict(list)
for c in cf:
    dcf[c['year']].append(c['title'])
print(dcf)# aaron abrams movie by yearwise
```

o/p:

```
defaultdict(<class 'list'>, {'2017': ['#FromJennifer', 'The Go-Getters'],
'2011': ['388 Arletta Avenue', 'Jesus Henry Christ', 'Jesus Henry Christ',
'Take This Waltz', 'The Chicago 8'], '2009': ['Amelia', 'At Home by
Myself... with You'], '2005': ['Cinderella Man', 'Sabah'], '2015':
['Closet Monster', 'Regression'], '2018': ['Code 8'], '2007': ['Firehouse
Dog', 'Young People Fucking'], '2008': ['Flash of Genius'], '2013': ['It
Was You Charlie'], '2004': ['Resident Evil: Apocalypse', 'Siblings'],
'2003': ['The In-Laws', 'The Visual Bible: The Gospel of John'], '2006':
['Zoom']})
```

or,

```
import numpy as np
e=np.identity(7)
print(e)
```

o/p:

```
[[1. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0.]
 [0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0.]
 [0. 0. 0. 0. 0. 0. 1.]]
```

Or,

```
import numpy as np
e=np.eye(6)
print(e)
```

o/p:

```
[[1. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0.]
 [0. 0. 1. 0. 0. 0.]
 [0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 0. 1. 0.]
 [0. 0. 0. 0. 0. 1.]]
```

Or,

```
import numpy as np
e=np.eye(3)
e2=e+2
```

```
print(e)
print('\n')
print(e2)
```

o/p:

```
[[1. 0. 0.]
 [0. 1. 0.]
 [0. 0. 1.]]
```

```
[[3. 2. 2.]
 [2. 3. 2.]
 [2. 2. 3.]]
```

Or,

```
import numpy as np
e=np.eye(3)
e2=e+2
o=np.ones_like(e2)
print(o)
```

o/p:

```
[[1. 1. 1.]
 [1. 1. 1.]
 [1. 1. 1.]]
```

Or,

```
import numpy as np
e=np.eye(3)
e2=e+2
o=np.ones_like(e2,dtype=int)
print(o)
```

o/p:

```
[[1 1 1]
 [1 1 1]
 [1 1 1]]
```

Or,

```
import numpy as np
e=np.eye(3)
e2=e+2
o=np.ones_like(e2)
oi=o.astype(np.int32)
print(oi)
```

o/p:

```
[[1 1 1]
 [1 1 1]
 [1 1 1]]
```

Or,

```
a=['1','2','3','4']
p=np.array(a,dtype=np.string_)
print(p)
```

o/p:

```
[b'1' b'2' b'3' b'4']
```

Or,

```
a=['1','2','3','4']
p=np.array(a,dtype=np.string_)
q=p.astype(np.float64)
print(q)
```

o/p:

```
[1. 2. 3. 4.]
```

Or,

```
import numpy as np
rand=np.random
x=rand.randint(0,10,(3,4))
print(x)
```

o/p:

```
[[5 7 4 3]
 [5 5 5 4]
 [8 5 3 1]]
```

Or,

```
import numpy as np
rand=np.random
x=rand.randn(5,6)
print(x)
```

o/p:

```
[[ 0.52265713  0.43570358 -1.39707046 -0.38340307  0.22064965  0.54563496]
 [-0.16394387  0.7220064   0.74543097  0.09289938  2.32814676  0.13342849]
 [-1.72504827  0.29014411 -1.9403379  -0.33293788 -1.078882   -0.78748683]
 [ 0.81816548  0.95596753  0.07880478  0.70884246  0.96651734 -2.00719102]]
```

```
[ 0.51963768 -0.61231733  0.91529582 -0.56187951 -0.49541153
 0.68628551]]
```

Or,

```
name=np.array(['a','b','c','a','c','d'])
name=='a'
```

o/p:

```
array([ True, False, False,  True, False, False])
```

or,

```
import numpy as np
rand=np.random
x=rand.randn(5,6)
name=np.array(['a','b','c','a','c'])
x[name!='a']
```

o/p:

```
array([[ 1.06227974,  0.68462962, -0.90441791,  0.85591429, -1.50012545,
         0.93697974],
       [ 0.17670384, -0.74918953, -0.32842217, -1.10621675,  0.27872409,
        -0.30761946],
       [-0.39742488, -0.10644757,  0.48750297,  1.69362694, -0.70539074,
         0.79044531]])
```

Or,

```
name=np.array(['b','b','c','a','c'])
(name=='b')|(name=='c')
```

o/p:

```
array([ True,  True,  True, False,  True])
```

or,

```
import numpy as np
rand=np.random
x=rand.randn(5,6)
print(x)
name=np.array(['b','b','c','a','c'])
x[(name=='b')|(name=='c')]
```

o/p:

```
[[-1.54545909  0.34562906  0.5777866  0.04906868 -0.29968268 -2.67826305]
 [-2.27729213 -0.73306646 -0.08096472  0.9244824  -1.60750473  1.65276998]]
```

```
[ 0.78624298 -0.90254861 -1.78731832 -0.18182593 -0.01535218 -1.12986856]
[-1.01747079 -0.25469292  1.05983196  0.08273958  0.92598587 -0.81466961]
[-0.58735201  0.85932793 -1.72403302 -1.55201251  1.58671908 -0.54504397]]
```

Out[28]:

```
array([[ -1.54545909,  0.34562906,  0.5777866 ,  0.04906868, -0.29968268,
        -2.67826305],
       [ -2.27729213, -0.73306646, -0.08096472,  0.9244824 , -1.60750473,
         1.65276998],
       [ 0.78624298, -0.90254861, -1.78731832, -0.18182593, -0.01535218,
        -1.12986856],
       [-0.58735201,  0.85932793, -1.72403302, -1.55201251,  1.58671908,
        -0.54504397]])
```

Or,

```
import numpy as np
rand=np.random
x=rand.randint(0,10,(5,6))
print(x)
x[(x>3)&(x<6)]
```

o/p:

```
[[7 3 9 4 5 2]
 [9 6 8 1 1 3]
 [4 1 5 0 7 1]
 [2 7 9 0 8 9]
 [3 5 6 3 9 6]]
```

Out[29]:

```
array([4, 5, 4, 5, 5])
```

or,

```
import numpy as np
rand=np.random
x=rand.randn(3,4)
print(x)
x[(x>0)&(x<2)]
```

o/p:

```
[[ 0.23095506  0.07604772  0.2299885  -0.2332424 ]
 [-0.1027375  -0.77895507  0.6209939  0.34065389]
 [ 0.36044389 -1.06525068 -0.68911341 -0.20097188]]
```

Out[32]:

```
array([0.23095506, 0.07604772, 0.2299885 , 0.6209939 , 0.34065389,
```



```
0.36044389))
```

Or,

```
import numpy as np
a=np.arange(30).reshape(5,6)
print('our original array:', '\n',a)
print('\n')
b=a[[2,0,1,3]]
print('array after selecting row using index', '\n',b)
```

o/p;

```
our original array:
[[ 0  1  2  3  4  5]
 [ 6  7  8  9 10 11]
 [12 13 14 15 16 17]
 [18 19 20 21 22 23]
 [24 25 26 27 28 29]]
```

```
array after selecting row using index
[[12 13 14 15 16 17]
 [ 0  1  2  3  4  5]
 [ 6  7  8  9 10 11]
 [18 19 20 21 22 23]]
```

Or,

```
import numpy as np
a=np.arange(30).reshape(5,6)
print('our original array:', '\n',a)
print('\n')
b=a[[2,4,1,3]]
print('array after selecting row using index', '\n',b)
print('\n')
c=b.transpose(1,0)
print('transpose of this matrix', '\n',c)
o/p:
```

```
our original array:
[[ 0  1  2  3  4  5]
 [ 6  7  8  9 10 11]
 [12 13 14 15 16 17]
 [18 19 20 21 22 23]
 [24 25 26 27 28 29]]
```

```
array after selecting row using index
[[12 13 14 15 16 17]
 [24 25 26 27 28 29]
 [ 6  7  8  9 10 11]]
```

```
[18 19 20 21 22 23]]
```

transpose of this matrix

```
[[12 24  6 18]
 [13 25  7 19]
 [14 26  8 20]
 [15 27  9 21]
 [16 28 10 22]
 [17 29 11 23]]
```

Or,

```
import numpy as np
a=np.arange(30).reshape(5,6)
print('our original array:', '\n',a)
print('\n')
b=a[[2,4,1,3]]
print('array after selecting row using index', '\n',b)
print('\n')
c=b.T
print('transpose of this matrix', '\n',a)
```

o/p:

our original array:

```
[[ 0  1  2  3  4  5]
 [ 6  7  8  9 10 11]
 [12 13 14 15 16 17]
 [18 19 20 21 22 23]
 [24 25 26 27 28 29]]
```

array after selecting row using index

```
[[12 13 14 15 16 17]
 [24 25 26 27 28 29]
 [ 6  7  8  9 10 11]
 [18 19 20 21 22 23]]
```

transpose of this matrix

```
[[12 24  6 18]
 [13 25  7 19]
 [14 26  8 20]
 [15 27  9 21]
 [16 28 10 22]
 [17 29 11 23]]
```

Or,

```
import numpy as np
a=np.arange(12).reshape(3,4)
rand=np.random
b=rand.randn(3,4)
c=np.concatenate([a,b])
print(c)

o/p:

[[ 0.          1.          2.          3.          ]
 [ 4.          5.          6.          7.          ]
 [ 8.          9.         10.         11.          ]
 [ 2.36950312 -0.81584835 -1.0564184  -1.04734759]
 [ 1.60548228 -0.0856831  0.24379618 -0.9140995  ]
 [-0.64765005  0.84161718  0.32903737 -0.08572807]]
```

Or,

```
import numpy as np
a=np.arange(12).reshape(3,4)
rand=np.random
b=rand.randn(3,4)
c=np.concatenate([a,b],axis=1)
print(c)

o/p:

[[ 0.          1.          2.          3.         -0.73168009 -0.59164133
  0.73869133 -1.69210929]
 [ 4.          5.          6.          7.         -0.66435598  0.7889323
  0.68781963 -0.91870774]
 [ 8.          9.         10.         11.         -0.66976651 -2.14361792
 -1.17080739 -0.38897354]]
```

Or,

```
import pandas as pd
a=pd.Series([20,25,24,26,23,24,21,26],\
            index=[['a','a','a','a','b','b','b','b'],\
                  ['obj1','obj1','obj1','obj1','obj1','obj1','obj1','obj1']])
print(a)

o/p:

a  obj1    20
   obj1    25
   obj1    24
   obj1    26
b  obj1    23
   obj1    24
   obj1    21
   obj1    26
```

```
dtype: int64
```

or,

```
import pandas as pd
x=pd.Series([25,21,26,23,24,21,69,24],\
            index=[['a','a','a','a','b','b','b','b'],\
                  ['obj1','obj2','obj3','obj4','obj1','obj2','obj3','obj4']])
x.index
o/p:
```

```
MultiIndex([('a', 'obj1'),
            ('a', 'obj2'),
            ('a', 'obj3'),
            ('a', 'obj4'),
            ('b', 'obj1'),
            ('b', 'obj2'),
            ('b', 'obj3'),
            ('b', 'obj4')],
           )
```

Or,

```
import pandas as pd
x=pd.Series([25,21,26,23,24,21,69,24],\
            index=[['a','a','a','a','b','b','b','b'],\
                  ['obj1','obj2','obj3','obj4','obj1','obj2','obj3','obj4']])
x['b']
```

o/p:

```
obj1    24
obj2    21
obj3    69
obj4    24
dtype: int64
```

or,

```
import pandas as pd
x=pd.Series([25,21,26,23,24,21,69,24],\
            index=[['a','a','a','a','b','b','b','b'],\
                  ['obj1','obj2','obj3','obj4','obj1','obj2','obj3','obj4']])
x[:, 'obj1']
```

o/p:

```
a    25
b    24
dtype: int64
```

In [7]:

Or,

```
import pandas as pd
a=pd.Series([21,22,23,25,24,26,27,28]\
            ,index=[['a','a','a','a','b','b','b','b'],\
                    ['obj1','obj2','obj3','obj4','obj1','obj2','obj3','obj4']])
print(a)
print('\n')
print(a.unstack())
print('\n')
print(a.unstack(0))
print('\n')
print(a.unstack(1))
```

o/p:

```
a  obj1    21
   obj2    22
   obj3    23
   obj4    25
b  obj1    24
   obj2    26
   obj3    27
   obj4    28
dtype: int64
```

	obj1	obj2	obj3	obj4
a	21	22	23	25
b	24	26	27	28

	a	b
obj1	21	24
obj2	22	26
obj3	23	27
obj4	25	28

	obj1	obj2	obj3	obj4
a	21	22	23	25
b	24	26	27	28

or,

```
import pandas as pd
a=pd.Series([21,22,23,25,24,26,27,28]\
            ,index=[['a','a','a','a','b','b','b','b'],\
                    ['obj1','obj2','obj3','obj4','obj1','obj2','obj3','obj4']])
```

```

print(a)
print('\n')
d=a.unstack()
print(d)
print('\n')
f=d.stack()
print(f)

```

o/p:

```

a  obj1    21
   obj2    22
   obj3    23
   obj4    25
b  obj1    24
   obj2    26
   obj3    27
   obj4    28
dtype: int64

```

	obj1	obj2	obj3	obj4
a	21	22	23	25
b	24	26	27	28

```

a  obj1    21
   obj2    22
   obj3    23
   obj4    25
b  obj1    24
   obj2    26
   obj3    27
   obj4    28
dtype: int64

```

or,

```

import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=['one','two','three','four'],columns=['red','green',
,'red'])
print(b)

```

o/p:

	red	green	red
one	0	1	2
two	3	4	5
three	6	7	8
four	9	10	11

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=['one','two','three','four'],columns=['red','green',
,'yellow'])
print(b)
```

o/p:

	red	green	yellow
one	0	1	2
two	3	4	5
three	6	7	8
four	9	10	11

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['num1','num2','num3'],['red','green','yellow']])
print(b)
```

o/p:

		num1	num2	num3
		red	green	yellow
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['num1','num2','num3'],['red','green','yellow']])
print(b)
print('\n')
print(b.index)
```

o/p:

		num1	num2	num3
		red	green	yellow
a	one	0	1	2

	two	3	4	5
b	three	6	7	8
	four	9	10	11

```
MultiIndex([('a', 'one'),
            ('a', 'two'),
            ('b', 'three'),
            ('b', 'four')],
           )
```

Or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'], ['one','two','three','four']],\
               columns=[['num1','num2','num3'], ['red','green','yellow']])
print(b)
print('\n')
print(b.columns)
```

o/p

		num1	num2	num3
		red	green	yellow
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

```
MultiIndex([('num1', 'red'),
            ('num2', 'green'),
            ('num3', 'yellow')],
           )
```

Or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'], ['one','two','three','four']],\
               columns=[['num1','num2','num3'], ['red','green','yellow']])
print(b)
print('\n')
b.index.names=['letter','number']
b.columns.names=['n','colour']
print(b)
```

o/p:



		num1	num2	num3
		red	green	yellow
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

n		num1	num2	num3
colour		red	green	yellow
letter	number			
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['num1','num2','num3'],['red','green','yellow']])
print(b)
print('\n')
b.index.names=['letter','number']
b.columns.names=['n','colour']
#print(b['num1'])
print(b.loc[:, 'num1'])
```

o/p:

		num1	num2	num3
		red	green	yellow
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

colour		red
letter	number	
a	one	0
	two	3
b	three	6
	four	9

or,

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

```

a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'], ['one','two','three','four']],\
               columns=[['num1','num2','num3'], ['red','green','yellow']])
print(b)
print('\n')
b.index.names=['letter','number']
b.columns.names=['n','colour']
print(b)
print('\n')
print(b.loc['a'])
print('\n')
print(b.iloc[0])

```

o/p:

		num1	num2	num3
		red	green	yellow
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

n		num1	num2	num3
colour		red	green	yellow
letter	number			
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

n	num1	num2	num3
colour	red	green	yellow
number			
one	0	1	2
two	3	4	5

n	colour	
num1	red	0
num2	green	1
num3	yellow	2

Name: (a, one), dtype: int32

Or,

```

import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'], ['one','two','three','four']],\
               columns=[['red','green','yellow'], ['num1','num2','num3']])
b.index.names=['letter','numerical']
b.columns.names=['colour','num']

```

```

print(b)
print('\n')
c=b.swaplevel('letter','numerical')
print(c)

```

o/p:

colour		red	green	yellow
num		num1	num2	num3
letter	numerical			
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

colour		red	green	yellow
num		num1	num2	num3
numerical	letter			
one	a	0	1	2
two	a	3	4	5
three	b	6	7	8
four	b	9	10	11

or,

```

import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['red','green','yellow'],['num1','num2','num3']])
b.index.names=['letter','numerical']
b.columns.names=['colour','num']
print(b)
print('\n')
c=b.sort_index(level='numerical')
print(c)

```

o/p:

colour		red	green	yellow
num		num1	num2	num3
letter	numerical			
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

colour		red	green	yellow
num		num1	num2	num3
letter	numerical			
b	four	9	10	11
a	one	0	1	2

b	three	6	7	8
a	two	3	4	5

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['red','green','yellow'],['num1','num2','num3']])
b.index.names=['letter','numerical']
b.columns.names=['colour','num']
print(b)
print('\n')
c=b.sum(level='letter')#sum in a+a and b+b: adda based on similar letter name
print(c)
```

o/p:

colour		red	green	yellow
num		num1	num2	num3
letter	numerical			
a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

colour	red	green	yellow
num	num1	num2	num3
letter			
a	3	5	7
b	15	17	19

or,

```
import numpy as np
import pandas as pd
a=np.arange(12).reshape(4,3)
b=pd.DataFrame(a,index=[['a','a','b','b'],['one','two','three','four']],\
               columns=[['red','green','red'],['num1','num2','num3']])
b.index.names=['letter','numerical']
b.columns.names=['colour','num']
print(b)
print('\n')
c=b.sum(level='colour',axis=1)#add based on similar color name
print(c)
```

o/p:

colour	red	green	red
num	num1	num2	num3
letter	numerical		

a	one	0	1	2
	two	3	4	5
b	three	6	7	8
	four	9	10	11

colour		red	green
letter	numerical		
a	one	2	1
	two	8	4
b	three	14	7
	four	20	10

or,

```
# convert dataframe to csv file
import pandas as pd
b={'name':['ram','shyam','jadu','madhu'],'roll
no':[25,23,24,26],'marks':[654,625,623,613]}
a=pd.DataFrame(b,index=['a','b','c','d'])
print(a)
a.to_csv('friend.csv')#this file is located in c/user/Shubhamay
```

Or,

```
df1=pd.DataFrame({'key1':['ram','shyam','jadu','madhu','jibon'],'data':range(5)},\
                  index=['player1','player2','player3','player4','player5'])
print(df1)
```

o/p:

	key1	data
player1	ram	0
player2	shyam	1
player3	jadu	2
player4	madhu	3
player5	jibon	4

or,

```
a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['c','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=0)
```

```
print('\n')
print(s4)
```

o/p:

```
a      0
b      1
dtype: int64
```

```
c      2
d      1
e      3
dtype: int64
```

```
e      0
f      3
dtype: int64
```

```
a      0
b      1
c      2
d      1
e      3
dtype: int64
```

or,

```
a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['c','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=0,keys=['i','ii'])
print('\n')
print(s4)
```

o/p:

```
a      0
b      1
dtype: int64
```

```
c      2
d      1
e      3
dtype: int64
```

```
e    0
f    3
dtype: int64
```

```
i    a    0
      b    1
ii   c    2
      d    1
      e    3
dtype: int64
```

or,

```
a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['c','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=1)
print('\n')
print(s4)
```

o/p:

```
a    0
b    1
dtype: int64
```

```
c    2
d    1
e    3
dtype: int64
```

```
e    0
f    3
dtype: int64
```

```
      0    1
a  0.0  NaN
b  1.0  NaN
c  NaN  2.0
d  NaN  1.0
e  NaN  3.0
```

or,

```

a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['c','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=1,keys=['i','ii'])
print('\n')
print(s4)

```

o/p:

```

a      0
b      1
dtype: int64

```

```

c      2
d      1
e      3
dtype: int64

```

```

e      0
f      3
dtype: int64

```

```

i   a      0
   b      1
ii  c      2
   d      1
   e      3
dtype: int64

```

or,

```

a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['c','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=1,join='inner')
print('\n')
print(s4)

```

o/p:



```
a      0
b      1
dtype: int64
```

```
c      2
d      1
e      3
dtype: int64
```

```
e      0
f      3
dtype: int64
```

```
Empty DataFrame
Columns: [0, 1]
Index: []
```

Or,

```
a=pd.Series([0,1],index=['a','b'])
b=pd.Series([2,1,3],index=['b','d','e'])
s3=pd.Series([0,3],index=['e','f'])
print(a)
print('\n')
print(b)
print('\n')
print(s3)
s4=pd.concat([a,b],axis=1,join='inner',keys=['i'])
print('\n')
print(s4)
```

o/p:

```
a      0
b      1
dtype: int64
```

```
b      2
d      1
e      3
dtype: int64
```

```
e      0
f      3
dtype: int64
```

```
      i
a      0
```

b 1

or,

```
df1=pd.DataFrame({'name':['ram','shyam','jadu','madhu','jibon'],'key1':['a','b','c','d','e'],'n':range(5)})
```

```
df2=pd.DataFrame({'name':['lalita','shyam','bristi'],'key2':['a','b','c'], 'p':range(3)})
```

```
print(a)
```

```
print('\n')
```

```
print(b)
```

```
print('\n')
```

```
print(s3)
```

```
s4=pd.concat([df1,df2],axis=1,join='inner',keys=['i','ii'])
```

```
print('\n')
```

```
print(s4)
```

o/p:

a 0

b 1

dtype: int64

b 2

d 1

e 3

dtype: int64

e 0

f 3

dtype: int64

	i			ii		
	name	key1	n	name	key2	p
0	ram	a	0	lalita	a	0
1	shyam	b	1	shyam	b	1
2	jadu	c	2	bristi	c	2

Or,

```
df1=pd.DataFrame({'name':['ram','shyam','jadu','madhu','jibon'],'key1':['a','b','c','d','e'],'n':range(5)})
```

```
df2=pd.DataFrame({'name':['lalita','shyam','bristi'],'key2':['a','b','c'], 'p':range(3)})
```

```
print(a)
```

```
print('\n')
```

```
print(b)
```

```
print('\n')
```

```
print(s3)
```

```
s4=pd.concat({'level1':df1,'level2':df2},axis=1,join='inner')
print('\n')
print(s4)
```

o/p:

```
a    0
b    1
dtype: int64
```

```
b    2
d    1
e    3
dtype: int64
```

```
e    0
f    3
dtype: int64
```

level1			level2			
	name	key1	n	name	key2	p
0	ram	a	0	lalita	a	0
1	shyam	b	1	shyam	b	1
2	jadu	c	2	bristi	c	2

Or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*3,'k2':range(6)},index=['a','b','c',
'd','e','f'])
print(a)
```

o/p:

	k1	k2
a	one	0
b	one	1
c	one	2
d	two	3
e	two	4
f	two	5

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*4+['two']*2,'k2':[1,1,2,2,3,5]},index=['a','b',
'c','d','e','f'])
print(a)#creating duplicate row using duplicate row.
a.duplicated()#true if row repeated
```

o/p:

	k1	k2
a	one	1
b	one	1
c	one	2
d	one	2
e	two	3
f	two	5

Out[10]:

a	False
b	True
c	False
d	True
e	False
f	False

dtype: bool

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*3,'k2':[1,1,2,2,3,5]},index=['a','b','c','d','e','f'])
print(a)#creating duplicate row using duplicate row.
a.duplicated()#true if row repeated
```

o/p:

	k1	k2
a	one	1
b	one	1
c	one	2
d	two	2
e	two	3
f	two	5

Out[11]:

a	False
b	True
c	False
d	False
e	False
f	False

dtype: bool

or,

```
import pandas as pd
```

```

a=pd.DataFrame({'k1':['one']*3+['two']*3,'k2':[1,1,2,2,3,5],'k3':['i']*1+['ii']*5},index=['a','b','c','d','e','f'])
print(a)#creating duplicate row using duplicate row.
a.duplicated()#true if row repeated

```

o/p:

	k1	k2	k3
a	one	1	i
b	one	1	ii
c	one	2	ii
d	two	2	ii
e	two	3	ii
f	two	5	ii

Out[12]:

```

a    False
b    False
c    False
d    False
e    False
f    False
dtype: bool

```

o/p:

```

import pandas as pd
a=pd.DataFrame({'k1':['one']*4+['two']*2,'k2':[1,1,2,2,3,5],'k3':['i']*1+['ii']*5},index=['a','b','c','d','e','f'])
print(a)#creating duplicate row using duplicate row.
a.duplicated()#true if row repeated

```

or,

	k1	k2	k3
a	one	1	i
b	one	1	ii
c	one	2	ii
d	one	2	ii
e	two	3	ii
f	two	5	ii

Out[13]:

```

a    False
b    False
c    False
d     True
e    False
f    False

```

dtype: bool

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*4+['two']*2,'k2':[1,1,2,2,3,5],'k3':['i']*1+['ii']*5},index=['a','b','c','d','e','f'])
print(a)#creating duplicate row using duplicate row.
print('\n')
print(a.drop_duplicates())#drop repeated row
```

o/p:

	k1	k2	k3
a	one	1	i
b	one	1	ii
c	one	2	ii
d	one	2	ii
e	two	3	ii
f	two	5	ii

	k1	k2	k3
a	one	1	i
b	one	1	ii
c	one	2	ii
e	two	3	ii
f	two	5	ii

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4]},index=['a','b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.drop_duplicates())#here keep the first duplicates entities
print('\n')
print(a.drop_duplicates(keep='last'))#here keep the last duplicates entities
```

o/p:

our original dataframe

	k1	k2
a	one	1
b	one	1
c	one	2
d	two	3
e	two	3
f	two	4
g	two	4

	k1	k2
a	one	1
c	one	2
d	two	3
f	two	4

	k1	k2
b	one	1
c	one	2
e	two	3
g	two	4

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4]},index=['a',
'b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.drop_duplicates(['k1']))
```

o/p:

our original dataframe

	k1	k2
a	one	1
b	one	1
c	one	2
d	two	3
e	two	3
f	two	4
g	two	4

	k1	k2
a	one	1
d	two	3

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4]},index=['a',
'b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.drop_duplicates(['k1','k2']))
```

o/p:

our original dataframe

	k1	k2
a	one	1
b	one	1
c	one	2
d	two	3
e	two	3
f	two	4
g	two	4

	k1	k2
a	one	1
c	one	2
d	two	3
f	two	4

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4]},\
                'k3':['i','ii','iii','iv','v','vi','vii']},\
                index=['a','b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.drop_duplicates(['k1','k2']))
```

o/p:

our original dataframe

	k1	k2	k3
a	one	1	i
b	one	1	ii
c	one	2	iii
d	two	3	iv
e	two	3	v
f	two	4	vi
g	two	4	vii

	k1	k2	k3
a	one	1	i
c	one	2	iii
d	two	3	iv
f	two	4	vi

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4]},\
                'k3':['i','ii','iii','iv','v','vi','vii']},\
                'k3':['data1','data2','data3','data4','data5','data6','data7']},\
                index=['a','b','c','d','e','f','g'])
```



```
print('our original dataframe','\n',a)
print('\n')
print(a.replace('one','1st'))
```

o/p:

our original dataframe

	k1	k2	k3
a	one	1	data1
b	one	1	data2
c	one	2	data3
d	two	3	data4
e	two	3	data5
f	two	4	data6
g	two	4	data7

	k1	k2	k3
a	1st	1	data1
b	1st	1	data2
c	1st	2	data3
d	two	3	data4
e	two	3	data5
f	two	4	data6
g	two	4	data7

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4],\
                'k3':['i','ii','iii','iv','v','vi','vii'],\
                'k3':['data1','data2','data3','data4','data5','data6','data7']},\
                index=['a','b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.replace(['one','two'],['first','second']))#replace one with first
and two with second
```

o/p:

our original dataframe

	k1	k2	k3
a	one	1	data1
b	one	1	data2
c	one	2	data3
d	two	3	data4
e	two	3	data5
f	two	4	data6
g	two	4	data7

	k1	k2	k3
a	first	1	data1
b	first	1	data2
c	first	2	data3
d	second	3	data4
e	second	3	data5
f	second	4	data6
g	second	4	data7

or,

```
import pandas as pd
a=pd.DataFrame({'k1':['one']*3+['two']*4,'k2':[1,1,2,3,3,4,4],\
               'k3':['i','ii','iii','iv','v','vi','vii'],\
               'k3':['data1','data2','data3','data4','data5','data6','data7']},\
               index=['a','b','c','d','e','f','g'])
print('our original dataframe','\n',a)
print('\n')
print(a.replace({'one':'mono','two':'double'}))# replace one with first
and two with second
```

o/p:

our original dataframe

	k1	k2	k3
a	one	1	data1
b	one	1	data2
c	one	2	data3
d	two	3	data4
e	two	3	data5
f	two	4	data6
g	two	4	data7

	k1	k2	k3
a	mono	1	data1
b	mono	1	data2
c	mono	2	data3
d	double	3	data4
e	double	3	data5
f	double	4	data6
g	double	4	data7

or,

```
c=pd.DataFrame({'data1':[5,6,4,7,8,9,6],'data2':[4,6,5,8,5,6,2],\
               'k1':['a','b','a','a','b','a','b'],\
               'k2':['one','two','two','one','two','one','two']})
print(c)
g=c['data1'].groupby(c['k1'])
print(g)
```

o/p:

	data1	data2	k1	k2
0	5	4	a	one
1	6	6	b	two
2	4	5	a	two
3	7	8	a	one
4	8	5	b	two
5	9	6	a	one
6	6	2	b	two

<pandas.core.groupby.generic.SeriesGroupBy object at 0x00000078372E57C8>

or:

```
c=pd.DataFrame({'data1':[5,6,4,7,8,9,6], 'data2':[4,6,5,8,5,6,2], \
                'k1':['a','b','a','a','b','a','b'], \
                'k2':['one','two','two','one','two','one','two']})
```

```
print(c)
```

```
g=c['data1'].groupby(c['k1'])
```

```
print(g.mean())#a: (9+7+4+5)/4=6.25 and b: (6+8+6)/3=6.66
```

o/p:

	data1	data2	k1	k2
0	5	4	a	one
1	6	6	b	two
2	4	5	a	two
3	7	8	a	one
4	8	5	b	two
5	9	6	a	one
6	6	2	b	two

k1

a 6.250000

b 6.666667

Name: data1, dtype: float64

Or,

```
c=pd.DataFrame({'data1':[5,6,4,7,8,9,6], 'data2':[4,6,5,8,5,6,2], \
                'k1':['a','b','a','a','b','a','b'], \
                'k2':['one','two','two','one','two','one','one']})
```

```
print(c)
```

```
g=c['data1'].groupby([c['k1'],c['k2']])# first groupby 'k1' next 'k2'
```

```
print(g.mean())#a: ('one'-(5+7+9)/3=7,two=4/1=4),b: ('one'-(8+6)/2=7,one=6/1=6)
```

o/p:

	data1	data2	k1	k2
0	5	4	a	one
1	6	6	b	two
2	4	5	a	two
3	7	8	a	one

```

4      8      5  b  two
5      9      6  a  one
6      6      2  b  one
k1  k2
a   one    7
    two    4
b   one    6
    two    7
Name: data1, dtype: int64

```

Or,

```

import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5], 'K1':
['a','a','b','b','a','b','a','b'],\
                  'K2':['one','one','two','one','two','two','one','one']})
grp1=df['Data2'].groupby([df['K1'],df['K2']])
grp1.mean()

```

o/p:

```

K1  K2
a   one    4.666667
    two    5.000000
b   one    5.500000
    two    5.000000
Name: Data2, dtype: float64

```

Iteration in group by:

```

import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5],\
                  'K1':['a','a','b','b','a','b','a','b'],\
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n', df)
print('\n')
grp1=df['Data1'].groupby(df['K1'])
print(grp1.mean())
print('\n')
for name, group in grp1:
    print(name)
    print(group)

```

o/p:

```

our original table:
   Data1  Data2 K1  K2
0      3      5  a  one
1      5      5  a  one
2      3      5  b  two
3      5      6  b  one
4      4      5  a  two

```

5	3	5	b	two
6	4	4	a	one
7	5	5	b	one

```
K1
a    4
b    4
Name: Data1, dtype: int64
```

```
a
0    3
1    5
4    4
6    4
Name: Data1, dtype: int64
b
2    3
3    5
5    3
7    5
Name: Data1, dtype: int64
```

Or,

```
import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5], \
                  'K1':['a','a','b','b','a','b','a','b'], \
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n', df)
print('\n')
grp1=df['Data1'].groupby([df['K1'], df['K2']])
print(grp1.mean())
print('\n')
for name, group in grp1:
    print(name)
    print(group)
```

o/p:

```
our original table:
  Data1  Data2 K1  K2
0     3     5  a  one
1     5     5  a  one
2     3     5  b  two
3     5     6  b  one
4     4     5  a  two
5     3     5  b  two
6     4     4  a  one
7     5     5  b  one
```

```

K1  K2
a   one    4
    two    4
b   one    5
    two    3
Name: Data1, dtype: int64

```

```

('a', 'one')
0     3
1     5
6     4
Name: Data1, dtype: int64
('a', 'two')
4     4
Name: Data1, dtype: int64
('b', 'one')
3     5
7     5
Name: Data1, dtype: int64
('b', 'two')
2     3
5     3
Name: Data1, dtype: int64

```

Or,

```

import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5], \
                  'K1':['a','a','b','b','a','b','a','b'], \
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n', df)
print('\n')
grp1=df['Data2'].groupby([df['K1'], df['K2']])
print(grp1.mean())
print('\n')
for name, group in grp1:
    print(name)
    print(group)

```

o/p:

```

our original table:
   Data1  Data2 K1  K2
0       3      5  a  one
1       5      5  a  one
2       3      5  b  two
3       5      6  b  one
4       4      5  a  two
5       3      5  b  two
6       4      4  a  one
7       5      5  b  one

```

```

K1  K2
a   one    4.666667
    two    5.000000
b   one    5.500000
    two    5.000000
Name: Data2, dtype: float64

```

```

('a', 'one')
0     5
1     5
6     4
Name: Data2, dtype: int64
('a', 'two')
4     5
Name: Data2, dtype: int64
('b', 'one')
3     6
7     5
Name: Data2, dtype: int64
('b', 'two')
2     5
5     5
Name: Data2, dtype: int64

```

Or,

```

import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5],\
                  'K1':['a','a','b','b','a','b','a','b'],\
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n',df)
print('\n')
grp1=df['Data1'].groupby([df['K1'],df['K2']])
print(grp1.mean())
print('\n')
for (K1,K2),group in grp1:
    print(K1,K2)
    print(group)

```

o/p:

```

our original table:
   Data1  Data2 K1  K2
0       3      5  a  one
1       5      5  a  one
2       3      5  b  two
3       5      6  b  one
4       4      5  a  two
5       3      5  b  two
6       4      4  a  one
7       5      5  b  one

```

```

K1  K2
a   one    4
    two    4
b   one    5
    two    3
Name: Data1, dtype: int64

```

```

a one
0    3
1    5
6    4
Name: Data1, dtype: int64
a two
4    4
Name: Data1, dtype: int64
b one
3    5
7    5
Name: Data1, dtype: int64
b two
2    3
5    3
Name: Data1, dtype: int64

```

Or,

```

import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5], \
                  'K1':['a','a','b','b','a','b','a','b'], \
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n', df)
print('\n')
grp1=df['Data1'].groupby(df['K1'])
print(grp1.max())

```

o/p:

```

our original table:
   Data1  Data2 K1  K2
0      3      5  a  one
1      5      5  a  one
2      3      5  b  two
3      5      6  b  one
4      4      5  a  two
5      3      5  b  two
6      4      4  a  one
7      5      5  b  one

```

K1



```
a      5
b      5
Name: Data1, dtype: int64
```

Or,

```
import pandas as pd
df=pd.DataFrame({"Data1":[3,5,3,5,4,3,4,5], 'Data2':[5,5,5,6,5,5,4,5],\
                  'K1':['a','a','b','b','a','b','a','b'],\
                  'K2':['one','one','two','one','two','two','one','one']})
print('our original table:', '\n', df)
print('\n')
grp1=df['Data1'].groupby([df['K1'], df['K2']])
print(grp1.max())
```

o/p:

```
our original table:
   Data1  Data2 K1  K2
0      3      5  a  one
1      5      5  a  one
2      3      5  b  two
3      5      6  b  one
4      4      5  a  two
5      3      5  b  two
6      4      4  a  one
7      5      5  b  one
```

```
K1  K2
a   one    5
    two    4
b   one    5
    two    3
Name: Data1, dtype: int64
```

Date-time-series:

```
import pandas as pd
import numpy as np
rng=pd.date_range('2011-03-01 10:15', periods=10, freq='M')
rng
```

or,

```
DatetimeIndex(['2011-03-31 10:15:00', '2011-04-30 10:15:00',
               '2011-05-31 10:15:00', '2011-06-30 10:15:00',
               '2011-07-31 10:15:00', '2011-08-31 10:15:00',
               '2011-09-30 10:15:00', '2011-10-31 10:15:00',
               '2011-11-30 10:15:00', '2011-12-31 10:15:00'],
              dtype='datetime64[ns]', freq='M')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range('01-03-2020 10:15',periods=13,freq='M')
rng
```

o/p:

```
DatetimeIndex(['2020-01-31 10:15:00', '2020-02-29 10:15:00',
               '2020-03-31 10:15:00', '2020-04-30 10:15:00',
               '2020-05-31 10:15:00', '2020-06-30 10:15:00',
               '2020-07-31 10:15:00', '2020-08-31 10:15:00',
               '2020-09-30 10:15:00', '2020-10-31 10:15:00',
               '2020-11-30 10:15:00', '2020-12-31 10:15:00',
               '2021-01-31 10:15:00'],
              dtype='datetime64[ns]', freq='M')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range('03-01-2019 10:15',periods=13,freq='MS')#MM-DD-YYYY
rng# in result YYYY-MM-DD
```

o/p:

```
DatetimeIndex(['2019-03-01 10:15:00', '2019-04-01 10:15:00',
               '2019-05-01 10:15:00', '2019-06-01 10:15:00',
               '2019-07-01 10:15:00', '2019-08-01 10:15:00',
               '2019-09-01 10:15:00', '2019-10-01 10:15:00',
               '2019-11-01 10:15:00', '2019-12-01 10:15:00',
               '2020-01-01 10:15:00', '2020-02-01 10:15:00',
               '2020-03-01 10:15:00'],
              dtype='datetime64[ns]', freq='MS')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range('03-02-2019 10:15',periods=13,freq='MS')#MM-DD-YYYY
rng# in result YYYY-MM-DD
```

o/p:

```
DatetimeIndex(['2019-04-01 10:15:00', '2019-05-01 10:15:00',
               '2019-06-01 10:15:00', '2019-07-01 10:15:00',
               '2019-08-01 10:15:00', '2019-09-01 10:15:00',
               '2019-10-01 10:15:00', '2019-11-01 10:15:00',
               '2019-12-01 10:15:00', '2020-01-01 10:15:00',
               '2020-02-01 10:15:00', '2020-03-01 10:15:00',
```

```
        '2020-04-01 10:15:00'],
        dtype='datetime64[ns]', freq='MS')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range('03-02-2019 10:15',periods=13,freq='D')#MM-DD-YYYY
rng# in result YYYY-MM-DD
```

o/p:

```
DatetimeIndex(['2019-03-02 10:15:00', '2019-03-03 10:15:00',
               '2019-03-04 10:15:00', '2019-03-05 10:15:00',
               '2019-03-06 10:15:00', '2019-03-07 10:15:00',
               '2019-03-08 10:15:00', '2019-03-09 10:15:00',
               '2019-03-10 10:15:00', '2019-03-11 10:15:00',
               '2019-03-12 10:15:00', '2019-03-13 10:15:00',
               '2019-03-14 10:15:00'],
              dtype='datetime64[ns]', freq='D')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range(start='13 june 2020 10:15',end='19 june 2020',freq='12H')
```

o/p:

```
DatetimeIndex(['2020-06-13 10:15:00', '2020-06-13 22:15:00',
               '2020-06-14 10:15:00', '2020-06-14 22:15:00',
               '2020-06-15 10:15:00', '2020-06-15 22:15:00',
               '2020-06-16 10:15:00', '2020-06-16 22:15:00',
               '2020-06-17 10:15:00', '2020-06-17 22:15:00',
               '2020-06-18 10:15:00', '2020-06-18 22:15:00'],
              dtype='datetime64[ns]', freq='12H')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range(start='13 june 2020 10:15',end='19 june
2020',freq='12H',tz="asia/kolkata")#MM-DD-YYYY
rng# in result YYYY-MM-DD
```

o/p:

```
DatetimeIndex(['2020-06-13 10:15:00+05:30', '2020-06-13 22:15:00+05:30',
               '2020-06-14 10:15:00+05:30', '2020-06-14 22:15:00+05:30',
               '2020-06-15 10:15:00+05:30', '2020-06-15 22:15:00+05:30',
               '2020-06-16 10:15:00+05:30', '2020-06-16 22:15:00+05:30',
               '2020-06-17 10:15:00+05:30', '2020-06-17 22:15:00+05:30',
               '2020-06-18 10:15:00+05:30', '2020-06-18 22:15:00+05:30'],
              dtype='datetime64[ns, Asia/Kolkata]', freq='12H')
```

or,

```
import pandas as pd
import numpy as np
rng=pd.date_range(start='13 june 2020 10:15',end='19 june
2020',freq='12H',tz="asia/kolkata")
rng.tz_convert('Australia/Sydney')
```

o/p:

```
DatetimeIndex(['2020-06-13 14:45:00+10:00', '2020-06-14 02:45:00+10:00',
               '2020-06-14 14:45:00+10:00', '2020-06-15 02:45:00+10:00',
               '2020-06-15 14:45:00+10:00', '2020-06-16 02:45:00+10:00',
               '2020-06-16 14:45:00+10:00', '2020-06-17 02:45:00+10:00',
               '2020-06-17 14:45:00+10:00', '2020-06-18 02:45:00+10:00',
               '2020-06-18 14:45:00+10:00', '2020-06-19 02:45:00+10:00'],
              dtype='datetime64[ns, Australia/Sydney]', freq='12H')
```

or

```
import pandas as pd
import numpy as np
rng=pd.date_range(start='13 june 2020 10:15',end='19 june
2020',freq='12H',tz="asia/kolkata")#MM-DD-YYYY
rng.tz_convert('Australia/Sydney')
len(rng)
```

o/p:

12

Or,

```
dd=['04/05/2012','02/12/2013','05/06/2018']
pd.to_datetime(dd)
```

o/p:

```
DatetimeIndex(['2012-04-05', '2013-02-12', '2018-05-06'],
              dtype='datetime64[ns]', freq=None)
```

Or,

```
dd=['04/05/2012','02/12/2013','05/06/2018']
list(pd.to_datetime(dd))
```

o/p:

```
[Timestamp('2012-04-05 00:00:00'),
 Timestamp('2013-02-12 00:00:00'),
 Timestamp('2018-05-06 00:00:00')]
```

Or,

```
dd=['04/05/2012','15/12/2013','05/06/2018']
list(pd.to_datetime(dd,dayfirst=True))
```

o/p:

```
[Timestamp('2012-05-04 00:00:00'),
 Timestamp('2013-12-15 00:00:00'),
 Timestamp('2018-06-05 00:00:00')]
```

Or,

```
dd=['4 th july 2015','13 th june 1994','15 th june 2012']
list(pd.to_datetime(dd))
```

o/p:

```
[Timestamp('2015-07-04 00:00:00'),
 Timestamp('1994-06-13 00:00:00'),
 Timestamp('2012-06-15 00:00:00')]
```

Or,

```
dd=['4 th july 2015','13 th june 1994','15 th june 2012']
list(pd.to_datetime(dd,dayfirst=True))
```

o/p:

```
[Timestamp('2015-07-04 00:00:00'),
 Timestamp('1994-06-13 00:00:00'),
 Timestamp('2012-06-15 00:00:00')]
```

Or,

```
import pandas as pd
pr=pd.Period('2012',freq="D")
pr.asfreq('D',"end")
pr
```

o/p:

```
Period('2012-01-01', 'D')
```

Or,

```
pr = pd.Period('2012', freq='M')
pr.asfreq('D', 'start')
pr
```

o/p:

```
Period('2012-01', 'M')
```

Or,

```
pr = pd.Period('2012', freq='M')
asf=pr.asfreq('D', 'start')
asf
```

o/p:

```
Period('2012-01-01', 'D')
```

Or,\

```
pr=pd.Period("2012",freq='A')
pr+1
```

o/p:

```
Period('2013', 'A-DEC')
```

Or,

```
pr=pd.Period("2012",freq='A')
prMonth=pr.asfreq('M')
prMonth
```

o/p:

```
Period('2012-12', 'M')
```

Or,

```
pr=pd.Period("2012",freq='A')
prMonth=pr.asfreq('M')
prMonth+1
```

o/p:

```
Period('2013-01', 'M')
```

Or,

```
pr=pd.Period("2012",freq='A')
prMonth=pr.asfreq('M')
prMonth-1
```

o/p:

```
Period('2012-11', 'M')
```

Or,

```
prg=pd.period_range('2012','2015',freq='A')
prg
```

o/p:

```
PeriodIndex(['2012', '2013', '2014', '2015'], dtype='period[A-DEC]',
freq='A-DEC')
```

Or,

```
import numpy as np
import pandas as pd
data=pd.Series(np.random.rand(len(prg)),index=prg)
data
```

o/p:

```
2012    0.200001
2013    0.065251
2014    0.537684
2015    0.173713
Freq: A-DEC, dtype: float64
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
dt=pd.to_datetime(dd)
dt
```

o/p:

```
DatetimeIndex(['2012-02-05', '2013-03-06', '2015-01-03'],
dtype='datetime64[ns]', freq=None)
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
dt=pd.to_datetime(dd)
prd=dt.to_period(freq="M")
prd
```

o/p:

```
PeriodIndex(['2012-02', '2013-03', '2015-01'], dtype='period[M]',
freq='M')
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
```

```
dt=pd.to_datetime(dd)
prd=dt.to_period(freq="M")
asfm=prd.asfreq('D')
asfm
```

o/p:

```
PeriodIndex(['2012-02-29', '2013-03-31', '2015-01-31'], dtype='period[D]',
freq='D')
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
dt=pd.to_datetime(dd)
prd=dt.to_period(freq="M")
asfy=prd.asfreq('Y')
asfy
```

o/p:

```
PeriodIndex(['2012', '2013', '2015'], dtype='period[A-DEC]', freq='A-DEC')
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
dt=pd.to_datetime(dd)
prd=dt.to_period(freq="M")
ff=prd.to_timestamp()
ff
```

o/p:

```
DatetimeIndex(['2012-02-01', '2013-03-01', '2015-01-01'],
dtype='datetime64[ns]', freq=None)
```

Or,

```
dd=['02-05-2012','03-06-2013','01-03-2015']
dt=pd.to_datetime(dd)
prd=dt.to_period(freq="M")
ff=prd.to_timestamp(how='end')
ff
```

o/p:

```
DatetimeIndex(['2012-02-29 23:59:59.999999999',
                '2013-03-31 23:59:59.999999999',
                '2015-01-31 23:59:59.999999999'],
dtype='datetime64[ns]', freq=None)
```



or,

```
pd.Timedelta('3 days')
```

o/p:

```
Timedelta('3 days 00:00:00')
```

Or,

```
pd.Timedelta('3M')
```

o/p:

```
Timedelta('0 days 00:03:00')
```

Or,

```
pd.Timedelta('4 days 3M')
```

o/p:

```
Timedelta('4 days 00:03:00')
```

Or,

```
rng=pd.date_range(start="13 th june 2012 10:15",end="13 th june  
2018",freq='12 H')
```

o/p:

```
DatetimeIndex(['2012-06-13 10:15:00', '2012-06-13 22:15:00',  
               '2012-06-14 10:15:00', '2012-06-14 22:15:00',  
               '2012-06-15 10:15:00', '2012-06-15 22:15:00',  
               '2012-06-16 10:15:00', '2012-06-16 22:15:00',  
               '2012-06-17 10:15:00', '2012-06-17 22:15:00',  
               ...,  
               '2018-06-08 10:15:00', '2018-06-08 22:15:00',  
               '2018-06-09 10:15:00', '2018-06-09 22:15:00',  
               '2018-06-10 10:15:00', '2018-06-10 22:15:00',  
               '2018-06-11 10:15:00', '2018-06-11 22:15:00',  
               '2018-06-12 10:15:00', '2018-06-12 22:15:00'],  
              dtype='datetime64[ns]', length=4382, freq='12H')
```

or,

```
rng=pd.date_range(start="13 th june 2012 10:15",end="13 th june  
2018",freq='12 H')  
rng+pd.Timedelta('1 days')
```

o/p:

```
DatetimeIndex(['2012-06-14 10:15:00', '2012-06-14 22:15:00',
               '2012-06-15 10:15:00', '2012-06-15 22:15:00',
               '2012-06-16 10:15:00', '2012-06-16 22:15:00',
               '2012-06-17 10:15:00', '2012-06-17 22:15:00',
               '2012-06-18 10:15:00', '2012-06-18 22:15:00',
               ...,
               '2018-06-09 10:15:00', '2018-06-09 22:15:00',
               '2018-06-10 10:15:00', '2018-06-10 22:15:00',
               '2018-06-11 10:15:00', '2018-06-11 22:15:00',
               '2018-06-12 10:15:00', '2018-06-12 22:15:00',
               '2018-06-13 10:15:00', '2018-06-13 22:15:00'],
              dtype='datetime64[ns]', length=4382, freq='12H')
```

or,

```
date=pd.date_range('12 th june 2012','5 th october 2012',freq="M")
date
```

o/p:

```
DatetimeIndex(['2012-06-30', '2012-07-31', '2012-08-31', '2012-09-30'],
              dtype='datetime64[ns]', freq='M')
```

Or,

```
date=pd.date_range('12 th june 2012','5 th october 2012',freq="M")
len(date)
```

o/p:

4

Or,

```
date=pd.date_range('12 th june 2012','5 th october 2012',freq="M")
ass=pd.Series([100,110,120,130],index=date)
ass
```

o/p:

```
2012-06-30    100
2012-07-31    110
2012-08-31    120
2012-09-30    130
Freq: M, dtype: int64
```

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
```

```
print(ass)
print('\n')
print(ass[2])
```

o/p:

```
2012-06-30    100
2012-07-31    110
2012-08-31    120
2012-09-30    130
Freq: M, dtype: int64
```

120

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
print(ass)
print('\n')
print(ass.index[2])
```

o/p:

```
2012-06-30    100
2012-07-31    110
2012-08-31    120
2012-09-30    130
Freq: M, dtype: int64
```

2012-08-31 00:00:00

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
print(ass)
print('\n')
idx=ass.index[2]
print(ass[idx])
```

o/p:

```
2012-06-30    100
2012-07-31    110
2012-08-31    120
2012-09-30    130
Freq: M, dtype: int64
```

120

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
dss=pd.Series([105,108,115,124],index=date)
sat=pd.DataFrame({"Aukland":ass,'Delhi':dss})
print(sat)
```

o/p:

	Aukland	Delhi
2012-06-30	100	105
2012-07-31	110	108
2012-08-31	120	115
2012-09-30	130	124

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
dss=pd.Series([105,108,115,124],index=date)
sat=pd.DataFrame({"Aukland":ass,'Delhi':dss})
sat["diff"]=sat["Aukland"]-sat["Delhi"]
print(sat)
```

o/p:

	Aukland	Delhi	diff
2012-06-30	100	105	-5
2012-07-31	110	108	2
2012-08-31	120	115	5
2012-09-30	130	124	6

Or,

```
import pandas as pd
date=pd.date_range('13 th june 2012','25 th october 2012',freq='M')
ass=pd.Series([100,110,120,130],index=date)
dss=pd.Series([105,108,115,124],index=date)
sat=pd.DataFrame({"Aukland":ass,'Delhi':dss})
sat["diff"]=sat["Aukland"]-sat["Delhi"]
del sat["diff"]
print(sat)
```

o/p:

	Aukland	Delhi
2012-06-30	100	105
2012-07-31	110	108
2012-08-31	120	115
2012-09-30	130	124

Or,

```
stocks=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv')
print(stocks.head())
d=stocks.date[0]
```

```
print('\n')
print(d)
```

o/p:

	Unnamed: 0	date	AA	GE	IBM	MSFT
0	0	2/1/1990 0:00	4.98	2.87	16.79	0.51
1	1	2/2/1990 0:00	5.04	2.87	16.89	0.51
2	2	2/5/1990 0:00	5.07	2.87	17.32	0.51
3	3	2/6/1990 0:00	5.01	2.88	17.56	0.51
4	4	2/7/1990 0:00	5.04	2.91	17.93	0.51

2/1/1990 0:00

Or,

```
df=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',
parse_dates=['date'],index_col='date')
print(df.head())
del df['Unnamed: 0']
print('\n')
print(df.head())
```

o/p:

	Unnamed: 0	AA	GE	IBM	MSFT
date					
1990-02-01	0	4.98	2.87	16.79	0.51
1990-02-02	1	5.04	2.87	16.89	0.51
1990-02-05	2	5.07	2.87	17.32	0.51
1990-02-06	3	5.01	2.88	17.56	0.51
1990-02-07	4	5.04	2.91	17.93	0.51

	AA	GE	IBM	MSFT
date				
1990-02-01	4.98	2.87	16.79	0.51
1990-02-02	5.04	2.87	16.89	0.51
1990-02-05	5.07	2.87	17.32	0.51
1990-02-06	5.01	2.88	17.56	0.51
1990-02-07	5.04	2.91	17.93	0.51

Or,

```
df=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',
parse_dates=['date'],index_col='date')
print(df.head())
del df['Unnamed: 0']
print('\n')
print(df.head())
df.index.name
```

o/p:

	Unnamed: 0	AA	GE	IBM	MSFT
date					
1990-02-01	0	4.98	2.87	16.79	0.51
1990-02-02	1	5.04	2.87	16.89	0.51
1990-02-05	2	5.07	2.87	17.32	0.51
1990-02-06	3	5.01	2.88	17.56	0.51
1990-02-07	4	5.04	2.91	17.93	0.51

	AA	GE	IBM	MSFT
date				
1990-02-01	4.98	2.87	16.79	0.51
1990-02-02	5.04	2.87	16.89	0.51
1990-02-05	5.07	2.87	17.32	0.51
1990-02-06	5.01	2.88	17.56	0.51
1990-02-07	5.04	2.91	17.93	0.51

Out[33]:

'date'

Or,

```
import pandas as pd
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'
],index_col="date")
print(dd.loc['1990-Feb-01':'1990-Feb-06'])
```

o/p:

	Unnamed: 0	AA	GE	IBM	MSFT
date					
1990-02-01	0	4.98	2.87	16.79	0.51
1990-02-02	1	5.04	2.87	16.89	0.51
1990-02-05	2	5.07	2.87	17.32	0.51
1990-02-06	3	5.01	2.88	17.56	0.51

Or,

```
import pandas as pd
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
print(dd.loc['1990-Feb'].head())
```

o/p:

Unnamed: 0	AA	GE	IBM	MSFT	
date					
1990-02-01	0	4.98	2.87	16.79	0.51
1990-02-02	1	5.04	2.87	16.89	0.51
1990-02-05	2	5.07	2.87	17.32	0.51
1990-02-06	3	5.01	2.88	17.56	0.51
1990-02-07	4	5.04	2.91	17.93	0.51

Or,

```
from datetime import datetime ,timedelta
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
start=datetime(1990,2,1)
mm=dd.loc[start:start+pd.offsets.Day(5)]
print(mm)
```

o/p:

Unnamed: 0	AA	GE	IBM	MSFT	
date					
1990-02-01	0	4.98	2.87	16.79	0.51
1990-02-02	1	5.04	2.87	16.89	0.51
1990-02-05	2	5.07	2.87	17.32	0.51
1990-02-06	3	5.01	2.88	17.56	0.51

Or,

```
from datetime import datetime ,timedelta
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
print(dd.head())
print('\n')
print(dd.index[0])
print('\n')
print(dd.loc[dd.index[0]])
```

o/p:

Unnamed: 0	AA	GE	IBM	MSFT
------------	----	----	-----	------

```

date
1990-02-01      0  4.98  2.87  16.79  0.51
1990-02-02      1  5.04  2.87  16.89  0.51
1990-02-05      2  5.07  2.87  17.32  0.51
1990-02-06      3  5.01  2.88  17.56  0.51
1990-02-07      4  5.04  2.91  17.93  0.51

```

```
1990-02-01 00:00:00
```

```

Unnamed: 0      0.00
AA              4.98
GE              2.87
IBM            16.79
MSFT           0.51
Name: 1990-02-01 00:00:00, dtype: float64

```

```

o
r,

```

```

from datetime import datetime ,timedelta
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
print(dd.head())
print('\n')
print(dd["AA"])

```

```
o/p:
```

```

           Unnamed: 0    AA    GE    IBM  MSFT
date
1990-02-01      0  4.98  2.87  16.79  0.51
1990-02-02      1  5.04  2.87  16.89  0.51
1990-02-05      2  5.07  2.87  17.32  0.51
1990-02-06      3  5.01  2.88  17.56  0.51
1990-02-07      4  5.04  2.91  17.93  0.51

```

```

date
1990-02-01      4.98
1990-02-02      5.04
1990-02-05      5.07
1990-02-06      5.01
1990-02-07      5.04
...
2011-10-10     10.09
2011-10-11     10.30
2011-10-12     10.05
2011-10-13     10.10
2011-10-14     10.26
Name: AA, Length: 5472, dtype: float64

```

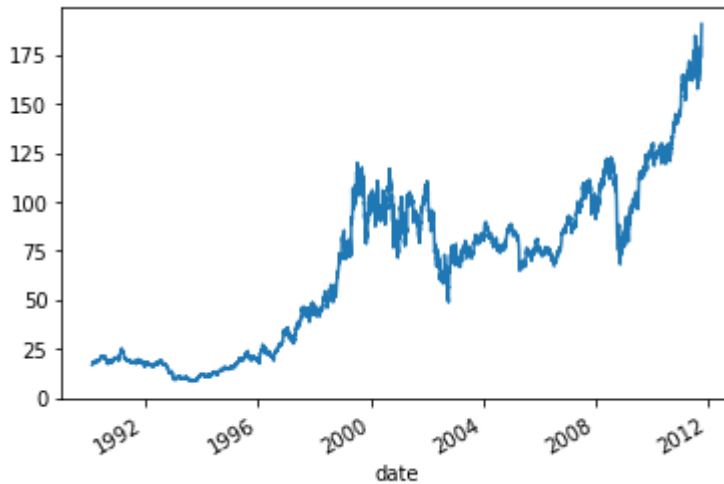


```

Or,
import matplotlib.pyplot as plt
import pandas as pd
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
dd.IBM.plot()
plt.show()

```

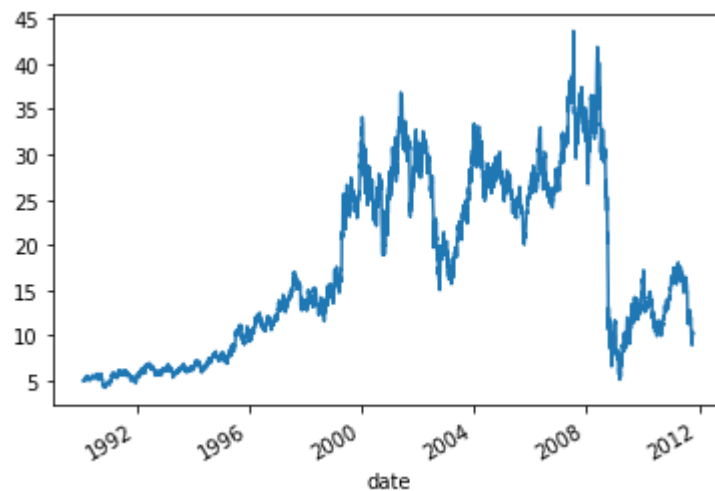
o/p:



```

Or,
import matplotlib.pyplot as plt
import pandas as pd
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
dd.AA.plot()

```



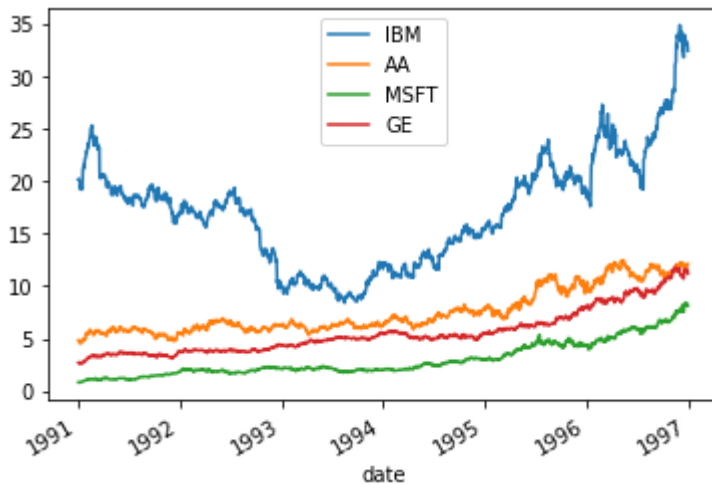
```
plt.show()
```

```

or,
import pandas as pd
import matplotlib.pyplot as plt
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
dd.loc['1991':'1996',['IBM','AA','MSFT','GE']].plot()
plt.show()

```

o/p:



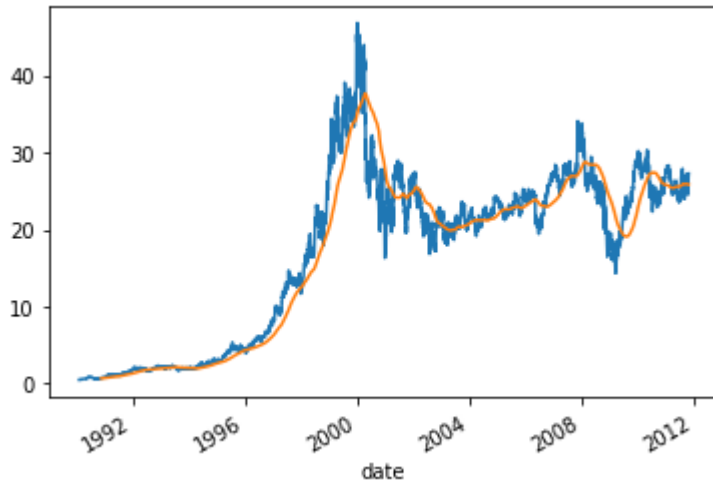
Or,

```

import pandas as pd
import matplotlib.pyplot as plt
import os
dd=pd.read_csv(r'C:\Users\Shubhamay\Documents\stocks.csv',parse_dates=['date'],index_col="date")
dd.MSFT.plot()
dd.MSFT.rolling(window=200).mean().plot()
plt.show()

```

o/p:



Read multiple file

```
import pandas as pd
pp=pd.read_csv(r'C:\Users\Shubhamay\Documents\yob1880.txt')
print(pp.head())
```

o/p:

	Mary	F	7065
0	Anna	F	2604
1	Emma	F	2003
2	Elizabeth	F	1939
3	Minnie	F	1746
4	Margaret	F	1578

Or,

```
import pandas as pd
pp=pd.read_csv(r'C:\Users\Shubhamay\Documents\yob1882.txt')
print(len(pp))
```

o/p:

2126

Or,

```
import pandas as pd
pp=pd.read_csv(r'C:\Users\Shubhamay\Documents\yob1882.txt',names=['Name','Gender','Birth Count'])
print(pp.head())
```

o/p:

	Name	Gender	Birth	Count
0	Mary	F		8149
1	Anna	F		3143
2	Emma	F		2303
3	Elizabeth	F		2187
4	Minnie	F		2004

```
Or,
import pandas as pd
pp=pd.read_csv(r'C:\Users\Shubhamay\Documents\yob1882.txt',names=['Name','Gender','birthcount'])
print(pp.groupby('Gender').birthcount.sum())
```

o/p

```
Gender
F    107851
M    113687
Name: birthcount, dtype: int64
```

Or,

```
years=range(1880,1887)
for year in years:
    print(year)
```

o/p:

```
1880
1881
1882
1883
1884
1885
1886
```

Or,

```
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
print(pieces)
```

o/p:

```
[          name gender  birthcount
```

0	Mary	F	7065
1	Anna	F	2604
2	Emma	F	2003
3	Elizabeth	F	1939
4	Minnie	F	1746
...	...	...	...
1995	Woodie	M	5
1996	Worthy	M	5
1997	Wright	M	5
1998	York	M	5
1999	Zachariah	M	5

[2000 rows x 3 columns],				name	gender	birthcount
0	Mary	F	6919			
1	Anna	F	2698			
2	Emma	F	2034			
3	Elizabeth	F	1852			
4	Margaret	F	1658			
...	...	...	...			
1930	William	M	5			
1931	Wilton	M	5			
1932	Wing	M	5			
1933	Wood	M	5			
1934	Wright	M	5			

[1935 rows x 3 columns],				name	gender	birthcount
0	Mary	F	8149			
1	Anna	F	3143			
2	Emma	F	2303			
3	Elizabeth	F	2187			
4	Minnie	F	2004			
...	...	...	...			
2122	Wirt	M	5			
2123	Woodson	M	5			
2124	Woody	M	5			
2125	Worley	M	5			
2126	Zed	M	5			

[2127 rows x 3 columns],				name	gender	birthcount
0	Mary	F	8012			
1	Anna	F	3306			
2	Emma	F	2367			
3	Elizabeth	F	2255			
4	Minnie	F	2035			
...	...	...	...			
2079	Titus	M	5			
2080	Toney	M	5			
2081	Verna	M	5			
2082	Winnie	M	5			
2083	Winthrop	M	5			

[2084 rows x 3 columns],				name	gender	birthcount
0	Mary	F	9217			
1	Anna	F	3860			

2	Emma	F	2587
3	Elizabeth	F	2549
4	Minnie	F	2243
...	...	...	...
2292	Williard	M	5
2293	Willy	M	5
2294	Winston	M	5
2295	York	M	5
2296	Zachariah	M	5

	name	gender	birthcount
[2297 rows x 3 columns],			
0	Mary	F	9128
1	Anna	F	3994
2	Emma	F	2728
3	Elizabeth	F	2582
4	Margaret	F	2204
...	...	...	...
2289	Wallie	M	5
2290	Willian	M	5
2291	Wirt	M	5
2292	Yee	M	5
2293	Zeb	M	5

	name	gender	birthcount
[2294 rows x 3 columns],			
0	Mary	F	9891
1	Anna	F	4283
2	Emma	F	2764
3	Elizabeth	F	2680
4	Minnie	F	2372
...	...	...	...
2387	Wood	M	5
2388	Worth	M	5
2389	Wright	M	5
2390	Xavier	M	5
2391	Yancy	M	5

[2392 rows x 3 columns]]

Or,

```

years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
    alldata=pd.concat(pieces,ignore_index=True)
print(alldata.head())

```

o/p:

	name	gender	birthcount	year
0	Mary	F	7065	1880
1	Anna	F	2604	1880
2	Emma	F	2003	1880
3	Elizabeth	F	1939	1880
4	Minnie	F	1746	1880

Or,

```
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
```

```
alldata=pd.concat(pieces,ignore_index=True)
print(alldata.head())
```

o/p:

	name	gender	birthcount
0	Mary	F	7065
1	Anna	F	2604
2	Emma	F	2003
3	Elizabeth	F	1939
4	Minnie	F	1746

Or,

```
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
alldata=pd.concat(pieces,ignore_index=True)
total_birth=alldata.pivot_table('birthcount',index=['year'],columns=['gender'],aggfunc=sum)
print(total_birth.head())
```

o/p:

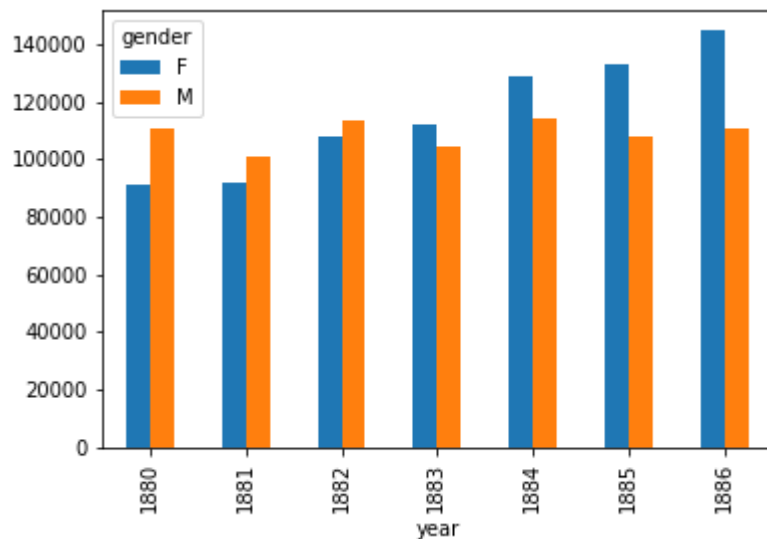
gender	F	M
year		
1880	90993	110493
1881	91955	100748
1882	107851	113687
1883	112322	104632

1884      129021   114445

Or,

```
import matplotlib.pyplot as plt
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
alldata=pd.concat(pieces,ignore_index=True)
total_birth=alldata.pivot_table('birthcount',index=['year'],columns=['gender'],aggfunc=sum)
total_birth.plot(kind='bar')
plt.show()
```

o/p:



Or,

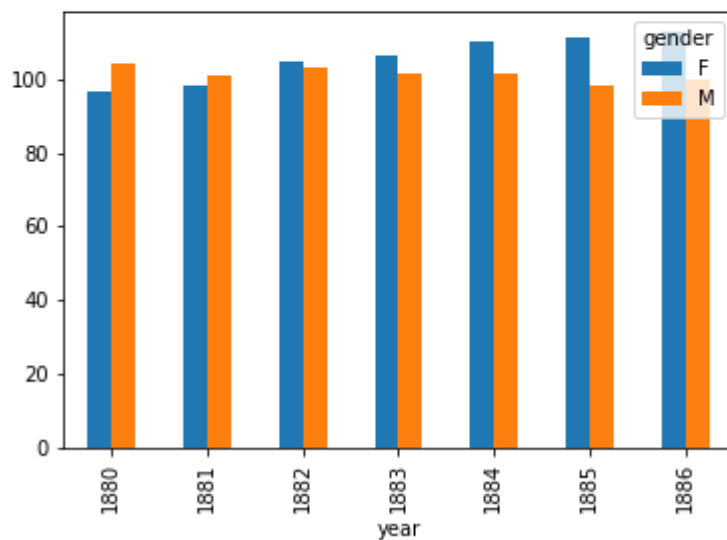
```
import matplotlib.pyplot as plt
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
alldata=pd.concat(pieces,ignore_index=True)
total_birth=alldata.pivot_table('birthcount',index=['year'],columns=['gender'])
print(total_birth)
```



```
total_birth.plot(kind='bar')
plt.show()
```

o/p:

gender	F	M
year		
1880	96.595541	104.435728
1881	98.033049	101.051153
1882	104.913424	103.445860
1883	106.567362	101.584466
1884	110.086177	101.728889
1885	111.157895	98.269827
1886	112.744150	99.806306



Or,

```
import matplotlib.pyplot as plt
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
alldata=pd.concat(pieces,ignore_index=True)
print(alldata.groupby(['year','gender']).sum().unstack('gender').head())
```

o/p:

	birthcount	
gender	F	M
year		

1880	90993	110493
1881	91955	100748
1882	107851	113687
1883	112322	104632
1884	129021	114445

Or,

```
import matplotlib.pyplot as plt
years = range(1880, 1887)
pieces = []
columns = ['name', 'gender', 'birthcount']
for year in years:
    path = r'C:\Users\Shubhamay\Documents\yob{}.txt'.format(year)
    df = pd.read_csv(path, names=columns)
    pieces.append(df)
    df['year']=year
alldata=pd.concat(pieces,ignore_index=True)
def addprop(group):
    births=group.birthcount
    group['prop']=births/births.sum()
    return group
print(alldata.groupby(['year','gender']).apply(addprop).head())
```

o/p:

	name	gender	birthcount	year	prop
0	Mary	F	7065	1880	0.077643
1	Anna	F	2604	1880	0.028618
2	Emma	F	2003	1880	0.022013
3	Elizabeth	F	1939	1880	0.021309
4	Minnie	F	1746	1880	0.019188