Data Collection,Modeling And Compilation ¶

In [1]:

```python
my_dict={'name':["a","b","c","d","e","f","g"],'age':[20,27,35,45,55,43,35],'designation':['
import pandas as pd
import numpy as np
df=pd.DataFrame(my_dict)
df
```

Out[1]:

	name	age	designation
0	a	20	VP
1	b	27	CEO
2	c	35	CFO
3	d	45	VP
4	e	55	VP
5	f	43	CEO
6	g	35	MD

In [2]:

```python
df.to_csv("example")
df
```

Out[2]:

	name	age	designation
0	a	20	VP
1	b	27	CEO
2	c	35	CFO
3	d	45	VP
4	e	55	VP
5	f	43	CEO
6	g	35	MD

```python
df_csv=pd.read_csv("example")
df_csv
```

Out[3]:

	Unnamed: 0	name	age	designation
0	0	a	20	VP
1	1	b	27	CEO
2	2	c	35	CFO
3	3	d	45	VP
4	4	e	55	VP
5	5	f	43	CEO
6	6	g	35	MD

In [4]:

```python
df.to_csv('ex',index=False)
df_csv=pd.read_csv("ex")
df_csv
```

Out[4]:

	name	age	designation
0	a	20	VP
1	b	27	CEO
2	c	35	CFO
3	d	45	VP
4	e	55	VP
5	f	43	CEO
6	g	35	MD

In [5]:

```python
import pandas as pd
df=pd.read_csv("student-mat - student-mat.csv",header=None)
df.head()
```

Out[5]:

	0	1	2	3	4	5	6	7	8	9	...	23	
0	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetir
1	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	
2	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	
3	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	
4	GP	F	15	U	GT3	T	4	2	health	services	...	3	

5 rows × 33 columns

In [6]:

```python
import pandas as pd
location = "student-mat - student-mat.csv"
df=pd.read_csv(location,names=['Rollno','Name','Grades'])
df.columns=['Rollno','Name','Grades']
df.head()
```

Out[6]:

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardia
		18	U	GT3	A	4	4	at_home	teacher	course	mothe
		17	U	GT3	T	1	1	at_home	other	course	fathe
GP	F			LE3	T	1	1	at_home	other	other	mothe
		15	U	GT3	T	4	2	health	services	home	mothe

```python
import pandas as pd
names=['bob','jessica','mary','john','mel']
grades=[76,95,77,78,99]
bsdegrees=[1,1,0,0,1]
msdegrees=[2,1,0,0,0]
phddgrees=[0,1,0,0,0]
degrees=zip(names,grades,bsdegrees,msdegrees,phddgrees)
columns=['Names','Grades','BS','MS','Phd']
df=pd.DataFrame(data=degrees,columns=columns)
df
```

Out[7]:

	Names	Grades	BS	MS	Phd
0	bob	76	1	2	0
1	jessica	95	1	1	1
2	mary	77	0	0	0
3	john	78	0	0	0
4	mel	99	1	0	0

In [8]:

```python
import pandas as pd
location='gradedata.xlsx'
df=pd.read_excel("gradedata.xlsx")
df.columns=['first','last','sex','age','exer','hrs','grd','addr']
df.head()
```

Out[8]:

	first	last	sex	age	exer	hrs	grd	addr
0	Marcia	Pugh	female	17.0	3.0	10.0	82.4	7379 Highland Rd. , Dublin, GA 31021
1	Kadeem	Morrison	male	18.0	4.0	4.0	78.2	8 Bayport St. , Honolulu, HI 96815
2	Nash	Powell	male	18.0	5.0	9.0	79.3	Encino, CA 91316, 3 Lilac Street
3	Noelani	Wagner	female	14.0	2.0	7.0	83.2	Riverview, FL 33569, 9998 North Smith Dr.
4	Noelani	Cherry	female	18.0	4.0	15.0	87.4	97 SE. Ocean Street , Bethlehem, PA 18015

In [9]:

```python
import pandas as pd
names=['bob','jessica','mary','john','mel']
grades=[76,95,77,78,99]
gradelist=zip(names,grades)
df=pd.DataFrame(data=gradelist,columns=['names','grades'])
writer=pd.ExcelWriter('dataframe.xlsx',engine='xlsxwriter')
df.to_excel(writer,sheet_name='sheet1')
writer.save()
```

```python
import sqlite3
con = sqlite3.connect("C:/Users/MSCIT/Downloads/portal_mammals.sqlite")
cur = con.cursor()

for row in cur.execute('Select * from species;'):
    print(row)
con.close()
```

```
('AB', 'Amphispiza', 'bilineata', 'Bird')
('AH', 'Ammospermophilus', 'harrisi', 'Rodent')
('AS', 'Ammodramus', 'savannarum', 'Bird')
('BA', 'Baiomys', 'taylori', 'Rodent')
('CB', 'Campylorhynchus', 'brunneicapillus', 'Bird')
('CM', 'Calamospiza', 'melanocorys', 'Bird')
('CQ', 'Callipepla', 'squamata', 'Bird')
('CS', 'Crotalus', 'scutalatus', 'Reptile')
('CT', 'Cnemidophorus', 'tigris', 'Reptile')
('CU', 'Cnemidophorus', 'uniparens', 'Reptile')
('CV', 'Crotalus', 'viridis', 'Reptile')
('DM', 'Dipodomys', 'merriami', 'Rodent')
('DO', 'Dipodomys', 'ordii', 'Rodent')
('DS', 'Dipodomys', 'spectabilis', 'Rodent')
('DX', 'Dipodomys', 'sp.', 'Rodent')
('EO', 'Eumeces', 'obsoletus', 'Reptile')
('GS', 'Gambelia', 'silus', 'Reptile')
('NL', 'Neotoma', 'albigula', 'Rodent')
('NX', 'Neotoma', 'sp.', 'Rodent')
('OL', 'Onychomys', 'leucogaster', 'Rodent')
('OT', 'Onychomys', 'torridus', 'Rodent')
('OX', 'Onychomys', 'sp.', 'Rodent')
('PB', 'Chaetodipus', 'baileyi', 'Rodent')
('PC', 'Pipilo', 'chlorurus', 'Bird')
('PE', 'Peromyscus', 'eremicus', 'Rodent')
('PF', 'Perognathus', 'flavus', 'Rodent')
('PG', 'Pooecetes', 'gramineus', 'Bird')
('PH', 'Perognathus', 'hispidus', 'Rodent')
('PI', 'Chaetodipus', 'intermedius', 'Rodent')
('PL', 'Peromyscus', 'leucopus', 'Rodent')
('PM', 'Peromyscus', 'maniculatus', 'Rodent')
('PP', 'Chaetodipus', 'penicillatus', 'Rodent')
('PU', 'Pipilo', 'fuscus', 'Bird')
('PX', 'Chaetodipus', 'sp.', 'Rodent')
('RF', 'Reithrodontomys', 'fulvescens', 'Rodent')
('RM', 'Reithrodontomys', 'megalotis', 'Rodent')
('RO', 'Reithrodontomys', 'montanus', 'Rodent')
('RX', 'Reithrodontomys', 'sp.', 'Rodent')
('SA', 'Sylvilagus', 'audubonii', 'Rabbit')
('SB', 'Spizella', 'breweri', 'Bird')
('SC', 'Sceloporus', 'clarki', 'Reptile')
('SF', 'Sigmodon', 'fulviventer', 'Rodent')
('SH', 'Sigmodon', 'hispidus', 'Rodent')
('SO', 'Sigmodon', 'ochrognathus', 'Rodent')
('SS', 'Spermophilus', 'spilosoma', 'Rodent')
('ST', 'Spermophilus', 'tereticaudus', 'Rodent')
('SU', 'Sceloporus', 'undulatus', 'Reptile')
('SX', 'Sigmodon', 'sp.', 'Rodent')
('UL', 'Lizard', 'sp.', 'Reptile')
('UP', 'Pipilo', 'sp.', 'Bird')
```

```
('UR', 'Rodent', 'sp.', 'Rodent')
('US', 'Sparrow', 'sp.', 'Bird')
('ZL', 'Zonotrichia', 'leucophrys', 'Bird')
('ZM', 'Zenaida', 'macroura', 'Bird')
```

In [11]:

```python
import sqlite3
#create a SQL connection to our SQlite database
con=sqlite3.connect("C:/Users/MSCIT/Downloads/portal_mammals.sqlite")
cur=con.cursor()
#Return all result of query
cur.execute('Select plot_id From plots Where plot_type="Control"')
print(cur.fetchall())
#Return First Result of query
cur.execute('Select species From species Where taxa="Bird"')
print(cur.fetchone())
#Be sure to close the connection
con.close()
```

```
[(2,), (4,), (8,), (11,), (12,), (14,), (17,), (22,)]
('bilineata',)
```

In [12]:

```python
import pandas as pd
import sqlite3

#read sqlite query results into a pandas dataframe
con=sqlite3.connect("C:/Users/MSCIT/Downloads/portal_mammals.sqlite")
df=pd.read_sql_query("Select*From surveys",con)

#verify that result of Sql query is the dataframe
print(df.head())
con.close()
```

```
   record_id  month  day  year  plot_id species_id sex  hindfoot_length  \
0          1      7   16  1977        2         NL   M             32.0
1          2      7   16  1977        3         NL   M             33.0
2          3      7   16  1977        2         DM   F             37.0
3          4      7   16  1977        7         DM   M             36.0
4          5      7   16  1977        3         DM   M             35.0

   weight
0     NaN
1     NaN
2     NaN
3     NaN
4     NaN
```

In [13]:

```python
from pandas import DataFrame
Cars={'Brand':['Honda Civic','Toyota Corolla','ford Focus','audi A4'],'Price':[22000,25000,
      }
df=DataFrame(Cars,columns=['Brand','Price'])
print(df)
```

```
           Brand  Price
0     Honda Civic  22000
1  Toyota Corolla  25000
2      ford Focus  27000
3         audi A4  35000
```

In [19]:

```python
import sqlite3
conn=sqlite3.connect('TestDB12.db')
c=conn.cursor()
c.execute('create table Carss(Brand text,Price number)')
conn.commit()
```

In [20]:

```python
df.to_sql('cars21',conn,if_exists='replace',index=False)
df
```

Out[20]:

	Brand	Price
0	Honda Civic	22000
1	Toyota Corolla	25000
2	ford Focus	27000
3	audi A4	35000

In [21]:

```python
c.execute('''
SELECT Brand,max(Price) from cars21
''')
```

Out[21]:

```
<sqlite3.Cursor at 0x1c17c470b20>
```

In [22]:

```python
df=DataFrame(c.fetchall(),columns=['Brand','Price'])
df
```

Out[22]:

	Brand	Price
0	audi A4	35000

```python
import pandas as pd
import os
import sqlite3 as lite
from sqlalchemy import create_engine
```

```python
studentId=["rj101","rj150","rj134","rj07"]
SName=["Shubham","Vaishu","Bittu","Vishal","Suraj"]
LName=["Singh","Yadav","Pawar","Shukla","Mishra"]
Department=["Bsc","Bcom","BMM","Msc","BBI"]
Email=["shum102@gmail.com","Vaishu12@gmail.com","bittu0908@gmail.com","vishal123@gmail.com"
```

```python
studata=zip(studentId,SName,LName,Department,Email)
```

```python
df=pd.DataFrame(data=studata,columns = ['studentId','SName','LName','Department','Email'])
df
```

	studentId	SName	LName	Department	Email
0	rj101	Shubham	Singh	Bsc	shum102@gmail.com
1	rj150	Vaishu	Yadav	Bcom	Vaishu12@gmail.com
2	rj134	Bittu	Pawar	BMM	bittu0908@gmail.com
3	rj07	Vishal	Shukla	Msc	vishal123@gmail.com

```python
df.to_csv('Csv ex')
df
```

	studentId	SName	LName	Department	Email
0	rj101	Shubham	Singh	Bsc	shum102@gmail.com
1	rj150	Vaishu	Yadav	Bcom	Vaishu12@gmail.com
2	rj134	Bittu	Pawar	BMM	bittu0908@gmail.com
3	rj07	Vishal	Shukla	Msc	vishal123@gmail.com

```python
writer=pd.ExcelWriter('dataframe1.xlsx',engine='xlsxwriter')
df.to_excel(writer,sheet_name='sheet1')
writer.save()
```

```
df.to_sql('studata',conn,if_exists='replace',index=False)
df
```

Out[29]:

	studentId	SName	LName	Department	Email
0	rj101	Shubham	Singh	Bsc	shum102@gmail.com
1	rj150	Vaishu	Yadav	Bcom	Vaishu12@gmail.com
2	rj134	Bittu	Pawar	BMM	bittu0908@gmail.com
3	rj07	Vishal	Shukla	Msc	vishal123@gmail.com

DATA PREPERATION

In [3]:

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

In [5]:

```
state=pd.read_csv("C:/Users/MSCIT/Downloads/US_violent_crime - US_violent_crime.csv")
state.head()
```

Out[5]:

	State	Murder	Assault	UrbanPop	Rape
0	Alabama	13.2	236	58	21.2
1	Alaska	10.0	263	48	44.5
2	Arizona	8.1	294	80	31.0
3	Arkansas	8.8	190	50	19.5
4	California	9.0	276	91	40.6

```
def some_func(y):
    return y*2
state.apply(some_func) # update each entry of data without any loop
```

Out[6]:

	State	Murder	Assault	UrbanPop	Rape
0	AlabamaAlabama	26.4	472	116	42.4
1	AlaskaAlaska	20.0	526	96	89.0
2	ArizonaArizona	16.2	588	160	62.0
3	ArkansasArkansas	17.6	380	100	39.0
4	CaliforniaCalifornia	18.0	552	182	81.2
5	ColoradoColorado	15.8	408	156	77.4
6	ConnecticutConnecticut	6.6	220	154	22.2
7	DelawareDelaware	11.8	476	144	31.6
8	FloridaFlorida	30.8	670	160	63.8
9	GeorgiaGeorgia	34.8	422	120	51.6
10	HawaiiHawaii	10.6	92	166	40.4
11	IdahoIdaho	5.2	240	108	28.4
12	IllinoisIllinois	20.8	498	166	48.0
13	IndianaIndiana	14.4	226	130	42.0
14	IowaIowa	4.4	112	114	22.6
15	KansasKansas	12.0	230	132	36.0
16	KentuckyKentucky	19.4	218	104	32.6
17	LouisianaLouisiana	30.8	498	132	44.4
18	MaineMaine	4.2	166	102	15.6
19	MarylandMaryland	22.6	600	134	55.6
20	MassachusettsMassachusetts	8.8	298	170	32.6
21	MichiganMichigan	24.2	510	148	70.2
22	MinnesotaMinnesota	5.4	144	132	29.8
23	MississippiMississippi	32.2	518	88	34.2
24	MissouriMissouri	18.0	356	140	56.4
25	MontanaMontana	12.0	218	106	32.8
26	NebraskaNebraska	8.6	204	124	33.0
27	NevadaNevada	24.4	504	162	92.0
28	New HampshireNew Hampshire	4.2	114	112	19.0
29	New JerseyNew Jersey	14.8	318	178	37.6
30	New MexicoNew Mexico	22.8	570	140	64.2
31	New YorkNew York	22.2	508	172	52.2
32	North CarolinaNorth Carolina	26.0	674	90	32.2
33	North DakotaNorth Dakota	1.6	90	88	14.6

	State	Murder	Assault	UrbanPop	Rape
34	OhioOhio	14.6	240	150	42.8
35	OklahomaOklahoma	13.2	302	136	40.0
36	OregonOregon	9.8	318	134	58.6
37	PennsylvaniaPennsylvania	12.6	212	144	29.8
38	Rhode IslandRhode Island	6.8	348	174	16.6
39	South CarolinaSouth Carolina	28.8	558	96	45.0
40	South DakotaSouth Dakota	7.6	172	90	25.6
41	TennesseeTennessee	26.4	376	118	53.8
42	TexasTexas	25.4	402	160	51.0
43	UtahUtah	6.4	240	160	45.8
44	VermontVermont	4.4	96	64	22.4
45	VirginiaVirginia	17.0	312	126	41.4
46	WashingtonWashington	8.0	290	146	52.4
47	West VirginiaWest Virginia	11.4	162	78	18.6
48	WisconsinWisconsin	5.2	106	132	21.6
49	WyomingWyoming	13.6	322	120	31.2

```
state.apply(lambda n:n*2) #lambda also works the same
```

	State	Murder	Assault	UrbanPop	Rape
0	AlabamaAlabama	26.4	472	116	42.4
1	AlaskaAlaska	20.0	526	96	89.0
2	ArizonaArizona	16.2	588	160	62.0
3	ArkansasArkansas	17.6	380	100	39.0
4	CaliforniaCalifornia	18.0	552	182	81.2
5	ColoradoColorado	15.8	408	156	77.4
6	ConnecticutConnecticut	6.6	220	154	22.2
7	DelawareDelaware	11.8	476	144	31.6
8	FloridaFlorida	30.8	670	160	63.8
9	GeorgiaGeorgia	34.8	422	120	51.6
10	HawaiiHawaii	10.6	92	166	40.4
11	IdahoIdaho	5.2	240	108	28.4
12	IllinoisIllinois	20.8	498	166	48.0
13	IndianaIndiana	14.4	226	130	42.0
14	IowaIowa	4.4	112	114	22.6
15	KansasKansas	12.0	230	132	36.0
16	KentuckyKentucky	19.4	218	104	32.6
17	LouisianaLouisiana	30.8	498	132	44.4
18	MaineMaine	4.2	166	102	15.6
19	MarylandMaryland	22.6	600	134	55.6
20	MassachusettsMassachusetts	8.8	298	170	32.6
21	MichiganMichigan	24.2	510	148	70.2
22	MinnesotaMinnesota	5.4	144	132	29.8
23	MississippiMississippi	32.2	518	88	34.2
24	MissouriMissouri	18.0	356	140	56.4
25	MontanaMontana	12.0	218	106	32.8
26	NebraskaNebraska	8.6	204	124	33.0
27	NevadaNevada	24.4	504	162	92.0
28	New HampshireNew Hampshire	4.2	114	112	19.0
29	New JerseyNew Jersey	14.8	318	178	37.6
30	New MexicoNew Mexico	22.8	570	140	64.2
31	New YorkNew York	22.2	508	172	52.2
32	North CarolinaNorth Carolina	26.0	674	90	32.2
33	North DakotaNorth Dakota	1.6	90	88	14.6
34	OhioOhio	14.6	240	150	42.8

	State	Murder	Assault	UrbanPop	Rape
35	OklahomaOklahoma	13.2	302	136	40.0
36	OregonOregon	9.8	318	134	58.6
37	PennsylvaniaPennsylvania	12.6	212	144	29.8
38	Rhode IslandRhode Island	6.8	348	174	16.6
39	South CarolinaSouth Carolina	28.8	558	96	45.0
40	South DakotaSouth Dakota	7.6	172	90	25.6
41	TennesseeTennessee	26.4	376	118	53.8
42	TexasTexas	25.4	402	160	51.0
43	UtahUtah	6.4	240	160	45.8
44	VermontVermont	4.4	96	64	22.4
45	VirginiaVirginia	17.0	312	126	41.4
46	WashingtonWashington	8.0	290	146	52.4
47	West VirginiaWest Virginia	11.4	162	78	18.6
48	WisconsinWisconsin	5.2	106	132	21.6
49	WyomingWyoming	13.6	322	120	31.2

```
state.transform(func=lambda x:x*10)
```

	State	Murder	Assault	UrbanPop	Rape
0	AlabamaAlabamaAlabamaAlabamaAlabamaAlabamaAlab...	132.0	2360	580	212.0
1	AlaskaAlaskaAlaskaAlaskaAlaskaAlaskaAlaskaAlas...	100.0	2630	480	445.0
2	ArizonaArizonaArizonaArizonaArizonaArizonaAriz...	81.0	2940	800	310.0
3	ArkansasArkansasArkansasArkansasArkansasArkans...	88.0	1900	500	195.0
4	CaliforniaCaliforniaCaliforniaCaliforniaCalifo...	90.0	2760	910	406.0
5	ColoradoColoradoColoradoColoradoColoradoColora...	79.0	2040	780	387.0
6	ConnecticutConnecticutConnecticutConnecticutCo...	33.0	1100	770	111.0
7	DelawareDelawareDelawareDelawareDelawareDelawa...	59.0	2380	720	158.0
8	FloridaFloridaFloridaFloridaFloridaFloridaFlor...	154.0	3350	800	319.0
9	GeorgiaGeorgiaGeorgiaGeorgiaGeorgiaGeorgiaGeor...	174.0	2110	600	258.0
10	HawaiiHawaiiHawaiiHawaiiHawaiiHawaiiHawaiiHawa...	53.0	460	830	202.0
11	IdahoIdahoIdahoIdahoIdahoIdahoIdahoIdahoIdahoI...	26.0	1200	540	142.0
12	IllinoisIllinoisIllinoisIllinoisIllinoisIllino...	104.0	2490	830	240.0
13	IndianaIndianaIndianaIndianaIndianaIndianaIndi...	72.0	1130	650	210.0
14	IowaIowaIowaIowaIowaIowaIowaIowaIowaIowa	22.0	560	570	113.0
15	KansasKansasKansasKansasKansasKansasKansasKans...	60.0	1150	660	180.0
16	KentuckyKentuckyKentuckyKentuckyKentuckyKentuc...	97.0	1090	520	163.0
17	LouisianaLouisianaLouisianaLouisianaLouisianaL...	154.0	2490	660	222.0
18	MaineMaineMaineMaineMaineMaineMaineMaineMaineM...	21.0	830	510	78.0
19	MarylandMarylandMarylandMarylandMarylandMaryla...	113.0	3000	670	278.0
20	MassachusettsMassachusettsMassachusettsMassach...	44.0	1490	850	163.0
21	MichiganMichiganMichiganMichiganMichiganMichig...	121.0	2550	740	351.0
22	MinnesotaMinnesotaMinnesotaMinnesotaMinnesotaM...	27.0	720	660	149.0
23	MississippiMississippiMississippiMississippiMi...	161.0	2590	440	171.0
24	MissouriMissouriMissouriMissouriMissouriMissou...	90.0	1780	700	282.0
25	MontanaMontanaMontanaMontanaMontanaMontanaMont...	60.0	1090	530	164.0
26	NebraskaNebraskaNebraskaNebraskaNebraskaNebras...	43.0	1020	620	165.0
27	NevadaNevadaNevadaNevadaNevadaNevadaNevadaNeva...	122.0	2520	810	460.0
28	New HampshireNew HampshireNew HampshireNew Ham...	21.0	570	560	95.0
29	New JerseyNew JerseyNew JerseyNew JerseyNew Je...	74.0	1590	890	188.0
30	New MexicoNew MexicoNew MexicoNew MexicoNew Me...	114.0	2850	700	321.0
31	New YorkNew YorkNew YorkNew YorkNew YorkNew Yo...	111.0	2540	860	261.0
32	North CarolinaNorth CarolinaNorth CarolinaNort...	130.0	3370	450	161.0
33	North DakotaNorth DakotaNorth DakotaNorth Dako...	8.0	450	440	73.0
34	OhioOhioOhioOhioOhioOhioOhioOhioOhioOhio	73.0	1200	750	214.0

	State	Murder	Assault	UrbanPop	Rape
35	OklahomaOklahomaOklahomaOklahomaOklahomaOklaho...	66.0	1510	680	200.0
36	OregonOregonOregonOregonOregonOregonOregonOreg...	49.0	1590	670	293.0
37	PennsylvaniaPennsylvaniaPennsylvaniaPennsylvan...	63.0	1060	720	149.0
38	Rhode IslandRhode IslandRhode IslandRhode Isla...	34.0	1740	870	83.0
39	South CarolinaSouth CarolinaSouth CarolinaSout...	144.0	2790	480	225.0
40	South DakotaSouth DakotaSouth DakotaSouth Dako...	38.0	860	450	128.0
41	TennesseeTennesseeTennesseeTennesseeTennesseeT...	132.0	1880	590	269.0
42	TexasTexasTexasTexasTexasTexasTexasTexasTexasT...	127.0	2010	800	255.0
43	UtahUtahUtahUtahUtahUtahUtahUtahUtahUtah	32.0	1200	800	229.0
44	VermontVermontVermontVermontVermontVermontVerm...	22.0	480	320	112.0
45	VirginiaVirginiaVirginiaVirginiaVirginiaVirgin...	85.0	1560	630	207.0
46	WashingtonWashingtonWashingtonWashingtonWashin...	40.0	1450	730	262.0
47	West VirginiaWest VirginiaWest VirginiaWest Vi...	57.0	810	390	93.0
48	WisconsinWisconsinWisconsinWisconsinWisconsinW...	26.0	530	660	108.0
49	WyomingWyomingWyomingWyomingWyomingWyomingWyom...	68.0	1610	600	156.0

```python
#usinggroupby
mean_purchase=state.groupby('State')['Murder'].mean().rename("user_mean").reset_index()
print(mean_purchase)
```

```
             State  user_mean
0          Alabama       13.2
1           Alaska       10.0
2          Arizona        8.1
3         Arkansas        8.8
4       California        9.0
5         Colorado        7.9
6      Connecticut        3.3
7         Delaware        5.9
8          Florida       15.4
9          Georgia       17.4
10          Hawaii        5.3
11           Idaho        2.6
12        Illinois       10.4
13         Indiana        7.2
14            Iowa        2.2
15          Kansas        6.0
16        Kentucky        9.7
17       Louisiana       15.4
18           Maine        2.1
19        Maryland       11.3
20   Massachusetts        4.4
21        Michigan       12.1
22       Minnesota        2.7
23     Mississippi       16.1
24        Missouri        9.0
25         Montana        6.0
26        Nebraska        4.3
27          Nevada       12.2
28   New Hampshire        2.1
29      New Jersey        7.4
30      New Mexico       11.4
31        New York       11.1
32  North Carolina       13.0
33    North Dakota        0.8
34            Ohio        7.3
35        Oklahoma        6.6
36          Oregon        4.9
37    Pennsylvania        6.3
38    Rhode Island        3.4
39  South Carolina       14.4
40    South Dakota        3.8
41       Tennessee       13.2
42           Texas       12.7
43            Utah        3.2
44         Vermont        2.2
45        Virginia        8.5
46      Washington        4.0
47   West Virginia        5.7
48       Wisconsin        2.6
49         Wyoming        6.8
```

```
mer=state.merge(mean_purchase)
mer
```

	State	Murder	Assault	UrbanPop	Rape	user_mean
0	Alabama	13.2	236	58	21.2	13.2
1	Alaska	10.0	263	48	44.5	10.0
2	Arizona	8.1	294	80	31.0	8.1
3	Arkansas	8.8	190	50	19.5	8.8
4	California	9.0	276	91	40.6	9.0
5	Colorado	7.9	204	78	38.7	7.9
6	Connecticut	3.3	110	77	11.1	3.3
7	Delaware	5.9	238	72	15.8	5.9
8	Florida	15.4	335	80	31.9	15.4
9	Georgia	17.4	211	60	25.8	17.4
10	Hawaii	5.3	46	83	20.2	5.3
11	Idaho	2.6	120	54	14.2	2.6
12	Illinois	10.4	249	83	24.0	10.4
13	Indiana	7.2	113	65	21.0	7.2
14	Iowa	2.2	56	57	11.3	2.2
15	Kansas	6.0	115	66	18.0	6.0
16	Kentucky	9.7	109	52	16.3	9.7
17	Louisiana	15.4	249	66	22.2	15.4
18	Maine	2.1	83	51	7.8	2.1
19	Maryland	11.3	300	67	27.8	11.3
20	Massachusetts	4.4	149	85	16.3	4.4
21	Michigan	12.1	255	74	35.1	12.1
22	Minnesota	2.7	72	66	14.9	2.7
23	Mississippi	16.1	259	44	17.1	16.1
24	Missouri	9.0	178	70	28.2	9.0
25	Montana	6.0	109	53	16.4	6.0
26	Nebraska	4.3	102	62	16.5	4.3
27	Nevada	12.2	252	81	46.0	12.2
28	New Hampshire	2.1	57	56	9.5	2.1
29	New Jersey	7.4	159	89	18.8	7.4
30	New Mexico	11.4	285	70	32.1	11.4
31	New York	11.1	254	86	26.1	11.1
32	North Carolina	13.0	337	45	16.1	13.0
33	North Dakota	0.8	45	44	7.3	0.8

	State	Murder	Assault	UrbanPop	Rape	user_mean
34	Ohio	7.3	120	75	21.4	7.3
35	Oklahoma	6.6	151	68	20.0	6.6
36	Oregon	4.9	159	67	29.3	4.9
37	Pennsylvania	6.3	106	72	14.9	6.3
38	Rhode Island	3.4	174	87	8.3	3.4
39	South Carolina	14.4	279	48	22.5	14.4
40	South Dakota	3.8	86	45	12.8	3.8
41	Tennessee	13.2	188	59	26.9	13.2
42	Texas	12.7	201	80	25.5	12.7
43	Utah	3.2	120	80	22.9	3.2
44	Vermont	2.2	48	32	11.2	2.2
45	Virginia	8.5	156	63	20.7	8.5
46	Washington	4.0	145	73	26.2	4.0
47	West Virginia	5.7	81	39	9.3	5.7
48	Wisconsin	2.6	53	66	10.8	2.6
49	Wyoming	6.8	161	60	15.6	6.8

In [17]:

```python
#checking for missing value
print(state.isnull().sum())
```

```
State       0
Murder      0
Assault     0
UrbanPop    0
Rape        0
dtype: int64
```

```
import pandas as pd
import numpy as np
cols=['col0','col1','col2','col3','col4']
rows=['row0','row1','row2','row3','row4']
data=np.random.randint(0,100,size=(5,5))
df= pd.DataFrame(data , columns=cols , index=rows)
df.head()
```

Out[25]:

	col0	col1	col2	col3	col4
row0	4	41	83	23	34
row1	93	2	34	39	80
row2	5	93	70	24	61
row3	31	74	7	9	52
row4	18	30	47	15	31

In [26]:

```
df.iloc[4,2]
```

Out[26]:

47

In [27]:

```
df.iloc[3,3]=0
df.iloc[1,2]=np.nan
df.iloc[4,0]=np.nan
df['col5']=0
df['col6']=np.nan
df.head()
```

Out[27]:

	col0	col1	col2	col3	col4	col5	col6
row0	4.0	41	83.0	23	34	0	NaN
row1	93.0	2	NaN	39	80	0	NaN
row2	5.0	93	70.0	24	61	0	NaN
row3	31.0	74	7.0	0	52	0	NaN
row4	NaN	30	47.0	15	31	0	NaN

In [28]:

```
df.loc[:,df.all()]# this function is used to remove all zero values
```

Out[28]:

	col0	col1	col2	col4	col6
row0	4.0	41	83.0	34	NaN
row1	93.0	2	NaN	80	NaN
row2	5.0	93	70.0	61	NaN
row3	31.0	74	7.0	52	NaN
row4	NaN	30	47.0	31	NaN

In [29]:

```
df.loc[:,df.any()]# if any values is integer then print all columns
```

Out[29]:

	col0	col1	col2	col3	col4
row0	4.0	41	83.0	23	34
row1	93.0	2	NaN	39	80
row2	5.0	93	70.0	24	61
row3	31.0	74	7.0	0	52
row4	NaN	30	47.0	15	31

In [30]:

```
df.loc[:,df.isnull().any()]
```

Out[30]:

	col0	col2	col6
row0	4.0	83.0	NaN
row1	93.0	NaN	NaN
row2	5.0	70.0	NaN
row3	31.0	7.0	NaN
row4	NaN	47.0	NaN

In [32]:

```
df.loc[:,df.isnull().all()]
```

Out[32]:

	col6
row0	NaN
row1	NaN
row2	NaN
row3	NaN
row4	NaN

In [33]:

```
df.loc[:,df.notnull().any()]
```

Out[33]:

	col0	col1	col2	col3	col4	col5
row0	4.0	41	83.0	23	34	0
row1	93.0	2	NaN	39	80	0
row2	5.0	93	70.0	24	61	0
row3	31.0	74	7.0	0	52	0
row4	NaN	30	47.0	15	31	0

In [38]:

```
df.dropna(how="all",axis=0)
```

Out[38]:

	col0	col1	col2	col3	col4	col5	col6
row0	4.0	41	83.0	23	34	0	NaN
row1	93.0	2	NaN	39	80	0	NaN
row2	5.0	93	70.0	24	61	0	NaN
row3	31.0	74	7.0	0	52	0	NaN
row4	NaN	30	47.0	15	31	0	NaN

```
df.dropna(how="all",axis=1)
```

Out[37]:

	col0	col1	col2	col3	col4	col5
row0	4.0	41	83.0	23	34	0
row1	93.0	2	NaN	39	80	0
row2	5.0	93	70.0	24	61	0
row3	31.0	74	7.0	0	52	0
row4	NaN	30	47.0	15	31	0

In [39]:

```
df.fillna(df.sum())
```

Out[39]:

	col0	col1	col2	col3	col4	col5	col6
row0	4.0	41	83.0	23	34	0	0.0
row1	93.0	2	207.0	39	80	0	0.0
row2	5.0	93	70.0	24	61	0	0.0
row3	31.0	74	7.0	0	52	0	0.0
row4	133.0	30	47.0	15	31	0	0.0

```python
#Demonstrate transfomr function using pandas in python
import pandas as pd
import numpy as np
import random
data=pd.DataFrame({
    'C':[random.choice(('b','a','d'))for i in range(1000000)],
    'A':[random.randint(1,10)for i in range(1000000)],
    'B':[random.randint(1,10)for i in range(1000000)]
})
data
```

Out[43]:

	C	A	B
0	d	8	1
1	a	9	3
2	b	9	8
3	b	9	10
4	d	7	6
...	...	...	...
999995	d	8	5
999996	a	3	6
999997	d	5	8
999998	a	6	2
999999	a	1	2

1000000 rows × 3 columns

In [47]:

```python
v=data.groupby('C')["A"].mean().rename("D").reset_index()
v
```

Out[47]:

	C	D
0	a	5.497722
1	b	5.498770
2	d	5.497221

```
df_1=data.merge(v)
df_1
```

	C	A	B	D
0	d	8	1	5.497221
1	d	7	6	5.497221
2	d	5	1	5.497221
3	d	5	4	5.497221
4	d	3	8	5.497221
...	...	...	...	...
999995	b	6	8	5.498770
999996	b	10	7	5.498770
999997	b	9	8	5.498770
999998	b	3	2	5.498770
999999	b	9	8	5.498770

1000000 rows × 4 columns