Assignment-1

Basic data frame operations

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [45]:
```
path=r"C:\Users\saiki\data science\datasets\winequality_red.xlsx"
```

In [46]:
```
pd.read_excel(path)
```

Out[46]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	qu
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
3193	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	
3194	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3195	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	
3196	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3197	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	11.0	

3198 rows × 12 columns

In [47]:
```
name=['Ramesh','Suresh','Satish','kishan','suman','nithin','sandesh','rohith']
age=[30,35,40,30,40,39,36,69,71]
name,age
```

Out[47]:
```
(['Ramesh',
  'Suresh',
  'Satish',
  'kishan',
  'suman',
  'nithin',
  'sandesh',
  'rohith'],
 [30, 35, 40, 30, 40, 39, 36, 69, 71])
```

$step - 1$

$create\ data frame$

In [48]:
```
pd.DataFrame()
```

Out[48]:

—

step − 2

creating dataframes from given list

In [49]:
```
pd.DataFrame(zip(name,age))
```

Out[49]:

	0	1
0	Ramesh	30
1	Suresh	35
2	Satish	40
3	kishan	30
4	suman	40
5	nithin	39
6	sandesh	36
7	rohith	69

step − 3

provide columns

In [50]:
```
pd.DataFrame(zip(name,age),columns=['name','age'])
```

Out[50]:

	name	age
0	Ramesh	30
1	Suresh	35
2	Satish	40
3	kishan	30
4	suman	40
5	nithin	39
6	sandesh	36
7	rohith	69

step − 4

Provide index

In [51]:
```
pd.DataFrame(zip(name,age),columns=['name','age'],index=['A','B','C','D','E','F','G'
```

Out[51]:

	name	age
A	Ramesh	30
B	Suresh	35
C	Satish	40
D	kishan	30
E	suman	40
F	nithin	39
G	sandesh	36
H	rohith	69

step − 5

Add new column

In [52]:
```
df=pd.DataFrame(zip(name,age),columns=['name','age'],index=['A','B','C','D','E','F',
city_names=["Hyd","Hyd","Hyd","Banglore","USA","Delhi","Mumbai","Pune"]
relations=['not friend','not friend','not friend','friend','school friend','inter fr
df["city"]=city_names
df["Relation"]=relations
df
```

Out[52]:

	name	age	city	Relation
A	Ramesh	30	Hyd	not friend
B	Suresh	35	Hyd	not friend
C	Satish	40	Hyd	not friend
D	kishan	30	Banglore	friend
E	suman	40	USA	school friend
F	nithin	39	Delhi	inter friend
G	sandesh	36	Mumbai	inter friend
H	rohith	69	Pune	inter friend

step − 6

update the exsisting column

In [53]:
```
df["name"]=["Ram","Suri","Sati","Kit","Biswas","Goka","Sandy","Rohit"]
df
```

Out[53]:

	name	age	city	Relation
A	Ram	30	Hyd	not friend
B	Suri	35	Hyd	not friend
C	Sati	40	Hyd	not friend
D	Kit	30	Banglore	friend
E	Biswas	40	USA	school friend
F	Goka	39	Delhi	inter friend
G	Sandy	36	Mumbai	inter friend
H	Rohit	69	Pune	inter friend

$step - 7$

drop the column

In [54]:
```
df.drop("Relation",axis=1,inplace=True)
df
```

Out[54]:

	name	age	city
A	Ram	30	Hyd
B	Suri	35	Hyd
C	Sati	40	Hyd
D	Kit	30	Banglore
E	Biswas	40	USA
F	Goka	39	Delhi
G	Sandy	36	Mumbai
H	Rohit	69	Pune

$step - 8$

drop the row

In [55]:
```
df.drop("H",axis=0,inplace=True)
df
```

Out[55]:

	name	age	city
A	Ram	30	Hyd
B	Suri	35	Hyd
C	Sati	40	Hyd
D	Kit	30	Banglore
E	Biswas	40	USA
F	Goka	39	Delhi
G	Sandy	36	Mumbai

step – 9

save the data frame

In [56]:
```
df.to_excel("output.xlsx")
```

In [57]:
```
pd.read_excel("output.xlsx")
```

Out[57]:

	Unnamed: 0	name	age	city
0	A	Ram	30	Hyd
1	B	Suri	35	Hyd
2	C	Sati	40	Hyd
3	D	Kit	30	Banglore
4	E	Biswas	40	USA
5	F	Goka	39	Delhi
6	G	Sandy	36	Mumbai

step – 10

remove the index

In [60]:
```
df.to_excel("output.xlsx",index=False)
pd.read_excel("output.xlsx")
```

Out[60]:

	name	age	city
0	Ram	30	Hyd
1	Suri	35	Hyd
2	Sati	40	Hyd
3	Kit	30	Banglore
4	Biswas	40	USA
5	Goka	39	Delhi
6	Sandy	36	Mumbai

Create dataframes using dictionary

In [62]:
```
d1={"NAME":['Ramesh','Suresh','Satish','kishan','suman','nithin','sandesh','rohith',
"AGE":[30,35,40,30,40,39,36,69,71]}
pd.DataFrame(d1)
```

Out[62]:

	NAME	AGE
0	Ramesh	30
1	Suresh	35
2	Satish	40
3	kishan	30
4	suman	40
5	nithin	39
6	sandesh	36
7	rohith	69
8	sandy	71

$step - 4$

In [6]:
```
d1={"NAME":['Ramesh','Suresh','Satish','kishan','suman','nithin','sandesh','rohith',
"AGE":[30,35,40,30,40,39,36,69,71]}
pd.DataFrame(d1,index=['A','B','C','D','E','F','G','H','I'])
```

Out[6]:

	NAME	AGE
A	Ramesh	30
B	Suresh	35
C	Satish	40
D	kishan	30
E	suman	40
F	nithin	39
G	sandesh	36
H	rohith	69
I	sandy	71

$step - 5$

In [8]:
```
df1=pd.DataFrame(d1,index=['A','B','C','D','E','F','G','H','I'])
city_names=["Hyd","Hyd","Hyd","Banglore","USA","Delhi","Mumbai","pondicherry","Pune"
df1["city"]=city_names
df1
```

Out[8]:

	NAME	AGE	city
A	Ramesh	30	Hyd
B	Suresh	35	Hyd
C	Satish	40	Hyd
D	kishan	30	Banglore
E	suman	40	USA
F	nithin	39	Delhi
G	sandesh	36	Mumbai
H	rohith	69	pondicherry
I	sandy	71	Pune

$step - 6$

In [10]:
```
df1["city"]=["hyd","hyd","hyd","Bgl","USA","Delhi","Mum","pond","Pune"]
df1
```

Out[10]:

	NAME	AGE	city
A	Ramesh	30	hyd
B	Suresh	35	hyd
C	Satish	40	hyd
D	kishan	30	Bgl
E	suman	40	USA
F	nithin	39	Delhi
G	sandesh	36	Mum
H	rohith	69	pond
I	sandy	71	Pune

$step - 7$

In [11]:
```
df1.drop("AGE",axis=1,inplace=True)
df1
```

Out[11]:

	NAME	city
A	Ramesh	hyd
B	Suresh	hyd
C	Satish	hyd
D	kishan	Bgl
E	suman	USA
F	nithin	Delhi
G	sandesh	Mum
H	rohith	pond
I	sandy	Pune

$step - 8$

In [12]:
```
df1.drop("I",axis=0,inplace=True)
df1
```

Out[12]:

	NAME	city
A	Ramesh	hyd
B	Suresh	hyd
C	Satish	hyd
D	kishan	Bgl
E	suman	USA
F	nithin	Delhi
G	sandesh	Mum
H	rohith	pond

$step - 9$

In [14]:
```
df1.to_excel("output.xlsx")
```

In [15]:
```
pd.read_excel("output.xlsx")
```

Out[15]:

	Unnamed: 0	NAME	city
0	A	Ramesh	hyd
1	B	Suresh	hyd
2	C	Satish	hyd
3	D	kishan	Bgl
4	E	suman	USA
5	F	nithin	Delhi
6	G	sandesh	Mum
7	H	rohith	pond

$step - 10$

In [16]:
```
df1.to_excel("output.xlsx",index=False)
```

In [17]:
```
pd.read_excel("output.xlsx")
```

Out[17]:

	NAME	city
0	Ramesh	hyd
1	Suresh	hyd
2	Satish	hyd
3	kishan	Bgl
4	suman	USA
5	nithin	Delhi
6	sandesh	Mum
7	rohith	pond

In [ ]:
```

```