1. Create 3 indexes with name, roll number, fees amount in a vector and print the data frame by taking values in indexes.

In [1]:
```
import pandas as pd
data = {'name' : ['ABC','DEF','GHI'],
        'roll number' : ['001','002','003'],
        'fees amount' : [20000, 22500, 21000]}
df=pd.DataFrame(data)
print(df)
```

```
  name roll number  fees amount
0  ABC         001        20000
1  DEF         002        22500
2  GHI         003        21000
```

2. Read mba.csv file and just fetch the name and fees and print it with the help of data frame.

In [5]:
```
import pandas as pd
df=pd.read_csv(r'C:\Users\Nirmalya Majhi\Desktop\Advanced IT Workshop\mba.csv')
print(df)
print(df[['Name','Fees']])
```

```
   Roll No.     Name    Fees
0         1  Anwesha   50000
1         2   Ipsita   80000
2         3    Soham   70000
3         4   Ritesh  100000
4         5    White  120000
      Name    Fees
0  Anwesha   50000
1   Ipsita   80000
2    Soham   70000
3   Ritesh  100000
4    White  120000
```

3. Fetch your answer for question 2 with the help of Pandas Series.

In [6]:
```
import pandas as pd
df=pd.read_csv(r'C:\Users\Nirmalya Majhi\Desktop\Advanced IT Workshop\mba.csv')
data=df.to_dict()
pd.Series(data,index=['Name','Fees'])
```

Out[6]:
```
Name    {0: 'Anwesha', 1: 'Ipsita', 2: 'Soham', 3: 'Ri...
Fees    {0: 50000, 1: 80000, 2: 70000, 3: 100000, 4: 1...
dtype: object
```

4. Copy the datas from csv file dirtydata.csv and replace all the null calories value to 120 and also try to change the null date format to a valid date.

In [13]:
```
import pandas as pd
df=pd.read_csv(r'C:\Users\Nirmalya Majhi\Desktop\Advanced IT Workshop\dirtydata.csv')
df1 = df
df1['Calories'] = df['Calories'].fillna(120)
df1['Date'] = pd.to_datetime(df1['Date'])
print(df1.to_string())
```

	Duration	Date	Pulse	Maxpulse	Calories
0	60	2020-12-01	110	130	409.1
1	60	2020-12-02	117	145	479.0
2	60	2020-12-03	103	135	340.0
3	45	2020-12-04	109	175	282.4
4	45	2020-12-05	117	148	406.0
5	60	2020-12-06	102	127	300.0
6	60	2020-12-07	110	136	374.0
7	450	2020-12-08	104	134	253.3
8	30	2020-12-09	109	133	195.1
9	60	2020-12-10	98	124	269.0
10	60	2020-12-11	103	147	329.3
11	60	2020-12-12	100	120	250.7
12	60	2020-12-12	100	120	250.7
13	60	2020-12-13	106	128	345.3
14	60	2020-12-14	104	132	379.3
15	60	2020-12-15	98	123	275.0
16	60	2020-12-16	98	120	215.2
17	60	2020-12-17	100	120	300.0
18	45	2020-12-18	90	112	120.0
19	60	2020-12-19	103	123	323.0
20	45	2020-12-20	97	125	243.0
21	60	2020-12-21	108	131	364.2
22	45	NaT	100	119	282.0
23	60	2020-12-23	130	101	300.0
24	45	2020-12-24	105	132	246.0
25	60	2020-12-25	102	126	334.5
26	60	2020-12-26	100	120	250.0
27	60	2020-12-27	92	118	241.0
28	60	2020-12-28	103	132	120.0
29	60	2020-12-29	100	132	280.0
30	60	2020-12-30	102	129	380.3
31	60	2020-12-31	92	115	243.0

5. Take a data frame df2 and replace all Nan rows with the mode value for the column calories.Data set is dirtydata.

In [27]:
```
import pandas as pd
df2=pd.read_csv(r'C:\Users\Nirmalya Majhi\Desktop\Advanced IT Workshop\dirtydata.csv')
x=df2['Calories'].mode()
print('mode of Calories column= ',x[0])
df2['Calories']=df2['Calories'].fillna(x[0])
print(df2)
```

```
mode of Calories column=  300.0
    Duration         Date  Pulse  Maxpulse  Calories
0         60  2020/12/01'    110       130     409.1
1         60  2020/12/02'    117       145     479.0
2         60  2020/12/03'    103       135     340.0
3         45  2020/12/04'    109       175     282.4
4         45  2020/12/05'    117       148     406.0
5         60  2020/12/06'    102       127     300.0
6         60  2020/12/07'    110       136     374.0
7        450  2020/12/08'    104       134     253.3
8         30  2020/12/09'    109       133     195.1
9         60  2020/12/10'     98       124     269.0
10        60  2020/12/11'    103       147     329.3
11        60  2020/12/12'    100       120     250.7
12        60  2020/12/12'    100       120     250.7
13        60  2020/12/13'    106       128     345.3
14        60  2020/12/14'    104       132     379.3
15        60  2020/12/15'     98       123     275.0
16        60  2020/12/16'     98       120     215.2
17        60  2020/12/17'    100       120     300.0
18        45  2020/12/18'     90       112     300.0
19        60  2020/12/19'    103       123     323.0
20        45  2020/12/20'     97       125     243.0
21        60  2020/12/21'    108       131     364.2
22        45          NaN    100       119     282.0
23        60  2020/12/23'    130       101     300.0
24        45  2020/12/24'    105       132     246.0
25        60  2020/12/25'    102       126     334.5
26        60     20201226    100       120     250.0
27        60  2020/12/27'     92       118     241.0
28        60  2020/12/28'    103       132     300.0
29        60  2020/12/29'    100       132     280.0
30        60  2020/12/30'    102       129     380.3
31        60  2020/12/31'     92       115     243.0
```

6. Take df2 and for all the index values for the column calories if it is > 250 replace the calorievalue with 250.

In [28]:
```
import pandas as pd
df=df2
for x in df.index:
    if df.loc[x,'Calories']>250:
        df.loc[x,'Calories']=250
print(df)
```

```
    Duration         Date  Pulse  Maxpulse  Calories
0         60  2020/12/01'    110       130     250.0
1         60  2020/12/02'    117       145     250.0
2         60  2020/12/03'    103       135     250.0
3         45  2020/12/04'    109       175     250.0
4         45  2020/12/05'    117       148     250.0
5         60  2020/12/06'    102       127     250.0
6         60  2020/12/07'    110       136     250.0
7        450  2020/12/08'    104       134     250.0
8         30  2020/12/09'    109       133     195.1
9         60  2020/12/10'     98       124     250.0
10        60  2020/12/11'    103       147     250.0
11        60  2020/12/12'    100       120     250.0
12        60  2020/12/12'    100       120     250.0
13        60  2020/12/13'    106       128     250.0
14        60  2020/12/14'    104       132     250.0
15        60  2020/12/15'     98       123     250.0
16        60  2020/12/16'     98       120     215.2
17        60  2020/12/17'    100       120     250.0
18        45  2020/12/18'     90       112     250.0
19        60  2020/12/19'    103       123     250.0
20        45  2020/12/20'     97       125     243.0
21        60  2020/12/21'    108       131     250.0
22        45          NaN    100       119     250.0
23        60  2020/12/23'    130       101     250.0
24        45  2020/12/24'    105       132     246.0
25        60  2020/12/25'    102       126     250.0
26        60     20201226    100       120     250.0
27        60  2020/12/27'     92       118     241.0
28        60  2020/12/28'    103       132     250.0
29        60  2020/12/29'    100       132     250.0
30        60  2020/12/30'    102       129     250.0
31        60  2020/12/31'     92       115     243.0
```

In [ ]:
```
```