

PRACTICAL - 04

4. Consider two excel files having attendance of a workshop's participants for two days. Each file has three fields 'Name', 'Time of joining', duration (in minutes) where names are unique within a file. Note that duration may take one of three values (30, 40, 50) only. Import the data into two dataframes .

CODE :

```
[1]: import numpy as np  
import pandas as pd
```

```
[3]: df=pd.read_csv("Prac_4_Sheet.csv")
```

```
[4]: df.T
```

[4]:	0	1	2	3	4	5	6	7
[17]:	<div>NameTime of JoiningDurationDay</div>							
0	Bradlyn McIntosh	10:30	50	1				
1	Kristian Sanders	11:30	40	1				
2	Emma Strong	12:30	30	1				
3	Axl Jaramillo	13:30	50	1				
4	Guadalupe Bush	14:30	40	1				
5	Kelsey Higgins	15:30	30	1				
6	Leighton Teng	16:30	50	1				
7	Rogelio Marin	17:30	40	1				
8	Celia Pena	9:00	30	1				
9	Marcus McDonald	10:00	50	1				
10	Daisy Villegas	11:00	40	1				
11	Kieran Golden	12:00	50	1				
12	Giuliana Reynolds	13:00	50	1				
13	Vincent Chung	14:00	50	1				
14	Rivka McBride	15:00	50	1				
15	Denver Compton	16:00	40	1				
16	Elina Hawkins	11:00	30	1				
17	Victor Santana	12:00	50	1				
18	Myra Holt	13:00	40	1				
19	Niko Pena	14:00	30	1				
20	Rachel Moran	8:30	30	1				
21	Tate Harper	9:30	50	1				
22	Aria Ward	10:30	50	1				
23	Jameson Chambers	11:30	40	1				
24	Makyla Bauer	12:30	50	1				
25	Kieran Blackwell	13:30	50	1				
26	Sabrina McCall	14:30	50	1				
27	Kean Gilbert	15:30	40	1				
28	Jocelyn Sawyer	16:30	30	2				
29	Jefferson Wilson	17:30	50	2				
30	Luna McCain	10:00	40	2				
31	Heath Mathis	11:00	30	2				
32	Anne Wells	12:00	50	2				
33	Max Snyder	13:00	40	2				
34	Callie Moses	14:00	30	2				
35	Niklaus Frost	15:00	50	2				
36	Paula Park	16:00	40	2				
37	Enzo Farley	17:00	30	2				
38	Wendley Cochran	10:15	50	2				

```
[5]: df1=df[df['Day']==1].drop('Day',axis=1)
df2=df[df['Day']==2].drop('Day',axis=1)
```

Day 1 Participants

[6]: df1

	Name	Time of Joining	Duration
0	Braelyn McIntosh	10:30	50
1	Kristian Sanders	11:30	40
2	Emma Strong	12:30	30
3	Axl Jaramillo	13:30	50
4	Guadalupe Rush	14:30	40
5	Kaiser Higgins	15:30	30
6	Leighton Tang	16:30	50
7	Rogelio Marin	17:30	40
8	Celia Pena	9:00	30
9	Marcus McDonald	10:00	50
10	Daisy Villegas	11:00	40
11	Kieran Golden	12:00	50
12	Giuliana Reynolds	13:00	50
13	Vincent Chung	14:00	50
14	Rivka McBride	15:00	50
15	Denver Compton	16:00	40
16	Eliana Hawkins	11:00	30
17	Victor Santana	12:00	50

18	Myra Holt	13:00	40
19	Niko Pena	14:00	30
20	Rachel Moran	8:30	30
21	Tate Harper	9:30	50
22	Ana Ward	10:30	50
23	Jameson Chambers	11:30	40
24	Makayla Bauer	12:30	50
25	Kieran Blackwell	13:30	50
26	Saoirse McCall	14:30	50

Day 2 Participants

```
[7]: df2
```

	Name	Time of Joining	Duration
28	Jocelyn Sawyer	16:30	30
29	Jefferson Wilson	17:30	50
30	Luna McCann	10:00	40
31	Heath Mathis	11:00	30
32	Anne Wells	12:00	50
33	Max Snyder	13:00	40
34	Callie Moses	14:00	30
35	Niklaus Frost	15:00	50
36	Paula Park	16:00	40
37	Enzo Farley	17:00	30
38	Wrenley Cochran	10:15	50
39	Danny Johnson	11:15	40
40	Emma Strong	12:15	30
41	Axl Jaramillo	13:15	30
42	Guadalupe Rush	14:15	50
43	Kaiser Higgins	15:15	40
44	Leighton Tang	16:15	30
45	Rogelio Marin	17:15	40
46	Celia Pena	13:00	30
47	Marcus McDonald	14:00	50
48	Daisy Villegas	15:00	40

a) Perform merging of the two dataframes to find the names of students who had attended the workshop on both days.

CODE :

```
[8]: pd.merge(df1,df2, 'inner', "Name")
```

SUBMITTED TO : DR. PRITI JAGWANI

NAME : GARIMA ROHILLA

ROLL NO : CSC/21/37

UNIVERSITY ROLL NO : 21059570010

[8]:	Name	Time of Joining_x	Duration_x	Time of Joining_y	Duration_y
0	Emma Strong	12:30	30	12:15	30
1	Axl Jaramillo	13:30	50	13:15	30
2	Guadalupe Rush	14:30	40	14:15	50
3	Kaiser Higgins	15:30	30	15:15	40
4	Leighton Tang	16:30	50	16:15	30
5	Rogelio Marin	17:30	40	17:15	40
6	Celia Pena	9:00	30	13:00	30
7	Marcus McDonald	10:00	50	14:00	50
8	Daisy Villegas	11:00	40	15:00	40

b) Find names of all students who have attended workshop on either of the days.

CODE :

```
[9]: pd.merge(df1,df2,'outer').Name
```

SUBMITTED TO : DR. PRITI JAGWANI
NAME : GARIMA ROHILLA
ROLL NO : CSC/21/37
UNIVERSITY ROLL NO : 21059570010

```
[9]: 0      Brwelyn McIntosh
      1      Kristian Sanders
      2      Emma Strong
      3      Axl Jaramillo
      4      Guadalupe Rush
      5      Kaiser Higgins
      6      Leighton Tang
      7      Rogelio Marin
      8      Celia Pena
      9      Marcus McDonald
     10      Daisy Villegas
     11      Kieran Golden
     12      Giuliana Reynolds
     13      Vincent Chung
     14      Rivka McBride
     15      Denver Compton
     16      Eliana Hawkins
     17      Victor Santana
     18      Myra Holt
     19      Niko Pena
     20      Rachel Moran
     21      Tate Harper
     22      Ana Ward
     23      Jameson Chambers
     24      Makayla Bauer
     25      Kieran Blackwell
     26      Saoirse McCall
     27      Kiaan Gilbert
     28      Jocelyn Sawyer
     29      Jefferson Wilson
     30      Luna McCann
     31      Heath Mathis
     32      Anne Wells
     33      Max Snyder
     34      Callie Moses
     35      Niklaus Frost
     36      Paula Park
     37      Enzo Farley
     38      Wrenley Cochran
     39      Danny Johnson
     40      Emma Strong
     41      Axl Jaramillo
     42      Guadalupe Rush
     43      Kaiser Higgins
     44      Leighton Tang
     45      Rogelio Marin
     46      Celia Pena
     47      Marcus McDonald
```

c) Merge two data frames row-wise and find the total number of records in the data frame.

CODE :

```
[10]: len(pd.concat([df2,df1],axis=0,ignore_index=1))

[10]: 49
```

SUBMITTED TO : DR. PRITI JAGWANI

NAME : GARIMA ROHILLA

ROLL NO : CSC/21/37

UNIVERSITY ROLL NO : 21059570010

- d) Merge two data frames and use two columns names and duration as multi-row indexes. Generate descriptive statistics for this multi-index.

CODE :

```
[11]: pd.concat([df1,df2],axis=0)
```

	Name	Time of Joining	Duration
0	Braelyn McIntosh	10:30	50
1	Kristian Sanders	11:30	40
2	Emma Strong	12:30	30
3	Aid Jaramillo	13:30	50
4	Guadalupe Rush	14:30	40
5	Kaiser Higgins	15:30	30
6	Leighton Tang	16:30	50
7	Rogelio Marin	17:30	40
8	Celia Pena	9:00	30
9	Marcus McDonald	10:00	50
10	Daisy Villegas	11:00	40
11	Kieran Golden	12:00	50
12	Giuliana Reynolds	13:00	50
13	Vincent Chung	14:00	50
14	Rivka McBride	15:00	50
15	Denver Compton	16:00	40
16	Elna Hawkins	11:00	30
17	Victor Santana	12:00	50
18	Myra Holt	13:00	40
19	Niko Pena	14:00	30
20	Rachel Moran	8:30	30
21	Tate Harper	9:30	50
22	Ana Ward	10:30	50
23	Jameson Chambers	11:30	40
24	Makayla Bauer	12:30	50
25	Kieran Blackwell	13:30	50
26	Saoirse McCall	14:30	50
27	Klaan Gilbert	15:30	40
28	Jocelyn Sawyer	16:30	30

SUBMITTED TO : DR. PRITI JAGWANI

NAME : GARIMA ROHILLA

ROLL NO : CSC/21/37

UNIVERSITY ROLL NO : 21059570010

```
[12]: df3=pd.concat([df1,df2],axis=0).set_index(['name','Duration'])
```

```
[13]: df3
```

```
[13]:
```

		Time of Joining	
	Name	Duration	
	Braelyn McIntosh	50	10:30
	Kristian Sanders	40	11:30
	Emma Strong	30	12:30
	Axl Jaramillo	50	13:30
	Guadalupe Rush	40	14:30
	Kaiser Higgins	30	15:30
	Leighton Tang	50	16:30
	Rogelio Marin	40	17:30
	Celia Pena	30	9:00
	Marcus McDonald	50	10:00
	Daisy Villegas	40	11:00
	Kieran Golden	50	12:00
	Giuliana Reynolds	50	13:00
	Vincent Chung	50	14:00
	Rivka McIride	50	15:00
	Denver Compton	40	16:00
	Elina Hawkins	30	11:00
	Victor Santana	50	12:00
	Myra Holt	40	13:00
	Niko Pena	30	14:00
	Rachel Moran	30	8:30
	Tate Harper	50	9:30
	Ana Ward	50	10:30
	Jameson Chambers	40	11:30
	Makayla Bauer	50	12:30
	Kieran Blackwell	50	13:30
	Saoirse McCall	50	14:30

SUBMITTED TO : DR. PRITI JAGWANI
NAME : GARIMA ROHILLA
ROLL NO : CSC/21/37
UNIVERSITY ROLL NO : 21059570010

```
[16]: Df3.describe()
```

```
[16]:
```

	Time of Joining
--	-----------------

count	49
-------	----

unique	27
--------	----

top	14:00
-----	-------

freq	4
------	---