Q4. 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 and do the following:

In [16]:
```
import numpy as np
import pandas as pd

dfDay1 = pd.read_excel('F:\CS\sem 5\Data Analysis and Visualisation\Practicals\Day1_register.x
dfDay2 = pd.read_excel('F:\CS\sem 5\Data Analysis and Visualisation\Practicals\Day2_register.x

dfDay1, dfDay2
```

Out[16]:
```
(         Name Time of Joining  Duration
 0   Christine        11:00:00        40
 1       Susan        11:04:00        30
 2    Margaret        11:08:00        30
 3      Judith        11:01:00        40
 4    Jennifer        11:12:00        50
 5        Mary        11:10:00        30
 6   Elizabeth        11:11:00        30
 7    Patricia        11:07:00        40
 8       Linda        11:08:00        50
 9     Barbara        11:15:00        30
 10    Lynette        11:02:00        40
 11      Robyn        11:03:00        30
 12       Anne        11:19:00        30
 13      Karen        11:13:00        40
 14      Helen        11:05:00        50,
          Name Time of Joining  Duration
 0   Christine        11:00:00        40
 1       Susan        11:06:00        30
 2     Lynette        11:10:00        40
 3      Judith        11:09:00        40
 4        Mary        11:10:00        50
 5      Sandra        11:12:00        30
 6   Elizabeth        11:08:00        30
 7    Patricia        11:08:00        40
 8       Wendy        11:13:00        40
 9     Barbara        11:14:00        30
 10      Janet        11:02:00        50
 11      Robyn        11:00:00        30
 12    Heather        11:08:00        30
 13      Karen        11:09:00        40
 14      Diane        11:06:00        30)
```

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

In [6]:
```
merged=pd.merge(dfDay1,dfDay2,how='inner',on='Name')
merged
```

Out[6]:

	Name	Time of Joining_x	Duration_x	Time of Joining_y	Duration_y
0	Christine	11:00:00	40	11:00:00	40
1	Susan	11:04:00	30	11:06:00	30
2	Judith	11:01:00	40	11:09:00	40
3	Mary	11:10:00	30	11:10:00	50
4	Elizabeth	11:11:00	30	11:08:00	30
5	Patricia	11:07:00	40	11:08:00	40
6	Barbara	11:15:00	30	11:14:00	30
7	Lynette	11:02:00	40	11:10:00	40
8	Robyn	11:03:00	30	11:00:00	30
9	Karen	11:13:00	40	11:09:00	40

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

In [7]:
```
either_day = pd.merge(dfDay1,dfDay2,how='outer',on='Name')
either_day
```

Out[7]:

	Name	Time of Joining_x	Duration_x	Time of Joining_y	Duration_y
0	Christine	11:00:00	40.0	11:00:00	40.0
1	Susan	11:04:00	30.0	11:06:00	30.0
2	Margaret	11:08:00	30.0	NaN	NaN
3	Judith	11:01:00	40.0	11:09:00	40.0
4	Jennifer	11:12:00	50.0	NaN	NaN
5	Mary	11:10:00	30.0	11:10:00	50.0
6	Elizabeth	11:11:00	30.0	11:08:00	30.0
7	Patricia	11:07:00	40.0	11:08:00	40.0
8	Linda	11:08:00	50.0	NaN	NaN
9	Barbara	11:15:00	30.0	11:14:00	30.0
10	Lynette	11:02:00	40.0	11:10:00	40.0
11	Robyn	11:03:00	30.0	11:00:00	30.0
12	Anne	11:19:00	30.0	NaN	NaN
13	Karen	11:13:00	40.0	11:09:00	40.0
14	Helen	11:05:00	50.0	NaN	NaN
15	Sandra	NaN	NaN	11:12:00	30.0
16	Wendy	NaN	NaN	11:13:00	40.0
17	Janet	NaN	NaN	11:02:00	50.0
18	Heather	NaN	NaN	11:08:00	30.0
19	Diane	NaN	NaN	11:06:00	30.0

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

In [8]:
```
either_day['Name'].count()
```

Out[8]:
```
20
```

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

In [9]:
```
both_days = pd.merge(dfDay1,dfDay2,how='outer',on=['Name','Duration']).copy() #  creates a co

both_days.fillna(value='-',inplace=True) # to fill out the missing values in the given series

both_days.set_index(['Name','Duration']) # a method to set a List as index of a Data Frame
```

Out[9]:

		Time of Joining_x	Time of Joining_y
Name	**Duration**		
Christine	**40**	11:00:00	11:00:00
Susan	**30**	11:04:00	11:06:00
Margaret	**30**	11:08:00	-
Judith	**40**	11:01:00	11:09:00
Jennifer	**50**	11:12:00	-
Mary	**30**	11:10:00	-
Elizabeth	**30**	11:11:00	11:08:00
Patricia	**40**	11:07:00	11:08:00
Linda	**50**	11:08:00	-
Barbara	**30**	11:15:00	11:14:00
Lynette	**40**	11:02:00	11:10:00
Robyn	**30**	11:03:00	11:00:00
Anne	**30**	11:19:00	-
Karen	**40**	11:13:00	11:09:00
Helen	**50**	11:05:00	-
Mary	**50**	-	11:10:00
Sandra	**30**	-	11:12:00
Wendy	**40**	-	11:13:00
Janet	**50**	-	11:02:00
Heather	**30**	-	11:08:00
Diane	**30**	-	11:06:00