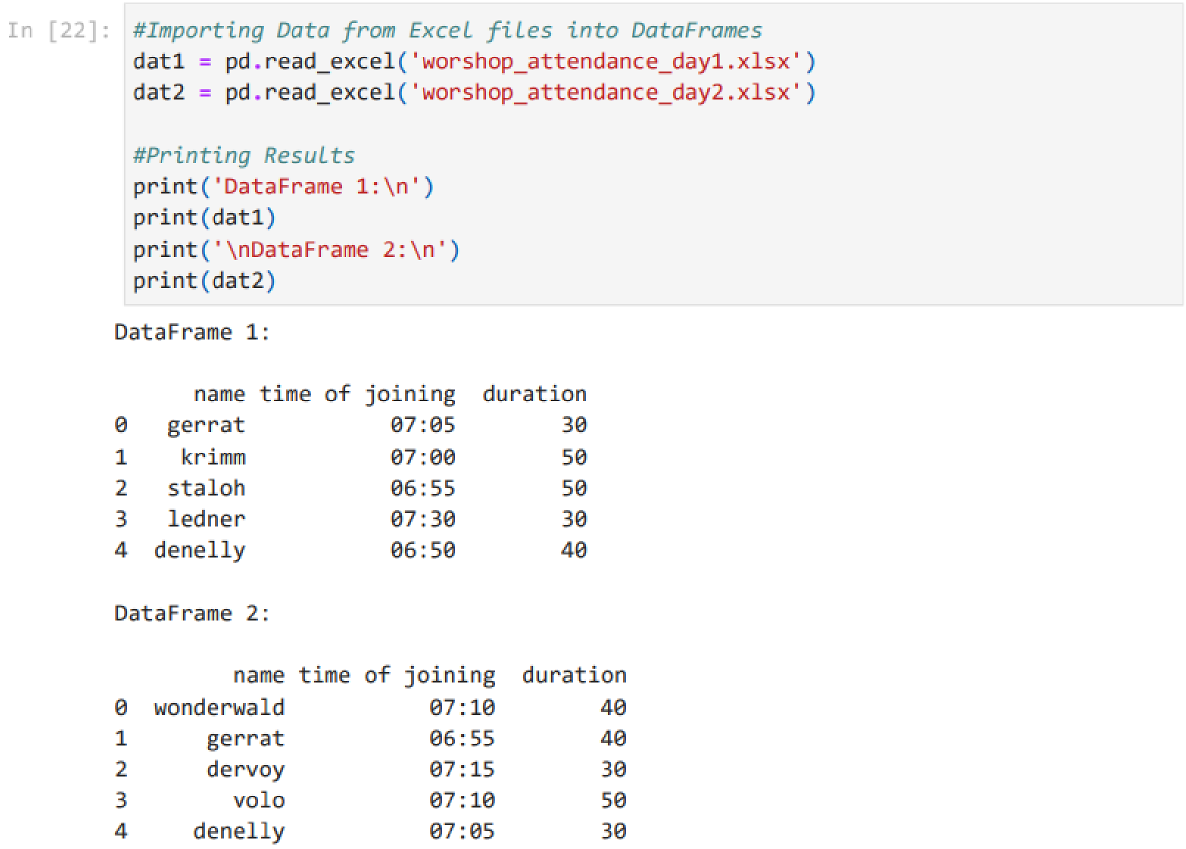
Data Analysis and Visualization Practical – 4

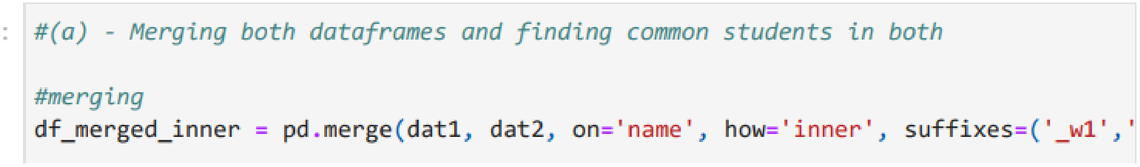
Shad Jamil

CSC/21/45

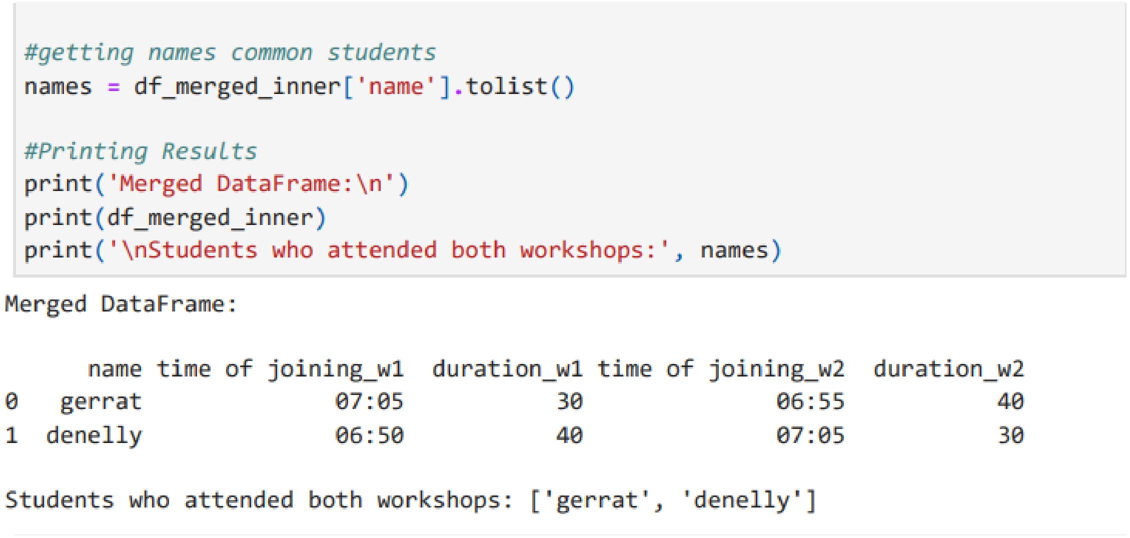


```
In [22]: #Importing Data from Excel files into DataFrames
         dat1 = pd.read_excel('worshop_attendance_day1.xlsx')
         dat2 = pd.read_excel('worshop_attendance_day2.xlsx')

         #Printing Results
         print('DataFrame 1:\n')
         print(dat1)
         print('\nDataFrame 2:\n')
         print(dat2)
```

```
DataFrame 1:

      name time of joining  duration
0   gerrat           07:05        30
1    krimm           07:00        50
2   staloh           06:55        50
3   ledner           07:30        30
4  denelly           06:50        40

DataFrame 2:

         name time of joining  duration
0  wonderwald           07:10        40
1      gerrat           06:55        40
2      dervoy           07:15        30
3        volo           07:10        50
4     denelly           07:05        30
```



```
#(a) - Merging both dataframes and finding common students in both

#merging
df_merged_inner = pd.merge(dat1, dat2, on='name', how='inner', suffixes=('_w1','
```

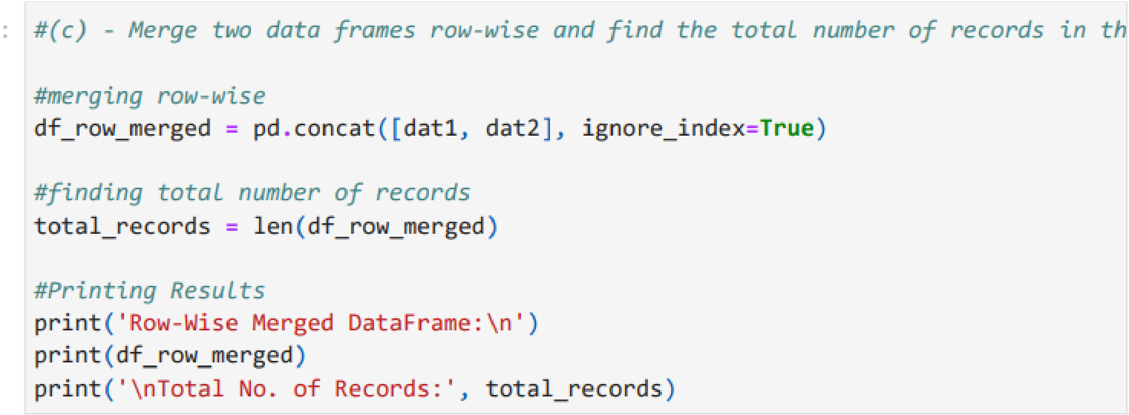


```

#getting names common students
names = df_merged_inner['name'].tolist()

#Printing Results
print('Merged DataFrame:\n')
print(df_merged_inner)
print('\nStudents who attended both workshops:', names)
```

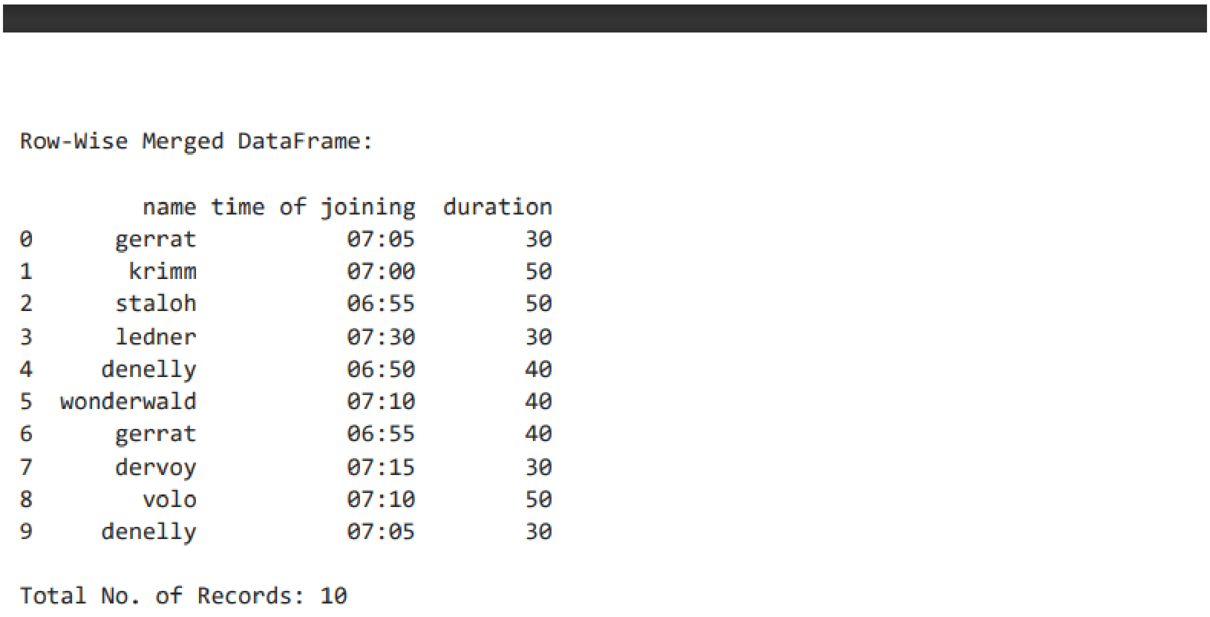
```
Merged DataFrame:

      name time of joining_w1  duration_w1 time of joining_w2  duration_w2
0   gerrat              07:05           30              06:55           40
1  denelly              06:50           40              07:05           30

Students who attended both workshops: ['gerrat', 'denelly']
```

```
#(b) - students who attended on either day

#merging
df_merged_outer = pd.merge(dat1, dat2, on='name', how='outer', suffixes=('_w1','

#getting names of desired students
names_attended_both = df_merged_outer['name'].tolist()

#Printing Results
print('Merged DataFrame:\n')
print(df_merged_outer)
print('\nStudents who attended either workshop:', names_attended_both)
```

```
Merged DataFrame:

         name time of joining_w1  duration_w1 time of joining_w2  duration_w2
0      gerrat              07:05         30.0              06:55         40.0
1       krimm              07:00         50.0                NaN          NaN
2      staloh              06:55         50.0                NaN          NaN
3      ledner              07:30         30.0                NaN          NaN
4     denelly              06:50         40.0              07:05         30.0
5  wonderwald                NaN          NaN              07:10         40.0
6      dervoy                NaN          NaN              07:15         30.0
7        volo                NaN          NaN              07:10         50.0

Students who attended either workshop: ['gerrat', 'krimm', 'staloh', 'ledner', 'd
enelly', 'wonderwald', 'dervoy', 'volo']
```



```
#(c) - Merge two data frames row-wise and find the total number of records in th

#merging row-wise
df_row_merged = pd.concat([dat1, dat2], ignore_index=True)

#finding total number of records
total_records = len(df_row_merged)

#Printing Results
print('Row-Wise Merged DataFrame:\n')
print(df_row_merged)
print('\nTotal No. of Records:', total_records)
```



```
Row-Wise Merged DataFrame:

         name time of joining  duration
0      gerrat           07:05        30
1       krimm           07:00        50
2      staloh           06:55        50
3      ledner           07:30        30
4     denelly           06:50        40
5  wonderwald           07:10        40
6      gerrat           06:55        40
7      dervoy           07:15        30
8        volo           07:10        50
9     denelly           07:05        30

Total No. of Records: 10
```

```
#(d) - Merge two data frames and use two columns names and duration as multi-row

#merging dataframe
df_merged_new = pd.merge(dat1, dat2, on=['name', 'duration'], how='outer', suffi
df_merged_new.set_index(['name', 'duration'], inplace=True)

#descriptive statistics
desc_stats = df_merged_new.describe()

#Printing Results
print('Merged DataFrames with Multi-Row Indexes:\n')
print(df_merged_new)
print('\nDescriptive Statistics:\n')
print(desc_stats)
```

```
Merged DataFrames with Multi-Row Indexes:

                    time of joining_w1 time of joining_w2
name       duration
gerrat     30                    07:05                NaN
krimm      50                    07:00                NaN
staloh     50                    06:55                NaN
ledner     30                    07:30                NaN
denelly    40                    06:50                NaN
wonderwald 40                      NaN              07:10
gerrat     40                      NaN              06:55
dervoy     30                      NaN              07:15
volo       50                      NaN              07:10
denelly    30                      NaN              07:05

Descriptive Statistics:

       time of joining_w1 time of joining_w2
count                   5                  5
unique                  5                  4
top                 07:05              07:10
freq                    1                  2
```