

subject

Machine learning

Titled homework

Collecting attendance minutes and knowing absence in more than one excels file using Python language

Supervisor

Abdalmohaymen Alesmaeel

Student

Wajeih MouRad Alsultan

```
import numpy as np
import pandas as pd
import glob
```

We called the three libraries:

- **numpy**: in order that this library allows you to deal with arrays in a better way than the existing Lists automatically.
- **Pandas**: In order that this library for data processing and analysis. It presents data structures and manipulations of numerical tables and time series.
- **glob**: They can be used to find files that fit your purpose. Search for files in Windows.

```
location = r'D:\Wajeeh\University\Machine Learning\Homework\One\*.xlsx'
excel_files = glob.glob(location)
```

- Here is the location of the folder that contains the Excel files that we will work with on my computer.

```
pd.set_option('display.max_rows', 91)
df1 = pd.DataFrame()
```

- Here is the Data structure also contains labeled axes (rows and columns). Arithmetic operations align on both row and column labels. Can be thought of as a dict-like container for Series objects. The primary panda's data structure.

```
for excel_file in excel_files:
    df2 = pd.read_excel(excel_file)
    df1 = pd.concat([df1, df2], ignore_index=True)
df1.fillna(value="N/A", inplace=True)
```

- Here all the files in the folder were read through a loop and also merged into one file.

```
df1.to_excel(r'D:\Wajeeh\University\Machine Learning\Homework\One\all.xlsx')
df1.head()
```

- Here the files were saved in one file on my computer and then read on the code editor in a beautiful way.

Out[9]:

	name	User Email	Date	attendance	Guest
0	ahmad alomar	gas@gmail.com	2022-11-13	51	Yes
1	خالد الطويل	gsd@gmail.com	2022-11-13	97	No
2	سمير معتوق	ghf@gmail.com	2022-11-13	96	Yes
3	mohammad alomar	gdd@gmail.com	2022-11-13	96	Yes
4	ahmad alomar	gas@gmail.com	2022-11-14	40	Yes

Sample Records

```
df1["name"] = df1["name"].astype("category")
df1.groupby(by="name").sum()
```

- Here, the name column was grouped into categories for each category name, and then the attendance column was collected for each category separately, and the following figure shows the information.

Out[5]:

	attendance
name	
ahmad alomar	251
mohammad alomar	457
خالد الطويل	293
سمير معتوق	328

Sum if for name

```
df1.groupby(by="name").count()
```

- Here, attendance days were counted for each person, assuming a number of sessions equal to five.

Out[6]:

	User Email	Date	attendance	Guest
name				
ahmad alomar	5	5	5	5
mohammad alomar	5	5	5	5
خالد الطويل	4	4	4	4
سمير معتوق	4	4	4	4

Count For attendance

The End