



الجامعة الدولية للعلوم والنهضة

كلية الهندسة المعلوماتية

# تقرير وظيفة تعلم الآلة

عمل الطالب: مصعب بويضاني

إشراف الاستاذ: عبد المهيمن الإسماعيل

بدأنا أولاً بقراءة عشر ملفات تم تسميتها بالتتالي من 2020-11-1 إلى 2022-11-10 وذلك بحلقة كالتالي:

```
3 files=[]
4 filesFolder='Excel files'
5 finalDF=pd.DataFrame()
6
7 for i in range(10):
8     df=pd.read_excel(f"{filesFolder}\\{i+1}-11-2022.xlsx")
9     files.append((1:df,2:f"{i+1}-11-2022"))
10
```

وتم تخزين محتوى الملف مع عنوانه في قاموس هيث الفهرس رقم 1 هو محتوى الملف والفهرس رقم 2 هو اسم الملف

بعدها قمنا بدمج الملفات العشر للحصول على كافة المعلومات الأساسية (الاسم الايميل و هل هو ضيف ام لا) كالتالي:

```
11 allBasicInfo=[]
12
13 for i in files:
14     allBasicInfo.append(i[1][['Name','Email','Guest']])
15
16 basicInfo=pd.concat(allBasicInfo)
17 basicInfo.drop_duplicates(subset="Name",inplace=True)
18
```

طبعا الان أصبح عندنا داتا فريم تحتوي على كافة المعلومات الأساسية لكن مكررة ولهذا سنقوم بتجميعها حسب الاسم

قمنا بهذه الخطوة بشكل أساسي لجمع كافة الأسماء في الملفات العشرة

بعد ذلك نأخذ كل اسم من نتج من الخطوة السابقة ونحاول الحصول على بياناته من جميع الملفات العشر

إذا كان موجود فنضيف الوقت الى متغير SUM وإذا لم يكن موجود فنقوم بزيادة عداد الغيابات

وخلال هذه العملية يتم حفظ الوقت حسب التاريخ أيضا في قاموس اسمه dates

وذلك كالتالي:

```
19 for n,e,g in zip(basicInfo["Name"],basicInfo["Email"],basicInfo["Guest"]):
20     sum=0
21     absent=0
22     dates={}
23     for f in files:
24         dataRow=f[1][f[1]["Name"] == n]
25         dataRow=dataRow.squeeze()
26         if(dataRow.empty==False):
27             sum+=dataRow["Time"]
28             dates[f[2]]=dataRow["Time"]
29
30     else:
31         absent+=1
32         dates[f[2]]=0
33
34
35     rowData = {'Name': n, 'Email': e, 'Guest': g, **dates, "Time":sum,"Absent":absent}
36     row = pd.Series(data=rowData)
37     row=row.to_frame().T
38     finalDF= pd.concat([finalDF,row],axis=0)
39
```

وأخيرا نحفظ الداتا فريم النهائية في ملف ايكسل كالتالي:

```
40 finalDF.to_excel("final.xlsx",index=False)
```

صورة للملف:

Name	Email	Guest	1-11-2022	2-11-2022	3-11-2022	4-11-2022	5-11-2022	6-11-2022	7-11-2022	8-11-2022	9-11-2022	10-11-2022	Time	Absent
ahmad alomar		Yes	65	98	51	51	51	0	28	98	98	51	591	1
mohammad alomar		Yes	69	69	0	0	0	0	68	0	69	0	275	6
سمير معشوق		Yes	76	55	85	85	85	65	24	26	55	85	641	0
jamal mahj@s.com		No	39	0	86	86	86	25	86	97	0	86	591	2
jamal kasejamka@gn		No	0	77	0	0	0	0	0	0	77	0	154	8
خالد الطويل	ghf@gmail	No	0	0	24	24	24	87	24	65	0	24	272	3
musab b		Yes	0	0	0	0	0	0	0	0	100	0	100	9

رابط الكود:

<https://github.com/Musab-boaidani/Pandas-Homework>