



للمبتدئين مكتبة باينداس بالعربي



Qusay AL-Btoush

[github.com/qusaybtoush](https://github.com/qusaybtoush)

<https://www.linkedin.com/in/qusayal-btoush>

<https://www.kaggle.com/qusaybtoush1990>

# پانداس / pandas

## للمبتدئين

مكتبة بانداس هي مكتبة برمجية مشهورة في لغة برمجة بايثون، وهي متخصصة في التعامل مع البيانات المنظمة والتحليلية. قوة هياكل بيانات ووظائف تسهل العمل مع بيانات منظمة في صورة جداول وسلاسل. لقد تقدم بانداس على نطاق واسع في مجالات مثل علم البيانات وتحليل النتائج والبرمجة العلمية.

### Qusay AL-Btoush

- <https://github.com/qusaybtoush> (http:)
- <https://www.linkedin.com/in/qusayal-btoush/> (http:)
- <https://www.kaggle.com/qusaybtoush1990> (http:)

In [17]: *# تحميل مكتبة بانداس*  
`import pandas as pd`

In [16]: *# تعريف متغير باستخدام القاموس*

```
di = {"Name": ["Ali", "Sahra", "Ahmad", "Mouhammed", "Ali", "Sahra", "Ali"],
      "Age": [ 7, 8, 14, 20, 7, 6, 11],
      "Mark": ["A", "A", "B", "A", "A", "D", "A"]
}
```

In [19]: *# نستخدم فنكشن داتا فريم للتعريف الداتا واستخدام الداتا من القاموس*

```
df = pd.DataFrame(data=di) # تم حفظ الداتا في متغير دي اف
df # print
```

Out[19]:

	Name	Age	Mark
0	Ali	7	A
1	Sahra	8	A
2	Ahmad	14	B
3	Mouhammed	20	A
4	Ali	7	A
5	Sahra	6	D
6	Ali	11	A

```
In [30]: # استخدام iloc or loc
# اسم وطريقه تحديد العنصر , رمز الفاصله بين الصفوف و الاعمده اولا صفوف ثانيا اعمده
# df.iloc[the rows from : to , than the number of column from : to]
# صف من الى , العמוד من الى
df.iloc[0:3 , 0:2] # اول ثلاث صفوف و اثنين عامود
```

Out[30]:

	Name	Age
0	Ali	7
1	Sahra	8
2	Ahmad	14

```
In [34]: # هنا احدد اسم العמוד
df.loc[0:3,["Name","Age"]]
```

Out[34]:

	Name	Age
0	Ali	7
1	Sahra	8
2	Ahmad	14
3	Mouhammed	20

```
In [38]: # تحديد عنصر بناء على الموقع اندكس
df.iloc[3,1]
```

Out[38]: 20

```
In [39]: # استطيع تغير القيمه في الداتا
df.iloc[3,1] = 30
df #print Data Frame
```

Out[39]:

	Name	Age	Mark
0	Ali	7	A
1	Sahra	8	A
2	Ahmad	14	B
3	Mouhammed	30	A
4	Ali	7	A
5	Sahra	6	D
6	Ali	11	A

In [40]: *# فنكشن هيد ياتي ب اول خميس قيم في الداتا و استطيع وضع عدد الصفوف بين الاقواس*  
`df.head()`

Out[40]:

	Name	Age	Mark
0	Ali	7	A
1	Sahra	8	A
2	Ahmad	14	B
3	Mouhammed	30	A
4	Ali	7	A

In [42]: *# فنكشن شيب ياتي ب شكل الداتا عدد الصفوف و عدد الاعمده*  
`df.shape` *# 7 rows and 3 columns*

Out[42]: (7, 3)

In [43]: *# فنكشن انفو تاتي ب الاعمده و نوع الداتا و معلومات عن الرام وغيرها*  
`df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7 entries, 0 to 6
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0    Name     7 non-null      object
1    Age       7 non-null      int64
2    Mark     7 non-null      object
dtypes: int64(1), object(2)
memory usage: 296.0+ bytes
```

In [65]: *# فنكشن يونيك ياتي في القيم الفريده من غير تكرار*  
`df["Mark"].unique()`

Out[65]: array(['A', 'B', 'D'], dtype=object)

In [47]: *# فنكشن از نل ياتي اذا كان لدي قيم مفقوده*  
`df.isnull().sum()` *# df["Name"].isnull().sum() اذا يوجد فيه قيم مفقوده*

Out[47]: Name 0  
Age 0  
Mark 0  
dtype: int64

In [48]: *# فنكشن دليكيشن يفحص اذا هناك تكرار في الداتا*  
`df.duplicated().sum()`

Out[48]: 1

In [50]: *فنكشن كولوم ياتي في اسماء الاعمده في الداتا*  
`df.columns`

Out[50]: Index(['Name', 'Age', 'Mark'], dtype='object')

In [52]: *فنكشن دسكرايب ياتي ب وصف ل الداتا الرقميه لاي*  
`df.describe() # show the statistics`

Out[52]:

	Age
count	7.000000
mean	11.857143
std	8.474050
min	6.000000
25%	7.000000
50%	8.000000
75%	12.500000
max	30.000000

In [53]: *استطيع عمل فلتر على اسم معي او اكثر من اسم*  
`df[df["Name"]== "Ahmad"]`

Out[53]:

	Name	Age	Mark
2	Ahmad	14	B

In [59]: *يمكن استخدام فنكشن الخاصه في المعلومات الاحصائيه*  
`print ("The Max Age :", df["Age"].max())`  
`print ("The Min Age :", df["Age"].min())`  
`print ("The Avg Age :", df["Age"].mean())`  
`print ("The Total Age :", df["Age"].sum())`

The Max Age : 30  
 The Min Age : 6  
 The Avg Age : 11.857142857142858  
 The Total Age : 83

**الحذف Drop**

```
In [60]: # Drop
#استطيع حذف اعمده او صفوف او قيم مفقوده او قيم مكرره

#drop column
df.drop(columns= "Name") # حذف عامود الاسم لكن يحفظ في الداتا الا اذا عملت انبليس صح
```

Out[60]:

	Age	Mark
0	7	A
1	8	A
2	14	B
3	30	A
4	7	A
5	6	D
6	11	A

```
In [61]: # حذف القيم المكرره
df.drop_duplicates() # حذف الصفوف المكرره لكن يحفظ في الداتا الا اذا عملت انبليس صح
```

Out[61]:

	Name	Age	Mark
0	Ali	7	A
1	Sahra	8	A
2	Ahmad	14	B
3	Mouhammed	30	A
5	Sahra	6	D
6	Ali	11	A

```
In [62]: # حذف القيم المفقوده
df.dropna()
```

Out[62]:

	Name	Age	Mark
0	Ali	7	A
1	Sahra	8	A
2	Ahmad	14	B
3	Mouhammed	30	A
4	Ali	7	A
5	Sahra	6	D
6	Ali	11	A

In [68]: *# استطيع تعبئه القيم المفقوده*  
*df["Age"].fillna(5) # اذا كان يوجد سوف تملأ برقم 5 او اي رقم انت تختاره*

Out[68]:

0	7
1	8
2	14
3	30
4	7
5	6
6	11

Name: Age, dtype: int64

### Group by جروب

In [64]: *# من خلال مكتبة بانداس تستطيع عمل جروب حسب عامود معين*  
*# هنا اريد عمل جروب على العلامه مع متوسط العمر*  
*df.groupby("Mark")["Age"].mean()*

Out[64]:

Mark	
A	12.6
B	14.0
D	6.0

Name: Age, dtype: float64

### استيراد و تصدير البيانات

- ... مكتبة بانداس تدعم الكثير من المصادر ل استيراد البيانات مثل الاكسل او الويب او قواعد البيانات وغيرها الكثير

In [69]: *# استيراد داتا من ملف سي اس في اولاً عرف متغير مثل دي اف و اكتب الفنكش التالي*  
*# df = pd.read\_csv("مكان الملف او اسمه اذا كان مرفوع على جيتور.csv")*

*بعد رفع الداتا تستطيع عمل التحليل وفحص البيانات وكل الخطوات السابقه و اكثر*

*# تصدير البيانات*

*بيانات او تم استيراد البيانات من الويب او اي مصدر وتريد تصدير البيانات يمكن من خلال الفنكش*

*#df.to\_csv() the name like df.to\_csv("new name") اسم الداتا*

- Qusay AL-Btoush : اذا كان لدي اي سؤال تواصل معنا
- <https://github.com/qusaybtoush> (http:)
- <https://www.linkedin.com/in/qusayal-btoush/> (http:)
- <https://www.kaggle.com/qusaybtoush1990> (http:)

In [ ]: