

# دوره جامع پایتون در یادگیری ماشین

قسمت هفتم، گروپ بار!

Pandas<sup>4</sup>

[Datalk.ir](https://datalk.ir)

Created by : Ali Arabshahi

Contact us : [Linkedin.com/in/mrAliArabshahi](https://www.linkedin.com/in/mrAliArabshahi)

## گروپ بای (Group By)

گروپ بای مفهوم خیلی پرکاربردی هست و معمولا خیلی کم پیش میاد باهاش سر و کار نداشته باشیم. به خصوص زمانی که می خوایم داده هامون رو تجمیع کنیم و یک دانشی ازشون به دست بیاریم.

اما کارش چیه؟ چند تا چیز شبیه هم رو می چسبونه به هم و یکیشون می کنه. این یکی کردن به چه دردی می خوره؟ می تونیم بعد از این یکی کردن یک سری تابع روشون اعمال کنیم. مثلا فرض کنیم ستونی از دیتا بیس شامل اسم افراد هست. علی پنج بار از فروشگاه شما خرید کرده یعنی در ستون مربوط به اسم ها پنج بار اسمش تکرار شده. هر بار خرید کردن نتیجه اش ثبت داده ای تحت عنوان مبلغ خرید در دیتا بیس مون هست. حالا می خوایم ببینیم علی کلا چه قدر خرید کرده. گروپ بای اینجا به کمکمون می یاد. به این صورت که اول روی ستون اسامی گروپ بای می زنیم و بعد تابع جمع رو روی مقادیر خرید اعمال می کنیم. 🙈

مشابه کاری که در تصویر پایین انجام شده. Funny ها یکی و در ستون power عملیات جمع انجام شده.

"CLOUD" table					t1_temp current outcome				
ID	NeighborhoodID	power	tag	ip	id	NeighborhoodID	power	tag	ip
1	1	5	funny		1	1	15	funny	
2	1	2	likeable		2	1	2	likeable	
3	4	1	funny		5	3	3	smart	
4	1	9	funny		6	3	8	nice	
5	3	2	smart		9	4	5	alright	
6	3	8	nice		8	8	3	okay	
7	3	1	smart						
8	8	3	okay						
9	4	5	alright						

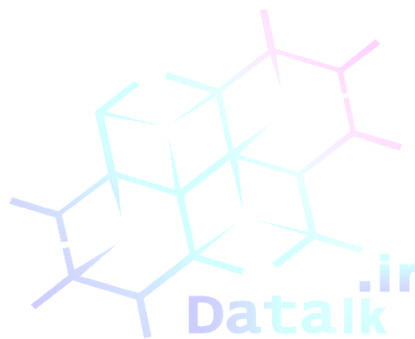
خب برای تمرین کتابخونه پانداس رو فراخونی می کنیم و یکسری داده ایجاد و اون ها رو به صورت دیتافریم در می آوریم.

```
import pandas as pd
data = {'Company':['GOOG','GOOG','MSFT','MSFT','FB','FB'],
        'Person':['Sam','Charlie','Amy','Vanessa','Carl','Sarah'],
        'Sales':[200,120,340,124,243,350]}
```

```
df = pd.DataFrame(data)
```

```
df
```

	Company	Person	Sales
0	GOOG	Sam	200
1	GOOG	Charlie	120
2	MSFT	Amy	340
3	MSFT	Vanessa	124
4	FB	Carl	243
5	FB	Sarah	350



حالا از دستور گروپ بای استفاده می کنیم. اما حواستون باشه که تا وقتی یک تابع رو روی اون فراخونی نکردیم نمیتونیم چیزی دریافت کنیم. برای این که انگار دستور ناقصه. پانداس می گه خب من داده ها رو بر اساس ستونی که بهم گفتین تجمیع کردم اما حالا از کجا باید بدونم، چه داده ای رو به شما برگردونم؟! 🤖

```
df.groupby('Company')
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001CF32785888>
```

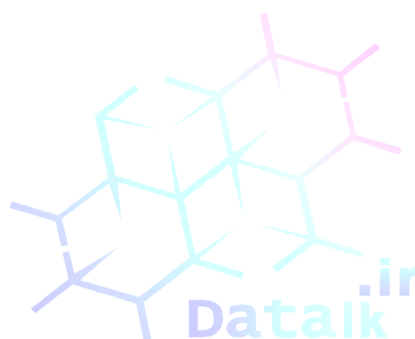
حالا گروپ بای انجام شده رو داخل یک متغیر می ریزیم:

```
by_comp = df.groupby("Company")
```

و در آخر یک تابع رو که در اینجا میانگین هست بر روی شیء گروپ بای شده مون اعمال می کنیم:

```
by_comp.mean()
```

Sales	
Company	
FB	296.5
GOOG	160.0
MSFT	232.0



می تونیم مراحل بالا رو در یک خط هم انجام بدیم:

```
df.groupby('Company').mean()
```

Sales	
Company	
FB	296.5
GOOG	160.0
MSFT	232.0

در ادامه توابع دیگه ای رو روی بر روی متغیرمون که حالا از جنس داده های گروپ بای شده هست، اعمال می کنیم.

این تابع انحراف معیار مقادیر مختلف رو بهمون بر می گردونه:

`by_comp.std()`

Sales	
Company	
FB	75.660426
GOOG	56.568542
MSFT	152.735065

می دونین دیگه: 😊

`by_comp.min()`

Person		Sales
Company		
FB	Carl	243
GOOG	Charlie	120
MSFT	Amy	124



`by_comp.max()`

Person		Sales
Company		
FB	Sarah	350
GOOG	Sam	200
MSFT	Vanessa	340

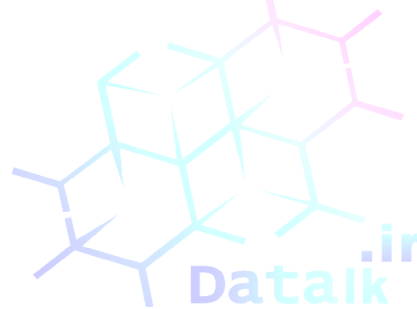
از هر مقدار چند تا وجود داشته:

```
by_comp.count()
```

Person Sales		
Company		
FB	2	2
GOOG	2	2
MSFT	2	2

این تابع خیلی باحال هم می یاد و کلی اطلاعات راجع به فیلد های عددی بهمون می ده :

```
by_comp.describe()
```



Sales								
Company								
	count	mean	std	min	25%	50%	75%	max
FB	2.0	296.5	75.660426	243.0	269.75	296.5	323.25	350.0
GOOG	2.0	160.0	56.568542	120.0	140.00	160.0	180.00	200.0
MSFT	2.0	232.0	152.735065	124.0	178.00	232.0	286.00	340.0

ترنسپووو ز! هم یک دستوری که جای سطر ها و ستون ها رو عوض می کنه. دقت کنین! همون دستور صفحه قبله و البته با فرمت جدید :

```
by_comp.describe().transpose()
```

	Company	FB	GOOG	MSFT
Sales	count	2.000000	2.000000	2.000000
	mean	296.500000	160.000000	232.000000
	std	75.660426	56.568542	152.735065
	min	243.000000	120.000000	124.000000
	25%	269.750000	140.000000	178.000000
	50%	296.500000	160.000000	232.000000
	75%	323.250000	180.000000	286.000000
	max	350.000000	200.000000	340.000000

و دستور بالا رو برای یک ستون دلخواه اجرا می کنیم:

```
by_comp.describe().transpose()['GOOG']
```

Sales	count	2.000000
	mean	160.000000
	std	56.568542
	min	120.000000
	25%	140.000000
	50%	160.000000
	75%	180.000000
	max	200.000000
Name: GOOG, dtype: float64		

به همین راحتی و به همین پر کاربردی!

