What can we do with Pandas Aggregates?

In this example, we'll be working with data from ShoeFly.com, a fictional online shoe store.

```
In [1]:  # Before we analyze anything, we need to import pandas
         # and load our data
         import pandas as pd

         df = pd.read_csv('shoefly_page_visits.csv')
```

Let's examine the first 10 rows of our data!

```
In [2]:  df.head(10)
```

Out[2]:

	id	first_name	last_name	email	month	utm_source
0	10043	Louis	Koch	LouisKoch43@gmail.com	3 - March	yahoo
1	10150	Bruce	Webb	BruceWebb44@outlook.com	3 - March	twitter
2	10155	Nicholas	Hoffman	Nicholas.Hoffman@gmail.com	2 - February	google
3	10178	William	Key	William.Key@outlook.com	3 - March	yahoo
4	10208	Karen	Bass	KB4971@gmail.com	2 - February	google
5	10260	Benjamin	Ochoa	Benjamin.Ochoa@outlook.com	1 - January	twitter
6	10271	Gerald	Aguilar	Gerald.Aguilar@gmail.com	3 - March	google
7	10278	Melissa	Lambert	Melissa.Lambert@gmail.com	2 - February	email
8	10320	Adam	Strickland	Adam.Strickland@gmail.com	3 - March	email
9	10389	Ethan	Payne	EthanPayne26@outlook.com	2 - February	facebook

Notice that there's a column called "utm_source". This column tells us the website that sent users to ShoeFly.com. There's also a column called "month", which tells us the month in which this user visited ShoeFly.com.

We want to know how our sources have changed from month to month. Let's investigate!

```
In [3]:  # This command shows us how many users visited the site from different sources in different months.
         df.groupby(['month', 'utm_source']).id.count().reset_index()
```

Out[3]:

	month	utm_source	id
0	1 - January	email	43
1	1 - January	facebook	404
2	1 - January	google	127
3	1 - January	twitter	164
4	1 - January	yahoo	262
5	2 - February	email	147
6	2 - February	facebook	263
7	2 - February	google	196
8	2 - February	twitter	154
9	2 - February	yahoo	240
10	3 - March	email	272
11	3 - March	facebook	156
12	3 - March	google	220
13	3 - March	twitter	97
14	3 - March	yahoo	255

```
In [4]:  # This command shows us how many users visited the site from different sources in different months.
         df.groupby(['month', 'utm_source']).id.count()\
             .reset_index()\
             .pivot(columns='month', index='utm_source', values='id')
```

Out[4]:

month	1 - January	2 - February	3 - March
utm_source			
email	43	147	272
facebook	404	263	156
google	127	196	220
twitter	164	154	97
yahoo	262	240	255

Over the course of these three months, it looks like we got more visits from "email" and "google", but fewer visits from "facebook" and "twitter". The number of visits from "yahoo" stayed mostly the same.