

data_challenge_2-1

February 27, 2019

```
In [2]: import pandas as pd
import numpy as np
```

```
In [24]: import matplotlib.pyplot as plt
```

```
In [3]: emails = pd.read_csv('yammer_emails.csv')
events = pd.read_csv('yammer_events.csv')
users = pd.read_csv('yammer_users.csv')
rollup = pd.read_csv('dimension_rollup_periods.csv')
```

First, I want to take a look at our data in general and get a sense of what I'm working with. I'll start with the emails table. What's in there, and what kind of actions are taken by email?

```
In [4]: emails.head()
```

```
Out[4]:
```

	user_id	occurred_at	action	user_type
0	0.0	2014-05-06 09:30:00	sent_weekly_digest	1.0
1	0.0	2014-05-13 09:30:00	sent_weekly_digest	1.0
2	0.0	2014-05-20 09:30:00	sent_weekly_digest	1.0
3	0.0	2014-05-27 09:30:00	sent_weekly_digest	1.0
4	0.0	2014-06-03 09:30:00	sent_weekly_digest	1.0

```
In [5]: print(f'Email actions were: {emails.action.unique()}')
print(f'user types are: {emails.user_type.unique()}')
```

```
Email actions were: ['sent_weekly_digest' 'email_open' 'email_clickthrough'
'sent_reengagement_email']
user types are: [1. 3. 2.]
```

Now, let's look at what's in events

```
In [6]: events.head()
```

```
Out[6]:
```

	user_id	occurred_at	event_type	event_name	location	\
0	10522.0	2014-05-02 11:02:39	engagement	login	Japan	
1	10522.0	2014-05-02 11:02:53	engagement	home_page	Japan	
2	10522.0	2014-05-02 11:03:28	engagement	like_message	Japan	
3	10522.0	2014-05-02 11:04:09	engagement	view_inbox	Japan	

```

4 10522.0 2014-05-02 11:03:16 engagement search_run Japan

              device user_type
0 dell inspiron notebook      3.0
1 dell inspiron notebook      3.0
2 dell inspiron notebook      3.0
3 dell inspiron notebook      3.0
4 dell inspiron notebook      3.0

```

Cool, there are a lot of different things that we can stratify on to see where the dip is coming from and if it's universal. Odds are that there's something in particular that's causing the decrease in engagement, and not an overall trend for everyone. But we'll have to test that hypothesis.

```

In [7]: print(f'Event types are:{events.event_type.unique()}')
        print(f'Event names are:{events.event_name.unique()}')
        print(f'Locations are:{events.location.unique()}')

```

```

Event types are:['engagement' 'signup_flow']
Event names are:['login' 'home_page' 'like_message' 'view_inbox' 'search_run'
'send_message' 'search_autocomplete' 'search_click_result_10'
'create_user' 'enter_email' 'enter_info' 'complete_signup'
'search_click_result_7' 'search_click_result_8' 'search_click_result_1'
'search_click_result_3' 'search_click_result_2' 'search_click_result_5'
'search_click_result_6' 'search_click_result_9' 'search_click_result_4']
Locations are:['Japan' 'Netherlands' 'Austria' 'Finland' 'United Kingdom' 'India'
'United States' 'France' 'Iran' 'Germany' 'Australia' 'Brazil' 'Thailand'
'Russia' 'Taiwan' 'Canada' 'Spain' 'Israel' 'Colombia' 'Iraq' 'Indonesia'
'Greece' 'Norway' 'United Arab Emirates' 'Korea' 'Venezuela' 'Belgium'
'Saudi Arabia' 'Poland' 'Sweden' 'Denmark' 'Mexico' 'Italy' 'Egypt'
'Nigeria' 'Pakistan' 'Portugal' 'Singapore' 'South Africa' 'Hong Kong'
'Switzerland' 'Turkey' 'Chile' 'Ireland' 'Argentina' 'Malaysia'
'Philippines']

```

Cool, so we can see that we have two types of events, and we're dealing with a decrease in engagement type events. Finally, let's get a look at what's in the users table before we get merging or further explore things.

```

In [8]: users.head()

```

```

Out[8]:   user_id      created_at  company_id language  activated_at \
0      0.0 2013-01-01 20:59:39      5737.0  english 2013-01-01 21:01:07
1      1.0 2013-01-01 13:07:46       28.0  english                NaN
2      2.0 2013-01-01 10:59:05       51.0  english                NaN
3      3.0 2013-01-01 18:40:36      2800.0   german 2013-01-01 18:42:02
4      4.0 2013-01-01 14:37:51      5110.0   indian 2013-01-01 14:39:05

      state
0  active

```

```

1 pending
2 pending
3 active
4 active

```

This one looks pretty straightforward- userID, what language they speak, what company they are with, and when they activated (finished registering and started using Yammer?)

```
In [11]: rollup.tail()
```

```

Out[11]:
      period_id      time_id      pst_start \
55997      2007.0  2015-12-31  19:00:00  2015-12-24  19:00:00
55998      2007.0  2015-12-31  20:00:00  2015-12-24  20:00:00
55999      2007.0  2015-12-31  21:00:00  2015-12-24  21:00:00
56000      2007.0  2015-12-31  22:00:00  2015-12-24  22:00:00
56001      2007.0  2015-12-31  23:00:00  2015-12-24  23:00:00

      pst_end      utc_start      utc_end
55997  2015-12-31  19:00:00  2015-12-25  03:00:00  2016-01-01  03:00:00
55998  2015-12-31  20:00:00  2015-12-25  04:00:00  2016-01-01  04:00:00
55999  2015-12-31  21:00:00  2015-12-25  05:00:00  2016-01-01  05:00:00
56000  2015-12-31  22:00:00  2015-12-25  06:00:00  2016-01-01  06:00:00
56001  2015-12-31  23:00:00  2015-12-25  07:00:00  2016-01-01  07:00:00

```

Ok, so now I've taken a look at what all we've got. I also went onto the Mode Analytics website where this chart was generated and pulled the SQL query so that I could know just what I'm looking at here.

```

SELECT DATE_TRUNC('week', e.occurred_at),
       COUNT(DISTINCT e.user_id) AS weekly_active_users
FROM tutorial.yammer_events e
WHERE e.event_type = 'engagement'
      AND e.event_name = 'login'
GROUP BY 1
ORDER BY 1

```

So, I'm no SQL expert, but this looks like it's not engagement, but just 'logins'. If it's just logins, we've got a different problem that we want to investigate. But I'm going to assume I don't know anything about SQL and that the underlying assumptions of the challenge are correct- IE this is a drop in **engagement** and not just logins.

The data goes into the future, so we will use the 'occurred at' values to select only those that are in our time of interest

```
In [9]: events.dtypes
```

```

Out[9]: user_id      float64
        occurred_at  object
        event_type   object
        event_name    object

```

```

location      object
device        object
user_type     float64
dtype: object

```

Because the 'occurred_at' columns are objects, we won't be able to properly select them. Thus, we will change them into pandas datetime objects.

```

In [10]: events.occurred_at = pd.to_datetime(events.occurred_at)
         emails.occurred_at = pd.to_datetime(emails.occurred_at)

```

```

In [11]: events_current = events[events['occurred_at'] < pd.to_datetime('2014-9-2')]
         emails_current = emails[emails['occurred_at'] < pd.to_datetime('2014-9-2')]

```

```

In [12]: # test = pd.merge(events_current, users, how='left', on='user_id')

```

1 What could be happening?!?

There are two likely culprits in this kind of case: First, something broke. If it's not that, then we might have a more complicated problem. It's also possible that this is just normal. Let's investigate the possibilities in a little more detail.

1. It's a seasonal variation and is totally normal
2. The website broke
3. The internet somewhere broke
4. We quit sending emails
5. We sent some offensive email
6. Nothing is wrong. It's literally just random noise

1.1 Hypthesis 1 - It's nothing but a seasonal variation

```

In [27]: ## Hypothesis 1 - It's seasonal variation and nothing is wrong- we could look at the
         ## previous year and see if a similar peak and dip occurred.
         events.head()

```

```

Out[27]:
  user_id  occurred_at  event_type  event_name location \
0  10522.0  2014-05-02  11:02:39  engagement      login    Japan
1  10522.0  2014-05-02  11:02:53  engagement  home_page    Japan
2  10522.0  2014-05-02  11:03:28  engagement  like_message  Japan
3  10522.0  2014-05-02  11:04:09  engagement  view_inbox   Japan
4  10522.0  2014-05-02  11:03:16  engagement  search_run   Japan

  device  user_type
0  dell inspiron notebook      3.0
1  dell inspiron notebook      3.0
2  dell inspiron notebook      3.0
3  dell inspiron notebook      3.0
4  dell inspiron notebook      3.0

```

Well, unfortunately we don't have the data from last year to tell us if it's just a seasonal variation. We could ask someone else at the company, but let's assume that the boss was here last year and judging by his strong response to the graph, this isn't actually normal. Next hypothesis.

1.2 Hypothesis 2: The website was down and people couldn't log in.

This one is a little hairy- we have the original data that says it's engagement, so if that's true then we want to know about logins specifically because those could be driving other engagement factors. A broken login means no one can engage in any other way. So we will test to see if log-ins preceded the drop in total engagement.

```
In [28]: ## The website broke. This means we would have seen a lot less log-ins than usual
logins = events_current[events_current['event_name']=='login']; logins.head()
```

```
Out [28]:
```

	user_id	occurred_at	event_type	event_name	location	\
0	10522.0	2014-05-02 11:02:39	engagement	login	Japan	
6	10612.0	2014-05-01 09:59:46	engagement	login	Netherlands	
18	10736.0	2014-05-09 17:52:45	engagement	login	Austria	
21	10965.0	2014-05-15 13:52:59	engagement	login	Finland	
23	11020.0	2014-05-08 09:15:35	engagement	login	Japan	

	device	user_type
0	dell inspiron notebook	3.0
6	iphone 5	1.0
18	iphone 4s	2.0
21	windows surface	3.0
23	macbook air	2.0

```
In [29]: by_loc = logins.groupby(by='location')
```

```
In [30]: test = logins.groupby(by='occurred_at').location.value_counts()
```

```
In [31]: ts = pd.Series(range(len(logins)), index=logins.occurred_at, )
# In [9]: idx = pd.date_range('2018-01-01', periods=5, freq='H')
```

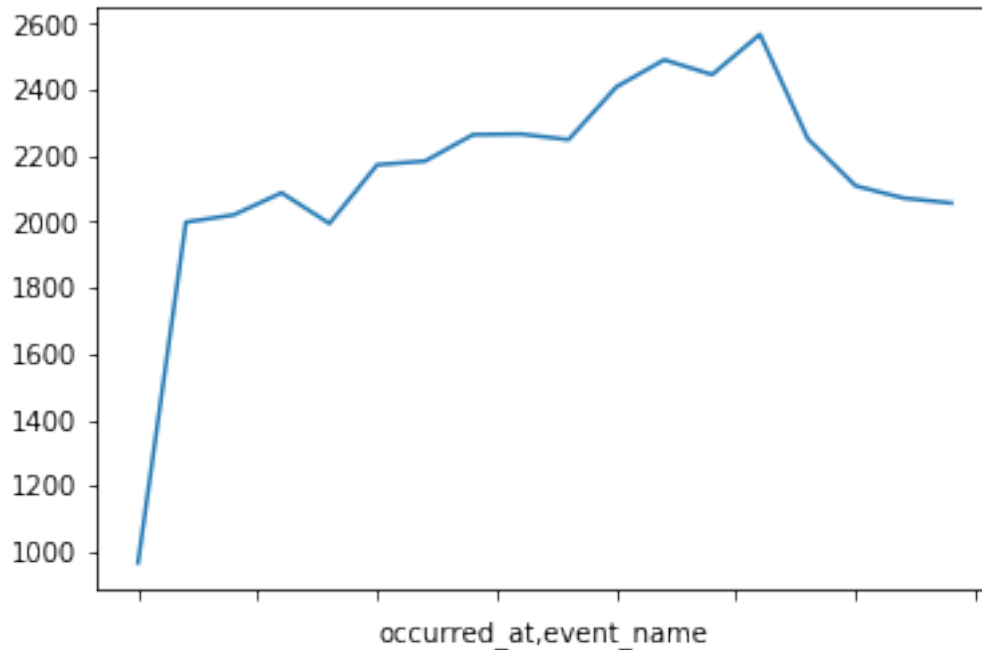
```
In [32]: # logins.occurred_at.resample('7D')
new = logins.set_index('occurred_at')

resamp = new.resample('W').nunique()
```

```
In [33]: resamp = new.resample('W')['event_name'].value_counts()
```

```
In [34]: # plt.plot(x=resamp.index[0], y=resamp)
resamp.plot()
```

```
Out [34]: <matplotlib.axes._subplots.AxesSubplot at 0x7f257619e208>
```



The above graph shows the number of logins, a specific type of engagement, by week. It looks pretty similar, but let's take a look at the numbers

In [35]: resamp

Out [35]:

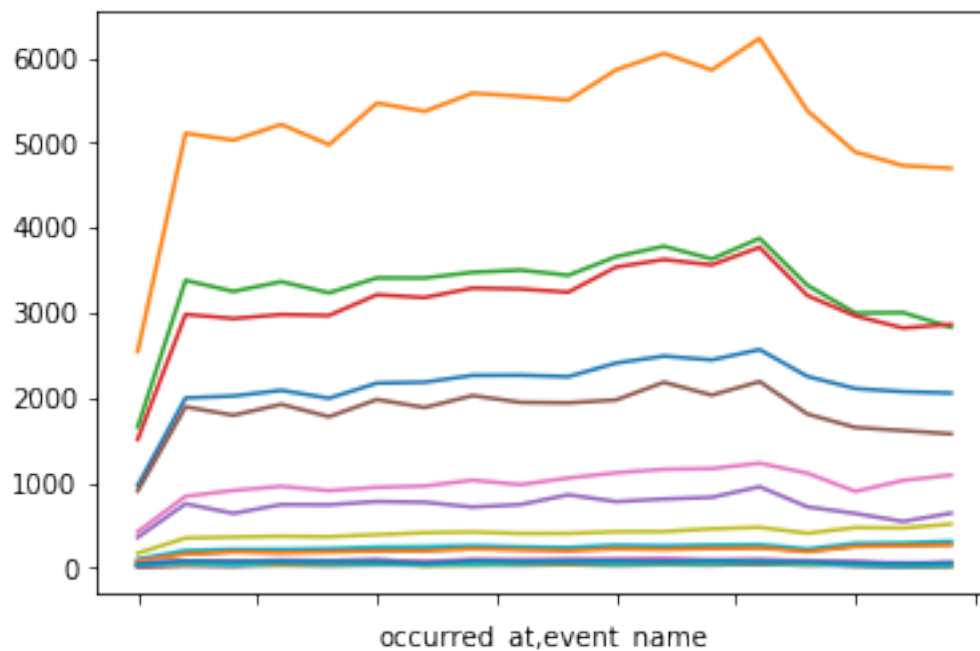
occurred_at	event_name	
2014-05-04	login	967
2014-05-11	login	1999
2014-05-18	login	2021
2014-05-25	login	2088
2014-06-01	login	1995
2014-06-08	login	2173
2014-06-15	login	2184
2014-06-22	login	2264
2014-06-29	login	2266
2014-07-06	login	2249
2014-07-13	login	2409
2014-07-20	login	2491
2014-07-27	login	2446
2014-08-03	login	2568
2014-08-10	login	2252
2014-08-17	login	2109
2014-08-24	login	2072
2014-08-31	login	2057

Name: event_name, dtype: int64

OK so this is kind of weird. We see the drop in engagement from the 27th to the 3rd, mostly. Then we see a decrease in logins the following week. So it looks like it's very possible that something broke, people couldn't use the product, and then they didn't log in the next week because they thought it would still be down. We can see that it wasn't a break in the ability to log in, which means that it was probably something with the functionality of the site.

```
In [48]: # new['counts'] = logins.event_name.value_counts()
         # let's look for each different type of event and see if any of them dropped to zero.
```

```
In [44]: ev_tests = [events.event_name.unique()]
         z=0
         for name in ev_tests[0]:
             stats = events_current[events_current['event_name']==name]
             new = stats.set_index('occurred_at')
             resamp = new.resample('W')['event_name'].value_counts()
             resamp.plot()
```



```
In [49]: #I didn't put a key on here because this is a quick and dirty look, and then I dive i
         # This cell shows what categories are displayed on the above chart in the various col
         print (ev_tests[0])
```

```
['login' 'home_page' 'like_message' 'view_inbox' 'search_run'
 'send_message' 'search_autocomplete' 'search_click_result_10'
 'create_user' 'enter_email' 'enter_info' 'complete_signup'
 'search_click_result_7' 'search_click_result_8' 'search_click_result_1'
 'search_click_result_3' 'search_click_result_2' 'search_click_result_5']
```

```
'search_click_result_6' 'search_click_result_9' 'search_click_result_4']
```

According to the above graph, it looks like this drop in engagement wasn't caused by some certain part of the website breaking. We've looked into each separate engagement metric and none of them went to zero or gave any indication that they aren't working. We can look at the underlying numbers below, but they aren't promising.

```
In [47]: ev_tests = [events.event_name.unique()]
        z=0
        for name in ev_tests[0]:
            stats = events_current[events_current['event_name']==name]
            new = stats.set_index('occurred_at')
            resamp = new.resample('W')['event_name'].value_counts()
            print(resamp)
```

occurred_at	event_name	
2014-05-04	login	967
2014-05-11	login	1999
2014-05-18	login	2021
2014-05-25	login	2088
2014-06-01	login	1995
2014-06-08	login	2173
2014-06-15	login	2184
2014-06-22	login	2264
2014-06-29	login	2266
2014-07-06	login	2249
2014-07-13	login	2409
2014-07-20	login	2491
2014-07-27	login	2446
2014-08-03	login	2568
2014-08-10	login	2252
2014-08-17	login	2109
2014-08-24	login	2072
2014-08-31	login	2057

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	home_page	2546
2014-05-11	home_page	5113
2014-05-18	home_page	5032
2014-05-25	home_page	5217
2014-06-01	home_page	4978
2014-06-08	home_page	5469
2014-06-15	home_page	5371
2014-06-22	home_page	5585
2014-06-29	home_page	5548
2014-07-06	home_page	5503
2014-07-13	home_page	5859

2014-07-20	home_page	6053
2014-07-27	home_page	5858
2014-08-03	home_page	6230
2014-08-10	home_page	5378
2014-08-17	home_page	4892
2014-08-24	home_page	4732
2014-08-31	home_page	4701

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	like_message	1657
2014-05-11	like_message	3382
2014-05-18	like_message	3252
2014-05-25	like_message	3367
2014-06-01	like_message	3235
2014-06-08	like_message	3411
2014-06-15	like_message	3410
2014-06-22	like_message	3477
2014-06-29	like_message	3503
2014-07-06	like_message	3442
2014-07-13	like_message	3661
2014-07-20	like_message	3784
2014-07-27	like_message	3634
2014-08-03	like_message	3877
2014-08-10	like_message	3324
2014-08-17	like_message	2998
2014-08-24	like_message	3003
2014-08-31	like_message	2831

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	view_inbox	1510
2014-05-11	view_inbox	2982
2014-05-18	view_inbox	2933
2014-05-25	view_inbox	2980
2014-06-01	view_inbox	2969
2014-06-08	view_inbox	3213
2014-06-15	view_inbox	3180
2014-06-22	view_inbox	3291
2014-06-29	view_inbox	3281
2014-07-06	view_inbox	3243
2014-07-13	view_inbox	3539
2014-07-20	view_inbox	3627
2014-07-27	view_inbox	3564
2014-08-03	view_inbox	3771
2014-08-10	view_inbox	3203
2014-08-17	view_inbox	2964
2014-08-24	view_inbox	2821
2014-08-31	view_inbox	2865

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_run	358
2014-05-11	search_run	753
2014-05-18	search_run	640
2014-05-25	search_run	744
2014-06-01	search_run	741
2014-06-08	search_run	780
2014-06-15	search_run	771
2014-06-22	search_run	715
2014-06-29	search_run	744
2014-07-06	search_run	859
2014-07-13	search_run	779
2014-07-20	search_run	809
2014-07-27	search_run	829
2014-08-03	search_run	954
2014-08-10	search_run	717
2014-08-17	search_run	639
2014-08-24	search_run	545
2014-08-31	search_run	642

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	send_message	904
2014-05-11	send_message	1894
2014-05-18	send_message	1795
2014-05-25	send_message	1923
2014-06-01	send_message	1778
2014-06-08	send_message	1980
2014-06-15	send_message	1886
2014-06-22	send_message	2026
2014-06-29	send_message	1947
2014-07-06	send_message	1941
2014-07-13	send_message	1975
2014-07-20	send_message	2183
2014-07-27	send_message	2032
2014-08-03	send_message	2191
2014-08-10	send_message	1809
2014-08-17	send_message	1651
2014-08-24	send_message	1614
2014-08-31	send_message	1576

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_autocomplete	424
2014-05-11	search_autocomplete	841
2014-05-18	search_autocomplete	910
2014-05-25	search_autocomplete	958
2014-06-01	search_autocomplete	908
2014-06-08	search_autocomplete	947
2014-06-15	search_autocomplete	962

2014-06-22	search_autocomplete	1031
2014-06-29	search_autocomplete	978
2014-07-06	search_autocomplete	1056
2014-07-13	search_autocomplete	1118
2014-07-20	search_autocomplete	1159
2014-07-27	search_autocomplete	1167
2014-08-03	search_autocomplete	1231
2014-08-10	search_autocomplete	1113
2014-08-17	search_autocomplete	898
2014-08-24	search_autocomplete	1028
2014-08-31	search_autocomplete	1091

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_10	15
2014-05-11	search_click_result_10	26
2014-05-18	search_click_result_10	19
2014-05-25	search_click_result_10	31
2014-06-01	search_click_result_10	24
2014-06-08	search_click_result_10	33
2014-06-15	search_click_result_10	29
2014-06-22	search_click_result_10	30
2014-06-29	search_click_result_10	29
2014-07-06	search_click_result_10	37
2014-07-13	search_click_result_10	26
2014-07-20	search_click_result_10	35
2014-07-27	search_click_result_10	30
2014-08-03	search_click_result_10	46
2014-08-10	search_click_result_10	27
2014-08-17	search_click_result_10	34
2014-08-24	search_click_result_10	12
2014-08-31	search_click_result_10	23

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	create_user	171
2014-05-11	create_user	350
2014-05-18	create_user	362
2014-05-25	create_user	371
2014-06-01	create_user	366
2014-06-08	create_user	390
2014-06-15	create_user	413
2014-06-22	create_user	421
2014-06-29	create_user	404
2014-07-06	create_user	405
2014-07-13	create_user	424
2014-07-20	create_user	426
2014-07-27	create_user	458
2014-08-03	create_user	476
2014-08-10	create_user	406

2014-08-17	create_user	473
2014-08-24	create_user	468
2014-08-31	create_user	514

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	enter_email	102
2014-05-11	enter_email	205
2014-05-18	enter_email	218
2014-05-25	enter_email	218
2014-06-01	enter_email	227
2014-06-08	enter_email	246
2014-06-15	enter_email	250
2014-06-22	enter_email	262
2014-06-29	enter_email	247
2014-07-06	enter_email	239
2014-07-13	enter_email	269
2014-07-20	enter_email	260
2014-07-27	enter_email	271
2014-08-03	enter_email	274
2014-08-10	enter_email	223
2014-08-17	enter_email	289
2014-08-24	enter_email	294
2014-08-31	enter_email	313

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	enter_info	86
2014-05-11	enter_info	169
2014-05-18	enter_info	194
2014-05-25	enter_info	189
2014-06-01	enter_info	199
2014-06-08	enter_info	210
2014-06-15	enter_info	212
2014-06-22	enter_info	231
2014-06-29	enter_info	220
2014-07-06	enter_info	214
2014-07-13	enter_info	231
2014-07-20	enter_info	229
2014-07-27	enter_info	235
2014-08-03	enter_info	244
2014-08-10	enter_info	199
2014-08-17	enter_info	258
2014-08-24	enter_info	269
2014-08-31	enter_info	283

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	complete_signup	81
2014-05-11	complete_signup	160
2014-05-18	complete_signup	186

2014-05-25	complete_signup	177
2014-06-01	complete_signup	186
2014-06-08	complete_signup	197
2014-06-15	complete_signup	198
2014-06-22	complete_signup	222
2014-06-29	complete_signup	210
2014-07-06	complete_signup	199
2014-07-13	complete_signup	223
2014-07-20	complete_signup	215
2014-07-27	complete_signup	228
2014-08-03	complete_signup	234
2014-08-10	complete_signup	189
2014-08-17	complete_signup	250
2014-08-24	complete_signup	259
2014-08-31	complete_signup	266

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_7	37
2014-05-11	search_click_result_7	36
2014-05-18	search_click_result_7	36
2014-05-25	search_click_result_7	35
2014-06-01	search_click_result_7	47
2014-06-08	search_click_result_7	49
2014-06-15	search_click_result_7	32
2014-06-22	search_click_result_7	38
2014-06-29	search_click_result_7	38
2014-07-06	search_click_result_7	55
2014-07-13	search_click_result_7	44
2014-07-20	search_click_result_7	36
2014-07-27	search_click_result_7	40
2014-08-03	search_click_result_7	50
2014-08-10	search_click_result_7	47
2014-08-17	search_click_result_7	32
2014-08-24	search_click_result_7	30
2014-08-31	search_click_result_7	27

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_8	14
2014-05-11	search_click_result_8	33
2014-05-18	search_click_result_8	30
2014-05-25	search_click_result_8	45
2014-06-01	search_click_result_8	42
2014-06-08	search_click_result_8	48
2014-06-15	search_click_result_8	36
2014-06-22	search_click_result_8	47
2014-06-29	search_click_result_8	42
2014-07-06	search_click_result_8	48
2014-07-13	search_click_result_8	52

2014-07-20	search_click_result_8	50
2014-07-27	search_click_result_8	42
2014-08-03	search_click_result_8	40
2014-08-10	search_click_result_8	43
2014-08-17	search_click_result_8	26
2014-08-24	search_click_result_8	24
2014-08-31	search_click_result_8	28

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_1	44
2014-05-11	search_click_result_1	82
2014-05-18	search_click_result_1	73
2014-05-25	search_click_result_1	77
2014-06-01	search_click_result_1	81
2014-06-08	search_click_result_1	98
2014-06-15	search_click_result_1	62
2014-06-22	search_click_result_1	91
2014-06-29	search_click_result_1	84
2014-07-06	search_click_result_1	99
2014-07-13	search_click_result_1	91
2014-07-20	search_click_result_1	94
2014-07-27	search_click_result_1	92
2014-08-03	search_click_result_1	96
2014-08-10	search_click_result_1	74
2014-08-17	search_click_result_1	69
2014-08-24	search_click_result_1	40
2014-08-31	search_click_result_1	66

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_3	41
2014-05-11	search_click_result_3	66
2014-05-18	search_click_result_3	48
2014-05-25	search_click_result_3	80
2014-06-01	search_click_result_3	68
2014-06-08	search_click_result_3	83
2014-06-15	search_click_result_3	58
2014-06-22	search_click_result_3	72
2014-06-29	search_click_result_3	64
2014-07-06	search_click_result_3	78
2014-07-13	search_click_result_3	62
2014-07-20	search_click_result_3	62
2014-07-27	search_click_result_3	67
2014-08-03	search_click_result_3	82
2014-08-10	search_click_result_3	55
2014-08-17	search_click_result_3	58
2014-08-24	search_click_result_3	39
2014-08-31	search_click_result_3	51

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_2	56
2014-05-11	search_click_result_2	78
2014-05-18	search_click_result_2	73
2014-05-25	search_click_result_2	93
2014-06-01	search_click_result_2	85
2014-06-08	search_click_result_2	86
2014-06-15	search_click_result_2	78
2014-06-22	search_click_result_2	98
2014-06-29	search_click_result_2	98
2014-07-06	search_click_result_2	92
2014-07-13	search_click_result_2	104
2014-07-20	search_click_result_2	114
2014-07-27	search_click_result_2	83
2014-08-03	search_click_result_2	91
2014-08-10	search_click_result_2	87
2014-08-17	search_click_result_2	67
2014-08-24	search_click_result_2	60
2014-08-31	search_click_result_2	56

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_5	26
2014-05-11	search_click_result_5	66
2014-05-18	search_click_result_5	43
2014-05-25	search_click_result_5	64
2014-06-01	search_click_result_5	67
2014-06-08	search_click_result_5	71
2014-06-15	search_click_result_5	52
2014-06-22	search_click_result_5	56
2014-06-29	search_click_result_5	64
2014-07-06	search_click_result_5	64
2014-07-13	search_click_result_5	61
2014-07-20	search_click_result_5	48
2014-07-27	search_click_result_5	65
2014-08-03	search_click_result_5	56
2014-08-10	search_click_result_5	53
2014-08-17	search_click_result_5	41
2014-08-24	search_click_result_5	36
2014-08-31	search_click_result_5	35

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_6	41
2014-05-11	search_click_result_6	49
2014-05-18	search_click_result_6	35
2014-05-25	search_click_result_6	48
2014-06-01	search_click_result_6	52
2014-06-08	search_click_result_6	50
2014-06-15	search_click_result_6	28

2014-06-22	search_click_result_6	29
2014-06-29	search_click_result_6	47
2014-07-06	search_click_result_6	54
2014-07-13	search_click_result_6	54
2014-07-20	search_click_result_6	52
2014-07-27	search_click_result_6	52
2014-08-03	search_click_result_6	60
2014-08-10	search_click_result_6	38
2014-08-17	search_click_result_6	50
2014-08-24	search_click_result_6	33
2014-08-31	search_click_result_6	33

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_9	25
2014-05-11	search_click_result_9	54
2014-05-18	search_click_result_9	32
2014-05-25	search_click_result_9	63
2014-06-01	search_click_result_9	48
2014-06-08	search_click_result_9	43
2014-06-15	search_click_result_9	40
2014-06-22	search_click_result_9	42
2014-06-29	search_click_result_9	50
2014-07-06	search_click_result_9	62
2014-07-13	search_click_result_9	47
2014-07-20	search_click_result_9	52
2014-07-27	search_click_result_9	47
2014-08-03	search_click_result_9	54
2014-08-10	search_click_result_9	48
2014-08-17	search_click_result_9	29
2014-08-24	search_click_result_9	19
2014-08-31	search_click_result_9	29

Name: event_name, dtype: int64

occurred_at	event_name	
2014-05-04	search_click_result_4	44
2014-05-11	search_click_result_4	78
2014-05-18	search_click_result_4	75
2014-05-25	search_click_result_4	77
2014-06-01	search_click_result_4	75
2014-06-08	search_click_result_4	74
2014-06-15	search_click_result_4	54
2014-06-22	search_click_result_4	84
2014-06-29	search_click_result_4	76
2014-07-06	search_click_result_4	77
2014-07-13	search_click_result_4	84
2014-07-20	search_click_result_4	74
2014-07-27	search_click_result_4	84
2014-08-03	search_click_result_4	75
2014-08-10	search_click_result_4	73


```

2014-08-17    search_click_result_4    55
2014-08-24    search_click_result_4    50
2014-08-31    search_click_result_4    55
Name: event_name, dtype: int64

```

1.3 Hypothesis 3: The internet broke somewhere

```

In [64]: # events_current.set_index('occurred_at', inplace=True)
         locs=events_current.groupby(by='location')
         resamp = locs.resample('W')#['location'].value_counts()

```

```

In [65]: locs.head()

```

```

Out [65]:

```

		user_id	event_type	event_name	location \
occurred_at					
2014-05-02 11:02:39	10522.0	engagement	login	Japan	
2014-05-02 11:02:53	10522.0	engagement	home_page	Japan	
2014-05-02 11:03:28	10522.0	engagement	like_message	Japan	
2014-05-02 11:04:09	10522.0	engagement	view_inbox	Japan	
2014-05-02 11:03:16	10522.0	engagement	search_run	Japan	
2014-05-01 09:59:46	10612.0	engagement	login	Netherlands	
2014-05-01 10:00:18	10612.0	engagement	like_message	Netherlands	
2014-05-01 10:00:53	10612.0	engagement	send_message	Netherlands	
2014-05-01 10:01:24	10612.0	engagement	home_page	Netherlands	
2014-05-01 10:01:52	10612.0	engagement	like_message	Netherlands	
2014-05-09 17:52:45	10736.0	engagement	login	Austria	
2014-05-09 17:53:16	10736.0	engagement	like_message	Austria	
2014-05-09 17:53:51	10736.0	engagement	send_message	Austria	
2014-05-15 13:52:59	10965.0	engagement	login	Finland	
2014-05-15 13:53:31	10965.0	engagement	home_page	Finland	
2014-05-01 07:25:08	11037.0	engagement	login	United Kingdom	
2014-05-01 07:25:39	11037.0	engagement	like_message	United Kingdom	
2014-05-01 07:26:04	11037.0	engagement	home_page	United Kingdom	
2014-05-01 07:26:23	11037.0	engagement	like_message	United Kingdom	
2014-05-01 07:26:58	11037.0	engagement	home_page	United Kingdom	
2014-05-08 11:13:33	11133.0	engagement	login	India	
2014-05-08 11:14:10	11133.0	engagement	home_page	India	
2014-05-08 11:14:42	11133.0	engagement	like_message	India	
2014-05-08 11:15:16	11133.0	engagement	home_page	India	
2014-05-08 11:15:44	11133.0	engagement	send_message	India	
2014-05-06 16:17:58	11194.0	engagement	login	United States	
2014-05-06 16:18:24	11194.0	engagement	home_page	United States	
2014-05-06 16:19:00	11194.0	engagement	view_inbox	United States	
2014-05-06 16:19:35	11194.0	engagement	home_page	United States	
2014-05-06 16:20:03	11194.0	engagement	send_message	United States	
...	
2014-05-09 10:44:31	12201.0	signup_flow	create_user	Singapore	

2014-05-09	11:27:30	12213.0	signup_flow	create_user	Iran
2014-05-09	15:26:43	12220.0	signup_flow	create_user	Ireland
2014-05-09	15:27:18	12220.0	signup_flow	enter_email	Ireland
2014-05-09	10:42:08	12226.0	signup_flow	create_user	Argentina
2014-05-09	04:38:15	12240.0	signup_flow	create_user	South Africa
2014-05-09	18:26:30	12247.0	signup_flow	create_user	Ireland
2014-05-09	18:26:54	12247.0	signup_flow	enter_email	Ireland
2014-05-09	18:27:08	12247.0	signup_flow	enter_info	Ireland
2014-05-12	09:40:12	12311.0	signup_flow	create_user	Malaysia
2014-05-12	09:40:50	12311.0	signup_flow	enter_email	Malaysia
2014-05-12	09:41:27	12311.0	signup_flow	enter_info	Malaysia
2014-05-12	09:41:54	12311.0	signup_flow	complete_signup	Malaysia
2014-05-12	09:41:54	12311.0	engagement	login	Malaysia
2014-05-13	15:11:59	12359.0	signup_flow	create_user	Philippines
2014-05-13	15:12:32	12359.0	signup_flow	enter_email	Philippines
2014-05-14	13:48:14	12431.0	signup_flow	create_user	Singapore
2014-05-14	12:42:33	12437.0	signup_flow	create_user	Philippines
2014-05-14	12:42:58	12437.0	signup_flow	enter_email	Philippines
2014-05-15	07:21:09	12487.0	signup_flow	create_user	Philippines
2014-05-15	17:40:11	12490.0	signup_flow	create_user	Argentina
2014-05-15	17:40:38	12490.0	signup_flow	enter_email	Argentina
2014-05-15	17:41:02	12490.0	signup_flow	enter_info	Argentina
2014-05-15	17:41:35	12490.0	signup_flow	complete_signup	Argentina
2014-05-15	05:15:06	12501.0	signup_flow	create_user	Iraq
2014-05-15	05:15:32	12501.0	signup_flow	enter_email	Iraq
2014-05-15	05:16:01	12501.0	signup_flow	enter_info	Iraq
2014-05-17	15:07:55	12614.0	signup_flow	create_user	Chile
2014-05-17	15:08:20	12614.0	signup_flow	enter_email	Chile
2014-05-19	08:38:30	12699.0	signup_flow	create_user	Singapore

			device	user_type
occurred_at				
2014-05-02	11:02:39	dell inspiron notebook		3.0
2014-05-02	11:02:53	dell inspiron notebook		3.0
2014-05-02	11:03:28	dell inspiron notebook		3.0
2014-05-02	11:04:09	dell inspiron notebook		3.0
2014-05-02	11:03:16	dell inspiron notebook		3.0
2014-05-01	09:59:46	iphone 5		1.0
2014-05-01	10:00:18	iphone 5		1.0
2014-05-01	10:00:53	iphone 5		1.0
2014-05-01	10:01:24	iphone 5		1.0
2014-05-01	10:01:52	iphone 5		1.0
2014-05-09	17:52:45	iphone 4s		2.0
2014-05-09	17:53:16	iphone 4s		2.0
2014-05-09	17:53:51	iphone 4s		2.0
2014-05-15	13:52:59	windows surface		3.0
2014-05-15	13:53:31	windows surface		3.0
2014-05-01	07:25:08	iphone 4s		1.0

2014-05-01 07:25:39	iphone 4s	1.0
2014-05-01 07:26:04	iphone 4s	1.0
2014-05-01 07:26:23	iphone 4s	1.0
2014-05-01 07:26:58	iphone 4s	1.0
2014-05-08 11:13:33	macbook pro	2.0
2014-05-08 11:14:10	macbook pro	2.0
2014-05-08 11:14:42	macbook pro	2.0
2014-05-08 11:15:16	macbook pro	2.0
2014-05-08 11:15:44	macbook pro	2.0
2014-05-06 16:17:58	ipad mini	2.0
2014-05-06 16:18:24	ipad mini	2.0
2014-05-06 16:19:00	ipad mini	2.0
2014-05-06 16:19:35	ipad mini	2.0
2014-05-06 16:20:03	ipad mini	2.0
...
2014-05-09 10:44:31	macbook pro	NaN
2014-05-09 11:27:30	iphone 5s	NaN
2014-05-09 15:26:43	lenovo thinkpad	NaN
2014-05-09 15:27:18	lenovo thinkpad	NaN
2014-05-09 10:42:08	asus chromebook	NaN
2014-05-09 04:38:15	samsung galaxy s4	NaN
2014-05-09 18:26:30	ipad air	NaN
2014-05-09 18:26:54	ipad air	NaN
2014-05-09 18:27:08	ipad air	NaN
2014-05-12 09:40:12	windows surface	NaN
2014-05-12 09:40:50	windows surface	NaN
2014-05-12 09:41:27	windows surface	NaN
2014-05-12 09:41:54	windows surface	1.0
2014-05-12 09:41:54	windows surface	1.0
2014-05-13 15:11:59	lenovo thinkpad	NaN
2014-05-13 15:12:32	lenovo thinkpad	NaN
2014-05-14 13:48:14	nexus 5	NaN
2014-05-14 12:42:33	lenovo thinkpad	NaN
2014-05-14 12:42:58	lenovo thinkpad	NaN
2014-05-15 07:21:09	nexus 5	NaN
2014-05-15 17:40:11	acer aspire notebook	NaN
2014-05-15 17:40:38	acer aspire notebook	NaN
2014-05-15 17:41:02	acer aspire notebook	NaN
2014-05-15 17:41:35	acer aspire notebook	3.0
2014-05-15 05:15:06	samsung galaxy s4	NaN
2014-05-15 05:15:32	samsung galaxy s4	NaN
2014-05-15 05:16:01	samsung galaxy s4	NaN
2014-05-17 15:07:55	nexus 5	NaN
2014-05-17 15:08:20	nexus 5	NaN
2014-05-19 08:38:30	iphone 4s	NaN

[235 rows x 6 columns]