Yammer Case Study with Mode Analytics

Brittany Bennett
June 1, 2018

Problem Background

Investigating a Drop in User Engagement

The problem

You show up to work Tuesday morning, September 2, 2014. The head of the Product team walks over to your desk and asks you what you think about the latest activity on the user engagement dashboards. You fire them up, and something immediately jumps out:

The above chart shows the number of engaged users each week. Yammer defines engagement as having made some type of server call by interacting with the product (shown in the data as events of type “engagement”). Any point in this chart can be interpreted as “the number of users who logged at least one engagement event during the week starting on that date.”

You are responsible for determining what caused the dip at the end of the chart shown above and, if appropriate, recommending solutions for the problem.

Getting oriented

Before you even touch the data, come up with a list of possible causes for the dip in retention shown in the chart above. Make a list and determine the order in which you will check them. Make sure to note how you will test each hypothesis. Think carefully about the criteria you use to order them and write down the criteria as well.

Also, make sure you understand what the above chart shows and does not show.

If you want to check your list of possible causes against ours, read the first part of the answer key.

Digging in

Once you have an ordered list of possible problems, it’s time to investigate.

For this problem, you will need to use four tables. The tables names and column definitions are listed below—click a table name to view information about that table. *Note: this data is fake and was generated for the purpose of this case study. It is similar in structure to Yammer's actual data, but for privacy and security reasons it is not real.*

- Table 1: Users
- Table 2: Events
- Table 3: Email Events
- Table 4: Rollup Periods

Making a recommendation

Start to work your way through your list of hypotheses in order to determine the source of the drop in engagement. As you explore, make sure to save your work. It may be helpful to start with the code that produces the above query, which you can find by clicking the link in the footer of the chart and navigating to the "query" tab.

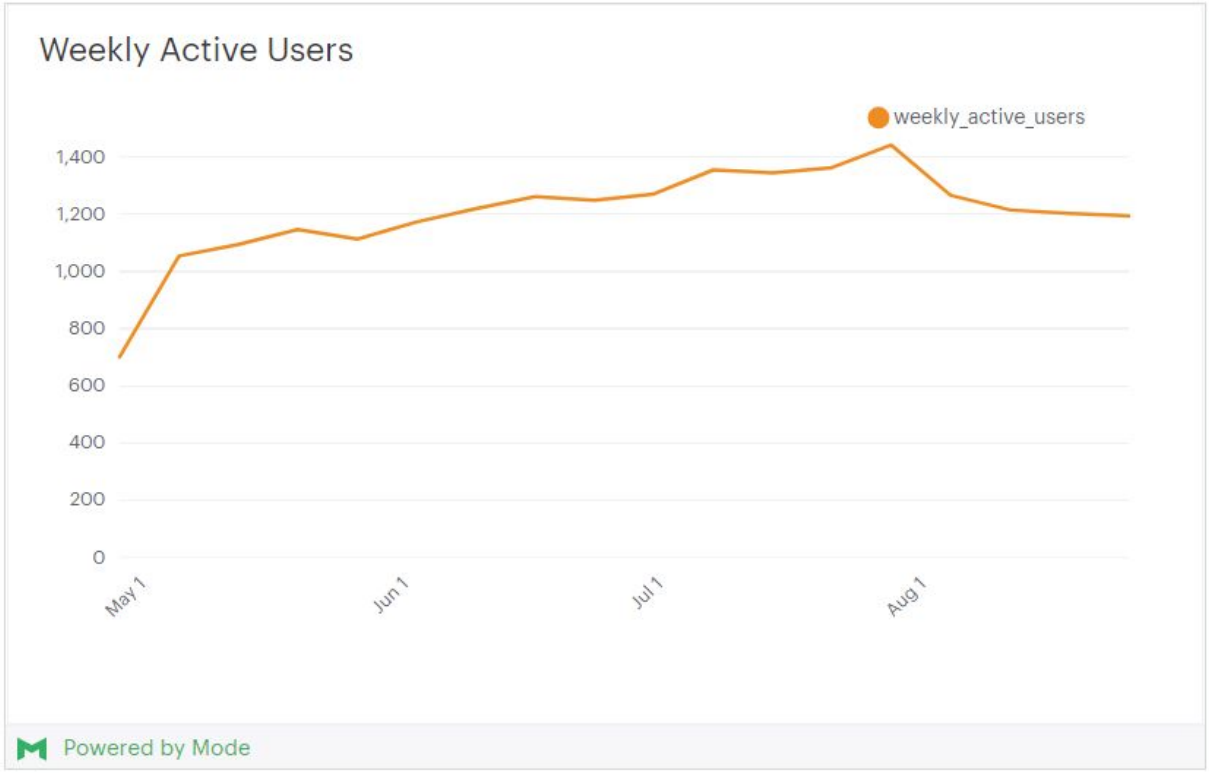
Answer the following questions:

- Do the answers to any of your original hypotheses lead you to further questions?
- If so, what are they and how will you test them?
- If they are questions that you can't answer using data alone, how would you go about answering them (hypothetically, assuming you actually worked at this company)?
- What seems like the most likely cause of the engagement dip?
- What, if anything, should the company do in response?

Answers

Initial Hypotheses

We are told that “Any point in this chart can be interpreted as “the number of users who logged at least one engagement event during the week starting on that date” (see below). According to this graph,t he dip occurs at July 28, 2014, starting with 1,442 users and dropping to 1,266 users on August 4, 2014.



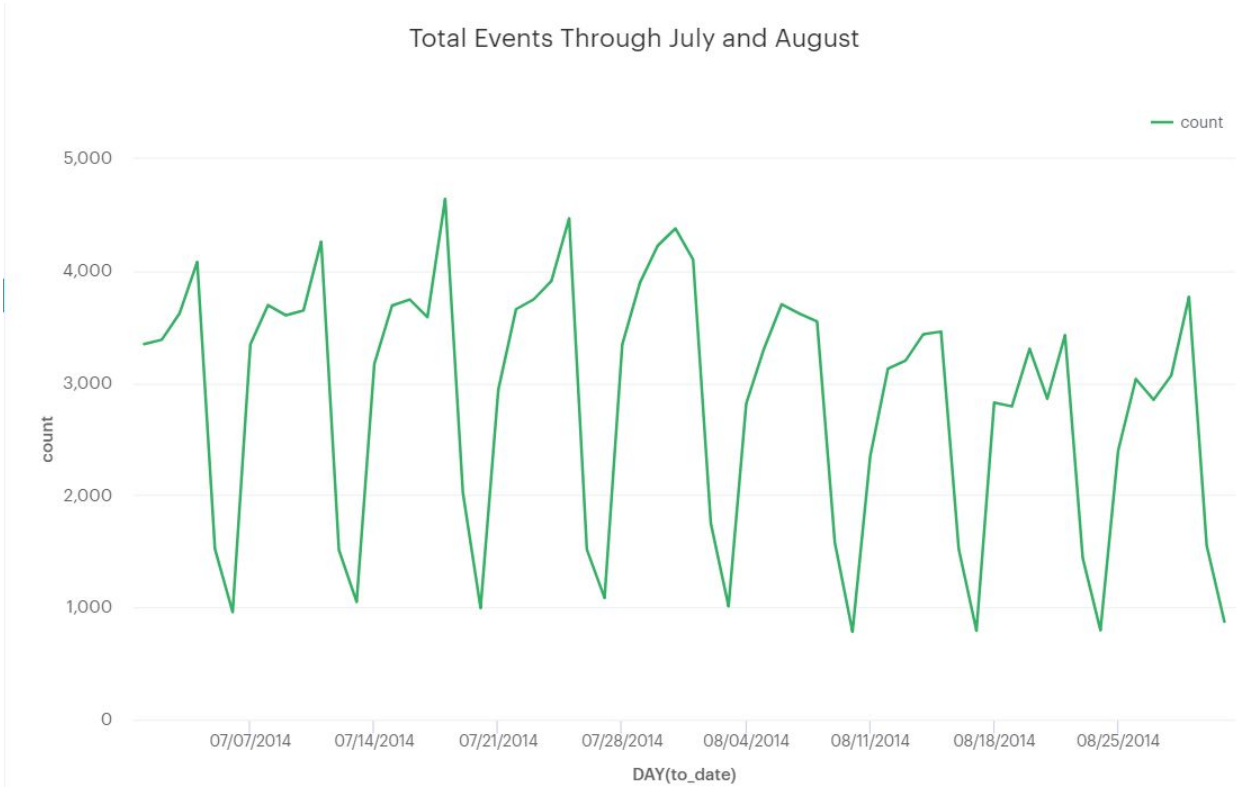
My initial hypothesis are as follows:

- There are fewer users on the site during the last week of August. We could pull all the users of that week and compare it to the weeks before.
- There might be a system error that is preventing users from using certain features of the site. We could verify this hypothesis by searching for error codes or dips in specific types of engagement.

Exploring Events

Below we plot daily logged events through July and August. Indeed, we see that engagement starts to decline at the end of July/beginning of August. We see a dip on August 6th, 2014.

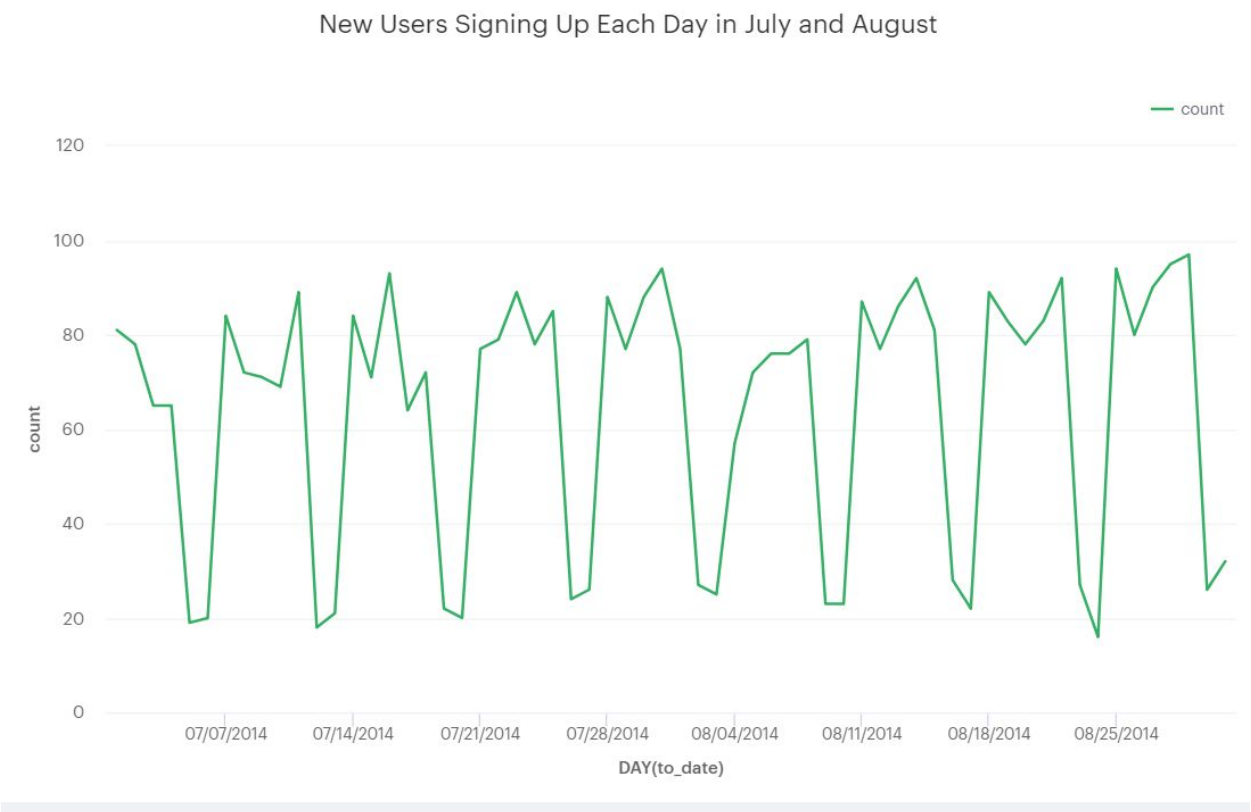
SELECT COUNT(*), to_date FROM(
SELECT device,event_name,event_type,location,user_id,user_type, to_date(cast(occurred_at as TEXT),'YYYY-MM-DD')
FROM tutorial.yammer_events)
as test
Where CAST(to_date AS text) LIKE '2014-07%' OR CAST(to_date AS text) LIKE '2014-08%'
GROUP BY to_date
ORDER BY to_date



Testing User Sign Ups

First, let's explore the number of new users signing up on the website. We see a clear up and down pattern week to week through July and August. However, we see that during the first week of August (around August 5th), the "up" does not go nearly as high as the other peaks. We see only ~75 users created as opposed to 80-90 users created at the highs of other weeks. Therefore, not as many new users signed up right before or during the dip.

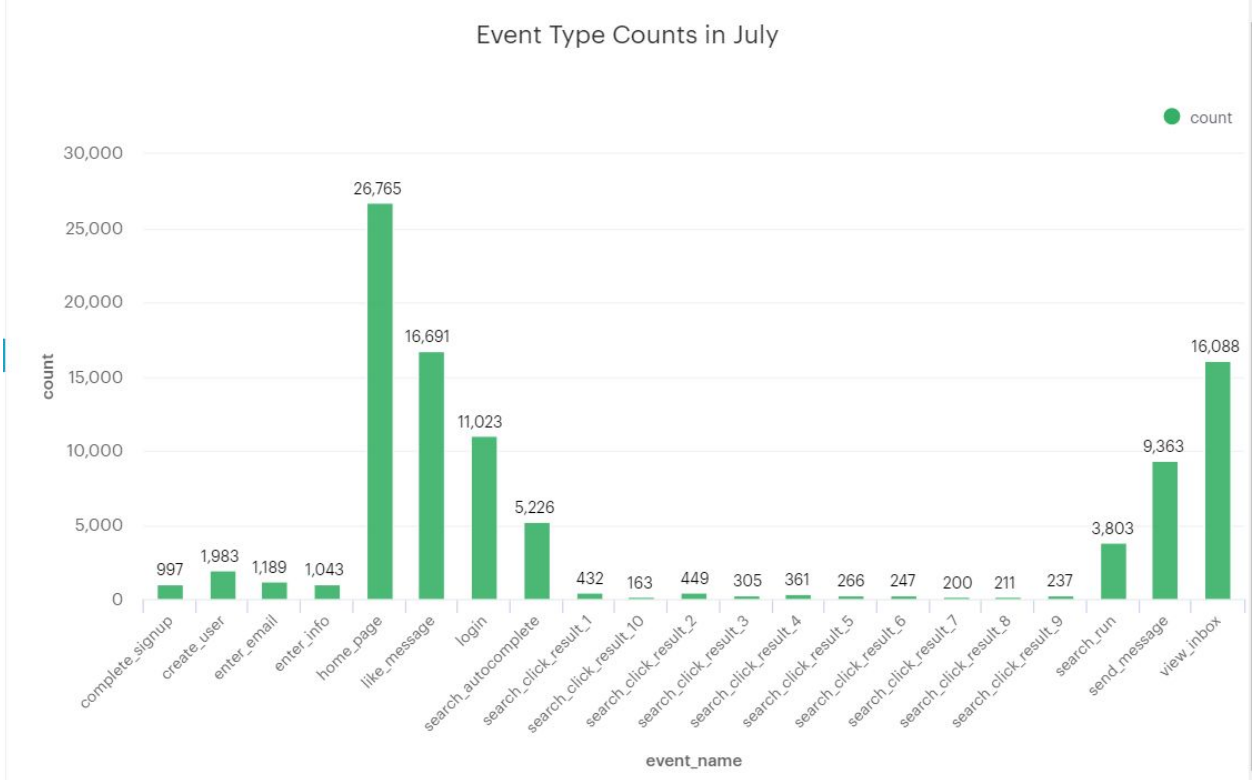
```
SELECT COUNT(*), to_date FROM(
SELECT company_id, language, state, user_id,  to_date(cast(created_at as TEXT),'YYYY-MM-DD')
FROM tutorial.yammer_users)
as test
Where CAST(to_date AS text) LIKE '2014-07%' OR CAST(to_date AS text) LIKE '2014-08%'
GROUP BY to_date
ORDER BY to_date
```

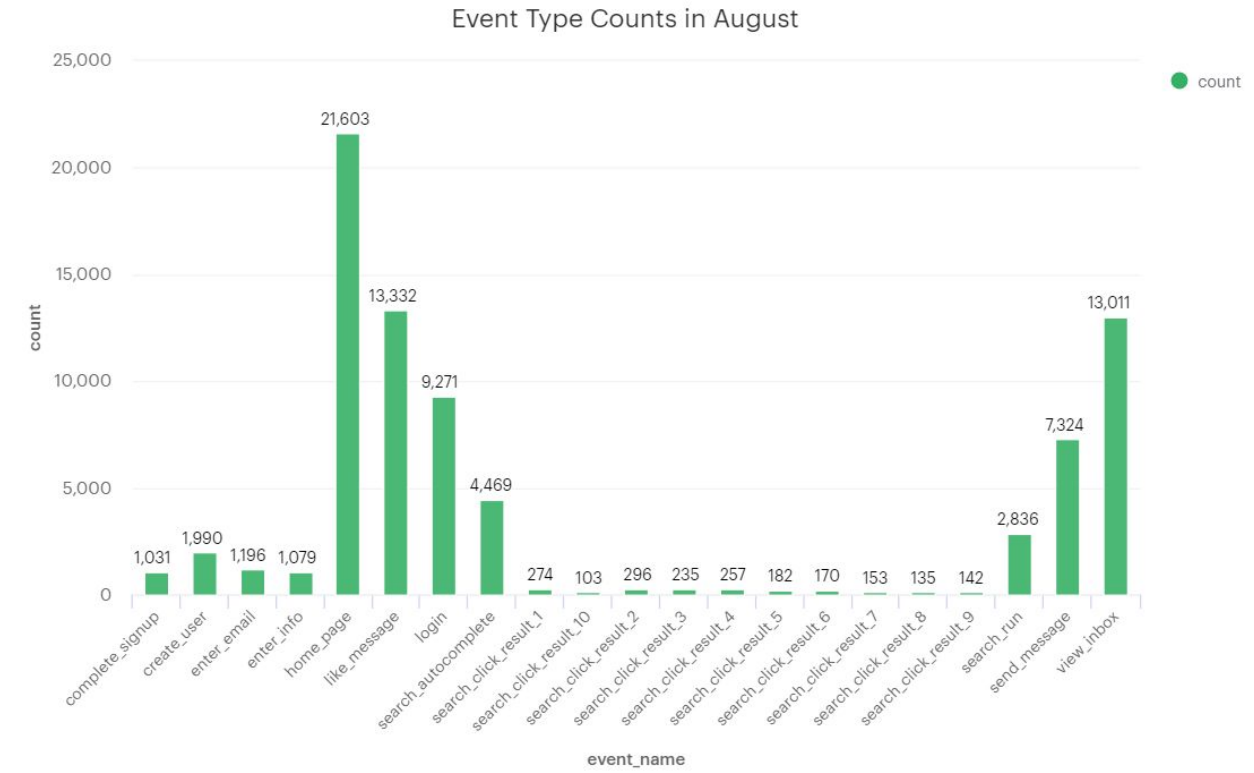


Types of Events

Below we compare the engagements counts for each type of event in July and August to determine if one specific feature was broken. Overall, we see lower engagement counts across the board. No one event stands out.

```
SELECT COUNT(*), to_date FROM(
SELECT device,event_name,event_type,location,user_id,user_type, to_date(cast(occurred_at as TEXT),'YYYY-MM-DD')   FROM tutorial.yammer_events)
as test
Where CAST(to_date AS text) LIKE '2014-07%'
GROUP BY to_date
ORDER BY to_date
```

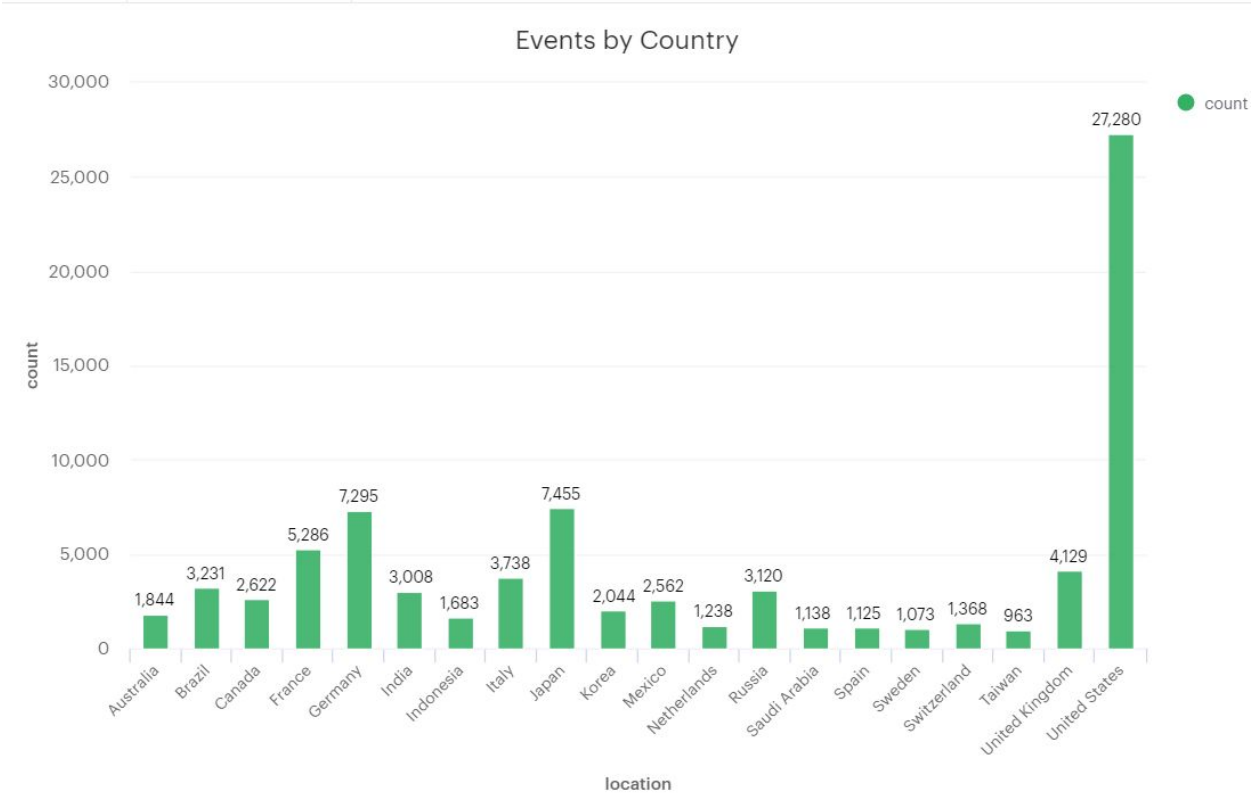


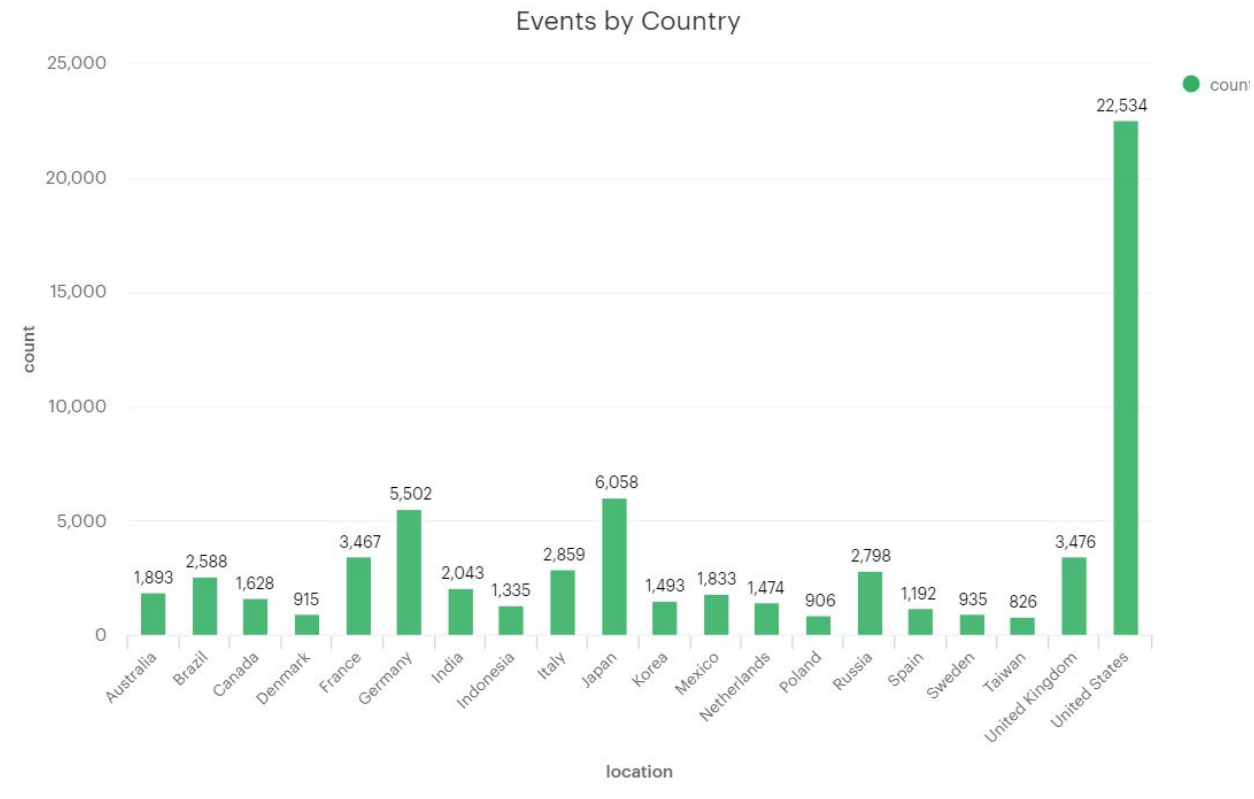

Event Type Counts in August
count
25,000
20,000
15,000
10,000
5,000
0
count
1,031
1,990
1,196
1,079
21,603
13,332
9,271
4,469
274
103
296
235
257
182
170
153
135
142
2,836
7,324
13,011
complete_signup
create_user
enter_email
enter_info
home_page
like_message
login
search_autocomplete
search_click_result_1
search_click_result_10
search_click_result_2
search_click_result_3
search_click_result_4
search_click_result_5
search_click_result_6
search_click_result_7
search_click_result_8
search_click_result_9
search_run
send_message
view_inbox
event_name

Location of Users

Below are two charts: one of locations of events for July and another for August. We see that the United States is by far the largest user base. While a hypothesis is that users in the United States may be on vacation during the last month of July, we see lower engagements across all locations. It seems more likely that there was a system issue.

```
SELECT COUNT(*), location FROM(
SELECT location,  to_date(cast(occurred_at as TEXT),'YYYY-MM-DD')   FROM tutorial.yammer_events)
as test
Where CAST(to_date AS text) LIKE '2014-07%'
GROUP BY location
ORDER BY count DESC
LIMIT 20
```


Events by Country
count
30,000
25,000
20,000
15,000
10,000
5,000
0
count
1,844
3,231
2,622
5,286
7,295
3,008
1,683
3,738
7,455
2,044
2,562
1,238
3,120
1,138
1,125
1,073
1,368
963
4,129
27,280
Australia
Brazil
Canada
France
Germany
India
Indonesia
Italy
Japan
Korea
Mexico
Netherlands
Russia
Saudi Arabia
Spain
Sweden
Switzerland
Taiwan
United Kingdom
United States
location


Events by Country
count
25,000
20,000
15,000
10,000
5,000
0
count
1,893
2,588
1,628
915
3,467
5,502
2,043
1,335
2,859
6,058
1,493
1,833
1,474
906
2,798
1,192
935
826
3,476
22,534
Australia
Brazil
Canada
Denmark
France
Germany
India
Indonesia
Italy
Japan
Korea
Mexico
Netherlands
Poland
Russia
Spain
Sweden
Taiwan
United Kingdom
United States
location

Emails!

Now we take a look at the email data. We finally found an interesting dip! On 8/04/2014, we see a drastic decrease in email click throughs. This tells us that maybe an email sent out on August 04 did not perform as well.

```
SELECT DATE_TRUNC('week', occurred_at) AS week_date, action, count(user_id) as user_actions
FROM   tutorial.yammer_emails emails
WHERE  action   = 'email_clickthrough'
GROUP  BY DATE_TRUNC('week',          occurred_at), action
ORDER  BY DATE_TRUNC('week',          occurred_at) ASC
```

Conclusion

Do the answers to any of your original hypotheses lead you to further questions?
None of my original hypotheses led me to any conclusions. I thought that somehow there had been less users overall, or some system error that prevented users from engaging, but these proved too simple guesses. After I explored the data, I had more questions about the other ways users engage with the site.

If so, what are they and how will you test them?
I wanted to know more about the email data, which I had not thought to originally explore. I tested this out by plotting to see various actions taken by users in the email table. I found a pattern in email clickthrough.

If they are questions that you can't answer using data alone, how would you go about answering them (hypothetically, assuming you actually worked at this company)?
I want to know what kind of email/s was/were sent out during the first week of August and how the marketing team thinks they preformed.

What seems like the most likely cause of the engagement dip?
A broken link in an email preventing people to clicking through.

What, if anything, should the company do in response?
Test all emails before release, paying close attention to ensuring all links are sound.