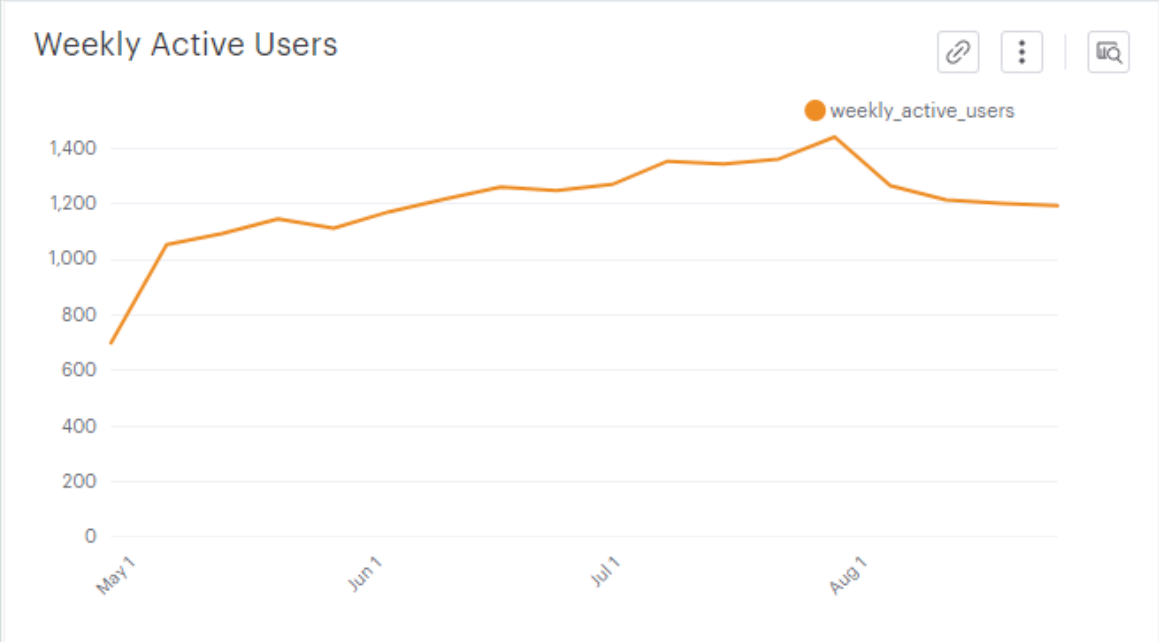
YAMMER ANALYSIS

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.

Data:

The following four tables are available for this user study.

(a) Users: This table contains descriptive information about the user's account.
(b) Events: This table includes one event where each event is an action that a user has taken on Yammer. These events include login events, messaging events, search events, events logged as users progress through a signup funnel, events around received emails.
(c) Email Events: This table contains events specific to the sending of the emails. It is similar in structure to the events table above.
(d) Rollup Periods: This table contains period, time, start and end time with respect to PST.

Possible Reasons:

Holiday: Assuming that Yammer is being used worldwide, holiday can play a major role in sudden ups and downs in user activity. Since different countries have different holiday seasons, it possible that one country might have less user engagement than others. If one country has much lower engagement than others, it's possible that this is the cause.

Technical Flaws: Technical difficulties are one of the major reason in a sudden drop of the user engagement. A technical flaw can be either a UI issue, integration or a back-end issue. Since we do not have enough information to identify the specificity of the technical flaw, it is hard to pinpoint our hypothesis and estimate its metrics.

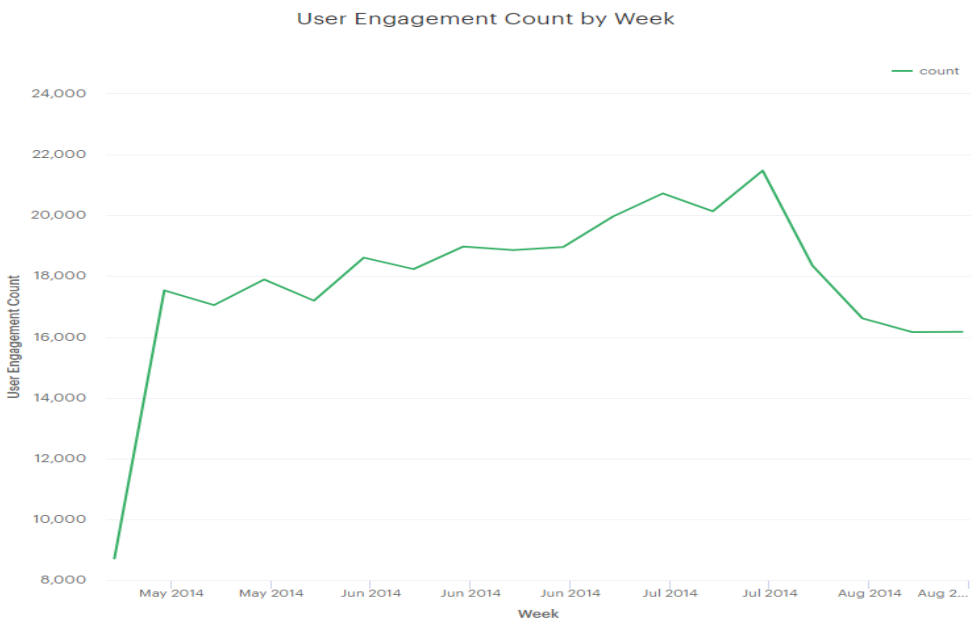
Site Traffic: It is also possible that the site traffic might have been extravagantly high due to which the user activity has suddenly dripped.

Bad Recent Update: Lack of delivery of app's fundamental purpose after the recent update can be a possible reason for the drastic change in user engagement.

Analysis:

More than 200 users have become inactive in the first week August. Since this is weekly analysis, quarter/yearly user counts can provide us with actual numbers of user engagement.

Just to make sure the graph provided in the problem statement is not the result of human error, we can take a look at the user engagement for the given time period. Since number of active users and user activity (events) go hand in hand, a drop in active numbers should also see a drop in user events.

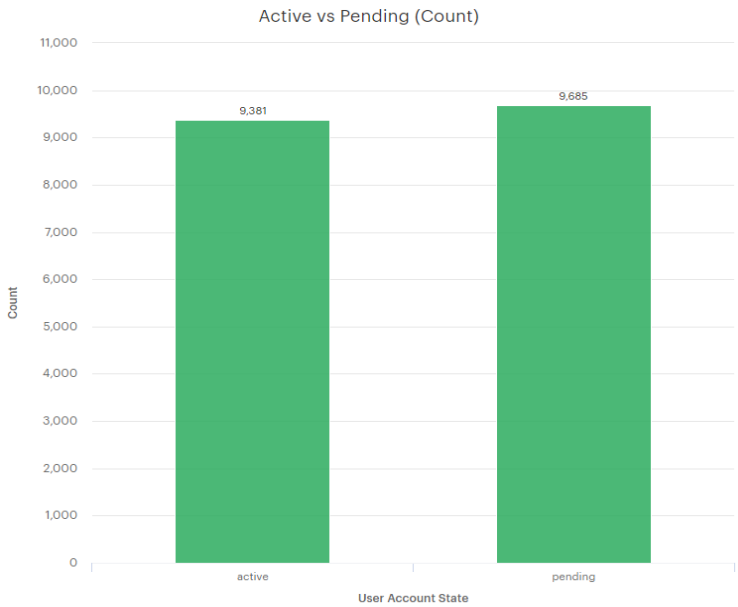


Graph 1: Graph showing the count of user engagement per week.

From the graph, it is evident that indeed the number of users has dropped in the first week of August.

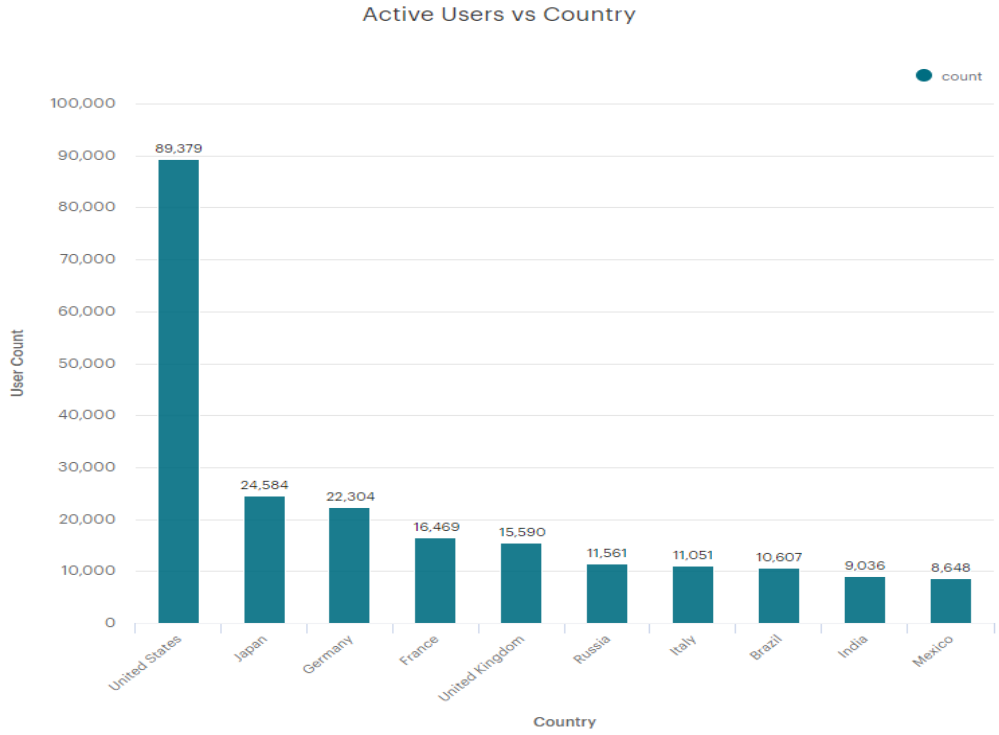
Hypothesis #1: Calendar Holidays have an effect on user engagement.

From the users table, we see that there are 9,381 active users and 9,685 pending. Both categories have almost equal numbers.



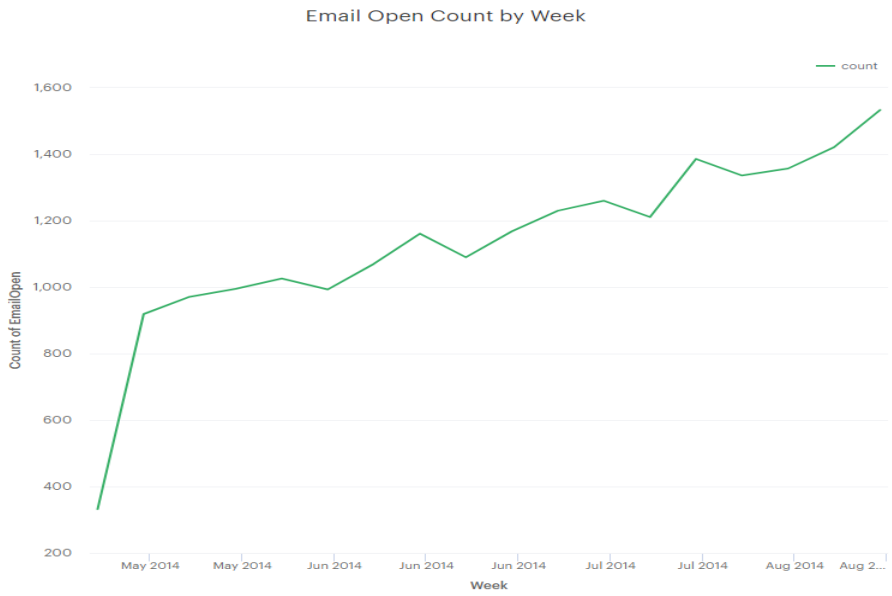
Graph 2: Graph showing active user count vs pending user count

Since Holidays varies at country-level, analyzing user location should give us better insights. It is evident that Unites States contributes about 40% of the users in the top 10 countries which use Yammer. Also, August happens to be Summer break month in United States which could possibly explain the sudden dip in the user engagement.



Graph 3: Graph 3 showing User counts per Country

For this time period, we can analyze user activity to either prove or disprove our hypothesis.

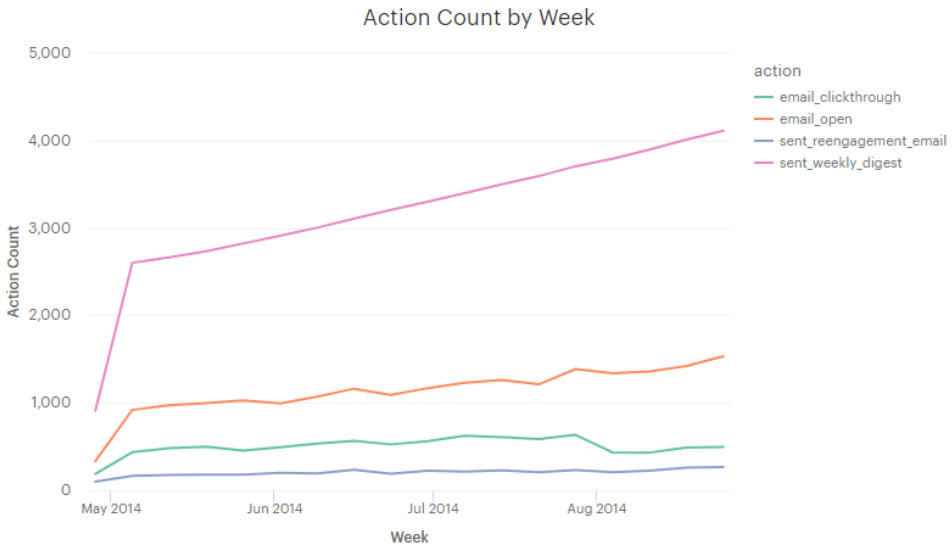


Graph 4: Graph showing count ‘email_open’ action per week

Though the number of users has dropped, it looks like the user engagement has not dropped which leads us to disprove our hypothesis.

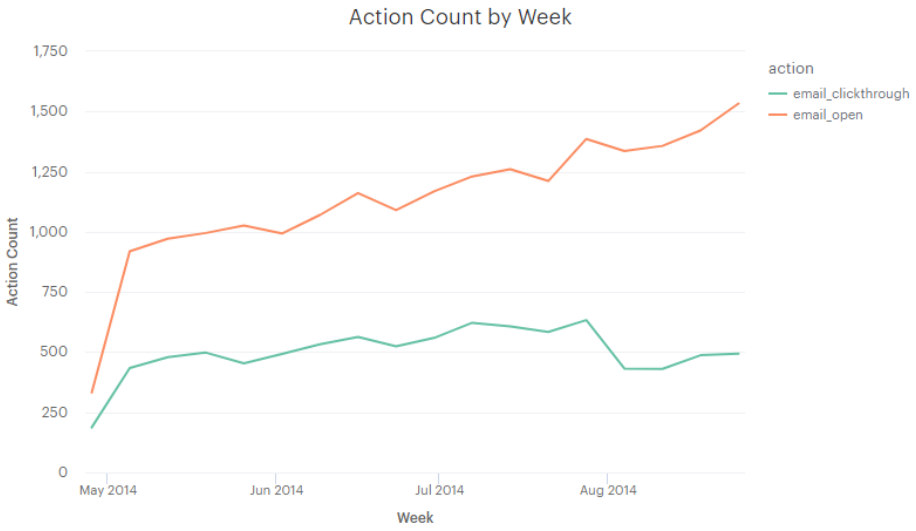
Hypothesis #2: A serious technical flaw has an effect on user engagement.

Looking at the action count by week, it is evident that 'email' specific events have consistently increased over time.



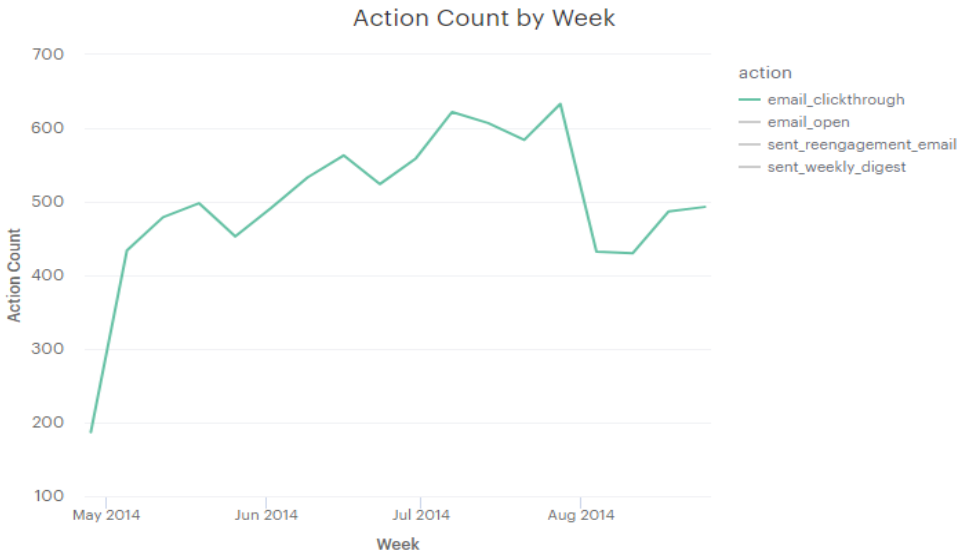
Graph 5: Graph showing count all user actions per week

Comparing four different classes of actions (email_clickthrough, email_open, sent_reengagement, sent_weekly_digest) against each other provides a closer look at the data.



Graph 6: Graph showing count 'email_clickthrough' and 'email_open' action per week

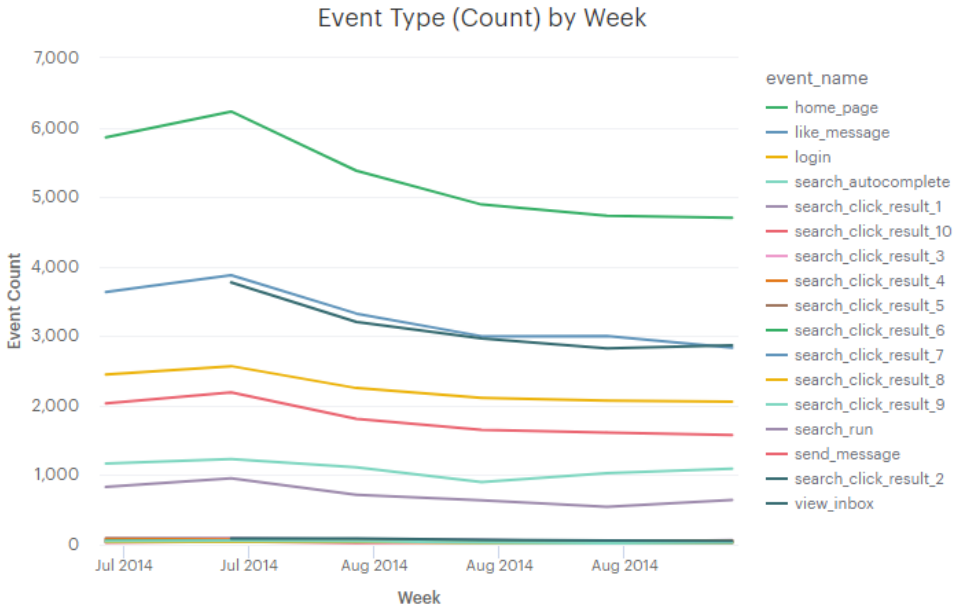
Though email_open has constantly been on rise, email_clickthrough has seen a gradual dip since August.



Graph 7: Graph showing count actions event per week

A close look at 'email_clickthrough' reveals that indeed there has been a sharp decline for first week of August.

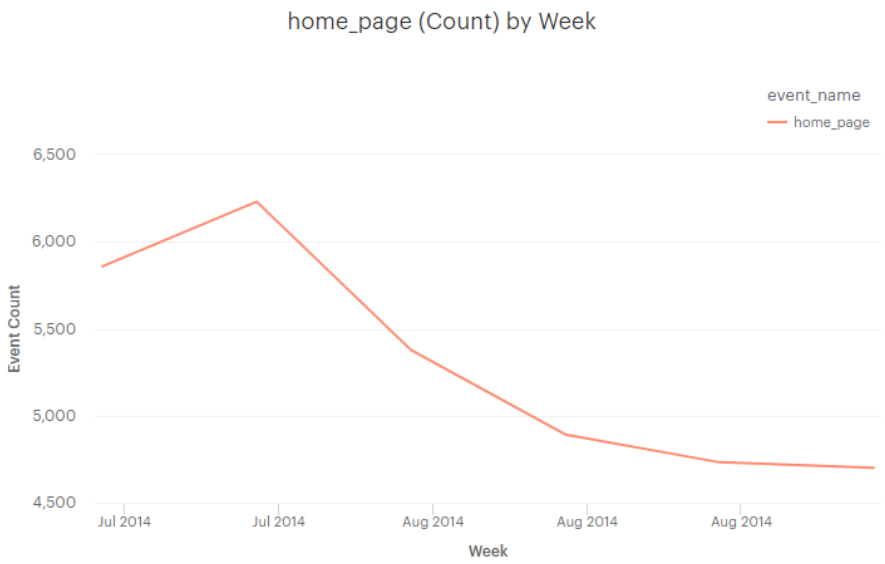
Analyzing other events could help us find any interesting insight specific to a particular user action.



Graph 8: Graph showing count events per week

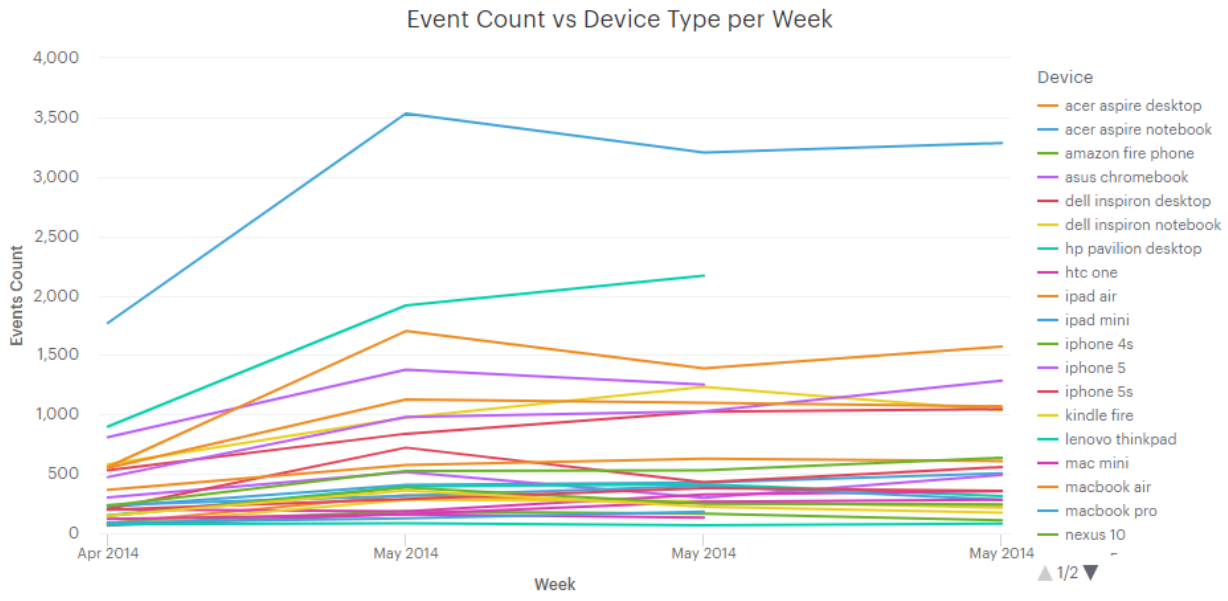
Looks like all the events have an almost similar behavior since late July.

However, a magnified look at each event reveals that ‘home_page’ event exhibits the same behavior as click through.



Graph 9: Graph showing count of ‘home_page’ event per week

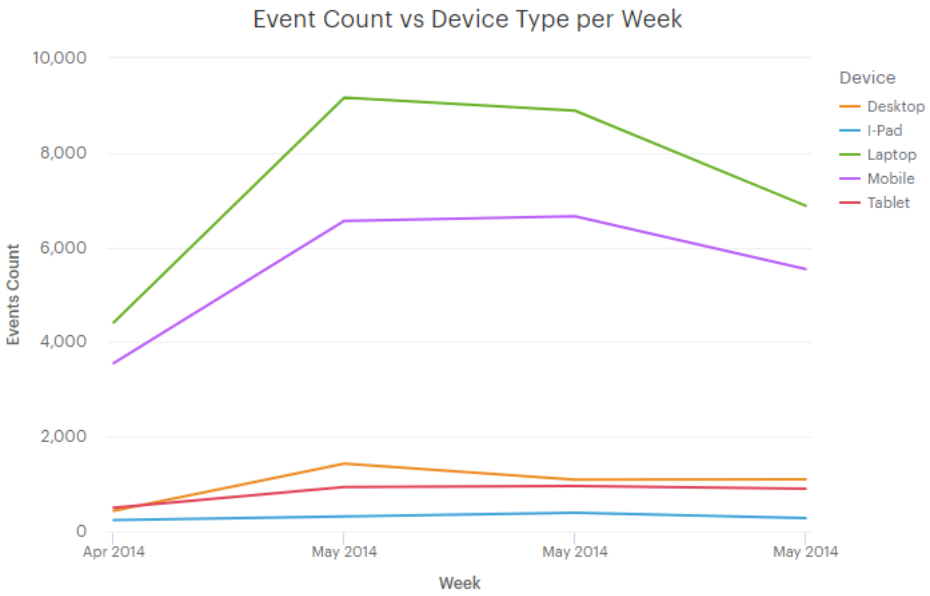
Hypothesis 3: A recent app update specific to a particular device might be correlated to the drop, in user engagement. Analyzing at the device level might give us a picture whether the user engagement drop is directly related to device type.



Graph 10: Graph showing event count by device per week

From the above graph it is evident that mobiles and laptops indeed have shown a negative rise. For a better view, we can take this a step further by categorizing like-devices (using case statements) into one category.

Below is the resultant graph for the event count by device type.



From the above graph, we can conclude that laptop and mobile devices have shown a huge drop in the user engagement.

SQL FOR YAMMER CASE STUDY

HYPOTHESIS I

```
-- Number of users
select * from tutorial.yammer_users;
select count(user_id) from tutorial.yammer_users;
select distinct event_type from tutorial.yammer_events;

--Time Series analysis of the events (user_engagement)
select date_trunc('week',occurred_at), count(event_name) from tutorial.yammer_events
where event_type = 'engagement'
group by 1 order by 1;

--active and pending
select distinct state from tutorial.yammer_users;
-- Count of different type of 'state'.
select state, count(user_id) from tutorial.yammer_users group by 1;

-- 47 different lcoations
select distinct location from tutorial.yammer_events;

-- User count at lcoation level
select a.location, count(b.user_id)
from tutorial.yammer_events a JOIN tutorial.yammer_users b on a.user_id = b.user_id
where state = 'active' and event_type = 'engagement'
group by 1 order by 2 desc limit 10;
```

```
-- 4 different actions
select distinct action from tutorial.yammer_emails;

--Time series for email open
SELECT DATE_TRUNC('week', occurred_at), count(user_id)
FROM tutorial.yammer_emails WHERE action = 'email_open'
GROUP BY 1 ORDER BY 1;

select * from tutorial.yammer_events ;

-- Count of active users
select date_trunc('day',occured_at), count(*),
count(case when activated_at is not null then u.user_id else null end)
from tutorial.yammer_users u
where created_at >= '2014-06-01' and created_at < '2014-09-01'
group by 1 order by 1
```

HYPOTHESIS II

```
-- Action ‘email_clickthrough’ analysis
select date_trunc('week', occurred_at), action, count(user_id)
from tutorial.yammer_emails emails where action = 'email_clickthrough'
group by 1, 2 order by 1 asc;

-- User engagement analysis
select date_trunc('week', occurred_at), event_name, count(user_id)
from tutorial.yammer_events events
where event_type = 'engagement'
group by 1, 2 order by 1 desc;
```

HYPOTHESIS III

```
-- Device analysis
select date_trunc('week', occurred_at), device, count(user_id)
from tutorial.yammer_events events
group by 1, 2 order by 1 asc;

-- Categorizing devices into device types.
select DATE_TRUNC('week', occurred_at), device,
(CASE WHEN device =    'acer aspire desktop'      THEN   'Desktop'
 WHEN device =         'dell inspiron desktop'    THEN   'Desktop'
 WHEN device =         'hp pavilion desktop'      THEN   'Desktop'
 WHEN device =         'ipad mini'          THEN   'I-Pad'
 WHEN device =         'macbook pro'    THEN   'Laptop'
 WHEN device =         'acer aspire notebook'    THEN   'Laptop'
 WHEN device =         'lenovo thinkpad'          THEN   'Laptop'
 WHEN device =         'dell inspiron notebook'  THEN   'Laptop'
 WHEN device =         'windows surface'          THEN   'Laptop'
 WHEN device =         'macbook air'     THEN   'Laptop'
 WHEN device =         'asus chromebook'          THEN   'Laptop'
 WHEN device =         'iphone 5s'          THEN   'Mobile'
 WHEN device =         'samsung galaxy note'    THEN   'Mobile'
 WHEN device =         'mac mini'          THEN   'Mobile'
 WHEN device =         'nokia lumia 635'          THEN   'Mobile'
 WHEN device =         'nexus 7'          THEN   'Mobile'
 WHEN device =         'nexus 10'          THEN   'Mobile'
 WHEN device =         'amazon fire phone'       THEN   'Mobile'
 WHEN device =         'nexus 5'          THEN   'Mobile'
 WHEN device =         'iphone 4s'          THEN   'Mobile'
 WHEN device =         'htc one'          THEN   'Mobile'
```

```
WHEN device =          'iphone 5'          THEN   'Mobile'
WHEN device =          'samsung galaxy s4'       THEN   'Mobile'
WHEN device =          'kindle fire'       THEN   'Tablet'
WHEN device =          'samsumg galaxy tablet'  THEN   'Tablet'
WHEN device =          'ipad air'          THEN   'Tablet'
else null END),
count(user_id)
from tutorial.yammer_events events
group by 1, 2 order by 1 asc;
```