

Investigating a Drop in User Engagement

Analytics cases: Yammer

Yammer is a social network for communicating with coworkers. Individuals share documents, updates, and ideas by posting them in groups. Yammer is free to use indefinitely, but companies must pay license fees if they want access to administrative controls, including integration with user management systems like ActiveDirectory.

Yammer has a centralized Analytics team, which sits in the Engineering organization. Their primary goal is to drive better product and business decisions using data. They do this partially by providing tools and education that make other teams within Yammer more effective at using data to make better decisions. They also perform ad-hoc analysis to support specific decisions.

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:

Weekly Active Users



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 Analysis Plan

Here are some of the possible causes for the dip in retention shown in the chart above.

Possible dip in user weekly engagement:

I'll address the issues in this order for ease, and stop when any significant issue has been identified.

1. User growth has diminished, and so user engagement as decreased
 - a. Weekly/Daily user growth
2. Users have changed service became inactive
 - a. Weekly/Daily user growth
3. User ID - Country specific issues
 - a. Weekly engagement (login) by region
 - i. If any issues are found region specific move into sub-region, countries
4. User ID - Device type issues
 - a. Weekly engagement (login) by device type
 - i. Desktop, tablet, cell phone
 1. Domain knowledge based on device
 - b. Manufacturer type
 - i. Apple, samsung, htc
 1. Split by domain knowledge/research based on device
5. User ID - Email action
 - a. Email action
 - i. Was user sent weekly digest
 - ii. Did user open email digest
 - iii. Did user interact with email - email_clickthrough
6. Engagement specific event issues
 - a. Liking events
 - b. Sending messages
 - c. Viewing inbox
 - d. Loading homepage
 - e. Searching
 - i. Autocompletes
 - ii. Run
 - iii. Click results

The engagement specific events will be tested last. The engagement dip is from login event level engagements. These events were included for full-expansion of possible reasons. The most likely cause will be from growth, device specific, region specific, or login portal (email events)

Answering Login Engagement decrease

1. User growth has diminished, and so user engagement as decreased
 - a. Weekly/Daily user growth

Users activated compared to all users



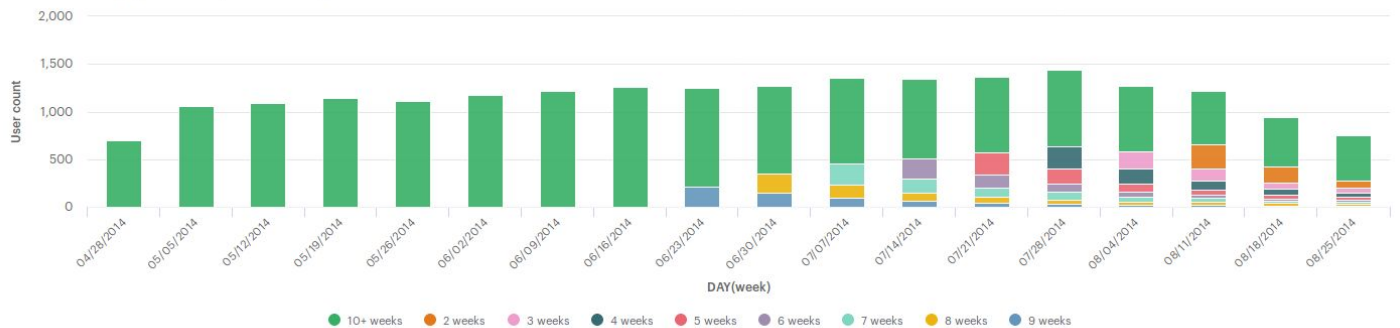
User growth trends



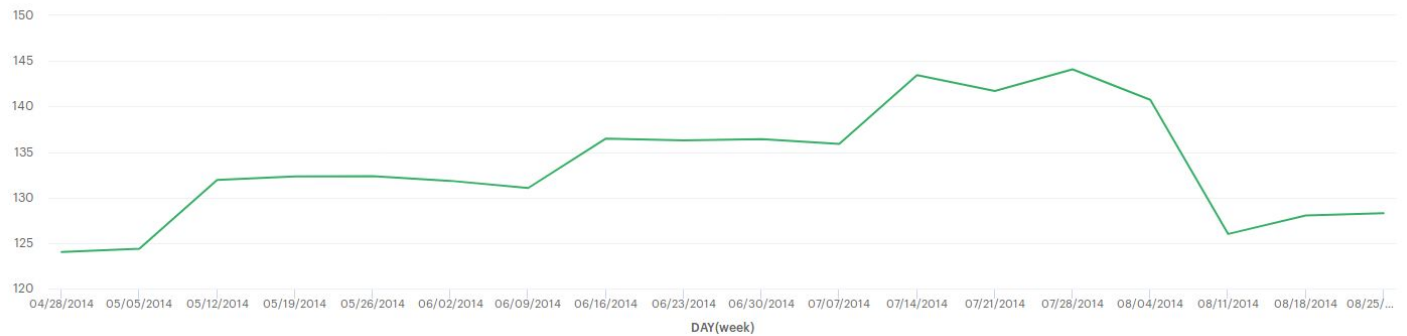
We can see from the growth trends that the weekly patterns of user and activated users remains the same. There is high activity during the week, and low activity during the week. Averaging it out over the week shows that there was a dip when we first notice the change in user engagement, but afterward the trendline continues upward. This is different than our login engagement trend line. We can say the dip isn't related to new customers but to our old user base.

2. User ID - User account age

Total user engagement by account age



Average age of accounts by week



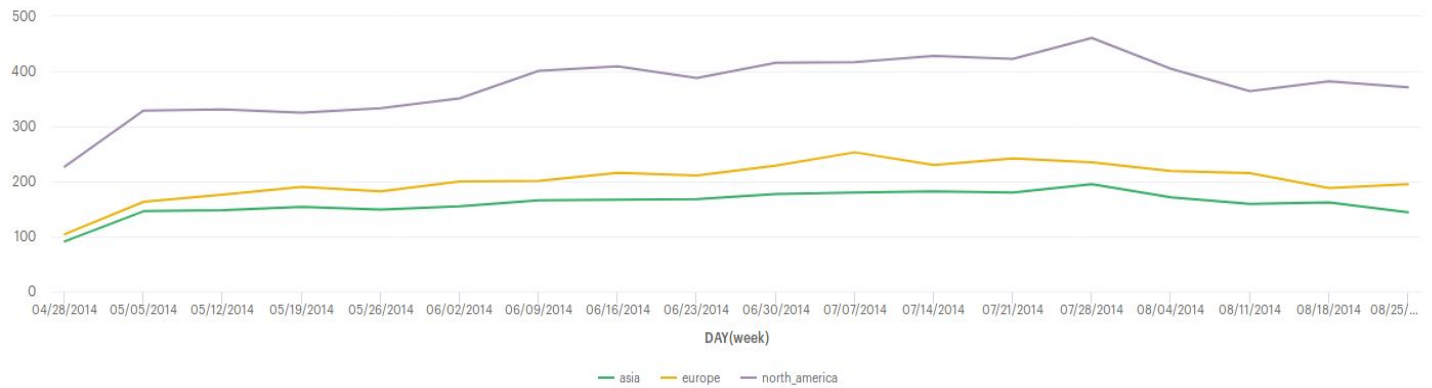
The account age piggy backs our user growth investigation. Among the existing users, it appears the older accounts are most affected by the drop in total user logins. The average age our accounts logging in takes a sharp dive after the 7/28 marker. This indicates the problem isn't with newer accounts having issues accessing the site.

An aside to marketing: looking through how the new accounts are processed and aged, there is an issue with new account churn. Each consecutive block segment from 6 weeks to 7 weeks, or 8 weeks to 9 weeks there is marked difference in size (logins for accounts that age). The blocks appear to shrink by at least 30%.

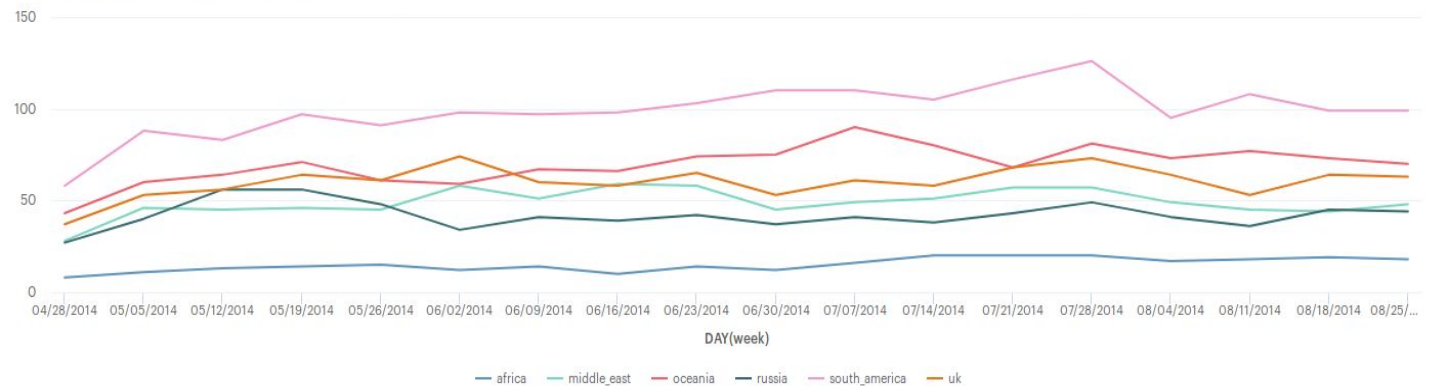
3. User ID - Country specific issues

a. Weekly engagement (login) by region

Weekly active users by top 3 regions



Weekly active users other regions

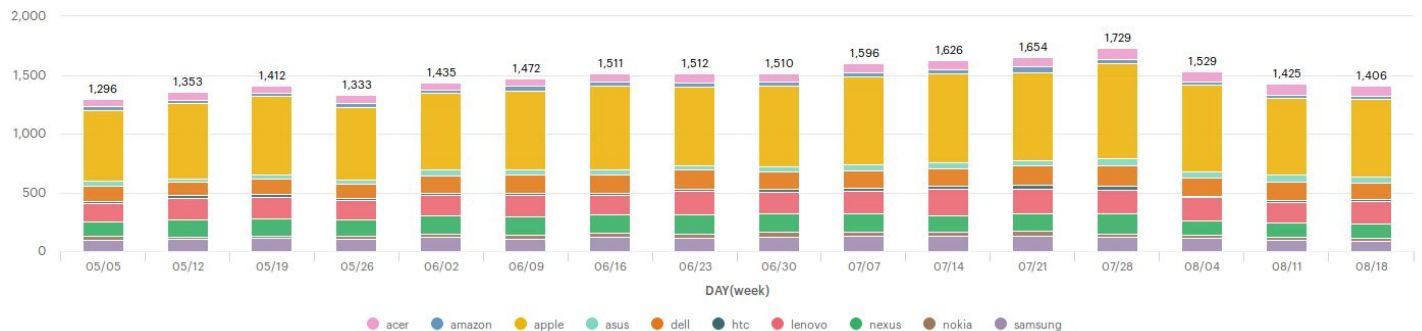


Looking at the week of 7/28 we can notice that there is a across the board decrease in engagement in all regions. This can help show that there isn't a specific region that we're seeing the decrease. We won't need to look deeper at independent countries to look for issues or reasons for the engagement drop. It appears to be agnostic of country/region.

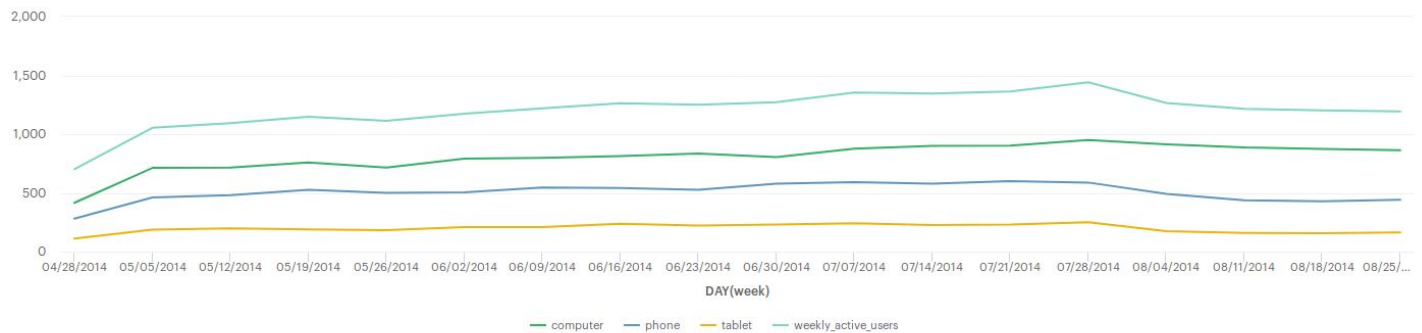
Europe appears to be the least affected, but they were already trending downward prior. This is a possible recommendation to the marketing team as an aside to this problem.

4. User ID - Device type issues
 - a. Weekly engagement (login) by device type
 - i. Desktop, tablet, cell phone
 - b. Manufacturer type
 - i. Apple, samsung, htc

User login engagement by device manufacture



User logins by device type



Looking at the logins by manufacturer type, you can see that there were specific manufacturers that were hit harder than others: Apple, Samsung, HTC. Most manufacturers were still hit by the lower engagement. As an aside for reporting and data collection, operating system for each device is a feature that could be added to show if there is a specific problem with, e.g. Android, or iOS. Looking at specific device type: Computer, Phone, Tablet, we can see that cell phone was hit the hardest, followed by tablet. This could be part of the problem, or the problem we're seeing with the engagement is through mobile application access.

5. User ID - Email action

a. Email action

- i. Was user sent weekly digest
- ii. Did user open email digest
- iii. Did user interact with email - email_clickthrough



2

Looking at a significant login portal, email interaction, shows a major problem that is existing. The weekly digest shows that it's being sent out. Our digest open indicator trendline is trending positive with our weekly digest. The issue appears in the email click-through. At 7/28 there is a noticeable downtrend in click-throughs. This has a high probability of being culprit to the downturn in engagement.

Summary:

After looking through the data I would report back that there is something wrong with the digest emails (i.e. possibly bad links). Also that the mobile (Cellphone, tablet) application(s) may have an issue going on with them too.

² <https://modeanalytics.com/modeanalytics/reports/d8c4d3c11ea3/runs/43cdd26cce26>

Description of data sets and features

The tables names and column definitions are listed below

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

This table includes one row per user, with descriptive information about that user's account.

This table name in Mode is tutorial.yammer_users

user_id:	A unique ID per user. Can be joined to user_id in either of the other tables.
created_at:	The time the user was created (first signed up)
state:	The state of the user (active or pending)
activated_at:	The time the user was activated, if they are active
company_id:	The ID of the user's company
language:	The chosen language of the user

Table 2: Events

This table includes one row per event, where an 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.

This table name in Mode is tutorial.yammer_events

user_id:	The ID of the user logging the event. Can be joined to user_id in either of the other tables.
occurred_at:	The time the event occurred.
event_type:	The general event type. There are two values in this dataset: "signup_flow", which refers to anything occurring during the process of a user's authentication, and "engagement", which refers to general product usage after the user has signed up for the first time.
event_name:	<p>The specific action the user took. Possible values include:</p> <ul style="list-style-type: none">create_user: User is added to Yammer's database during signup processenter_email: User begins the signup process by entering her email addressenter_info: User enters her name and personal information during signup processcomplete_signup: User completes the entire signup/authentication processhome_page: User loads the home pagelike_message: User likes another user's messagelogin: User logs into Yammersearch_autocomplete: User selects a search result from the autocomplete listsearch_run: User runs a search query and is taken to the search results pagesearch_click_result_X: User clicks search result X on the results page, where X is a number from 1 through 10.send_message: User posts a messageview_inbox: User views messages in her inbox
location:	The country from which the event was logged (collected through IP address).
device:	The type of device used to log the event.

Table 3: Email Events

This table contains events specific to the sending of emails. It is similar in structure to the events table above.

This table name in Mode is tutorial.yammer_emails

user_id:	The ID of the user to whom the event relates. Can be joined to user_id in either of the other tables.
occurred_at:	The time the event occurred.
action:	The name of the event that occurred. "sent_weekly_digest" means that the user was delivered a digest email showing relevant conversations from the previous day. "email_open" means that the user opened the email. "email_clickthrough" means that the user clicked a link in the email.

Table 4: Rollup Periods

The final table is a lookup table that is used to create rolling time periods. Though you could use the INTERVAL() function, creating rolling time periods is often easiest with a table like this. You won't necessarily need to use this table in queries that you write, but the column descriptions are provided here so that you can understand the query that creates the chart shown above.

This table name in Mode is benn.dimension_rollup_periods

period_id:	This identifies the type of rollup period. The above dashboard uses period 1007, which is rolling 7-day periods.
time_id:	This is the identifier for any given data point — it's what you would put on a chart axis. If time_id is 2014-08-01, that means that it represents the rolling 7-day period leading up to 2014-08-01.
pst_start:	The start time of the period in PST. For 2014-08-01, you'll notice that this is 2014-07-25 — one week prior. Use this to join events to the table.
pst_end:	The end time of the period in PST. For 2014-08-01, the end time is 2014-08-01. You can see how this is used in conjunction with pst_start to join events to this table in the query that produces the above chart .
utc_start:	The same as pst_start, but in UTC time.
utc_end:	The same as pst_end, but in UTC time.