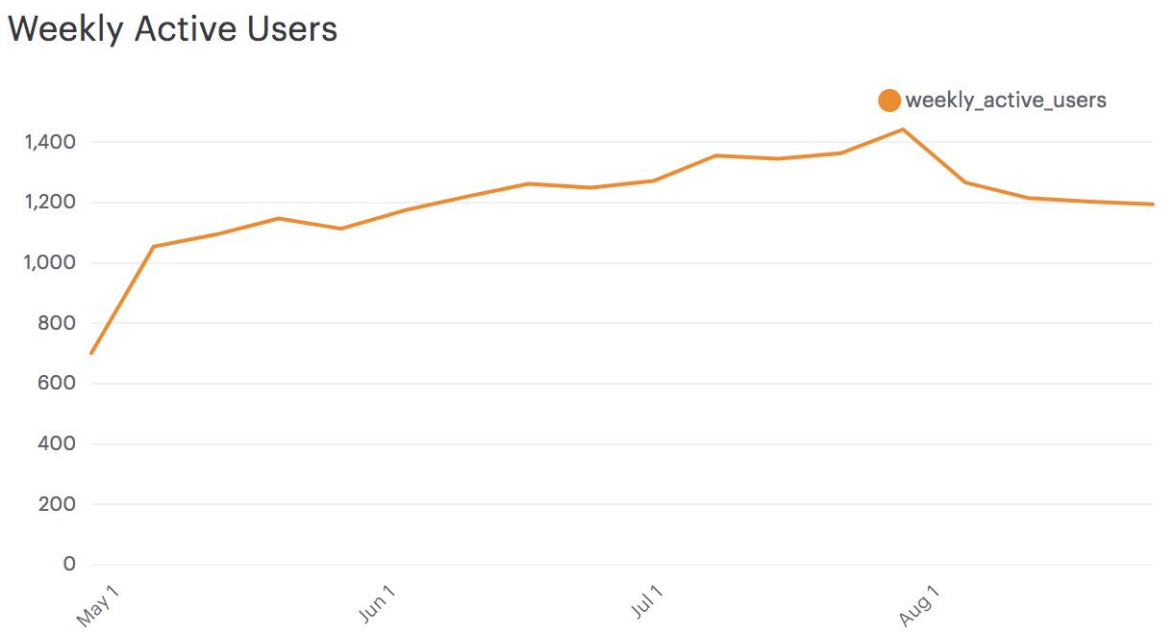
Yammer

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. You sit in their centralized Analytics team, which sits in the Engineering organization.

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 Problem

What caused the drop in user engagement? Based on what you observe in the data, what recommendations would you make to the organization?

The Data

Table 1: Users

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

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.

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	The specific action the user took. Possible values include: • **create_user:** User is added to Yammer's database during signup process • **enter_email:** User begins the signup process by entering her email address • **enter_info:** User enters her name and personal information during signup process • **complete_signup:** User completes the entire signup/authentication process • **home_page:** User loads the home page • **like_message:** User likes another user's message • **login:** User logs into Yammer • **search_autocomplete:** User selects a search result from the autocomplete list

	• **search_run:** User runs a search query and is taken to the search results page • **search_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 message • **view_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.

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. You won't necessarily need to use this table, but the column descriptions are provided here nonetheless.

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 is 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.
utc_start	The same as pst_start, but in UTC time.
pst_start	The same as pst_end, but in UTC time.