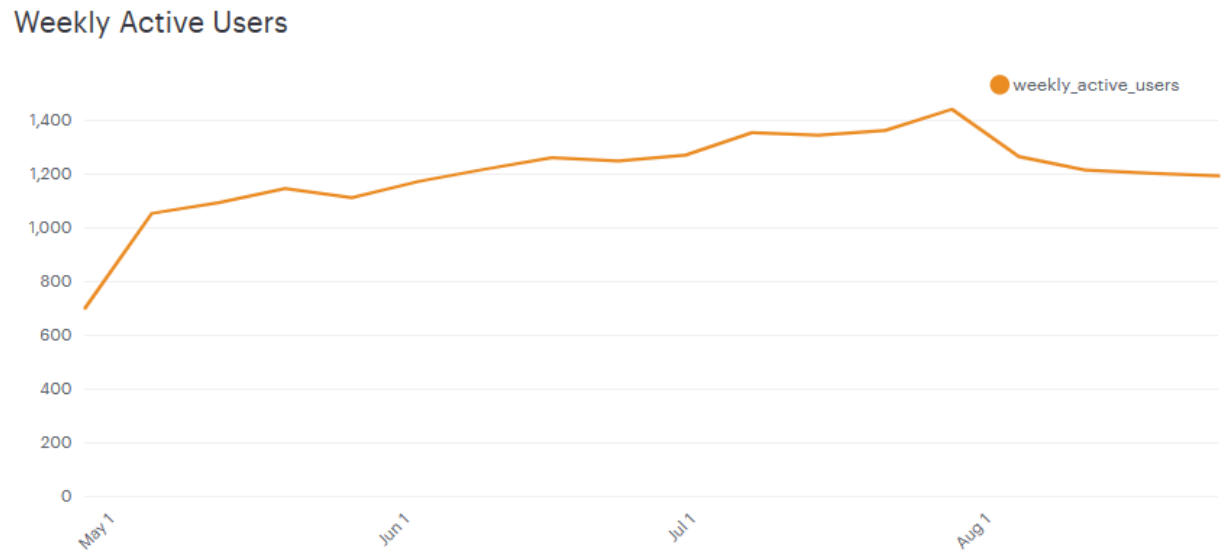
Project Description: [Investigating a Drop in User Engagement | SQL Analytics Training - Mode]
This project was completed using fictional datasets for an app named Yammer. In August 2014, it suffered a 200 person drop in monthly active users. What happened to user engagement, and what solutions might mitigate further drops?



To be clear, engagement is defined as “the number of users who logged at least one engagement event during the week starting on that date.” Such events include logins, visiting Yammer’s homepage, viewing your inbox, sending/liking messages, and running search queries.

Hypothesis 1: User engagement declined in specific categories.

To investigate this hypothesis, I crafted a SQL query to ascertain if there was a notable decrease in engagement within specific event categories. While activities such as logging in are expected to consistently attract high user interaction, other actions, like executing search queries, may not sustain the same level of popularity.

```
SELECT EXTRACT ('month' FROM occurred_at) AS month,
  COUNT (CASE WHEN event_name  = 'login' THEN 1 END) AS login,
  COUNT (CASE WHEN event_name  = 'home_page' THEN 1 END) AS home_page,
  COUNT (CASE WHEN event_name  = 'view_inbox' THEN 1 END) AS view_inbox,
  COUNT (CASE WHEN event_name  = 'send_message' THEN 1 END) AS
send_message,
  COUNT (CASE WHEN event_name  = 'like_message' THEN 1 END) AS
like_message,
  COUNT (CASE WHEN event_name  = 'search_autocomplete' THEN 1 END) AS
search_autocomplete,
```

```
  COUNT (CASE WHEN event_name  = 'search_run' THEN 1 END) AS search_run,
  COUNT (CASE WHEN event_name ilike 'search_click_result_%' THEN 1 END) AS
search_click_result_X,
  COUNT (event_name ) AS total_events
  FROM tutorial.yammer_events
  GROUP BY 1
```

To address the limitations of space, the results have been visualized in an image named "**output1**." Analysis reveals a decrease in user engagement across all categories, suggesting that the decline wasn't due to a diminished interest in specific features of Yammer's service. Rather, it appears that the platform is being utilized for its intended purposes just as before (see "**output2**"). To provide further clarity, the following SQL query converts the previous event counts into percentages. This enables a comparison that highlights consistent user interaction with the service's features, mirroring engagement percentage levels observed in July. The only difference between July and August is that fewer people are using Yammer, but why?

```
SELECT sub.month,
  100 * login/total_engagement_events AS percent_login,
  100 * home_page/total_engagement_events AS percent_home_page,
  100 * view_inbox/total_engagement_events AS percent_view_inbox,
  100 * send_message/total_engagement_events AS percent_send_message,
  100 * like_message/total_engagement_events AS percent_like_message,
  100 * search_autocomplete/total_engagement_events AS
percent_search_autocomplete,
  100 * search_run/total_engagement_events AS percent_search_run,
  100 * search_click_result_X/total_engagement_events AS
percent_search_click_result_X
  FROM (
    SELECT EXTRACT ('month' FROM occurred_at) AS month,
    COUNT (CASE WHEN event_name  = 'login' THEN 1 END) AS login,
    COUNT (CASE WHEN event_name  = 'home_page' THEN 1 END) AS home_page,
    COUNT (CASE WHEN event_name  = 'view_inbox' THEN 1 END) AS view_inbox,
    COUNT (CASE WHEN event_name  = 'send_message' THEN 1 END) AS
send_message,
    COUNT (CASE WHEN event_name  = 'like_message' THEN 1 END) AS
like_message,
    COUNT (CASE WHEN event_name  = 'search_autocomplete' THEN 1 END) AS
search_autocomplete,
    COUNT (CASE WHEN event_name  = 'search_run' THEN 1 END) AS search_run,
    COUNT (CASE WHEN event_name  ilike 'search_click_result_%' THEN 1 END)
AS search_click_result_X,
    COUNT (CASE WHEN event_type = 'engagement' THEN 1 END) AS
```

```
total_engagement_events
    FROM tutorial.yammer_events
    GROUP BY 1
  ) sub
```

Hypothesis 2: Yammer has a problem retaining users.

The Yammer datasets only record user activity on the platform between May 2014 and August 2014. If a user created an account in May, and Yammer logged a user's activity in June, that means the individual was retained for a one month period. This query provides a table below showing that only 20% of employees are retained from one month to the next.

```
SELECT EXTRACT ('month' FROM users.created_at) AS month,
  COUNT (DISTINCT users.user_id) AS total_users_created,
  COUNT (DISTINCT e.user_id) AS total_users_retained_next_month,
  COUNT(DISTINCT e.user_id) * 100.0 / COUNT(DISTINCT users.user_id) AS
percent_active_next_month
  FROM tutorial.yammer_users users
  LEFT JOIN tutorial.yammer_events e
  ON e.user_id = users.user_id
  AND EXTRACT ('month' FROM e.occurred_at) = EXTRACT('month' FROM
users.created_at) + 1
  WHERE EXTRACT ('month' FROM users.created_at) BETWEEN 5 AND 8
  GROUP BY 1
```

month	total_users_created	total_users_retained_next_month	percent_active_next_month
5	2083	440	21.1234
6	2213	503	22.7293
7	2591	513	19.7993
8	2626	0	0

This query looks at retention over a two month period. Here, retention is worse. I should also note that some of these percentages are 0 because the dataset only covers four months.

```
SELECT EXTRACT ('month' FROM users.created_at) AS month,
  COUNT (DISTINCT users.user_id) AS total_users_created,
  COUNT (DISTINCT e.user_id) AS total_users_retained_next_month,
  COUNT(DISTINCT e.user_id) * 100.0 / COUNT(DISTINCT users.user_id) AS
percent_active_next_month
  FROM tutorial.yammer_users users
  LEFT JOIN tutorial.yammer_events e
```

```
  ON e.user_id = users.user_id
  AND EXTRACT ('month' FROM e.occurred_at) = EXTRACT('month' FROM
users.created_at) + 2
  WHERE EXTRACT ('month' FROM users.created_at) BETWEEN 5 AND 8
  GROUP BY 1
```

month	total_users_created	total_users_retained_next_month	percent_active_next_month
5	2083	247	11.8579
6	2213	197	8.9019
7	2591	0	0
8	2626	0	0

Although Yammer may have over two thousand new users every month, It's abundantly clear that Yammer has a problem retaining them.

Hypothesis 3: Yammer has a problem retaining users in certain parts of the globe.

```
SELECT sub.location,
  sub.july_events_logged,
  sub.august_events_logged
  FROM (
    SELECT location,
    COUNT (CASE WHEN EXTRACT ('month' FROM occurred_at) = 7 THEN 1 END) AS
july_events_logged,
    COUNT (CASE WHEN EXTRACT ('month' FROM occurred_at) = 8 THEN 1 END) AS
august_events_logged
    FROM tutorial.yammer_events
    GROUP BY 1
  ) sub
  WHERE sub.august_events_logged < sub.july_events_logged
ORDER BY sub.august_events_logged -  sub.july_events_logged
```

location	july_events_logged	august_events_logged
United States	27280	22534
France	5286	3467
Germany	7295	5502
Japan	7455	6058
Canada	2622	1628
India	3008	2043
Italy	3738	2859
Mexico	2562	1833
Switzerland	1368	711
United Kingdom	4129	3476
Brazil	3231	2588
Korea	2044	1493
Belgium	865	452
Indonesia	1683	1335
Saudi Arabia	1138	806
Russia	3120	2798
Austria	927	608
Venezuela	683	392
Singapore	628	340
Thailand	529	259
Norway	532	303
Chile	485	306
Greece	425	285
Sweden	1073	935
Taiwan	963	826
Colombia	622	541
Finland	520	455
Nigeria	469	408
Philippines	414	390

The biggest drop is in the United States, where it experienced 5000 fewer event engagements. We can also see the breakdowns along device usage.

```
SELECT sub.location,
  sub.device,
  sub.july_events_logged,
  sub.august_events_logged
  FROM (
    SELECT location,
    device,
    COUNT (CASE WHEN EXTRACT ('month' FROM occurred_at) = 7 THEN 1 END) AS
july_events_logged,
    COUNT (CASE WHEN EXTRACT ('month' FROM occurred_at) = 8 THEN 1 END) AS
august_events_logged
    FROM tutorial.yammer_events
    GROUP BY 1,2
  ) sub
  WHERE sub.august_events_logged < sub.july_events_logged
  ORDER BY sub.august_events_logged - sub.july_events_logged
```

location	device	july_events_logged	august_events_logged
United States	macbook pro	4861	3525
United States	samsung galaxy s4	1961	1180
United States	iphone 4s	908	475
United States	macbook air	2127	1791
United States	ipad air	870	558
United States	dell inspiron desktop	617	317
Japan	iphone 5	714	430
Japan	lenovo thinkpad	769	487
France	lenovo thinkpad	738	467
United Kingdom	iphone 5	482	214
Brazil	macbook air	426	181
France	dell inspiron desktop	307	67
United States	htc one	437	199
France	macbook air	379	152
United States	nexus 5	1109	884
Korea	iphone 5	386	162
Germany	lenovo thinkpad	935	711

Despite the decline in usage, this does not identify the problem I'm investigating because the drop is happening worldwide.

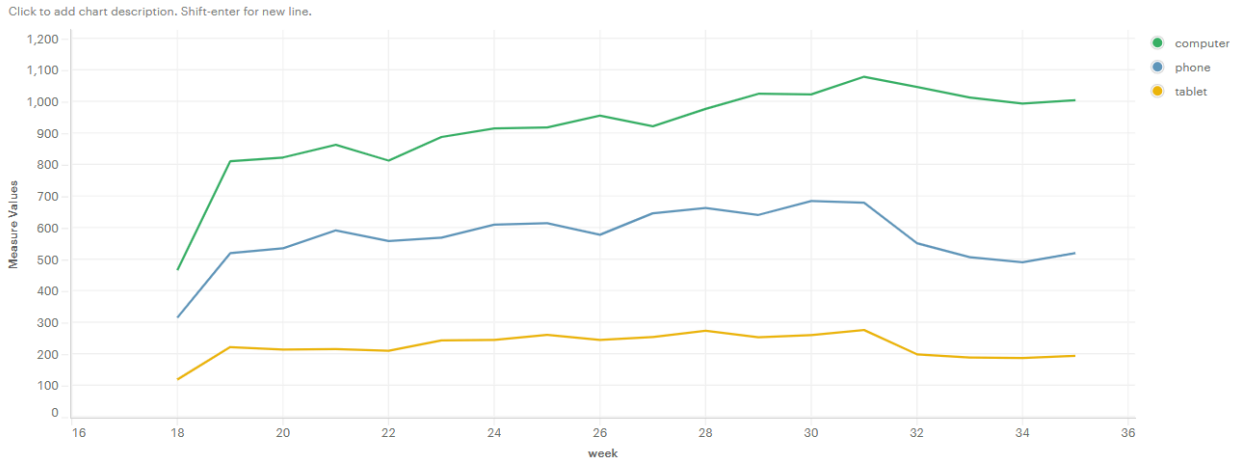
I would want to sort the devices by whether they are a phone, tablet, or laptop because, by doing this, you may catch a technical glitch.

It may also be easier to see a drop if we count the number of engagement events logged weekly, not monthly.

```
SELECT EXTRACT('week' FROM occurred_at) AS week,
  COUNT(DISTINCT CASE WHEN device IN ('macbook pro','lenovo
thinkpad','macbook air','dell inspiron notebook',
          'asus chromebook','dell inspiron desktop','acer aspire
notebook','hp pavilion desktop','acer aspire desktop','mac mini')
          THEN user_id ELSE NULL END) AS computer,
  COUNT(DISTINCT CASE WHEN device IN ('iphone 5','samsung galaxy s4','nexus
5','iphone 5s','iphone 4s','nokia lumia 635',
       'htc one','samsung galaxy note','amazon fire phone') THEN user_id
ELSE NULL END) AS phone,
  COUNT(DISTINCT CASE WHEN device IN ('ipad air','nexus 7','ipad
mini','nexus 10','kindle fire','windows surface',
        'samsumg galaxy tablet') THEN user_id ELSE NULL END) AS tablet
  FROM tutorial.yammer_events
  GROUP BY 1
```

	week	computer	phone	tablet
1	18	465	314	118
2	19	810	519	221
3	20	822	534	213
4	21	862	591	215
5	22	812	557	209
6	23	887	568	242
7	24	914	609	244
8	25	917	614	260
9	26	955	577	244
10	27	921	645	253
11	28	976	662	273
12	29	1024	640	252

This data is hard to read, so I turned it into a graph.



GREEN = computer
BLUE = phone
ORANGE = tablet

Since usage is declining across the board, I would expect there is a competitor taking some of the users away. Still, phone usage dropped dramatically. Either there is just a competitor taking users away, or there is an additional technical problem with Yammer's services on phones. It's not clear why that would be the case, but you would need to investigate that further with a software team.

Overall Conclusion: The inability to retain users is likely why Yammer saw a decline in event engagement after July. Given that the change is so sudden, especially in the United States, I would also expect there to be an alternative in the marketplace that is gaining attention - as well as a technical bug in the phones.