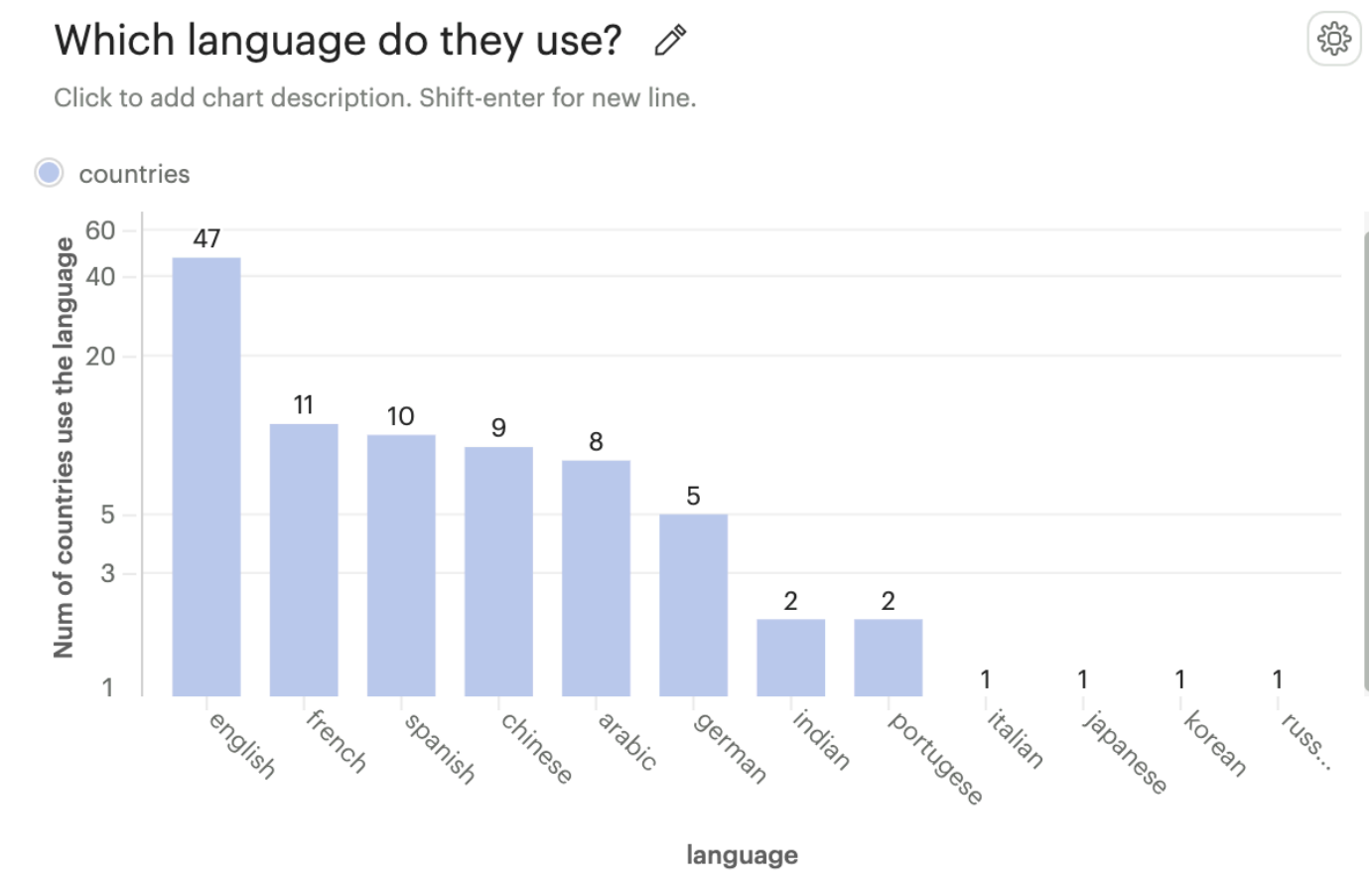
Yammer Analysis

Analyze Users

1. Which language do they use?

```
SELECT language,COUNT(DISTINCT(location)) AS countries
FROM tutorial.yammer_users AS users
LEFT JOIN tutorial.yammer_events AS events
USING(user_id)
GROUP BY language
ORDER BY countries DESC;
```

	language	countries
1	english	47
2	french	11
3	spanish	10
4	chinese	9
5	arabic	8
6	german	5
7	portugese	2
8	indian	2
9	russian	1
10	korean	1
11	italian	1
12	japanese	1



2. How many companies are in the database?

```
SELECT COUNT(DISTINCT(company_id)) AS num_of_companies_in_db
FROM tutorial.yammer_users AS users
ORDER BY num_of_companies_in_db DESC;
```

	num_of_companies_in_db
1	13198

****

3. Which is the company with the most users?

```
SELECT company_id,
COUNT(user_id) AS num_of_users_per_company,
SUM(CASE WHEN state='active' THEN 1 ELSE 0 END) AS active_users,
SUM(CASE WHEN state='pending' THEN 1 ELSE 0 END) AS pending_users
FROM tutorial.yammer_users AS users
WHERE state IS NOT NULL
GROUP BY company_id
ORDER BY num_of_users_per_company DESC
LIMIT 10;
```

	company_id	num_of_users_per_company	active_users	pending_users
1	1	20721	20141	580
2	2	10640	10373	267
3	3	6747	6585	162
4	4	4197	4061	136
5	6	3529	3446	83
6	5	3132	3043	89
7	8	2924	2855	69
8	9	2133	2084	49
9	7	2109	2039	70
10	13	1852	1812	40

Analyze Events

1. From which location are most of the events?

```
SELECT location,
COUNT(event_type) AS interactions
FROM tutorial.yammer_events AS events
GROUP BY location
ORDER BY interactions DESC
LIMIT 10;
```

	location	interactions
1	United States	94728
2	Japan	26046
3	Germany	23524
4	France	17364
5	United Kingdom	16475
6	Russia	12226
7	Italy	11790
8	Brazil	11240
9	India	9620
10	Canada	9126

****

2. Which are the most frequent events?

```
SELECT event_name,
COUNT(event_name) As action_count
FROM tutorial.yammer_events AS events
GROUP BY event_name
ORDER BY action_count DESC;
```

	event_name	action_count
1	home_page	94065
2	like_message	59248
3	view_inbox	55936
4	login	38610
5	send_message	33105
6	search_autocomplete	17820
7	search_run	13019
8	create_user	7298
9	enter_email	4407
10	enter_info	3872
11	complete_signup	3680
12	search_click_result_2	1499
13	search_click_result_1	1413
14	search_click_result_4	1264
15	search_click_result_3	1134
16	search_click_result_5	968
17	search_click_result_6	805
18	search_click_result_9	784
19	search_click_result_7	709
20	search_click_result_8	690
21	search_click_result_10	506

****

3. Which devices are used?

```
SELECT DISTINCT(device)
FROM tutorial.yammer_events AS events
ORDER BY device ASC;
```

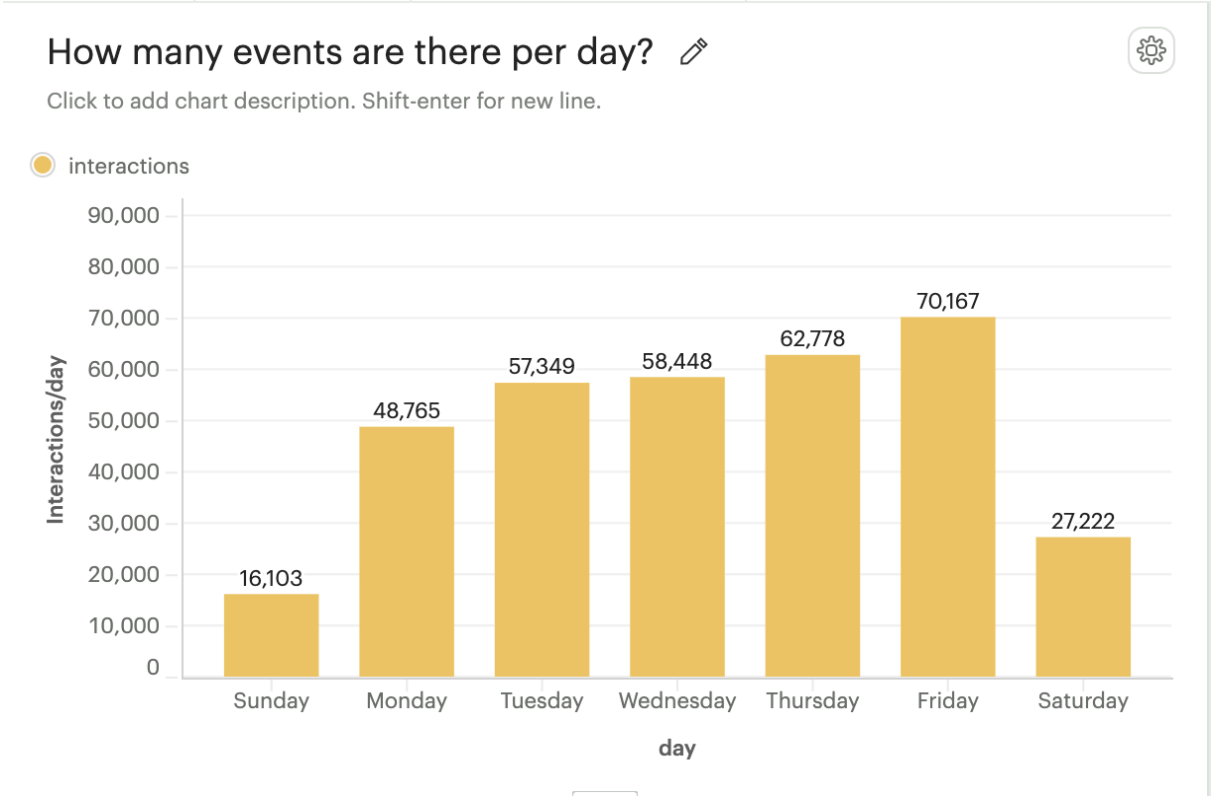
	device
1	acer aspire desktop
2	acer aspire notebook
3	amazon fire phone
4	asus chromebook
5	dell inspiron desktop
6	dell inspiron notebook
7	hp pavilion desktop
8	htc one
9	ipad air
10	ipad mini
11	iphone 4s
12	iphone 5
13	iphone 5s
14	kindle fire
15	lenovo thinkpad
16	macbook air
17	macbook pro
18	mac mini
19	nexus 10
20	nexus 5
21	nexus 7
22	nokia lumia 635
23	samsumg galaxy tablet
24	samsung galaxy note
25	samsung galaxy s4
26	windows surface

4. How many events are there per day?

```
SELECT DISTINCT day_of_week,
CASE
WHEN subquery.day_of_week=0 THEN 'Sunday'
WHEN subquery.day_of_week=1 THEN 'Monday'
WHEN subquery.day_of_week=2 THEN 'Tuesday'
WHEN subquery.day_of_week=3 THEN 'Wednesday'
WHEN subquery.day_of_week=4 THEN 'Thursday'
WHEN subquery.day_of_week=5 THEN 'Friday'
WHEN subquery.day_of_week=6 THEN 'Saturday'
END AS day,
COUNT(event_name) AS interactions
FROM
(
SELECT occurred_at,
    EXTRACT(DOW FROM occurred_at) AS day_of_week,
    event_name
FROM tutorial.yammer_events
) AS subquery
GROUP BY day_of_week
ORDER BY interactions DESC;
```

****

5. Create a chart for the events per day



Analyze Users and Events

1. Which company has the most logins?

```
SELECT sub.company_id,
            sub.login_count
FROM (
   SELECT company_id,
        SUM(CASE WHEN event_name = 'login' THEN 1 ELSE 0 END) AS login_count
   FROM tutorial.yammer_users AS users
   LEFT JOIN tutorial.yammer_events AS events
   USING(user_id)
   GROUP BY company_id
) AS sub
WHERE sub.login_count = (
   SELECT MAX(login_count)
   FROM (
      SELECT company_id,
           SUM(CASE WHEN event_name = 'login' THEN 1 ELSE 0 END) AS login_count
      FROM tutorial.yammer_users AS users
      LEFT JOIN tutorial.yammer_events AS events
      USING(user_id)
      GROUP BY company_id
   ) AS max_sub
)
ORDER BY sub.login_count DESC;
```

	company_id	login_count
1	1	2265

*

-For me this was tricky because I tried to use the max() on Sub.login_count in the bigger SELECT but always returned a list of login count for each company.

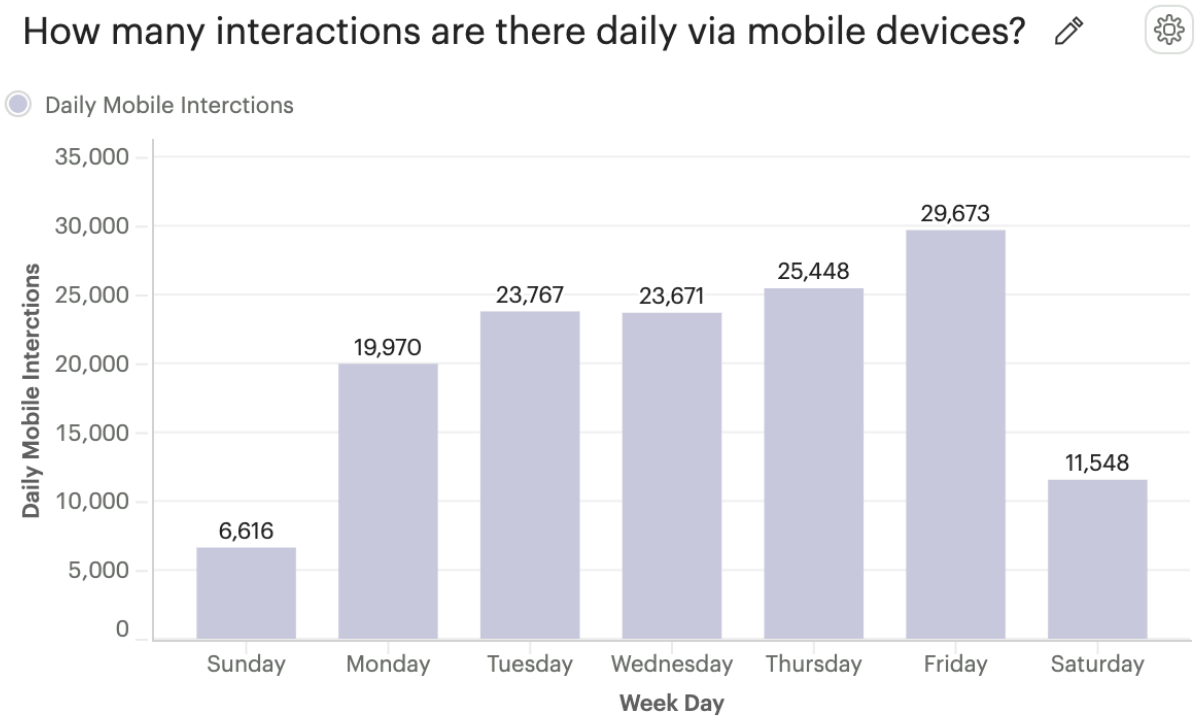
-Of Course I was able to select the max(login_count) value itself without the company name as MAX() With no GROUP BY returns one row. But then, how would we know which Company is this.

-Finally i figured that i have to get the max(login_count)in a separate nested query.

*e

2. How many interactions are there daily via mobile devices?

```
SELECT
CASE
WHEN subquery.day_of_week=0 THEN 'Sunday'
WHEN subquery.day_of_week=1 THEN 'Monday'
WHEN subquery.day_of_week=2 THEN 'Tuesday'
WHEN subquery.day_of_week=3 THEN 'Wednesday'
WHEN subquery.day_of_week=4 THEN 'Thursday'
WHEN subquery.day_of_week=5 THEN 'Friday'
WHEN subquery.day_of_week=6 THEN 'Saturday'
END AS day,
COUNT(event_type)
FROM
(
SELECT occurred_at,
    device,event_type,
    EXTRACT(DOW FROM occurred_at) AS day_of_week,
    event_name
FROM tutorial.yammer_events
WHERE NOT device ILIKE
 ANY(ARRAY['%desktop%','%chrom%book','%thinkpad%','%macbook%',
'%mac%','%notebook%','%windows%'])
) AS subquery
GROUP BY day,subquery.day_of_week
ORDER BY subquery.day_of_week ASC;
```

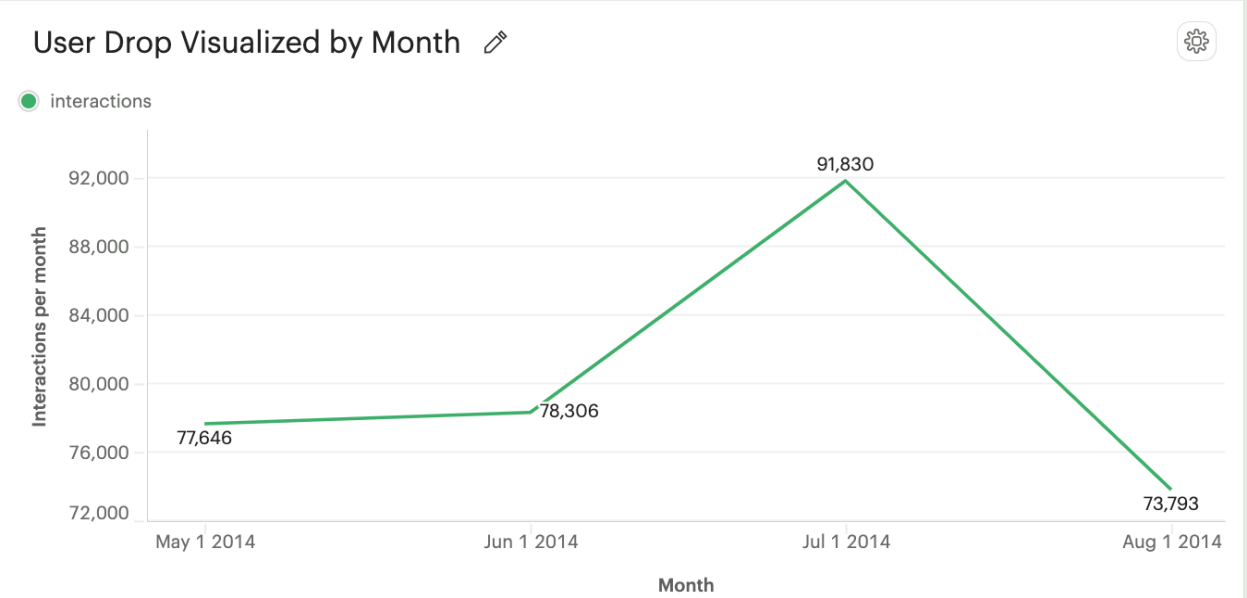


What is interesting to you?

Problem:
To investigate why there is a drop in user engagement.

[STEP 1] Visualizing the drop in users per month to detect when interactions started to drop

```
SELECT DATE_TRUNC('month',occurred_at) AS month,
  count(event_type) AS interactions
  FROM tutorial.yammer_events AS events
  WHERE event_type='engagement'
  GROUP BY month
  ORDER BY month ASC;
```



We can notice that the drop happened in August

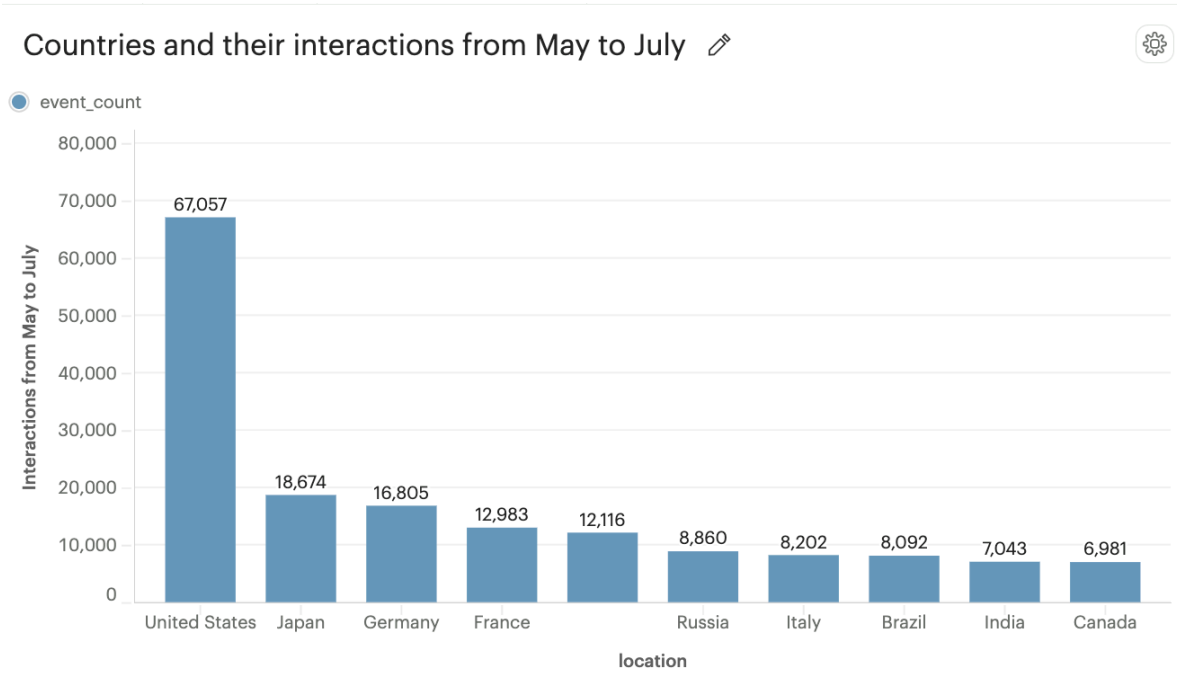
[STEP 2] Investigate the location to know in which country/countries caused that drop

```
SELECT location,
COUNT(event_name)  AS event_count
FROM tutorial.yammer_events AS events
WHERE event_type='engagement'
GROUP BY location
ORDER BY event_count DESC
 LIMIT 10;
```

	location	event_count
1	United States	89379
2	Japan	24584
3	Germany	22304
4	France	16469
5	United Kingdom	15590
6	Russia	11561
7	Italy	11051
8	Brazil	10607
9	India	9036
10	Mexico	8648

```
SELECT
  location,
  COUNT(event_name)  AS event_count
  FROM tutorial.yammer_events AS events
  WHERE event_type='engagement'
  AND  occurred_at >= '2014-05-1'
  AND occurred_at <= '2014-7-31'
 GROUP BY location
 ORDER BY event_count DESC
 LIMIT 10;
```

	location	event_count
1	United States	67057
2	Japan	18674
3	Germany	16805
4	France	12983
5	United Kingdom	12116
6	Russia	8860
7	Italy	8202
8	Brazil	8092
9	India	7043
10	Canada	6981



Results show that United States has the most interactions in this data set, so The investigation moves towards finding the reasons interactions drop in USA in August