Drop in yammer's User Engagement Analysis Using Postgre Sql.

Arunkumar

links: github | linkedin

1) Introduction:

/*
* Before using the data let us make the assumptions what made the engagement drop in active users in yammer
* we can see the chart of the engagement drop in the below link
https://app.mode.com/modeanalytics/reports/cbb8c291ee96/runs/7925c979521e/viz1/cfcdb6b78885
* Dataset Can be Accessed freely in Mode's public warehouse.

My raw assumptions are :

By learning the overview we know that yammer is a social network for communicating with co-workers

1) workers used when remote but not when they started working in onsite

2) Any Unexpected events like company getting closed or company no longer using the yammer

3) Any big promotions attracted the users and they didnt stayed after using it.

4) Issues in the yammer user experience like bugs in signing up , activation process features and other initiatives turns into backslashes

5) By overviewing the data i came to know it can also be the region specific , device specific, language specific.

6) lastly the data recorded or data logged itself a faulty one.
*/

-- Now let us confirm that the data we have are consistent before working with it.
-- we have 3 Tables let us run the each query separately

```
SELECT *
FROM tutorial.yammer_users
```

```
SELECT *
FROM tutorial.yammer_emails
```

```
SELECT *
FROM tutorial.yammer_events
```

-- Now after running the each query separately we know that the data is consistent
-- now we can move forward

2) **Trends of active & pending users:**

Drop in yammer's User Engagement Analysis Using Postgre Sql.

-- let us know the period of the data we are having

```
SELECT *
FROM tutorial.yammer_users
ORDER BY created_at
```

-- we are having the data from 2013-01-01

```
SELECT *
FROM tutorial.yammer_users
ORDER BY created_at DESC
```

-- we are having the data till 2014-08-31

-- Now i am trying to find any trends or patterns to be found in the active and pending users

-- users_count

```
SELECT COUNT( DISTINCT user_id)AS users_count,
    EXTRACT('month' FROM created_at) AS month,
    EXTRACT('year' FROM created_at) AS year
FROM tutorial.yammer_users
GROUP BY 2,3
ORDER BY 3,2
```

-- active_users_count

```
SELECT COUNT( DISTINCT user_id) AS active_users_count,
    EXTRACT('month' FROM created_at) AS month,
    EXTRACT('year' FROM created_at) AS year
FROM tutorial.yammer_users
WHERE state = 'active'
GROUP BY 2,3
ORDER BY 3,2
```

-- pending_users_count

```
SELECT COUNT( DISTINCT user_id) AS pending_users_count,
    EXTRACT('month' FROM created_at) AS month,
    EXTRACT('year' FROM created_at) AS year
FROM tutorial.yammer_users
WHERE state = 'pending'
GROUP BY 2,3
ORDER BY 3,2
```

-- active and pending users count

```
SELECT s1.month,
    s1.year,
    s1.pending_users_count,
    s2.active_users_count,
    s1.join_id
FROM
   (SELECT COUNT( DISTINCT user_id) AS pending_users_count,
       EXTRACT('month' FROM created_at) AS month,
       EXTRACT('year' FROM created_at) AS year,
       ROW_NUMBER() OVER (ORDER BY EXTRACT('year' FROM created_at) ) AS join_id
   FROM tutorial.yammer_users
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

```
   WHERE state = 'pending'
   GROUP BY 2,3
   ORDER BY 3,2)s1
JOIN
  (SELECT COUNT( DISTINCT user_id) AS active_users_count,
      EXTRACT('month' FROM created_at) AS month,
      EXTRACT('year' FROM created_at) AS year,
      ROW_NUMBER() OVER (ORDER BY EXTRACT('year' FROM created_at) ) AS join_id
  FROM tutorial.yammer_users
  WHERE state = 'active'
  GROUP BY 2,3
  ORDER BY 3,2)s2
ON s1.join_id = s2.join_id
```

-- percentage of active and pending users to the total users

-- USING WITH clause for simplified view

```
 WITH pending_active AS(
       SELECT s1.month,
           s1.year,
           s1.pending_users_count,
           s2.active_users_count,
           s1.join_id
       FROM
          (SELECT COUNT( DISTINCT user_id) AS pending_users_count,
              EXTRACT('month' FROM created_at) AS month,
              EXTRACT('year' FROM created_at) AS year,
              ROW_NUMBER() OVER (ORDER BY EXTRACT('year' FROM created_at) ) AS join_id
          FROM tutorial.yammer_users
          WHERE state = 'pending'
          GROUP BY 2,3
          ORDER BY 3,2)s1
       JOIN
         (SELECT COUNT( DISTINCT user_id) AS active_users_count,
             EXTRACT('month' FROM created_at) AS month,
             EXTRACT('year' FROM created_at) AS year,
             ROW_NUMBER() OVER (ORDER BY EXTRACT('year' FROM created_at) ) AS join_id
         FROM tutorial.yammer_users
         WHERE state = 'active'
         GROUP BY 2,3
         ORDER BY 3,2)s2
       ON s1.join_id = s2.join_id),
   users_total_count AS(
       SELECT COUNT( DISTINCT user_id)AS users_count,
        EXTRACT('month' FROM created_at) AS month,
        EXTRACT('year' FROM created_at) AS year,
        ROW_NUMBER() OVER (ORDER BY EXTRACT('year' FROM created_at) ) AS join_id
        FROM tutorial.yammer_users
        GROUP BY 2,3
        ORDER BY 3,2)
SELECT use.users_count ,
    pen.month,
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

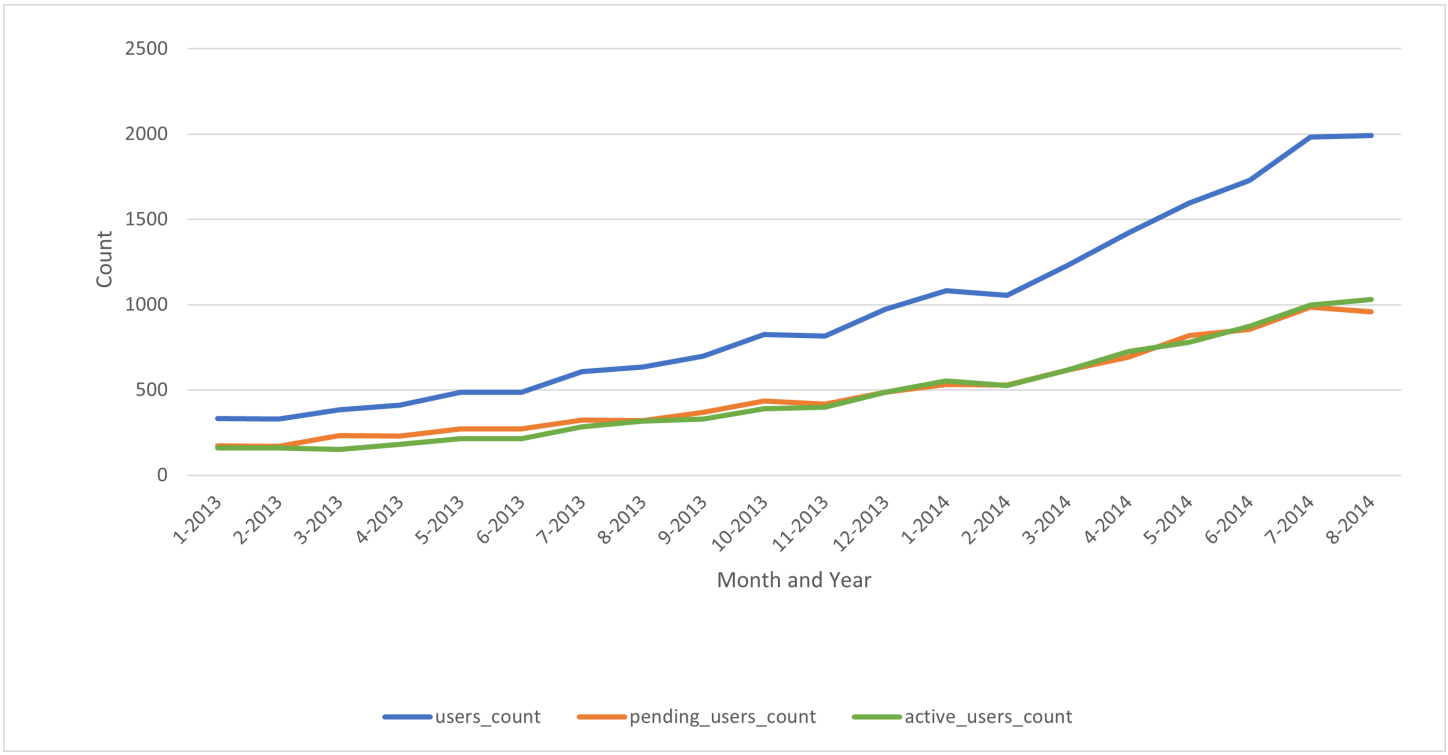
```
    pen.year,
    pen.pending_users_count,
    pen.active_users_count,
    (pen.active_users_count::float / use.users_count::float) * 100 AS active_users_percentage,
    ( pen.pending_users_count::float/ use.users_count::float) * 100 AS pending_users_percentage
FROM pending_active pen
JOIN users_total_count use
ON pen.join_id =use.join_id
```

/*

* after finding the active users and pending users percentage i came to know that there is no significant trends or patterns, more or like it same.

* so i can now can explore the data or process the data and analyse the data only for the latest engagement rise and fall which is from 2014-07-01 to 2014-08-31

* From this i came to know that my assumption that the remote workers turned to the office site becomes invalid because
there is no significant trends or patterns , more or like it same.



Drop in yammer's User Engagement Analysis Using Postgre Sql.

Data Fields Source Export Copy

	users count	month	year	pending users count	active users count	active users percentage	pending users percentage
1	332	1	2013	172	160	48.1928	51.8072
2	328	2	2013	168	160	48.7805	51.2195
3	383	3	2013	233	150	39.1645	60.8355
4	410	4	2013	229	181	44.1463	55.8537
5	486	5	2013	272	214	44.0329	55.9671
6	485	6	2013	272	213	43.9175	56.0825
7	608	7	2013	324	284	46.7105	53.2895
8	636	8	2013	320	316	49.6855	50.3145
9	699	9	2013	369	330	47.2103	52.7897
10	826	10	2013	436	390	47.2155	52.7845
11	816	11	2013	417	399	48.8971	51.1029
12	972	12	2013	486	486	50	50
13	1083	1	2014	531	552	50.9695	49.0305
14	1054	2	2014	529	525	49.8102	50.1898
15	1231	3	2014	616	615	49.9594	50.0406
16	1419	4	2014	693	726	51.1628	48.8372
17	1597	5	2014	818	779	48.7790	51.2210
18	1728	6	2014	855	873	50.5208	49.4792
19	1983	7	2014	986	997	50.2774	49.7226
20	1990	8	2014	959	1031	51.8090	48.1910

3) signup problem for activation:

/*
By overviewing the data in the table events we came to know that the we can know the how many times the users stuck in the signup process by using the column event_type
*/
-- now iam trying ot know how many event_types are there,

```
SELECT distinct event_type
FROM tutorial.yammer_events
```

/*
now i know there area only 2 event_types
1) signup_flow and
2) engagement
now i want to know how many times the user struck in the signup_flow itself because this itself leads to bad user experience .
*/

```
SELECT distinct user_id,
    COUNT(*) as no_of_times_user_signed_up
FROM tutorial.yammer_events
WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- highest a user faces to signup is 4 in july

```
SELECT distinct user_id,
    COUNT(*) as no_of_times_user_signed_up
FROM tutorial.yammer_events
WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

-- highest a user faces to signup is 4 in august

-- now let us know how much the users faced the signup issue in both the months categorised by the no_of_signups
-- For july month

```
SELECT COUNT( DISTINCT user_id) AS no_of_users_in_july,
    no_of_times_user_signed_up
FROM
   (SELECT DISTINCT user_id,
       COUNT(*) as no_of_times_user_signed_up
   FROM tutorial.yammer_events
   WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
   GROUP BY 1
   ORDER BY 2 DESC)s3
GROUP BY 2
ORDER BY 1 DESC
```

-- for august month

```
SELECT COUNT( DISTINCT user_id) AS no_of_users_in_august,
    no_of_times_user_signed_up
FROM
   (SELECT DISTINCT user_id,
       COUNT(*) as no_of_times_user_signed_up
   FROM tutorial.yammer_events
   WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
   GROUP BY 1
   ORDER BY 2 DESC) s4
GROUP BY 2
ORDER BY 1 DESC
```

-- Now atlast i want to compare both the months how much users are there

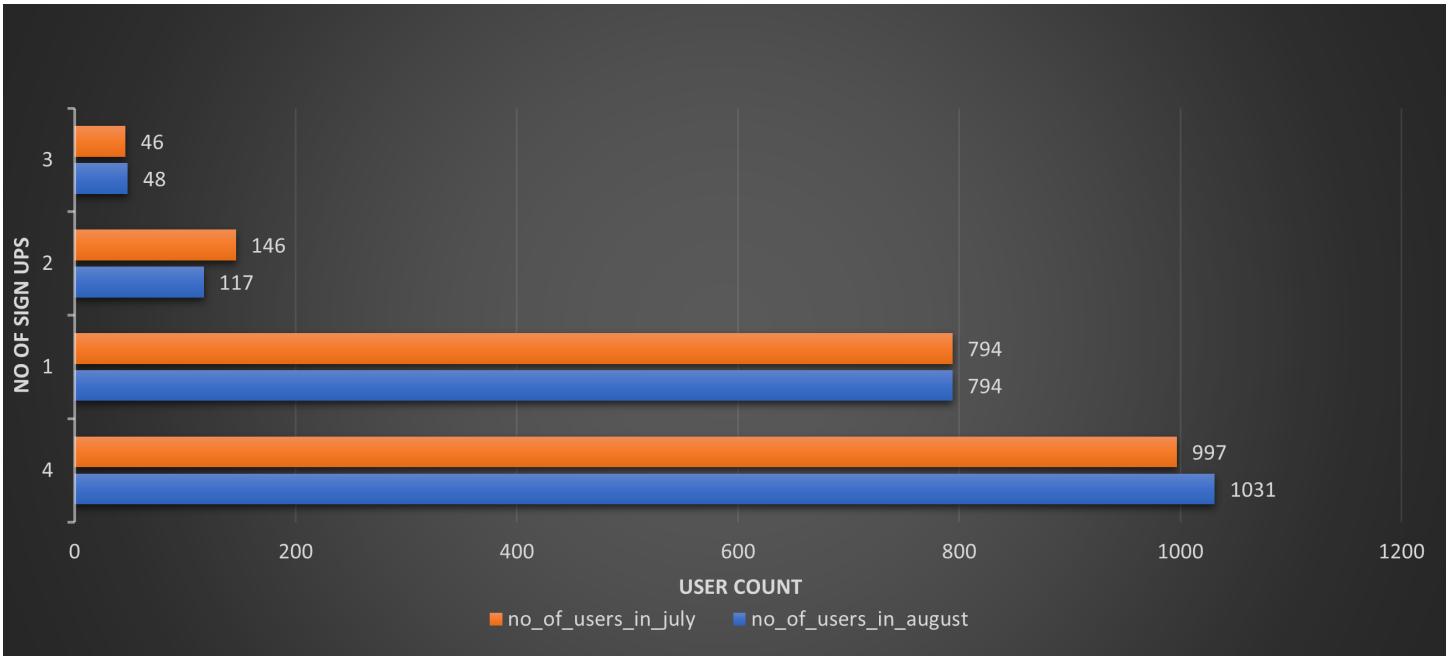
```
WITH july_users AS (
       SELECT COUNT( DISTINCT user_id) AS no_of_users_in_july,
           no_of_times_user_signed_up
       FROM
          (SELECT DISTINCT user_id,
              COUNT(*) as no_of_times_user_signed_up
          FROM tutorial.yammer_events
          WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2
014
          GROUP BY 1
          ORDER BY 2 DESC)s3
          GROUP BY 2
          ORDER BY 1 DESC ),
   august_users AS(
          SELECT COUNT( DISTINCT user_id) AS no_of_users_in_august,
           no_of_times_user_signed_up
          FROM
             (SELECT DISTINCT user_id,
                 COUNT(*) as no_of_times_user_signed_up
             FROM tutorial.yammer_events
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

```
            WHERE event_type = 'signup_flow' AND EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at)
= 2014
            GROUP BY 1
            ORDER BY 2 DESC) s4
         GROUP BY 2
         ORDER BY 1 DESC )
SELECT aug.no_of_users_in_august ,
    jul.no_of_users_in_july,
    jul.no_of_times_user_signed_up
FROM july_users jul
JOIN august_users aug
ON jul. no_of_times_user_signed_up = aug. no_of_times_user_signed_up
```

Data | Fields | Source

	no of users in august	no of users in july	no of times user signed up
1	1031	997	4
2	794	794	1
3	117	146	2
4	48	46	3



/*
Now my assumption which is due to user experience in specific to signup process is invalid beacause by comparing i came to know that there is no significant difference in the no of users experienced the the signup more in august than july it is almost the same.
*/

4) Checking company left out the yammer:

-- now lets check whether there are any company left out of yammer recently

Drop in yammer's User Engagement Analysis Using Postgre Sql.

-- JULY MONTH

```
SELECT COUNT(DISTINCT company_id)AS july_month
FROM tutorial.yammer_users
WHERE EXTRACT('month' FROM created_at) = 7 AND EXTRACT('year' FROM created_at) = 2014
```

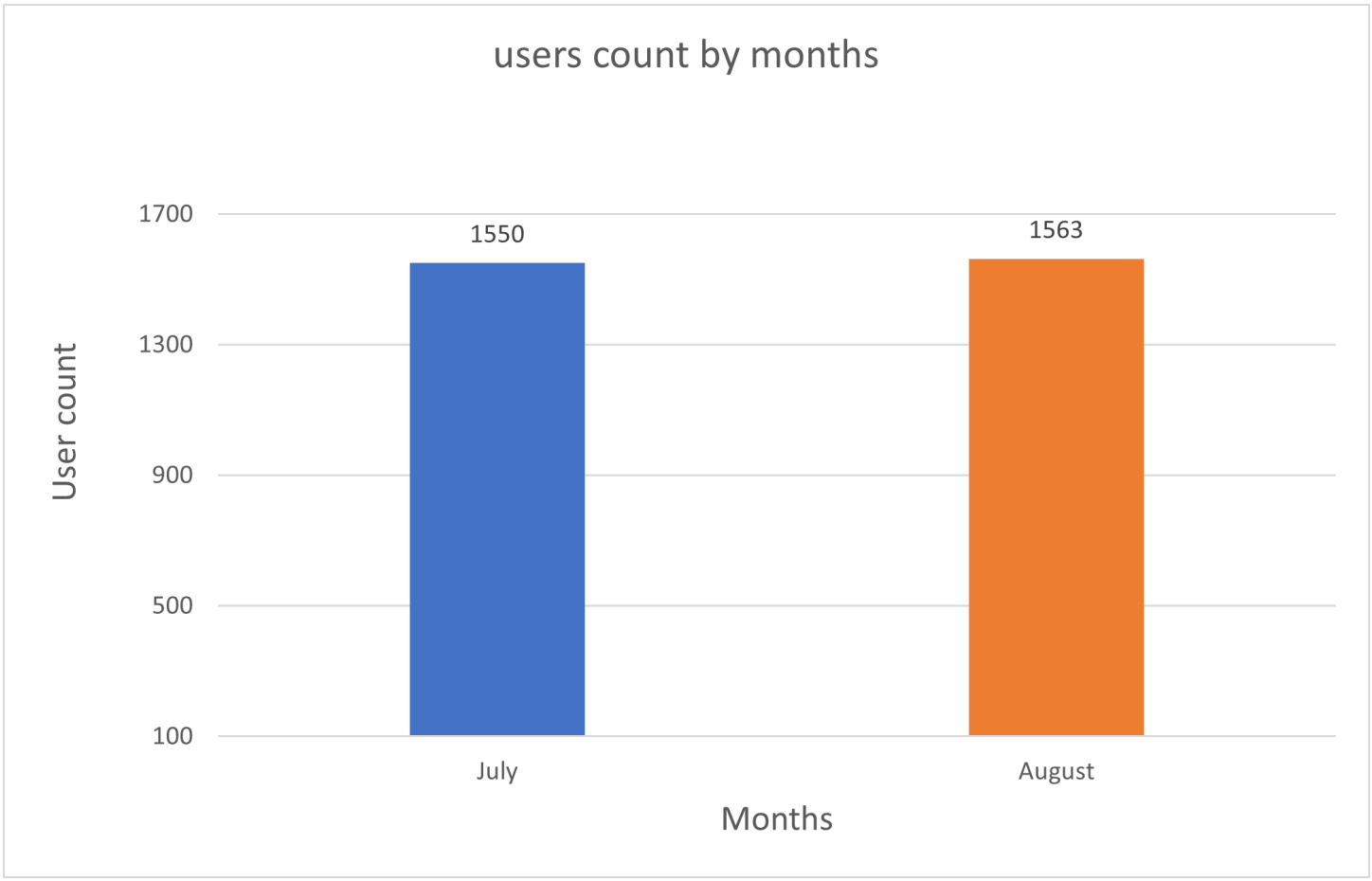
-- AUGUST MONTH

```
SELECT COUNT(DISTINCT company_id)AS august_month
FROM tutorial.yammer_users
WHERE EXTRACT('month' FROM created_at) = 8 AND EXTRACT('year' FROM created_at) = 2014
```

Data Fields Source

	july_month	august_month
1	1550	1563

Drop in yammer's User Engagement Analysis Using Postgre Sql.



/*
* Precisely the companies gets increased slightly in august = 1563 than the july = 1550.

* Now we dont want to check the users in the company because we already know the total count in our query while checking the
percentage of users in active and pending where there were pretty much similar.

* if we encounter any significant differnce in company we would have asked the higher offical to get the data of the company
but here it is nullfied because of increase in company.
*/

5) types of events:

-- Now i am trying to get how many times a event occurred in july and august
--july events

```
SELECT event_name,
    COUNT(user_id) no_of_times_events_occurred
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

--august events

```
SELECT event_name,
    COUNT(user_id) no_of_times_events_occurred
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

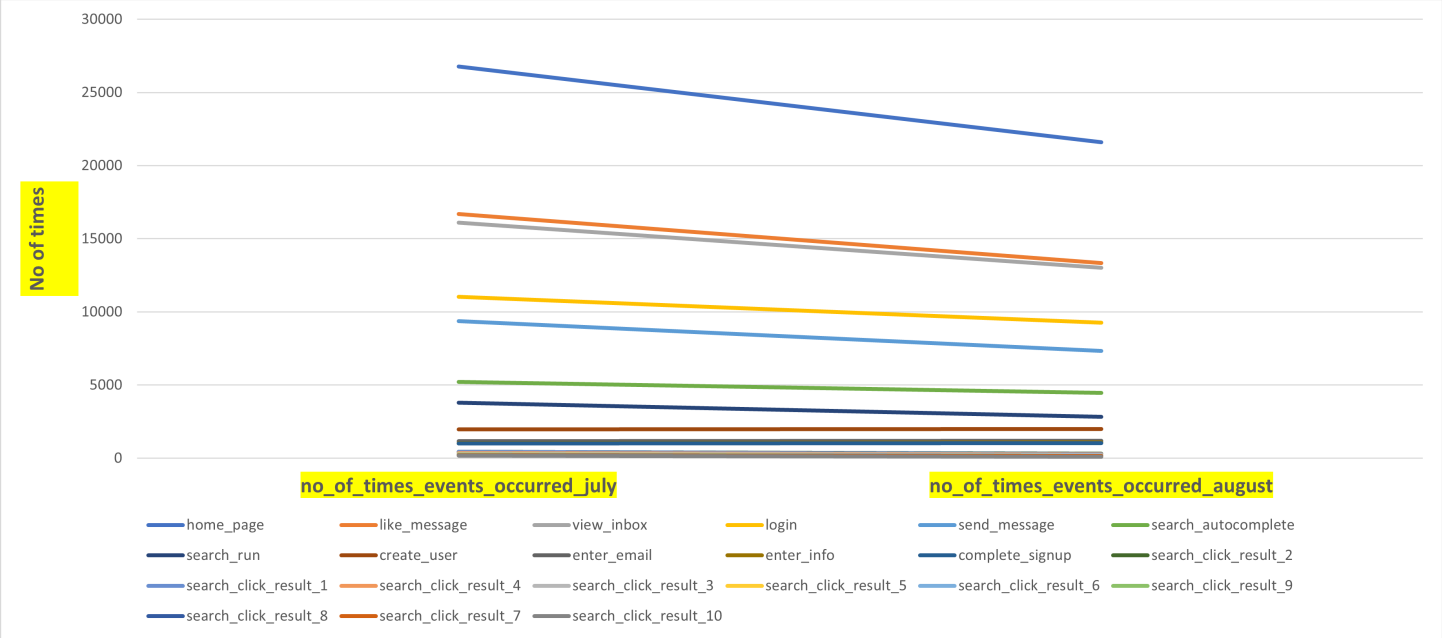
-- here i am trying to comapare the no of times the event occurred in july and august
-- to get an idea whether the event specific features is caused anything bad

```
WITH july_events AS (
        SELECT event_name AS event_name,
            COUNT(user_id) AS no_of_times_events_occurred_july
        FROM tutorial.yammer_events
        WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC ),
   august_events AS(
        SELECT event_name AS event_name,
            COUNT(user_id) AS no_of_times_events_occurred_august
        FROM tutorial.yammer_events
        WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC)
SELECT jul.event_name AS event_name,
    jul.no_of_times_events_occurred_july,
    aug.no_of_times_events_occurred_august,
   ( (aug.no_of_times_events_occurred_august - jul.no_of_times_events_occurred_july):: FLOAT/
    aug.no_of_times_events_occurred_august::FLOAT ) * 100 AS percentage_of_events_occurred_in_aug_with_july
FROM july_events jul
JOIN august_events aug
ON jul.event_name = aug.event_name
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

Data Fields Source

	event_name	no_of_times_events_occurred_july	no_of_times_events_occurred_august	percentage_of_events_occurred_in_aug_with_july
1	home_page	26765	21603	-23.8948
2	like_message	16691	13332	-25.1950
3	view_inbox	16088	13011	-23.6492
4	login	11023	9271	-18.8976
5	send_message	9363	7324	-27.8400
6	search_autocomplete	5226	4469	-16.9389
7	search_run	3803	2836	-34.0973
8	create_user	1983	1990	0.3518
9	enter_email	1189	1196	0.5853
10	enter_info	1043	1079	3.3364
11	complete_signup	997	1031	3.2978
12	search_click_result_2	449	296	-51.6892
13	search_click_result_1	432	274	-57.6642
14	search_click_result_4	361	257	-40.4669
15	search_click_result_3	305	235	-29.7872
16	search_click_result_5	266	182	-46.1538
17	search_click_result_6	247	170	-45.2941
18	search_click_result_9	237	142	-66.9014
19	search_click_result_8	211	135	-56.2963
20	search_click_result_7	200	153	-30.7190
21	search_click_result_...	163	103	-58.2524



/*

* Yes after comparing, there was good drop in engagement in august in comparison with july with event types which are the features
of the yammer.

* but it is not significant enough to tell this is a cause for engagement and it is worth noting it. Also it seems proportional
because active users drop leads to event or yammer's features usage drop.

*/

Drop in yammer's User Engagement Analysis Using Postgre Sql.

6) language specific issue is there ?

/*
* now let us know any language specific features leads to users drop in engagement

* i am trying to compare the users using the language in july and august.
*/

-- july users

```
SELECT language,
    COUNT(distinct user_id) no_of_users_in_july
FROM tutorial.yammer_users
WHERE EXTRACT('month' FROM created_at) = 7 AND EXTRACT('year' FROM created_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- august users

```
SELECT language,
    COUNT(distinct user_id) no_of_users_in_august
FROM tutorial.yammer_users
WHERE EXTRACT('month' FROM created_at) = 8 AND EXTRACT('year' FROM created_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

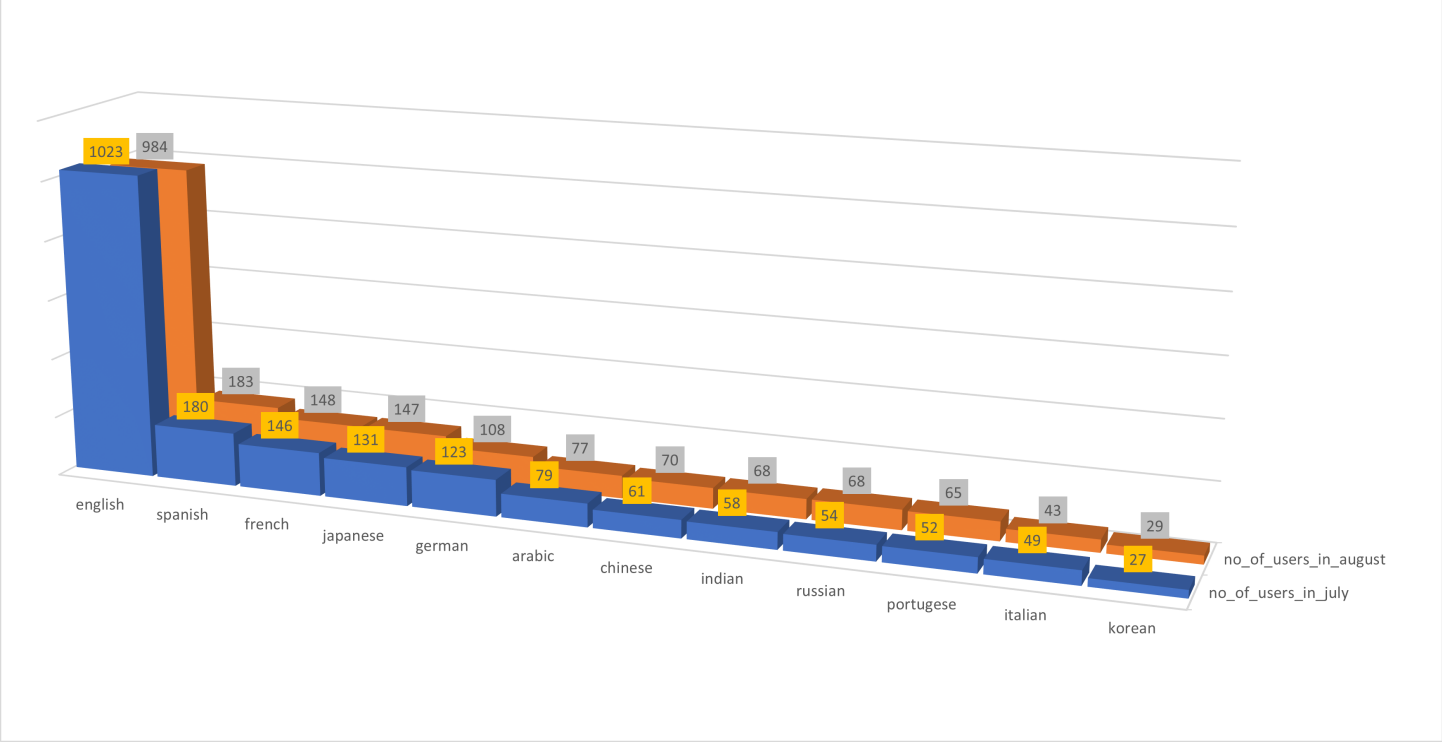
-- comparison between the no of users using the language in july and august

```
WITH july_users AS(
        SELECT language,
            COUNT(distinct user_id) no_of_users_in_july
        FROM tutorial.yammer_users
        WHERE EXTRACT('month' FROM created_at) = 7 AND EXTRACT('year' FROM created_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC ) ,
   august_users AS (
        SELECT language,
            COUNT(distinct user_id) no_of_users_in_august
        FROM tutorial.yammer_users
        WHERE EXTRACT('month' FROM created_at) = 8 AND EXTRACT('year' FROM created_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC )
SELECT jul.language,
    jul.no_of_users_in_july ,
    aug.no_of_users_in_august,
    ((aug.no_of_users_in_august - jul.no_of_users_in_july)::FLOAT
    /aug.no_of_users_in_august::FLOAT) * 100 AS percentage_of_difference_in_users_languages
FROM july_users jul
FULL JOIN august_users aug
ON jul.language = aug.language
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

Data | Fields | Source — Export | Copy

	language	no of users in july	no of users in august	percentage of difference in users languages
1	english	1023	984	-3.9634
2	spanish	180	183	1.6393
3	french	146	148	1.3514
4	japanese	131	147	10.8844
5	german	123	108	-13.8889
6	arabic	79	77	-2.5974
7	chinese	61	70	12.8571
8	indian	58	68	14.7059
9	russian	54	68	20.5882
10	portugese	52	65	20
11	italian	49	43	-13.9535
12	korean	27	29	6.8966



/*
* After comaparing there is no significant difference in the users using their language so our language specific issue is not
there and therefore it is invalid.
*/

7) Location specific issue:

/*
Here i am trying to check whether there is a issue arises in location or region specific
which can forces users to not use the yammer.
*/

Drop in yammer's User Engagement Analysis Using Postgre Sql.

-- july users

```
SELECT location,
    COUNT(DISTINCT user_id)
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- august users

```
SELECT location,
    COUNT(DISTINCT user_id)
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- comparison between no of users in the countries in july and august

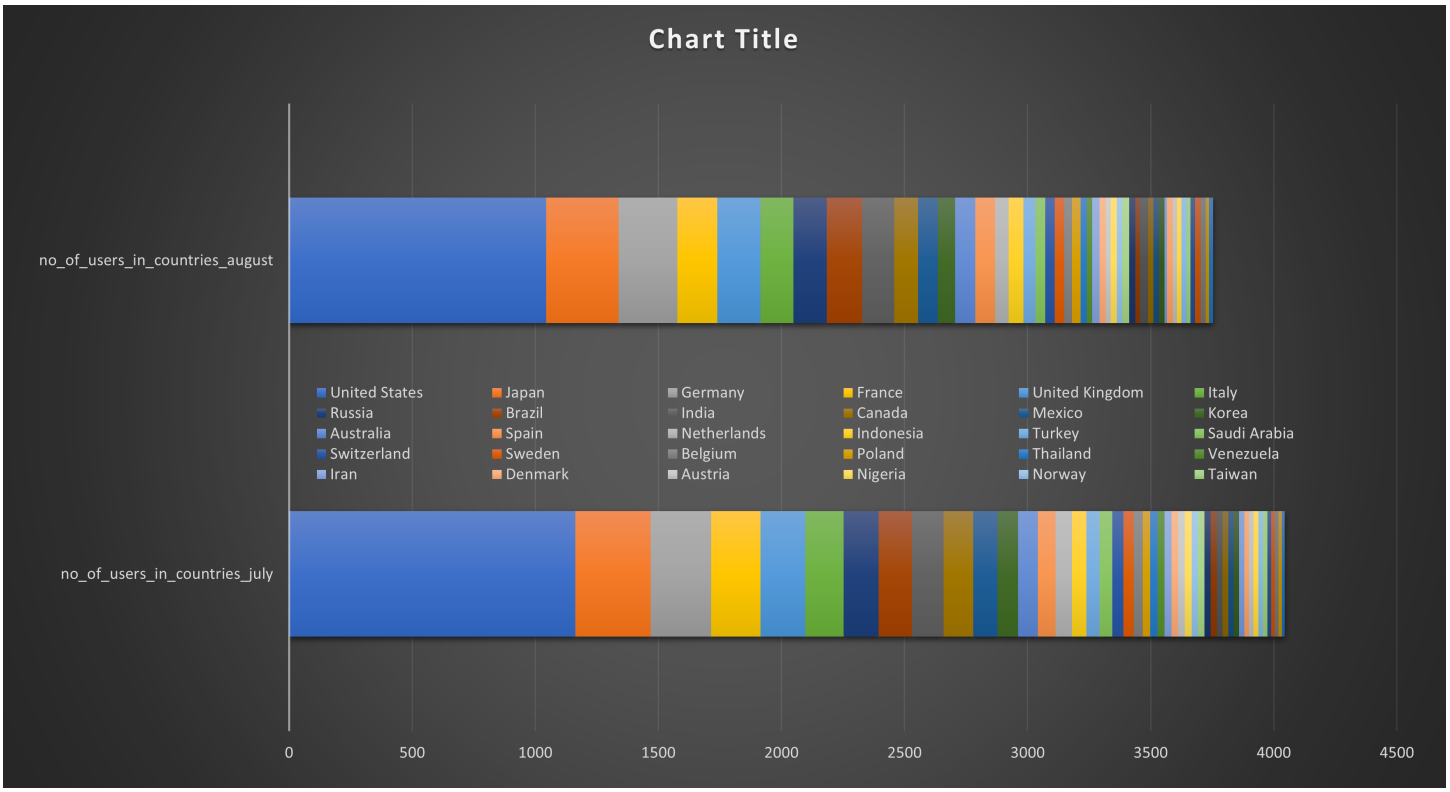
```
WITH july_users AS (
        SELECT location,
           COUNT(DISTINCT user_id) AS no_of_users_in_countries_july
        FROM tutorial.yammer_events
        WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC ) ,
   august_users AS (
        SELECT location,
            COUNT(DISTINCT user_id)AS no_of_users_in_countries_august
        FROM tutorial.yammer_events
        WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC )
SELECT jul.location,
    jul.no_of_users_in_countries_july,
    aug.no_of_users_in_countries_august,
   ( (aug.no_of_users_in_countries_august - jul.no_of_users_in_countries_july)::FLOAT
    /aug.no_of_users_in_countries_august::FLOAT ) * 100
    AS percentage_diff_in_no_of_users_in_the_countries_in_july_and_august
FROM july_users jul
FULL JOIN august_users aug
ON jul.location = aug.location
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

Data Fields Source

	location	no_of_users_in_countries_july	no_of_users_in_countries_august	percentage_diff_in_no_of_users_in_the_countries_in_july_and_aug
1	United States	1163	1045	-11.2919
2	Japan	305	293	-4.0956
3	Germany	246	240	-2.5
4	France	201	162	-24.0741
5	United Kingdom	183	175	-4.5714
6	Italy	155	133	-16.5414
7	Russia	142	136	-4.4118
8	Brazil	135	144	6.25
9	India	129	131	1.5267
10	Canada	120	96	-25
11	Mexico	99	81	-22.2222
12	Korea	83	70	-18.5714
13	Australia	81	82	1.2195
14	Spain	72	79	8.8608
15	Netherlands	67	56	-19.6429
16	Indonesia	58	60	3.3333
17	Turkey	54	48	-12.5
18	Saudi Arabia	51	40	-27.5
19	Switzerland	46	39	-17.9487
20	Sweden	41	39	-5.1282
21	Belgium	36	32	-12.5
22	Poland	31	34	8.8235
23	Thailand	30	26	-15.3846
24	Venezuela	29	22	-31.8182
25	Iran	28	29	3.4483
26	Denmark	27	26	-3.8462
27	Austria	27	20	-35
28	Nigeria	27	25	-8
29	Norway	26	21	-23.8095
30	Taiwan	26	28	7.1429
31	United Arab Emirat...	26	25	-4
32	Finland	24	19	-26.3158
33	Malaysia	24	34	29.4118
34	Colombia	23	22	-4.5455
35	Israel	22	21	-4.7619
36	South Africa	22	23	4.3478
37	Chile	21	10	-110
38	Argentina	19	24	20.8333
39	Singapore	19	17	-11.7647
40	Egypt	19	18	-5.5556
41	Philippines	19	21	9.5238
42	Hong Kong	18	15	-20
43	Iraq	16	19	15.7895
44	Greece	15	24	37.5
45	Portugal	14	19	26.3158
46	Pakistan	13	14	7.1429
47	Ireland	12	17	29.4118

Drop in yammer's User Engagement Analysis Using Postgre Sql.



/*

* After checking out i came to know that there is a slight drop in users in countries when i see it in absolute numbers but when i saw in percentage it shows like a huge difference but i should not take the percentage because the distinct no of users itself small .

* so even slight change can leads to big difference in users base.

* so i would say location specific issue is not significant enough to cause the user engagement drop.

*/

8) device specific issue:

/*

Here i am , trying to check is there any device specific issue and it is related to user engagement drop.

*/

--july_devices

```
SELECT device,
    COUNT(*) AS no_of_times_device_used_jul
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- august devices

Drop in yammer's User Engagement Analysis Using Postgre Sql.

```
SELECT device AS device ,
    COUNT(*) AS no_of_times_device_used_aug
FROM tutorial.yammer_events
WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

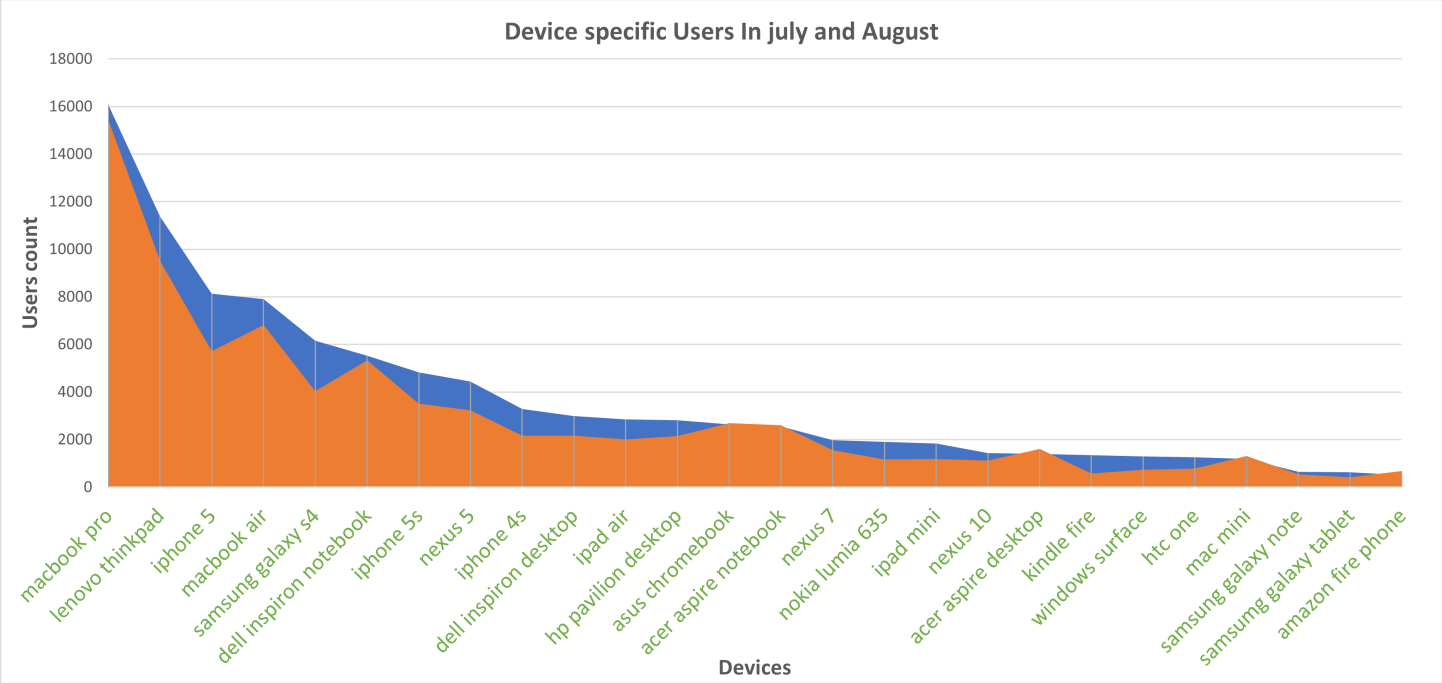
-- now i am going to comapare the no of times each device used in july and august

```
WITH july_devices AS (
        SELECT device AS device,
            COUNT(*) AS no_of_times_device_used_jul
        FROM tutorial.yammer_events
        WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC ),
   august_devices AS (
          SELECT device,
              COUNT(*) AS no_of_times_device_used_aug
          FROM tutorial.yammer_events
          WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
          GROUP BY 1
          ORDER BY 2 DESC )
SELECT jul.device,
    jul.no_of_times_device_used_jul,
    aug.no_of_times_device_used_aug,
    ((aug.no_of_times_device_used_aug - jul.no_of_times_device_used_jul)::decimal
   /aug.no_of_times_device_used_aug::decimal ) * 100 AS percentage_diff_in_devices_used_jul_aug
FROM july_devices jul
FULL JOIN august_devices aug
ON jul.device = aug.device
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

Data | Fields | Source

	device	no of times device used jul	no of times device used aug	percentage diff in devices used jul aug
1	macbook pro	16085	15462	-4.0292
2	lenovo thinkpad	11376	9510	-19.6215
3	iphone 5	8137	5711	-42.4794
4	macbook air	7902	6825	-15.7802
5	samsung galaxy s4	6156	4033	-52.6407
6	dell inspiron notebook	5531	5333	-3.7127
7	iphone 5s	4832	3514	-37.5071
8	nexus 5	4448	3229	-37.7516
9	iphone 4s	3289	2159	-52.3390
10	dell inspiron desktop	2991	2173	-37.6438
11	ipad air	2857	2010	-42.1393
12	hp pavilion desktop	2820	2145	-31.4685
13	asus chromebook	2643	2700	2.1111
14	acer aspire notebook	2558	2608	1.9172
15	nexus 7	1981	1560	-26.9872
16	nokia lumia 635	1903	1164	-63.4880
17	ipad mini	1831	1185	-54.5148
18	nexus 10	1434	1112	-28.9568
19	acer aspire desktop	1401	1604	12.6559
20	kindle fire	1353	581	-132.8744
21	windows surface	1284	737	-74.2198
22	htc one	1261	791	-59.4185
23	mac mini	1191	1310	9.0840
24	samsung galaxy note	641	540	-18.7037
25	samsumg galaxy tabl...	633	417	-51.7986
26	amazon fire phone	504	676	25.4438



Drop in yammer's User Engagement Analysis Using Postgre Sql.

/*
After getting the results i came to know the absolute number is huge so i decided to go with percentage difference and
while going through it i encountered that there is a significant difference in devices used in iphone 5 and samsung galaxy s4 which is 42 % approximately , 2426 in absolute numbers and 52 % approximately , 2123 in absolute numbers.
*/

9) email issue:

-- Atlast i want to check is there any issue with emails send to the users.

-- july_emails

```
SELECT action,
    COUNT(*) AS no_of_times_emails_sent_jul
FROM tutorial.yammer_emails
WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

-- august_emails

```
SELECT action,
    COUNT(*) AS no_of_times_emails_sent_aug
FROM tutorial.yammer_emails
WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
GROUP BY 1
ORDER BY 2 DESC
```

/*
* by above the results from both the months first i need to check the clickthrough rates to the email
sent and email openned in both months.
* After that i want to compare with both months.
*/

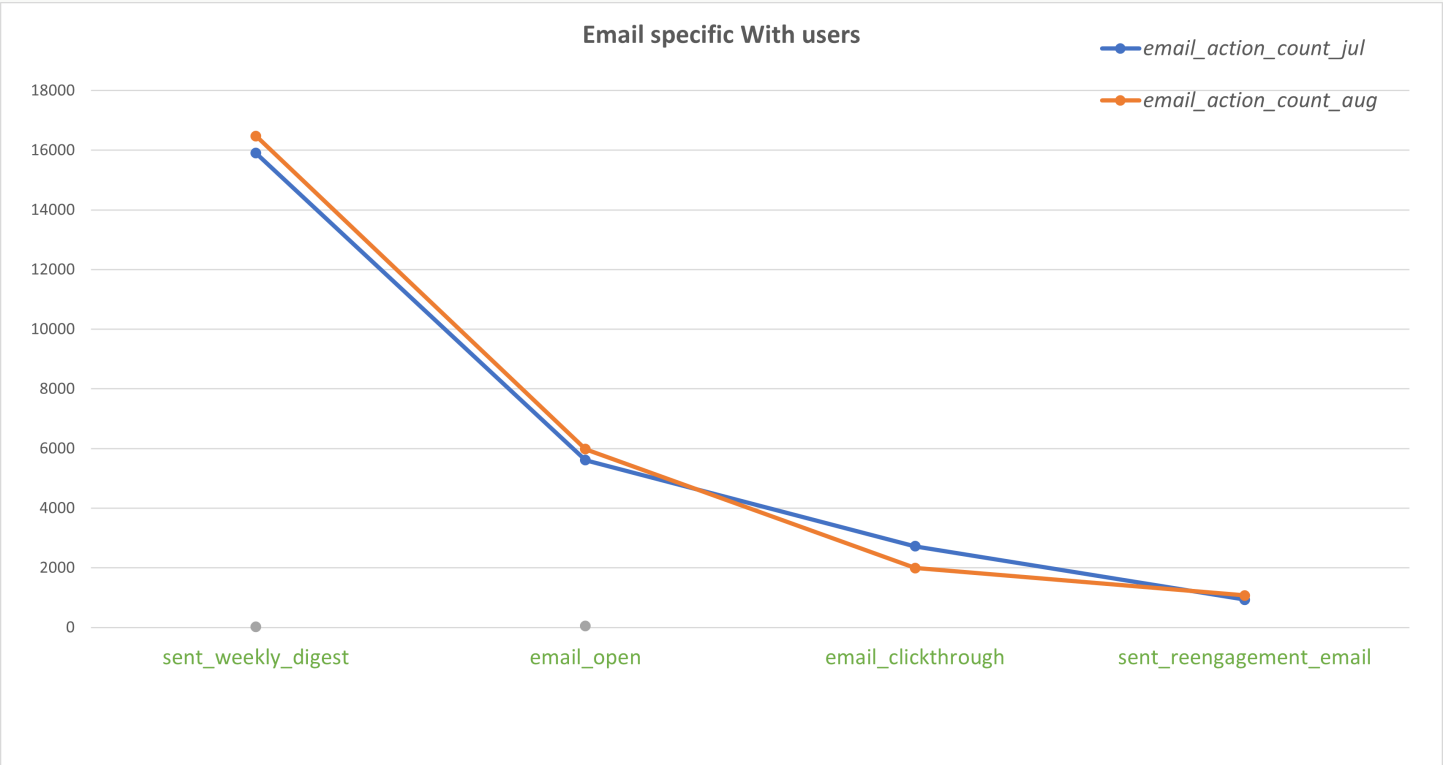
```
WITH july_emails AS (
        SELECT action,
            COUNT(*) AS email_action_count_jul,
            CASE WHEN action ='sent_weekly_digest'OR action = 'email_open' THEN 2721 ELSE NULL
                                END AS click_through_jul
        FROM tutorial.yammer_emails
        WHERE EXTRACT('month' FROM occurred_at) = 7 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
        ORDER BY 2 DESC ) ,
   august_emails AS (
        SELECT action,
            COUNT(*) AS email_action_count_aug,
            CASE WHEN action ='sent_weekly_digest'OR action = 'email_open' THEN 1992 ELSE NULL
                                END AS click_through_aug
        FROM tutorial.yammer_emails
        WHERE EXTRACT('month' FROM occurred_at) = 8 AND EXTRACT('year' FROM occurred_at) = 2014
        GROUP BY 1
```

Drop in yammer's User Engagement Analysis Using Postgre Sql.

```
        ORDER BY 2 DESC )
SELECT jul.action,
    jul.email_action_count_jul,
    aug.email_action_count_aug,
    (jul.click_through_jul::float / jul.email_action_count_jul::float) * 100 AS click_through_rate_to_action_jul,
    (aug.click_through_aug::float / aug.email_action_count_aug::float) * 100 AS click_through_rate_to_action_aug
FROM july_emails jul
FUll JOIN august_emails aug
ON jul.action = aug.action
```

Data | Fields | Source — Export | Copy

	action	email_action_count_jul	email_action_count_aug	click_through_rate_to_action_jul	click_through_rate_to_action_aug
1	sent_weekly_digest	15902	16480	17.1111	12.0874
2	email_open	5611	5978	48.4940	33.3222
3	email_clickthrough	2721	1992		
4	sent_reengagement_em...	933	1073		



/*
1)now after comparing july and august month we found out that click-through rate to the sent emails were dropped from 17 % to
12 % and click_through rate to the opened emails were dropped from 48 % to 33 % approximately.

2)so i would say this can be one of the issue for the user engagement drop in august.
*/

10) Conclusion:

Drop in yammer's User Engagement Analysis Using Postgre Sql.

Findings Of My Analysis:

1) Users experiencing the signups for 4 or more times and the bad part about this more than 50% of the people experiencing which is greater than the people combined in experiencing (2,3,4) times signups.

2) There was a device specific issue in iphone 5 and samsung galaxy s4 ,where the active users are dropped by 42 % approximately 2426 in absolute numbers (iphone-5) and 52 % approximately , 2123 in absolute numbers (Samsung Galaxy s4).

3) Now after comparing july and august month we found out that click-through rate to the sent emails were dropped from 17 % to 12 % and click_through rate to the opened emails were dropped from 48 % to 33 % approximately.

Recommendations:

1) Have to fix the issue in signups and optimize the user experience .

2) Must be active in responding to the user reviews and act upon it.

3) Have to create more engaging and relevant emails and sent to people for increasing the click through rate.

*/