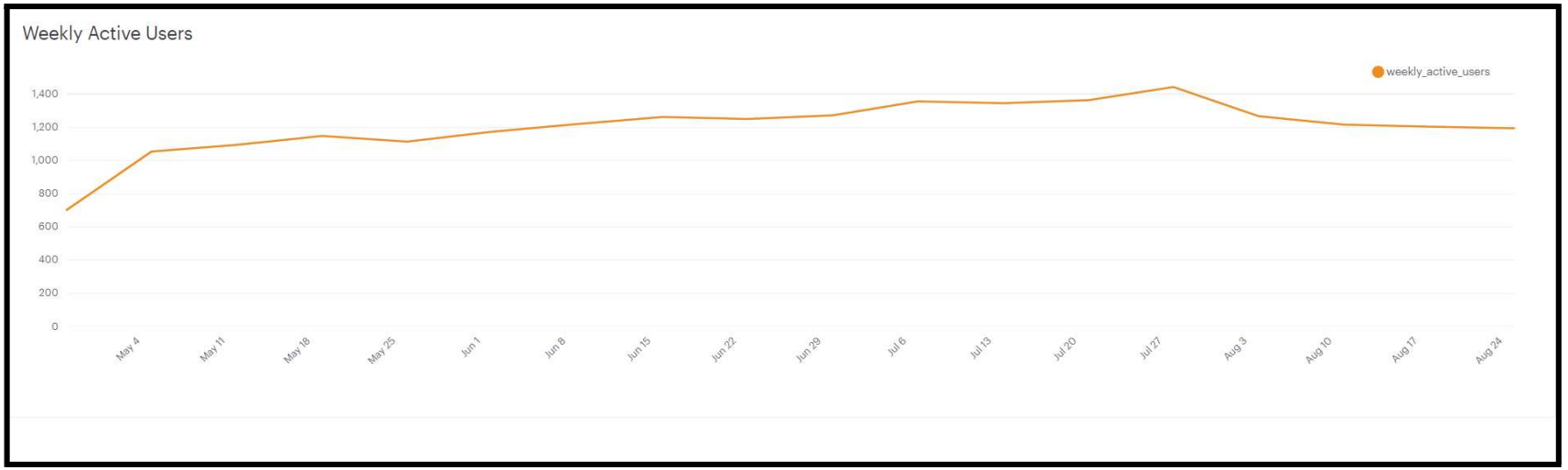
**Project: Investigating a Drop in User Engagement - Yammer**




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."

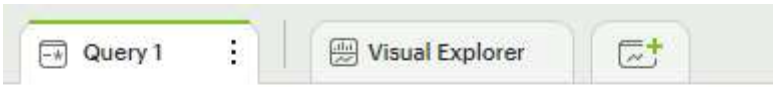
You are responsible for determining what caused the dip at the end of the chart shown above and, if appropriate, recommending solutions for the problem.

1. **Understand User Growth Rate**

Query:

```
SELECT
DATE_TRUNC('day',created_at) AS day,
COUNT(*) as all_users,
COUNT(CASE WHEN state = 'active' THEN user_id END) AS active_users
FROM tutorial.yammer_users
WHERE created_at >= '2014-06-01 00:00:00' AND created_at < '2014-09-01 00:00:00'
GROUP BY 1
```

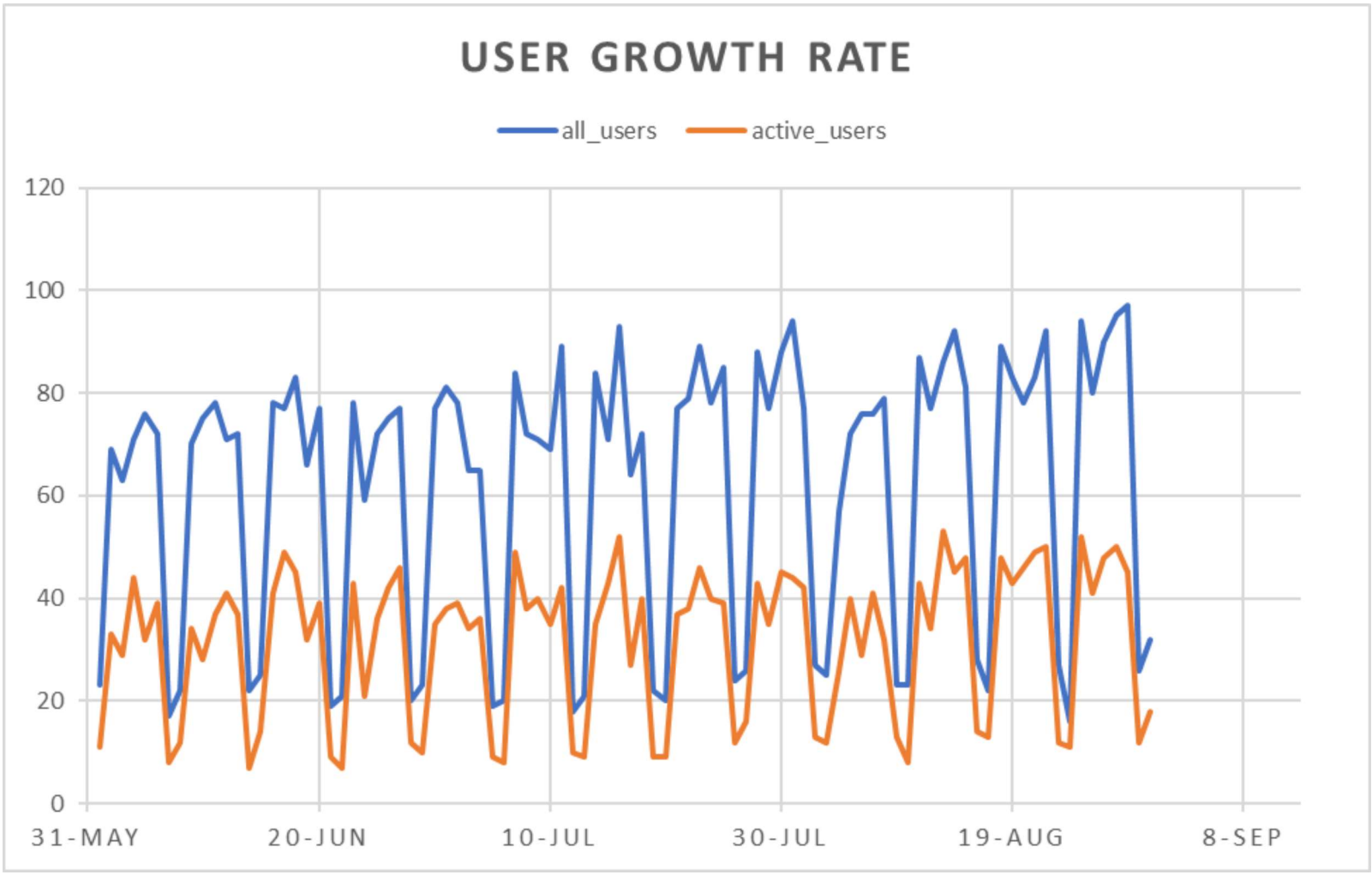
Output:




Data | Fields | Source

| | day | all_users | active_users |
|---|---|---|---|
| 1 | 2014-06-01 00:00:00 | 23 | 11 |
| 2 | 2014-06-02 00:00:00 | 69 | 33 |
| 3 | 2014-06-03 00:00:00 | 63 | 29 |
| 4 | 2014-06-04 00:00:00 | 71 | 44 |
| 5 | 2014-06-05 00:00:00 | 76 | 32 |
| 6 | 2014-06-06 00:00:00 | 72 | 39 |
| 7 | 2014-06-07 00:00:00 | 17 | 8 |
| 8 | 2014-06-08 00:00:00 | 22 | 12 |
| 9 | 2014-06-09 00:00:00 | 70 | 34 |
| 10 | 2014-06-10 00:00:00 | 75 | 28 |
| 11 | 2014-06-11 00:00:00 | 78 | 37 |
| 12 | 2014-06-12 00:00:00 | 71 | 41 |
| 13 | 2014-06-13 00:00:00 | 72 | 37 |
| 14 | 2014-06-14 00:00:00 | 22 | 7 |
| 15 | 2014-06-15 00:00:00 | 25 | 14 |
| 16 | 2014-06-16 00:00:00 | 78 | 41 |
| 17 | 2014-06-17 00:00:00 | 77 | 49 |
| 18 | 2014-06-18 00:00:00 | 83 | 45 |
| 19 | 2014-06-19 00:00:00 | 66 | 32 |
| 20 | 2014-06-20 00:00:00 | 77 | 39 |

Chart:




Findings:

Nothing has changed for growth rate - user growth rate remains high during weekday and low during weekends

**2. Understand the Engagement between User Age - Cohorting the users based on when they signed up**

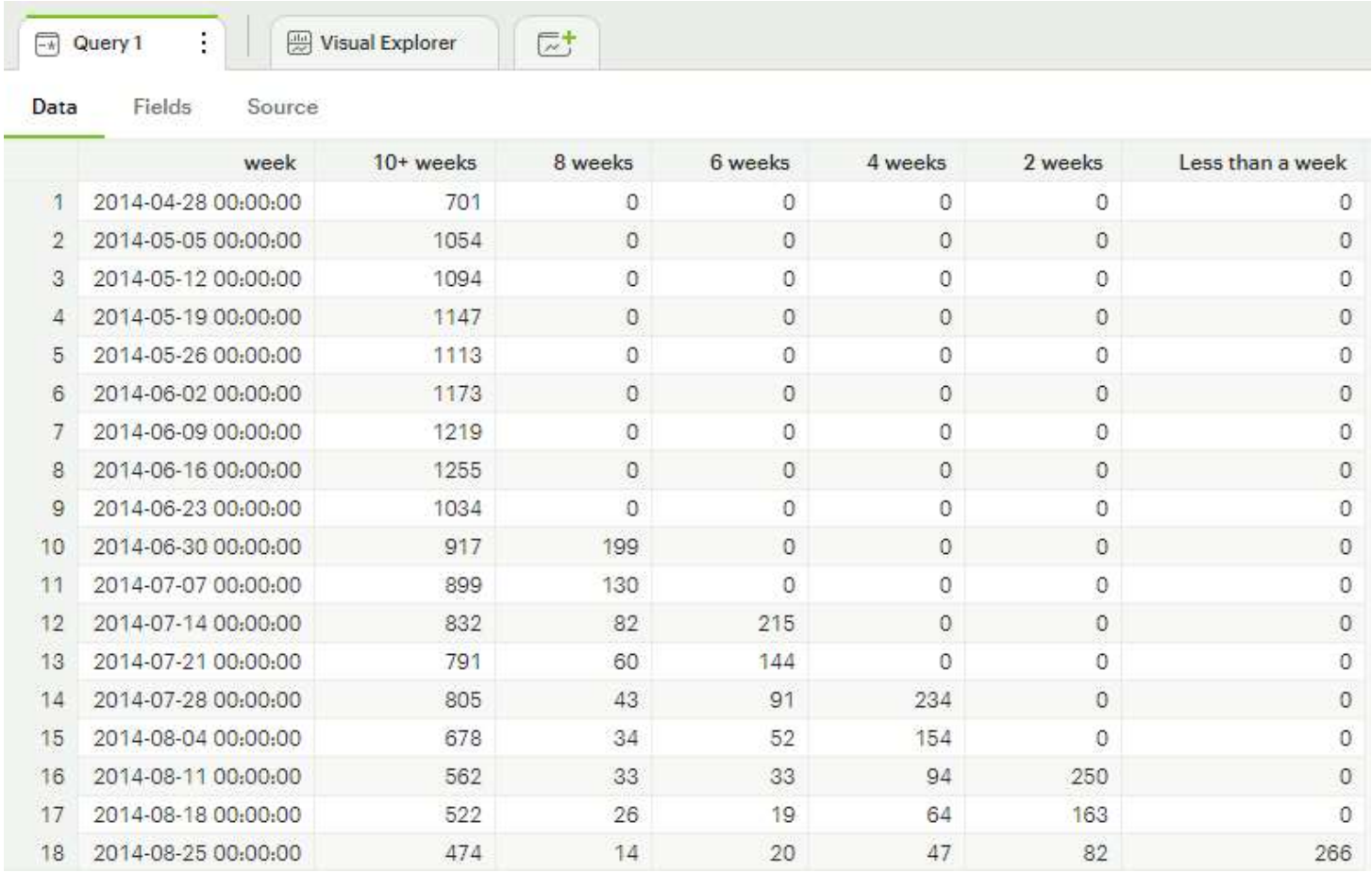
Query:

```
SELECT DATE_TRUNC('week',z.occurred_date) AS "week",
      COUNT(DISTINCT CASE WHEN z.user_age > 70 THEN z.user_id ELSE NULL END) AS "10+ weeks",
      COUNT(DISTINCT CASE WHEN z.user_age < 63 AND z.user_age >= 56 THEN z.user_id ELSE NULL END) AS "8 weeks",
      COUNT(DISTINCT CASE WHEN z.user_age < 49 AND z.user_age >= 42 THEN z.user_id ELSE NULL END) AS "6 weeks",
      COUNT(DISTINCT CASE WHEN z.user_age < 35 AND z.user_age >= 28 THEN z.user_id ELSE NULL END) AS "4 weeks",
      COUNT(DISTINCT CASE WHEN z.user_age < 21 AND z.user_age >= 14 THEN z.user_id ELSE NULL END) AS "2 weeks",
      COUNT(DISTINCT CASE WHEN z.user_age < 7 THEN z.user_id ELSE NULL END) AS "Less than a week"
FROM
(
SELECT
u.user_id,
DATE_TRUNC('week',e.occurred_at) AS occurred_date,
DATE_TRUNC('week',u.activated_at) AS activated_week,
DATE_TRUNC('day',e.occurred_at - u.activated_at) AS user_age_during_event,
EXTRACT('day'FROM '2014-09-01 00:00:00' - u.activated_at) AS user_age
       FROM tutorial.yammer_users u
       JOIN tutorial.yammer_events e
       ON e.user_id = u.user_id
        AND e.event_type = 'engagement'
          AND e.event_name = 'login'
          AND e.occurred_at >= '2014-05-01'
          AND e.occurred_at < '2014-09-01'
        WHERE u.activated_at IS NOT NULL
ORDER BY 2
```

*) z*

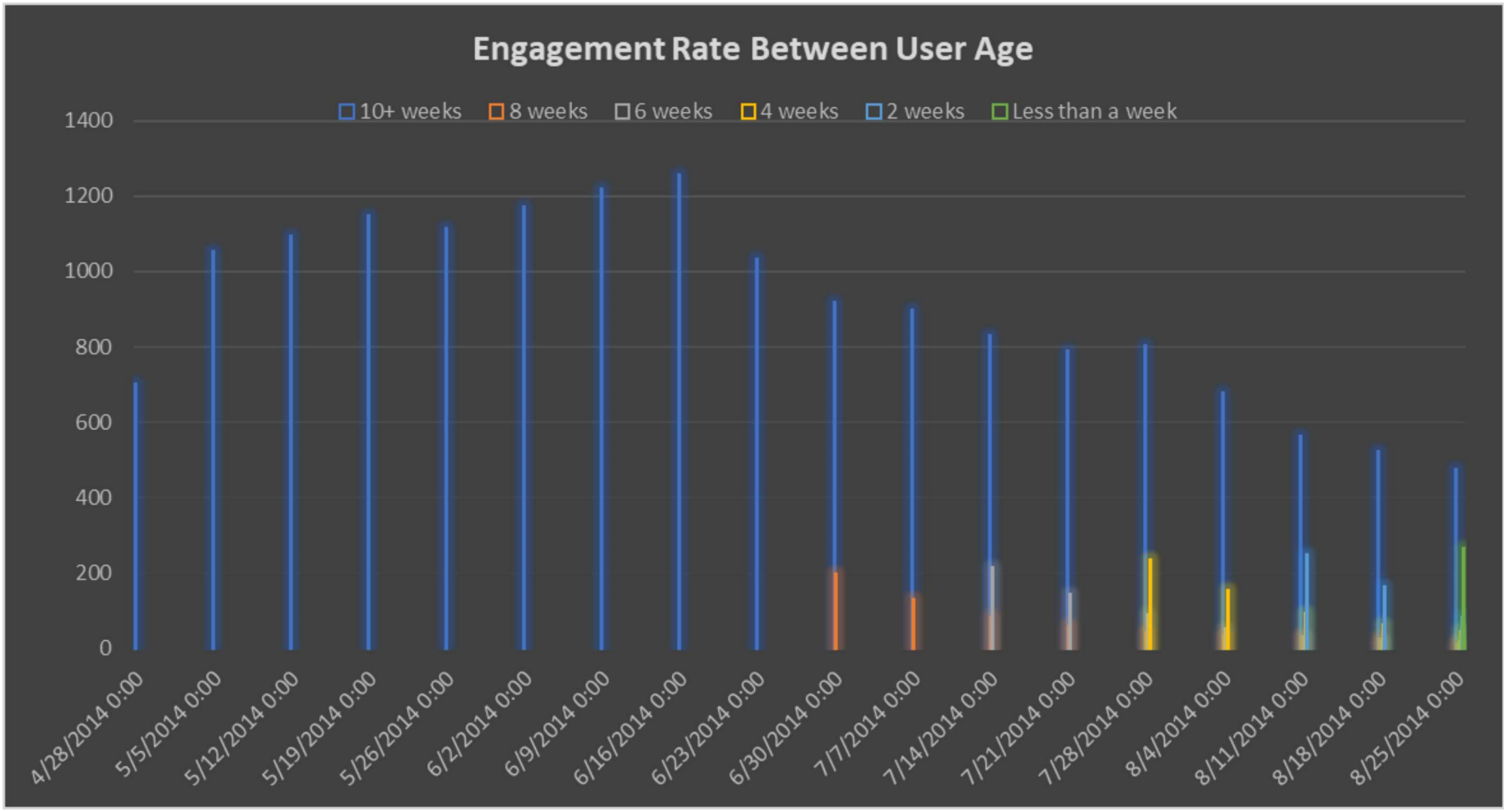
*GROUP BY 1*
*ORDER BY 1*

Output:




| | week | 10+ weeks | 8 weeks | 6 weeks | 4 weeks | 2 weeks | Less than a week |
|---|---|---|---|---|---|---|---|
| 1 | 2014-04-28 00:00:00 | 701 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2014-05-05 00:00:00 | 1054 | 0 | 0 | 0 | 0 | 0 |
| 3 | 2014-05-12 00:00:00 | 1094 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2014-05-19 00:00:00 | 1147 | 0 | 0 | 0 | 0 | 0 |
| 5 | 2014-05-26 00:00:00 | 1113 | 0 | 0 | 0 | 0 | 0 |
| 6 | 2014-06-02 00:00:00 | 1173 | 0 | 0 | 0 | 0 | 0 |
| 7 | 2014-06-09 00:00:00 | 1219 | 0 | 0 | 0 | 0 | 0 |
| 8 | 2014-06-16 00:00:00 | 1255 | 0 | 0 | 0 | 0 | 0 |
| 9 | 2014-06-23 00:00:00 | 1034 | 0 | 0 | 0 | 0 | 0 |
| 10 | 2014-06-30 00:00:00 | 917 | 199 | 0 | 0 | 0 | 0 |
| 11 | 2014-07-07 00:00:00 | 899 | 130 | 0 | 0 | 0 | 0 |
| 12 | 2014-07-14 00:00:00 | 832 | 82 | 215 | 0 | 0 | 0 |
| 13 | 2014-07-21 00:00:00 | 791 | 60 | 144 | 0 | 0 | 0 |
| 14 | 2014-07-28 00:00:00 | 805 | 43 | 91 | 234 | 0 | 0 |
| 15 | 2014-08-04 00:00:00 | 678 | 34 | 52 | 154 | 0 | 0 |
| 16 | 2014-08-11 00:00:00 | 562 | 33 | 33 | 94 | 250 | 0 |
| 17 | 2014-08-18 00:00:00 | 522 | 26 | 19 | 64 | 163 | 0 |
| 18 | 2014-08-25 00:00:00 | 474 | 14 | 20 | 47 | 82 | 266 |

Chart:




Findings:

Understood that the dip in engagement is localized within 10+ weeks user age - older users.

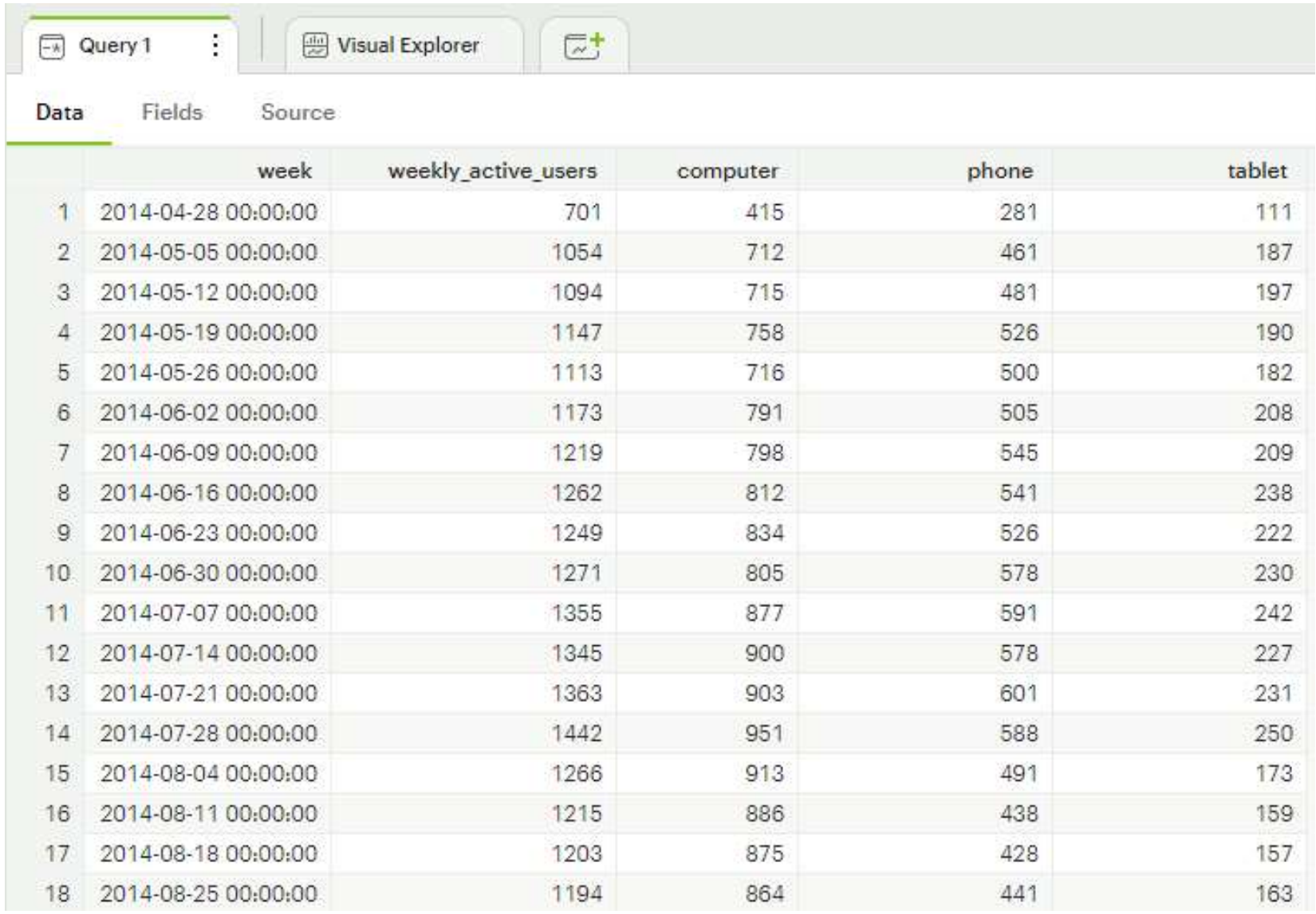
Hence leading us to believe that the issue does not lie within search engine ranks, marketing spike or site blocked.

**3. To understand whether the issue is localized to any device in particular**

Query:

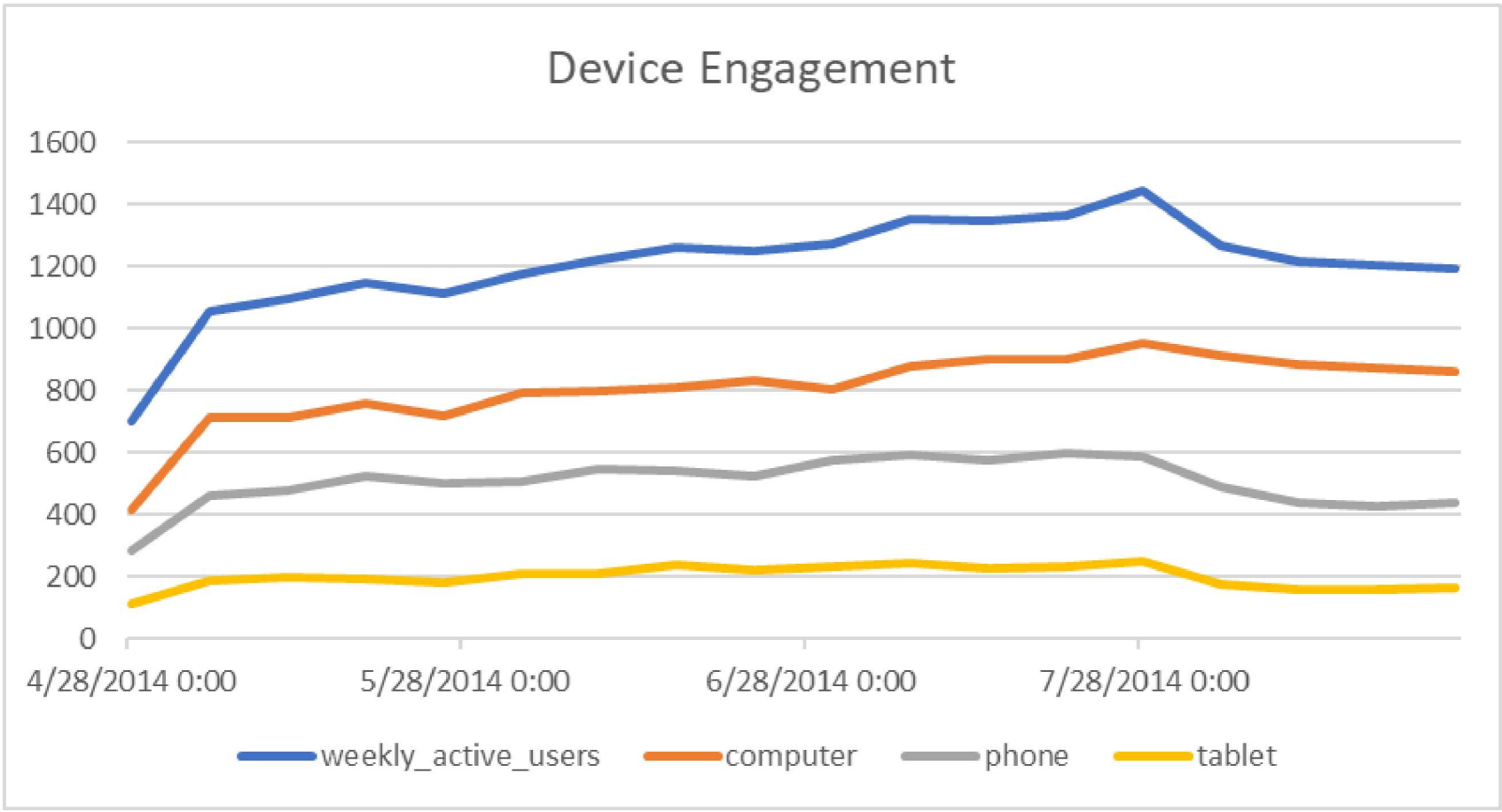
```
SELECT DATE_TRUNC('week', occurred_at) AS week,
      COUNT(DISTINCT user_id) AS weekly_active_users,
      COUNT(DISTINCT CASE WHEN device IN ('macbook pro','lenovo thinkpad','macbook air','dell inspiron notebook',
        'asus chromebook','dell inspiron desktop','acer aspire notebook','hp pavilion 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
WHERE event_type = 'engagement'
  AND event_name = 'login'
GROUP BY 1
ORDER BY 1
```

Output:




| | week | weekly_active_users | computer | phone | tablet |
|---|---|---|---|---|---|
| 1 | 2014-04-28 00:00:00 | 701 | 415 | 281 | 111 |
| 2 | 2014-05-05 00:00:00 | 1054 | 712 | 461 | 187 |
| 3 | 2014-05-12 00:00:00 | 1094 | 715 | 481 | 197 |
| 4 | 2014-05-19 00:00:00 | 1147 | 758 | 526 | 190 |
| 5 | 2014-05-26 00:00:00 | 1113 | 716 | 500 | 182 |
| 6 | 2014-06-02 00:00:00 | 1173 | 791 | 505 | 208 |
| 7 | 2014-06-09 00:00:00 | 1219 | 798 | 545 | 209 |
| 8 | 2014-06-16 00:00:00 | 1262 | 812 | 541 | 238 |
| 9 | 2014-06-23 00:00:00 | 1249 | 834 | 526 | 222 |
| 10 | 2014-06-30 00:00:00 | 1271 | 805 | 578 | 230 |
| 11 | 2014-07-07 00:00:00 | 1355 | 877 | 591 | 242 |
| 12 | 2014-07-14 00:00:00 | 1345 | 900 | 578 | 227 |
| 13 | 2014-07-21 00:00:00 | 1363 | 903 | 601 | 231 |
| 14 | 2014-07-28 00:00:00 | 1442 | 951 | 588 | 250 |
| 15 | 2014-08-04 00:00:00 | 1266 | 913 | 491 | 173 |
| 16 | 2014-08-11 00:00:00 | 1215 | 886 | 438 | 159 |
| 17 | 2014-08-18 00:00:00 | 1203 | 875 | 428 | 157 |
| 18 | 2014-08-25 00:00:00 | 1194 | 864 | 441 | 163 |

Chart:




Findings:

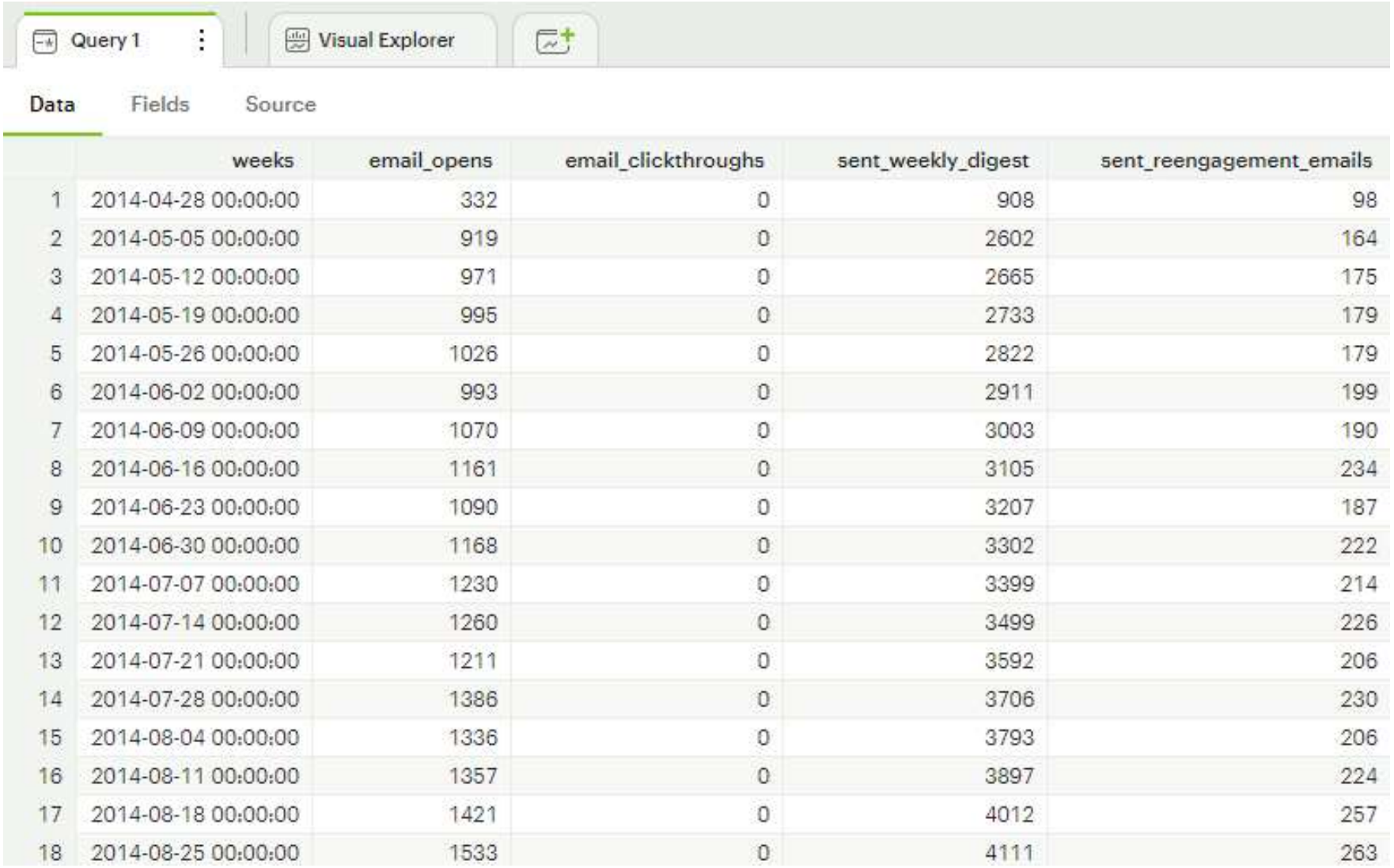
The chart shows a steep drop in phone devices in particular. There could be something that has been changed in the phone interface that's causing a disruption. Since it's localized to older users - another area which we can dive deeper is email digestion.

**4. Look into email digestion patterns**

Query:

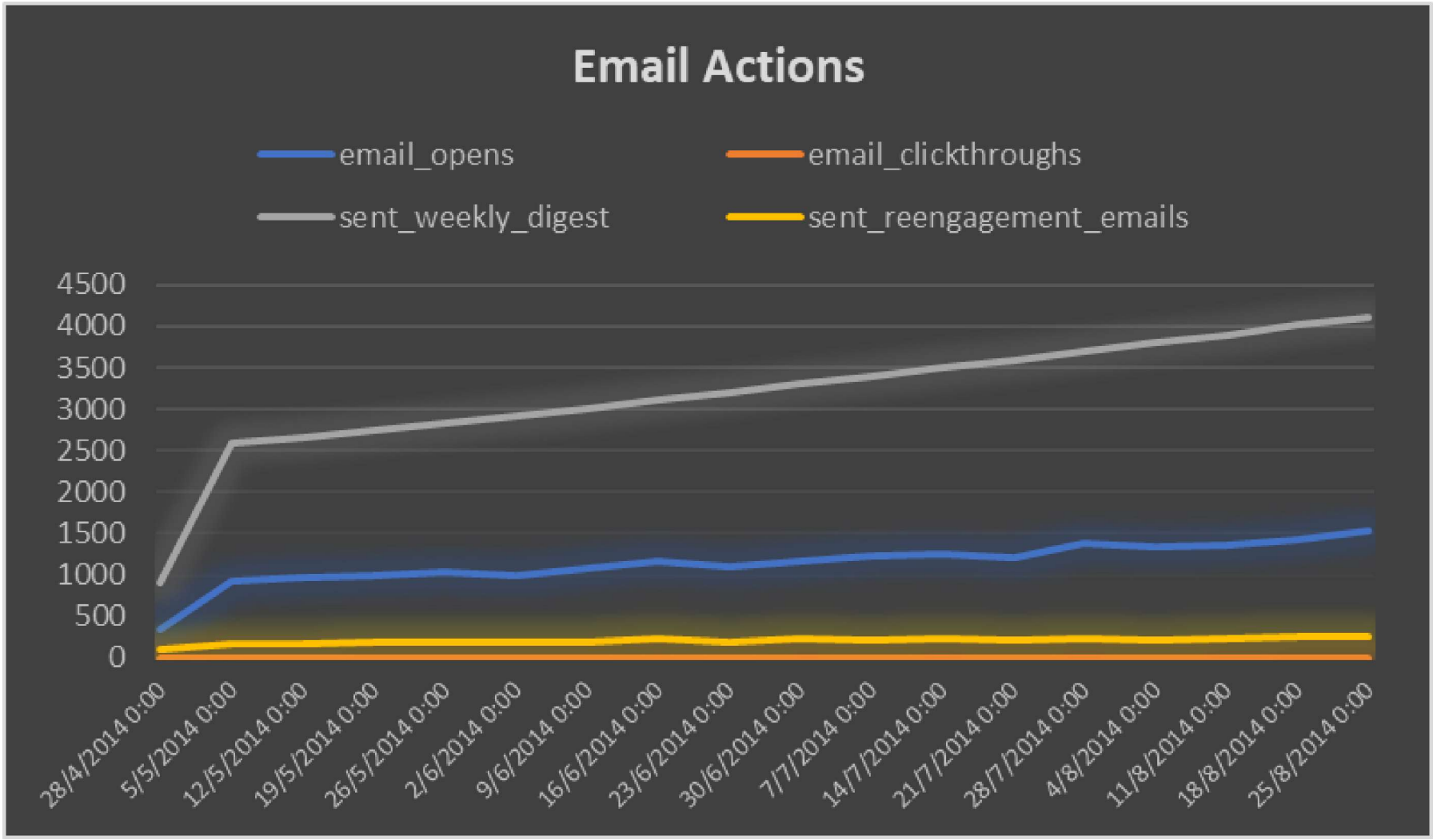
```
SELECT
DATE_TRUNC ('week', occurred_at) AS weeks,
COUNT(CASE WHEN action = 'email_open' THEN user_id END) AS email_opens,
COUNT(CASE WHEN action = 'email_clickthroughs' THEN user_id ELSE NULL END) AS email_clickthroughs,
COUNT(CASE WHEN action = 'sent_weekly_digest' THEN user_id ELSE NULL END) AS sent_weekly_digest,
COUNT(CASE WHEN action = 'sent_reengagement_email' THEN user_id ELSE NULL END) AS sent_reengagement_emails
FROM tutorial.yammer_emails
GROUP BY 1
ORDER BY 1
```

Output:




| | weeks | email_opens | email_clickthroughs | sent_weekly_digest | sent_reengagement_emails |
|---|---|---|---|---|---|
| 1 | 2014-04-28 00:00:00 | 332 | 0 | 908 | 98 |
| 2 | 2014-05-05 00:00:00 | 919 | 0 | 2602 | 164 |
| 3 | 2014-05-12 00:00:00 | 971 | 0 | 2665 | 175 |
| 4 | 2014-05-19 00:00:00 | 995 | 0 | 2733 | 179 |
| 5 | 2014-05-26 00:00:00 | 1026 | 0 | 2822 | 179 |
| 6 | 2014-06-02 00:00:00 | 993 | 0 | 2911 | 199 |
| 7 | 2014-06-09 00:00:00 | 1070 | 0 | 3003 | 190 |
| 8 | 2014-06-16 00:00:00 | 1161 | 0 | 3105 | 234 |
| 9 | 2014-06-23 00:00:00 | 1090 | 0 | 3207 | 187 |
| 10 | 2014-06-30 00:00:00 | 1168 | 0 | 3302 | 222 |
| 11 | 2014-07-07 00:00:00 | 1230 | 0 | 3399 | 214 |
| 12 | 2014-07-14 00:00:00 | 1260 | 0 | 3499 | 226 |
| 13 | 2014-07-21 00:00:00 | 1211 | 0 | 3592 | 206 |
| 14 | 2014-07-28 00:00:00 | 1386 | 0 | 3706 | 230 |
| 15 | 2014-08-04 00:00:00 | 1336 | 0 | 3793 | 206 |
| 16 | 2014-08-11 00:00:00 | 1357 | 0 | 3897 | 224 |
| 17 | 2014-08-18 00:00:00 | 1421 | 0 | 4012 | 257 |
| 18 | 2014-08-25 00:00:00 | 1533 | 0 | 4111 | 263 |

Chart:


Email Actions
email_opens
email_clickthroughs
sent_weekly_digest
sent_reengagement_emails
4500
4000
3500
3000
2500
2000
1500
1000
500
0
28/4/2014 0:00
5/5/2014 0:00
12/5/2014 0:00
19/5/2014 0:00
26/5/2014 0:00
2/6/2014 0:00
9/6/2014 0:00
16/6/2014 0:00
23/6/2014 0:00
30/6/2014 0:00
7/7/2014 0:00
14/7/2014 0:00
21/7/2014 0:00
28/7/2014 0:00
4/8/2014 0:00
11/8/2014 0:00
18/8/2014 0:00
25/8/2014 0:00

Findings:

We have identified email click-throughs as the main factor contributing to low email engagement rates. While we require input from the IT team to fully understand the overall situation, there are a few areas that we are aware of as areas of weakness:

1. Older user cohorts - 10 weeks +
2. Localized to phones
3. Email click throughs

The data does not provide a clear indication of the specific problem or the optimal solution.

However, the ability to pinpoint problem areas can greatly benefit other teams by saving them significant time in identifying the areas they should investigate.