databrickseLinks_ReasonsofDropinUserEngagement

Case Introduction

eLinks is an enterprise networking company that provides a platform for organizations to communicate and collaborate. Many companies in diverse areas of business including, for example, manufacturers, producers, suppliers, retailers, transportation, and others use eLinks. Basic membership of eLinks is free though the company provides many other value-added services at modest cost upfront or through annual subscription.

Now the management of eLinks has noticed that user engagement with the company's platform appears to have dropped in the most recent days. The management is unsure whether this is actually the case, and if so, what possible reasons for drop in user activity may be. Hence, we are going to look into this issue and advise the management team. The available data is from three tables, including USERS, EVENTS and EMAILS. Specifically, eLinks defines user activity as an engagement with its online portal, i.e., the customers (users) having made some type of server call by interacting with the company's website/web server. Such events are listed as "engagement" in the event_type column of the EVENTS table.

Problem-solving Thoughts

First it's necessary to make sure whether there indeed exists a drop in user engagement through our users' activity data. If yes, we explore the possible reasons.

The possible reasons might lie in both external and internal sides. For the external environment, we should check if some changes have happend or are happening in the whole field which leads to this decrease. Does it happen nationaly or only in some regions? Do the competitors also face this issue or not? Is there anything conducted by the competitors that makes our consumers shift to their products?

All mentioned above is important parts. But here due to data limitation, this analysis does not focus on the external aspect. In the following analysis we will employ the afforded three tables to look into this issue only from the inner side of eLinks in detail.

Daily Engagement (2014.05.01-2014.08.31)

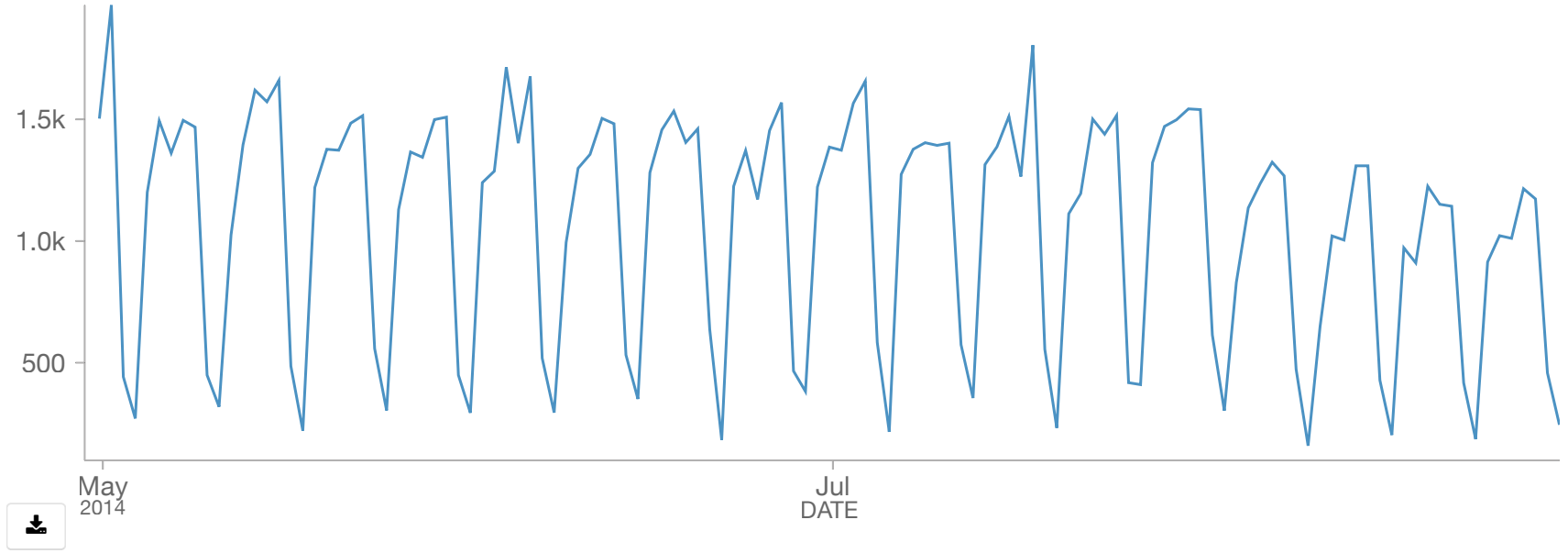
Find out the first and last date of EVENTS table

```
%sql
select max(occurred_at) as last_date_of_activity, min(occurred_at) as first_date_of_activity
from events
```

last_date_of_activity
2014-08-31 20:40:32

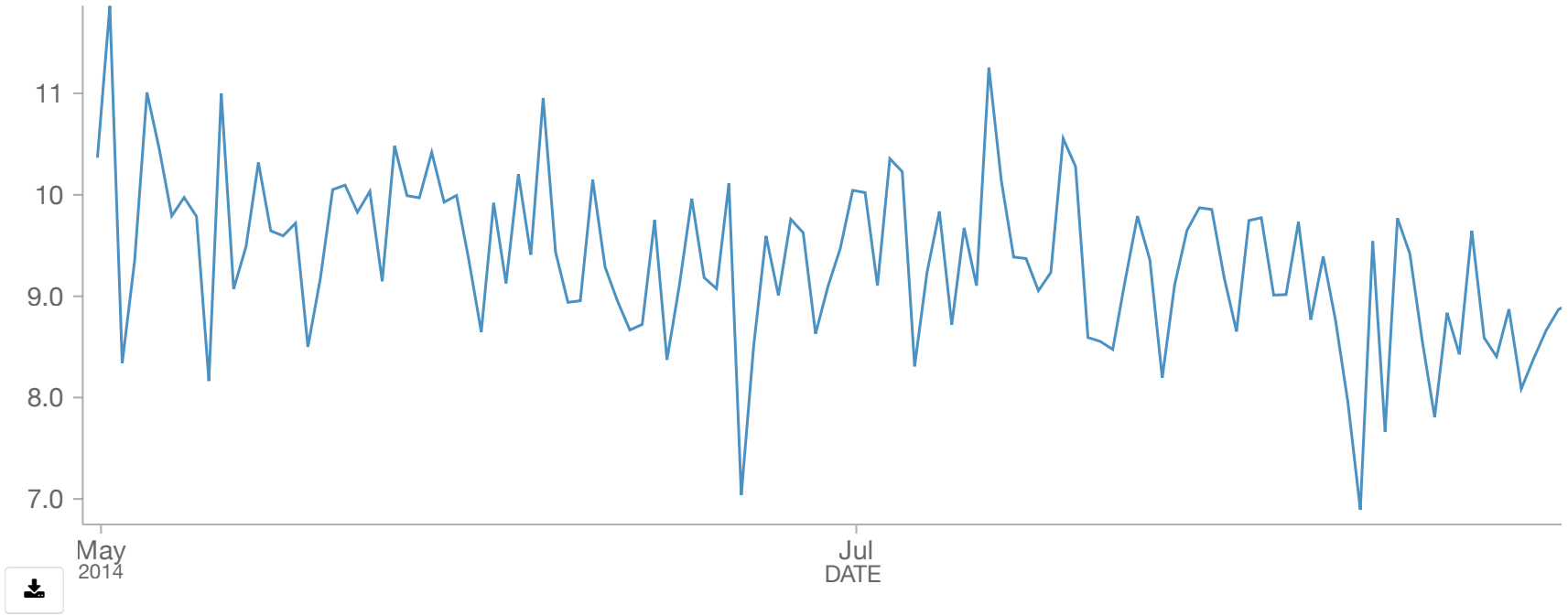
Daily total engagement

```
%sql
select date_trunc('day', occurred_at) as date, count(*) as engagement
from events
where event_type == 'engagement'
group by 1
order by 1
```



Daily engagement per person

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as enga_per_user
from events
where event_type == 'engagement'
group by 1
order by 1
```

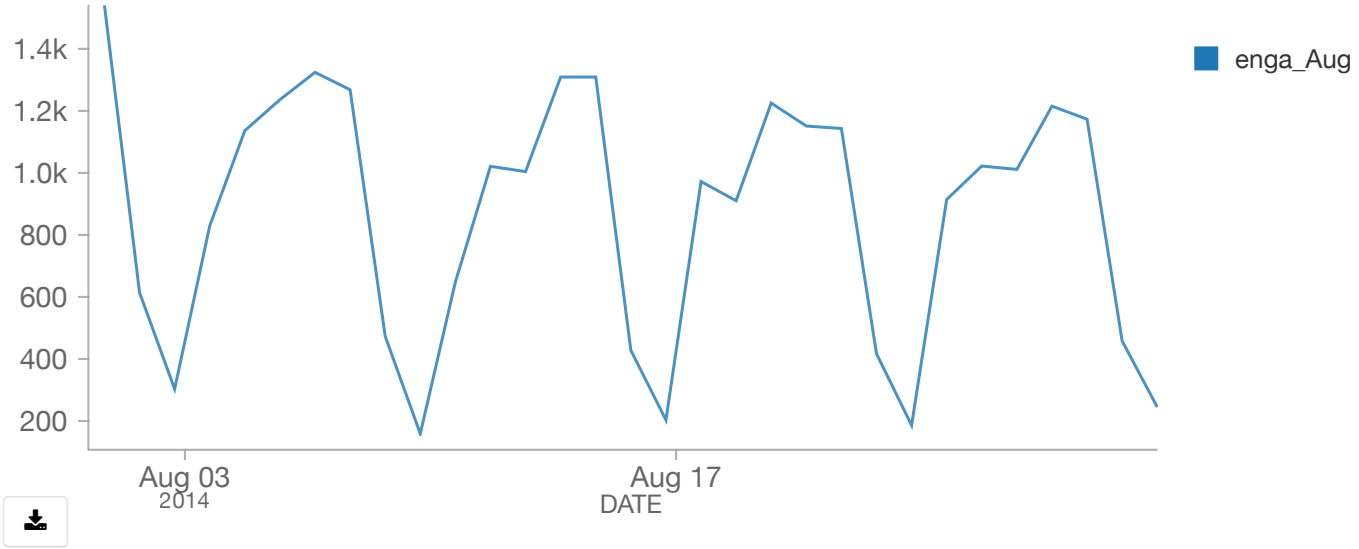


The first figure shows that there are weekly fluctuations. Generally, users had a high activity at weekdays. The highest engagement levels were usually in the midweeks, i.e., either Tuesday or Wednesday or Thursday. The number of users who had an engagement with eLinks' online portal was least on Saturday. This distribution is consistent with the fact that eLinks' customers are various companies or organizations whose business hours are at weekdays.

In addition, the engagement in the midweeks of August is slightly lower than that of the previous months. This is similar to the distribution of daily engagement per person (the second figure above). In order to figure out if there is a trend that the user engagement are dropping recently, I'm going to explore the daily number of engaged users seperately in August and other months.

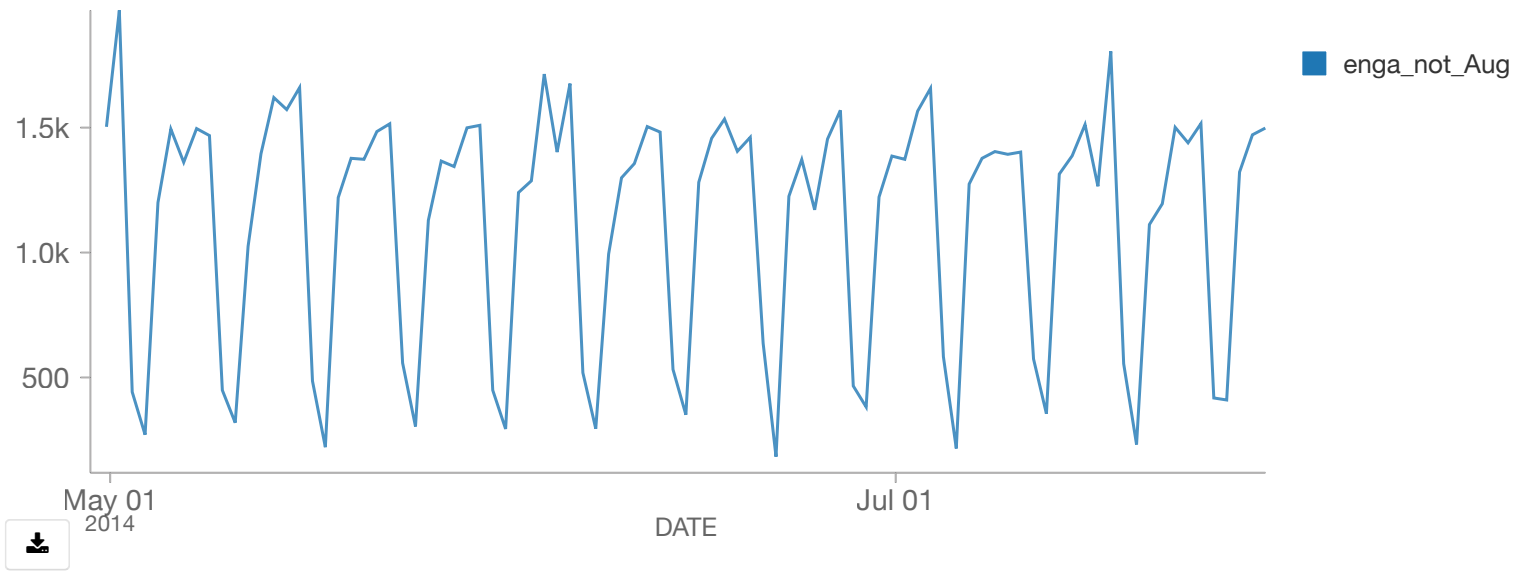
August

```
%sql
select date_trunc('day', occurred_at) as date, count(*) as enga_Aug
from events
where event_type == 'engagement'
and occurred_at >= '2014-08-01'
and occurred_at < '2014-09-01'
group by 1
order by 1;
```



May - July

```
%sql
select date_trunc('day', occurred_at) as date, count(*) as enga_not_Aug
from events
where event_type == 'engagement'
and occurred_at >= '2014-05-01'
and occurred_at < '2014-07-31'
group by 1
order by 1
```



From the above two figures, we can identify clearly that the daily active volumes on weekdays of August is around 1000-1200 while during May-July the daily engagement on weekdays is always around 1500. What's more, the activity level on the weekends of August also decreased from around 500 in May-July to around 200. As a result, we conclude that it is actually the case that user engagement with the company's platform has dropped recently.

Possible reasons

Observing the data on hand, we come up with three possible reasons as follows:

- The net customer churn in August may be more serious than other months.
- There is a possibility that users don't like some specific features (or events) of the platform. So they decrease their activities on this platform.
- There may be some technical problems with the products, such as uncompatible with some mobile devices, broken website, etc. It will affect user experience significantly, which may resulst in a drop of user engagement.

Reason 1: Net Customer Churn

In general, there are new users who start to activate their accounts in each month. At the same time some existing users would stop their activties on the platform in every month. So I focused on the net customer churn. The customer churn of a month is the number of users who engage actively in previous months but don't have any activity in current month. The new engagement of a month refers to the number of new engaged users who creat and activate their accounts in the current month. Therefore, the net customer churn of a month means the customer churn minus the new engagement.

If the net customer churn of August is more than that of other months, the engagement of August may be less than that of other months. So in the following, I calculated the net customer churn of each month.

Net customer churn of August

```
%sql
select customer_churn_in_Aug, new_activated_in_Aug, customer_churn_in_Aug - new_activated_in_Aug as
net_customer_churn_in_Aug
from (select
  (select count(*)
    from (
          select distinct user_id
          from events
          where occurred_at between '2014-05-01' and '2014-07-31'
            and event_type == 'engagement'
            and user_id not in (
                select distinct user_id
                from events
                where occurred_at between '2014-08-01' and '2014-08-31'
                  and event_type == 'engagement'))) as customer_churn_in_Aug,
    (select count(user_id)
      from users
      where created_at between '2014-08-01' and '2014-08-31'
       and state == 'active') as new_activated_in_Aug
   );
```

customer_churn_in_Aug	new_a
3224	1013

Net customer churn of July

```
%sql
select customer_churn_in_Jul, new_activated_in_Jul, customer_churn_in_Jul - new_activated_in_Jul as
net_customer_churn_in_Jul
from (select
  (select count(*)
    from (
          select distinct user_id
          from events
          where occurred_at between '2014-05-01' and '2014-06-31'
            and event_type == 'engagement'
            and user_id not in (
                select distinct user_id
                from events
                where occurred_at between '2014-07-01' and '2014-07-31'
                  and event_type == 'engagement'))) as customer_churn_in_Jul,
    (select count(user_id)
      from users
      where created_at between '2014-07-01' and '2014-07-31'
       and state == 'active') as new_activated_in_Jul
  );
```

customer_churn_in_Jul	new_a
2050	953

Net customer churn of June

```
%sql
select customer_churn_in_Jun, new_activated_in_Jun, customer_churn_in_Jun - new_activated_in_Jun as
net_customer_churn_in_Jun
from (select
  (select count(*)
    from (
          select distinct user_id
          from events
          where occurred_at between '2014-05-01' and '2014-05-31'
            and event_type == 'engagement'
            and user_id not in (
                select distinct user_id
                from events
                where occurred_at between '2014-06-01' and '2014-06-31'
                  and event_type == 'engagement'))) as customer_churn_in_Jun,
    (select count(user_id)
      from users
      where created_at between '2014-06-01' and '2014-06-31'
       and state == 'active') as new_activated_in_Jun
   );
```

customer_churn_in_Jun	new_a
1024	873

The above calculations show that the net customer churn of August was, as I expected, indeed more seriou than other months. The new activated users of August were more than 1000, slightly higer than June and July. But its customer churn had a hugh growth compared with the previous months, reaching more than 3200 users. As a result, the net customer churn of August was up to 2000 users, much higher than that of previous two months. It could lead to the drop of user engagement in August.

What's more, not only the total engagement but the engagement per person in August also declined significantly. Then why did so many users leave this platform in August? Why did the daily activity of per user who still stay on this platform also go down? To answer these questions, a futher exploration about its products is required.

Reason 2: Specific Feature

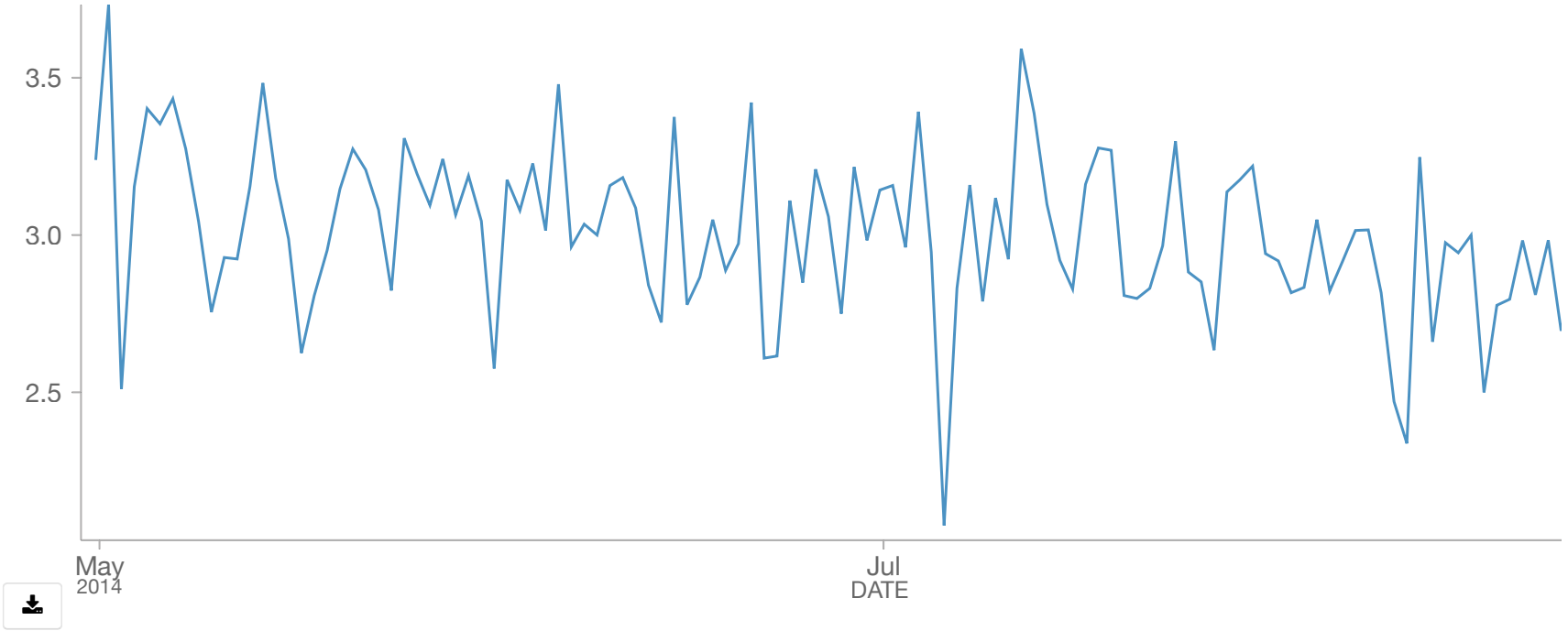
On one side, the reason why users give up a product is likely to be the fact that they are not satisfied with some specific features of that product. But the data on this aspect is absent. Considering that we have the number of times a specific event performed by users, we could aquire its change trend from 2014-05-01 to 2014-08-31.

Specifically speaking, we focused on 8 specific events relating to user engagement according to the attribute of 'event_name' in Table EVENTS. We could observe whether the daily number of times each of these 8 specific events performed by users drops in August. Based on this, we could know that which feature of the product is used less than before.

Besides, the net customer churn was more than zero which means the total users in August were less than that in previous months. To eliminate the impact of falling headcount, we concentrated on the daily number of times a specific event engaged by per person.

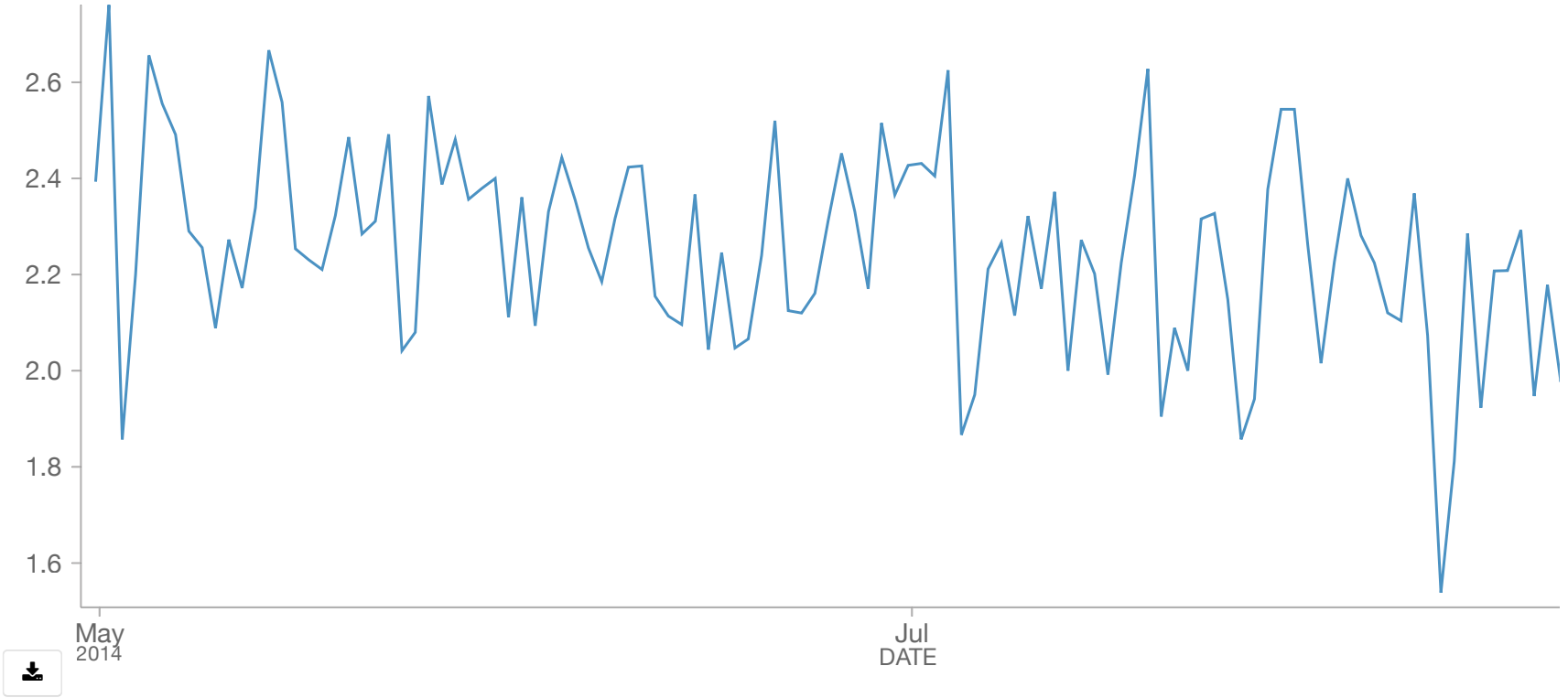
Event 1: Home_page

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
home_page_per_person_engaged
from events
where event_name == 'home_page'
group by 1
order by 1
```



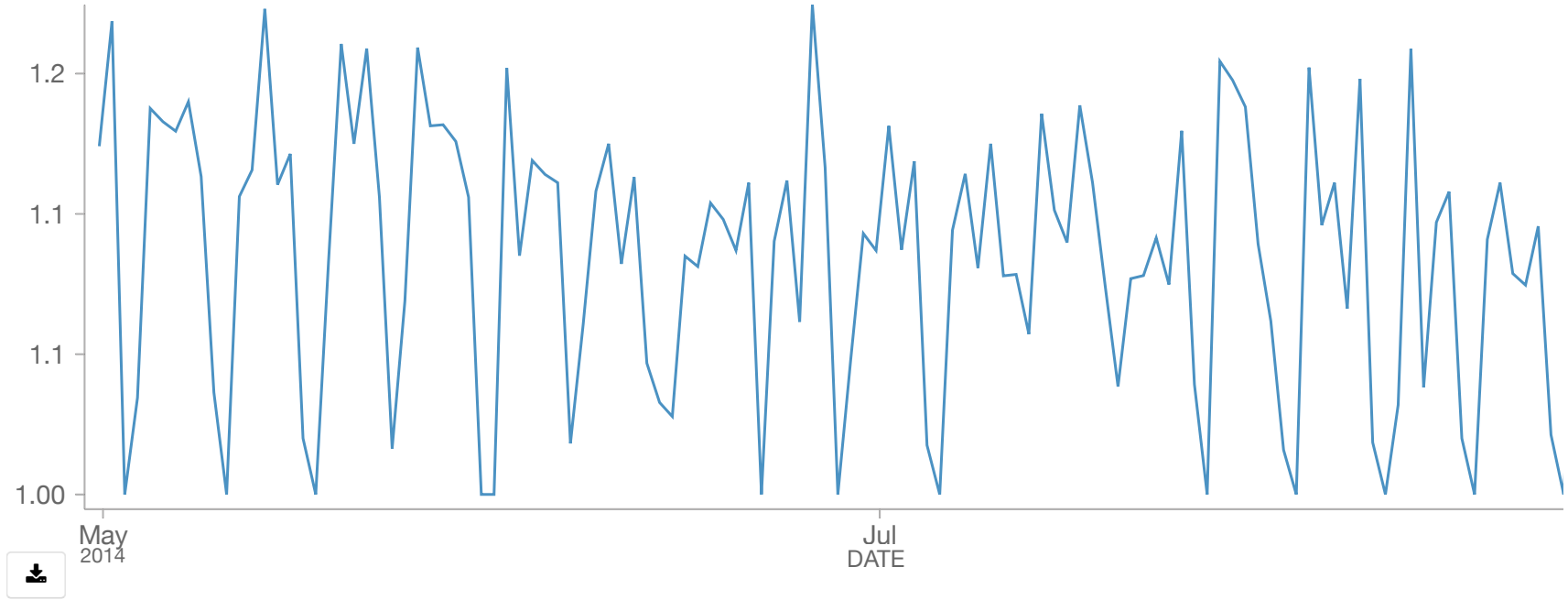
Event 2: Like_message

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
like_message_per_person_engaged
from events
where event_name == 'like_message'
group by 1
order by 1
```



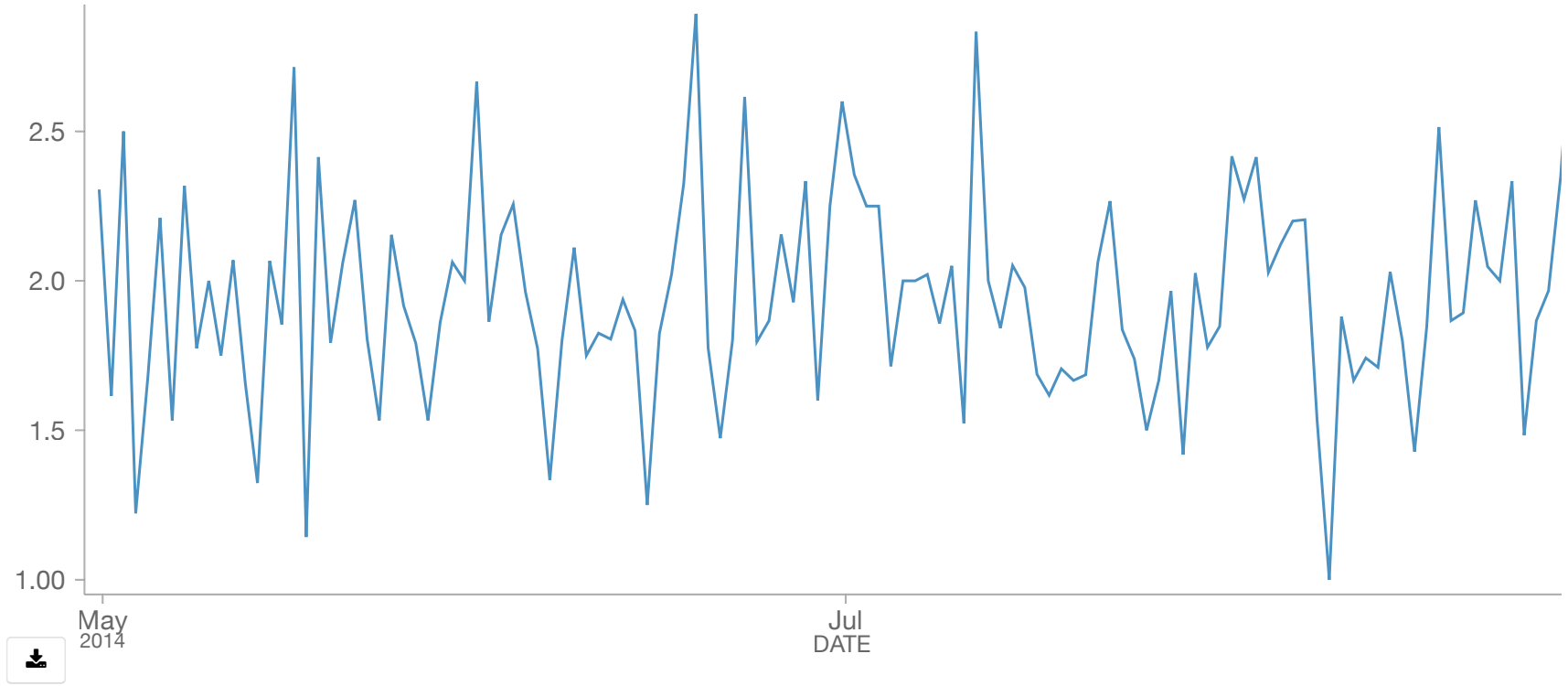
Event 3: Login

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as login_per_person_engaged
from events
where event_name == 'login'
group by 1
order by 1
```



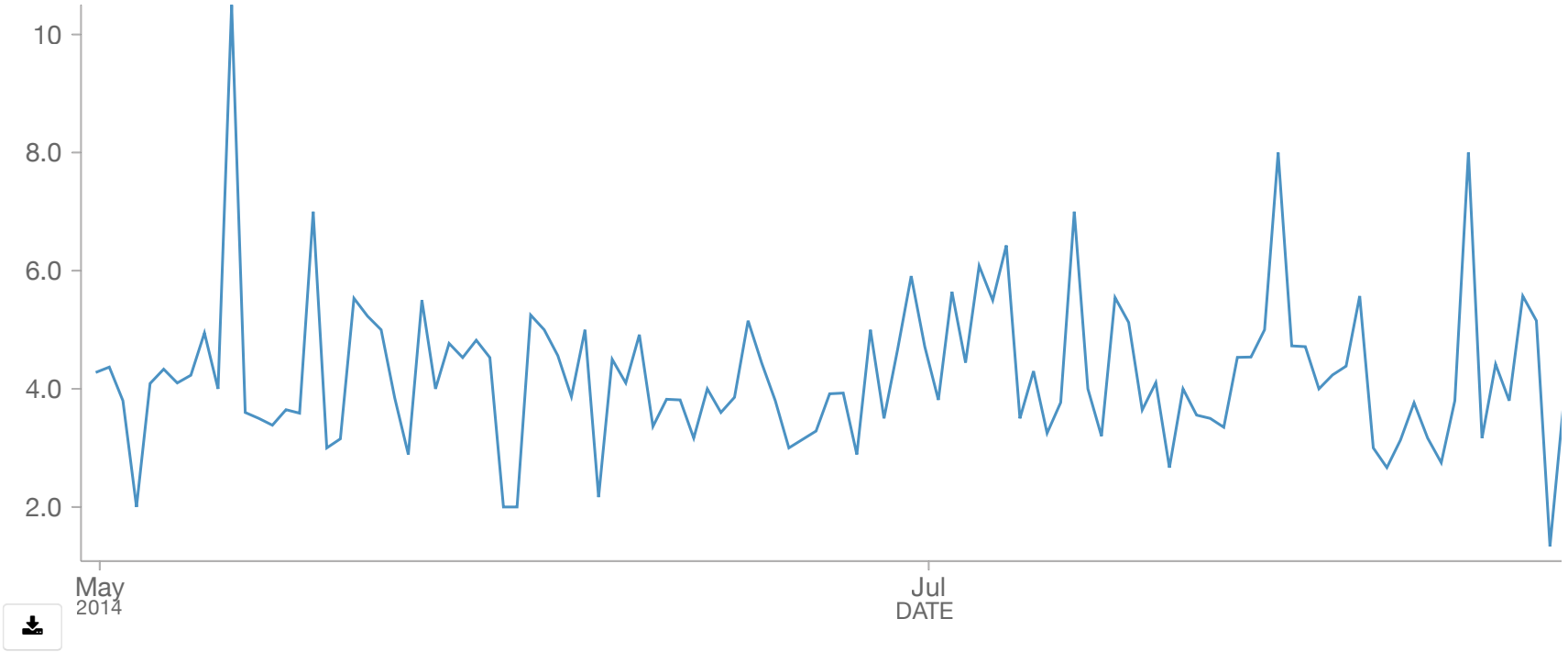
Event 4: Search_autocomplete

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
srch_autocompl_per_person_engaged
from events
where event_name == 'search_autocomplete'
group by 1
order by 1
```



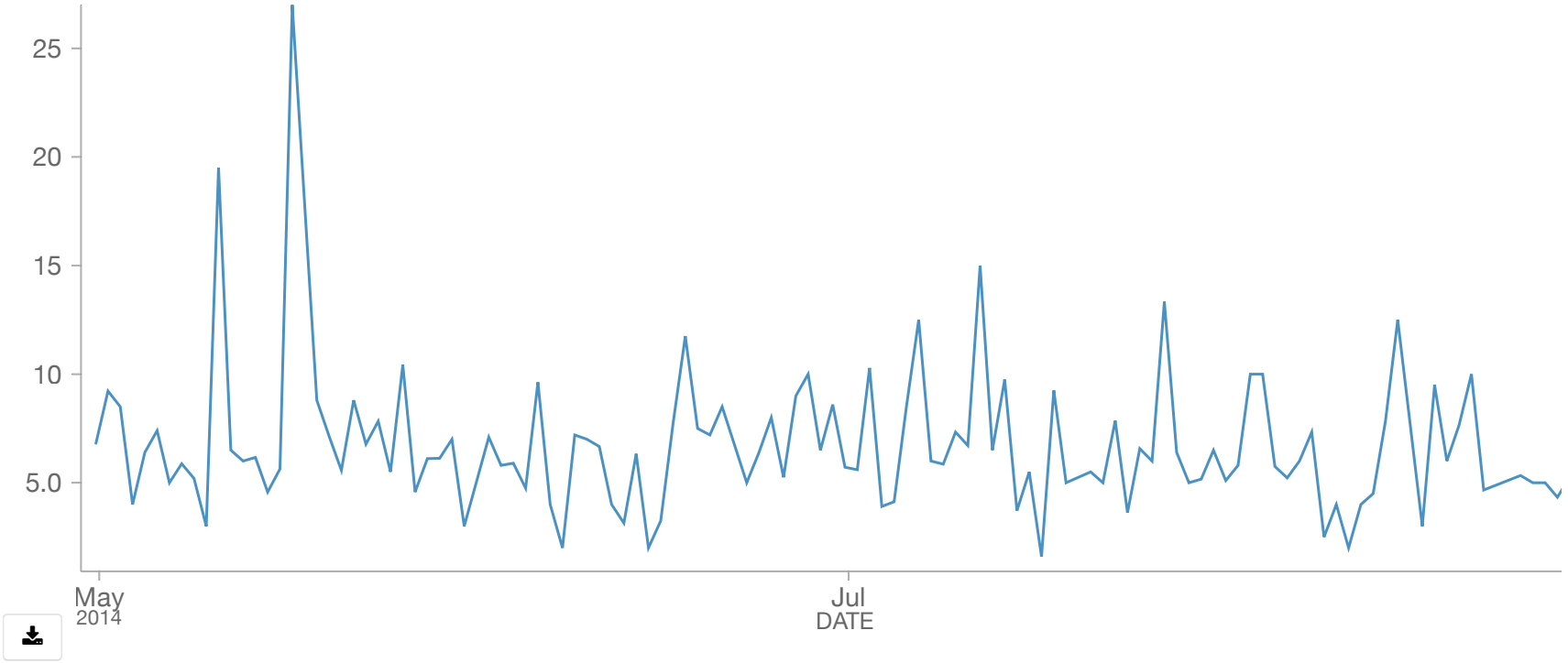
Event 5: Search_run

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
search_run_per_person_engaged
from events
where event_name == 'search_run'
group by 1
order by 1
```



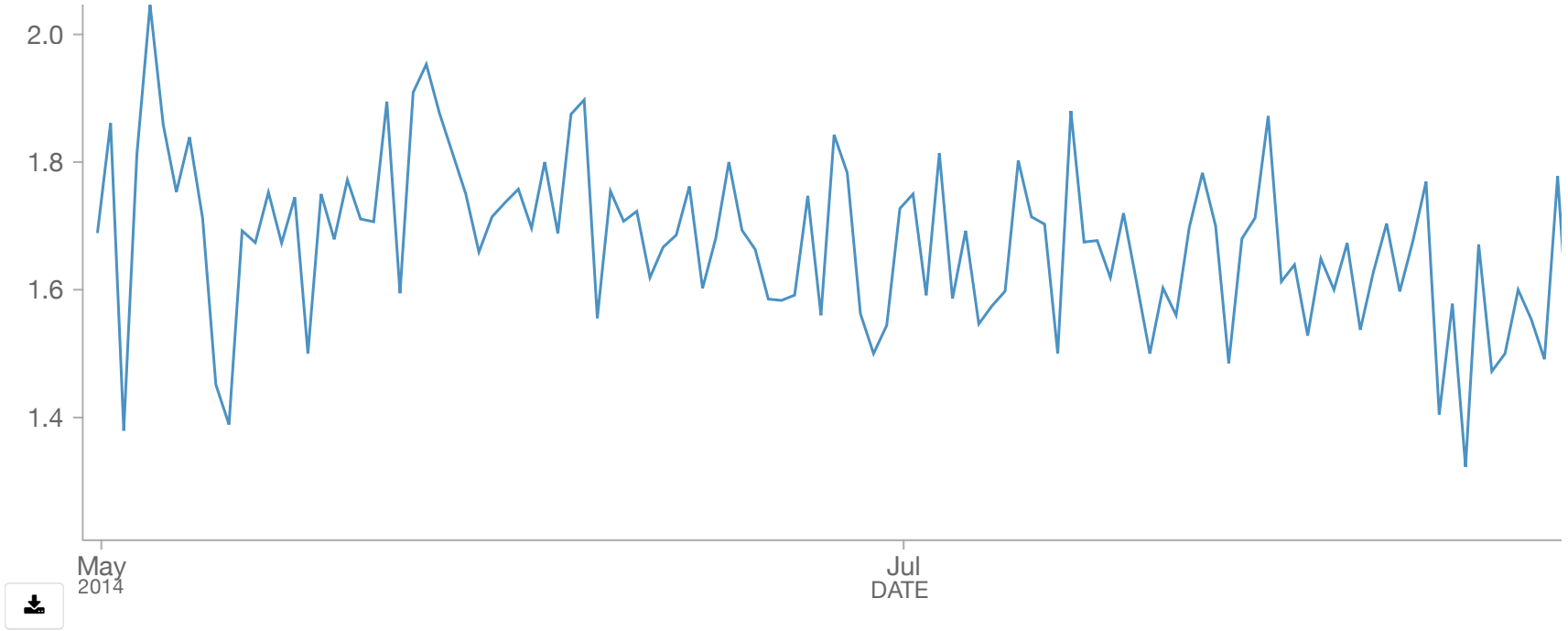
Event 6: Search_click_result_X

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
srch_click_resul_X_per_person_engaged
from events
where event_name in ('search_click_result_1', 'search_click_result_2', 'search_click_result_3',
'search_click_result_4', 'search_click_result_5', 'search_click_result_6', 'search_click_result_7',
'search_click_result_8', 'search_click_result_9')
group by 1
order by 1
```



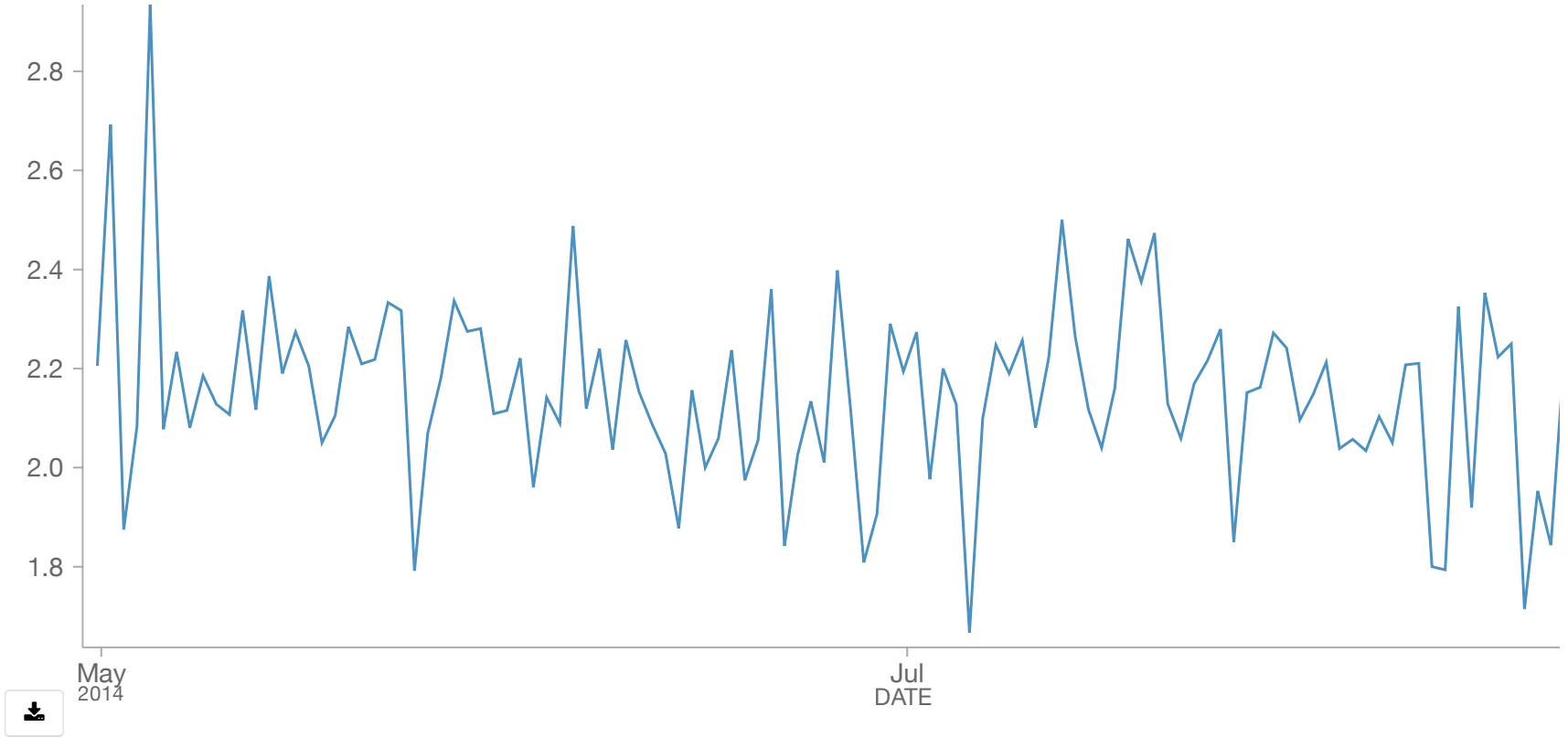
Event 7: Send_message

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
send_message_per_person_engaged
from events
where event_name == 'send_message'
group by 1
order by 1
```



Event 8: View_inbox

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as
view_inbox_per_person_engaged
from events
where event_name == 'view_inbox'
group by 1
order by 1
```


2.8
2.6
2.4
2.2
2.0
1.8
May
2014
Jul
DATE

The above analysis shows that in August the frequency of the three specific events, i.e., home_page loading, like_message clicking and sending message, happened less than before while there is no any difference about the other five events in August compared with the previous months. The drop of these three events engagement may partly explain the fall of user activity in August.

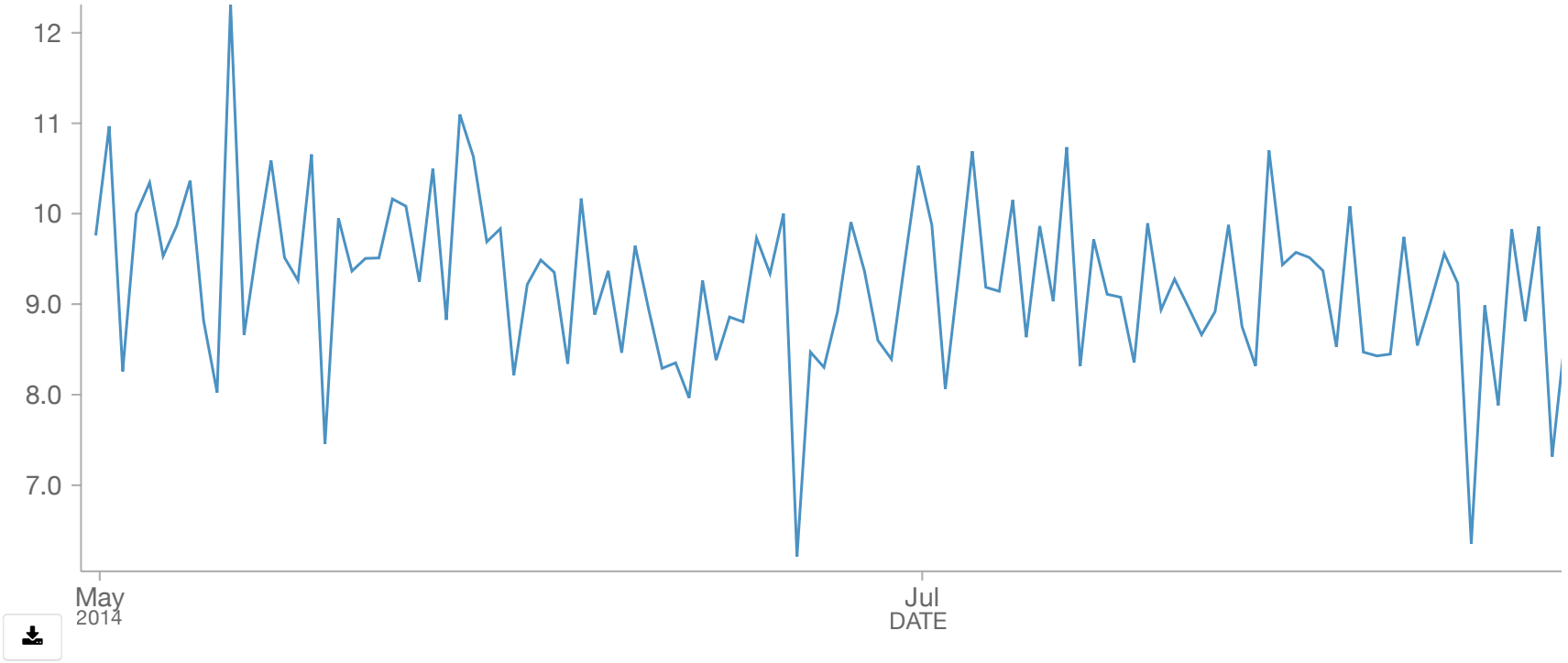
However, this infomation is not sufficient to answer whether the declining engagement is resulted from users unsatisfy with these three features of this platform. Because other than unsatisfication, many other things also may cause the decrease. For example, if some technical failures about the three features happen at some time in August, users cannot use them and then the users' engagement with these three features would go down.

Therefore, if we want to figure out the detailed reasons we need more data and a further research. For instance, the company could conduct a survey which focuses on the users attitutes about these three features. Or we could cooperate with IT department and investigate if there are technical failures of these three features in August.

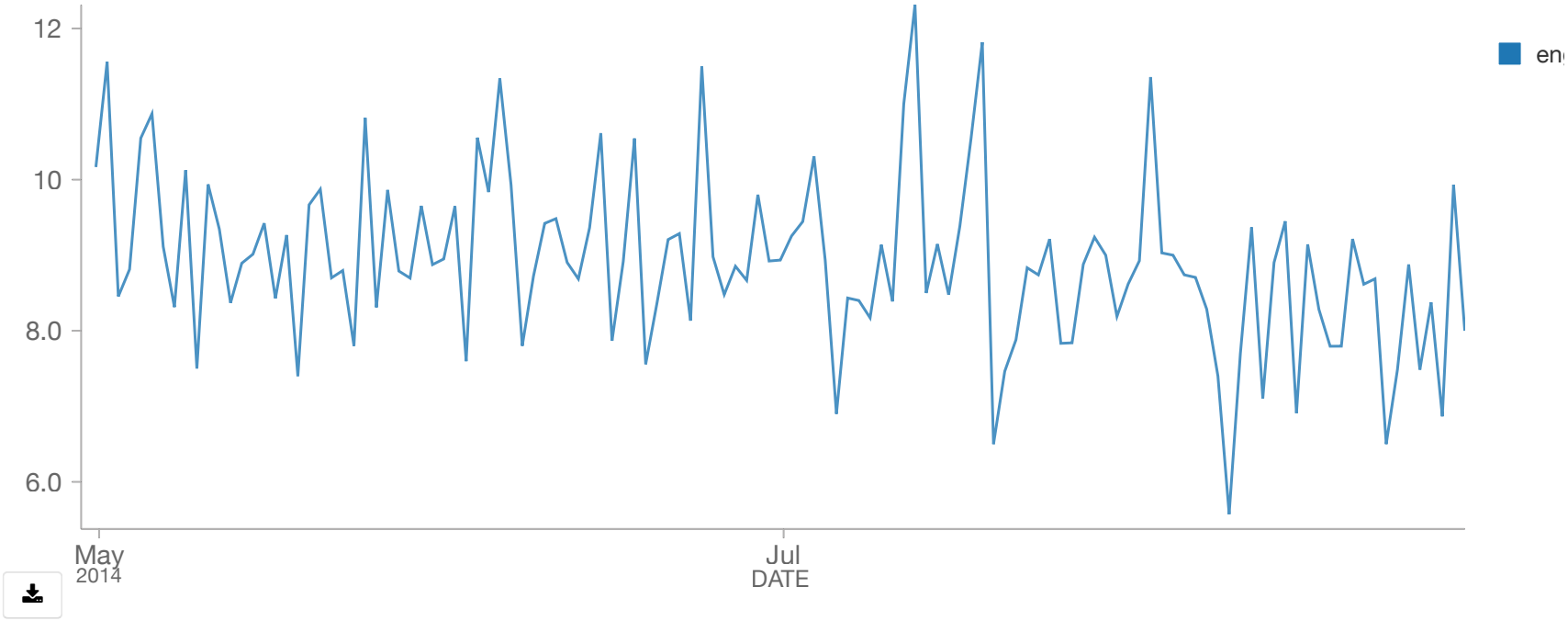
Reason 3: Technical Failure

On the other side, users giving up a product may be partly due to their technical instability. We noticed that this platform has two versions, of which one is for PC/Mac and the other is for mobile devices like cell phones or tablets. Then is it possible that there is a difference between the change trend of daily engagement of the two different versions of the platform? If the engagement with one of the two versions remains stable but the other declines, it may means there is a technical failure on the latter one which makes it impossible for people to use it normally. So to figure out this point, we drew the daily engagement per person on the PC version and on the mobile version according to the attribute of "device" in Table EVENTS.

```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as enga_with_PC_per_person
from events
where event_type == 'engagement'
  and device in ('hp pavilion desktop', 'dell inspiron desktop', 'acer aspire desktop', 'windows surface',
'macbook pro', 'macbook air', 'mac mini', 'lenovo thinkpad', 'dell inspiron notebook', 'asus chromebook',
'acer aspire notebook')
group by 1
order by 1
```




```
%sql
select date_trunc('day', occurred_at) as date, count(*)/count(distinct user_id) as enga_with_mobile_per_person
from events
where event_type == 'engagement'
  and device not in ('hp pavilion desktop', 'dell inspiron desktop', 'acer aspire desktop', 'windows surface',
'macbook pro', 'macbook air', 'mac mini', 'lenovo thinkpad', 'dell inspiron notebook', 'asus chromebook',
'acer aspire notebook')
group by 1
order by 1
```



The above two pictures illustrate that the daily engagement per person with the PC version platform remained stable from May to August while the engagement with the mobile version platform had a significant decrease in August compared with the previous months. It means that the mobile version maybe has some technical failures in August. But what we have now is not enouth to get that conclusion. We need more information to explore the underliying reasons. We could look up for the data to see if there are any technical problems about the mobile version happening in August, like APP crashing, ect.

Conclusions

On the basis of existing data, we first analyzed the change trend of users' daily total engagement as well as the daily engagement per user from May to August. Through this we confirmed that user engagement started to drop from the beginning of August. The we came up with three possible reasons:

1. The net customer churn in August may be more serious than other months.

2. There is a possibility that users don't like some specific features (or events) of the platform. So they decrease their activities on this platform.

3. There may be some technical problems with the products, such as uncompatible with some mobile devices, broken website, etc. It will affect user experience significantly, which may resulst in a drop of user engagement.

After relevant calculations and explorations, we could conclude that the **net customer churn** in August was indeed **more serious** than other months, which is the direct reason of engagement dropping. Then we conducted a further investigation and found out users decreased their engagement on some specific aspects. Specifically speaking, users' engagement in home_page loading, like_message clicking and sending message declined in August. And users' activity happenning on the mobile version platform also dropped in August. However, we **cannot** reach a conclusion that these specific engagement drop is caused by users not satisfy with some features of the platform or some technical failures of the platform. In order to figure out what makes it happen, we need more data to do further research, such as cooperation with IT department to see if there are any technical failure records in August, or a survey focusing on users' attitudes to the platform, etc.