Identifying Unreliable Data + Nulls

Exercise 1: Using any methods you like determine if you can trust this events table.

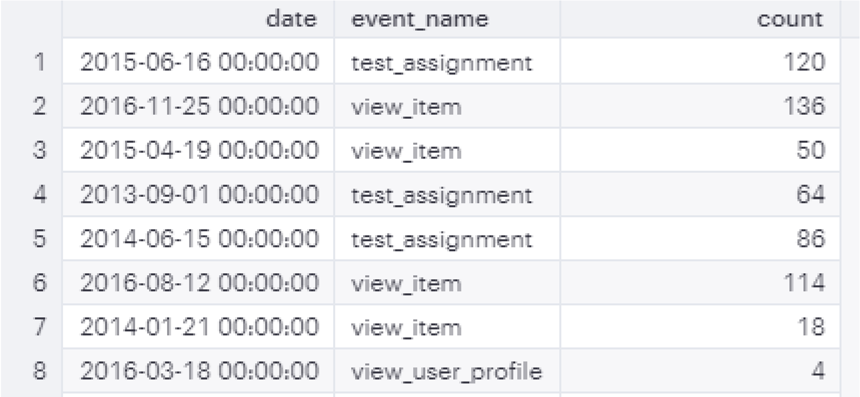
```
-- some days are missing!

select date (event_time) as date,
       count(*) as rows
from dsv1069.events_201701
group by 1
```

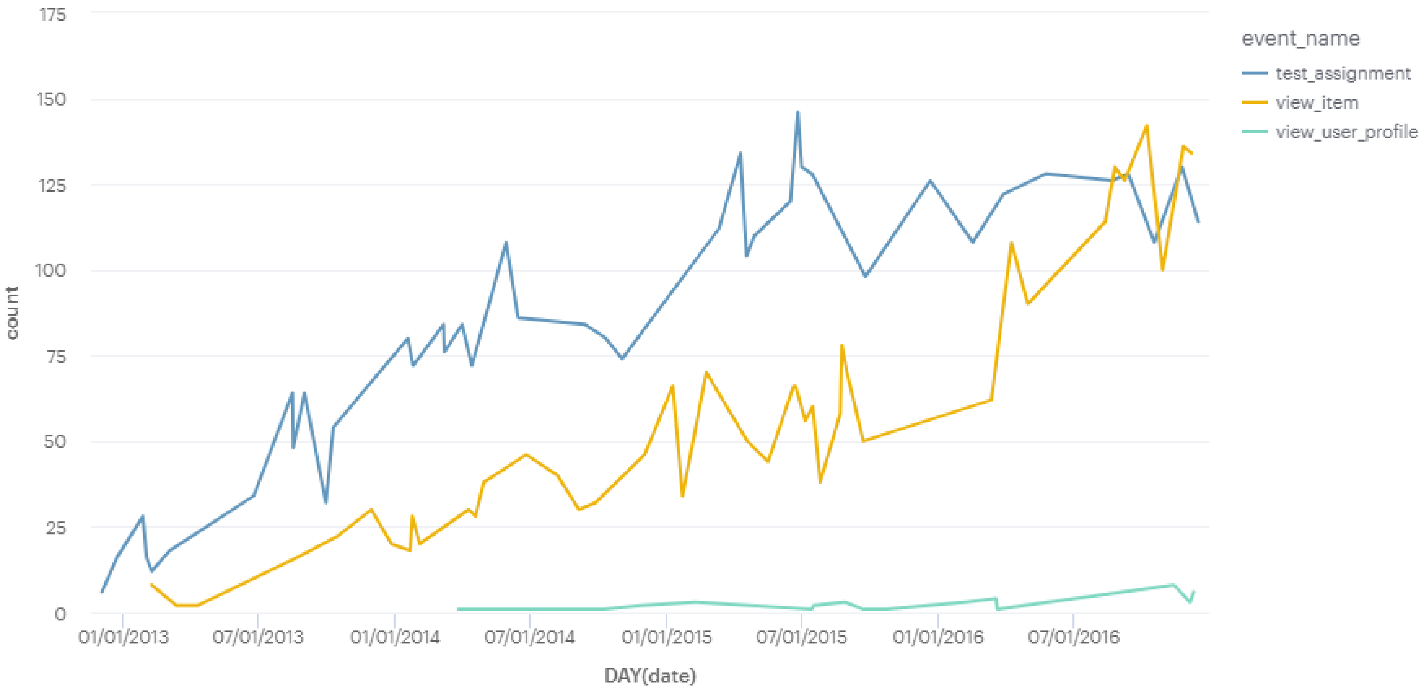
Exercise 2: Using any methods you like, determine if you can you trust this events table. (HINT: When did we start recording events on mobile)

```
-- view_user_profile started late!

select date (event_time) as date,
             event_name,
             count(*)
from dsv1069.events_ex2
group by 1,2
```

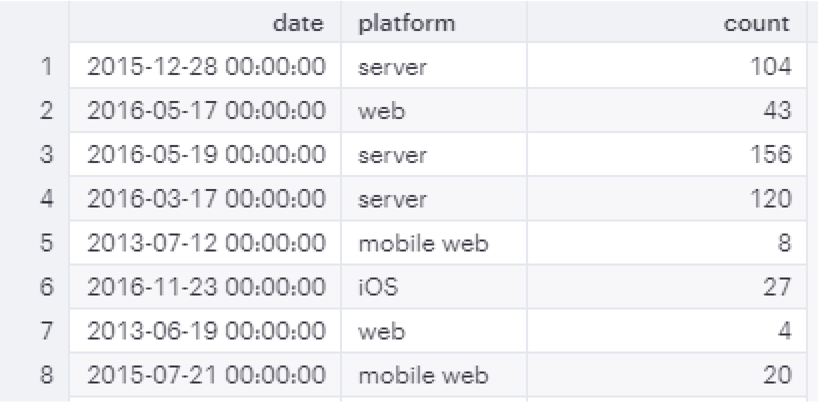


	date	event_name	count
1	2015-06-16 00:00:00	test_assignment	120
2	2016-11-25 00:00:00	view_item	136
3	2015-04-19 00:00:00	view_item	50
4	2013-09-01 00:00:00	test_assignment	64
5	2014-06-15 00:00:00	test_assignment	86
6	2016-08-12 00:00:00	view_item	114
7	2014-01-21 00:00:00	view_item	18
8	2016-03-18 00:00:00	view_user_profile	4

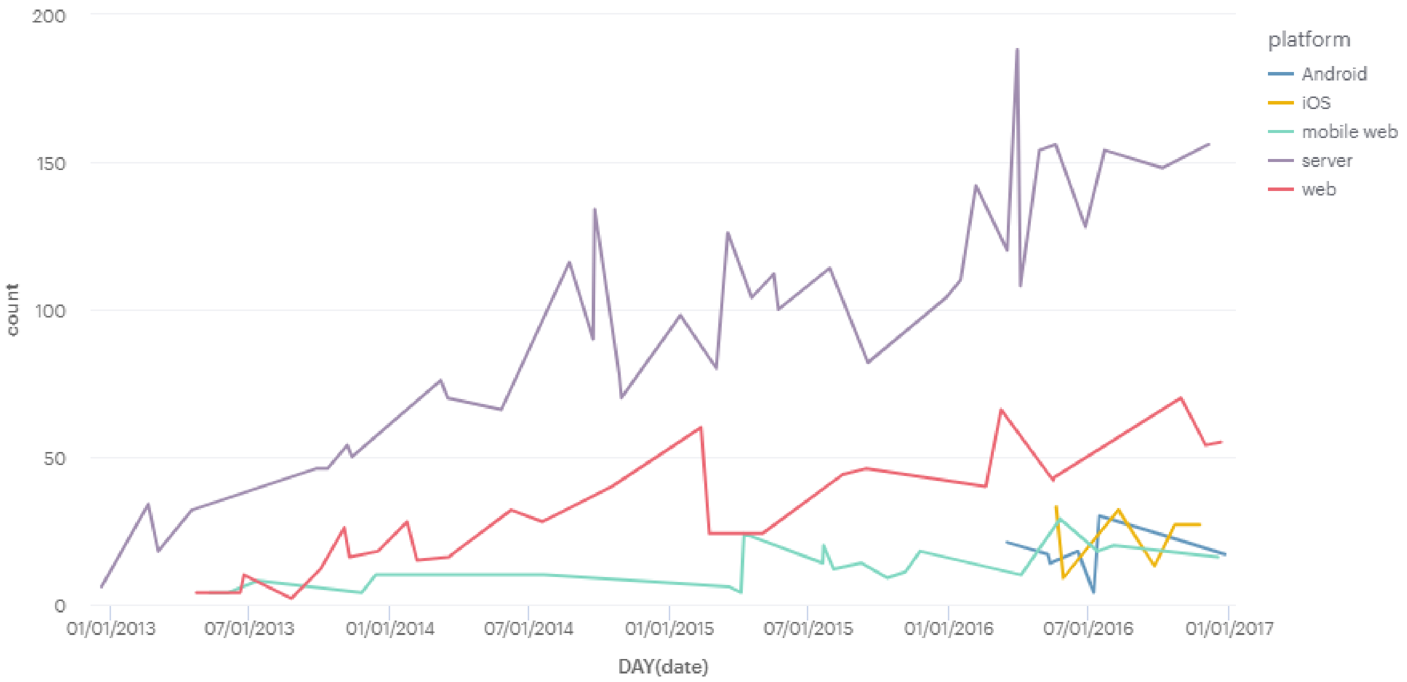


```sql
-- mobile logging hasn't started until recently!

select date (event_time) as date,
       platform,
       count(*)
from dsv1069.events_ex2
group by 1,2
```



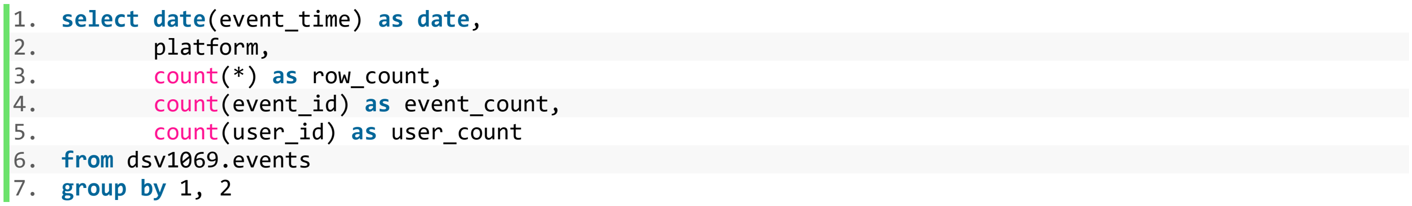
	date	platform	count
1	2015-12-28 00:00:00	server	104
2	2016-05-17 00:00:00	web	43
3	2016-05-19 00:00:00	server	156
4	2016-03-17 00:00:00	server	120
5	2013-07-12 00:00:00	mobile web	8
6	2016-11-23 00:00:00	iOS	27
7	2013-06-19 00:00:00	web	4
8	2015-07-21 00:00:00	mobile web	20



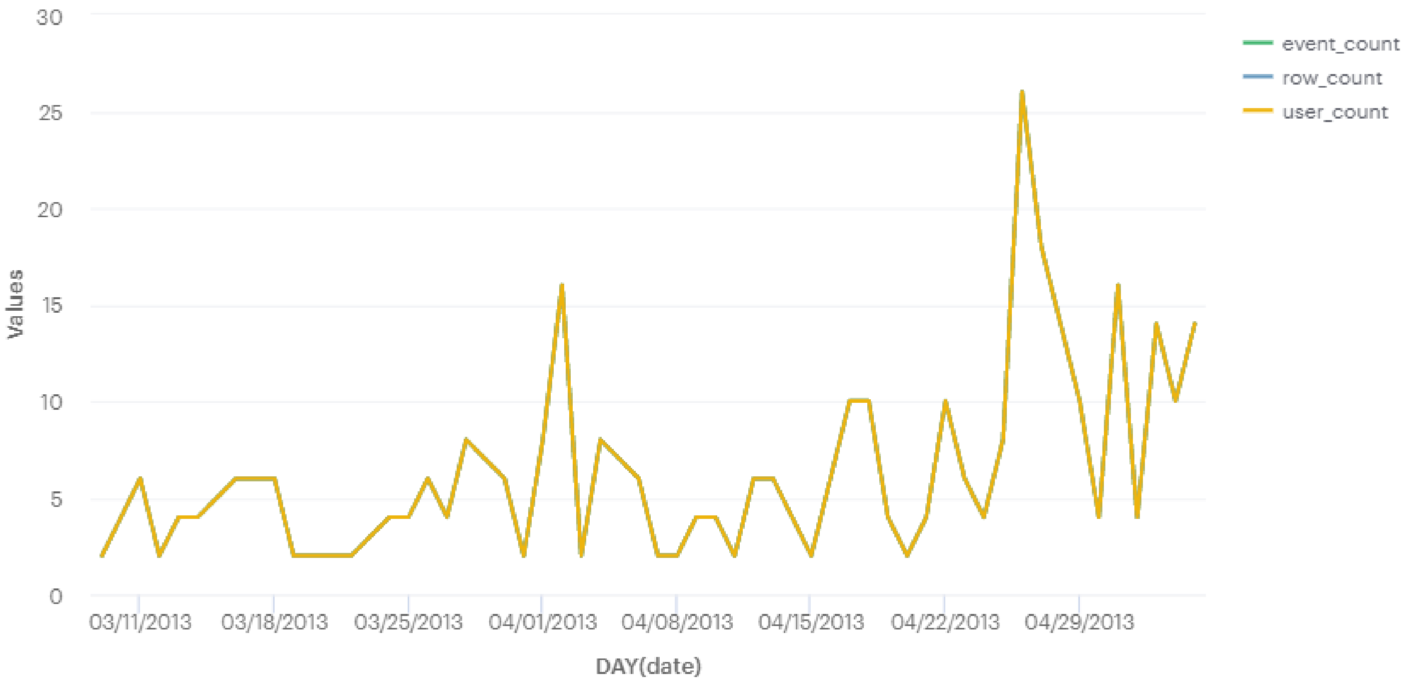
Exercise 3: Imagine that you need to count item views by day. You found this table item_views_by_category_temp - should you use it to answer your question? NO.

```sql
select date(event_time) as date,
       count(*) as row_count
from dsv1069.events
group by 1
```

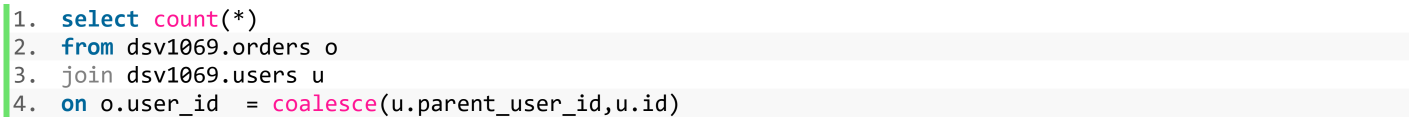
Exercise 4: Using any methods you like, decide if this table is ready to be used as a source of truth. YES.



```sql
select date(event_time) as date,
       platform,
       count(*) as row_count,
       count(event_id) as event_count,
       count(user_id) as user_count
from dsv1069.events
group by 1, 2
```



Exercise 5: Is this the right way to join orders to users? Is this the right way this join.



```sql
select count(*)
from dsv1069.orders o
join dsv1069.users u
on o.user_id  = coalesce(u.parent_user_id,u.id)
```