Mode Analytics Case Study 2: Search Functionality

Problem Statement

The product team at Yammer is interested in determining whether they should work on search functionality and, if so, how they should modify it

Problem Statement 1: Is there a problem with search functionality?

Problem Statement 2: If so, how to fix it?

Solution Steps:

Step 1: Identification of user session

SQL Query:

```
--SQL code to identify each user session. Sessions are tagged different if the time difference between their events is >10 minutes

select sub.*,
        case when sub.time is null or sub.time >= interval '10 minute' then sub.sessionid
             else 0 end as "sessionno"
from(
    select user_id, occurred_at,
      occurred_at - LAG(occurred_at, 1) over (partition by user_id order by occurred_at) as "time",
      event_name, row_number() OVER () AS "sessionid"
    from tutorial.yammer_events
    where event_type='engagement'
    order by user_id, occurred_at
   ) sub
```

Step 1: Output

```
36  from(
37      select user_id, occurred_at,
38          occurred_at - LAG(occurred_at, 1) over (partition by user_id order by occurred_at) as "tir
39          event_name, row_number() OVER () AS "sessionid"
40
```

Ready

Export Copy Chart Pivot

100 rows returned

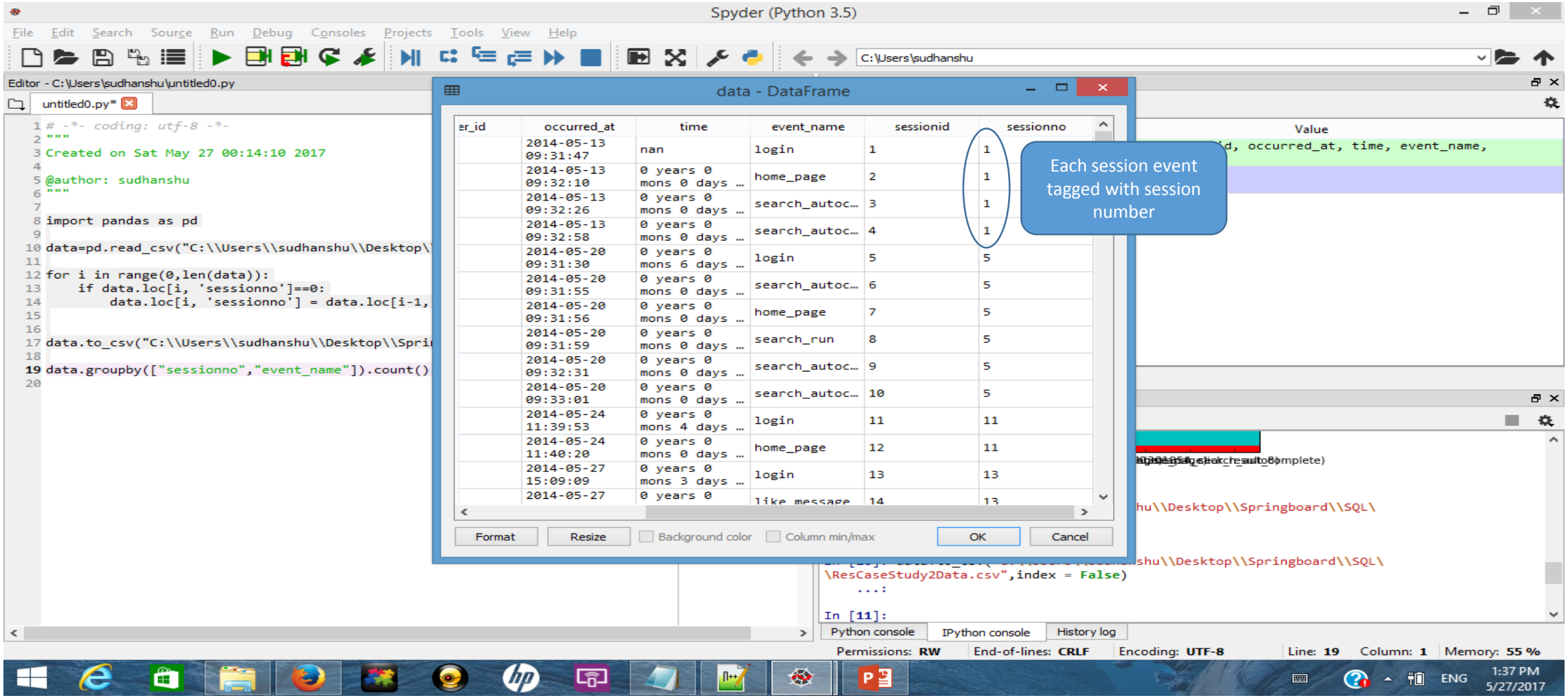
	occurred_at	time	event_name	sessionid	sessionno
1	2014-05-13 09:31:47		login	1	1
2	2014-05-13 09:32:10	0 years 0 mons 0 days 0 hours 0 mins 23.00 secs	home_page	2	0
3	2014-05-13 09:32:26	0 years 0 mons 0 days 0 hours 0 mins 16.00 secs	search_autocomplete	3	0
4	2014-05-13 09:32:58	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	search_autocomplete	4	0
5	2014-05-20 09:31:30	0 years 0 mons 6 days 23 hours 58 mins 32.00 secs	login	5	5
6	2014-05-20 09:31:55	0 years 0 mons 0 days 0 hours 0 mins 25.00 secs	search_autocomplete	6	0
7	2014-05-20 09:31:56	0 years 0 mons 0 days 0 hours 0 mins 1.00 secs	home_page	7	0
8	2014-05-20 09:31:59	0 years 0 mons 0 days 0 hours 0 mins 3.00 secs	search_run	8	0

Uniquely identified session

Step 2: Tagging each session event with session number

```
-- Python Code to tag each event against a session

import pandas as pd
#Reading in Data
data=pd.read_csv("C:\\Users\\sudhanshu\\Desktop\\Springboard\\SQL\\CaseStudy2Data.csv")

#Looping over data to tag event names with respective user sessions

for i in range(0,len(data)):
    if data.loc[i, 'sessionno']==0:
        data.loc[i, 'sessionno'] = data.loc[i-1, 'sessionno']

#Exporting Data
data.to_csv("C:\\Users\\sudhanshu\\Desktop\\Springboard\\SQL\\ResCaseStudy2Data.csv",index = False)
```

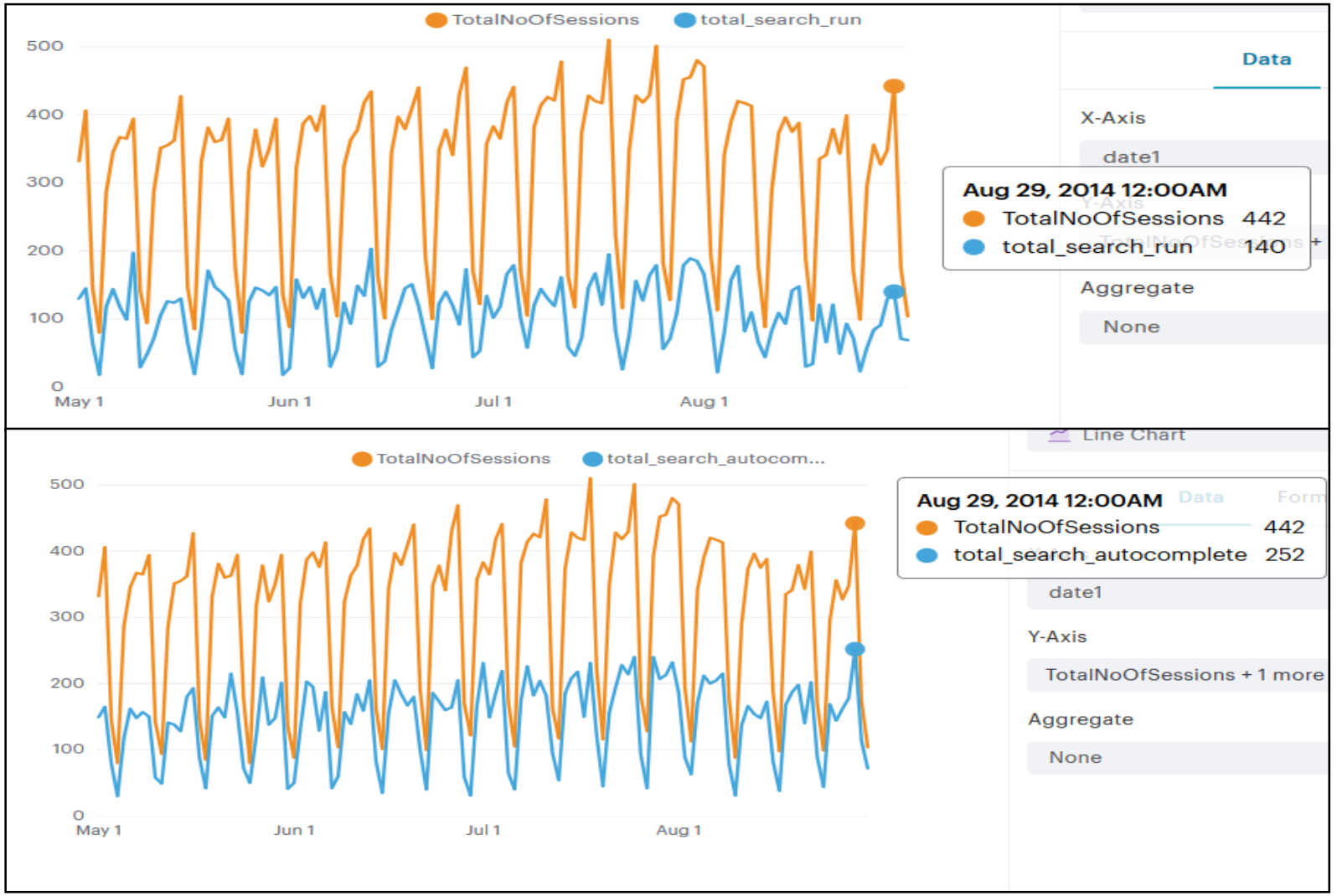
Step 2: Output



Step 3: Analysing Search Events per session over time

```
-- Query for analysing search potency w.r.t total user sessions over time

select sub.date1,
    count(distinct(sub.sessionno)) as "TotalNoOfSessions",
                SUM(CASE WHEN sub.event_name = 'search_run' THEN sub.total ELSE NULL END) AS "total_search_run",
    SUM(CASE WHEN sub.event_name = 'search_autocomplete' THEN sub.total ELSE NULL END) AS "total_search_autocomplete",
    SUM(CASE WHEN sub.event_name = 'search_click_result_1' THEN sub.total ELSE NULL END) AS "search_result1_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_2' THEN sub.total ELSE NULL END) AS "search_result2_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_3' THEN sub.total ELSE NULL END) AS "search_result3_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_4' THEN sub.total ELSE NULL END) AS "search_result4_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_5' THEN sub.total ELSE NULL END) AS "search_result5_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_6' THEN sub.total ELSE NULL END) AS "search_result6_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_7' THEN sub.total ELSE NULL END) AS "search_result7_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_8' THEN sub.total ELSE NULL END) AS "search_result8_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_9' THEN sub.total ELSE NULL END) AS "search_result9_click",
                SUM(CASE WHEN sub.event_name = 'search_click_result_10' THEN sub.total ELSE NULL END) AS "search_result10_click"
from(
                select DATE_TRUNC('day', occurred_at) as "date1",
                                                sessionno,
                                                event_name,
                                                count(event_name) as "total"
                from sudhanshu_chib.rescasestudy2data
                group by 1,2,3
                order by 1
                ) sub
group by 1
```

Step 3: Output (1/2)



Number of search_run w.r.t number of total sessions have remained stable over the time period

Number of autocomplete_search w.r.t number of total sessions have also remained stable over the time period

Step 3: Output (2/2)

Number of autocomplete_search have increased w.r.t search_run over the time period

When autocomplete_search is initiated not many results are clicked

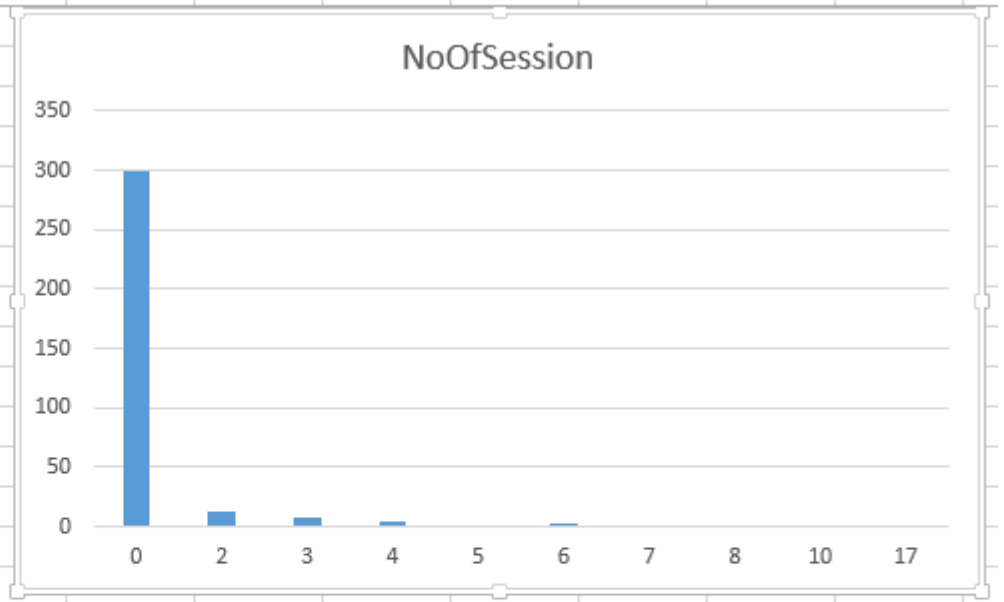
Step 4: Finding importance of Search_Run per user session

How many session perform 1 or more search_run event

```
select sub1.date1, sub1.total_search_run, count(sub1.sessionno) as "NoOfSession"
from (
      select sub.date1,
            sub.sessionno,
                                                SUM(CASE WHEN sub.event_name = 'search_run' THEN sub.total ELSE NULL END) AS "total_search_run",
                                                SUM(CASE WHEN sub.event_name = 'search_autocomplete' THEN sub.total ELSE NULL END) AS "total_search_autocomplete",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_1' THEN sub.total ELSE NULL END) AS "search_result1_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_2' THEN sub.total ELSE NULL END) AS "search_result2_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_3' THEN sub.total ELSE NULL END) AS "search_result3_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_4' THEN sub.total ELSE NULL END) AS "search_result4_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_5' THEN sub.total ELSE NULL END) AS "search_result5_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_6' THEN sub.total ELSE NULL END) AS "search_result6_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_7' THEN sub.total ELSE NULL END) AS "search_result7_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_8' THEN sub.total ELSE NULL END) AS "search_result8_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_9' THEN sub.total ELSE NULL END) AS "search_result9_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_10' THEN sub.total ELSE NULL END) AS "search_result10_click"
                                    from(
                                                select DATE_TRUNC('day', occurred_at) as "date1",
                                                                        sessionno,
                                                                        event_name,
                                                                        count(event_name) as "total"
                                                from sudhanshu_chib.rescasestudy2data
                                                group by 1,2,3
                                                order by 1
                                                ) sub
                                    group by 1,2
                        )sub1
group by 1, 2
order by 1, 2
```

Step 4: Output

	date1	total_sear	NoOfSess	on
2	5/1/2014 0:00	0	300	
3	5/1/2014 0:00	2	12	
4	5/1/2014 0:00	3	7	
5	5/1/2014 0:00	4	5	
6	5/1/2014 0:00	5	1	
7	5/1/2014 0:00	6	3	
8	5/1/2014 0:00	7	1	
9	5/1/2014 0:00	8	1	
10	5/1/2014 0:00	10	1	
11	5/1/2014 0:00	17	1	
12	5/2/2014 0:00	0	372	
13	5/2/2014 0:00	1	2	
14	5/2/2014 0:00	2	5	
15	5/2/2014 0:00	3	9	
16	5/2/2014 0:00	4	3	
17	5/2/2014 0:00	5	3	
18	5/2/2014 0:00	6	7	
19	5/2/2014 0:00	7	4	
20	5/2/2014 0:00	9	1	
21	5/3/2014 0:00	0	130	
22	5/3/2014 0:00	1	3	
23	5/3/2014 0:00	2	3	



Search_Run is evoked in typically 10% of total sessions in a day

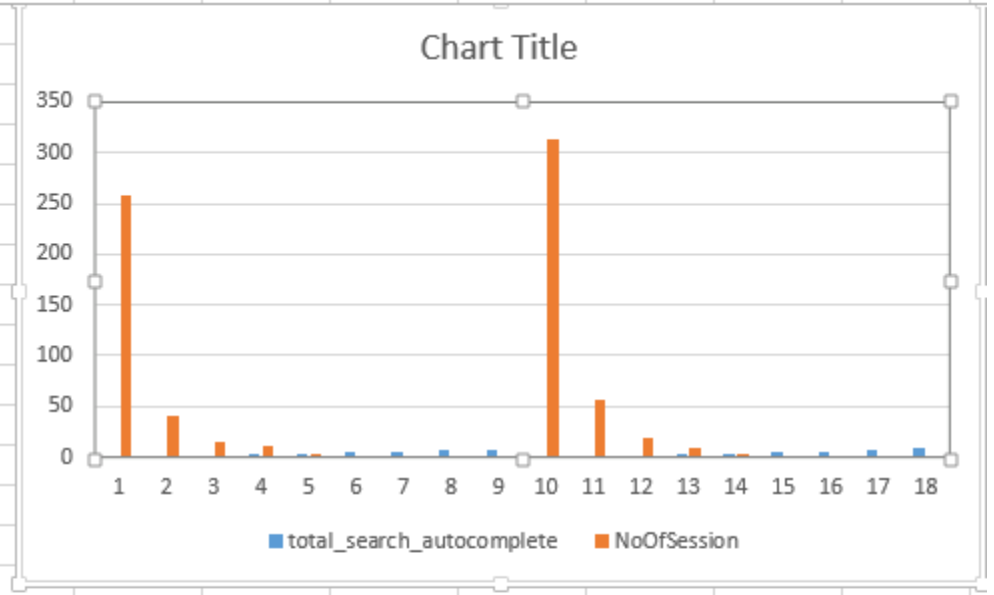
Step 5: Finding importance of search_autocomplete per user session

How many session perform 1 or more search_autocomplete event

```
select sub1.date1, sub1.total_search_autocomplete, count(sub1.sessionno) as "NoOfSession"
from (
      select sub.date1,
            sub.sessionno,
                                                SUM(CASE WHEN sub.event_name = 'search_run' THEN sub.total ELSE NULL END) AS "total_search_run",
                                                SUM(CASE WHEN sub.event_name = 'search_autocomplete' THEN sub.total ELSE NULL END) AS "total_search_autocomplete",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_1' THEN sub.total ELSE NULL END) AS "search_result1_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_2' THEN sub.total ELSE NULL END) AS "search_result2_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_3' THEN sub.total ELSE NULL END) AS "search_result3_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_4' THEN sub.total ELSE NULL END) AS "search_result4_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_5' THEN sub.total ELSE NULL END) AS "search_result5_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_6' THEN sub.total ELSE NULL END) AS "search_result6_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_7' THEN sub.total ELSE NULL END) AS "search_result7_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_8' THEN sub.total ELSE NULL END) AS "search_result8_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_9' THEN sub.total ELSE NULL END) AS "search_result9_click",
                                                SUM(CASE WHEN sub.event_name = 'search_click_result_10' THEN sub.total ELSE NULL END) AS "search_result10_click"
                                    from(
                                                select DATE_TRUNC('day', occurred_at) as "date1",
                                                                        sessionno,
                                                                        event_name,
                                                                        count(event_name) as "total"
                                                from sudhanshu_chib.rescasestudy2data
                                                group by 1,2,3
                                                order by 1
                                                ) sub
                                    group by 1,2
                        )sub1
group by 1, 2
order by 1, 2
```

Step 5: Output

date1	total_search_autocomplete	NoOfSession
5/1/2014 0:00	0	258
5/1/2014 0:00	1	40
5/1/2014 0:00	2	14
5/1/2014 0:00	3	11
5/1/2014 0:00	4	4
5/1/2014 0:00	5	1
5/1/2014 0:00	6	2
5/1/2014 0:00	7	1
5/1/2014 0:00	8	1
5/2/2014 0:00	0	314
5/2/2014 0:00	1	56
5/2/2014 0:00	2	18
5/2/2014 0:00	3	10
5/2/2014 0:00	4	4
5/2/2014 0:00	5	1
5/2/2014 0:00	6	1
5/2/2014 0:00	7	1
5/2/2014 0:00	9	1
5/3/2014 0:00	0	113
5/3/2014 0:00	1	15
5/3/2014 0:00	2	7
5/3/2014 0:00	3	3



Autocomplete_search is evoked in typically 26% of total sessions in a day

Insights and Recommendation:

- Search_Run and Autocomplete_Search are important functionality of the Yammer application as they are part of approx. 26% and 10% sessions respectively
- Use of Autocomplete_Search has increased over Search_Run over the time period but not many users are clicking the results shown by autocomplete_search
- Identified problem area:

    a. Search_Run functionality is losing ground indicating users are not finding it useful

    b. Autocomplete_Search is being frequently used but users don't tend to click on the subsequent search results

Recommendation:

a. Improve the search algorithm for search_autocomplete so that more important results are displayed upfront resulting in user clicks

b. Search_Run functionality is losing ground and its algorithm needs to be upgraded

Problem with Mode Analytics Solution

The logic used by mode analytics in their query to identify a user session will work only if the session ends with a search. So if in a session a user after login performs a search run or triggers an auto complete search and proceeds to like or click a message then all the intermediate search steps are lost.

Hence the solution posted on the website will not be completely correct

The same is highlighted in the next slides:

Problem with Mode Analytics Solution

```
SELECT user_id,
       event_type,
       event_name,
       occurred_at,
       occurred_at - LAG(occurred_at,1) OVER (PARTITION BY user_id ORDER BY occurred_at) AS last_event,
       LEAD(occurred_at,1) OVER (PARTITION BY user_id ORDER BY occurred_at) - occurred_at AS next_event,
       ROW_NUMBER() OVER () AS id
  FROM tutorial.yammer_events e
 WHERE e.event_type = 'engagement'
 ORDER BY user_id,occurred_at
```

✓ Succeeded in 1s

Export Copy Chart Pivot

100 rows returned

One session details

	user_id	event_type	event_name	occurred_at	last_event	next_event	id
13	4	engagement	login	2014-05-27 15:09:09	0 years 0 mons 3 days 3 hours 28 mins 49.00 secs	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	13
14	4	engagement	like_message	2014-05-27 15:09:36	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	0 years 0 mons 0 days 0 hours 0 mins 4.00 secs	14
15	4	engagement	search_run	2014-05-27 15:09:40	0 years 0 mons 0 days 0 hours 0 mins 4.00 secs	0 years 0 mons 0 days 0 hours 0 mins 26.00 secs	15
16	4	engagement	search_run	2014-05-27 15:10:06	0 years 0 mons 0 days 0 hours 0 mins 26.00 secs	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	16
17	4	engagement	search_click_result_6	2014-05-27 15:10:38	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	0 years 0 mons 0 days 0 hours 0 mins 0.00 secs	17
18	4	engagement	search_run	2014-05-27 15:10:38	0 years 0 mons 0 days 0 hours 0 mins 0.00 secs	0 years 0 mons 0 days 0 hours 0 mins 19.00 secs	18
19	4	engagement	search_click_result_10	2014-05-27 15:10:57	0 years 0 mons 0 days 0 hours 0 mins 19.00 secs	0 years 0 mons 0 days 0 hours 0 mins 4.00 secs	19
20	4	engagement	search_run	2014-05-27 15:11:01	0 years 0 mons 0 days 0 hours 0 mins 4.00 secs	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	20
21	4	engagement	search_click_result_2	2014-05-27 15:11:28	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	0 years 0 mons 0 days 0 hours 0 mins 30.00 secs	21
22	4	engagement	search_click_result_7	2014-05-27 15:11:58	0 years 0 mons 0 days 0 hours 0 mins 30.00 secs	0 years 0 mons 0 days 0 hours 0 mins 31.00 secs	22
23	4	engagement	search_click_result_1	2014-05-27 15:12:29	0 years 0 mons 0 days 0 hours 0 mins 31.00 secs	0 years 0 mons 0 days 0 hours 0 mins 29.00 secs	23
24	4	engagement	search_click_result_5	2014-05-27 15:12:58	0 years 0 mons 0 days 0 hours 0 mins 29.00 secs	0 years 0 mons 0 days 20 hours 8 mins 4.00 secs	24
25	4	engagement	login	2014-05-28 11:21:02	0 years 0 mons 0 days 20 hours 8 mins 4.00 secs	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	25

Problem with Mode Analytics Solution

```
90                    event_type,
91                    event_name,
92                    occurred_at,
93                    occurred_at - LAG(occurred_at,1) OVER (PARTITION BY user_id ORDER BY occurred_at) AS last_event,
94                    LEAD(occurred_at,1) OVER (PARTITION BY user_id ORDER BY occurred_at) - occurred_at AS next_event,
95                    ROW_NUMBER() OVER () AS id
96                FROM tutorial.yammer_events e
97               WHERE e.event_type = 'engagement'
98               ORDER BY user_id,occurred_at
99             ) bounds
100  WHERE last_event >= INTERVAL '10 MINUTE'
101     OR next_event >= INTERVAL '10 MINUTE'
102    OR last_event IS NULL
103     OR next_event IS NULL
```

Cause of error

Ready

Export Copy Chart Pivot

100 rows returned

	user_id	event_type	event_name	occurred_at	last_event	next_event	id	session
1	4	engagement	login	2014-05-13 09:31:47		0 years 0 mons 0 days 0 hours 0 mins 23.00 secs	1	1
2	4	engagement	search_autocomplete	2014-05-13 09:32:58	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	0 years 0 mons 6 days 23 hours 58 mins 32.00 secs	4	1
3	4			2014-05-20 09:31:30	0 years 0 mons 6 days 23 hours 58 mins 32.00 secs	0 years 0 mons 0 days 0 hours 0 mins 25.00 secs	5	5
4	4			09:33:01	0 years 0 mons 0 days 0 hours 0 mins 30.00 secs	0 years 0 mons 4 days 2 hours 6 mins 52.00 secs	10	5
5	4			11:39:53	0 years 0 mons 4 days 2 hours 6 mins 52.00 secs	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	11	11
6	4	engagement	home_page	2014-05-24 11:40:20	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	0 years 0 mons 3 days 3 hours 28 mins 49.00 secs	12	11
7	4	engagement	login	2014-05-27 15:09:09	0 years 0 mons 3 days 3 hours 28 mins 49.00 secs	0 years 0 mons 0 days 0 hours 0 mins 27.00 secs	13	13
8	4	engagement	search_click_result_5	2014-05-27 15:12:58	0 years 0 mons 0 days 0 hours 0 mins 29.00 secs	0 years 0 mons 0 days 20 hours 8 mins 4.00 secs	24	13
9	4	engagement	login	2014-05-28 11:21:02	0 years 0 mons 0 days 20 hours 8 mins 4.00 secs	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	25	25
10	4	engagement	home_page	2014-05-28 11:22:40	0 years 0 mons 0 days 0 hours 0 mins 32.00 secs	0 years 0 mons 1 days 22 hours 58 mins 38.00 secs	28	25
11	4	engagement	login	2014-05-30 10:21:18	0 years 0 mons 1 days 22 hours 58 mins 38.00 secs	0 years 0 mons 0 days 0 hours 0 mins 25.00 secs	29	29
12	4	engagement	home_page	2014-05-30 10:21:46	0 years 0 mons 0 days 0 hours 0 mins 3.00 secs	0 years 0 mons 1 days 5 hours 37 mins 0.00 secs	31	29
13	4	engagement	login	2014-05-31 15:58:46	0 years 0 mons 1 days 5 hours 37 mins 0.00 secs	0 years 0 mons 0 days 0 hours 0 mins 26.00 secs	32	32

Details captured by mode query:
Lot of search events are lost