

# Understanding Yammer Search Functionality

This report aims at understanding the current performance of the Yammer's search functionality and aims at answering if it requires any modification or improvement.

## Introduction

The complete problem statement can be found at:

<https://community.modeanalytics.com/sql/tutorial/understanding-search-functionality/>  
(<https://community.modeanalytics.com/sql/tutorial/understanding-search-functionality/>)

Yammer is a corporate social networking website used internally within organisations and the product team is trying to decide if it is worth investing engineering resources into improving or modifying the search functionality. To prepare this report and test various hypotheses, we have been provided with two tables: tutorial.yammer\_users and tutorial.yammer\_events.

The former table consists of user information from the Yammer database. We have run a few preliminary SQL commands to get a picture of the structure of the data present in this table.

Yammer Users Table

	user_id	created_at	company_id	language	activated_at	state
1	0	2013-01-01 20:59:39	5737	english	2013-01-01 21:01:07	active
2	1	2013-01-01 13:07:46	28	english		pending
3	2	2013-01-01 10:59:05	51	english		pending
4	3	2013-01-01 18:40:36	2800	german	2013-01-01 18:42:02	active
5	4	2013-01-01 14:37:51	5110	indian	2013-01-01 14:39:05	active
6	5	2013-01-01 13:39:51	2463	spanish		pending
7	6	2013-01-01 18:37:27	11699	english	2013-01-01 18:38:45	active
8	7	2013-01-01 16:19:01	4765	french	2013-01-01 16:20:28	active
9	8	2013-01-01 04:38:30	2698	french	2013-01-01 04:40:10	active
10	9	2013-01-01 08:04:17	1	french		pending

Since we are dealing with users who can perform search, we will be dealing only with users whose state is 'active'. Although this information is not critical to the analysis, it helps us in knowing the total number of users currently accessing the search feature.

Number of Active Users

	count
1	9381

Next, we examine the Yammer Events table. This table contains the bulk of the data that will be required to perform our analysis and test various hypotheses. The Events table contains informations about a variety of events that Yammer users have triggered over time.

However, we are interested only in 3 events recorded by this table; namely search\_autocomplete, search\_run and search\_click\_result\_x

These events record the following:

- **search\_autocomplete:** This is logged when a user clicks on a search option from autocomplete
- **search\_run:** This is logged when a user runs a search and sees the search results page.
- **search\_click\_X:** This is logged when a user clicks on a search result. X, which ranges from 1 to 10, describes which search result was clicked.

The query result below gives us a picture of the table with the appropriate data (as described above).

	user_id	occurred_at	event_type	event_name	location	device	user_type
1	10522	2014-05-02 11:03:1...	engagemen...	search_run	Japan	dell inspiron noteboo...	3
2	10522	2014-05-02 11:03:4...	engagemen...	search_run	Japan	dell inspiron noteboo...	3
3	11037	2014-05-01 07:25:3...	engagemen...	search_autocomplet...	United Kingdo...	iphone 4s	1
4	11194	2014-05-06 16:18:2...	engagemen...	search_autocomplet...	United States	ipad mini	2
5	11212	2014-05-06 07:30:2...	engagemen...	search_autocomplet...	France	macbook pro	3
6	11212	2014-05-06 07:30:5...	engagemen...	search_autocomplet...	France	macbook pro	3
7	11212	2014-05-06 07:31:2...	engagemen...	search_autocomplet...	France	macbook pro	3
8	11212	2014-05-06 13:42:3...	engagemen...	search_autocomplet...	France	nexus 7	3
9	11212	2014-05-06 13:43:0...	engagemen...	search_autocomplet...	France	nexus 7	3
10	11212	2014-05-06 13:43:3...	engagemen...	search_autocomplet...	France	nexus 7	3

# Hypotheses

To gauge the searching experience of users, it is important that we formulate certain concrete hypotheses and test them with the data available. In this report, we will resort to answering the following questions.

## What fraction of the searches are autocomplete?

If the majority of the searches are autocomplete, we can assume, to a certain degree, that the search is returning results that are satisfactory to the user. On the other hand, if a considerable amount of searches are directing users to a search page, the experience needs to be improved.

## Do the users return to using the search feature?

A return would mean that the user is enjoying using the feature. In contrast, a non-return would indicate frustration.

## How good is the Search Run ordering algorithm?

If the searches are returning good results, it would mean that the number of times users click on the top results would be substantially higher than the number of times they click on the results towards the bottom.

## How many times does the search feature get used on average?

To a certain degree, a greater number of searches would imply that the search feature is really important to the users. Although care has to be taken to detect cases where frustrated users are just using the feature multiple times due to unsatisfactory results.

Before performing any analysis, we take a look at the amount of data available with us.

### Number of Searches

	count	duration
1	40611	0 years 0 mons 122 days 19 hours 53 mins 42.00 secs

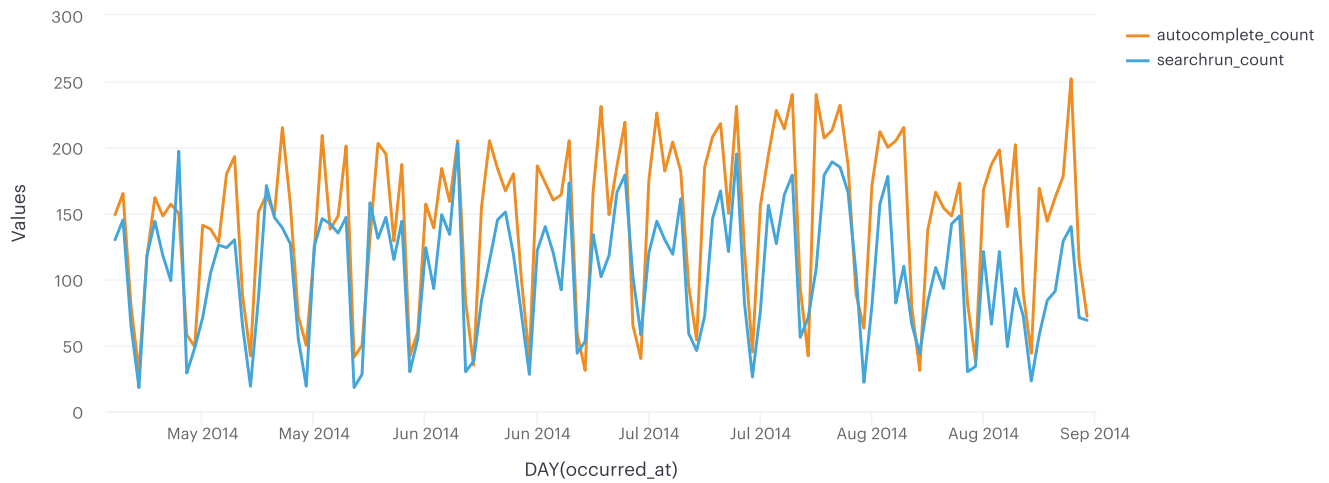
The above query result shows that roughly 20,000 searches were made in a span of approximately 120 days with an average of 332 searches per day.

An average of 330 searches does indicate that search is a reasonably popular feature which demands a good user experience. This, in a way, validates the report being done on it. However, since we do not have complementary data to gauge this against, a definitive conclusion regarding the popularity of the feature cannot be reached.

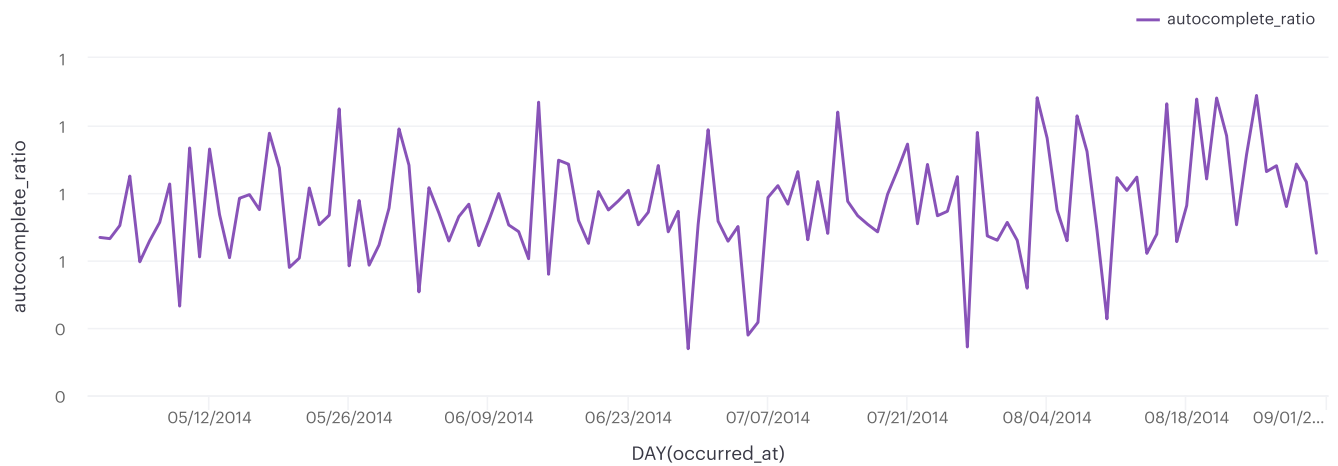
# Autocomplete v/s Search Run

The chart gives us a comparison of the number of autocompletes and search runs done by users on a daily basis.

## Comparison of Searches



## Ratio of Autocompletes



From the above charts, the following deductions can be made:

On average, the number of autocomplete searches are slightly higher than the number of search runs. This implies that users are able to get what they are looking for while typing in more than half of the cases.

That being said, it is extremely worthwhile to invest resources in improving the quality of search results to increase this gap. Present day users expect to get results of what they're searching midway through typing and a 50% hit rate simply isn't sufficient.

## Do Users Return to Using the Search Feature?

An important statistic to examine if users are returning to using the Search feature. The number of returning users is a very powerful metric to gauge the popularity of both the search features: autocomplete and search run.

To test this hypothesis, we examine the number of times the user has used a particular feature over the months with the data that has been provided to us.

Search Statistics by User and Month

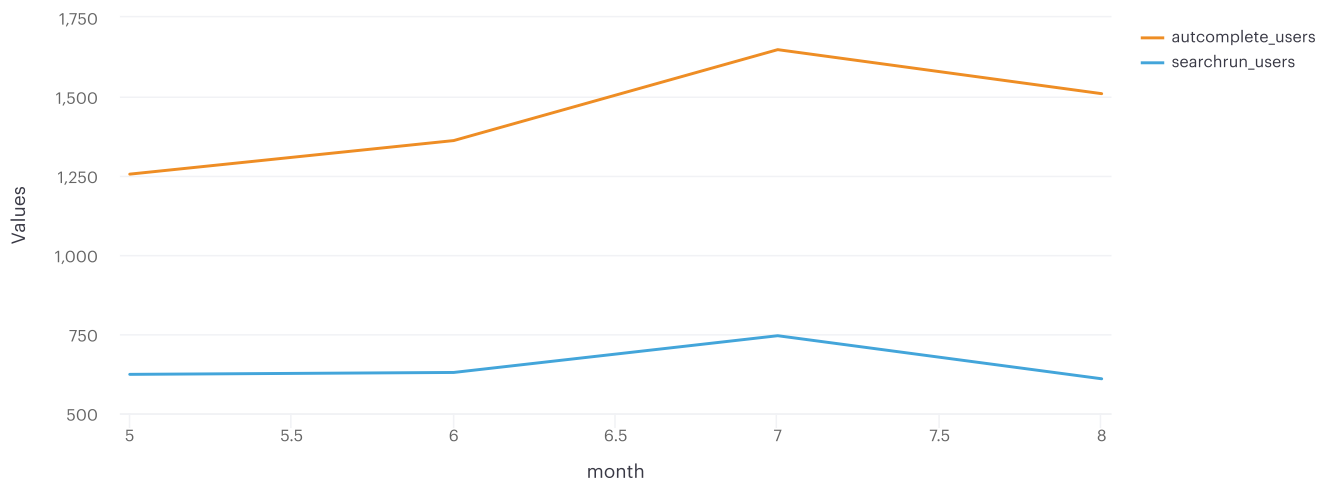
	user_id	month	autocomplete_count	searchrun_count
1	4	5	7	5
2	4	6	4	7
3	4	7	0	0
4	8	5	9	0
5	8	7	0	0
6	11	6	3	0
7	11	7	1	0
8	11	8	0	0
9	17	7	2	3
10	17	8	4	0
11	19	6	0	6
12	19	7	1	0
13	20	5	1	0

The search statistics shown above gives us little to no insight into the popularity of both the search features. Therefore, it would be better if we could count the number of distinct users every subsequent month. An increase would indicate good user experience whereas a decrease would indicate otherwise.

Number of Distinct Users

	autocomplete_users	month	searchrun_users
1	1255	5	623
2	1361	6	629
3	1648	7	745
4	1509	8	609

### Distinct Users



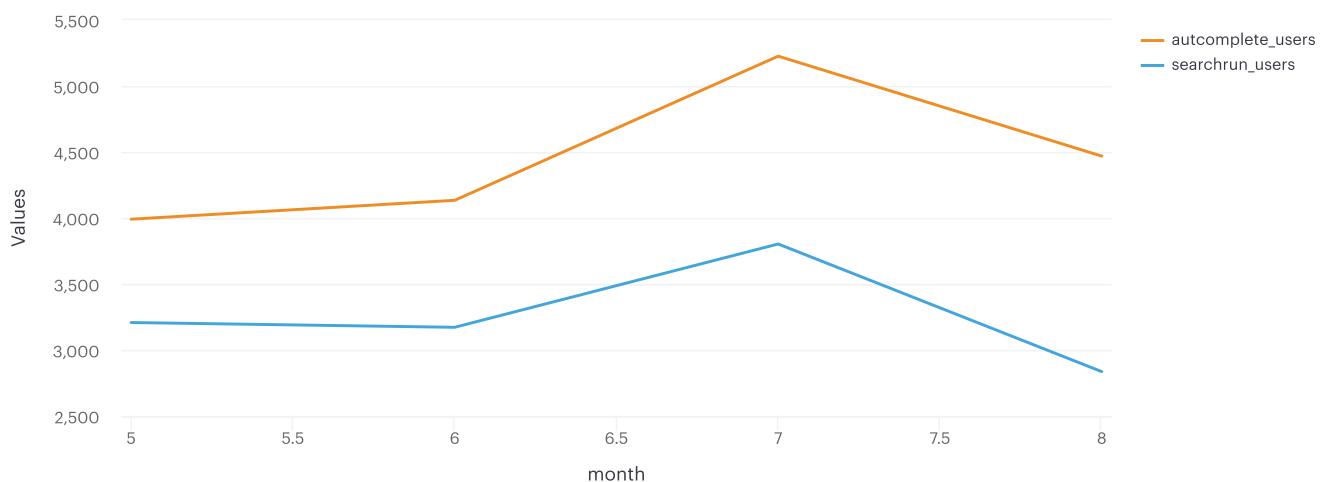
From the graph above, it can be clearly seen that the number of distinct autocomplete users has increased, as a whole, in the span of four months for which the data has been provided. In contrast, the number of search run users has experienced a dip in the same period of time. Therefore, it would be logical to conclude that the Autocomplete Search Feature has a substantially better experience with more returning users whereas the search runs may have led to some frustration on part of the user.

The other statistic we are interested in is the number of queries performed with each subsequent month. This would help us in strengthening the hypothesis formed with the aforementioned query results.

### Number of Queries

	autocomplete_users	month	searchrun_users
1	3991	5	3208
2	4134	6	3172
3	5226	7	3803
4	4469	8	2836

### Queries

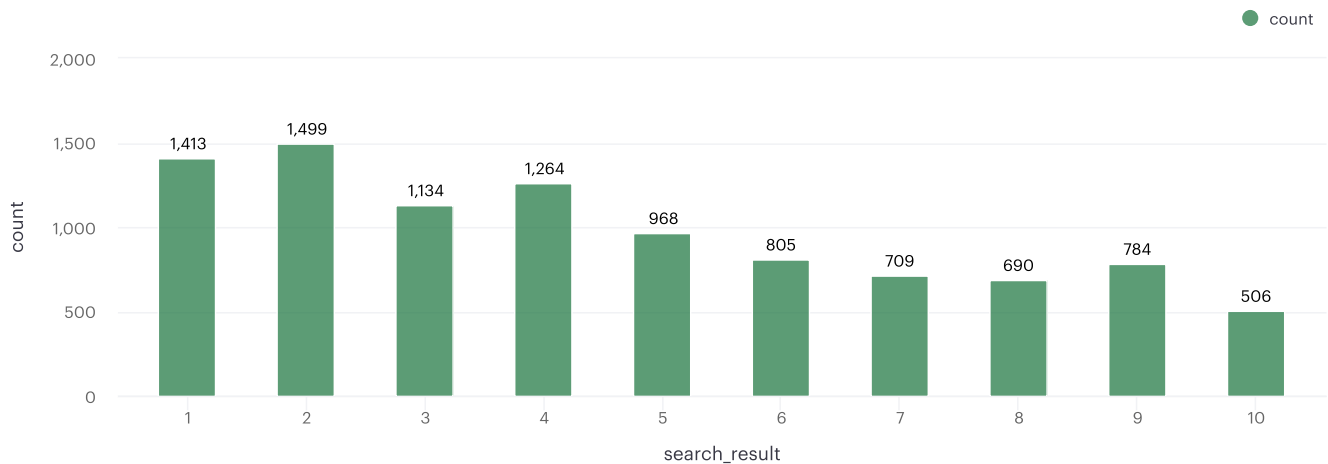


From the chart above, it can be clearly seen that the number of queries has increased over the months for Autocomplete but has decreased for Search Runs. Our hypothesis that Autocompletes are far more popular than the Search Runs is thus established.

## How good is the Search Run ordering algorithm?

To test the potency of the Search Run ordering algorithm, we bring up the number of times a user has clicked on an nth result. If the number of clicks decreases as n increases, it would indicate a good ordering. If not, then this would be an area that would require some improvement.

Search Ordering Algorithm Results



The bar chart shows us a very uneven distribution of the click result distribution. Furthermore, there is no trace of a systematic decrease in the number of clicks as the search result rank decreases.

This clearly indicates that the ordering algorithm is not doing a very good job of returning results that the user is expecting.

Search Run and Click Result Counts

searchrun_count		clickresult_count	
1	13019		9772

Furthermore, it can be seen that the number of Search Click Results is substantially lower than the number of Search Runs. This indicates that in a considerable number of cases, the user didn't end up finding what s/he was looking for.

# Conclusions

From the analysis conducted in this report, the following conclusions and inferences can be derived:

1. The search functionality is an important feature for Yammer Users. More than 40,000 searches in 120 days with an average of 330 search events a day. It is therefore essential that search be optimised and its user experience be made painless.
2. Autocompletes account for roughly half the searches. Since it is clearly the more popular and better functioning feature (described later), it makes sense to improve the autocomplete algorithm to ensure users get what they are looking for whilst typing without having to go into a separate Search Run page.
3. The Autocomplete feature is clearly more popular than the Search Run Feature. It has more returning users and more queries made to it as a whole in the four months of data observed. Therefore, it makes sense to incentivise users to use Autocomplete more or improve the UX of Search Runs.
4. The Search Run Ordering Algorithm is poor. The Click Results show a very uneven distribution. Furthermore, in a considerable fraction of the cases, the users do not click on a result at all. This may be directly contributing to the unpopularity of the Search Run feature.