

Dataset reference: <https://github.com/metgauss/Discovery-Hiring-Analyst-2016>

Executive Summary

The purpose of this analysis is to find out the performance of the metrics by using the data that is available. As a quick introduction, this data includes 67979 unique sessions that the server captured between 01 and 08 of March 2016. Each session has, on average, 6 events in the server. The dataset is also divided in A and B Groups.

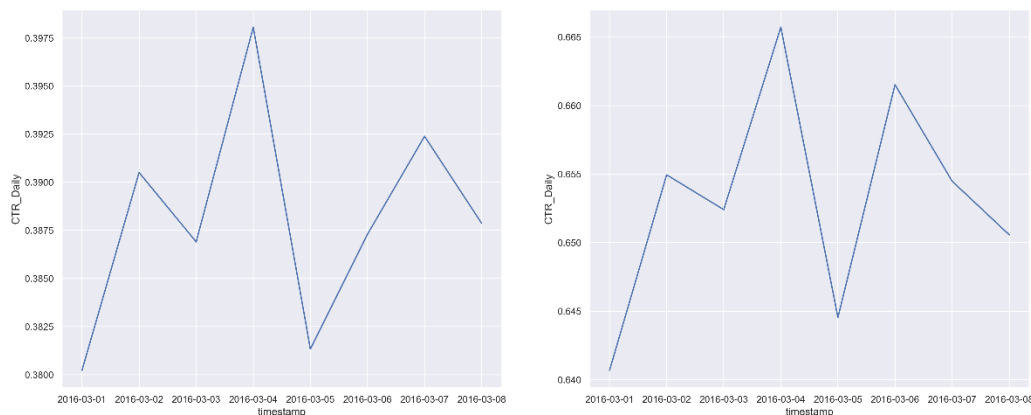
One of the relevant insights found in the dataset is that **Group B** contains all the sessions that only included searches performed in the SERP with no any extra action. This covers **40.5%** of the total sessions. Within the Group B, 72% of the total sessions are only search in the result. Without any doubt, this affects the web metrics that Group B might have.

These events that only included searches in SERP are not coming from a particular date, but from all the 7 days. Therefore, it is not easy to draw conclusions from where this data might come from. In order to have clear metrics, the summary contains metric comparison including and not including this irregular data.

- What is our daily overall clickthrough rate?

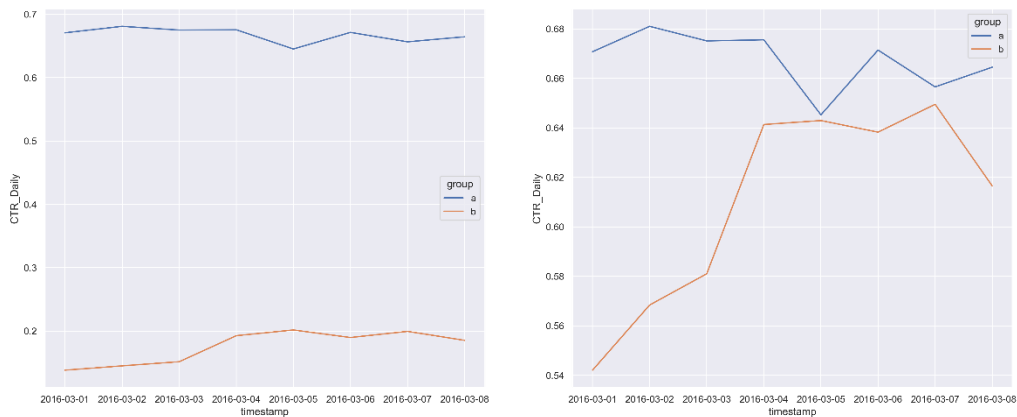
The daily CTR maintains steady throughout the 7 days. If you remove events with only SERP searches, the CTR is approximately between **64%** and **67%**. However, if you include the irregular data, the CTR goes down to 39% approximately. See the graph below.

Overall Daily CTR (Not cleaned dataset vs cleaned dataset)



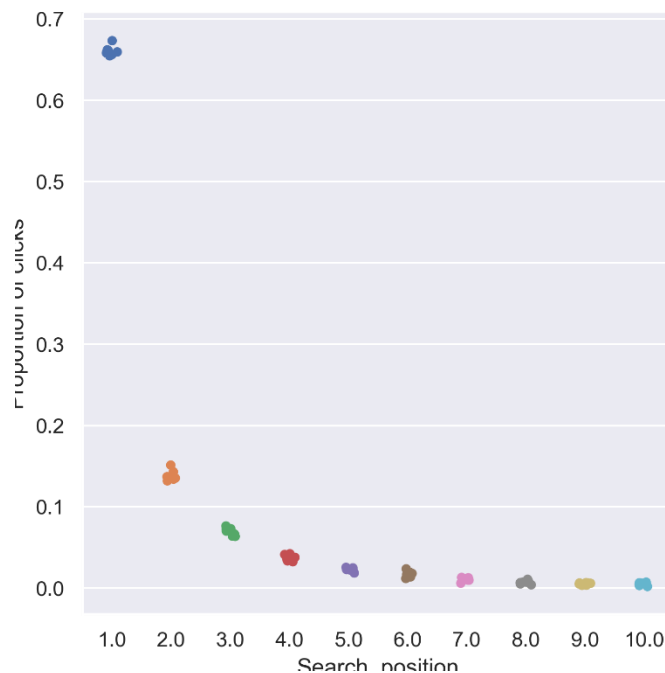
When it comes comparing Group A with Group B, they are quite similar (Group A slightly better) removing the irregular sessions. If we include this last data, Group A outperforms significantly Group B.

Daily CTR by Group (Not cleaned dataset vs cleaned dataset)



- **Which results do people tend to try first? How does it change day-to-day?**

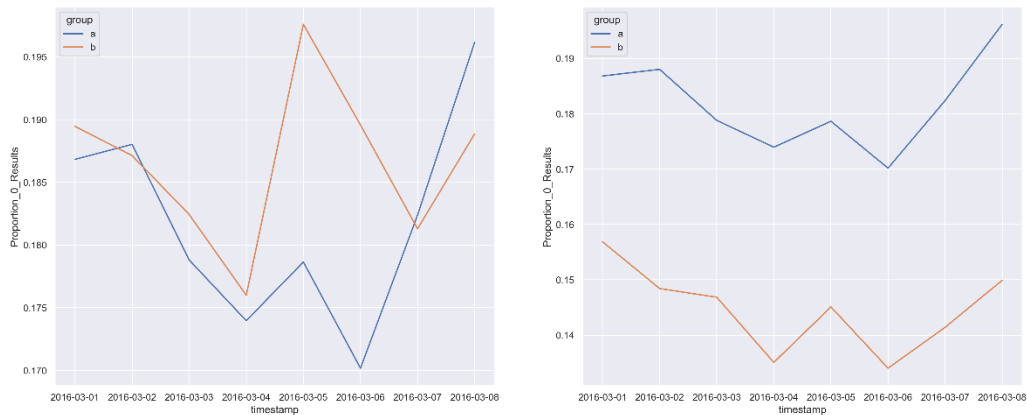
Based on the data available, **65%** of the users tend to click in the first result, whereas approximately 15 of the users click the second result. Finally, less than 10% of the users click in 3rd to 10th position. It is clear to say the importance of being the first result in the SERP as **65%** of the users choose it. As you can see below, this split does not vary significantly day-over-day in the dataset. Each color dot represents a different day.



- **What is our daily overall zero results rate? How does it vary between the groups?**

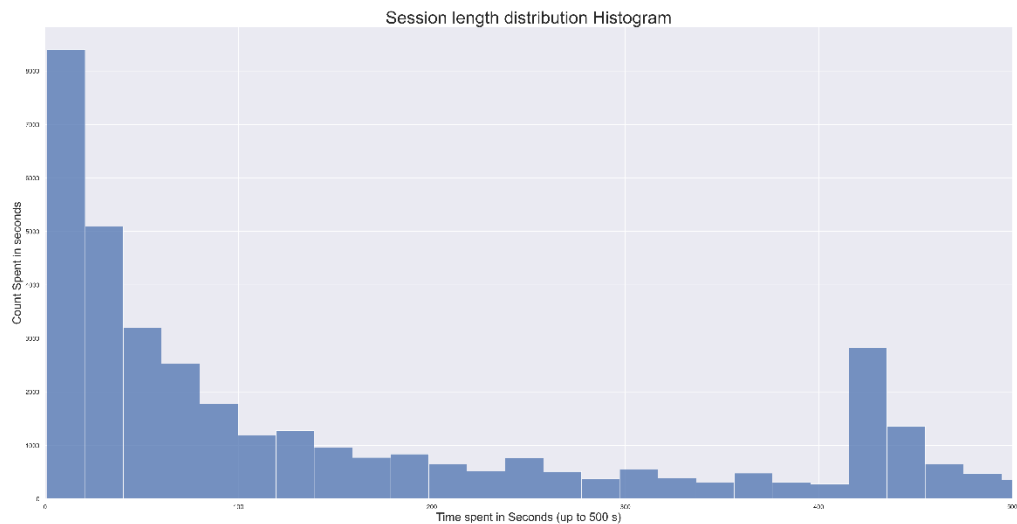
The daily ZRR maintains quite steady across the 7 days. If we take the 7 days, the overall rate is 18%. Group B has higher % with the whole dataset and Group A outperforms Group B if we clean the irregular sessions from the dataset.

ZRR by Group (Not cleaned dataset vs cleaned dataset)



- **Insights on Session length**

As it can be seen in the graph below, if we filter out sessions with 0 time spent, most of the sessions have a length of approximately less than 100 seconds in total. Surprisingly, there is a spike of time length count between 400 and 500 seconds (6 and 8 minutes). In other words, the count of sessions is higher between 6-8 minutes than 2 to 6 minutes.



If we compare the Result position with the session length, we see that regardless of the time length the first position in the SERP predominates in the entire dataset. **With this information, we can conclude saying that the session on each result position number is very similar as it varies proportionally across the session length.**

