

Coding Assessment

1. My observations about the data are as follows:

- a. **Data structure:** There are 469308 entries spread across 5 columns in the data. The period of data collection is 8/7/2022–14/7/2022.
- b. **Column information:** 'keyword_id' and 'keyword_rank' appear to represent identifiers and ranking positions, respectively. 'date' represents the date of the observations. 'searches' indicates the number of searches for a particular keyword on a specific date. 'search_engine' is a categorical variable representing the search engine used.
- c. **Data Summary**
The search_engine column ranges from 0 to 25, representing different search engines.
- d. **Data type:**
Keyword_id, keyword_rank, searches, search_engine have data type as 'int' and date has 'object'.
- e. **Descriptive analysis:** I did a descriptive analysis, these statistics helped me to understand the central tendencies, variability, and distribution of datasets helping to understand the characteristics of each column.
 - A. Count: All columns have the same count 469308 indicating no missing value.
 - B. Standard deviation: This helped me to understand amount of variation in a set of values.
 - C. Similarly, min and max values were calculated to understand the max and min values for each column.

I also noticed variations in keyword rankings for the same keyword_id. Specifically, there are instances where identical keyword_ids exhibit different ranks while sharing the same search engine. This suggests that keyword rankings are not consistent across all occurrences of a particular keyword_id, indicating potential fluctuations or changes in ranking positions for the same keyword across different time periods.

2. What You Attempted and Why

I began my search for keywords with a rank of 1, but soon found out that there were no keywords matching that rank. Consequently, I did the following.

I conducted a comprehensive analysis focused on a specified search engine. This involves aggregating the number of searches ("searches" column) for each "keyword_id" across various dates within the chosen search engine. Subsequently, the analysis aimed to pinpoint the specific "keyword_id" that exhibits the maximum total searches within the designated search engine. The keyword_id with maximum searches is considered to be '**rank=1**' for individual search_engines. By undertaking this targeted examination, I gained a valuable insight into the performance and popularity of individual keyword_id's within a particular search engine over time. The cumulative search count for each keyword is evaluated, enabling the identification of keyword_id's that resonate most with users of the chosen search engine. **This analytical approach not only facilitates a nuanced understanding of keyword dynamics but also sheds light on the competitive landscape within the specified search engine domain.**

I opted for this method due to the fluctuating nature of keyword_rank, where the same keyword_id has different keyword_ranks. The keyword_ranks are assigned in a random order, making it impractical to use them directly for the analysis to identify instances where rank equals 1. However, upon closer examination, I observed that specific keyword_ids were consistently recurring only within certain search_engines. With the help of this observation, I made the decision to proceed with my analysis, focusing on the 'keyword_id',

'date', 'searches', and 'search_engine' variables to determine the rank 1 for each individual search_engine. This tailored approach allowed me to navigate the variability in keyword_ranks and pinpoint specific keyword performance within distinct search_engine contexts.

3. Visualizations:

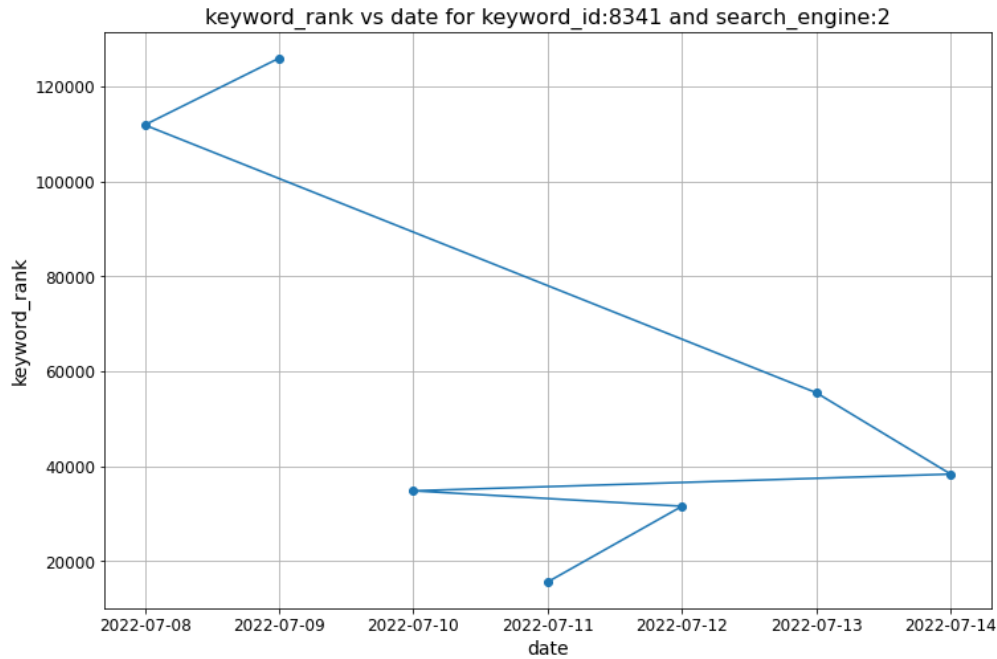


Fig 1: Keyword_rank Vs date for keyword_id: 8341 and search_engine:2

The Fig:1 illustrates how “keyword_rankings” vary with “dates” for a specific “keyword_id” and “search_engine”. Here in the above plot, I am considering “keyword_id: 8341” and “search_engine: 0” for illustration. Recognizing that keyword rankings are influenced by the number of searches, it became evident that to conduct a meaningful analysis on the number of searches for a top-ranked keyword (rank ‘1’) in a particular engine, I needed to aggregate the searches across all dates for that specific keyword and search engine. After aggregating the searches for a specific “search_engine” across all the dates, we can find the keyword with maximum searches, and I refer to it as rank ‘1’ which automatically becomes the top-ranked keyword for that specific engine. Furthermore, since there is no specific “keyword_id” explicitly defined with a rank of ‘1’ and considering that the given rankings are not unique throughout the entire dataset (i.e., rankings are repeated), I opted to identify the top-ranked keyword for each engine. Subsequently, I plotted the number of searches for this top-ranked keyword (rank ‘1’) across all 26 search engines.

4. Model(s) and Evaluation

For the model, I used the loop that iterates over the unique values in the 'search_engine' column of the DataFrame. For each iteration, it creates a subset of the DataFrame (engine_data) containing only the rows corresponding to the current search engine. Then, using the groupby function, it calculates the sum of searches for each unique keyword_id within that particular search engine.

The variable keyword_sum holds the aggregated sum of searches for each keyword_id in the current search engine. The code then uses the idxmax() function to find the keyword_id with the maximum sum of searches and retrieves the corresponding maximum search count.

Overall, this code efficiently performs a per-search-engine analysis, calculating the sum of searches for each keyword and identifying the keyword with the highest total searches within each specified search engine.

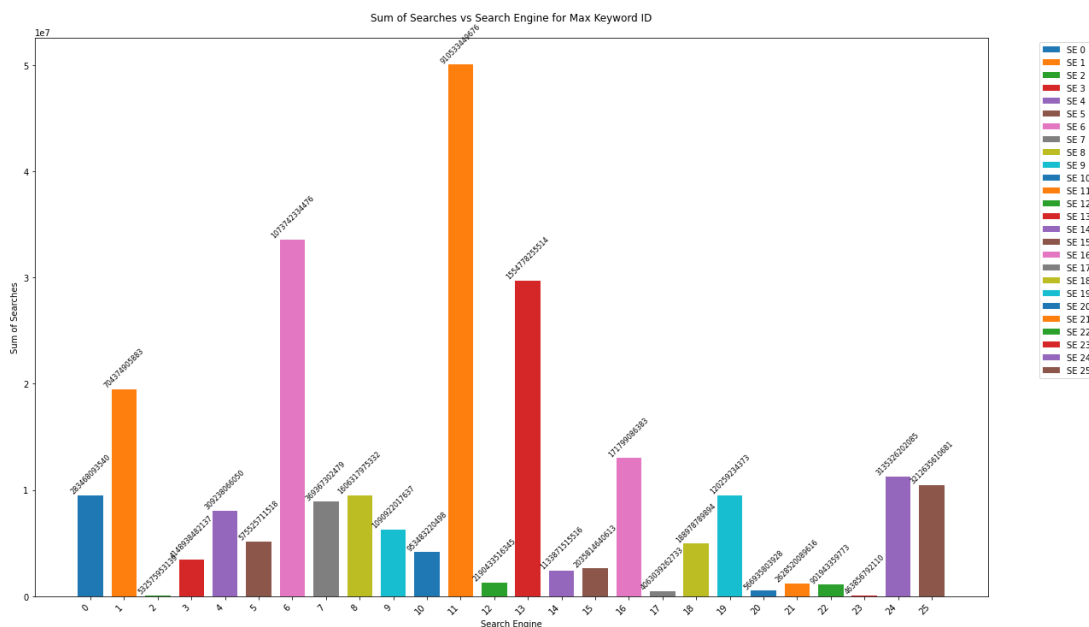


Fig 2: Total searches for max keyword_id in search_engines with rank 1

Evaluation,

Figure 2 illustrates the cumulative searches for the maximum keyword_id within search engines holding rank 1. The y-axis represents the sum of searches, while the x-axis denotes the search_engine.

- Notably, each search engine is linked to unique keyword_ids, as evident from the figure 2.
- Among the search engines, Search_engine 11 stands out with the highest search count, attributed to keyword_Id 910533449676, whereas Search_engine 2 reports the lowest searches.
- Search_engines 11, 6, 13, and 1 exhibit slightly higher search counts compared to their counterparts, while Search_engines 2, 23, 17, and 20 register lower search volumes.
- The distinct rank 1 values obtained for each search_engine underscore significant variations in associated search counts, emphasizing the diverse performance of keyword_ids across different search engines.

5. Conclusion:

The following conclusions were made.

a. Search Engine Performance Variation:

The analysis reveals notable variations in the cumulative searches for the maximum keyword_id across search engines with rank 1. Search engines exhibit diverse levels of performance, as evidenced by the range of search counts represented on the y-axis.

b. Distinct Keyword_IDs per Search Engine:

Each search engine is associated with unique keyword_ids, emphasizing the specificity of search engine dynamics. This implies that certain keywords may be more prominent or relevant within specific search engines compared to others.

c. Performance Disparity Among Search Engines:

Search_engine 11 emerges as the standout performer with the highest search count, linked to keyword_Id 910533449676. Conversely, Search_engine 2 reports the lowest searches, highlighting a significant performance disparity among different search engines.

d. Variation in Search Counts:

Certain search engines, namely 11, 6, 13, and 1, exhibit slightly higher search counts, indicating potentially higher user interest or search activity for keywords associated with these engines. Conversely, Search_engines 2, 23, 17, and 20 register lower search volumes, suggesting comparatively lower popularity or relevance of keywords in these engines.

e. Rank 1 Values and Associated Search Counts:

The distinct rank 1 values for each search_engine underscore substantial variations in the associated search counts. This indicates that the top-ranking keywords differ significantly across search engines, highlighting the need for tailored strategies when optimizing keywords for diverse search engine environments.

In conclusion, the findings emphasize the importance of understanding the specific dynamics of each search engine, tailoring keyword strategies accordingly, and recognizing the diverse performance of keyword_ids across different search engines. These insights can inform more effective search engine optimization (SEO) strategies and enhance the overall performance of keywords in online searches.