Device-dependent click-through rate estimation in Google organic search results based on clicks and impressions data

Journal:	*Aslib Journal of Information Management*
Manuscript ID	AJIM-04-2023-0107.R2
Manuscript Type:	Research Paper
Keywords:	click-through rate, organic search results, clicks, desktops, smartphones, tablets

Device-dependent click-through rate estimation in Google organic search results based on clicks and impressions data

Abstract

Purpose

The landscape of search engine usage has evolved since the last known data was used to calculate click through rate (CTR) values. The objective was to provide a replicable method for accessing data from the Google search engine using programmatic access and calculating CTR values from the retrieved data to show how the CTRs have changed since the last studies were published.

Design/methodology/approach

In this study, we present the estimated CTR values in organic search results based on actual clicks and impressions data, and establish a protocol for collecting this data using Google programmatic access. For this study, we collected data on 416,386 clicks, 31,648,226 impressions, and 8,861,416 daily queries.

Findings

Our results show that CTRs have decreased from previously reported values in both academic research and industry benchmarks. Our estimates indicate that the top-ranked result in Google's organic search results features a CTR of 9.28%, followed by 5.82% and 3.11% for positions two and three, respectively. We also demonstrate that CTRs vary across various types of devices. On desktop devices, the CTR decreases steadily with each lower ranking position. On smartphones, the CTR starts high but decreases rapidly, with an unprecedented increase from position 13 onwards. Tablets have the lowest and most variable CTR values.

Practical implications

The theoretical implications include the generation of a current dataset on search engine results and user behavior, made available to the research community, creation of a unique methodology for generating new datasets, and presenting the updated information on CTR trends. The managerial implications include the establishment of the need for businesses to focus on optimizing other forms of Google search results in addition to organic text results, and the possibility of application of this study's methodology to determine CTRs for their own websites.

Originality/value

This study provides a novel method to access real CTR data and estimates current CTRs for top organic Google search results, categorized by device.

Keywords: clicks; click-through rate; desktops; organic search results; clicks; smartphones; tablets

Device-dependent click-through rate estimation in Google organic search results based on clicks and impressions data

Introduction

Web search engines have been well-known systems used on the internet since their inception (Haider and Sundin, 2019). The first web search engines were established in 1990 (Mager, 2012). Since then, the Google search engine has become synonymous with searching and is used billions of times a day by billions of users[i]. ~~According to Statista, in February 2023, its global search market share was 92.3%, followed by Bing with a share of 3.03% and Yahoo! with a share of 1.21%[ii].~~ Search engines are embedded services in almost every web browser and serve as navigational tools for people using various devices connected to the internet, such as desktop computers, smartphones, and tablets (Levene, 2010). Search engines allow for searching of different types of content (Halavais, 2017). In the beginning of their creation, they were typically text search engines (Brin and Page, 2012), but later on, they were extended to cover other forms of content that can be searched, such as images (Jones and Oyen, 2023), videos (Wildemuth *et al.*, 2019), news (Karimi *et al.*, 2018), and books (Pechenick *et al.*, 2015). ~~Contemporary search engines also enable their users to search for particular information, such as stock prices, hotel reservations (Cezar and Ögüt, 2016), travel information (Fesenmaier *et al.*, 2011), currency rates, and more.~~ Search engines crawl the internet, download data, store it in data centers, create indexes for this data, and present the results using a search engine interface.

Search engines typically present a list of results for a query provided by a user (Lewandowski, 2023). There are various types of search results appearing on the search engine results page such as sponsored search results, map results, products, news, videos, images, direct answers, organic text results and others (Miklosik *et al.*, 2019). The structure of the search engine results page is changing to best match the type of search query

(Zhitomirsky-Geffet *et al.*, 2016). ~~The well-known and common structure of organic search results consists of 10 text results (Kammerer and Gerjets, 2012). Each result is composed of a title, two lines of description, and a web address, also known as a Uniform Resource Locator (URL) (Jamali and Shahbaztabar, 2017). The title is always presented at the top, but the order of the description and URL has changed over the years. The title usually presents the content of the *title* tag from the source code of the crawled website (Kattenbeck and Elsweiler, 2019). Sometimes, instead of presenting content from the *title* tag, the search engine chooses to present the content of *heading 1* or switches the order of words in the *title* tag. The two lines of description usually contain the content found in the *meta description* tag in the source code of the website, but the search engine can change it and present other content found in the source that is better connected to the provided query (Kim *et al.*, 2013). Two lines are conventionally present in the text result, but sometimes there may be more than two or just one. The presented URL, if short, can be displayed in full, but long URLs are often shortened and contain the full domain name and extracted keywords from the URL address to present them in the form of breadcrumbs (Gudivada *et al.*, 2015).~~

When a user submits a query on a web search engine, whether using a manual keyboard, smartphone keypad, or their voice, they are presented with a list of results (Sachse, 2019). If the user clicks on one of the results, the search engine records this action and takes note of the click for that search result. If the user only browses the results list without clicking on any of them, the search engine only records the impression of the search results list (Taghavi *et al.*, 2012). An impression is one display of a particular set of results presented on the search engine results page. Data about clicks and impressions are stored internally by search engines (Balakrishnan *et al.*, 2016). Since search engines, especially Google, are used billions of times a day, the amount of data is very large.

By having all the data about the number of clicks and impressions, a search engine can determine the actual click-through rate for using the search engine interface. The click-through rate is the number of clicks divided by the number of impressions, expressed as a percentage (Jerath *et al.*, 2014). The click-through rate (CTR) can be calculated generally for all the number of clicks and impressions, but it is more useful to calculate it for different segments of search services. The CTR can be calculated for organic search results, which are crawled over the internet and presented based on the search engine index. The CTR can also be calculated for sponsored search results, which are created by advertisers and presented in the sponsored area of the search engine results page (Agarwal *et al.*, 2015). Additionally, CTR can be separately calculated for every search service, such as news, images, videos, and others.

CTRs ~~represent~~ show the level of interest in the presented results. It is expected that results placed at the top of the search engine results page will receive more clicks (Baye *et al.*, 2016). This assumption comes from the fact that the search engine algorithm, which is responsible for choosing and ordering the best results for a particular query, will place the most suitable results at the top (Brin and Page, 2012; Luh *et al.*, 2016). Over the years, people using search engines have become accustomed to checking only the first page of search results, rarely going to the second or later pages (Chitika Insights, 2013). ~~Estimating~~ Calculating the CTR is valuable as knowing what CTR is associated with each position from one to 10 on the first page of search results makes it possible to predict the amount of traffic that a website will receive (Glick *et al.*, 2014).

For the first two decades of using search engines, users primarily used desktop computers to submit queries and check results. In the last decade, the landscape of internet-connected devices has changed, resulting in more people using mobile devices, such as smartphones (Danovitch, 2019) or tablets (Jayroe and Wolfram, 2012) to search the internet.

Tablets have smaller screens than desktop devices, they are portable like smartphones devices, but have larger screens than typical smartphones. Users' behavior when using these three types of devices varies, and the CTR for search engine results can also vary accordingly (Song *et al.*, 2013).

Little is known about the current CTRs for Google search results. The most recent data available for analysis is from 2006 when Pass *et al.* (2006) published a data log from the AOL search engine. However, the CTR values estimated in that paper are no longer valid due to significant changes in the search results landscape, such as the integration of different search services and the use of various devices. There are some industry benchmarks published, e.g. by Statista who estimated that in 2020 the organic CTR was 33.59%[iii]. However, such data lacks detail, the collection methodology cannot be used by researchers nor is the data available for further studies.

Our study contributes to the current knowledge about CTRs in organic search results in three ways. Firstly, we developed a method for accessing real data on clicks and impressions available through Google's programmatic access. This allowed us to collect data for a specific website domain indexed in Google's search engine. Secondly, we estimated the current CTRs for the top 20 results on Google's search engine results page using data collected from 416,386 clicks, 31,648,226 impressions, and 8,861,416 daily queries over a period of 12 months in 2022. Thirdly, since search behavior differs across devices, we estimated the CTRs separately for desktops, smartphones, and tablets.

This study is organized as follows. The introduction section offers details on ~~what is found on search engine results pages,~~ how data on clicks and impressions are gathered, why they matter, and which devices are used for searching and browsing the search engine results pages. The literature review section covers ~~two types of~~ studies ~~: those~~ that use real log data from search engines and those that ~~predict~~ simulate the click-through rate. The method section explains how the data about clicks and impressions was accessed, used, and how the CTR was estimated. The next section presents the results of the CTR estimation in Google search engine, segmented by device type and period type, followed by a discussion section in which our results are compared with those of other studies. The novelty and contribution of the study is emphasized in this section. The conclusions section states the limitations of the study and provides directions for future research.

Literature review

Recent literature on estimating CTRs in organic search results focuses on three main areas: (1) using real log data from search engines to calculate CTR, (2) estimating CTR based on simulation or eye tracking studies, and (3) estimating CTR by utilizing industry benchmarks.

~~**Real log data**~~

~~The first group of~~Some studies use real log data from search engines to calculate the factual CTR. ~~CTR in organic search results refers to the percentage of users who click on a particular website link after conducting a search on a search engine.~~ Using historical data allows to determine the average CTR for a particular query or a domain name. To make valid conclusions about the CTR, the available data needs to contain the information about the clicks for all search queries. ~~However, in several~~ There are papers where data logs from search engine were released, but the information about clicks was missing. ~~These are log datasets described by Silverstein et al. (1999), Spink et al. (2001), Lempel and Moran (2003), Agichten et al. (2006), and Zhang and Moffat (2007).~~

~~Silverstein et al. (1999) analyzed a query log retrieved from the AltaVista search engine, which contained almost a billion queries spanning 43 days. Although the authors provided important insights into the distribution of queries and sessions, and correlations between query terms, they did not present any findings related to the click-through rate. The~~

~~investigation conducted by Spink et al. (2001) scrutinized a query log consisting of more than one million queries from Excite. The results of the study revealed that most queries are brief and users tend to examine only a small number of answer pages. However, the authors did not present any data regarding CTR. Lempel and Moran (2003) employed an AltaVista query log comprising approximately 7.7 million queries as part of their study aimed at enhancing search engine throughput by caching frequently accessed query results. However, the results of their study did not include any specific statistics regarding CTR.~~ Agichten *et al.* (2006) developed a model for forecasting the click-through rate in organic search results. The researchers randomly selected 3,500 queries from query logs and gathered click data from over 120,000 searches conducted over a three-week span. However, the initial function they proposed for estimation suggested a 100% chance that the result in the first position would be clicked, and the tenth result had roughly a 15% chance of being clicked. These values are no longer attainable at present. Zhang and Moffat (2007) utilized a dataset from MSN consisting of approximately 15 million queries, along with the corresponding number of clicks for each position. However, their chart only displayed relative values for the number of clicks and did not include specific CTR values. The same dataset form MSN seems to be used by Bendersky and Croft (2009) for analysis of long queries.

Pass *et al.* (2006) published a real-world data log from AOL (the search engine used was Google). Although the dataset has been removed after 3 days due to privacy issues (it was possible to connect search queries with real people), it was used in several later studies for analyzing queries and CTRs. The dataset published by Pass *et al.* (2006) showed that the first search result attracted 42.3% of all clicks. The second result accounted only for 11.92% of the total number of clicks, the result in the third position achieved 8.44% of all clicks. Later, other studies have used the same dataset. Chen *et al.* (2008) used it to detect events from search engine click data, whereas Wang et at. (2012) used it to explore queries and clicks in healthcare and information security. Schaefer and Sapi (2023) analyzed archival data from Yahoo!, released in 2010, and calculated CTR values that were quite similar to those published by Pass et al. (2006). However, the Yahoo! search engine no longer exists as it was acquired by Microsoft and merged into Bing.

~~**Simulating CTR**~~

~~Papers in the second line of research focusing on estimating CTR in organic search results used eye-tracking studies or simulation to test which results receive the highest CTR.~~ Joachims *et al.* (2017) ~~estimated~~ simulated that the first result received over 40% of the clicks, CTR of second result was around 17%, and the third results had CTR about 11%. However, such studies were conducted in controlled environments and may not accurately represent real-world click-through rates. ~~Strzelecki (2020) and Lewandowski and Kammerer (2021) provide more information on eye-tracking studies in web search engines in their systematic literature reviews. Eye-tracking studies are conducted with the use of desktop or mobile tracking devices and have shown that there is a difference in search behavior between these two device types.~~ Park and Cho (2021) ~~calculated~~ simulated a total CTR of 16.07% on desktop and 20.90% on mobile for a Korean shopping search engine. However, the study lacks information regarding the distribution of CTR for each position and does not provide a dataset for replication.

~~**Industry benchmarks**~~

~~The third method for estimating the CTR is based on utilizing industry benchmarks.~~ Some organizations publish industry benchmarks for CTR in organic search results. These benchmarks can provide a rough estimate of what a typical CTR might be for a particular keyword or search query. ~~iProspect (2008) reported that the majority of search engine users tend to select a search result from the first page of results, with 68% of users doing so, and that as many as 92% of users click on a result within the first three pages of search results.~~

~~The importance of appearing high in the search results has steadily increased over time. A clear trend was observed in the past data which revealed that in 2008, more search engine users clicked on the first page of results (68%) compared to 2006 (62%), 2004 (60%), and 2002 (48%) (iProspect, 2008). Conversely, fewer search engine users were willing to click on results past the third page in 2008 (8%) compared to 2006 (10%), 2004 (13%), and 2002 (19%) (iProspect, 2008).~~ Chitika Insights (2013) has analyzed tens of millions of online ad impressions resulting from a user being referred to a web page from Google search. From the search engine referring URL, Chitika was able to extract the position that the page was on within the prior search results page. From this, Chitika was able to measure what percentage of Google traffic comes from each position of the search results page. The first position in the search results has been found to have the click-through rate of 32.5%, whereas the second position has the click-through rate of only 17.6%. The common result from the past studies is that estimating CTR in organic search results depends on many factors, including the position of the results on the search engine results page, the relevance of the website to the search query, the query popularity (Jerath *et al.*, 2014), and user experience (Dupret and Liao, 2010).

Maillé *et al.* (2022) employed CTR for detecting and mitigating bias. However, they utilized industry benchmarks from 2017 that are no longer accessible for verification. They reported that the CTR for the first position was 36.4%, for the second position was 12.5%, and for the third position was 9.5%. Nagpal & Petersen (2021) referenced an industry benchmark that suggests the first three links on a search engine results page receive approximately 60% of all clicks, while the first page receives about 90% of clicks. However, they did not conduct their own evaluation. Di Caprio *et al.* (2022) designed a model to study the behavior of satisficing and impatient users in online search environments, and simulated CTRs for one million searches. The simulation revealed that patient users have similar CTRs to another industry benchmark where the first position has 21.7%, the second has 24.7%, and the third has 18.7%.

~~Closely related to the above reviewed three lines of research on CTR estimation are studies on CTR prediction. These are mostly conducted for online advertising and marketing purposes and do not entail organic search results. Such prediction involves estimating the likelihood of a user clicking on an advertisement or a link, based on various features and contextual information. According to Edizel et al. (2017), in order to predict potential revenue, a commercial search engine must accurately predict the probability of a user clicking on an ad for a given query. These studies are usually conducted with the use of statistical models and machine learning algorithms (Yang and Zhai, 2022), however, they are out of the scope of this study.~~

The literature review reveals that the dataset published by Pass *et al.* (2006) is the most recent real log data available to researchers for calculating CTR. Since then, no newer datasets have been made available to calculate CTR for organic search engine results pages. To fill this gap, the objective of the research presented in this paper was to provide a replicable method for accessing data from the Google search engine using programmatic access and calculating CTR from the retrieved data to show how the CTRs have changed since the last studies were published.

Method

The data used in this study was obtained from the Google search analytics service, which can be accessed through the Google Search Console (GSC) web service or the GSC application programming interface (API) (Google, 2023). GSC is a free service that enables authorized users to monitor the performance of verified websites in Google search. The service is available through APIs, which provide programmatic access to this data. The GSC web interface enables observation and downloading of four metrics (clicks, impressions, CTR, and position) along with one dimension, which can be the query, page, country, device, or date. However, this web interface has a limitation of downloading only 1000 rows of data. In contrast, the GSC API allows for the retrieval of all four metrics and all five dimensions, with a current limit of 1,000,000 rows per download.

Data selection for this study is based on a verified domain available in the ~~Google Search Console (~~GSC~~)~~. Only with administrative access to the domain name is it possible to verify its ownership or management rights. The domain can be verified through several methods, such as uploading a specific file to a host server, pasting a verification line of code into the domain source code, or adding a verification text record into the domain name system zone. Once the domain is verified, GSC begins to collect data from search results on a daily basis, labeled according to local time in California. For this study, the authors obtained authorized access to the GSC API to retrieve search data about a domain in a one-year period from January 2022 to December 2022.

Dataset

We utilized the Clusteric Search Auditor software, which includes a connector for the GSC API and a protocol previously established by Strzelecki (2019), to retrieve and collect data. GSC API authorized access was used to retrieve search data for a specific domain in monthly batches. The software permits the data to be saved in a CSV or XLSX format, and we opted for the latter. Our dataset comprises 416,386 clicks, 31,648,226 impressions, and 8,861,416 daily queries, which are broken down by daily queries. Each daily query includes search engine-reported data on the device type used, impressions, clicks, average CTR, and average ranking position for the query on a specific date. The data is structured and easily interpretable, with numerical data for position, CTR, clicks, and impressions and textual data for device and query, as well as dates in date format. There are no missing data, outliers, or other issues that may affect the validity or accuracy of the results. The data does not include any sensitive information, as personal or sensitive queries are not tracked (Erola and Castellà-Roca, 2014). The dataset comprises 12 files, one for each month, and includes the following data: date, device type, query, number of impressions, number of clicks, CTR for this query on this day, and average position on the search engine results page. The authors have made this dataset openly available on Zenodo (https://doi.org/10.5281/zenodo.7687893). All files are compressed to one archive and the size of the archive is 314.3MB. The collected dataset includes the following statistics presented in Table 1.

Table I

Statistics of the dataset

Period	Clicks	Impressions	Daily-queries	Unique queries
2022-01	50 428	3 158 994	830 151	88 974
2022-02	39 739	2 652 369	748 705	88 910
2022-03	44 492	3 044 945	826 904	92 948
2022-04	39 449	2 847 345	795 931	91 199
2022-05	39 022	2 935 569	826 670	88 105
2022-06	29 636	2 280 431	683 438	72 226
2022-07	31 764	2 515 581	691 559	70 979
2022-08	26 685	2 384 455	709 028	75 424
2022-09	26 195	2 276 715	694 573	73 744
2022-10	32 081	2 496 917	723 712	73 463
2022-11	31 511	2 713 319	727 946	76 256
2022-12	25 384	2 341 586	602 799	72 783
Total:	416 386	31 648 226	8 861 416	

Data processing

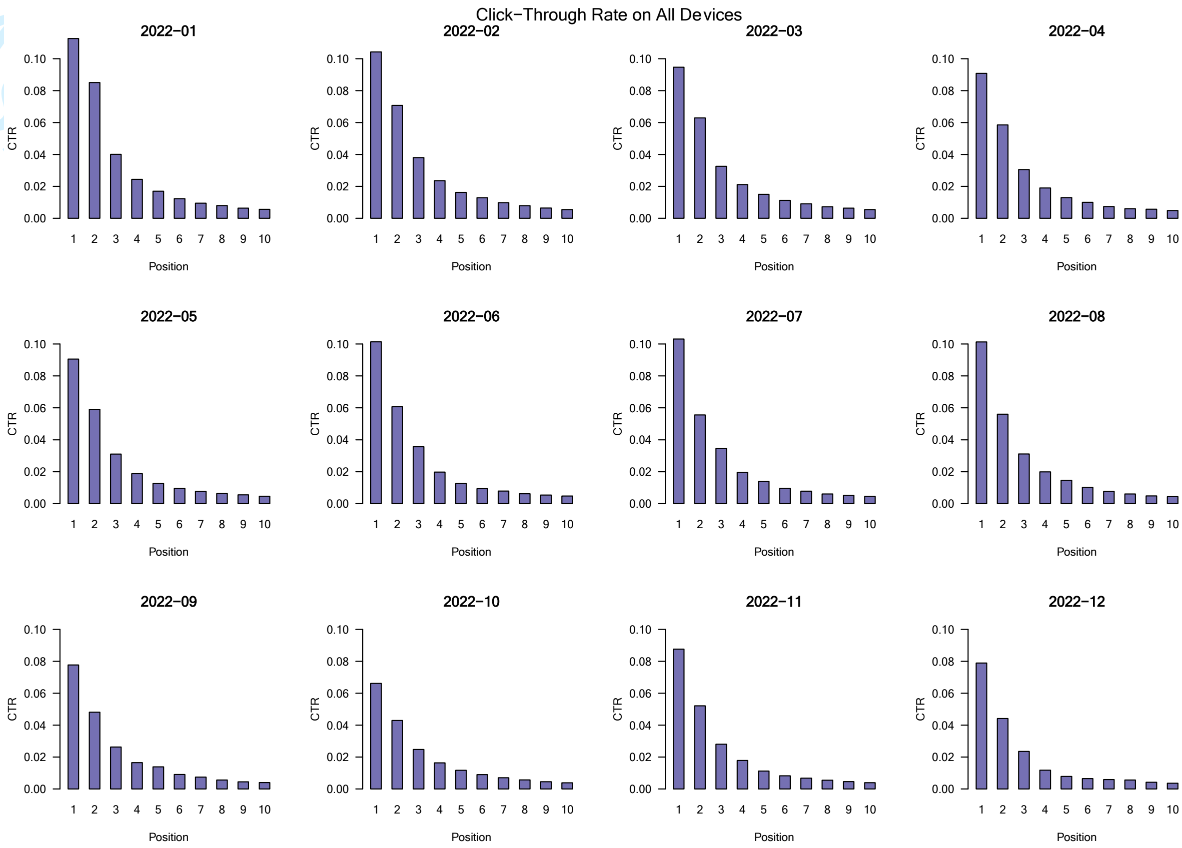
For each set of daily queries, there are several associated features, such as the number of clicks, impressions, average position in search results, and CTR for that set. To analyze the data, we created a pivot table for each time period with the following settings: rows for position, columns for the sum of clicks, sum of impressions, average CTR, and filtering for device. Using the pivot table, we obtained the total number of clicks and impressions for each corresponding position. Since the average position was recorded with two decimal places, we grouped the rows accordingly, so that positions 1.0 to 1.5 were grouped into position 1, positions 1.51 to 2.5 were grouped into position 2, and so on up to position 100. For each position, we calculated the corresponding sum of clicks and impressions. By dividing clicks by impressions, we obtained the CTR for each position. By filtering the results by device, we were able to compare CTRs for different device types, including desktop, smartphone, and tablet.

Results

The following results are presented in the bar plots, which were created in R studio (version 4.2.2). Figure 1 displays the overall CTR scores for each period and all devices. The CTR value is on the Y-axis and the position on the search engine results page is on the X-axis. CTR for the results on the first position is in the range of 6.61% to 11.26%. On the second position, CTR ranges between 4.17% to 8.51%, and on the third position, CTR is between 2.33% and 4.01%. The ranking positions of seven and below have a CTR lower than 1%.

Figure 1

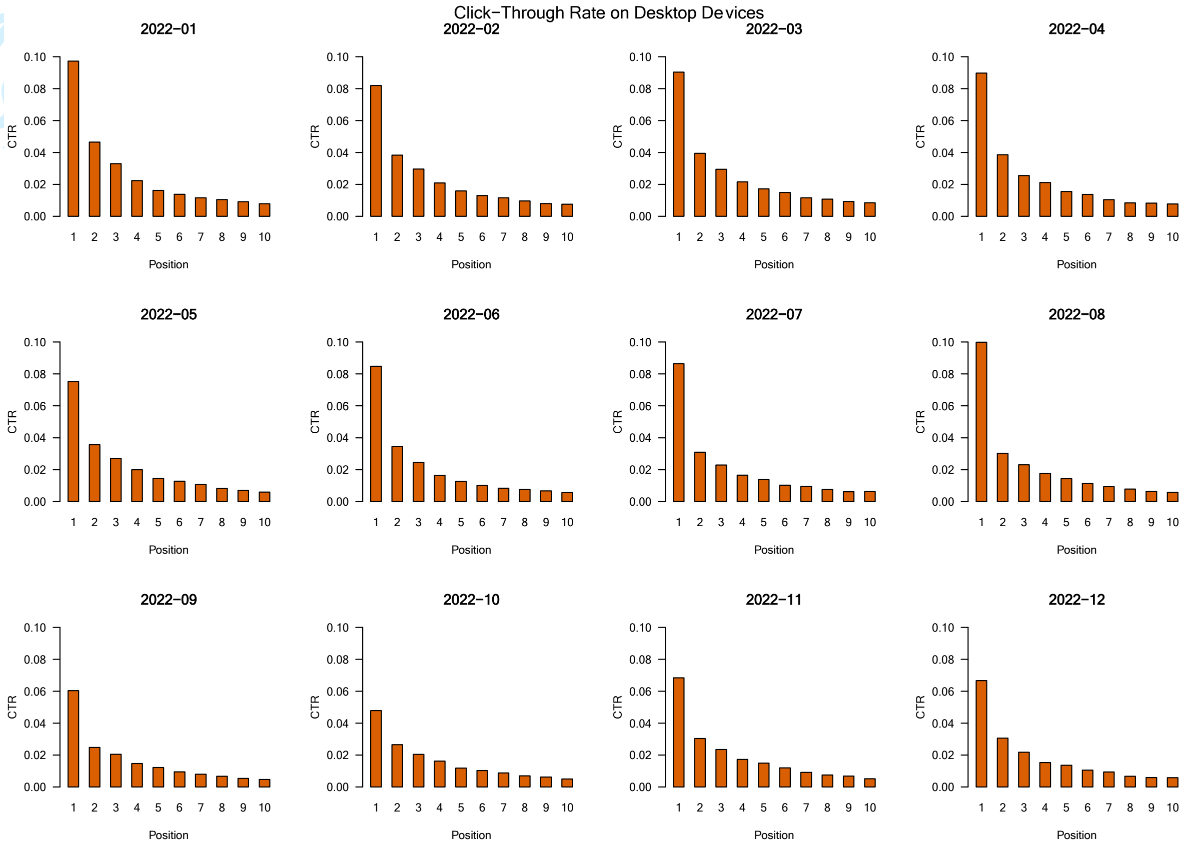
Click-through rate on all device in Google search engine for positions 1 to 10



The findings diverge when the data is divided by devices. Figure 2 showcases the CTR values only for search engine results pages viewed on desktop devices. The CTR for results on the first position ranges from 4.78% to 9.72%, while for the second position, it varies between 2.48% to 4.65%, and for the third position, it ranges from 2.04% to 3.30%. All CTR results for position seven and lower are slightly higher than the overall result for all devices.

Figure 2

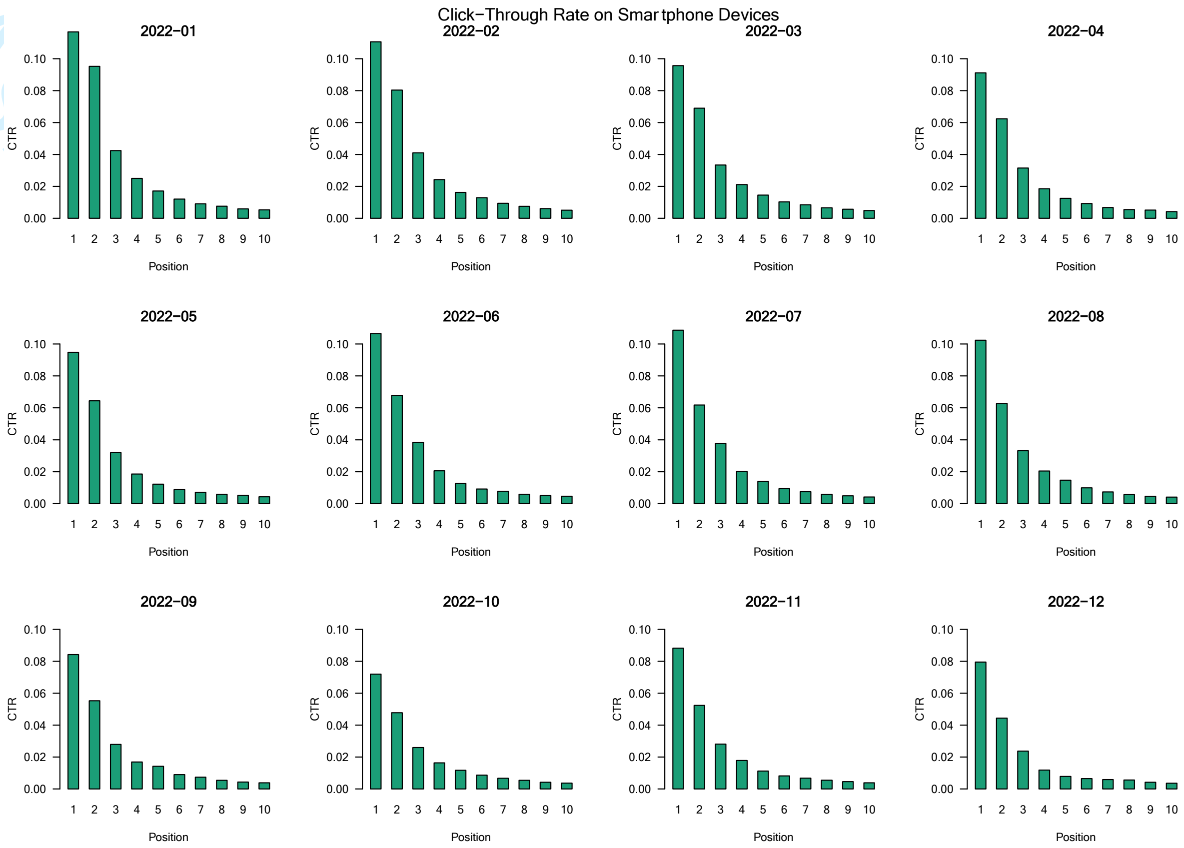
Click-through rate on desktop devices in Google search engine for positions 1 to 10



The bar plot in Figure 3 illustrates the CTR scores only for search engine results pages viewed on smartphones. The CTR for results on the first position ranges from 7.20% to 11.68%, on the second position CTR ranges between 4.44% to 9.52%, and on the third position, CTR is between 2.37% and 4.25%. The results indicate that CTR scores for positions one to four receive higher CTR on smartphones compared to desktop devices. Position number five has almost identical CTR scores on desktop and smartphones, whereas desktop devices have higher CTR scores for positions six to ten.

Figure 3

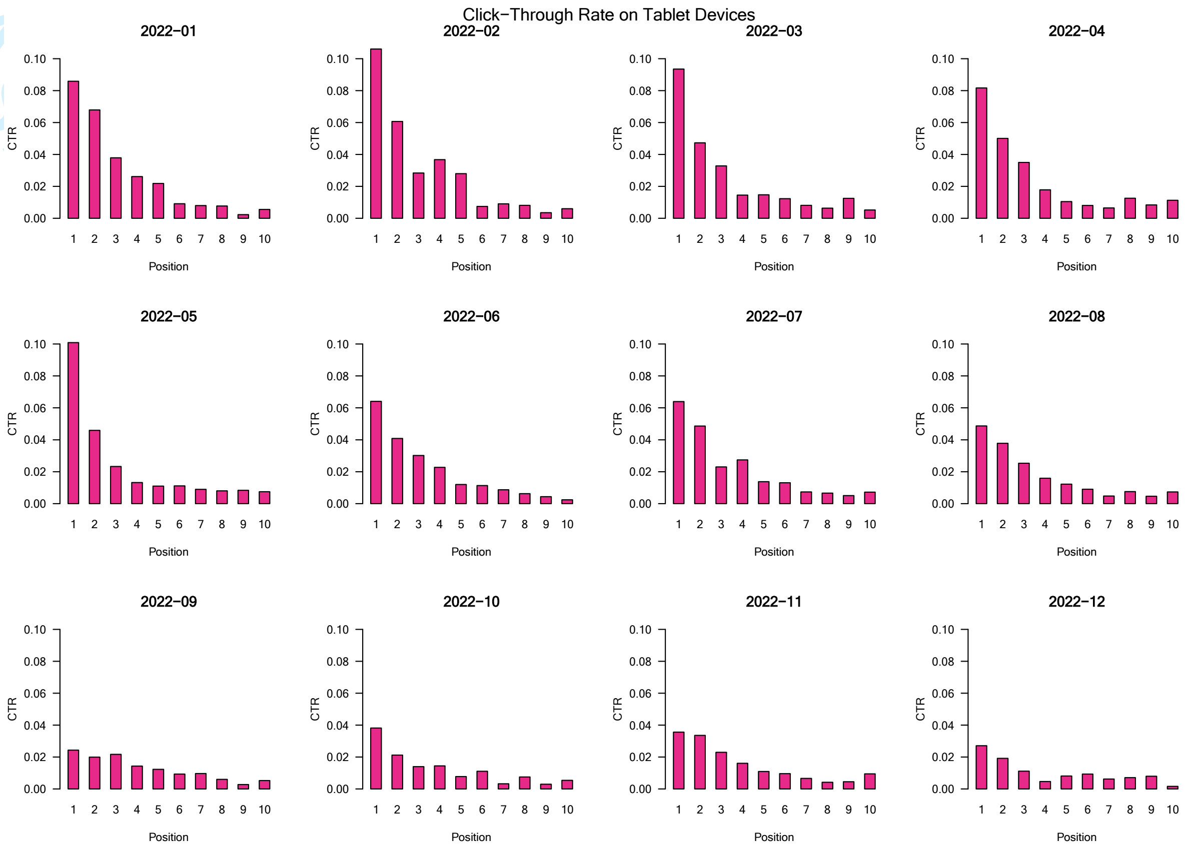
Click-through rate on smartphones in Google search engine for positions 1 to 10



In Figure 4, the CTR scores for search engine results pages viewed on tablet devices are displayed. The CTR for results on the first position ranges from 2.43% to 10.61%, on the second position between 1.92% to 6.79%, and CTR for the results in the third position is between 1.12% and 3.79%. It is observed that the CTR scores on tablet devices vary more compared to the results on other devices. Over some periods, tablets achieve higher CTR scores than desktop devices, while during other periods, CTRs are lower.

Figure 4

Click-through rate on tablets in Google search engine for positions 1 to 10



Based on the collected data, there is evidence that the CTR scores are slightly lower for the latter part of the study period. However, this decrease is observed at different periods for different devices. Additionally, an unexpected trend was observed for smartphones with respect to ranking positions 11 to 20. In contrast to desktop devices where CTR scores decrease with lower positions in ranking results, the trend on smartphones is opposite. Figure 5 illustrates that the CTR score for positions 11 to 20 in each study period is lower than 1%, and there is a decreasing trend. Conversely, on smartphones for ranking positions 11 to 20 (Figure 6), the CTR score starting at position 11 increases, reaching its highest score around positions 15 and 16.

Figure 5

Click-through rate on desktop devices in Google search engine for positions 11 to 20

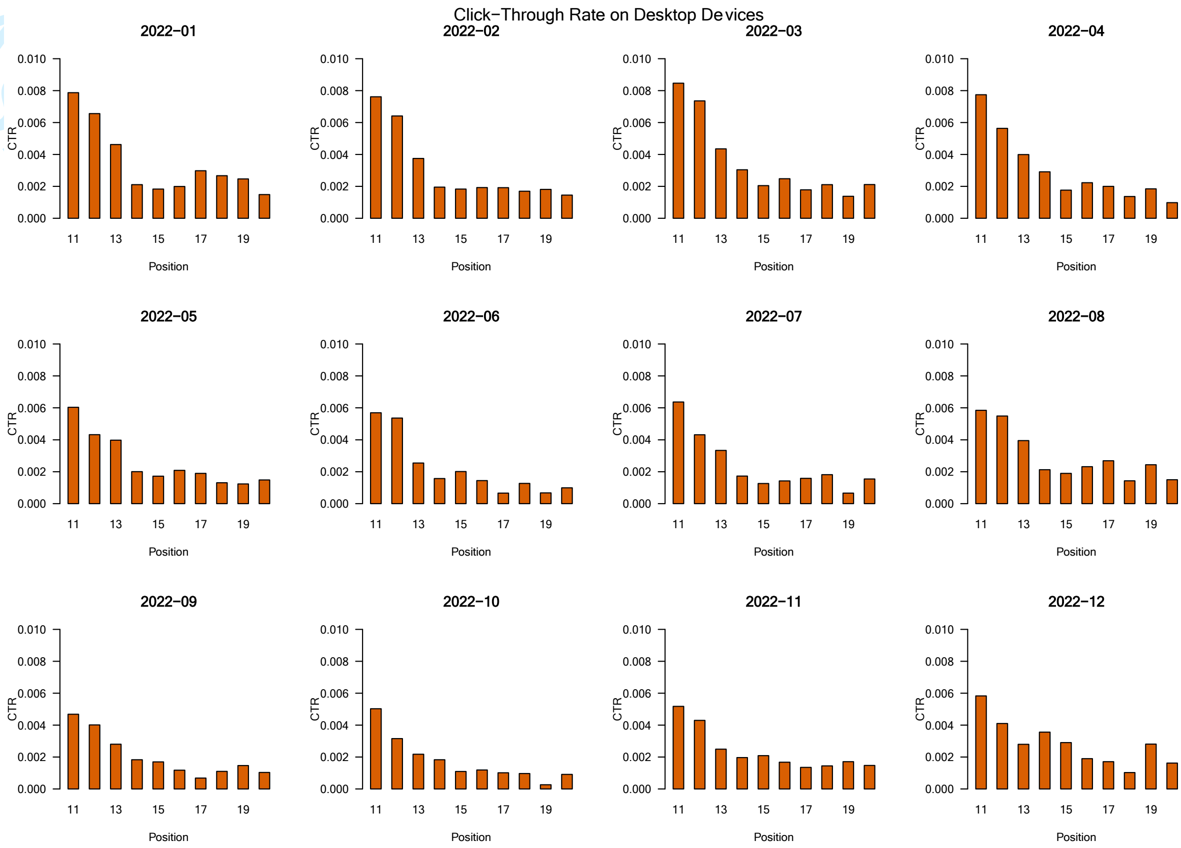
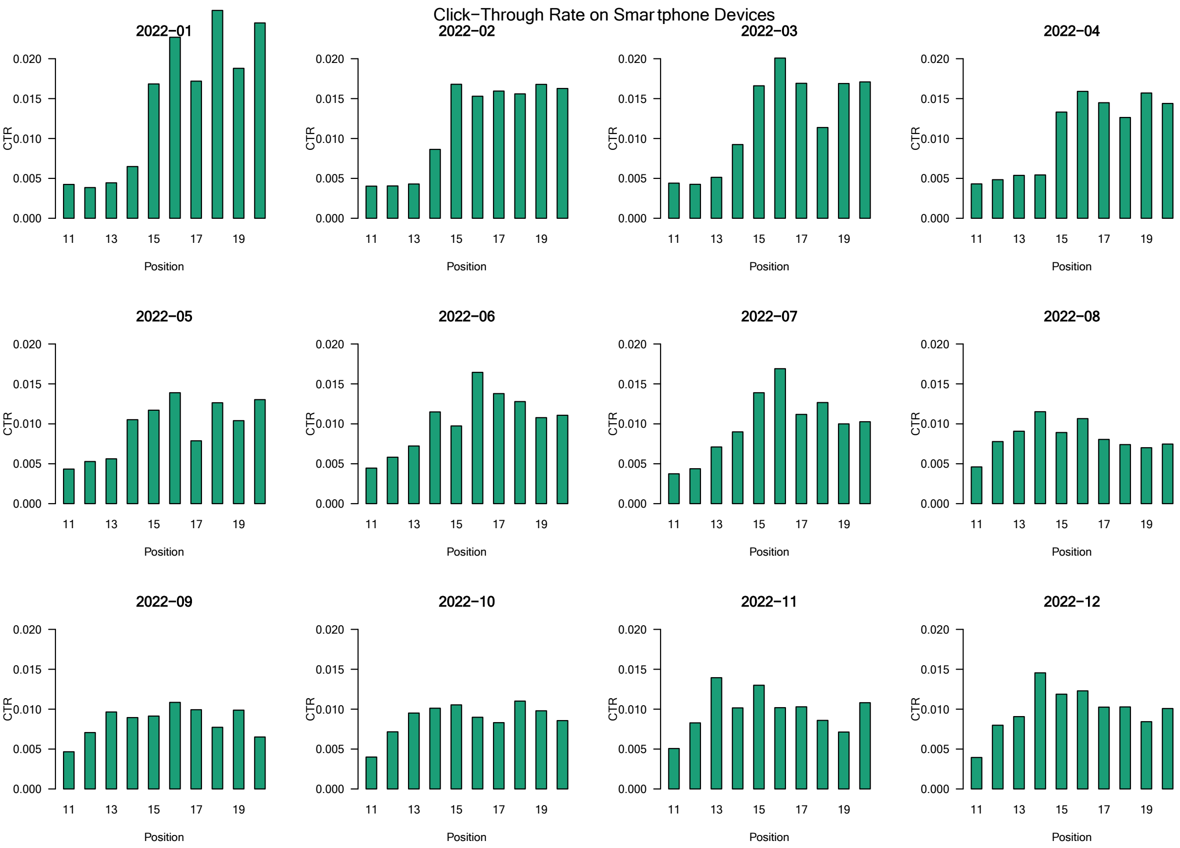


Figure 6

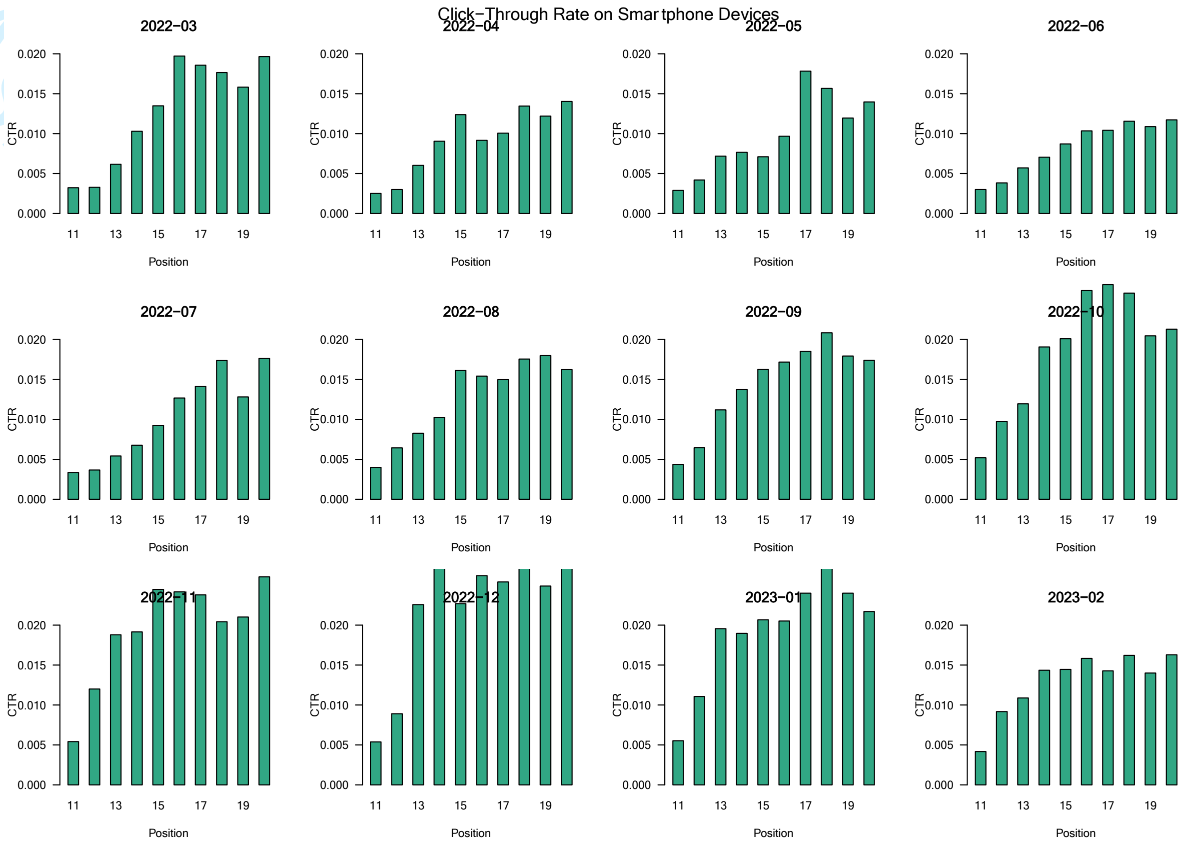
Click-through rate on smartphones in Google search engine for positions 11 to 20



To verify the observed increase in the CTR on the second page of search results and to determine whether it is an anomaly or a consistent trend, we compared the CTRs for smartphones with data generated for a second website domain. This second domain is similar to the original one in terms of the size of organic Google traffic. On smartphones, the second domain received approximately 902 thousand clicks and had 63.5 million impressions over a 12-month period. We were able to confirm the increase in CTR on the second page of search results specifically for smartphone devices. The results are presented in Figure 7. This comparison data is not included in the generated dataset made available for researchers and was only generated to validate the detected ~~anomaly/~~/trend.

Figure 7

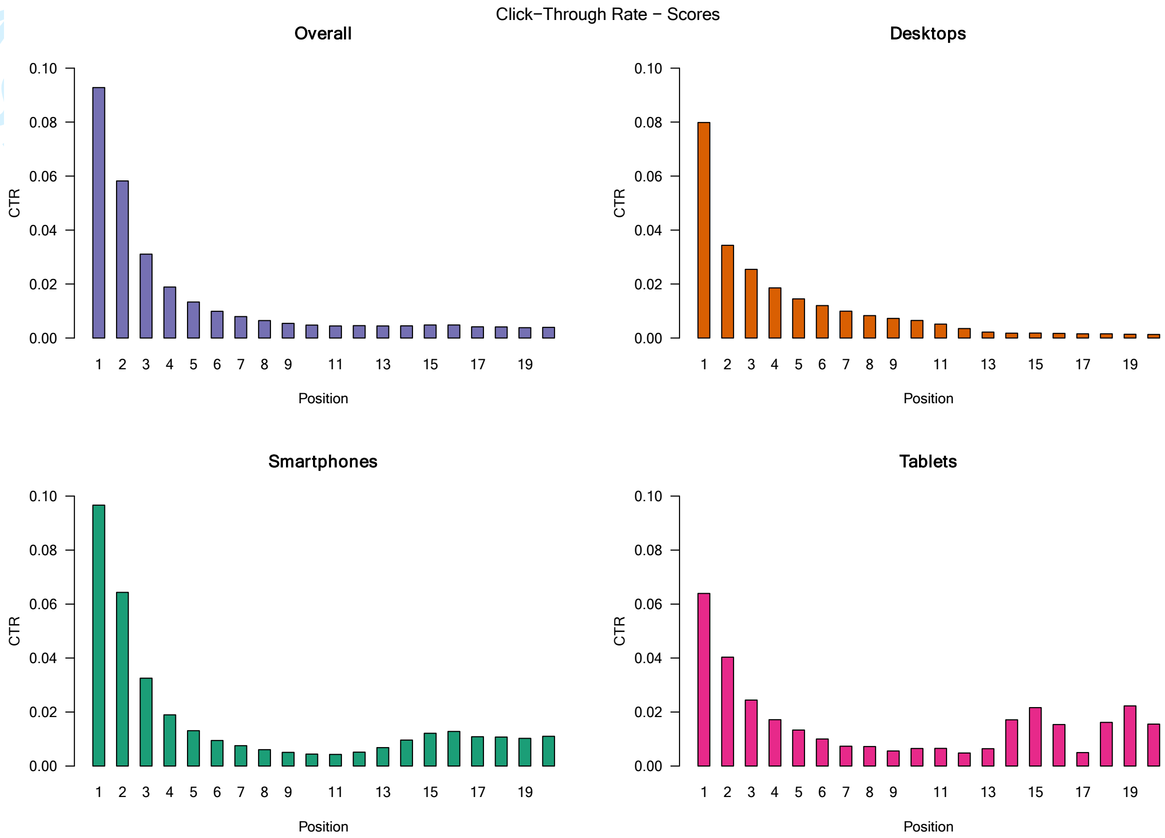
Click-through rate on smartphones in Google search engine for positions 11 to 20 – second domain name



The final results are presented in Figure 8, which includes four charts representing the average CTR scores for all devices, and for desktops, smartphones, and tablets separately. Based on these charts, the final overall CTR metrics for each of the first twenty ranking positions can be determined. The CTR for the first position is 9.28%, for the second position 5.82%, and for the third position the CTR is 3.11%. For desktop devices, the CTR decreases with each subsequent ranking position. For smartphone devices, the situation is similar up to position 11, after which the CTR increases; at first slightly and then more notably for position 14 onwards. For tablets, again, there are some increases in CTR from position 13 onwards, with some irregularities in the pattern.

Figure 8

Click-through rate – overall scores for all devices, desktops, smartphones and tables in Google search engine for positions 1~~1~~ to 20



Discussion

The created methodology was deployed to get a current dataset for a particular website that would include impressions and clicks, along with other relevant data such as date, device type, query, CTR for this query on this day, and average position on the search engine results page. The process of extracting the data has been described in detail in the Methodology section of this paper. Thanks to this, other academics and practitioners can now create their own precise datasets usable for further analysis of CTR trends and other phenomena. By creating and presenting the methodology, the authors have fulfilled the first main contribution of this research.

The results presented in this paper and the unique methodology developed for the purpose of this research have helped the authors to meet the set objectives of this study. The second contribution of this paper was to offer insights into the current click-through rates in organic Google search. These are long overdue as the previous complete dataset that was used to investigate a similar paradigm dates back to 2006 (Pass *et al.*, 2006). The determined CTRs calculated from the data presented in this paper are very different to those from 2006. For the first position, the average CTRs have dropped from 42.3% to current 9.28%. For the second position, CTRs also decreased, although the difference is lower compared to position number one – the CTR went down from 11.92% in 2006 to 5.82%. For search results on the third position, the CTR decreased from 8.44% in 2006 to current 3.11%.

The data shows that the biggest absolute difference lies in the CTR for the first position. 17 years ago, more than 4 out of 10 people clicked on the first organic text result. Currently, this is below 1 in 10 people – the CTRs for the first organic text results are less than one fourth of the value back in 2006. We can argue that the fact that other types of content and search results often take prime position on the search engine results page, such as sponsored search results, images (Jones and Oyen, 2023), videos (Wildemuth *et al.*, 2019), news (Karimi *et al.*, 2018), map results etc., is one of the main factors causing this shift. Our results can also be compared against the industry benchmark published by Statista[iii], who claim that the CTR for organic search is 33.59%. However, it is unclear how many ranking positions were considered in the calculation or what the individual CTRs for each position are. Such data is also biased as only clicks and impressions of selected users are included in the data.

Our findings also reveal the differences in CTRs between desktop computers, tablets and smartphones, which is the third contribution of this paper. The highest CTR for the first position is on smartphones, followed by desktop and tablets. On average, the CTRs are decreasing with the increasing ranking position: the better ranking (lower position), the higher CTR. However, on smartphones, a paradox has been detected where CTRs for positions 14 – 20 are much higher than those for positions 11 – 13. This is well illustrated on data presented in Figure 6. We have also confirmed the observed paradox on data from another website domain, with similar Google organic traffic than the original domain (Figure 7). The reason behind this behavior could lie in the infinite scroll of search results on mobile devices, however, this assumption needs to be confirmed by further research. Another possible explanation for this anomaly might be found in recent research presented in two independent papers. Gleason et al. (2023) examined 12 components on the search engine results page (SERP) that either amplify or attenuate CTR in Google search results. Similarly, Fubel et al. (2023) discovered that SERP features significantly influence CTR. The presence of various SERP components alongside regular organic results increases the likelihood of affecting the CTR. This could be further investigated, for example, by checking whether fewer SERP features on the second page of results cause this anomaly. Our data also confirmed that smartphone usage is larger than desktop usage. In a previous study, it was confirmed that mobile usage is higher than desktop usage. Specifically, there were more mobile sessions than desktop sessions and the number of mobile queries was more than double that of desktop queries (Park and Cho, 2021).

Theoretical implications

The theoretical implications of this work are threefold. Firstly, the provided dataset that has been made publicly available fills the gap of researchers not having exact data to analyze and draw conclusions on various aspects of organic search engine results including the CTR. The last available dataset was from 2006 (Pass *et al.*, 2006) ~~(Pass et al, 2006)~~ and since, the technology has evolved and there have been substantial changes in the search engine's algorithm, the features and results type on the search engine results page, the devices used to search the Internet, and the behavior of users when using online search. Therefore, the dataset created for the purpose of this dataset that has been made available on Zenodo (https://doi.org/10.5281/zenodo.7687893) can be used by researchers to study the current landscape of search engine results pages and user behavior in regards to CTRs and other parameters.

Secondly, the unique methodology created for the purpose of this research and described in detail makes it possible for researchers to generate new datasets that can be used for the purpose of determining CTR for various websites in various industries. This creates an opportunity for the academia to start producing results that will shed more light on the issue of CTR in different segments and their development in time. Thanks to this replicable methodology that has been described in detail in this paper, by using Google programmatic access, researchers will be able to draw fresh search engine-reported data containing the device type used, impressions, clicks, average CTR, and average ranking position for the query on a specific date. This makes our study different from previous studies, which did not provide a complete methodology enabling the generation of datasets for the academic community.

Thirdly, the provided results from the analyses of data from our dataset has shown how the CTR for organic text results decreased dramatically, thus updating the old data and results that the academic community was using from 2006. The average CTRs have dropped from 42.3% in 2006 to current 9.28% for the first search result. For the second result, CTR fell from 11.92% in 2006 to current 5.82% and for the third search result, the CTR decreased from 8.44% in 2006 to current 3.11%. This updates the knowledge base regarding the online user search behavior.

Managerial implications

There are two main implications of this study for business practice. Firstly, the results have confirmed that focusing on achieving the number one ranking for relevant keywords in organic Google text search results is not sufficient nowadays. Organizations working towards increasing their organic Google search visibility need to recognize that the number of users and sessions from organic search would have dropped by 400% over these years even if they were able to retain their top organic Google ranking. To compensate for this loss and be able to compete in today's highly competitive environment where hundreds of websites of various of organizations and individuals are competing for valuable organic Google traffic, other forms of Google search results need to be considered and optimized for. These include videos, images, products, news, maps, direct answers, etc.

Secondly, organizations and their digital marketing agencies can use the methodology created by the authors and presented in this article to determine the actual CTRs for their own websites. They can use these as key performance indicators when evaluating the results of their work on improving organic search visibility, while utilizing the data presented in this article as benchmarks to compare their results against. Knowing and using the current CTRs is crucial for marketing managers and has further implications on areas such as search engine management, online marketing in general, calculating return on investment for various marketing and sales channels or sales predictions.

Limitations

~~Conclusions~~

This work is not without limitations. By deploying the methodology created and presented in this article, one can only gather unique data for organic text results in Google and only for those websites, which they have verified in their Google Search Console. From the perspective of the data analyzed and results made available in this article, the limitation is that they only related to one particular website for which the dataset has been created. It can be expected that the CTRs may (and will) differ for other websites from the same industry and websites from other industries. In the future work, brand related searches may be filtered out from this, or similar datasets and the relative number of brand searches will have effect on the average CTRs, with brand searches attracting more clicks that regular non branded searches.

There is also ~~is~~ a limitation that relates to the lower quantity of click data for tablets for positions 11 and lower, which makes the calculated CTRs for these positions less reliable. Lastly, there are some searches that are most probably done by automated software (e.g. rank checkers) that are included in the GSC data, which was already noted in the work of Zhang and Moffat (2007). We have not attempted to identify and exclude them, which also represents a limitation of the research presented in this article.

Future work

This original work opens multiple avenues for future research. The ones most relevant for academia and business practice can be summed up as follows: i) Apply the methodology developed and presented in this article to create unique datasets for websites from the same or other industries, operating in one or more countries, to be able to compare the results in time and across industries and countries; ii) Study the differences between various types of devices deeper and explain the detected paradox of rising CTRs for search results on the 14^{th} – 20^{th} position for searches on smartphones, iii) Develop a methodology that would enable to determine what the CTRs are for other types of organic Google search results that include video, images, maps, news, products etc.; iv) Determine whether similar datasets can be created for other relevant search engines incl. Bing and Yahoo to enable for comparisons of results between various search engines.

Conclusions

Our study found that CTRs for organic search results in Google have significantly decreased from previously reported values. The current top-ranked result features a CTR of 9.28%, followed by 5.82% and 3.11% for the second and third positions, respectively. The research highlighted notable variations in CTR across different devices. Desktop devices showed a steady decrease in CTR with each lower ranking position. On smartphones, the CTR starts high but decreases rapidly, with a surprising increase from the 13th position onwards. Tablets exhibited the lowest and most variable CTR values. The study's results indicate a substantial decline in CTRs compared to data from 2006. For instance, the CTR for the first search result dropped from 42.3% in 2006 to 9.28% in the current study. The findings suggest a shift in user search behavior, potentially influenced by the prominence of other types of content (e.g., sponsored results, images, videos) on SERPs. This shift indicates a need for businesses to optimize for diverse forms of Google search results. The paper introduced a novel methodology for accessing real CTR data, allowing researchers and practitioners to generate datasets for further analysis of CTR trends and user behaviors on search engines. The research provides insights for businesses and digital marketing agencies, emphasizing the importance of not solely relying on top organic text results for visibility. It suggests a broader approach to optimizing various forms of Google search results, including videos, images, and maps.

Data availability

The data used to support the findings of this study have been deposited in the Zenodo repository (https://doi.org/10.5281/zenodo.7687893).

Declaration of Conflicting Interests

"The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article."

Funding

"This paper is an output of the research project VEGA 1/0109/23 Framework for systematisation of digital transformation in organizations with the focus on marketing and business processes."

References

Agarwal, A., Hosanagar, K. and Smith, M.D. (2015), "Do Organic Results Help or Hurt Sponsored Search Performance?", *Information Systems Research*, Vol. 26 No. 4, pp.

695–713, doi: 10.1287/isre.2015.0593.

Agichtein, E., Brill, E., Dumais, S. and Ragno, R. (2006), "Learning user interaction models for predicting web search result preferences", *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Vol. 2006, pp. 3–10, doi: 10.1145/1148170.1148175.

Balakrishnan, V., Ahmadi, K. and Ravana, S.D. (2016), "Improving retrieval relevance using users' explicit feedback", *Aslib Journal of Information Management*, Vol. 68 No. 1, pp. 76–98, doi: 10.1108/AJIM-07-2015-0106.

Baye, M.R., los Santos, B. and Wildenbeest, M.R. (2016), "Search Engine Optimization: What Drives Organic Traffic to Retail Sites?", *Journal of Economics and Management Strategy*, Blackwell Publishing Inc., Vol. 25 No. 1, pp. 6–31, doi: 10.1111/jems.12141.

Bendersky, M. and Croft, W.B. (2009), "Analysis of long queries in a large scale search log", *Proceedings of the 2009 Workshop on Web Search Click Data*, ACM, New York, NY, USA, pp. 8–14, doi: 10.1145/1507509.1507511.

Brin, S. and Page, L. (2012), "Reprint of: The anatomy of a large-scale hypertextual web search engine", *Computer Networks*, Vol. 56 No. 18, pp. 3825–3833, doi: 10.1016/j.comnet.2012.10.007.

Di Caprio, D., Santos-Arteaga, F.J. and Tavana, M. (2022), "An information retrieval benchmarking model of satisficing and impatient users' behavior in online search environments", *Expert Systems with Applications*, Vol. 191, p. 116352, doi: 10.1016/j.eswa.2021.116352.

Chen, L., Hu, Y. and Nejdl, W. (2008), "DECK: Detecting Events from Web Click-Through Data", *2008 Eighth IEEE International Conference on Data Mining*, IEEE, pp. 123–132, doi: 10.1109/ICDM.2008.78.

Chitika Insights. (2013), "The Value of Google Result Positioning", *Chitika*, Chitika.

Danovitch, J.H. (2019), “Growing up with Google: How children’s understanding and use of internet-based devices relates to cognitive development”, *Human Behavior and Emerging Technologies*, Vol. 1 No. 2, pp. 81–90, doi: 10.1002/hbe2.142.

Dupret, G. and Liao, C. (2010), “A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine”, *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ACM, New York, NY, USA, pp. 181–190, doi: 10.1145/1718487.1718510.

Erola, A. and Castellà-Roca, J. (2014), “Using Search Results to Microaggregate Query Logs Semantically”, in Garcia-Alfaro, J., Lioudakis, G., Cuppens-Boulahia, N., Foley, S. and Fitzgerald, W. (Eds.), *Data Privacy Management and Autonomous Spontaneous Security*, Springer Berlin Heidelberg, pp. 148–161, doi: 10.1007/978-3-642-54568-9_10.

Glick, M., Richards, G., Sapozhnikov, M. and Seabright, P. (2014), “How Does Ranking Affect User Choice in Online Search?”, *Review of Industrial Organization*, Vol. 45 No. 2, pp. 99–119, doi: 10.1007/s11151-014-9435-y.

Google. (2023), “Google Search Console”, available at: https://search.google.com/search-console/about (accessed 17 February 2023).

Haider, J. and Sundin, O. (2019), *Invisible Search and Online Search Engines*, Routledge, London, doi: 10.4324/9780429448546.

Halavais, A. (2017), *Search Engine Society*, John Wiley & Sons.

Jayroe, T.J. and Wolfram, D. (2012), “Internet searching, tablet technology and older adults”, *Proceedings of the American Society for Information Science and Technology*, Vol. 49 No. 1, pp. 1–3, doi: 10.1002/meet.14504901236.

Jerath, K., Ma, L. and Park, Y.-H. (2014), “Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity”, *Journal of Marketing Research*, Vol. 51 No. 4, pp. 480–486, doi: 10.1509/jmr.13.0099.

Joachims, T., Granka, L., Pan, B., Hembrooke, H. and Gay, G. (2017), "Accurately Interpreting Clickthrough Data as Implicit Feedback", *ACM SIGIR Forum*, Vol. 51 No. 1, pp. 4–11, doi: 10.1145/3130332.3130334.

Jones, S.M. and Oyen, D. (2023), "Abstract Images Have Different Levels of Retrievability Per Reverse Image Search Engine", in Karlinsky, L., Michaeli, T. and Nishino, K. (Eds.), *Computer Vision – ECCV 2022 Workshops*, Springer, Cham, pp. 203–222, doi: 10.1007/978-3-031-25085-9_12.

Karimi, M., Jannach, D. and Jugovac, M. (2018), "News recommender systems – Survey and roads ahead", *Information Processing & Management*, Vol. 54 No. 6, pp. 1203–1227, doi: 10.1016/j.ipm.2018.04.008.

Levene, M. (2010), *An Introduction to Search Engines and Web Navigation*, Wiley, doi: 10.1002/9780470874233.

Lewandowski, D. (2023), "Search Result Presentation", *Understanding Search Engines*, Springer International Publishing, Cham, pp. 137–163, doi: 10.1007/978-3-031-22789-9_7.

Luh, C.-J., Yang, S.-A. and Huang, T.-L.D. (2016), "Estimating Google's search engine ranking function from a search engine optimization perspective", *Online Information Review*, Emerald Group Publishing Ltd., Vol. 40 No. 2, pp. 239–255, doi: 10.1108/OIR-04-2015-0112.

Mager, A. (2012), "Algorithmic ideology: How capitalist society shapes search engines", *Information, Communication & Society*, Vol. 15 No. 5, pp. 769–787, doi: 10.1080/1369118X.2012.676056.

Maillé, P., Maudet, G., Simon, M. and Tuffin, B. (2022), "Are Search Engines Biased? Detecting and Reducing Bias using Meta Search Engines", *Electronic Commerce Research and Applications*, p. 101132, doi: 10.1016/j.elerap.2022.101132.

Miklosik, A., Evans, N., Zak, S. and Lipianska, J. (2019), "A framework for constructing optimisation models to increase the visibility of organizations' information in search engines", *Information Research*, Vol. 24 No. 1, p. 808.

Nagpal, M. and Petersen, J.A. (2021), "Keyword Selection Strategies in Search Engine Optimization: How Relevant is Relevance?", *Journal of Retailing*, Elsevier Ltd, Vol. 97 No. 4, pp. 746–763, doi: 10.1016/j.jretai.2020.12.002.

Park, S. and Cho, K. (2021), "Mobile vs desktop user search behaviours of the 1300K site, a Korean shopping search engine", *The Electronic Library*, Vol. 39 No. 2, pp. 239–257, doi: 10.1108/EL-09-2020-0261.

Pass, G., Chowdhury, A. and Torgeson, C. (2006), "A picture of search", *Proceedings of the 1st International Conference on Scalable Information Systems - InfoScale '06*, ACM Press, New York, New York, USA, pp. 1-es, doi: 10.1145/1146847.1146848.

Pechenick, E.A., Danforth, C.M. and Dodds, P.S. (2015), "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution", edited by Barrat, A.*PLOS ONE*, Vol. 10 No. 10, p. e0137041, doi: 10.1371/journal.pone.0137041.

Sachse, J. (2019), "The influence of snippet length on user behavior in mobile web search", *Aslib Journal of Information Management*, Emerald Publishing Limited, Vol. 71 No. 3, pp. 325–343, doi: 10.1108/AJIM-07-2018-0182.

Schaefer, M. and Sapi, G. (2023), "Complementarities in learning from data: Insights from general search", *Information Economics and Policy*, p. 101063, doi: 10.1016/j.infoecopol.2023.101063.

Song, Y., Ma, H., Wang, H. and Wang, K. (2013), "Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance", *Proceedings of the 22nd International Conference on World Wide Web*, ACM, New York, NY, USA, pp.

1201–1212, doi: 10.1145/2488388.2488493.

Strzelecki, A. (2019), "Google Web and Image Search Visibility Data for Online Store", *Data*, Vol. 4 No. 3, p. 125, doi: 10.3390/data4030125.

Taghavi, M., Patel, A., Schmidt, N., Wills, C. and Tew, Y. (2012), "An analysis of web proxy logs with query distribution pattern approach for search engines", *Computer Standards & Interfaces*, Vol. 34 No. 1, pp. 162–170, doi: 10.1016/j.csi.2011.07.001.

Wang, J., Xiao, N. and Rao, H.R. (2012), "An exploration of risk information search via a search engine: Queries and clicks in healthcare and information security", *Decision Support Systems*, Elsevier B.V., Vol. 52 No. 2, pp. 395–405, doi: 10.1016/j.dss.2011.09.006.

Wildemuth, B.M., Marchionini, G., Fu, X., Oh, J.S. and Yang, M. (2019), "The usefulness of multimedia surrogates for making relevance judgments about digital video objects", *Information Processing & Management*, Vol. 56 No. 6, p. 102091, doi: 10.1016/j.ipm.2019.102091.

Zhang, Y. and Moffat, A. (2007), "Some observations on user search behavior", *Australian Journal of Intelligent Information Processing Systems*, Vol. 9 No. 2, pp. 1–8.

Zhitomirsky-Geffet, M., Bar-Ilan, J. and Levene, M. (2016), "Testing the stability of 'wisdom of crowds' judgments of search results over time and their similarity with the search engine rankings", *Aslib Journal of Information Management*, Vol. 68 No. 4, pp. 407–427, doi: 10.1108/AJIM-10-2015-0165.

[i] https://gs.statcounter.com/search-engine-market-share
~~[ii] https://au.oberlo.com/blog/google-search-statistics~~
[iii] https://www.statista.com/statistics/1240608/google-search-ctr-worldwide