THE CURIOSITY CUP 2024
A Global SAS® Student Competition

Deciphering the Influence of Ad Rank on Customer Engagement in Search Engine Marketing using SAS®

WestCoastWizards: Mrigangka Medhi, Shimeng (Summer) Wang, and Peilin (Penny) Zhao, University of California, Los Angeles (UCLA)

ABSTRACT

This study examines the impact of ad rank on keyword auction performance within search engine marketing, utilizing a dataset provided by Yahoo. The dataset, which spans a four-month period, includes data points such as anonymized advertiser account ID, anonymized key-phrase, rank, impressions, and clicks. This information forms the foundation for analyzing the relationship between ad rank and customer engagement. In this study, we crafted and analyzed several metrics of customer engagement, including the volume of clicks generated, the number of impressions obtained, and the click-through rate (defined as the ratio of clicks to impressions). Our primary goal is to explore the relationship between enhancements in ad rank and their subsequent impact on customer engagement levels. The results of this analysis could provide valuable insights for businesses aiming to optimize their search engine advertising strategies. It is particularly relevant for those considering the benefits of increasing their ad spend to secure higher ad positions, thereby enhancing ad visibility and potentially driving better campaign performance.

INTRODUCTION

Search Engine Marketing (SEM) represents a critical part of the digital marketing world, offering an effective avenue for businesses to increase their visibility online. It works by allowing advertisers to bid on keywords that people might use on search engines like Google and Bing when they're searching for specific products or services. This bidding process enables advertisers' ads to show up along with the search results for those keywords.

Yahoo's advertising system works on an auction-based model, just like many other major search engines. In this system, advertisers bid the highest amount they're willing to pay for a click on their ads, which are linked to specific keywords they've chosen. When someone searches for one of these keywords, it starts a keyword auction where different advertisers' bids compete for spots. The search engine then ranks the ads based on several factors: how much was bid, how relevant the ad is to the keyword, the quality of the ad's content, and the quality score of the landing page. If someone clicks on one of these ads, the advertiser pays a fee, which is based on how the ad performed in terms of exposure, rankings, click-through rates, conversion rates, and the revenue it generated.

The position of an ad, especially if it's at the top of the page, can greatly affect how visible it is to consumers, considering people typically read from left to right and top to bottom. In our study, we aim to delve deeper and explore the impact of incremental changes in ad rank on key customer engagement metrics, specifically the number of clicks, the volume of impressions, and the click-through rate (calculated as the ratio of clicks to impressions). By focusing specifically on the effect of ad rank, we intend to shed light on whether securing higher positions on the page really leads to a better metric score and greater customer exposure. Our findings would support advertisers with valuable insights, helping them enhance their online presence and establish stronger connections with potential customers.

The paper is organized as follows: First, we identify and ingest the data acquired from Yahoo. This step is followed by a thorough cleaning and exploratory analysis, during which we justify data adjustments and visually scrutinize trends. The subsequent section delves into data analysis, leveraging regression models and the variables employed. Findings are then presented, examining the influence of ad rank on performance metrics, and offering insights. The final section describes the risks and assumptions to provide a clear understanding of the study's limitations. The analysis, specifically data cleaning, transformation, exploration, and development of regression models was carried out through SAS Studio in SASOnDemand for Academics.

METHODOLOGY

IDENTIFYING DATA

Our dataset is derived from the publicly available keyword auction data provided by Yahoo! (Search Marketing Advertiser Bid-Impression-Click data on Competing Keywords). It encompasses the daily bidding activities of 16,268 advertisers, with a total of 77,850,271 entries over a four-month span in 2008, distributed across 75,358 keyword categories. The dataset includes variables such as the day number (ranging from 1 to 123, each representing a consecutive day within the four-month period), anonymized advertiser account IDs, anonymized key phrases (each potentially comprising multiple keywords), the bid value, ad rank, and the count of impressions and clicks. Please note that an 'impression' refers to the number of times an ad appears on a search engine results page, and a 'click' indicates when a user actively selects the ad.

CLEANING & EXPLORING DATA

Our data cleaning process was carried out to ensure the integrity and relevance of the information for our analysis. We addressed the challenge of managing the vast volume of over 77 million records. Given the sheer size of the dataset posed significant computational demands, we initially considered filtering the data based on days to mitigate this. This approach, however, presented the risk of encountering biases since an advertiser could place bids on the same key phrase over multiple days. To ensure a more representative sample of the population, we refined our strategy by focusing on key phrases. By randomly selecting 1,000 key phrases, we were able to distill the dataset to 84,774 records. This reduction not only made the dataset more manageable but also captured a more accurate representation of the diverse bidding activities compared to filtering on specific days. We also checked the summary statistics and distribution of all columns in the filtered dataset, and the statistics were in a similar range as the original dataset.

In addition, we enriched the dataset with new variables, seeking a comprehensive analysis of how ad rank correlates with crucial performance indicators. We derived the click-through rate (CTR) by finding the ratio of clicks to impressions. In some cases, we observed the CTR exceeded 1. Subsequent desk research suggested a plausible explanation: browsers caching search results and advertisements. In such scenarios, a single ad impression is counted per search, while multiple clicks can be recorded if a user clicks on the ad more than once during a cached browser session. Therefore, as CTR greater than 1 presents an inflated picture of customer engagement, we determined to cap such values to 1. Besides, we extracted keywords from the key-phrases and generated dummy variables for all keywords present in the filtered dataset. We also calculated the number of competitors bidding on the same key-phrase on a given day, to get a sense of how highly sought after certain keywords were.

Then we conducted a preliminary exploration of the relationship between ad ranks and our specified customer engagement metrics. The trends in Figures 1 - 3 reveal a uniform right-skewed distribution, suggesting a decline in user engagement as ad rank increases. While Figure 2 exhibits a unique surge in average impressions at ranks 11 to 15, this likely occurs because users often bypass the lower-ranked ads on the bottom of the first page, preferring to start fresh at the top of the second page. These observations provide a solid foundation for the subsequent regression analysis that we will conduct subsequently. Additional plots can be found in the Appendix.

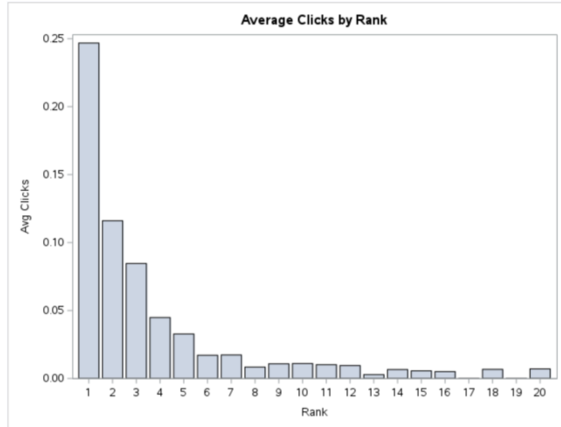


Figure 1. Dist. of Avg. Clicks by Rank

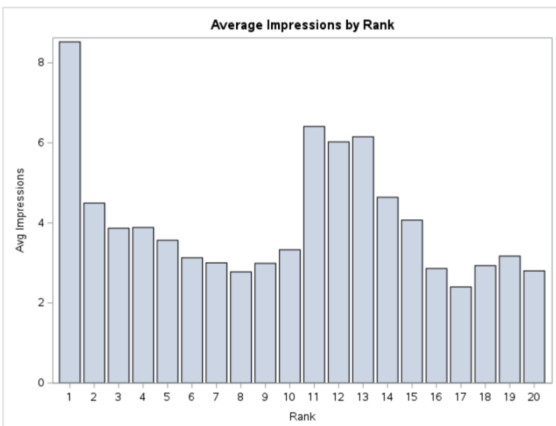


Figure 2. Dist. of Avg. Impressions by Rank

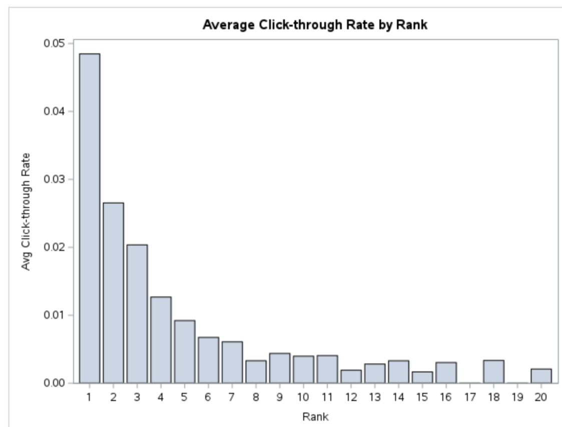


Figure 3. Dist. of Average CTR by Rank

ANALYZING DATA

To further understand the effect of ad rank on customer engagement metrics, we employed linear regression analysis, considering ad rank along with other relevant variables as independent variables. This approach enabled us to quantify the extent to which changes in ad rank can affect these key metrics.

To make a causal inference regarding the impact of rank on customer engagement, it was essential to minimize omitted variable bias. This type of bias could mistakenly attribute the indirect effects of variables correlated with rank to the customer engagement metrics as a direct effect. Therefore, instead of merely conducting a regression of the number of clicks, number of impressions, or click-through rate on rank, we introduced control variables. These variables account for time-fixed effects, such as day of the week, to adjust for temporal variations (for instance, the volume of bidding might increase on holidays, potentially affecting rankings and metrics). Additionally, we controlled for unit-fixed effects like account ID to account for distinct company characteristics, including size and brand recognition. We also included dummy variables for keywords to assess how sensitive clicks are to specific keywords that might target particular products or product segments. By accounting for these time-fixed and unit-fixed effects, we attempted to control for unobserved factors that vary over time or across accounts and products. This approach is captured in regression model 1 below.

To mitigate the potential omitted variable bias due to factors that might remain after controlling for time-fixed and unit-fixed effects, we developed regression model 2. This model incorporates the number of competitors to assess the competitive environment's influence on ad performance. In regression model 3, we introduced average bid amounts to discern whether monetary differences in bids, independent of rank, affected the number of clicks. Each control variable was selected for its ability to reveal the complex dynamics affecting keyword performance, thereby ensuring a comprehensive analysis.

Moreover, we conducted a regression treating ranks as categorical variables, not numerical, to examine the discrete impact of transitioning between adjacent ranks. This was under the premise that the change in engagement from rank 1 to 2 might differ from the change between ranks 4 and 5. Instead of evaluating the average effect of a one-rank advancement regardless of the current rank, this analysis intended to identify the particular effects at specific ranks.

RESULTS & INSIGHTS

By incorporating the number of competitors and average bid amount into models 2 and 3, the ad rank's coefficient slightly decreased, suggesting a more substantial negative impact on impressions and clicks, although the click-through rates remain pretty much the same. This refinement indicates that considering these additional factors may yield a more precise assessment of the rank's influence on customer engagement metrics (Figure 4). Specifically, the coefficient of -0.0021 for rank in the third chart indicates that, for every 1,000 impressions, an increase in rank leads, on average, to 2 fewer clicks.

Hierarchical Regression of Impressions on Rank, Competition, and Bid Amount

Independent Variables	Coefficients		
	Model 1	Model 2	Model 3
Rank	-0.3067***	-0.3203***	-0.3165***
Number of Competitors	-	0.2798***	0.2841***
Average Bid Amount	-	-	0.0005**

Hierarchical Regression of Clicks on Rank, Competition, and Bid Amount

Independent Variables	Coefficients		
	Model 1	Model 2	Model 3
Rank	-0.0116***	-0.0118***	-0.0118***
Number of Competitors	-	0.0055***	0.0055***
Average Bid Amount	-	-	4.866E-06*

Hierarchical Regression of Click-through Rates on Rank, Competition, and Bid Amount

Independent Variables	Coefficients		
	Model 1	Model 2	Model 3
Rank	-0.0021***	-0.0021***	-0.0021***
Number of Competitors	-	0.0006***	0.0006***
Average Bid Amount	-	-	7.82E-08

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Figure 4. Hierarchical Regression Model Results

Including ad ranks as categorical variables reveals distinct impacts for each rank change. A drop from rank 1 to 2 sees the most significant decrease across all metrics. The effects diminish with lower rank changes, but each transition between ranks still shows a statistically significant decrease, indicating that even small changes in rank can influence customer engagement with ads (Figure 5).

Regression of Impressions on Change in Ranks		Regression of Clicks on Change in Ranks		Regression of Click-through Rates on Change in Ranks	
Change in Rank	**Coefficients**	**Change in Rank**	**Coefficients**	**Change in Rank**	**Coefficients**
1 -> 2	-5.3255***	1 -> 2	-0.1435***	1 -> 2	-0.0194***
2 -> 3	-0.8482***	2 -> 3	-0.0338***	2 -> 3	-0.0065***
3 -> 4	-0.2632***	3 -> 4	-0.0442***	3 -> 4	-0.0077***
4 -> 5	-0.5534***	4 -> 5	-0.0122***	4 -> 5	-0.0037***
5 -> 10	-0.4435***	5 -> 10	-0.0218***	5 -> 10	-0.0058***

* p<0.1, ** p<0.05, *** p<0.01

Figure 5. Regression Model Results on Change in Ranks

RISK & ASSUMPTIONS

We extracted a smaller subset of data—1,000 randomly chosen unique key-phrases from a total of more than 77 million records—and assumed that this sample accurately represents the entire dataset. We also presumed that the sequence of keywords within bids does not significantly influence our analysis's outcomes. Although this assumption simplifies our study, it might overlook the nuanced effects that keyword order could have. Furthermore, we tried to account for all potential omitted variable biases by incorporating several control variables into our analysis, but it is recognized that there may still be unaccounted variables that could affect ad ranking, which were beyond our control. Therefore, these assumptions are important to consider, as they could potentially limit the broader applicability of our findings.

CONCLUSION

Our study investigated the relationship between ad rank and customer engagement, unveiling how ad rank impacts clicks, impressions, and click-through rates in the dynamic world of search engine marketing. By illuminating the significant influence of ad positioning on consumer behavior, our findings would empower advertisers with actionable insights to elevate ad visibility, and ultimately unlock superior campaign performance. Future research could explore the integration of machine learning algorithms to predict ad rank outcomes more accurately, considering external factors such as economic trends. Additionally, studies might also focus on long-term impacts of ad rank on brand loyalty and consumer trust, providing deeper insights into strategic advertising planning.

REFERENCES

Yahoo! Webscope dataset. "A3 - Yahoo! Search Marketing Advertiser Bid-

Impression-Click data on Competing Keywords, version 1.0 (845 MB)." Accessed February 3, 2024. https://webscope.sandbox.yahoo.com/catalog.php?datatype=a

Shi, Savannah Wei, and Dong, Xiaojing. 2014. "The Effects of Bid-Pulsing on Keyword Performance in Search Engines." Forthcoming at the International Journal of Electronic Commerce.

Agarwal, Ashish, and Mukhopadhyay, Tridas. 2016. "The Impact of Competing Ads on Click Performance in Sponsored Search." Information Systems Research.

Pal, Pallavi. 2018. "Revenue-Maximizing Number of Ads per Page in the Presence of Market Externalities." Social Science Research Network.

SAS® Viya® Platform Programming Documentation. "Controlling Bar Chart Patterns, Colors, and Images." Accessed February 19, 2024. https://documentation.sas.com/doc/en/pgmsascdc/v_048/graphref/n0w4iwk5xb39rqn17ensco5vbt19.htm

ACKNOWLEDGMENTS

We express our gratitude to Prof. Francisco Castro for his invaluable guidance and support throughout our study.

CONTACT INFORMATION

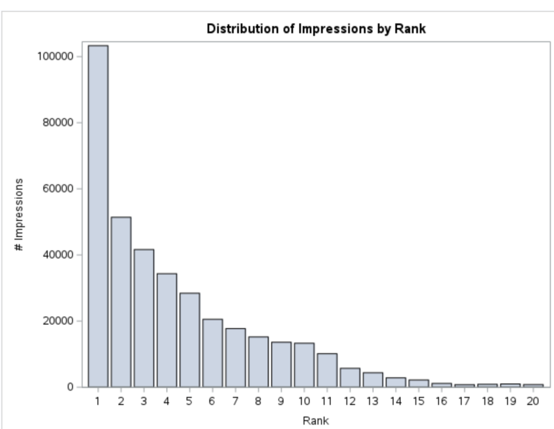
Your comments and questions are valued and encouraged. Contact the author at:

Mrigangka Medhi,
mrigangka.medhi.2024@anderson.ucla.edu

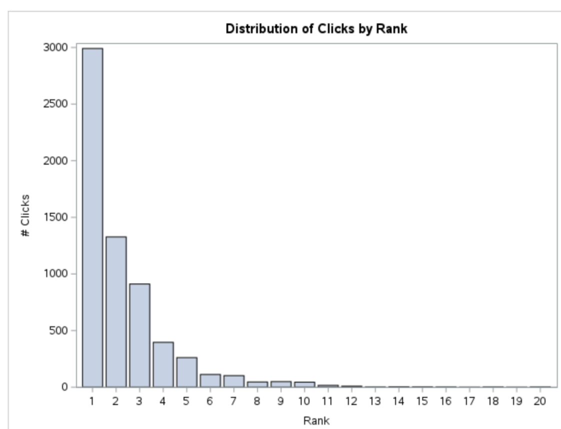
Shimeng (Summer) Wang
shimeng.wang.2024@anderson.ucla.edu

Peilin (Penny) Zhao
peilin.zhao.2024@anderson.ucla.edu

APPENDIX



A1. Dist. of Impressions by Rank



A2. Dist. of Clicks by Rank

```
/* Create macro variable that includes all independent variables */
proc sql;
select name into :ivars separated by ' '
from dictionary.columns
where libname eq 'WORK'      /*library name        */
  and memname eq 'IMPORT'  /*data set name       */
  and name    ne 'ctr'    /*exlude dep variables */ ;
  and name    ne 'impressions'    /*exlude dep variables */ ;
  and name    ne 'clicks'    /*exlude dep variables */ ;
quit;

/* Regression of impressions on rank and other control variables*/
proc reg;
  model impressions = &ivars;
run;


/* Regression of clicks on rank and other control variables*/
proc reg;
  model clicks = &ivars;
run;


/* Regression of impressions on CTR and other control variables*/
proc reg;
  model ctr = &ivars;
run;
```

A3. SAS Code for Regression Model