Assignment #2 - Bazaar

Daniel Smetana, Joseph Boykin, Trent Hannack

4/28/2021

# Executive Summary

Digital sponsored search advertisements are an integral component of Bazaar’s overall marketing strategy. As such, the management team would like to get a sense of its true return on investment in the space, with particular concern surrounding the value of advertisements that result from “branded keyword searches” (searches that directly reference Bazaar, i.e. “Bazaar shoes”) from consumers online. This specific interest in “branded” search ads stems from the intuition that within the current marketing model these ads may be redundant to some degree and in turn unnecessary expenditures. For example, it is reasonable to assume that a consumer searching for a product using the Bazaar name would in most scenarios reach the Bazaar.com site even in the absence of a sponsored banner ad being present, but at the same time might opt to utilize the sponsored ad link simply due to the ad’s placement within the browser results being more convenient than the organic site link. Based upon this reality, it is paramount for a carefully constructed ROI analysis to be run, one that accounts for all of the confounding factors rather than simply assessing the raw numbers of users who click on sponsored ads.

In order to properly estimate the effectiveness of sponsored search advertisements, this report focuses on analyzing a natural experiment that arose during a recent 12-week advertising program where an unforeseen technical glitch resulted in zero sponsored search ads being shown on Google’s search engine only for the final three weeks of the program. To produce evidence of the effect of concern, both a simple regression comparing Google results before and after the glitch and a more comprehensive “Difference-in-Differences (DiD)” regression were run. In both cases, a statistically significant drop in total traffic attributable to the pause in sponsored search ads was found, though the magnitude of the effect differed slightly across the two analyses. Leaning on the “DiD” estimate specifically (which is likely to be the more accurate of the two as it accounts for all of the other factors outside of the glitch that may have influenced the results) this report finds evidence that sponsored ads uniquely account for ~10% of all total traffic to Bazaar.com.

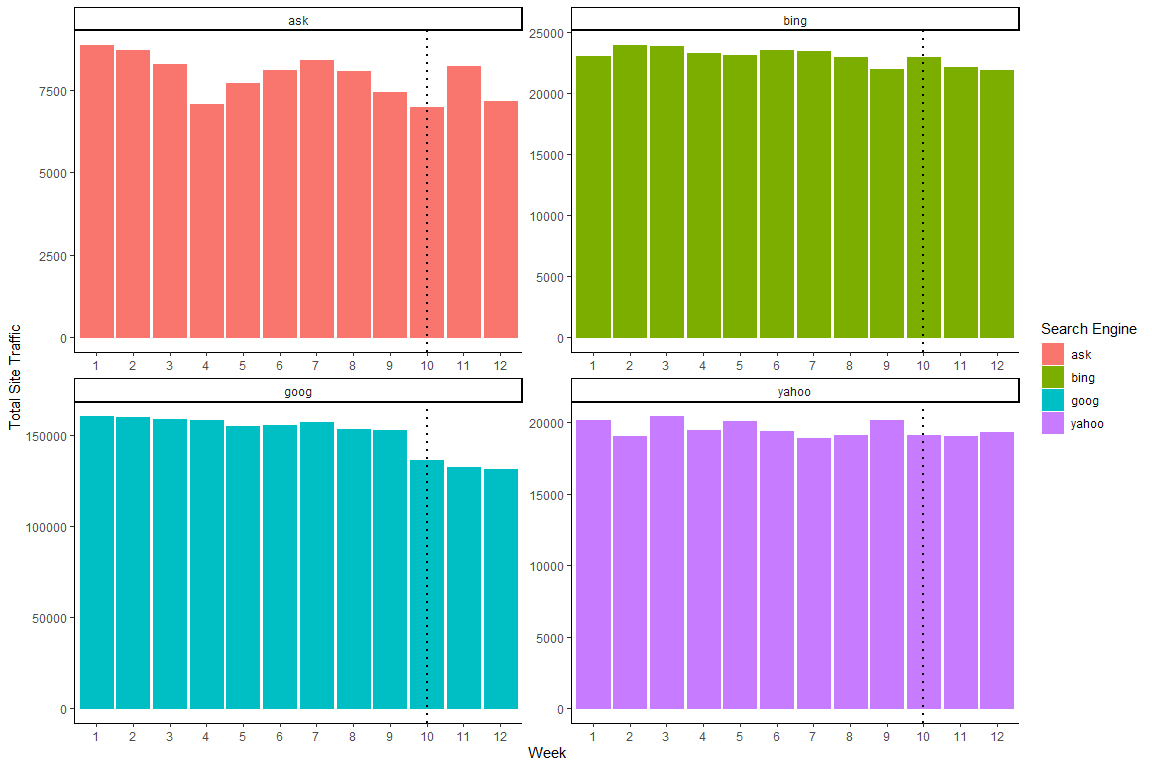
Using this estimate of the true value add of sponsored advertisements, this report calculates the actual ROI of sponsored advertisements resultant from branded searches to be approximately 125%. Although more muted than originally thought, this figure still represents a significant positive return on Bazaar’s sponsored ad spend and is furthermore a figure that can be interpreted with confidence due to the robustness of the analysis performed. Based upon this ROI calculation, the recommendation of this report is to continue utilizing sponsored advertisements as they are effective in drawing new consumers to Bazaar.com.

# Introduction

Bazaar is a leading online retailer in the United States that utilizes both display and search engine advertising to draw consumers into its site. With regards to the search engine ads in particular, the marketing team segments the ads further into two broad categories – “branded” and “nonbranded”. While the ads shown in both categories are fundamentally the same to the end consumer, the generation of these ads does differ slightly and is important to note. Specifically, branded ads appear to those consumers who search for products using the Bazaar name (i.e. “Bazaar shoes”), whereas nonbranded ads appear to consumers who simply search using a generic keyword (i.e. “shoes”) on their web browser of choice. An intuitive concern of the marketing team about the current ad spend structure is that sponsored ads shown to consumers who make branded searches, which cost the company $.60 per click if utilized, are redundant and not an effective use of the firm’s budget. The thought behind this concern is that if these customers are well aware of Bazaar’s existence and are intentionally seeking it out with their search, then they would find their way to the organic link that appears within the search results anyways at no cost to the firm. In order to address this anxiety, 12 weeks of recently collected Bazaar site traffic data tracking both organic and sponsored ad clicks for branded keyword searches via Goggle, Ask, Yahoo and Bing was analyzed.

Beneficial to answering the question of concern is that a natural experiment actually occurred within the data collected. During weeks 10-12 of the studied period a glitch systemically turned off the Bazaar sponsored banner ad displays on the Google search engine, but left the ads on the other three search engines unaffected. Due to the fact that this glitch was entirely random, and furthermore that the ads/consumer bases are fundamentally identical across each site, we can use this break in the ads to our advantage to compare site traffic before and after the event occurred in order to isolate the true effect of the sponsored ads. Once properly identified, this effect estimate can then be utilized to formulate a robust ROI calculation for the advertisements.

In using the below visualization to explore the data, it is clear that there was an observable decrease in overall site traffic through Google once the glitch occurred. The decrease shown represents a difference of ~23,398 consumers weekly, with the average number of weekly site visits through Google during Weeks 1-9 being 156,632 compared to 133,234 during Weeks 10-12. However, it is also evident that there was some noticeable variability (albeit to a lesser degree) in the traffic through Bing, Yahoo and Ask as well during the same time, in turn giving pause to simply taking the data as is when evaluating the true value of the sponsored ads. Due to this nuance, this report will explore both “simple” and “Difference-in-Differences” regressions to properly isolate how much of the decrease in site traffic through Google can truly be attributed to the disappearance of sponsored ads.



# Analysis

## What is wrong with Bob’s ROI calculation?

Core to this report is the desire to produce an accurate ROI estimate of sponsored advertisements resulting from branded web searches for Bazaar.com products. As Myra has previously alluded to, Bob’s original ROI calculation is unfortunately too simplistic to provide real value and in turn must be reevaluated. One of the main issues with Bob’s ROI estimation is that it does not account for other factors that might be driving conversion via sponsored search clicks. Namely, Bob’s calculation doesn’t account for the cases where a customer would have clicked on the organic search result and made it to the site, even if a sponsored link was not present. Intuitively, it is easy to fathom that a consumer that searched specifically for the Bazaar name would have made it to the site regardless of the sponsored ad’s presence, as they are already familiar with the brand and are actively looking for it, and that some of the sponsored ad clicks versus organic clicks are just due to the location/nature of the ad itself.

Said a different way, Bob’s calculation does not account for the correlated confound(s) likely to be present in the data. In order to analyze the true ROI, we must perform a separate Difference-in-Differences (“DiD”) analysis that eliminates any fixed differences not related to the treatment of sponsoring search ads. Such analysis will allow us to calculate the actual value add of the sponsored advertisements, which in turn can be used to calculate an accurate ROI figure

## Defining the Treatment and Control Groups of the Experiment

In order to produce the aforementioned “DiD” analysis, a natural experiment is needed. Thankfully, the technical glitch that prevented sponsored search ads from appearing during Weeks 10-12 created just that. We can use the data surrounding total site traffic before and after the glitch to produce analysis that allows us to understand what impact the treatment of sponsored ads actually had on user clicks, and ultimately, the ROI of these ads to the company.

Before running this analysis, we must first define and understand what we are trying to observe - the unit of observation. In this case, we care primarily concerned with total traffic originating from searches (both sponsored and organic) since that is what the company is trying to impact with its sponsored search ad investment. Next, it is imperative to identify which is the control group in the experiment and which is the treatment group to be able to identify which group we expect the treatment to have had some effect on. In this case, Google is the treatment group given the pause in ads after Week 9 and all other search engines that Bazaar advertises on make up the control group. Comparing the total traffic for Google with the total traffic on the other search engines will show us what effect the unintentional pause in advertising on Google had on traffic, and in turn will allow us to estimate the true lift of the sponsored ads.

## Sponsored Ads Effect

With a detailed background on the experiment having been established, we can now turn towards exploring what causal effect, if any, sponsored ads from branded keyword searches have on overall site traffic at Bazaar. One potential approach to isolate this effect is to simply compare the Google web traffic figures before and after the random systemic glitch occurred. In theory, the only tangible difference across the two time periods was the visibility of the advertisements, so logically a basic regression that compares the “before” and “after” Google data should provide good insight into the effect of interest. In following this idea, the below model was created.

MyData$after <- (MyData$week > 9)  
MyData$treat <- (MyData$search\_engine == "goog")  
  
MyData\_Google <- subset(MyData, MyData$search\_engine == "goog")

##   
## Call:  
## lm(formula = log(tot\_traffic) ~ after, data = MyData\_Google)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.024436 -0.011497 -0.003054 0.014618 0.022675   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.961524 0.005975 2001.99 < 2e-16 \*\*\*  
## afterTRUE -0.161791 0.011950 -13.54 9.32e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01792 on 10 degrees of freedom  
## Multiple R-squared: 0.9483, Adjusted R-squared: 0.9431   
## F-statistic: 183.3 on 1 and 10 DF, p-value: 9.318e-08

## (Intercept) afterTRUE   
## 1.566105e+05 -1.493813e-01

As one might expect, the above regression evidences with a high degree of confidence the existence of a causal relationship between sponsored ads being visible to prospective consumers and overall Bazaar site traffic. In interpreting the output, this simplistic model attributes nearly 14.94% of all site traffic to the existence of sponsored ads, a figure that when considered by itself provides strong support of the value of sponsored ads to the company. However, even though this model is at its core is logically sound, it is likely too simplistic to effectively produce a reasonable relationship estimate and as a result should be used with some significant caution when considered on its own.

The main fault of the above model is that it does not take into account the material differences that could have and likely did exist across the weeks of the study before and after the glitch occurred. It is very reasonable to assume that Weeks 1-9 were fundamentally different than Weeks 10-12, due to factors outside of just the presence sponsored ads, and furthermore that some of these unaccounted for factors played a real role in the number of consumers who visited the Bazaar site through Google. By not specifically accounting for these differences, the above model is compromised in that it confounds these additional effects with the effect of the sponsored ads, likely overstating the actual effect of the ads themselves. This is not to say that the above evidence of a relationship existing is in itself invalid, but rather that it is probable the magnitude of the estimate found is skewed and in turn should not be utilized directly.

In order to isolate the true size of the effect that the sponsored advertisements have on aggregate site traffic at Bazaar.com, we must augment the previous model to take into account the impactful time-variant variables that may have existed throughout the study. In order to accomplish this, the below “Difference-in-Differences (DiD)” regression was created. In practice, “DiD” models eliminate any fixed differences not related to the relationship of interest, and as a result the below model allows us to properly analyze the cause and effect relationship of the sponsored ads being removed from Google without worry of other factors skewing the results.

## Twoways effects Within Model  
##   
## Call:  
## plm(formula = log(tot\_traffic) ~ treat \* after, data = MyData,   
## effect = "twoway", model = "within", index = c("week", "search\_engine"))  
##   
## Balanced Panel: n = 12, T = 4, N = 48  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.09062701 -0.01695157 0.00041492 0.02115800 0.06878258   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## treatTRUE:afterTRUE -0.113539 0.032353 -3.5093 0.001358 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 0.078277  
## Residual Sum of Squares: 0.056524  
## R-Squared: 0.2779  
## Adj. R-Squared: -0.060579  
## F-statistic: 12.3154 on 1 and 32 DF, p-value: 0.001358

## treatTRUE:afterTRUE   
## -0.1073302

In looking at the output above, we once again see a statistically significant relationship between sponsored ads and total site traffic, confirming at a base level that some relationship does indeed exist even when considering all of the other potential factors at play. Furthermore, and liken to the first regression, the removal of sponsored ads was once again found to have a distinctly negative impact on Bazaar.com visits, validating the directionality of previous findings, as well. Where difference can be found, however, is in the magnitude of the effect estimator. The above “DiD” regression estimates the effect of sponsored ads on total traffic to be around 10.73% of all clicks, which is a 4.21% decrease when compared to the prior, more simplistic, model. Fundamentally, this result makes sense as the “DiD” model is effectively removing unobserved noise from the relationship estimator, creating a more conservative and realistic estimate of the true effect of sponsored ads on site traffic. Within this context, this estimate is likely to far better represent the actual effect sponsored ads have on site traffic and in turn would be the figure this report would recommend leaning on for business needs.

As with any analysis it is important to be confident in the results found. In order to confirm that the relationship found above is indeed real, and not just a result of coincidence within the data, the below “Placebo Test” was run. Essentially, this test creates a pseudo “after” variable (“after1” in this case) which cuts the data before the actual glitch occurred, and then regresses the data on this new variable with hopes that no relationship will be found. As can be seen below, no statistically significant effect was found, evidencing that the relationship effect discussed previously is indeed a substantial one and in turn can be used and interpreted with confidence.

#placebo test  
MyDataPre <- MyData[MyData$after == 0,]  
max(MyDataPre$week)

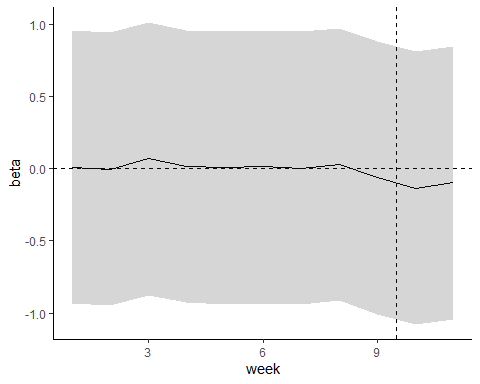
## [1] 9

MyDataPre$after1 <- MyDataPre$week > 5  
did\_placebo <- lm(log(tot\_traffic)~treat\*after1, data = MyDataPre)  
summary(did\_placebo )

##   
## Call:  
## lm(formula = log(tot\_traffic) ~ treat \* after1, data = MyDataPre)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.7892 -0.1564 0.1073 0.2993 0.4329   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9.652497 0.110836 87.088 < 2e-16 \*\*\*  
## treatTRUE 2.319702 0.221671 10.465 7.38e-12 \*\*\*  
## after1TRUE -0.018277 0.166254 -0.110 0.913   
## treatTRUE:after1TRUE -0.005741 0.332507 -0.017 0.986   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4293 on 32 degrees of freedom  
## Multiple R-squared: 0.8601, Adjusted R-squared: 0.847   
## F-statistic: 65.57 on 3 and 32 DF, p-value: 9.289e-14

Additionally important, we must also validate that the key assumptions of a “DiD” model were satisfied by the natural experiment analyzed. Without these assumptions, the effect estimate produced by the model would come with extreme caveats, leaving it largely useless to the business. The first essential assumption to the model is that no interference between the test (Google) and control (all other sites) groups occurred during the course of the studied period Logically, there is no real issue with this assumption. There is no reason to believe that a random glitch on Google would have influenced either the operations of the other sites or the habits of the consumers using the platforms, and in turn it is highly unlikely that any interference existed that would have skewed the data. The second key assumption to the “DiD” model is that “parallel trends” prevailed prior to the treatment (the glitch) taking place. In the context of this analysis, parallel trends can be defined as both the test and control groups behaving similarly prior to the glitch occurring. In order to evaluate this assumption, the below analysis and subsequent summary chart was produced. In looking at the chart specifically, we can see that the data is relatively stagnant at a beta of zero prior to the glitch (represented by the dashed vertical line in the visual). This result is consistent with what is desired to validate the parallel trends assumption, providing evidence to the fact that there was no discernable difference between the control and test groups prior to the glitch occurring. The chart below also validates that the lack of sponsored ads had a sustained negative effect on total site traffic, represented by the beta values existing below zero following the glitch. In short, when considering both assumptions key to “DiD” models, no issues can be found within the studied period and in turn the effect estimate produced by the regression can continue to be interpreted with confidence.

#dynamic did  
did\_dyn <- lm(log(tot\_traffic) ~ treat\*factor(week), data = MyData)  
  
  
# Let's retrieve the coefficients and standard errors, and create confidence intervals  
model = summary(did\_dyn)  
coefs\_ses = as.data.frame(model$coefficients[14:24,c("Estimate", "Std. Error")])  
colnames(coefs\_ses) = c("beta", "se")  
coefs\_ses = coefs\_ses %>%  
 mutate(ub70 = beta + 1\*se,  
 lb70 = beta - 1\*se,  
 week = 1:nrow(coefs\_ses))  
  
  
# Let's connect the estimates with a line and include a ribbon for the CIs.   
ggplot(coefs\_ses, aes(x = week, y = beta)) +   
 geom\_line() +   
 geom\_hline(yintercept=0,linetype="dashed") +   
 geom\_vline(xintercept= 9.5,linetype="dashed") +   
 geom\_ribbon(aes(ymin = lb70, ymax = ub70), alpha = 0.2) +   
 theme\_classic()



## New ROI Calculation

Having established that sponsored ads that appear from branded keyword searches do in fact have some positive impact on site traffic, and furthermore that they account for a tangible percentage of total Bazaar.com clicks week over week, we can now turn towards estimating the actual ROI of these advertisements based upon the data available.

In order to accurately calculate ROI, we must first isolate how many clicks actually resulted from sponsored ads. Rather than relying on the recorded figures from the studied period for sponsored ad clicks, which as mentioned previously likely over estimates the true value add, we can use the effect estimator produced by the “DiD” regression to estimate how many site visits out of the total clicks (both organic and sponsored) were truly unique to the existence of the sponsored ads. This calculated figure should give us an estimate of what the actual gain in clicks were due the sponsored ads week over week, absent of the noise inherently present in the recorded figure. This figure can then be multiplied by the average value add per click, $2.52 (based upon a 12% conversion rate of consumers and a $21 value if converted), to best estimate the real return of sponsored ads.

Conversely, when considering the cost of sponsored ads it is most accurate to use the recorded figure rather than the estimated figure. This is because even if a site visit would have happened in the absence of the sponsored ad, if a consumer clicks the sponsored ad rather than the organic link it is still a cost to the company. Ignoring this redundancy reality would result in a gross overestimate of the true return on investment, and in turn should be avoided.

A breakdown of the aforementioned analysis and the relevant figures can be seen in the below table. Of note, the true ROI on sponsored ads can be estimated to be around 125%. Although this figure is less than that of what was previously calculated, it is still an extremely strong return and would lend this report to recommend continuing the spend on sponsored ads at this time.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Recorded Average Total Clicks (Google Weeks 1-9) | Recorded Average Total Sponsored Ad Clicks (Google Weeks 1-9) | Estimated Actual Clicks Gained from Sponsored Ads (10.73% \* Recorded Total Clicks) | Sponsored Ads Cost (Recorded Total Sponsored Ad Clicks \* $.60) | Sponsored Ads Return (Total Estimated Actual Clicks Gained \* $2.52) | ROI |
| 156,633 | 31,334 | 16,812 | $18,800.40 | $42,366.24 | 125% |

# Conclusion

The primary objective of this report was to provide a robust analysis of the true effect that sponsored search ads resulting from branded keyword searches have on total site traffic at Bazaar.com, culminated through a succinct ROI calculation that could be easily interpreted by the business. Although previous discussions surrounding the ROI of sponsored ads have occurred, they have fallen short in that they were too simplistic and furthermore did not properly account for the exterior factors that may influence site traffic week over week, inherently overstating the true impact of the ads and in turn leaving a need for additional analysis.

In order to address the main points of concern, two separate analyses were completed – a “simple” regression comparing Google traffic before and after the ads were halted by the random glitch and a more comprehensive “DiD” analysis that accounted for the fixed effects present within the data studied. With regards to the “simple” regression, a statistically significant effect of sponsored ads on total site traffic was found, with the model attributing nearly 14.94% of all site traffic to sponsored advertisements. Although this figure is more reasonable than what the raw recorded data would suggest (~20%), it is still likely to be an overestimate of the true effect given that it does not account for any of the potential time variant factors which may have created inherent differences between Weeks 1-9 and Weeks 10-12 in the data. To control for these time variant variables, and as a result better estimate the real causal relationship that exists, a “DiD” regression was performed. Again with statistical significance, the “DiD” regression found the actual effect to be better estimated as 10.73% of all site traffic being attributable to sponsored ads, a decrease from the previous estimate but nonetheless still a tangible percentage of total traffic. To provide confidence in the effect shown by the “DiD” model, both a Placebo Test and furthermore an evaluation of the “DiD” model’s key assumptions were explored, with neither producing any pause for concern.

This “DiD” estimate was then used to properly calculate the ROI of sponsored ads for Bazaar, with the analysis illuminating a 125% return on the money spent. Although this figure is substantially lower than the initial estimates from previous discussions, it is still a very healthy return and in turn provides strong evidence that sponsored ads do indeed have a discernible positive impact on site visits to Bazaar.com. Given this reality, this report recommends to continue investing in sponsored search advertisements, even within the context of “branded” keyword searches, in order to reach new customers.