

# Bazaar.com

## Assessing the Effectiveness of Sponsored Search Engine Searches

*Anthony Meyers, Danny Moncada, Jack Quick, Kevin Grady, Michael Brucek*

*April 08, 2020*

### Executive Summary

The Bazaar.com sponsored search ROI is about 205%, not 320% as reported earlier. This is based on the treatment effect of having removed sponsored ad searches from Google.com. A 320% ROI calculation includes revenue generated from all ad sources and the expenses of the sponsored ad search. Estimating the true portion of clicks generated by the sponsored ads must take into account that sponsored searches and organic searches are substitute options for some users. This can be done using a difference in difference model.

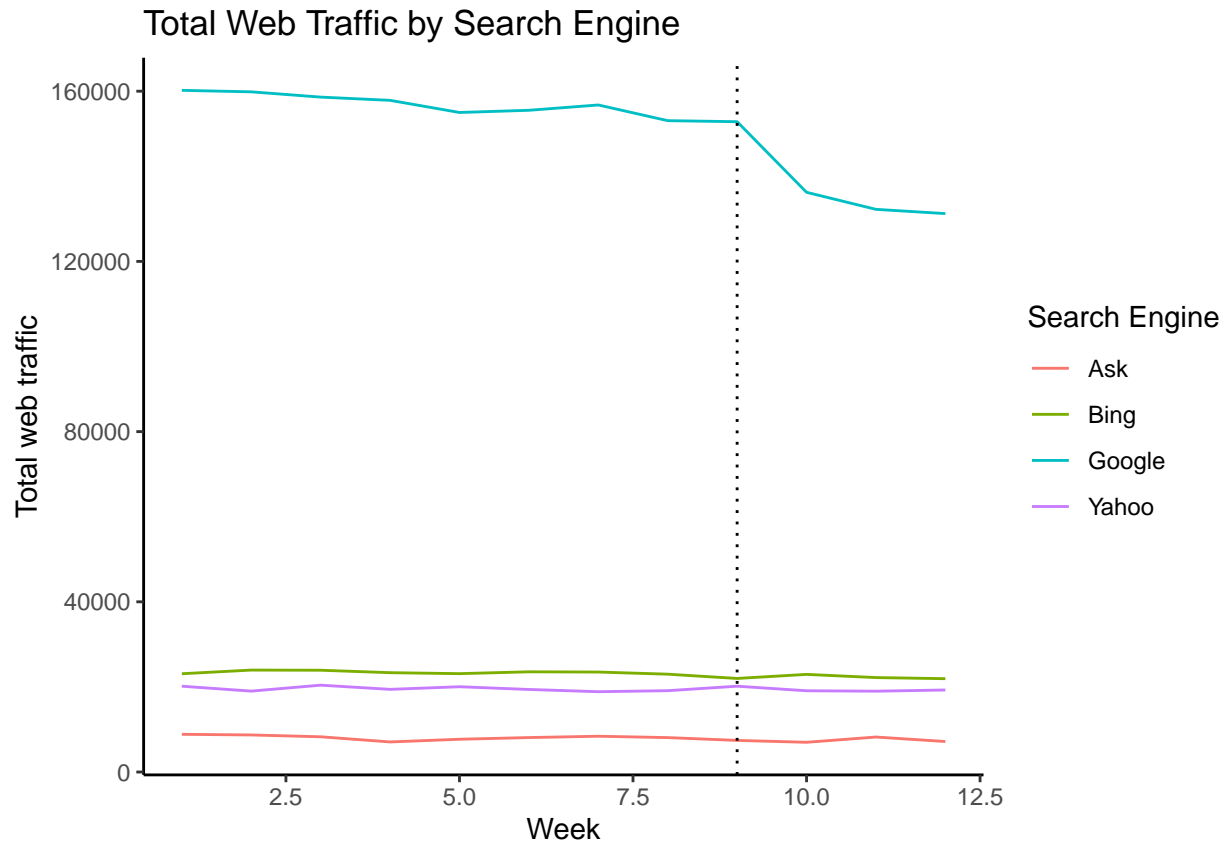
### Problem Statement

The experiment conducted by Bazaar intended to measure the effectiveness of sponsored advertisements across an assortment of search engines. Prior ad analysis provided several key findings. First, the average cost of a sponsored ad click is \$0.60. Also, the average conversion rate of customers who land on the site is approximately 12%, regardless of how they got there. Finally, the average net revenue per conversion is \$21. Bob's ROI calculation incorrectly uses these three measures to calculate ROI by overstating the revenue from users who would have clicked on the Bazaar.com result regardless of a sponsored ad, thus drastically inflating the ROI calculation.

Bazaar suspended their sponsored advertisement campaigns on the Google search engine at the beginning of the 10th week of data collection due to a technical glitch. The panel data Bazaar captured can be leveraged as an experimental design to estimate the true impact of sponsored ads. The **treatment** for this experiment is the *absence* of sponsored advertisements, while the **control** will be the continued presence of sponsored advertisements. The unit of observation are the search engines on which Bazaar is currently paying to advertise. Google is the treatment group due to the previously mentioned technical glitch which resulted in the suspension of sponsored advertisements. Yahoo!, Ask.com, and Bing are control group since sponsored advertisements were run consistently on these sites without interruption for the duration of the panel.

### First Difference Estimate

One approach for estimating post-treatment change would be to calculate the first difference. First difference would be defined as the % change in web traffic to Bazaar.com arriving from Google. This estimate is the pre vs. post difference in the treated cohort. Estimating the treatment effect could be done using a regression. However, it is unwise to solely rely on this estimate as our estimate of the treatment effect because it only quantifies the pre and post treatment differences. We need to account for the differences across the websites, as our treatment is Google and our control groups are the remaining search engines.



From the visual provided above, it is obvious that total web traffic on Bazaar.com that originated with Google decreases after the sponsored advertisement campaign is suspended after Week 9.

```
goog <- ads %>% filter(search_engine == 'goog')
goog_pre <- goog %>% filter(after_treatment == 0)
goog_post <- goog %>% filter(after_treatment == 1)

(mean(goog_post$tot_traffic) - mean(goog_pre$tot_traffic)) / mean(goog_pre$tot_traffic) * 100
```

Google-originated traffic to Bazaar's website decreased by 14.9% once the sponsored advertisement campaign was suspended after Week 9.

Using a linear model, one can see that total traffic from Google to Bazaar.com decreased by an average of about 23,399 visits per week once the sponsored campaign was suspended.

```
summary(lm1)
```

```
##
## Call:
## lm(formula = tot_traffic ~ as.factor(after_treatment), data = goog)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3802  -1723   -437    2232   3570
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      156632.7      903.8  173.30 < 2e-16 ***
## as.factor(after_treatment)1 -23398.6      1807.7  -12.94 1.43e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2712 on 10 degrees of freedom
## Multiple R-squared:  0.9437, Adjusted R-squared:  0.938
## F-statistic: 167.5 on 1 and 10 DF,  p-value: 1.429e-07
```

```
# Calculate the difference in total traffic pre & post treatment
# This is our first difference.
mean(goog_post$tot_traffic) - mean(goog_pre$tot_traffic)
```

```
## [1] -23398.57
```

It is important to estimate the value of sponsored advertisements in an experiment where the treatment group did not receive any sponsored advertisements post-treatment. Organic traffic was used to estimate the value of sponsored advertisements. Any increase or decrease in organic visits will be attributed directly to visits that would have normally come from sponsored links. The assumption is that the process that results in organic site visits *does not change* after the introduction of the treatment.

## Difference-in-Difference Model

A difference-in-difference model can be used to examine the increase in organic search traffic within the treatment period. This allows for estimation of web traffic that would have found bazaar.com through either a sponsored ad search or an organic search. The coefficient from the treatment:after\_treatment component in the model is 8,545. This represents the change in Google organic search traffic during the treatment period. In other words, this traffic would have found Bazaar.com either way and shouldn't be included in the revenue portion of the ROI calculation for paid search advertising because their purchase isn't predicated on viewing an ad. That said, the expenses incurred as a result of their clicks on a paid ad must be kept in the expense portion of the ROI calculation because Bazaar was charged for those clicks.

```
lm_org_did <- lm(avg_org ~ treatment*after_treatment, data = ads)
summary(lm_org_did)
```

```
##
## Call:
## lm(formula = avg_org ~ treatment * after_treatment, data = ads)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6244   -3212    2034    2807    3804
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13043.2       706.8  18.454 < 2e-16 ***
## treatment        112255.9      1413.6  79.413 < 2e-16 ***
## after_treatment     -610.0      1413.6  -0.432  0.66818
## treatment:after_treatment  8545.0      2827.1   3.022  0.00417 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3673 on 44 degrees of freedom
## Multiple R-squared:  0.995, Adjusted R-squared:  0.9947
## F-statistic: 2914 on 3 and 44 DF,  p-value: < 2.2e-16
```

A difference in difference regression model can be used to estimate the true number of users that were affected by the paid search program. This model considers two unique differences. First, it looks at the before and after change in total web traffic to Bazaar.com originating from Google and the control groups before and after the the treatment. This shows the impact of the Google search engine specifically. The output below shows the estimate of this effect is 22,785 visits per week.

```
lm_did <- lm(tot_traffic ~ factor(search_engine)*after_treatment, data = ads)
summary(lm_did)
```

```
##
## Call:
## lm(formula = tot_traffic ~ factor(search_engine) * after_treatment,
##     data = ads)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3801.7  -483.3    -4.1    555.6   3570.5
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)         8073.8      480.9  16.791
## factor(search_engine)bing         15203.2      680.0  22.357
## factor(search_engine)goog        148558.8      680.0 218.458
## factor(search_engine)yahoo         11551.7      680.0  16.987
## after_treatment          -612.8      961.7   -0.637
## factor(search_engine)bing:after_treatment      -303.5     1360.1   -0.223
## factor(search_engine)goog:after_treatment    -22785.7     1360.1  -16.753
## factor(search_engine)yahoo:after_treatment      110.4     1360.1    0.081
##              Pr(>|t|)
## (Intercept)          <2e-16 ***
## factor(search_engine)bing          <2e-16 ***
## factor(search_engine)goog          <2e-16 ***
## factor(search_engine)yahoo          <2e-16 ***
## after_treatment              0.528
## factor(search_engine)bing:after_treatment      0.825
## factor(search_engine)goog:after_treatment    <2e-16 ***
## factor(search_engine)yahoo:after_treatment      0.936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1443 on 40 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9994
## F-statistic: 1.128e+04 on 7 and 40 DF,  p-value: < 2.2e-16
```

## Conclusion

The original ROI calculation fails to account for the web traffic from Google that would eventually find its way to Bazaar.com. A portion of this web traffic, estimated at 8,454 clicks, would have found bazaar.com

through either paid search or organic search.

The treatment effect from the Google searches should be used to determine the true ROI for paid ads. The true number of searches driven by the paid search advertising is 22,785, per the difference-in-difference model provided in this report.

Previous analysis provided a 12% web conversion rate, an average net profit of \$21 per web conversion, a \$0.60 cost per paid search click. This analysis adds that the true average weekly amount of clicks affected by the paid search is 22,785. Also, the average number of paid search traffic clicks per week is as well as 31,333. Combining all these variables into a single ROI calculation becomes:

$$\frac{22,785 * .12 * 21}{0.6 * 31,333} = 2.054$$

The true ROI of paid ads is 205.4%