ROI Case Study

Anusha Aligeti, Kushal Acharya, Akshat Shah, Nittesh Kumar

2024-10-07

ROI Case Summary:

The case study analysis at Bazaar.com explores the effectiveness and return on investment (ROI) of running sponsored search ads for branded keywords on Google and Bing. The marketing team, consisting of Bob, Myra, and Sunil, debates whether the incremental value gained from these ads justifies the expenditure, given that branded keywords suggest that users are already aware of the company. The initial ROI calculation shows a high value of 320%, calculated based on the average cost per click, conversion probability, and profit margin. However, Myra raises concerns that these figures may overstate the true ROI, as many users might have clicked on organic links if the sponsored ads were not present, leading to the same conversions at no additional cost.

Experiment:

The experiment arose as a result of the campaign suspension, which created a natural experiment to compare traffic and conversions when sponsored ads were running (weeks 1-9) and when they were not (weeks 10-12). Bob performs an initial ROI analysis using data from the first nine weeks, which shows a high ROI of 320% for the sponsored ads. However, Myra expresses concerns that these results might not fully capture the incremental value of paid ads for branded keywords. The team suggests comparing the data from the period when ads were suspended to understand the actual impact of not running the ads. They also discuss potential biases and systematic differences between the two periods, considering variables like user behavior or external factors that could affect their findings.

(a) What is Wrong with Current RoI Calculation?

Revenue Misattribution: The current ROI calculation incorrectly attributes all revenue from users who clicked on sponsored ads to the ads, without considering that many of these users might have purchased through organic links if the ads weren't shown.

Inflated ROI: By aggregating revenue from both sponsored and organic channels, the true impact of ads is overestimated, making the ROI appear much higher than it actually is.

Skewed Decision-Making: This misattribution leads to skewed conclusions about ad effectiveness and can result in over-investment in ads, believing they generate more incremental revenue than they truly do.

The main problem with the current ROI calculation is revenue misattribution due to the aggregation of revenue from both sponsored and organic clicks. The marketing team is attributing all sales from users who clicked on sponsored ads directly to the ads, without considering that many of those users might have visited the website and made purchases through organic links, even if they hadn't seen the ad. This approach inflates the effectiveness of the ads by counting conversions that would have happened organically, regardless of the advertisement. By aggregating revenue from both sources without distinguishing between ad-driven and organic conversions, the team overestimates the true impact of sponsored ads on revenue. As a result, the ROI appears artificially high, leading to skewed conclusions about the actual return on ad spend.

Example:

Suppose 100 users visit the website through sponsored ads, and the marketing team calculates the ROI assuming all 100 users were influenced by the ad. If the conversion rate is 12%, this means 12 users are expected to make a purchase. With an average margin of $21 per customer, the total revenue is $252. The cost of the sponsored ad clicks is $0.60 per click, so the total ad spend is $60. Based on this, the ROI is calculated as:

```
ROI = ((21 * 0.12 * 100) - (100*0.6) )/ (100 * 0.6)
ROI
```

```
## [1] 3.2
```

However, if we take into account that many of these users would have visited the website organically, regardless of the ad, the true number of conversions caused by the ad is lower. For instance, if only 60 out of the 100 users visited because of the ad, and the rest would have visited organically, the actual revenue attributable to the ad would be $151.20 (12% of 60 users multiplied by $21). This means the ROI calculation should look like this:

```
ROI = ((21 * 0.12 * 60) - (100*0.6) )/ (100 * 0.6)
ROI
```

```
## [1] 1.52
```

This demonstrates how aggregating revenue from both sponsored and organic sources can significantly overestimate the true return on investment.

(b)What is the treatment and control of the experiment?

Unit of Observation: The unit of observation in this analysis is the weekly traffic for users visiting the website. This data includes metrics such as the number of users who click on sponsored ads and the overall revenue generated from those clicks, segmented by platform (Google and its competitors)

Treatment: The specific treatment being analyzed is the discontinuation of sponsored search ads on Google during weeks 10, 11, and 12.It's a natural experiment. This intervention allows researchers to evaluate the impact of ad removal on user traffic and conversions, as well as to compare these outcomes against the consistent advertising on the control platforms.

Treatment Group: The treatment group in this study is Google, where the sponsored search ads were halted during weeks 10, 11, and 12. This cessation allows for an analysis of the impact of stopping ads on overall traffic and revenue.

Control Group: The control groups consist of Bing. This platform continued their sponsored ads during the same period, providing a baseline for comparison. By observing the performance of these platforms, we can assess the relative effect of stopping ads on Google.

(c) Consider a First Difference Estimate:

```r
model <- lm(Total_Traffic ~ treatment + after, data = traffic_data)
log_model <-lm(log(Total_Traffic) ~ treatment + after, data = traffic_data)
stargazer(model,log_model,
          title = "Estimates",
          column.labels = c("Total Traffic", "Log(Total Traffic)"),
          type = "text")
```

```
##
## Estimates
## ===============================================================
##                                     Dependent variable:
##                               ---------------------------------
##                               Total_Traffic  log(Total_Traffic)
##                               Total Traffic  Log(Total Traffic)
##                                    (1)               (2)
## ---------------------------------------------------------------
## treatment                     135,134.900***      2.074***
##                                 (986.097)         (0.006)
##
## after                         -4,202.500***      -0.041***
##                                (1,045.913)        (0.006)
##
## Constant                      20,824.170***       9.888***
##                                 (779.578)         (0.005)
##
## ---------------------------------------------------------------
## Observations                        24                24
## R2                                0.999             1.000
## Adjusted R2                       0.999             1.000
## Residual Std. Error (df = 21)   2,415.434           0.014
## F Statistic (df = 2; 21)       9,398.086***    62,279.370***
## ===============================================================
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

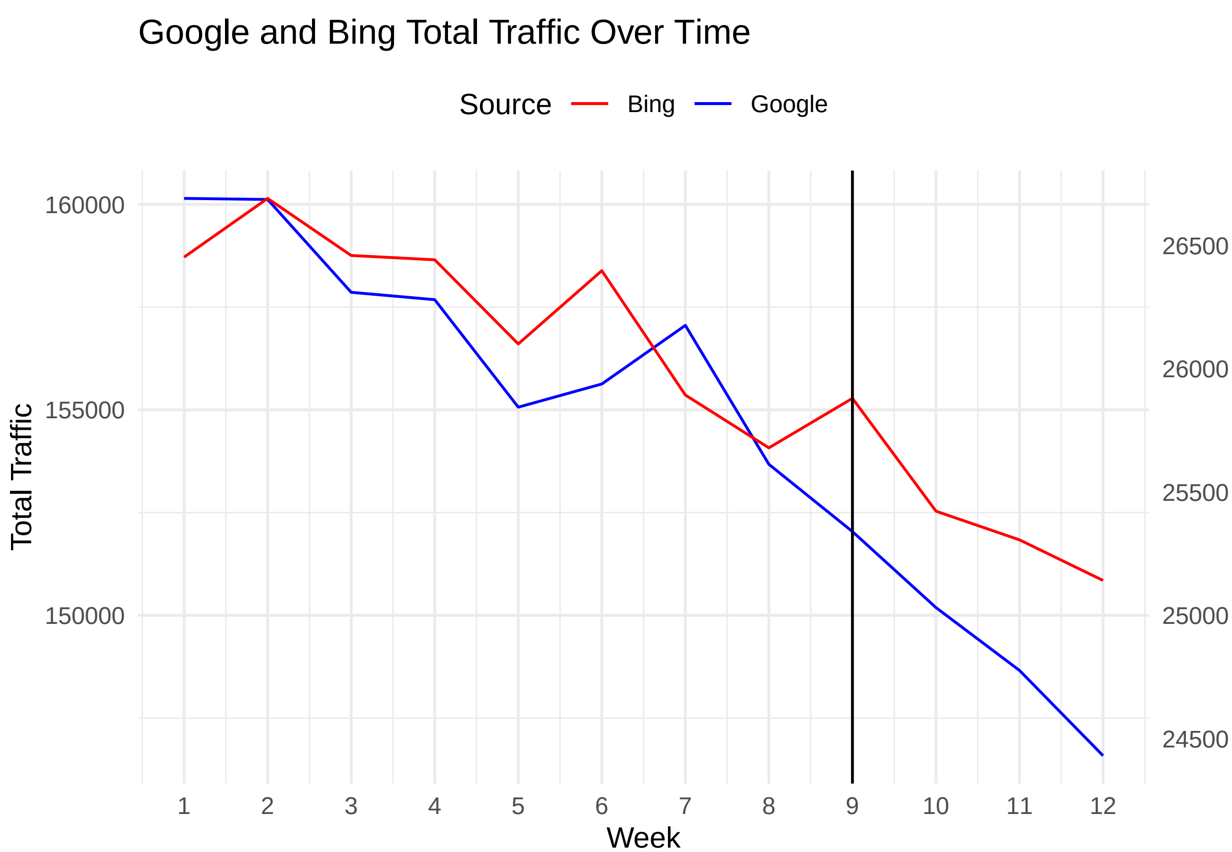
Our analysis indicates that the discontinuation of sponsored search ads on Google led to an approximate 4000 unit decline in weekly total traffic to the website. This reduction was statistically significant, with a p-value of less than 0.001, suggesting that the impact of stopping the ads is not likely a result of random variation.

The regression model focused on pre-post differences within the treated cohort, assuming that changes in traffic were solely due to the removal of sponsored ads. While this provides useful insights, it is important to recognize that this estimated treatment effect merely reflects observed differences rather than confirming a causal relationship. Relying solely on this pre-post estimate is problematic because it assumes a stable market environment and overlooks external factors that could affect traffic, such as shifts in consumer behavior, competing marketing strategies, or seasonal changes, So, A more comprehensive analytical approach that accounts for these variables would be necessary to better establish a causal relationship.

Parallel Trends

```r
scaling_factor <- max(google_data$Total_Traffic) / max(bing_data$Total_Traffic)
ggplot() +
```

```r
  geom_line(aes(x = Week, y = Sponsored + Organic, color = "Google"),data = google_data) +
  geom_line(aes(x = Week, y = (Sponsored+ Organic)*scaling_factor, color = "Bing"),data = bing_data) +
  geom_vline(xintercept = 9) +
  scale_y_continuous(sec.axis = sec_axis(~./6,)) +
  scale_x_continuous(breaks = seq(1, 12, by = 1)) +
  scale_color_manual(values = c("Google" = "blue", "Bing" = "red")) +
  labs(x = "Week", y = "Total Traffic", color = "Source",
       title = "Google and Bing Total Traffic Over Time") +
  theme_minimal() +
  theme(legend.position = "top")
```



(d)Calculate the Difference-in-Differences

Assuming that before treatment, the control and the treatment groups have parallel trends, we are doing difference in differences method:

```r
library(stargazer)
did_basic <- lm(data = traffic_data, Total_Traffic ~ treatment + after+ treatment * after)
did_log_basic <- lm(data = traffic_data, log(Total_Traffic + 1) ~ treatment + after+ treatment * after)
stargazer(did_basic, did_log_basic,
          title = "DiD Estimates",
          column.labels = c("Total Traffic", "Log(Total Traffic)"),
          type = "text")
```

```
##
## DiD Estimates
## ===================================================================
##                                       Dependent variable:
##                               -------------------------------------
##                               Total_Traffic  log(Total_Traffic + 1)
##                               Total Traffic    Log(Total Traffic)
##                                    (1)                (2)
## -------------------------------------------------------------------
## treatment                     137,524.500***       2.080***
##                                 (821.579)          (0.007)
##
## after                            -618.125         -0.032***
##                                (1,006.225)         (0.009)
##
## treatment:after               -7,168.750***        -0.019
##                                (1,423.017)         (0.012)
##
## Constant                      19,629.380***        9.885***
##                                 (580.944)          (0.005)
##
## -------------------------------------------------------------------
## Observations                       24                 24
## R2                                1.000              1.000
## Adjusted R2                       0.999              1.000
## Residual Std. Error (df = 20)   1,643.159            0.014
## F Statistic (df = 3; 20)      13,547.240***     44,364.660***
## ===================================================================
## Note:                                   *p<0.1; **p<0.05; ***p<0.01
```

The interaction between treatment and the post-treatment period indicates a decrease in effectiveness, with a reduction of about 7,168 units in traffic after the ads were implemented. This treatment effect, derived from the interaction term in the difference-in-differences regression analysis, highlights the causal impact of the absence of ads by comparing the behavior of the treatment group with that of the control group.

The new treatment effect estimate indicates that the absence of sponsored ads on Google results in an average decline of 7,168 clicks per week for Bazaar.com. In contrast, the pre-post estimate, which merely compares total traffic before and after the absence of ads, suggested a decline of about 4000 clicks per week.

(e) Fix RoI Calculation, given New Treatment Effect Estimate:

```r
did_org <- lm(Organic ~ treatment + after + treatment * after, data=traffic_data)
summary(did_org)
```

```
##
## Call:
## lm(formula = Organic ~ treatment + after + treatment * after,
##     data = traffic_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20129.8   -197.2    101.6    504.9   8421.3
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       15699.4     1890.0   8.306 6.49e-08 ***
## treatment        109940.5     2672.9  41.131  < 2e-16 ***
## after              -479.6     3273.6  -0.147  0.88498
## treatment:after   16606.5     4629.6   3.587  0.00184 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5346 on 20 degrees of freedom
## Multiple R-squared:  0.993,  Adjusted R-squared:  0.9919
## F-statistic: 941.3 on 3 and 20 DF,  p-value: < 2.2e-16
```

The original ROI calculation assumed that all traffic generated through sponsored ads was incremental, without considering the possibility that some users might have clicked on organic links if the ads were not present. This led to an overestimated ROI of 320%. To address this, a more accurate estimate of the proportion of traffic causally driven by the ads was needed.

The linear model analysis showed that in the absence of sponsored ads, 2,293 users would have used organic search results to visit Bazaar.com. After estimating the total traffic truly motivated by the ads (7169) and those who would still visit via organic search (16,606), we calculated the proportion of true traffic(7169/16606+7169) causally driven by the ads to be approximately 30.6%. Using this corrected estimate, the new ROI calculation incorporates only the true incremental traffic.

The updated ROI formula is:

(Margin per conversion * probability of click * proportion caused by ads - cost per click) / cost per click

```r
ROI = (21*0.12*0.306-0.6)/0.6
ROI
```

```
## [1] 0.2852
```

The corrected ROI of 28.5% shows the true financial return from the ads, after removing the effect of traffic that would have come in without the ads. This means that, while the ads still bring a positive return, the profitability is much lower than previously estimated. The marketing team should consider whether it's worth continuing to invest in these branded keyword ads at the same level.