Measuring ROI on Sponsored Search Ads - Bazzar.com

Ssu Hsien Lee, Tsu Jung Liu

2023-11-10

Business Problem Overview

Bazaar, the leading online retailer in the United States, employs a dual strategy involving the display of advertisements and the use of search engine advertising, running paid search ads on both Google and Bing. Bazaar classifies its ads into two broad categories based on keywords: Branded and Nonbranded. Branded keywords include terms such as "Bazaar," "Bazaar shoes," and "Bazaar guitar," while Nonbranded keywords include terms like "shoes" and "guitar" without containing the Bazaar brand name.

To evaluate the assess the efficacy of these ads, Bob, a colleague in the data analytics team at Bazaar, calculated the Return on Investment (ROI) metric. However, the reported ROI was exceptionally high at 320%, leading Myra, another colleague, to express skepticism about this figure. To comprehensively understand the causal relationship between sponsored ads and their effectiveness, we will leverage the available data to address the following questions:

1. What is Wrong with Bob's ROI Calculation?
2. Define the Treatment and Control
3. Consider a First Difference Estimate.
4. Calculate the Difference-in-Differences.
5. Given the Treatment Effect Estimate, Fix Bob's ROI Calculation.

Experiment

To examine the impact of sponsored ads, we will employ the difference-in-differences technique to estimate the causal effect observed before and after implementing the treatment.

Data pre-processing

We start by classifying the data into treatment and control groups, add pre-post indicator for weeks before and after the sponsored search ads are stopped, and combining web traffic from both organic and sponsored search ads.

```r
df <- df %>% mutate(total_traffic = avg_spons + avg_org,
                    treatment = ifelse(platform == "goog", 1, 0),
                    after = ifelse(week %in% c(10, 11, 12), 1, 0))
```

Address the Questions:

(a) What's wrong with Bob's ROI Calculation?

ROI is a valuable metric for evaluating the effectiveness and profitability of an investment. In this context, we can use ROI to inform decisions about the viability of displaying ads on Google and Bing as a strategic approach. However, it's essential to approach the calculation and the figures used for ROI with caution to prevent drawing misleading conclusions.

In the article, Bob reported a 320% ROI. If this figure is accurate, it suggests that sponsoring ads on Google and Bing constitutes a highly successful investment. However, it's important to note that such exceptionally high ROI numbers are not typical. Therefore, it's crucial to thoroughly understand the approach Bob used to calculate this figure.

Let's first recall how Bob got 320% ROI:

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

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

The error in the calculation stems from an oversight in the conversion rate, specifically the lack of differentiation. Bob's calculation assumed that all website visitors making purchases did so through the ad links. However, this assumption overlooked the fact that some customers actually arrived via organic clicks. In other words, failing to distinguish the source of customers can result in an overestimation of the ROI.

For instance, if only 30% of customers were genuinely attracted by the ads and clicked through the link, the accurate calculation of ROI would be as follows. Factoring in the actual conversion rate in this scenario, the ROI decreases to 26%, a significant deviation from the initially reported extreme figure of 320%.

```
ROI_fixed = (0.12*21*0.3-0.6) / 0.6
ROI_fixed
```

```
## [1] 0.26
```

(b) Define the Treatment and Control

To establish a causal relationship between sponsored ads and their effectiveness, it is crucial to define our treatment and control groups:

Treatment: The treatment group comprises the last three weeks (9, 10, 11) of Google sponsored search ads. This is define as the treatment period because there was no sponsorship for Google ads during these three weeks.

Control: The control group consists of Bing, Yahoo, and Ask. Bazaar continued sponsoring all ads on these three platforms during the last three weeks, making them the control group for comparison.

(c) Consider a First Difference Estimate

```r
summary(lm(total_traffic ~ after, data=filter(df, platform == "goog")))
```

```
##
## Call:
## lm(formula = total_traffic ~ after, data = filter(df, platform ==
##     "goog"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7003.9 -2630.1  -172.5  2088.4  8625.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     8390       1598   5.252 0.000373 ***
## after          -1846       3195  -0.578 0.576238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4793 on 10 degrees of freedom
## Multiple R-squared:  0.0323, Adjusted R-squared:  -0.06447
## F-statistic: 0.3337 on 1 and 10 DF,  p-value: 0.5762
```

We run a regression with 'total traffic' regressed on 'after'. The findings reveal an average web traffic decrease of 1846 after discontinuing sponsored search ads, reflecting a 22% reduction in weekly total traffic. However, this approach is not suitable because it assumes that the total traffic remains stable over time, not influenced by factors like changes in user behavior or economic conditions. This assumption can introduce potential confounding factors, compromising the accurate measurement of causal effects.

(d) Calculate the Difference-in-Differences.

Before proceeding with the difference-in-differences analysis, it is essential to validate the parallel trends assumption and argue the SUTVA assumption.

- SUTVA Assumption

The SUTVA is not violated since the treatment is specific to the treated unit and does not spill over to influence the outcomes of other units. In addition, the effect of the treatment is constant over the entire study period.

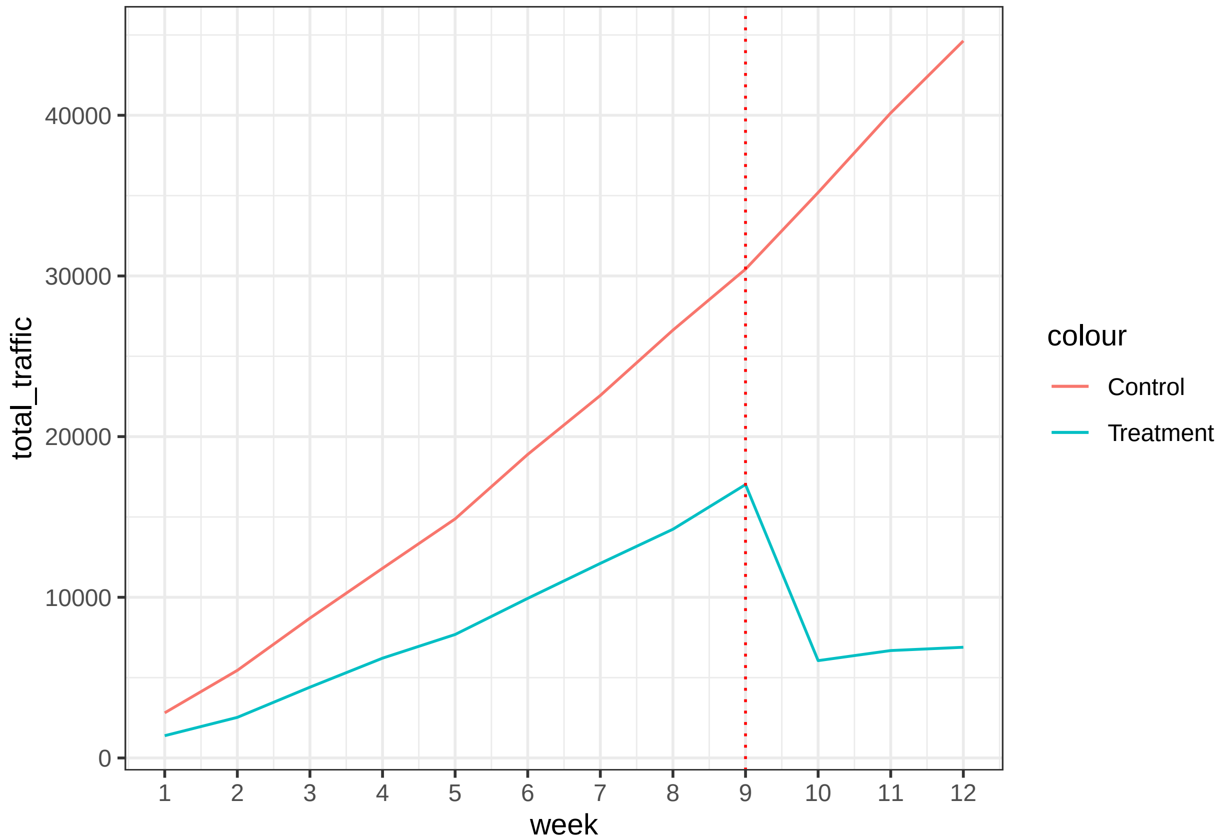
- Parallel Trends Assumption

According to this assumption, the treatment and control groups should exhibit parallel patterns during the pre-treatment week (the week before Week9).

```r
control = df %>%
  group_by(week) %>%
  filter(platform %in% c('bing', 'yahoo', 'ask')) %>%
  summarize(total_traffic = sum(total_traffic)) %>%
  ungroup()
```

```r
treatment = df %>% filter(platform %in% c('goog')) %>% select(week, total_traffic)

ggplot(control,aes(x = week, y = total_traffic , color = 'Control')) +
  geom_line() +
  geom_line(aes(x = week, y = total_traffic , color = 'Treatment'), data = treatment) +
  geom_vline(xintercept = 9 , linetype = 'dotted',color = 'red') +
  scale_x_continuous(breaks = seq(1, 12, by = 1)) +
  theme_bw()
```



From the chart above, it appears that the trends from Week1 to Week9 are relatively parallel, indicating that the parallel trends assumption is not being violated.

```r
summary(lm(total_traffic ~ treatment*factor(week), data = df))
```

```
##
## Call:
## lm(formula = total_traffic ~ treatment * factor(week), data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8710.7  -111.8    87.3  1422.3  6586.3
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    936.3     2465.2   0.380 0.707414
## treatment                      449.7     4930.3   0.091 0.928087
```

```
## factor(week)2              881.3     3486.3   0.253 0.802574
## factor(week)3             1964.7     3486.3   0.564 0.578291
## factor(week)4             2998.3     3486.3   0.860 0.398274
## factor(week)5             4023.3     3486.3   1.154 0.259840
## factor(week)6             5361.0     3486.3   1.538 0.137190
## factor(week)7             6584.7     3486.3   1.889 0.071069 .
## factor(week)8             7940.0     3486.3   2.278 0.031955 *
## factor(week)9             9204.3     3486.3   2.640 0.014337 *
## factor(week)10           10794.3     3486.3   3.096 0.004932 **
## factor(week)11           12445.3     3486.3   3.570 0.001550 **
## factor(week)12           13940.3     3486.3   3.999 0.000529 ***
## treatment:factor(week)2    259.7     6972.5   0.037 0.970600
## treatment:factor(week)3   1055.3     6972.5   0.151 0.880960
## treatment:factor(week)4   1826.7     6972.5   0.262 0.795571
## treatment:factor(week)5   2274.7     6972.5   0.326 0.747075
## treatment:factor(week)6   3187.0     6972.5   0.457 0.651723
## treatment:factor(week)7   4140.3     6972.5   0.594 0.558196
## treatment:factor(week)8   4909.0     6972.5   0.704 0.488177
## treatment:factor(week)9   6424.7     6972.5   0.921 0.365997
## treatment:factor(week)10 -6122.3     6972.5  -0.878 0.388613
## treatment:factor(week)11 -7146.3     6972.5  -1.025 0.315616
## treatment:factor(week)12 -8437.3     6972.5  -1.210 0.238030
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4270 on 24 degrees of freedom
## Multiple R-squared:  0.6819, Adjusted R-squared:  0.3771
## F-statistic: 2.237 on 23 and 24 DF,  p-value: 0.0278
```

From the parallel trend analysis above, the p-value of the interaction terms before the treatment (week 1 to week 9) are all not statistically significant and further support the parallel trends assumption. Therefore, we can proceed to conduct difference-in-differences analysis.

- Difference-in-Differences Analysis

```
summary(lm(total_traffic ~ treatment*after , data = df))
```

```
##
## Call:
## lm(formula = total_traffic ~ treatment * after, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8437.7 -3231.0  -510.5  3591.6  8630.0
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5265.0      882.5   5.966 3.79e-07 ***
## treatment         3124.9     1765.0   1.770  0.08357 .
## after             8064.7     1765.0   4.569 3.94e-05 ***
## treatment:after  -9910.6     3530.0  -2.808  0.00741 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4586 on 44 degrees of freedom
## Multiple R-squared:  0.3274, Adjusted R-squared:  0.2816
## F-statistic: 7.141 on 3 and 44 DF,  p-value: 0.0005211
```

The p-value is statistically significant and indicates that the average total traffic declined by 9910.6 without the sponsored search ads. This means that Google search ads are indeed helpful in introducing more web traffic. Compared with the first difference analysis, this approach is more robust as it compares the changes within each platform and considers the changes over time.

(e) Given the Treatment Effect Estimate, Fix Bob's ROI Calculation

The primary error in Bob's ROI calculation lies in the omission of customers who might enter the website via organic clicks. To modify this, we need to calculate the true conversion rate from sponsored ads links. According to the results from the Difference-in-Differences analysis, the total traffic decreased by 9910, indicating that sponsored ads can introduce 9910 traffic. Additionally, we conducted a linear regression to estimate the traffic from organic clicks, and the result was 2293.

```
summary(lm(avg_org ~ treatment*after , data = df))
```

```
##
## Call:
## lm(formula = avg_org ~ treatment * after, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1928.78  -847.92   -52.67   825.00  2067.33
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1489.7      215.4   6.917 1.51e-08 ***
## treatment          777.0      430.7   1.804   0.0781 .
## after             1984.1      430.7   4.607 3.49e-05 ***
## treatment:after   2293.2      861.4   2.662   0.0108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1119 on 44 degrees of freedom
## Multiple R-squared:  0.6043, Adjusted R-squared:  0.5773
## F-statistic:  22.4 on 3 and 44 DF,  p-value: 5.881e-09
```

The new ROI is 240%.

```
conversion_rate = 9910 / (9910 + 2293)
New_ROI = (0.12*21*conversion_rate-0.6) / 0.6
New_ROI
```

```
## [1] 2.410801
```