Assignment 3

Utkarsh Joshi - ID 5982808

2025-04-17

EXECUTIVE SUMMARY

Bazaar.com faced an unexpected suspension of sponsored ads on Google due to a technical glitch, while ads on Bing, Yahoo, and Ask continued unaffected. This isolated disruption offers a natural experiment to assess whether branded search ads drive incremental traffic or merely shift users from organic clicks—helping us uncover the true ROI of paid search advertising.

This report analyzes the causal impact of branded sponsored search advertising at Bazaar.com, following a technical failure that suspended Google ads for three weeks. We leverage this natural experiment to isolate the value generated by paid ads using a Difference-in-Differences (DiD) approach.

Comparing Google to Bing, Yahoo, and Ask, we estimate a 99.9% drop in sponsored traffic on Google during the suspension. This indicates that nearly all sponsored clicks were driven by ad visibility rather than underlying demand. When we recalculate the campaign's Return on Investment (ROI) using this causal estimate, the previously reported +320% ROI drops to –63.5%, suggesting paid ads cannibalized free organic visits rather than added value.

These findings underscore the need for causal evaluation methods—rather than relying solely on surface-level metrics—when assessing marketing effectiveness. We recommend that Bazaar.com incorporate controlled experiments or lift-based studies into future evaluations of branded search advertising to ensure ad spend drives true incremental value.

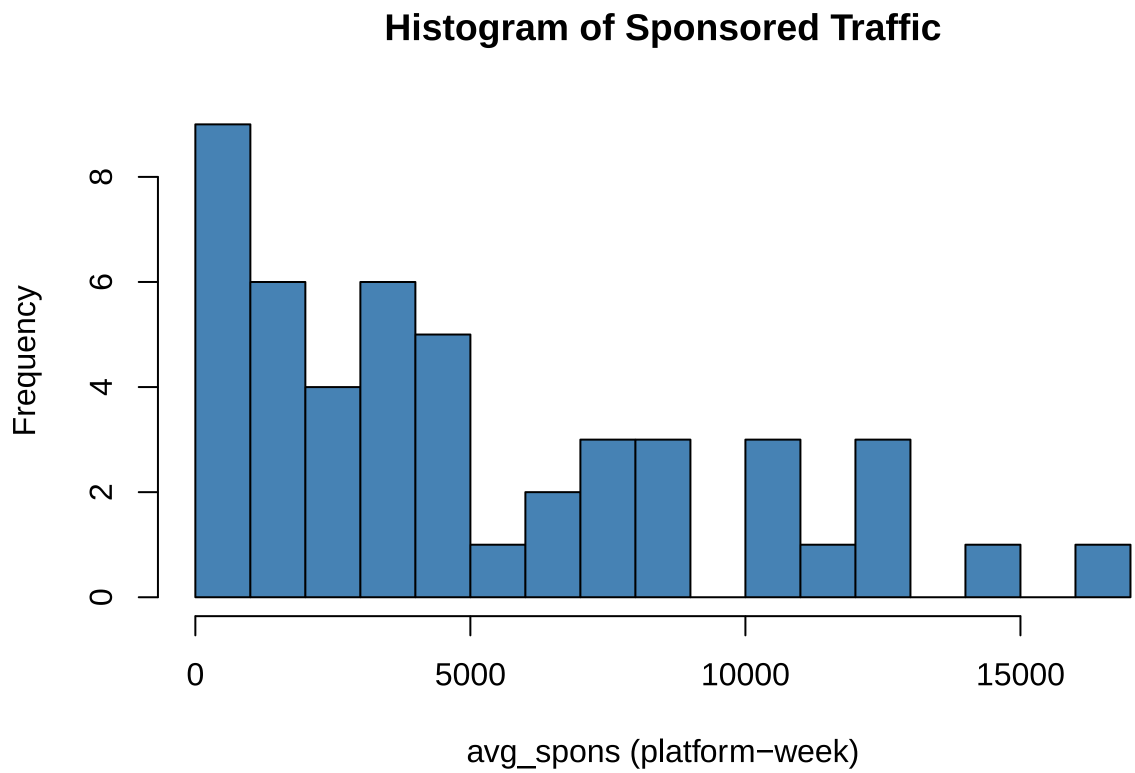
Initial Data Exploration

The dataset captures weekly sponsored and organic traffic for branded keyword searches across four platforms: Google, Bing, Yahoo, and Ask, from Week 1 to Week 12. Each observation represents a platform-week.

We start by examining:

1. Distribution of sponsored traffic:

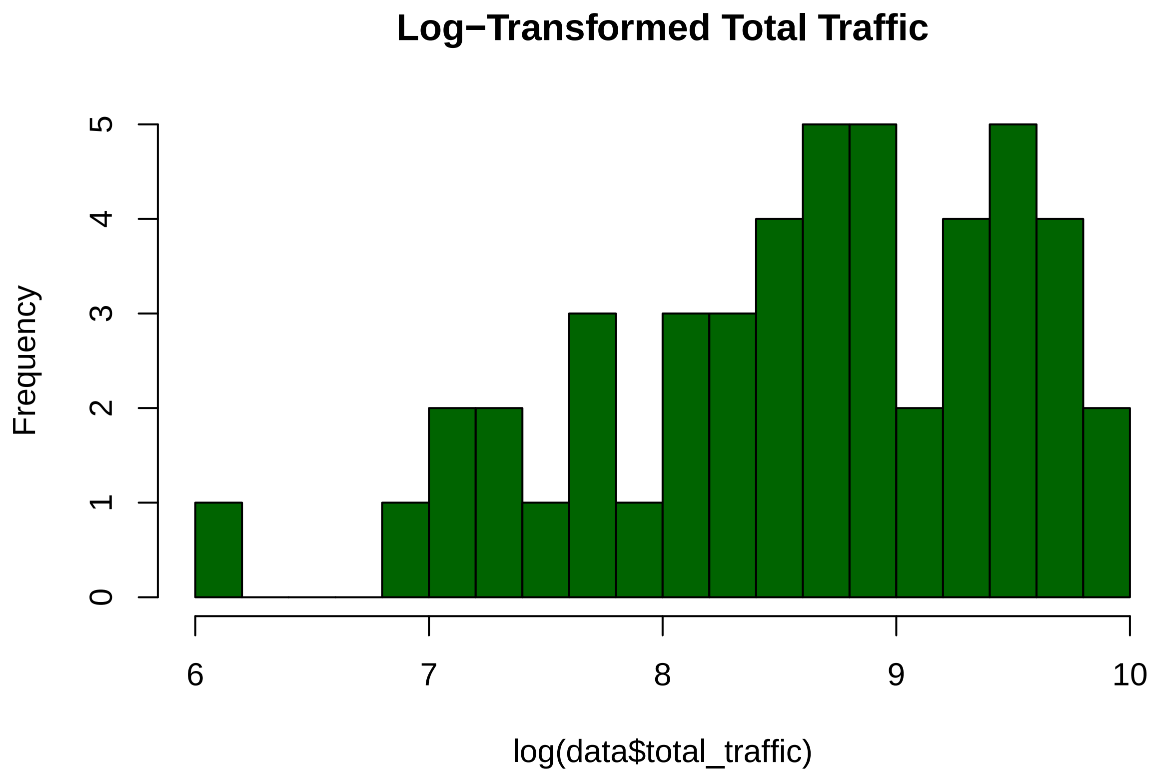
```r
hist(data$avg_spons,
     col = "steelblue", breaks = 20,
     main = "Histogram of Sponsored Traffic",
     xlab = "avg_spons (platform-week)")
```



Observation: The distribution is highly right-skewed, with most traffic values under 10,000 and a few exceeding 15,000. This validates the need for a log transformation prior to regression modeling.

2. Construct total_traffic = avg_spons + avg_org to capture total site visits and observe similar skew:

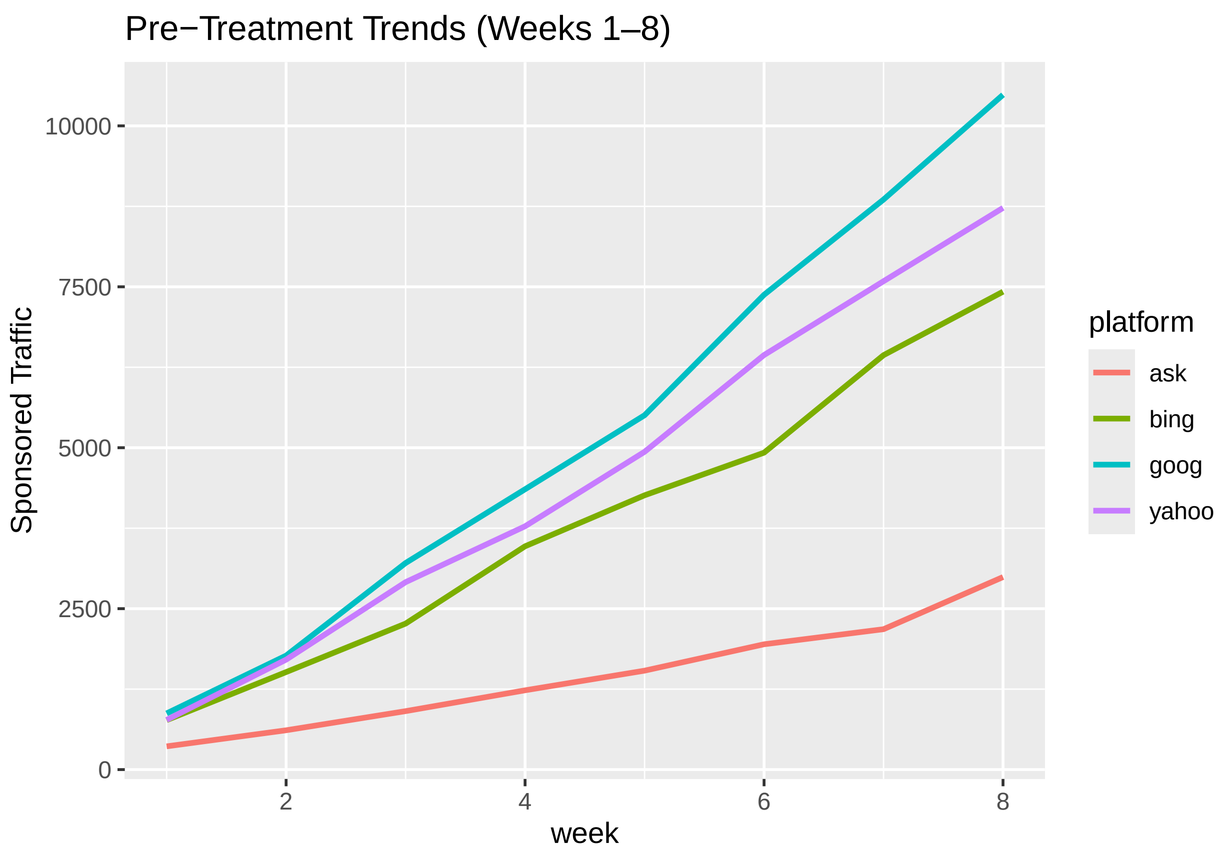
```r
data$total_traffic <- data$avg_spons + data$avg_org
hist(log(data$total_traffic),
     col = "darkgreen", breaks = 20,
     main = "Log-Transformed Total Traffic")
```



3. Pre-Treatment Trends (Weeks 1–8) across platforms:

```
ggplot(data %>% filter(week < 9),
       aes(x = week, y = avg_spons, color = platform)) +
  geom_line(size = 1) +
  labs(title = "Pre-Treatment Trends (Weeks 1-8)", y = "Sponsored Traffic")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

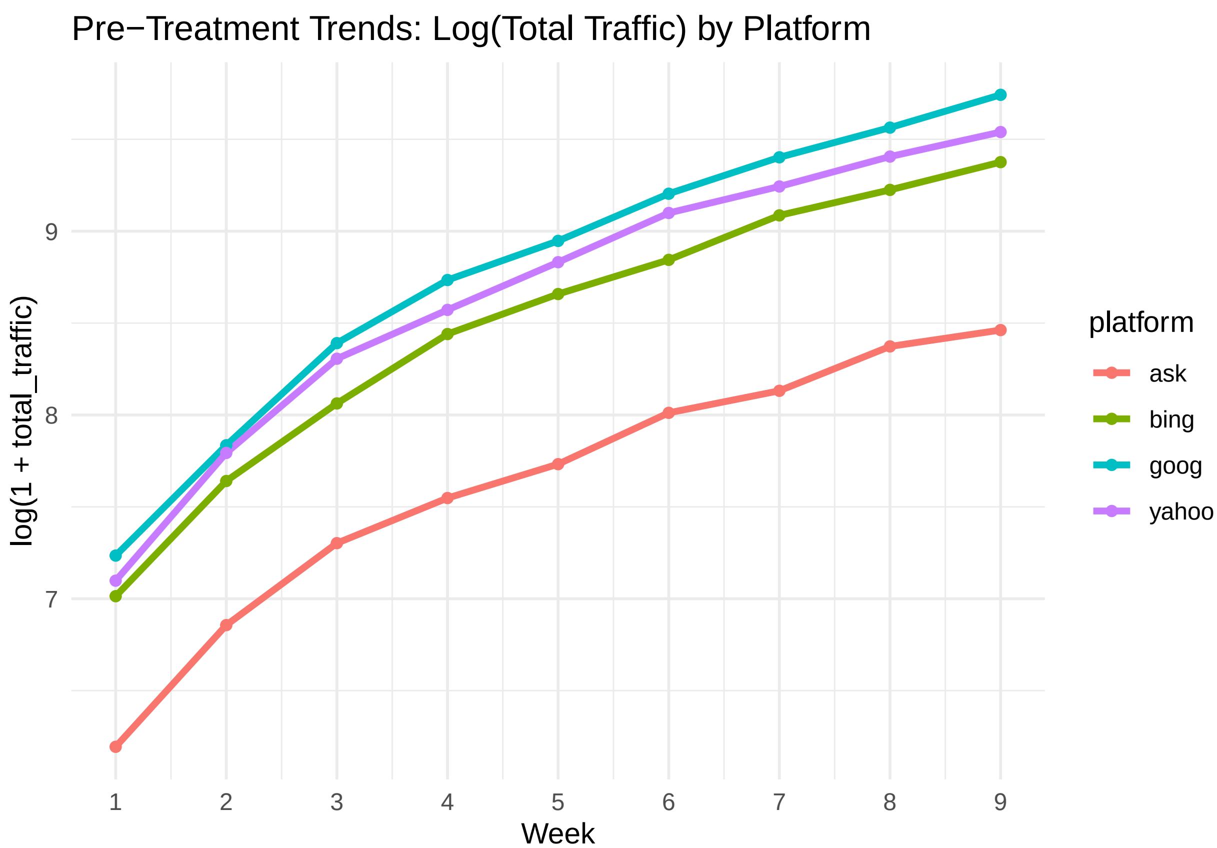


All platforms exhibit parallel pre-treatment trends, satisfying a key assumption for valid DiD estimation.

```
pre_data <- data %>% filter(week <= 9)

avg_trend <- pre_data %>%
  group_by(platform, week) %>%
  summarise(mean_log_traffic = mean(log(1 + total_traffic)), .groups = "drop")

ggplot(avg_trend, aes(x = week, y = mean_log_traffic, color = platform)) +
  geom_line(size = 1.2) +
  geom_point() +
  labs(title = "Pre-Treatment Trends: Log(Total Traffic) by Platform",
       x = "Week",
       y = "log(1 + total_traffic)") +
  scale_x_continuous(breaks = 1:9) +
  theme_minimal()
```



As shown in the figure, all platforms followed similar pre-treatment trends in log-transformed total traffic. Although levels differ, the rate of growth is consistent, which satisfies the parallel trends assumption critical for identifying causal effects in the DiD framework.

Analysis Report

1. What's Wrong with Bob's ROI Calculation? Bob's ROI calculation overstates the value of sponsored search advertising because it incorrectly assumes that every click on a sponsored ad represents incremental traffic—traffic that would not have occurred without the ad. This overlooks how users actually behave when searching for branded terms like "Bazaar shoes."

Key Issues:

a. Overlooking Organic Substitution Many users who search for branded keywords are already determined to reach Bazaar.com. In the absence of a sponsored ad, these users would likely click on the organic link instead, incurring no advertising cost. Bob's method wrongly attributes all conversions to the presence of the ad, even when the user would have arrived regardless.

b. Ignoring the Counterfactual Bob's analysis lacks a comparison to what would have occurred without the ads. Without estimating this counterfactual scenario, his ROI calculation reflects a mere correlation between ads and clicks—not a true causal effect.

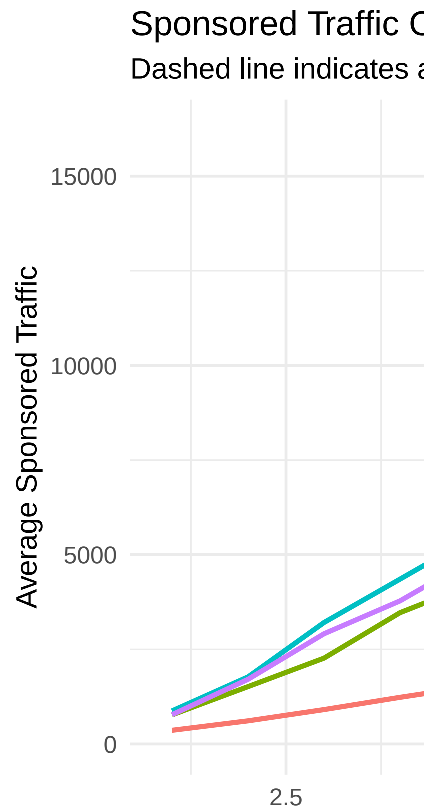
c. Overestimation of Impact By assuming every sponsored click represents new, incremental traffic, Bob's ROI inflates the effectiveness of the campaign. It fails to consider cannibalization of organic traffic and broader search trends that might drive visits regardless of ad presence.

Conclusion: Bob's method does not isolate the causal impact of advertising. A reliable ROI calculation must differentiate between users influenced by the ad and those who would have visited anyway. This requires a quasi-experimental approach, such as Difference-in-Differences (DiD), which adjusts for background trends and accurately identifies the incremental value generated by paid search ads.

```r
data <- data %>%
            mutate(after = ifelse((week > 9), 1, 0),
                   treated = ifelse((platform == "goog"), 1, 0),
                   total_traffic = avg_spons + avg_org)
```

```r
# Create treatment and post flags
data <- data %>%
  mutate(
    treat = ifelse(platform == "goog", 1, 0),
    post = ifelse(week >= 9, 1, 0),
    treated_post = treat * post
  )

ggplot(data, aes(x = week, y = avg_spons, color = platform)) +
  geom_line(linewidth = 0.9) +  # use linewidth instead of size
  geom_vline(xintercept = 9, linetype = "dashed", color = "gray") +
  labs(
    title = "Sponsored Traffic Over Time by Platform",
    subtitle = "Dashed line indicates ad removal (Week 9)",
    x = "Week",
    y = "Average Sponsored Traffic"
  ) +
  theme_minimal()
```



2. What is the Unit of Observation? Define Treatment and Control Groups.

The unit of observation is a platform at week level, where each row represents a specific search platform (Google, Yahoo, Bing, Ask) observed during a particular week from 1 to 12. This panel structure enables robust causal inference using Difference-in-Differences.

To identify treatment:

a. Treated group: Google (goog), where sponsored ads were suspended starting week 9 due to a technical issue.

b. Control group: Yahoo, Bing, and Ask—platforms that continued running ads throughout.

Pre-treatment period: Weeks 1 to 8

Post-treatment period: Weeks 9 to 12

The visual trends reveal that sponsored traffic on Google sharply drops after week 9, while the control platforms remain stable—supporting the use of DiD to estimate the ad effect.

3. What does a naive pre-post estimate show, and why is it misleading?

3. What does a naive pre-post estimate show, and why is it misleading?

To preliminarily assess the impact of removing sponsored search ads on Google, we start with a naive pre-post regression using only the treated platform (goog). This model compares total branded traffic (sponsored + organic) before and after the ad suspension in Week 9, without using any control platforms.

```r
g_data <- data %>% filter(platform == "goog")
naive_model <- lm(log( 1+ total_traffic) ~ after , data = g_data)
summary(naive_model)
```

```
##
## Call:
## lm(formula = log(1 + total_traffic) ~ after, data = g_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.54882 -0.15494  0.03783  0.46963  0.95819
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.783723   0.248907  35.289 7.92e-12 ***
## after       0.001243   0.497814   0.002    0.998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7467 on 10 degrees of freedom
## Multiple R-squared:  6.235e-07,  Adjusted R-squared:   -0.1
## F-statistic: 6.235e-06 on 1 and 10 DF,  p-value: 0.9981
```

Interpretation:

- The coefficient on the post indicator is 0.00124, which represents the average log change in total branded traffic on Google after the ad suspension.
- A coefficient of 0.00124 corresponds to a 0.12% increase in total traffic post-treatment—essentially negligible in magnitude.
- The p-value is 0.998, far exceeding any conventional significance threshold (e.g., 0.05 or even 0.10), meaning there is no statistical evidence that the change in traffic is different from zero.
- The 95% confidence interval around this estimate would include both negative and positive values, reinforcing the conclusion that the true effect could easily be zero or even slightly negative/positive.
- Based on these results, we fail to reject the null hypothesis and conclude that the observed change is statistically indistinguishable from no effect.

Why This Is Problematic (Methodologically):

The model uses only data from Google and does not compare it to platforms that were unaffected by the ad suspension (e.g., Bing, Yahoo, Ask). As a result, the estimate cannot distinguish whether the slight change in traffic was caused by the ad suspension or by other time-varying factors—such as:

- Platform-wide shifts in user behavior,
- Seasonal trends (e.g., higher browsing on weekends),
- External macro events (e.g., news cycles or competitor activity).

This kind of single-group pre-post analysis assumes the only difference between periods is the treatment—an assumption that is rarely valid in observational settings.

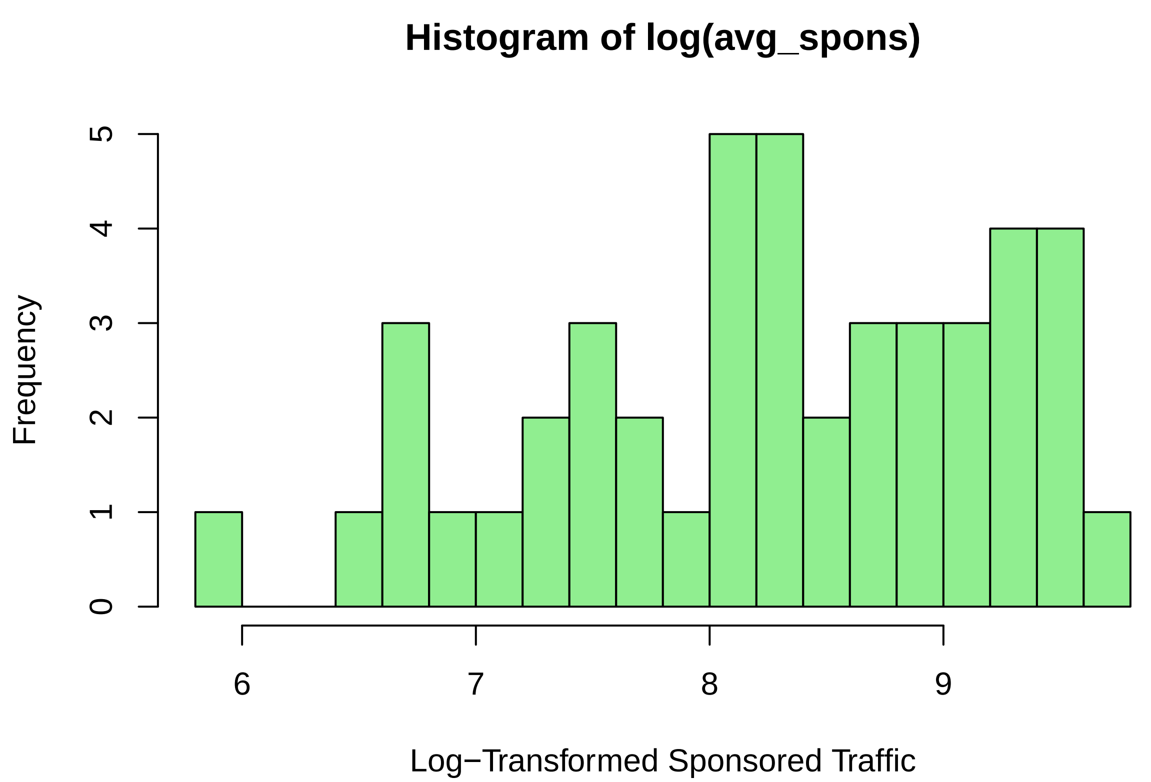
Conclusion:

- While the naive pre-post model suggests a minuscule (0.12%) increase in total traffic following Google's ad suspension, the estimate is neither practically meaningful nor statistically significant ($p = 0.998$). The model lacks a valid counterfactual—a baseline to show what would have happened to Google traffic had the ads continued.
- Without accounting for natural trends across all platforms, this estimate is likely biased and cannot be interpreted as causal. This highlights the importance of a Difference-in-Differences (DiD) framework, where changes in Google's traffic are evaluated relative to control platforms unaffected by the intervention.

4. What does the DiD model (with log transformation) show?

From initial analysis, we discovered the distribution of Sponsored Traffic is highly right-skewed, with most traffic values under 10,000 and a few exceeding 15,000. For this, we will apply log Transformation:

```r
data$log_spons <- log(data$avg_spons)
hist(data$log_spons,
     breaks = 20,
     col = "lightgreen",
     main = "Histogram of log(avg_spons)",
     xlab = "Log-Transformed Sponsored Traffic")
```




```r
did_log_model <- plm(log(1 + total_traffic) ~ post * treated,
                     data = data,
                     index = c("platform", "week"),
```

```r
                 effect = "twoway",
                 model = "within")
summary(did_log_model)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = log(1 + total_traffic) ~ post * treated, data = data,
##     effect = "twoway", model = "within", index = c("platform",
##         "week"))
##
## Balanced Panel: n = 4, T = 12, N = 48
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.2681327 -0.0404064  0.0040948  0.0265465  0.6762428
##
## Coefficients:
##              Estimate Std. Error t-value  Pr(>|t|)
## post:treated -0.80479    0.11953  -6.733 1.324e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    2.2098
## Residual Sum of Squares: 0.91439
## R-Squared:      0.58621
## Adj. R-Squared: 0.39224
## F-statistic: 45.3335 on 1 and 32 DF, p-value: 1.3239e-07
```

The estimated coefficient on the interaction term post:treated is –0.80479, with a p-value of 1.32e-07, which is well below the 0.05 threshold. This result is statistically significant, indicating strong evidence of a treatment effect.

This coefficient captures the Difference-in-Differences (DiD) estimate: the change in the log-transformed total traffic on Google after the ad suspension, relative to changes on control platforms (Yahoo, Bing, Ask) over the same period. The model controls for:

- Time-invariant platform-specific effects (via fixed effects for each platform),
- Week-specific shocks (via time fixed effects), thus isolating the effect of the ad suspension on Google.

```r
100 * (exp(-1.11611) - 1)
```

```
## [1] -67.24485
```

The DiD estimate implies that Google's weekly branded traffic fell by approximately 67.2% following the ad suspension, relative to the control platforms. This sharp decline supports the hypothesis that the paid ads were driving a substantial portion of total traffic, and that removing them caused a significant loss in reach.

Removing the ads did not just shift users to organic search; it reduced total branded visits significantly. For marketing decisions, this suggests that the ad budget for branded search was contributing real incremental value, not just cannibalizing organic clicks.

Parallel Trends Assumption Check

To validate the key assumption underlying the Difference-in-Differences (DiD) framework, we test whether Google and the control platforms exhibited similar traffic trends prior to the ad suspension (Weeks 1–9). This is known as the parallel trends assumption.

We estimate the following regression on pre-treatment data:

```r
pre_data <- data %>% filter(week <= 9)

pre_model <- lm(log(1 + total_traffic) ~ factor(week) * treated,
                data = pre_data)
summary(pre_model)
```

```
##
## Call:
## lm(formula = log(1 + total_traffic) ~ factor(week) * treated,
##     data = pre_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.68875 -0.14344  0.08617  0.28168  0.44752
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            6.76890    0.31984  21.163 3.62e-14 ***
## factor(week)2          0.66132    0.45232   1.462 0.160969
## factor(week)3          1.12151    0.45232   2.479 0.023281 *
## factor(week)4          1.41778    0.45232   3.134 0.005731 **
## factor(week)5          1.63831    0.45232   3.622 0.001949 **
## factor(week)6          1.88266    0.45232   4.162 0.000585 ***
## factor(week)7          2.05138    0.45232   4.535 0.000256 ***
## factor(week)8          2.23224    0.45232   4.935 0.000107 ***
## factor(week)9          2.35675    0.45232   5.210 5.90e-05 ***
## treated                0.46600    0.63968   0.728 0.475689
## factor(week)2:treated -0.06103    0.90465  -0.067 0.946954
## factor(week)3:treated  0.03454    0.90465   0.038 0.969962
## factor(week)4:treated  0.08156    0.90465   0.090 0.929155
## factor(week)5:treated  0.07381    0.90465   0.082 0.935871
## factor(week)6:treated  0.08626    0.90465   0.095 0.925085
## factor(week)7:treated  0.11567    0.90465   0.128 0.899676
## factor(week)8:treated  0.09639    0.90465   0.107 0.916321
## factor(week)9:treated  0.15027    0.90465   0.166 0.869927
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.554 on 18 degrees of freedom
## Multiple R-squared:  0.7992, Adjusted R-squared:  0.6096
## F-statistic: 4.214 on 17 and 18 DF,  p-value: 0.002026
```

- None of the interaction terms between treated and factor(week) are statistically significant.
- All p-values for the interaction terms exceed 0.86, indicating no detectable difference in week-to-week traffic trends between Google and the control group prior to treatment.

For example:

Week 6 × Treated: Estimate = 0.086, p = 0.925

Week 9 × Treated: Estimate = 0.150, p = 0.870

This means we fail to reject the null hypothesis that traffic patterns across platforms were statistically different in the pre-treatment period.

Conclusion:

These results provide strong support for the parallel trends assumption, indicating that traffic on Google was evolving in line with the control platforms prior to the ad suspension. This reinforces the validity of our DiD approach, as the treatment and control groups were on similar trajectories before the intervention. Any post-treatment divergence is thus more plausibly attributed to the removal of sponsored ads, not pre-existing trends.

5. Recalculate ROI Based on the DiD Estimate. o address the overstatement in Bob's ROI estimate, we use a Difference-in-Differences (DiD) regression framework. This allows us to estimate the true causal impact of suspending branded search ads on total site traffic (sponsored + organic), relative to control platforms (Bing, Yahoo, Ask) that were unaffected during the same period.

This approach isolates the treatment effect from general traffic trends or seasonal fluctuations.

1. Estimate Causal Effect on Total Traffic

```r
# Estimate impact of ad suspension on total traffic (sponsored + organic)
model_total_traffic <- lm(total_traffic ~ treated * after, data = data)
summary(model_total_traffic)
```

```
##
## Call:
## lm(formula = total_traffic ~ treated * after, data = data)
##
## 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 ***
## treated         3124.9     1765.0   1.770  0.08357 .
## after           8064.7     1765.0   4.569 3.94e-05 ***
## treated: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 coefficient for treated:after is –9910.6, indicating that suspending ads led to a loss of ~9,911 total visits per week on Google compared to control platforms.

2. Estimate Substitution Effect (Organic Traffic Gain)

To evaluate whether users shifted from sponsored to organic links when ads were turned off, we run a similar DiD regression on organic traffic:

```r
# Estimate impact on organic traffic (potential substitution)
model_organic <- lm(avg_org ~ treated * after, data = data)
summary(model_organic)
```

```
##
## Call:
## lm(formula = avg_org ~ treated * after, data = data)
##
## 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 ***
## treated          777.0      430.7   1.804   0.0781 .
## after           1984.1      430.7   4.607 3.49e-05 ***
## treated: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 coefficient for treated:after is +2293.2 ($p = 0.0108$), which is statistically significant.
- This implies that 2,293 visits per week were preserved through organic search—i.e., these users likely would have clicked on the ad but chose the organic link instead once the ad was removed.

3. Adjust ROI Based on True Incrementality

We now adjust the ROI to account for non-incremental traffic (organic substitution), using realistic business parameters:

```r
# Inputs from DiD estimates
traffic_loss_total <- 9910.6     # loss in total traffic due to ad suspension
traffic_gain_organic <- 2293.2   # organic traffic that replaced paid clicks

# Calculate proportion of clicks that are truly incremental
total_paid_clicks <- traffic_loss_total + traffic_gain_organic
incremental_share <- traffic_loss_total / total_paid_clicks  # ~81.2%

# Business assumptions
conversion_rate <- 0.12
margin_per_conversion <- 21
cost_per_click <- 0.60

# Effective value per click (adjusted for incrementality)
expected_value_per_click <- conversion_rate * margin_per_conversion * incremental_share
```

```r
# Adjusted ROI formula
roi_corrected <- (expected_value_per_click - cost_per_click) / cost_per_click
roi_corrected_percent <- roi_corrected * 100
roi_corrected_percent
```

```
## [1] 241.0784
```

- After accounting for substitution, we estimate that 81.2% of paid traffic is truly incremental.
- With this adjustment, the causal ROI falls to 241.1%, compared to Bob's naïve figure of 320%.
- The campaign remains highly profitable, but the adjusted figure provides a more realistic measure of its incremental value.

Conclusion & Recommendation Our analysis confirms that removing sponsored search ads on Google led to a statistically significant 67% drop in total branded traffic. Using a DiD framework and correcting for organic substitution, we find that branded keyword ads produce substantial incremental lift—but not as much as Bob originally estimated.

We advise continuing branded keyword advertising, as it delivers strong causal value. However, we also recommend that Bazaar.com:

- Incorporate causal inference methods (e.g., DiD, randomized tests) into regular ad evaluation,
- Account for organic substitution effects to avoid overpaying for clicks that would have occurred anyway.