

Case 7 Bazaar.com

Jonathan Ratschat / Franziska Bülck

22 12 2019

Libraries

```
#Libraries
library(readxl)
library(ggplot2)
library(reshape2)
library(dplyr)
```

Prepare data

```
#Import data
Data <- read_excel("Bazaar data.xlsx")
head(Data)

## # A tibble: 6 x 4
##   Week Sponsored Organic Origin
##   <dbl>      <dbl>   <dbl> <chr>
## 1     1      32269  127876 Google
## 2     2      31951  128169 Google
## 3     3      32143  125717 Google
## 4     4      31417  126264 Google
## 5     5      31194  123871 Google
## 6     6      31576  124053 Google

#Transform variables
str(Data)

## Classes 'tbl_df', 'tbl' and 'data.frame':   24 obs. of  4 variables:
##  $ Week      : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Sponsored: num  32269 31951 32143 31417 31194 ...
##  $ Organic   : num  127876 128169 125717 126264 123871 ...
##  $ Origin    : chr   "Google" "Google" "Google" "Google" ...

Data$Origin <- as.factor(Data$Origin)

#Create column with total clicks
Data$Total <- Data$Sponsored + Data$Organic

#Create time indicator (before treatment = 0, after treatment = 1)
Data$Time <- ifelse(Data$Week < 10, 0, 1)
Data$Time <- as.factor(Data$Time)

#Data set used for analysis
head(Data)

## # A tibble: 6 x 6
##   Week Sponsored Organic Origin Total Time
```

```
##      <dbl>      <dbl>      <dbl> <fct>      <dbl> <fct>
## 1         1        32269  127876 Google 160145 0
## 2         2        31951  128169 Google 160120 0
## 3         3        32143  125717 Google 157860 0
## 4         4        31417  126264 Google 157681 0
## 5         5        31194  123871 Google 155065 0
## 6         6        31576  124053 Google 155629 0
```

```
str(Data)
```

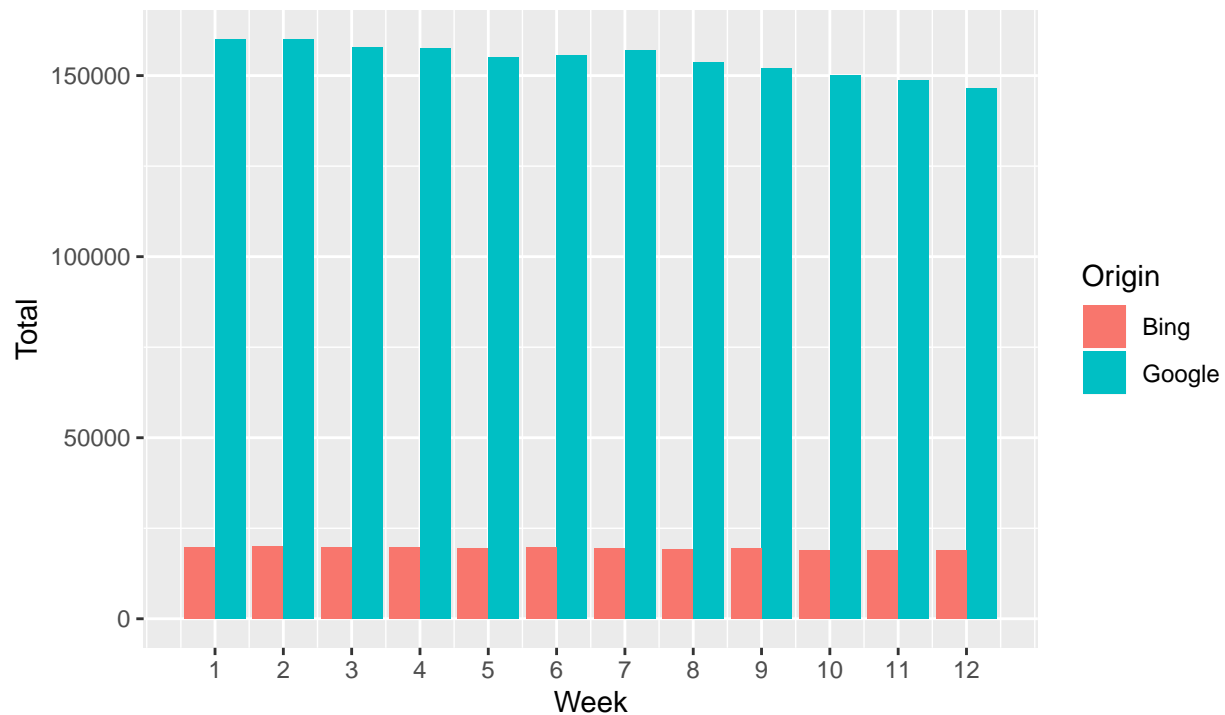
```
## Classes 'tbl_df', 'tbl' and 'data.frame':   24 obs. of  6 variables:
## $ Week      : num  1 2 3 4 5 6 7 8 9 10 ...
## $ Sponsored: num  32269 31951 32143 31417 31194 ...
## $ Organic   : num  127876 128169 125717 126264 123871 ...
## $ Origin    : Factor w/ 2 levels "Bing","Google": 2 2 2 2 2 2 2 2 2 2 ...
## $ Total     : num  160145 160120 157860 157681 155065 ...
## $ Time      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
```

Exploratory analysis

```
#Total clicks by Week and Origin
ggplot(Data, aes(x = Week, y = Total, fill = Origin)) +
  geom_bar(position = "dodge", stat = "identity") +
  ggtitle("Total clicks by Week and Origin", subtitle = "Number of total clicks shows a decreasing trend over time") +
  labs(caption = "Ads were only used in the first nine weeks on Google (whole timespan on Bing).") +
  xlab("Week") +
  ylab("Total") +
  theme(plot.title = element_text(color = "blue", face = "bold")) +
  scale_colour_manual(name="Origin") +
  scale_x_continuous(breaks = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), labels = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12"))
```

Total clicks by Week and Origin

Number of total clicks shows a decreasing trend



Ads were only used in the first nine weeks on Google (whole timespan on Bing).

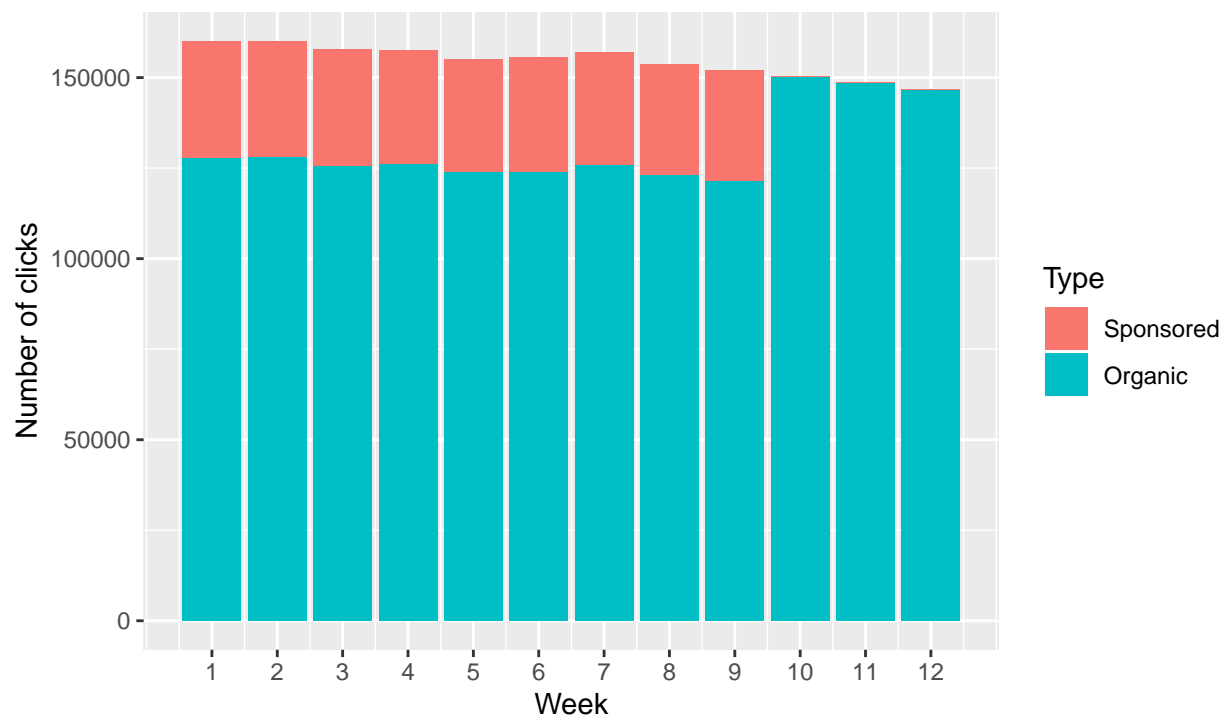
```
#Google Subset
dataGoogle <- subset(Data, Data$Origin == "Google")

#Google clicks by week
dataGoogleLong <- melt(dataGoogle %>%
  select(-Total, -Time),
  id = c("Week", "Origin"))

ggplot(dataGoogleLong, aes(x = Week, y = value, fill = variable)) +
  geom_bar(position = "stack", stat = "identity") +
  ggtitle("Google clicks by Week", subtitle = "When sponsored ads are turned off, many clicks are")
xlab("Week") +
  ylab("Number of clicks") +
  labs(caption = "Ads were only used in the first nine weeks on Google.") +
  theme(plot.title = element_text(color = "blue", face = "bold")) +
  guides(fill=guide_legend(title="Type")) +
  scale_x_continuous(breaks = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), labels = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12"))
```

Google clicks by Week

When sponsored ads are turned off, many clicks are substituted by organic clicks



Ads were only used in the first nine weeks on Google.

Difference-in-difference analysis

We perform difference-in-difference analysis to find out if using paid search ads (brandkeywords) has a significant effect on $\log(\text{Total})$. The log is chosen to make the results of Bing and Google more comparable.

Two independent variables: Time: before and after treatment (treatment: turn brandkeywords off) Origin: treatment group and control group (treatment = Google, control = Bing)

```
mod1 <- lm(log(Total) ~ Time*Origin, data = Data)
summary(mod1)
```

```
##
## Call:
## lm(formula = log(Total) ~ Time * Origin, data = Data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.029334 -0.010347  0.002204  0.008539  0.022615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.883078   0.004840 2042.014 < 2e-16 ***
## Time1         -0.036118   0.009680  -3.731  0.00132 **
## OriginGoogle    2.078141   0.006845 303.617 < 2e-16 ***
## Time1:OriginGoogle -0.016968  0.013689  -1.239  0.22951
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01452 on 20 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 4.082e+04 on 3 and 20 DF,  p-value: < 2.2e-16
```

Time1:OriginGoogle coefficient is not significant. Zero hypothesis that the usage of brandkeywords has no effect on dependent variable cannot be rejected.

Conclusion

The conducted difference-in-difference analysis confirms the statistically insignificant effect of paid search ads. Therefore, we recommend to stop spending money for brand keyword ads.