RocketFuel

Manish Grewal

17/07/2020

## Load libraries

library(dplyr)  
library(ggplot2)

## Reading the dataset

setwd("C:/Users/manish.grewal/emdp/R/RocketFuel")  
dat <- read.csv("RocketFuel.csv")

# 1. Was the advertising campaign effective? Did additional consumers convert as a result of the ad campaign?

#### We can compare the conversion rate of control group vs. the conversion rate of exposed group to see effectiveness

control <- filter(dat, test == 0)  
num\_control <- nrow(control)  
num\_control

[1] 23524

control\_conv <- filter(dat, test == 0, converted == 1)  
num\_control\_conv <- nrow(control\_conv)  
num\_control\_conv

[1] 420

exposed <- filter(dat, test == 1)  
num\_exposed <- nrow(exposed)  
num\_exposed

[1] 564577

exposed\_conv <- filter(dat, test == 1, converted == 1)  
num\_exposed\_conv <- nrow(exposed\_conv)  
num\_exposed\_conv

[1] 14423

#### Conversion rate for control group

rate\_control\_conv = num\_control\_conv \* 100 / num\_control  
rate\_control\_conv

[1] 1.785411

#### Conversion rate for exposed group

rate\_exposed\_conv = num\_exposed\_conv \* 100 / num\_exposed  
rate\_exposed\_conv

[1] 2.554656

#### **Answer 1**

#### As the conversion rate of exposed group 2.554656% is greater than the conversion rate for control group 1.7854106%, we can conclude that the campaign is effective.

#### Yes, rate\_exposed\_conv - rate\_control\_conv = 0.7692453% additional consumers converted as a result of the ad campaign.

# 2. Was the campaign profitable?

## a. How much more money did TaskaBella make by running the campaign (excluding advertising costs)?

#### Number of users in exposed group = 564577

#### Profit per conversion = $40

#### Additional profit = ((Conversion rate of exposed group) - (Conversion rate of control group)) \* (Number of users in exposed group) \* (Profit per conversion) / 100

addl\_profit <- (rate\_exposed\_conv - rate\_control\_conv) \* num\_exposed \* 40 / 100  
addl\_profit

[1] 173719.3

## b. What was the cost of the campaign?

#### Cost per 1000 (CPM) = $9

#### Cost of campaign

cpm <- 9  
tot\_impr <- sum(dat$tot\_impr)  
cost\_campaign <- tot\_impr \* cpm / 1000  
cost\_campaign

[1] 131374.6

## c. Calculate the ROI of the campaign. Was the campaign profitable?

roi <- addl\_profit / cost\_campaign \* 100   
roi

[1] 132.232

paste("ROI: ", roi, "%")

[1] "ROI: 132.231980601722 %"

### Yes, Campaign was profitable as ROI was good

## d. What was the opportunity cost of including a control group; how much more could have TaskaBella made with a smaller control group or not having a control group at all?

opp\_cost <- num\_control \* (rate\_exposed\_conv - rate\_control\_conv) \* 40 / 100  
opp\_cost

[1] 7238.291

paste("Opportunity cost in $:", opp\_cost)

[1] "Opportunity cost in $: 7238.29075573394"

### If the control group had been shown ads for the product instead of PSA, an additional profit of $7238.2907557 could have been obtained.

# 3. How did the number of impressions seen by each user influence the effectiveness of advertising?

## a. Create a chart of conversion rates as a function of the number of ads displayed to users. Plot conversion rates for those who were in the control group and for those who were exposed to the ad. Group together number of impressions as necessary to obtain a meaningful plot. (Conversion rate means the percentage of unique users who made a purchase.)

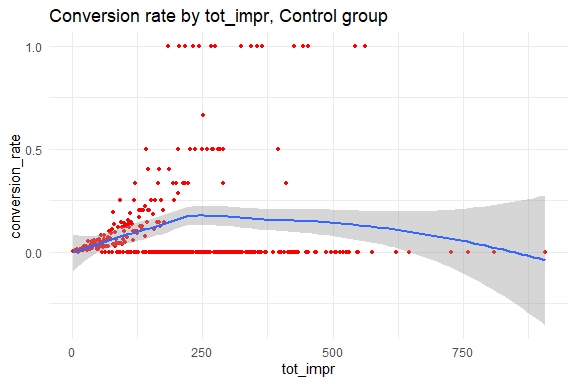
#### control group

ctrl <- filter(dat, test == 0) # control group  
expd <- filter(dat, test == 1) # exposed group  
  
by\_tot\_impr\_ctrl <- group\_by(ctrl, tot\_impr)  
summ <- summarise(by\_tot\_impr\_ctrl, conversion\_rate = mean(converted))

`summarise()` ungrouping output (override with `.groups` argument)

ggplot(summ) +  
 aes(x = tot\_impr, y = conversion\_rate) +  
 geom\_point(size = 1L, colour = "red") +  
 geom\_smooth(method = 'loess') +  
 theme\_minimal() +  
 ggtitle("Conversion rate by tot\_impr, Control group")

`geom\_smooth()` using formula 'y ~ x'



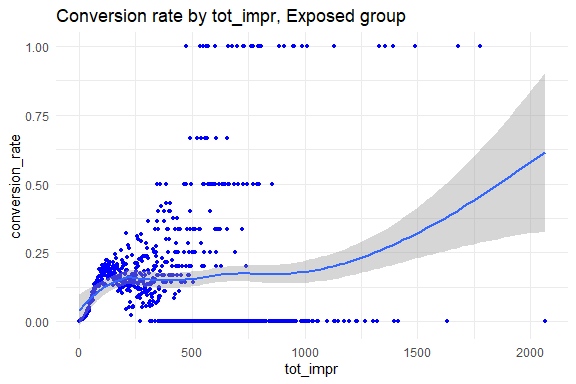
#### exposed group

by\_tot\_impr\_expd <- group\_by(expd, tot\_impr)  
summ <- summarise(by\_tot\_impr\_expd, conversion\_rate = mean(converted))

`summarise()` ungrouping output (override with `.groups` argument)

ggplot(summ) +  
 aes(x = tot\_impr, y = conversion\_rate) +  
 geom\_point(size = 1L, colour = "blue") +  
 geom\_smooth(method = 'loess') +  
 theme\_minimal() +  
 ggtitle("Conversion rate by tot\_impr, Exposed group")

`geom\_smooth()` using formula 'y ~ x'



## b. What can you infer from the charts? In what region is advertising most effective?

#### The conversion rate increases as the number of impressions is increased. However, as the number of impressions is increased beyond a limit (appears to be 250), we see a flattening of the smoothing curve and also a concentration of non conversions. Need to zoom in to the chart to get a better picture.

## c. What do the above figures imply for the design of the next campaign assuming that consumer response would be similar?

#### total impressions can be limited to up to 500 to ensure cost savings. At higher number of impression, conversion rate can dip unexpectedly.

# 4. How does consumer response to advertising vary on different days of the week and at different times of the day?

## a. Create a chart with the conversion rates for the control group and the exposed group as a function of the day of week when they were shown the most impressions.

by\_mode\_day\_ctrl <- group\_by(ctrl, mode\_impr\_day)  
summ\_ctrl <- summarise(by\_mode\_day\_ctrl, conversion\_rate = mean(converted))

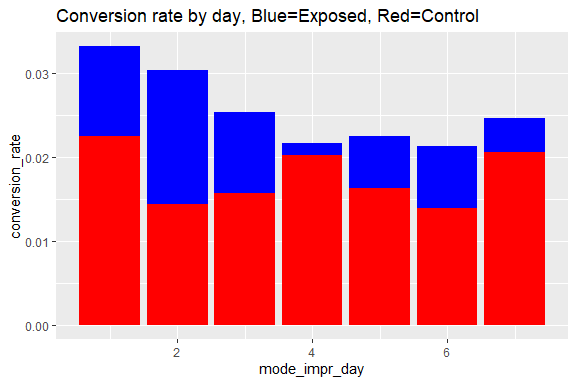
`summarise()` ungrouping output (override with `.groups` argument)

by\_mode\_day\_expd <- group\_by(expd, mode\_impr\_day)  
summ\_expd <- summarise(by\_mode\_day\_expd, conversion\_rate = mean(converted))

`summarise()` ungrouping output (override with `.groups` argument)

summ\_day <- summ\_expd

ggplot() +  
 geom\_bar(data = summ\_expd, aes(x = mode\_impr\_day, y = conversion\_rate), stat = "identity", fill = "blue") +   
 geom\_bar(data = summ\_ctrl, aes(x = mode\_impr\_day, y = conversion\_rate), stat = "identity", fill = "red") +  
 ggtitle("Conversion rate by day, Blue=Exposed, Red=Control")



## b. Create the same chart for hours within a day (excluding the period between midnight and 8 a.m.).

dat1 <- filter(dat, mode\_impr\_hour > 7)  
ctrl <- filter(dat1, test == 0)  
expd <- filter(dat1, test == 1)  
  
by\_mode\_hour\_ctrl <- group\_by(ctrl, mode\_impr\_hour)  
summ\_ctrl <- summarise(by\_mode\_hour\_ctrl, conversion\_rate = mean(converted))

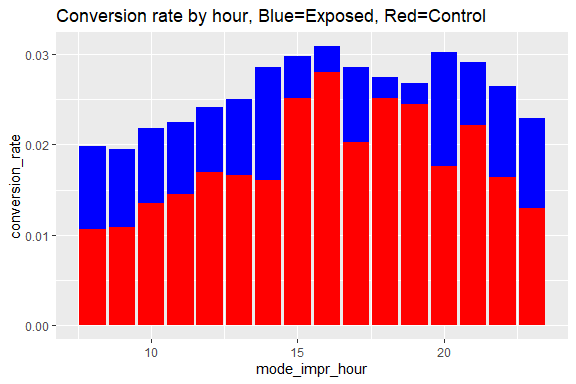
`summarise()` ungrouping output (override with `.groups` argument)

by\_mode\_hour\_expd <- group\_by(expd, mode\_impr\_hour)  
summ\_expd <- summarise(by\_mode\_hour\_expd, conversion\_rate = mean(converted))

`summarise()` ungrouping output (override with `.groups` argument)

summ\_hour <- summ\_expd

ggplot() +   
 geom\_bar(data = summ\_expd, aes(x = mode\_impr\_hour, y = conversion\_rate), stat = "identity", fill = "blue") +  
 geom\_bar(data = summ\_ctrl, aes(x = mode\_impr\_hour, y = conversion\_rate), stat = "identity", fill = "red") +  
 ggtitle("Conversion rate by hour, Blue=Exposed, Red=Control")



## c. What days/hours is advertising most/least effective?

#### From the graph, we can read the highest and lowest values for the exposed group. We can also cross check by arranging summ\_day and summ\_hour variables by conversion rate

### Most effective day

tail(arrange(summ\_day, conversion\_rate), 1)

# A tibble: 1 x 2  
 mode\_impr\_day conversion\_rate  
 <int> <dbl>  
1 1 0.0332

#### => Most effective on Mondays.

### Least effective day

head(arrange(summ\_day, conversion\_rate), 1)

# A tibble: 1 x 2  
 mode\_impr\_day conversion\_rate  
 <int> <dbl>  
1 6 0.0213

#### => Least effective on Saturdays

### Most effective hour

tail(arrange(summ\_hour, conversion\_rate), 1)

# A tibble: 1 x 2  
 mode\_impr\_hour conversion\_rate  
 <int> <dbl>  
1 16 0.0309

#### => Most effective between 4 to 5 pm

### Least effective hour

head(arrange(summ\_hour, conversion\_rate), 1)

# A tibble: 1 x 2  
 mode\_impr\_hour conversion\_rate  
 <int> <dbl>  
1 9 0.0195

#### => Least effective between 9 to 10 am