Assessing The Effectiveness of Display Advertising

1. Introduction

Star Digital designed an online display advertising campaign run on six websites with the primary objective of increasing subscription package sales. It also was interested in website visits as a proxy for measuring brand impact. Having been aware of the issues with click-through and view-through-based ad effectiveness metrics, Star Digital designed an experiment that measured the incremental impact of advertising on the outcome of interest, i.e. sales.

2. Experimental Design Overview

2. A. Experiment Setup

1. Before the first campaign ad was to be served to a user, s/he was assigned randomly to either the test group or the control group.

    a. The consumers in the control group were shown an advertisement for a charity organization in place of the advertisement for Star Digital, whenever their online actions prompted the serving of an ad.
    b. Banner advertising is typically served through ad-serving software,which determines whether a consumer browsing on a particular web page is eligible to be shown that ad or on.

2. The ad-serving software was programmed to then check if the consumer was in the predetermined control group or the test group.

3. If the consumer was in the test group, the advertisement for Star Digital would be shown.

4. If the consumer was in the control group, the software would intervene and replace the Star Digital advertisement with an advertisement for a charity organization.

2. B. Group Size Considerations

Random assignment of consumers - expensive option

1. since charity ads cost the advertiser as much as the campaign ads, since the advertiser paid for advertisements served to the control group as well.

2. There was also an opportunity cost incurred when a consumer who could have been served an advertisement for Star Digital's campaign, and who could have been influenced by these advertisements to sign up for its subscription service, was shown charity advertisements instead.

Factors based on which fraction of users assigned to the control group

1. Baseline Conversion rate

If users naturally convert at high rates, then the fraction can be small because the control users will generate enough conversion events to detect statistically significant campaign impact.

2. campaign Reach

Showing advertisements to more users increases the number of control group conversions, assuming a fixed baseline conversion rate, and thus reduces the control group's fraction.

3. The minimum lift that the advertiser cares to detect

Choosing the smallest lift yielding positive ROI minimizes the cost of the charity ads.

4. Power of the experiment

An experiment achieves greater power if a larger number of consumers are placed in the control group.

Star Digital considered these factors and determined that in its experiment, 10 percent of the consumers would be placed in the control group, and the remaining 90 percent would be placed in the test group.

2. C. Sampling

For the purpose of analysis, a sample was drawn since the original database was very large. Of the analysis sample, 50 percent consisted of people who had chosen to purchase the subscription package of Star Digital, while the remaining 50 percent consisted of those who had not purchased the package.

However, whether the person belonged to the control group or test group was random in this sample.

3. Threats to Causal Inference

3. A. Selection Bias

we negate the effects of selection bias, as the selection process is random in both treatment-control assignments and sample data selection.

As a part of our analysis, we will further validate this by randomized testing.

3. B. Omitted Variable Bias

Omitted variables bias may occur in the experiment, when an external variable is correlated with the impressions. For example: Internet speed connectivity can be correlated with the number of impressions a particular user encounters and their purchase activity.

3. C. Simultaneity Bias

Since there is no reason to believe that an user that purchases a subscription would also cause an increase in views of the ad. It is safe to say there is no simultaneity bias in the experiment.

3. D. Measurement Error

Impressions are the only variable that is captured at an user level and which is not that difficult to track. It is safe to say there is no measurement error.

3. E. SUTVA Assumption Violation

We assume that consumers are unaware that they are participating in the experiment because showing them ads that grab their attention does not provide value to them.

4. Data Exploration and Overview

4. A. Importing relevant packages

```
library(pwr)
library(Hmisc)
library(ggplot2)
library(gmodels)
library(reshape)
library(dplyr)
```

4. B. Importing the data set

```
star_digital = read.csv("star_digital.csv")
```

4. C. Data description

```
head(star_digital)

##   X      id purchase test imp_1 imp_2 imp_3 imp_4 imp_5 imp_6
## 1 0  545716        1    1     0     1     0     0     0     0
## 2 1  893524        1    1     1     0     0    17     0     1
## 3 2 1372718        1    1     0     0     0    10     0     0
## 4 3  971359        1    1    14    37     1     7     0     7
## 5 4   59999        1    1     0     0     0    13     0     0
## 6 5  842034        1    0     0     1     0     0     0     0

str(star_digital)

## 'data.frame':    25303 obs. of  10 variables:
##  $ X       : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ id      : int  545716 893524 1372718 971359 59999 842034 731724 49425
1351203 357681 ...
```

```
##  $ purchase: int  1 1 1 1 1 1 0 0 1 0 ...
##  $ test    : int  1 1 1 1 1 0 1 0 1 1 ...
##  $ imp_1   : int  0 1 0 14 0 0 2 0 97 0 ...
##  $ imp_2   : int  1 0 0 37 0 1 272 0 214 0 ...
##  $ imp_3   : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ imp_4   : int  0 17 10 7 13 0 18 0 11 0 ...
##  $ imp_5   : int  0 0 0 0 0 0 0 0 3 0 ...
##  $ imp_6   : int  0 1 0 7 0 0 2 1 13 2 ...
```

4. D. Missing Values Inspection

```
colSums(is.na(star_digital))

##        X       id purchase     test    imp_1    imp_2    imp_3    imp_4 
##        0        0        0        0        0        0        0        0 
##    imp_5    imp_6 
##        0        0
```

As observed from the above, none of the columns in our data frame contains missing values.

4. E. Data Transformations

After loading the data set, we observed that we have not calculated overall total impressions that a treatment or control group user have seen across all the six websites.

Hence, we created a new variable "total_imp" which captures the overall impression for the user.

```
star_digital$total_imp <- star_digital$imp_1 + star_digital$imp_2 +
  star_digital$imp_3 + star_digital$imp_4 + star_digital$imp_5 +
star_digital$imp_6
```

We will also create another variable "imp_1to5" which captures the overall impression for an user for sites 1-5.

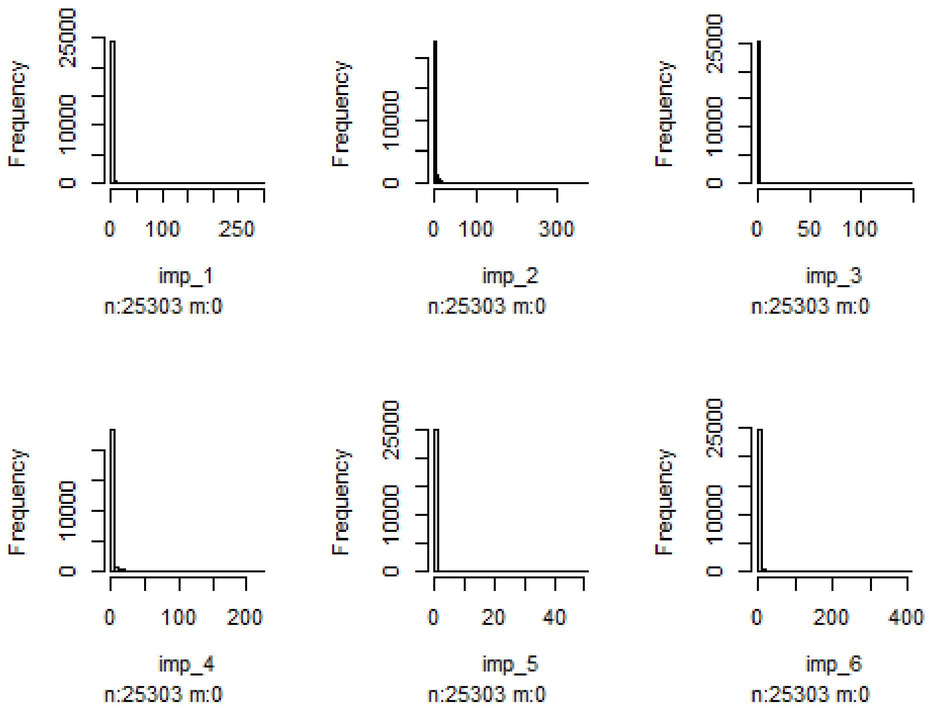
```
star_digital$imp_1to5 <- star_digital$imp_1 + star_digital$imp_2 +
  star_digital$imp_3 + star_digital$imp_4 + star_digital$imp_5
```

Hence, we have created two new variables "total_imp" which catches overall impression for all the sites and "imp_1to5" which catches overall impression of all the sites except site 6.

4. F. Distribution detection:

Checking the distributions of variables across the data to look for any anomaly

```
star_digital_hist <- star_digital[c("imp_1","imp_2","imp_3","imp_4"
,"imp_5","imp_6")]
hist.data.frame(star_digital_hist)
```

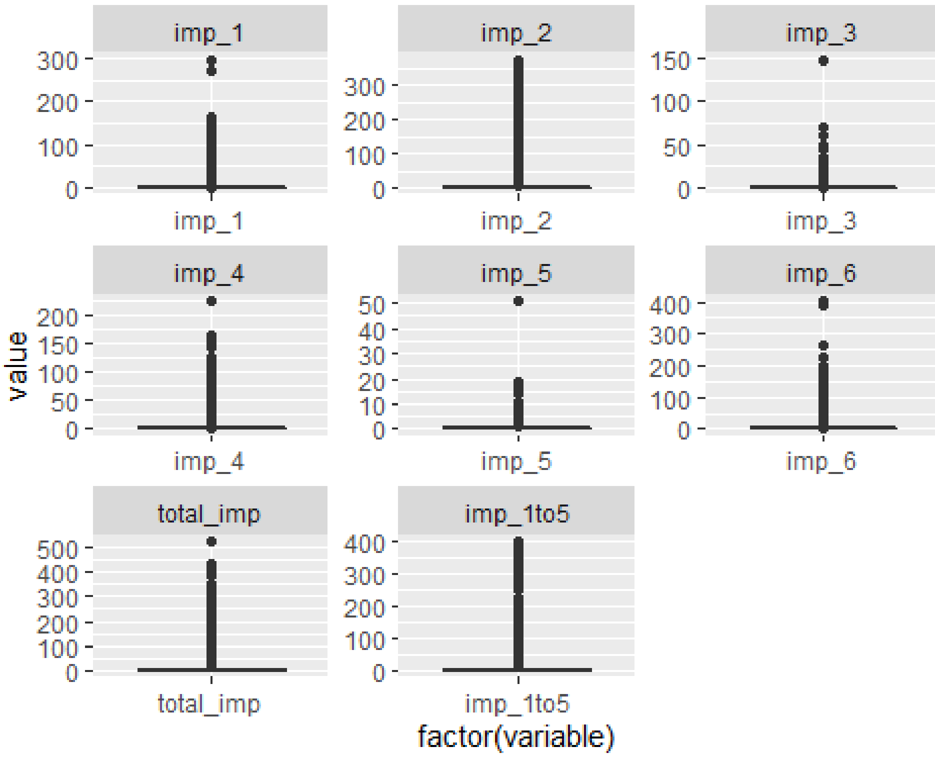


We can see that all the variables are right skewed with no abnormal distribution.

4. G. Outlier Treatment:

Let's look for outliers in out data.

```
meltData <- melt(select(star_digital, -c(1:4)))
p <- ggplot(meltData, aes(factor(variable), value))
p + geom_boxplot() + facet_wrap(~variable, scale="free")
```



We did observe some outliers in the impression data, but we are not applying any treatment for the same as the treatment might distort the proportion between test and control supports and may lead to non-interpretative results.

5. Experiment Quality Check

5. A. Checking for validity of Randomization

The first thing that needs to be done, when we receive the results of a particular experiment is to check whether the randomization has been done properly or not. Specifically, we have to check if the control group and the treatment group are identical, except for one being treated.

Here, we check whether an average person in each group saw the same number of impressions(those in control group saw the "charity" ad whereas those in treatment saw an advertisement for Star Digital's campaign.)

```
control = star_digital[star_digital$test == 0, ]
treatment = star_digital[star_digital$test == 1, ]

t.test(control$total_imp, treatment$total_imp)

##
##  Welch Two Sample t-test
##
## data:  control$total_imp and treatment$total_imp
```

```
## t = 0.12734, df = 3204.4, p-value = 0.8987
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8658621  0.9861407
## sample estimates:
## mean of x mean of y
##  7.929217  7.869078
```

As observed from the above, an customer in control group has on an average 7.929 impressions, whereas an customer in treatment group has on an average 7.869 impressions.

These numbers are different, but since p-value is large, we don't have to worry about the difference ; in other words, the averages are not statistically different which implies that our randomization is probably fine.

5. B. Power of the test

Size of the control and treatment group:

```
control_count = nrow(control)
control_count

## [1] 2656

treatment_count = nrow(treatment)
treatment_count

## [1] 22647
```

For the current experiment, the treatment and control group sizes are around 23K and 2.6K respectively.

Given the current sample size, we can identify the minimum effect size that can be observed.

Assumptions:

1. alpha(i.e.incorrectly rejecting the assumption that the two groups are indeed different) : 0.05
2. power(probability of correctly rejecting the null hypothesis) : 0.8

Here, we are using a two sample two proportions test

```
pwr.2p2n.test(n1 = control_count , n2 = treatment_count , sig.level = .05 ,
              power = .8)

##
##      difference of proportion power calculation for binomial distribution
(arcsine transformation)
##
```

```
##               h = 0.05748084
##              n1 = 2656
##              n2 = 22647
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: different sample sizes
```

Based on the above conditions, the minimum lift that can be detected is at 5.7%.

6. Data Analysis

PART 1: Measuring effectiveness of ads

Next, we want to see if our ads are effective. To do so, we compare the purchase rate of an average customer in control to an average customer in treatment.

If our ads are effective, the average purchase rate in treatment should be higher than that of control, and the difference should be significant:

Model : Purchase = a + b(test)

Variable a is the intercept and corresponds to purchase rate of those in control. Coefficient b is the coefficient of treatment and shows us how much treatment increases the purchase rate of the customers. If b is positive and significant, we can conclude that our ads are working.

Now, lets run this model.

```
summary(glm(purchase ~ test, data = star_digital ))

##
## Call:
## glm(formula = purchase ~ test, data = star_digital)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5049  -0.5049   0.4951   0.4951   0.5143
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.485693   0.009701  50.064   <2e-16 ***
## test        0.019186   0.010255   1.871   0.0614 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.249977)
##
##     Null deviance: 6325.5  on 25302  degrees of freedom
```

```
## Residual deviance: 6324.7  on 25301  degrees of freedom
## AIC: 36731
##
## Number of Fisher Scoring iterations: 2
```

As we can see, the average purchase rate of customers in control is 0.486 whereas as the average purchase rate of customers is treatment is 0.505.

Since the p-value is small (it is smaller than 0.1, but it's not as small as we ideally like it to be i.e. 0.05), we conclude that the difference is somewhat significant.

Therefore, online advertising seem to be effective.

PART 2: Measuring effect of ad frequency

We should "control" for the number of impressions, and then see how much being exposed to real ads increases the purchase probability on top of (in addition to) just receiving more impressions. Therefore, our model becomes

Model : purchase = a + b1(test) + b2(total_imp) + b3(treatment * total_imp)

In this formulation, b2 shows how much "active customers" customers that receive more impressions (charity ads or real ads) are more likely to purchase the product. Coefficient b3 shows how much those who saw relevant real ads become more likely to purchase the product as the number of ads increases.

Now, lets run this model.

```
summary(glm(purchase ~ test * total_imp, data = star_digital))

##
## Call:
## glm(formula = purchase ~ test * total_imp, data = star_digital)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.89562  -0.47994  -0.05711   0.51280   0.53228
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4651265  0.0101335  45.900  < 2e-16 ***
## test           0.0111885  0.0107209   1.044   0.2967
## total_imp      0.0025937  0.0004131   6.278 3.49e-10 ***
## test:total_imp 0.0010362  0.0004408   2.351   0.0188 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2442377)
##
##     Null deviance: 6325.5  on 25302  degrees of freedom
## Residual deviance: 6179.0  on 25299  degrees of freedom
```

```
## AIC: 36145
##
## Number of Fisher Scoring iterations: 2
```

As we can see, both coefficients are positive and significant. The fact that the coefficient of total_imp is positive and significant implies that even customers who saw charity ad are more likely to purchase.

But, we also get a positive(i.e. 0.001) and significant(p-value < 0.05) coefficient for test:total_imp (i.e., treatment * tot_imp) which implies that when the ads are real, showing more ads increases the probability of purchase even more (on top of the "active user" effect that is already captured in the model).

So, yes, showing more ads is effective.

PART 3: Comparing Sites 1-5 to Site 6

Finally, we compare the effect of ads on site 6 to ads on sites 1-5.

To do this, we use a linear model and look at the coefficient of an impression on sites 1-5, and see if the coefficient is larger than the coefficient of site 6.

purchase = a + b1(#impression for site 1-5) + b2(#impression for site 6)

Note: The model below is regressed on filtered data i.e. test = 1, i.e treated group.

```r
summary(glm(purchase ~ imp_1to5 + imp_6, data = treatment))
```

```
##
## Call:
## glm(formula = purchase ~ imp_1to5 + imp_6, data = treatment)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9664  -0.4822   0.0494   0.5131   0.5202
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4774579  0.0035254 135.432  < 2e-16 ***
## imp_1to5    0.0038081  0.0001684  22.610  < 2e-16 ***
## imp_6       0.0023735  0.0005095   4.659  3.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2439281)
##
##     Null deviance: 5661.2  on 22646  degrees of freedom
## Residual deviance: 5523.5  on 22644  degrees of freedom
## AIC: 32322
##
## Number of Fisher Scoring iterations: 2
```

Among the users, that were subjected to Star Digital Ad, keeping site 6 impressions constant, each additional ad impression in the site 1-5 increases the odds of purchase by 0.0038.

Among the users, that were subjected to Star Digital Ad, keeping site 1-5 impressions constant, each additional ad impression in the site 6 increases the odds of purchase by 0.0023.

Calculating ROI

a. cost for ads in site1-5 : $25 / 1000 impressions
b. cost for ads in site 6 : $20 / 1000 impressions

And, each purchase results in a lifetime contribution of $1200 for Star Digital.

For Sites 1-5:

Revenue for 1000 impressions = 1000 * 0.0038 * 1200 = $4560 Cost = $25 ROI = $182

For Site 6:

Revenue for 1000 impressions = 1000 * 0.00237 * 1200 = $2844 Cost = $20 ROI = $143

Ads on sites 1-5 gives a better return on investment as compared to on site 6. So, it would be a wise decision for Star Digital to put its advertising dollars in site 6.

7. Conclusion

1. Online advertising is effective.

   Hence, showing relevant ads (ads related to Star Digital) is a wise decision to boost sales.

2. Showing more ads is effective.

   Hence, the same advertisement could be shown multiple times to enhance the odds of purchase and in turn sales.

3. Showing ads on site 1-5 is more effective than ads on site 6.