Case Study Report

Author: Feiyi Su (feiyi.su@aalto.fi)

Background

On the home page, there are many campaigns running in the showstopper areas on a daily basis. Most of the showstoppers are basically image with a white box that includes text that directs the user to explore the catalog. This report is to measure the "success" of the new type of showstopper - an animated gif (or a video) instead of an image.
I will make analysis firstly by modelling and **then from the statistics perspective analysing three top KIPs (please focus on this part)**. Finally, I also made a supplementary analysis. I analyzed the click rate according to the time dimension.

Programming language: R language

Data description

A/B test result: Recorded customer behaviors (view and click) form Sept.2016 to Jan. 2017.

Control Group: No specific experiment running for this group; thus, they see the original image campaign showstopper.

Experiment Group: A showstopper with a video is shown to this group.

Define Goals

The Top 3 important KPIs to measure is: firstly, which I think is the most important one, **view-to-click conversion rate**. **Secondly**, it's **average time on page**, i.e. how much time on average spent on the page for those who have click action. **Thirdly**, it is the **counts of homepage visitors**.

Data analysis

Data cleaning

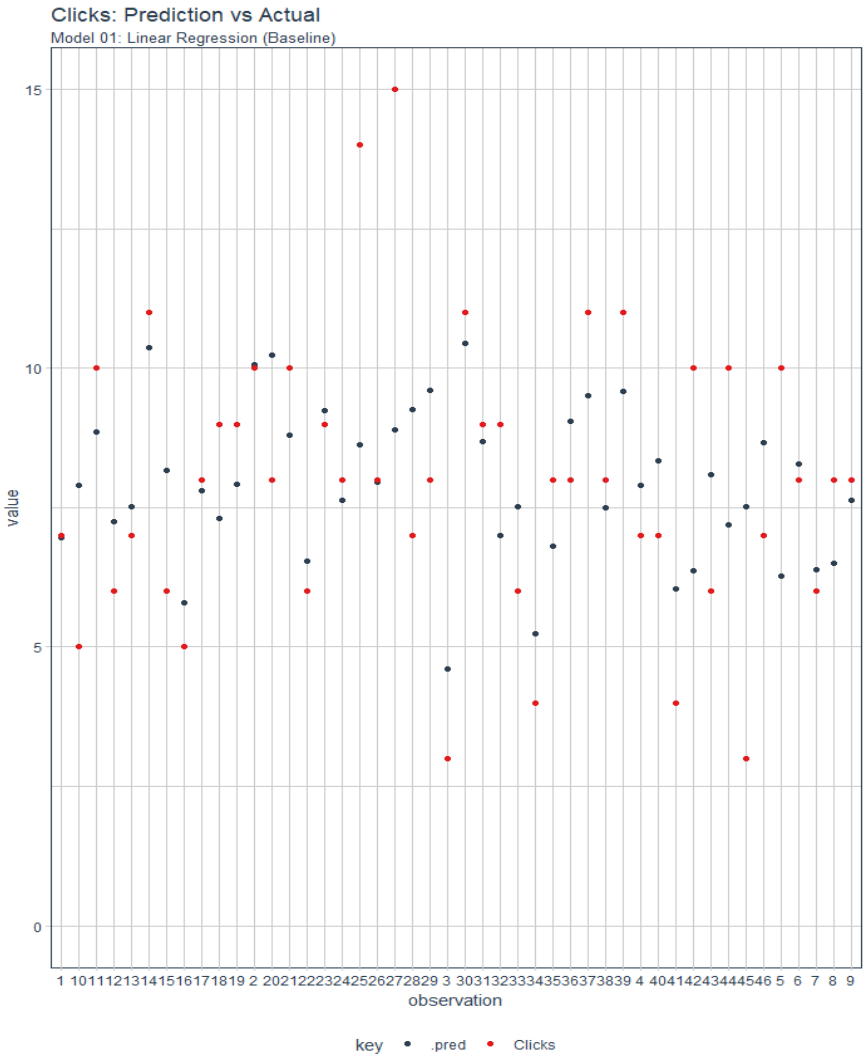
Missing value rows deletion
Date conversion and other data format conversion

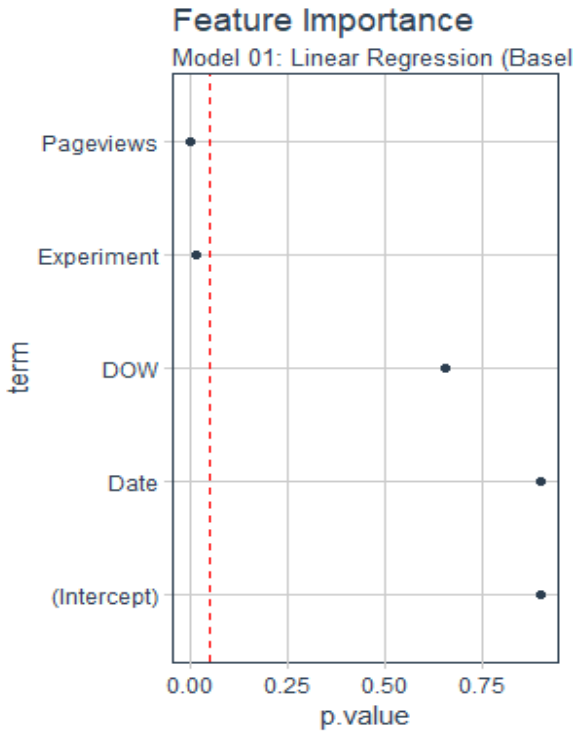
Modeling

(With the data formatted properly for analysis, I now separate into training and testing sets using an 80% / 20% ratio)

Linear Regression

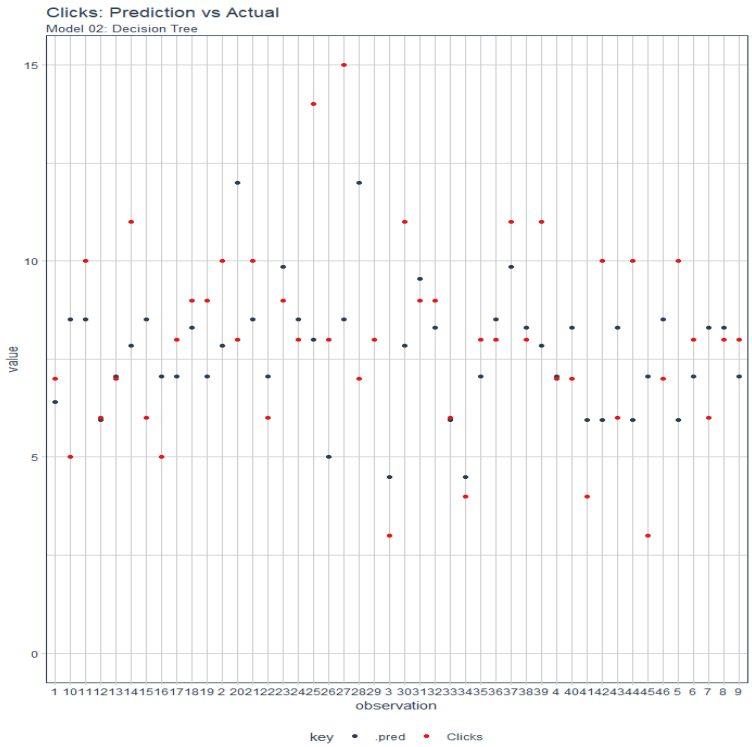
term	estimate	std.error	statistic	p.value
Pageviews	0.2592653	0.0288036	9.0011416	0.0000000
Experiment	0.8715673	0.3452024	2.5248007	0.0124259
DOW	0.0370374	0.0821735	0.4507217	0.6527237
Date	0.0000000	0.0000001	0.1308954	0.8960019
(Intercept)	-10.9806254	86.3526635	-0.1271602	0.8989532





Findings & Conclusions: As we can see in the graph, Pageviews and Experiment are judged strong predictors with a p-value less than 0.05. Also, we note that the coefficient of Experiment is 0.8715673, and because the term is binary (0 or 1) this can be interpreted as increasing clicks by 0. 8715673 per day when the Experiment is run.

Decision Trees



.metric	.estimator	.estimate
rmse	standard	2.0414520
rsq	standard	0.3170865
mae	standard	1.5346462

.metric	.estimator	.estimate
rmse	standard	2.510079
rsq	standard	0.088926
mae	standard	1.912689

(top is from linear regression model; bottom is decision trees model)

Findings & Conclusion: The MAE of the predictions of the linear model is at 1.5 clicks per day. The MAE of the predictions of decision tree is at 1.9 clicks per day.
Attention: the two modelling (decision tree & linear regression) didn't show a good predictive performance.

Top 3 KPIs

1. view-to-click conversion rate

Goal: Aimed to test if conversion rate improved as page goes from "control" to "experiment" version.

Steps:
I calculate the conversion rate:
Pa is the control group. Here Pa=21.84%
Pb is the experiment group, Pb= 23.64%
It seems conversion rate improves as page goes from "control" to "experiment" version. I'll use hypothesis testing to figure it out:

Hypothesis testing:
Here,
H_0: Pb-Pa<=0.
H_1: Pb-Pa >0
Here, establish the z-score as:

$$Z = \frac{\hat{P}_b - \hat{P}_a}{\sqrt{\frac{\hat{P}_b(1-\hat{P}_b)}{n} + \frac{\hat{P}_a(1-\hat{P}_a)}{n}}}$$

(reference from: https://en.wikipedia.org/wiki/Test_statistic)

z-score here is equal to 2.69; and 1-dnorm(z) is equal to 0.99.
So, from above result, we fail to reject the null hypothesis and accept H_0.

Findings & Conclusion: **Conversion rate doesn't improve significantly** as page goes from "control" to "experiment" version.

2. Average time on page

Goal: Aimed to test if average time on page shortened because of the GIF.

Steps:

I calculated average & total time spent between viewing and clicking on webpage;

	group	mean.time	sum.time
	<chr>	<dbl>	<dbl>
1	control	49.4	72467.
2	experiment	49.5	69161.

Findings: It appears that experiment group and control group doesn't differ because people spent as much time on page before clicking and being directed to subcategories. To further prove the hypothesis that experiment group doesn't differ control group in terms of time spend, here I use T-test.

Before T-test, test of homoscedasticity is needed (i.e. are the variances homogenous?)

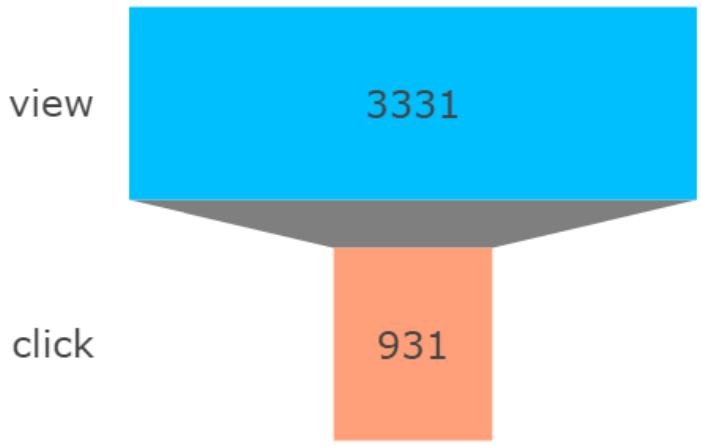
```
        F test to compare two variances

data:  CG$Duration and TG$Duration
F = 1.012, num df = 1467, denom df = 1396, p-value = 0.822
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9122625 1.1224679
sample estimates:
ratio of variances
           1.01199
```

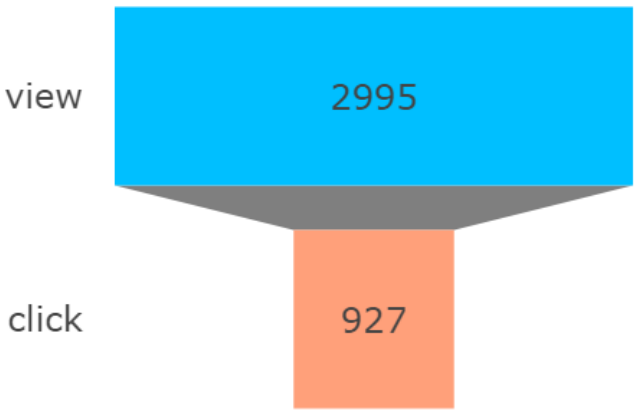
Findings & Conclusion: Since the p-value of 0.822 is greater than the significance level of 0.05, thus, we fail to reject the H_0. That is to say, **time spent on page doesn't differ between these two groups.**

3. Counts of homepage visitors— (Funnel Plot)

Control group



Experiment group



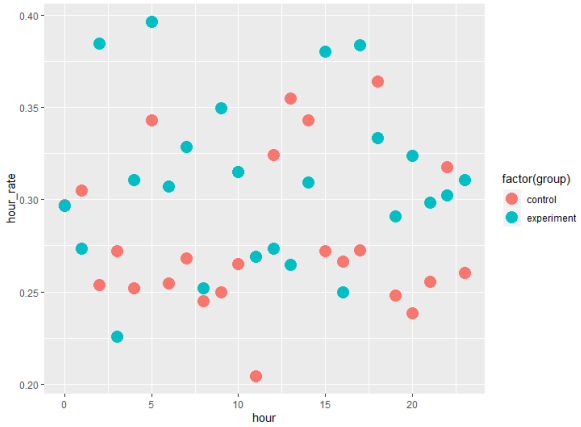
Conclusion

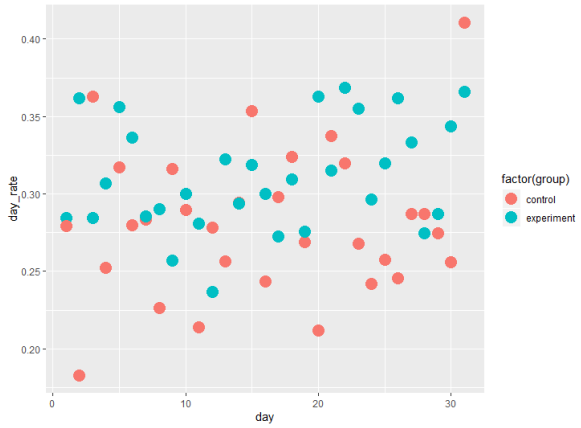
From above 3 KPIs analysis, it cannot say that new type of showstopper is “success”. There are no **significant** improvements through above three KPIs. So, I think we need to extend the observation time or modify the new function to test it again.

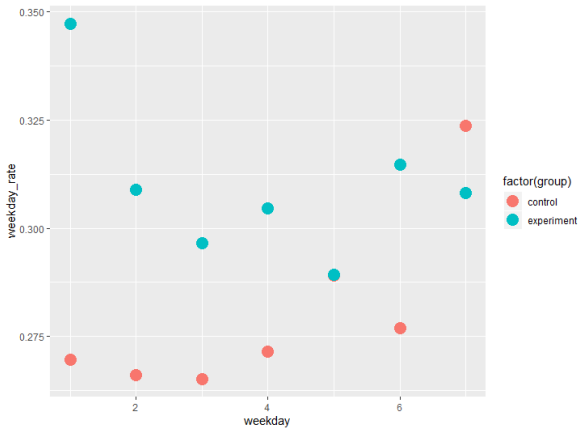
Supplementary analysis

Goal: I would explore deeply about click rate based on time dimension in this part as my supplementary analysis.
I defined weekday (weekend day & work day), and worktime (8:00am - 18:00pm) & work relax time (19:00pm – 7:00am).

Plots







Linear Models

I built five linear models and the below three have good performance.

```
> lm3 <- lm(train$month_rate~group+month,data=train)
> summary(lm3)

Call:
lm(formula = train$month_rate ~ group + month, data = train)

Residuals:
      Min        1Q    Median        3Q       Max
-0.021839 -0.006954  0.000742  0.014099  0.060985

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.882e-01  5.865e-04  491.46   <2e-16 ***
groupexperiment 3.005e-02  4.081e-04   73.65   <2e-16 ***
month          -9.364e-04  5.523e-05  -16.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0162 on 6323 degrees of freedom
Multiple R-squared:  0.4741,    Adjusted R-squared:  0.4739
F-statistic:  2850 on 2 and 6323 DF,  p-value: < 2.2e-16

> lm4 <- lm(train$work_relax_rate~group+work_relax_time,data=train)
> summary(lm4)

Call:
lm(formula = train$work_relax_rate ~ group + work_relax_time,
    data = train)

Residuals:
      Min        1Q    Median        3Q       Max
-0.005330 -0.004156 -0.004156  0.004603  0.004770

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              2.778e-01  9.831e-05 2825.40   <2e-16 ***
groupexperiment          3.003e-02  1.183e-04  253.87   <2e-16 ***
work_relax_timeworktime 3.746e-03  1.184e-04   31.63   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.004697 on 6323 degrees of freedom
Multiple R-squared:  0.9119,    Adjusted R-squared:  0.9118
F-statistic: 3.271e+04 on 2 and 6323 DF,  p-value: < 2.2e-16


> lm5 <- lm(train$weekend_rate~group+weekend_time,data=train)
> summary(lm5)

Call:
lm(formula = train$weekend_rate ~ group + weekend_time, data = train)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0006456 -0.0002742  0.0002460  0.0002460  0.0007139

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.966e-01  1.122e-05   26425   <2e-16 ***
groupexperiment     2.998e-02  1.058e-05    2835   <2e-16 ***
weekend_timeworkday -2.365e-02  1.180e-05   -2004   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00042 on 6323 degrees of freedom
Multiple R-squared:  0.9995,    Adjusted R-squared:  0.9995
F-statistic: 6.036e+06 on 2 and 6323 DF,  p-value: < 2.2e-16
```

Findings & Conclusion: The above three models shows the click rate correlates the time dimension of the two groups. From the above three models, the experiment group improves the click rate compared to the control group.

Besides, we can also find on work days (lm5), the click rate becomes negative increase, which accords with a general life rule (people are busy). While on their work relax time (lm4), the click rate increases, which is also normal.