Impact of Email Personalization on Click Rate

Srujana Gali, Anmol Srivastava, Syed Ghalib Qadri, Sarmad Kahut, Mugdha Khairnar

Introduction:

In today's digital age, organizations are constantly seeking new and innovative ways to capture the attention of consumers and drive engagement. One promising approach is the use of personalized notifications, emails, and SMS messages, which are tailored to the individual preferences and behaviors of each customer. By delivering targeted and relevant messages directly to a consumer's device, organizations hope to increase the likelihood that they will visit their website and explore their offerings.

But does this strategy work? In this experiment, we will investigate the impact of personalized emails on user engagement and explore whether this approach is an effective way to drive business growth. Through careful analysis of user behavior and response rates, we hope to shed light on this important question and help organizations make informed decisions about how best to reach and engage their customers.

Goal of the Experiment:

The goal of the experiment was to observe the effect of personalized emails on the likelihood of a person engaging with the survey. We decided to conduct the experiment on graduate students at Boston University. We collected the email addresses of our classmates as well as other students in the Masters programs at Questrom to conduct our experiment.

Survey Design:

Since our goal was to observe the effect of personalized emails on whether a person would fill out our survey, we decided to create a Culinary survey to ask the about respondent's food preferences (Halal, Kosher, Vegetarian, etc.) What respondents filled in the survey did not matter to us, however, we are interested in whether participants clicked on our email link, and whether any personalisation in the email impacted the rate at which they clicked the email.

Thus, to effectively gauge the impact of different levels of personalization (email header, subject etc.), we decided to utilize 2 separate treatment arms (Treatment Arm 1 and Treatment Arm 2). The email text body for all 3 groups contained the same invitation to complete the survey, links to the survey, and mention of a chance to win a $50 voucher if they fill out the survey. The differences were that Treatment1 also had a personalized greeting, while Treatment 2 had both a personalized greeting as well as personalized subject with mention of the $50 voucher. The details of the treatment and treatment arms are:

1) Control Number of Emails: 30 Description: Standard email text body with no personalization (static email).

2) Treatment Arm 1 Number of Emails: 30 Description: Standard email text body with personalized greeting.

3) Treatment Arm 2 Number of Emails: 30 Description: Standard email text body with personalized greeting and personalized subject including the chance to win $50.

   Sample emails for the above are attached in the appendix.

We decided to randomly assign the participants to the treatment and control groups by using the below function in R-Studio. This function randomly assigned our survey participants into control or one of the treatment arms.

```r
# Read in the data from the original file
emails <- read_csv("Emails.csv")

# Modify the data by adding a new column of randomly generated values
emails$values <- sample(0:2, nrow(emails), replace=TRUE)

# Write the modified data to a new file
write_csv(emails, "Emails_modified.csv")
```

Sending out the email:

To send out the emails at the same time, and to track and monitor the emails and click rate, we used Yet Another Mail Merge (YAAM), which is an extension to Google sheets which allows us to track and send emails to multiple recipients at the same time. Through the free version, we were able to track 3 things:

1) Whether the email was sent to the recipient

2) Whether the email was opened by the recipient
3) Whether the recipient clicked on the link

Furthermore, we were able to track whether the recipients filled out the emails through Google forms.

To ensure uniformity, we sent the emails to all 3 groups at the same time, at 11 am on Saturday (11th March 2023), we gave each respondent 72 hours to track respond to the email before we stopped collecting responses (11 am 14th March). We did this to ensure that the variations in the time of the day do not impact the response rates for each of the control and treatment arms.

To ensure that the chances of 'Failure to Treat' were minimized (due to emails being sent directly to participants spam folders), we sent the emails using BU email addresses, to people who are already studying in BU. This precluded the possibility of the emails being consigned to spam folders, as they are within organization emails.

Exploratory Data Analysis

Once the 72 hour data collection period had passed, we stopped collecting responses, and compiled our results into one csv file ("Compiled Data (1).csv") which we then opened in R-studio for further analysis.

After loading our data, we first performed exploratory data analysis to ascertain how many people in each category (control, treatment arm 1, treatment arm 2) opened our emails and clicked on the links.

```r
email_data <- read.csv("Compiled Data.csv")
head(email_data)

##   ID Group Open Click Gender
## 1  1     2    1     0      0
## 2  2     2    1     1      1
## 3  3     2    1     0      0
## 4  4     2    1     0      0
## 5  5     2    1     0      1
## 6  6     2    1     0      0

table(email_data$Group)

##
##  0  1  2
## 30 30 30
```

From the above output, we can see that number of observations taken for Control Group, Treatment Group 1 and Treatment Group 2 are the same (i.e., 30 observations).

```r
aggregate(email_data[,3:4],by=list(email_data$Group),sum)

##   Group.1 Open Click
## 1       0   26     6
## 2       1   25     8
## 3       2   27    12
```

We can see that the total number of people who opened the mail in the control group are 26, Treatment Group 1 are 25 and Treatment 2 is 27. And when it comes to Click Rate, Control group has 6 clicks, Treatment 1 has 8 and Treatment 2 has 12.

```r
fail <- sum(email_data$Open == 0)
cat("Number of people who did not open the email: ", fail, "\n")

## Number of people who did not open the email:  12

Per_fail<- fail/(sum(email_data$Open == 0)+sum(email_data$Open == 1))*100

cat(Per_fail, "percentage of people  did not open the email")

## 13.33333 percentage of people  did not open the email
```

The above information shows that out of 90 survey participants, 12 did not open the emails. However, this could not be assigned to 'failure to treat' as we do not know whether these 12 people did not view the emails, or whether they viewed the emails but decided to open them. As mentioned earlier, we had taken steps to ensure that emails would not be sent to spam folders by ensuring that both the sender and receiver emails are BU email ids

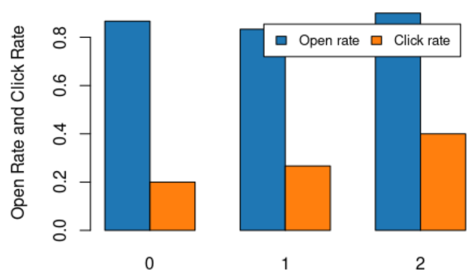
Open and Click Rates:

```
open_rate <- aggregate(email_data$Open, by=list(email_data$Group), mean)
names(open_rate)[2] <- "Open Rate"
click_rate <- aggregate(email_data$Click, by=list(email_data$Group), mean)
names(click_rate)[2] <- "Click Rate"

open_rates <- aggregate(Open ~ Group, data = email_data, mean)
click_rates <- aggregate(Click ~ Group, data = email_data, mean)

rates <- merge(open_rates, click_rates, by = "Group")

# Create a bar plot of the open and click rates by group
barplot(t(as.matrix(rates[, 2:3])), beside = TRUE,
        names.arg = rates$Group,
        ylab = "Open Rate and Click Rate",
        col = c("#1F77B4", "#FF7F0E"),
        legend.text = c("Open rate", "Click rate"),
        args.legend = list(x = "topright", ncol = 2, cex = 0.8, inset = 0.05))
```



Open Rates:

The above bar chart and tables shows us that the Open-rate for all three subsets are roughly similar, (Control: 0.8666667, Treatment 1: 0.8333333, Treatment 2: 0.9000000). If the email was not opened, the Control group and Treatment 1 saw the same standardized email heading, while Treatment 2 saw the personalized version with the mention of the $50 voucher. The results show that those who saw Treatment 2 had a higher probability of opening the email, but we will need to check whether this result is statistically significant, which we will do later on.

Click Rates:

What we see instead for click rates is that the Control has the lowest click rate for the link contained in the email (at 0.2000000), while Treatment 1 had a higher click rate (at 0.2666667) and Treatment 2 had the highest (at 0.4000000). This tells us that 40% of the people who received Treatment 2 opened clicked on the email.

So what we see from these results is that while most respondents (83-90%) opened the emails regardless of whether they were in Control or any of the treatment arms, those who were in Treatment 1 had higher click rates than the Control. This shows the impact of having a personalized email greetings.

What is even more interesting is that Treatment 2 had the highest Click rate, despite having a similar Open rate to the Control. This means that while all 3 groups opened at similar rates, those in Treatment 2 were far more likely to click on the links than the other 2 groups. This might be due to all 3 groups only reading the heading, as the text body for all 3 emails mentioned the $50 gift voucher, but only Treatment 2 mentioned it in the subject.

Now we will explore the impact in more detail:

Gender Differences:

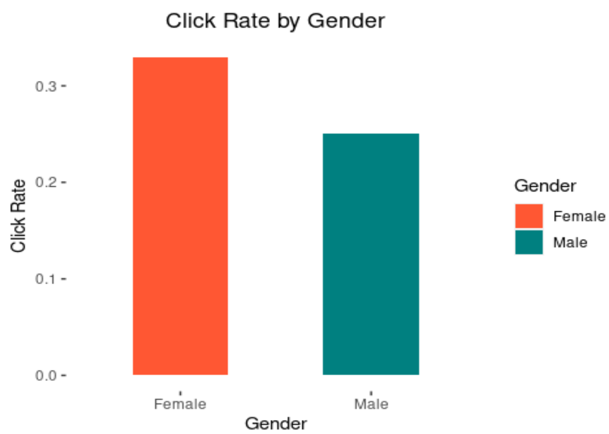
```
male_click <- mean(email_data[email_data$Gender == 0, "Click"])
female_click <- mean(email_data[email_data$Gender == 1, "Click"])
cat("Average click rate for males: ", male_click, "\n")

## Average click rate for males:  0.2222222

cat("Average click rate for females: ", female_click, "\n")

## Average click rate for females:  0.3555556
```

```r
click_rate <- data.frame(Gender = c("Male", "Female"), Click_Rate = c(0.25, 0.33))
ggplot(click_rate, aes(x=Gender, y=Click_Rate, fill=Gender)) +
  geom_bar(stat="identity", width=0.5) +
  ggtitle("Click Rate by Gender") +
  xlab("Gender") +
  ylab("Click Rate") +
  scale_fill_manual(values = c("#FF5733", "#008080")) +
  theme(plot.title = element_text(hjust = 0.5),
        panel.background = element_rect(fill = "white"))
```



This shows that women are more likely to click the email link than men are (0.35 vs 0.22). This follows more general trends where women have higher response rates to surveys than men. For example, an article by the National Institute of Health shows that women are more likely to participate in health surveys than men. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7714285/. Capturing this difference was one of the reasons why we blocked for gender, ensuring there were equal number of men and women in each group.

Estimated Average Treatment Effect:

```r
ate_arm1 <- with(email_data, mean(Click[Group == 1]) - mean(Click[Group == 0]))
ate_arm2 <- with(email_data, mean(Click[Group == 2]) - mean(Click[Group == 0]))
print(paste0("ATE for Treatment Arm 1: ", round(ate_arm1, 3)))

## [1] "ATE for Treatment Arm 1: 0.067"

print(paste0("ATE for Treatment Arm 2: ", round(ate_arm2, 3)))

## [1] "ATE for Treatment Arm 2: 0.2"
```

The Average Treatment Effect of Treatment group 1 is 0.067 which means that those who were assigned to treatment group 1 were more likely to click on the survey link than the control group. Similarly, the ATE of treatment group 2 is 0.2 which means that they are more likely to click on the survey link than the control group.

However, as shown earlier there were differences in click rates between men and women. Thus, we will now look at the conditional average treatment effect to explore these differences.

Conditional Average Treatment Effect

For Treatment 1 by Gender:

```r
# Subset the data for male and female separately
male <- email_data[email_data$Gender == 0, ]
female <- email_data[email_data$Gender == 1, ]

# Calculate the CATE for male and female separately
m_ate <- mean(male[male$Group == 1, "Click"]) - mean(male[male$Group == 0, "Click"])
f_ate <- mean(female[female$Group == 1, "Click"]) - mean(female[female$Group == 0, "Click"])

# Print the CATE for male and female
cat("CATE for male: ", m_ate, "\n")

## CATE for male:  -0.1333333
```

```
cat("CATE for female: ", f_ate, "\n")

## CATE for female:  0.2666667
```

When we account for gender we see that for treatment group 1 the ATE of 0.067 is not homogeneous. When looking at the males separately, the CATE of males is -0.1333333 which means that for men the effect of treatment group 1 is negative. On the other hand for women, the CATE of treatment 1 is 0.2666667 which shows that the impact on click rate is positive for women.

For Treatment 2 for Gender:

```
# Subset the data for male and female separately
male <- email_data[email_data$Gender == 0, ]
female <- email_data[email_data$Gender == 1, ]

# Calculate the CATE for male and female separately
m_ate <- mean(male[male$Group == 2, "Click"]) - mean(male[male$Group == 0, "Click"])
f_ate <- mean(female[female$Group == 2, "Click"]) - mean(female[female$Group == 0, "Click"])

# Print the CATE for male and female
cat("CATE for male: ", m_ate, "\n")

## CATE for male:  0

cat("CATE for female: ", f_ate, "\n")

## CATE for female:  0.4
```

When observing Treatment group 2, for men we see that the CATE is 0, which means that there is no impact of treatment 2 compared to the control group.

Women on the other hand have a CATE of 0.4, which means that Treatment group 2 has a positive impact for women. Thus the ATE of 0.2 for Treatment group 2 is driven primarily by women when comparing to the control.

Thus, overall we see that our treatment groups had a much higher impact on women, who were then driving the overall ATE caused by the treatment effects.

Simple Regression

```
model <- lm(Click ~ Group*Gender, data = email_data)
summary(model)

##
## Call:
## lm(formula = Click ~ Group * Gender, data = email_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5556 -0.2222 -0.2222  0.4444  0.8444
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.222e-01  1.045e-01   2.128   0.0362 *
## Group         1.730e-16  8.091e-02   0.000   1.0000
## Gender       -6.667e-02  1.477e-01  -0.451   0.6529
## Group:Gender  2.000e-01  1.144e-01   1.748   0.0840 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4432 on 86 degrees of freedom
## Multiple R-squared:  0.08654,	Adjusted R-squared:  0.05467
## F-statistic: 2.716 on 3 and 86 DF,  p-value: 0.04968
```

Here, we see that the impact of treatment group is that when we regress the click rate by the treatment group, the results are not significant. Similarly if we take the impact of click rate by gender the results are not significant. However, if we combine the two the results become marginally significant.

Now to calculate the effect size, we will have to calculate the Cohens' D. However before that, we should check there is a statistically significant difference between the two groups. For that purpose, we will carry out the Welch Two Sample t-test on both the treatment groups compared to the control.

The magnitude of the t value is large which means that there is the difference between the two groups(Treatment group 1 and the control group).This observation is supported by the fact that the p-value is <0.05 which is less than the significance level, therefore we reject the null hypothesis and conclude that there is significant difference between the means of the two groups.

We have done the t-test for Treatment group 1. We will now do the t-test for Treatment group 2 as well:

```
t.test(email_data[email_data$Group==2, email_data$Click], email_data[email_data$Group==0,
email_data$Click])

##
##  Welch Two Sample t-test
##
## data:  email_data[email_data$Group == 2, email_data$Click] and email_data[email_data$Group == 0,
email_data$Click]
## t = -136.81, df = 1558, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -60.86024 -59.13976
## sample estimates:
## mean of x mean of y
##      15.5      75.5
```

The same results hold when we compare the control group with treatment group 2. The p-value is <0.05 and the magnitude of the t-value is large which means that there is a significant difference between the two groups.

Randomization Check

Before moving forward, we want to ascertain whether there are no differences between the groups other than the treatments. For that, we will be checking the randomization using prop test:

```
Randomization_model <- lm(Gender ~ Group, data = email_data)
summary(Randomization_model)

##
## Call:
## lm(formula = Gender ~ Group, data = email_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -0.5   -0.5    0.0    0.5    0.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.000e-01  8.427e-02   5.933 5.75e-08 ***
## Group       1.433e-17  6.528e-02   0.000        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5056 on 88 degrees of freedom
## Multiple R-squared:  1.297e-31,  Adjusted R-squared:  -0.01136
## F-statistic: 1.141e-29 on 1 and 88 DF,  p-value: 1
```

From the above regression model, we can see that the p-value is 1, which means that we cannot reject the null hypothesis. The randomization was done properly.

Statistical Power

The Cohens' D measures the standardized effect size of the difference between the mean of treatment and the mean of control. We define Cohens' D as the standard which is used to measure the difference between the means of two groups. In this case, we will be comparing each of the treatment groups to the control.

```
cohens_d_1 <- ate_arm1/sd(email_data$Click)
cohens_d_1

## [1] 0.1462677

cohens_d_2 <- ate_arm2/sd(email_data$Click)
cohens_d_2

## [1] 0.438803
```

The Cohen's D value for the treatment group 1 is 0.146 which means that it has none to slight effect. While the Cohen's D value for the treatment group 2 is 0.438 which means that it has moderate effect.

We can see that the power is 0.08 which is significantly small indicating that study is under powered to detect effect size. The power is 0.08 which means that there is a very small chance that the experiment will detect an effect when there is an effect.

```
pwr.t2n.test(n1 = 30, n2 = 30, d = cohens_d_2,
sig.level = .05, power = NULL)

##
##      t test power calculation
##
##              n1 = 30
##              n2 = 30
##               d = 0.438803
##       sig.level = 0.05
##           power = 0.386576
##     alternative = two.sided
```

Since the effect size is significantly moderate, and statistical power is 0.386 which is considered low, it means that there is only 38.6% chance of detecting true effect.

With the treatment group 1 of this experiment, if we want to have a higher statistical power as high as 0.9 we need 983 observations.

```
pwr.t.test(n = NULL, d = cohens_d_2, sig.level = .05, power = 0.9)

##
##      Two-sample t test power calculation
##
##               n = 110.11
##               d = 0.438803
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

With the treatment group 2 of this experiment, if we want to have a higher statistical power as high as 0.9 we need 110 observations.

Limitations:

The Experiment we have done might have only investigated the impact of using names in emails and not consider other personalisation factors, such as geographic location.The results of this experiment may not be generalisable to other populations or contexts beyond Boston University. The sample size of 90 participants is relatively small, which may limit the statistical power of the analysis.The participants in this experiment were selected based on their availability and willingness to participate, which may introduce selection bias.

Conclusion:

Through this experiment, we have seen that the open rates for the different groups were roughly the same, however the click rates were different. While the ATEs were 0.067 for treatment arm 1, and 0.2 for treatment arm 2 (compared to the control) we saw that this difference was primarily driven by the women. This was validated further by the CATE values. When we conducted the regression, we can say that the outcome approaches statistical significance when we combine the effects of gender and treatment groups. We have also performed randomization check using lm regression model, to reassure if our randomization was performed correctly. We then conducted a Welch two t sample test and calculated the Cohens' D to observe a weak effect for treatment 1 and a moderate effect for treatment 2. Thus, personalization in both the subject and greetings of an email leads a moderate effect size (d = 0.438).

Appendix:

Control Group:

Invitation To Participate in Culinary Survey YAMM - Invitation To Participate in Culinary Survey ×

Boston University Survey <sarmadk@bu.edu> Fri, Mar 3, 11:32 AM (13 days ago)

to Visaaln

Dear Students,

We hope you're doing well. As part of our ongoing efforts to improve your experience at BU dining and events, we'd like to invite you to participate in a brief survey regarding food options on campus.

The survey will take just a minute of your time, and your feedback will help us ensure that we are inclusive of everyone's dietary preferences. By sharing your thoughts on the variety and quality of food offered at our events and dining, you will be helping us better serve you in the future.

As a thank you for your participation, each survey participant will have a chance to win a $50 GrubHub gift card.

To begin the survey, simply click on the following link or copy and paste the URL below into your internet browser:

Take the Survey

https://docs.google.com/forms/d/e/1FAIpQLScLpykG_Sr0O6GZdvlKlWXUL-s7N1i8Ce1X49GixAFwdyZi1g/viewform

Thank you for your time and valuable feedback. We look forward to your response.

Sarmad Iqbal Kahut
Boston University

Treatment Group 1:

Invitation To Participate in Culinary Survey YAMM - Invitation To Participate in Culinary Survey ×

Boston University Survey <sarmadk@bu.edu> Fri, Mar 3, 11:36 AM (13 days ago)

to Mahesh

Dear Mahesh,

We hope you're doing well. As part of our ongoing efforts to improve your experience at BU dining and events, we'd like to invite you to participate in a brief survey regarding food options on campus.

The survey will take just a minute of your time, and your feedback will help us ensure that we are inclusive of everyone's dietary preferences. By sharing your thoughts on the variety and quality of food offered at our events and dining, you will be helping us better serve you in the future.

As a thank you for your participation, each survey participant will have a chance to win a $50 GrubHub gift card.

To begin the survey, simply click on the following link or copy and paste the URL below into your internet browser:

Take the Survey

https://docs.google.com/forms/d/e/1FAIpQLScLpykG_Sr0O6GZdvlKlWXUL-s7N1i8Ce1X49GixAFwdyZi1g/viewform

Thank you for your time and valuable feedback. We look forward to your response.

Sarmad Iqbal Kahut
Boston University

Treatment Group 2:

Gokul, Here is Your Chance to Win 50$ GrubHub Voucher

YAMM - Gokul, Here is Your Chance to Win 50$ GrubHub Voucher ×

Boston University Survey <sarmadk@bu.edu> Fri, Mar 3, 11:39 AM (13 days ago)

to Gokulp

Dear Gokul,

We hope you're doing well. As part of our ongoing efforts to improve your experience at BU dining and events, we'd like to invite you to participate in a brief survey regarding food options on campus.

The survey will take just a minute of your time, and your feedback will help us ensure that we are inclusive of everyone's dietary preferences. By sharing your thoughts on the variety and quality of food offered at our events and dining, you will be helping us better serve you in the future.

As a thank you for your participation, each survey participant will have a chance to win a $50 GrubHub gift card.

To begin the survey, simply click on the following link or copy and paste the URL below into your internet browser:

Take the Survey

https://docs.google.com/forms/d/e/1FAIpQLScLpykG_Sr0O6GZdvlKlWXUL-s7N1i8Ce1X49GixAFwdyZi1g/viewform

Thank you for your time and valuable feedback. We look forward to your response.

Sarmad Iqbal Kahut
Boston University