1 Game Fun: Customer Acquisition through Digital Advertising

Disclaimer: “No similarity to actual persons or companies is intended or should be inferred.”

Game Fun is one of the world's top developers of casual mobile games and spends millions of dollars every year on digital advertising. Of particular interest to Game Fun is their efforts on improving their customer acquisition. It is a common belief in the mobile gaming industry that the more paid traffic you have, the better your organic traffic will be. This is known as K-Factor (see appendix for an explanation).

However, measuring the causal effect of online digital advertising has proven to be an extremely challenging task. Towards this end, Game Fun decided to run an A/B experiment.

2 Display Banners Experiment

Game Fun ran an online display banners advertising campaign with the primary objective of increasing its sales on gaming subscription packages. To attract new users, the display ad advertised their most popular game and offered the user a promotion of $25 signing-up bonus. The credits would appear in the customer’s game account and could be used to purchase any further in-app features. Based on historical data, a new customer subscription brings a revenue of $37.5 on average. This results in a net inflow of $12.5 after the $25 credit for the users acquired through this promotion.

The A/B experiment worked as follows. Before the start of the digital ad campaign, Game Fun chose two different websites (“content publishers”) to run the experiment on. The content publishers have randomly assigned their web users to test and control groups. As users browsed on the two websites, the advertising servers checked whether a given user should be show a Game-Fun ad. If the user qualified for a Game-Fun ad, then the ad server checked whether the user was assigned to the test or the control group. If the user belonged to the test set, a Game-Fun ad was displayed to the user. Otherwise, a completely irrelevant ad was displayed to the user.

However, Game-Fun had to pay the content publishers for these irrelevant ads, as well. This raised a tension between the management and the data scientist teams. The management team had two concerns. First, paying for other companies ads is directly decreasing their marketing budget. Second, they didn’t like the fact that some users who saw an irrelevant ad might have signed up for the Game Fun game in the first place, had they been shown their gaming ad (indirect effect – opportunity cost missed). For these two reasons, the management team asked their data scientist team to carefully decide on the best proportion of users to assign to the test and control groups, while at the same time maintaining a statistically valid comparison.

The data scientist team ran a statistical power analysis of the experiment, and decided to allocate 70% of users to test group and the rest 30% to the control group (see appendix for an explanation on power analysis).

3 Analysis of the Experiment

After the completion of the experiment, the results came in and you (the data analytics team) is being asked to analyze them. The data are provided in the file “GameFun.xls”. Each row in the data belongs to an individual customer.

- The first column is the anonymized customer id.
- The second column, “test”, indicates whether the user was part of the treatment group (test =1) or the control group (test=0).
- There are three demographic variables, “gender” (male =1, female =0), “income” (this is measured in thousands), and “gamer” (gamer = 1, if user is a gaming enthusiast; 0, otherwise).
- The website that a customer visited is in the variable “site”.
- The variable “impressions” contains the number of advertising impressions that a customer received. If a customer is in the test group, then all of this customer’s impressions are for the Game-Fun’s ad; if a customer is in the control group, then all of this customer’s impressions are for the irrelevant ad.
- Last, the column “purchase” is the dependent variable, and it indicates if the customer purchased anything within 30 days after her/his conversation to the game (30 days is the expected customer lifetime duration for a mobile gamer). If a customer purchased, then purchased =1; 0, otherwise.

4 Questions

In [1]:
```
library(tidyverse)
library(readxl)
```
executed in 4.65s, finished 20:20:11 2020-04-08

```
Registered S3 methods overwritten by 'ggplot2':
  method         from
  [.quosures     rlang
  c.quosures     rlang
  print.quosures rlang
Registered S3 method overwritten by 'rvest':
  method            from
  read_xml.response xml2
-- Attaching packages --------------------------------------- tidyverse 1.2.1 --
v ggplot2 3.1.1     v purrr   0.3.2
v tibble  2.1.1     v dplyr   0.8.3
v tidyr   0.8.3     v stringr 1.4.0
v readr   1.3.1     v forcats 0.4.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

In [2]:
```
setwd("D:/Spring 2020/BAX423BigDataAnalytics/Homework 1")
df<-read_xlsx('GameFun.xlsx')
head(df)
```
executed in 348ms, finished 20:20:13 2020-04-08

id	test	purchase	site	impressions	income	gender	gamer
1956	0	0	site1	0	100	1	0
45821	1	0	site1	20	70	1	0
59690	1	0	site1	22	100	1	0
18851	0	0	site1	13	90	1	0
60647	1	0	site1	12	60	1	0
30908	1	0	site1	0	40	1	0

4.1 Question 1: Check Probabilistically Equivalent for Groups

Before evaluating the effect of an experiment, it is important to make sure that the experiment was executed correctly. Check whether the test and control groups are probabilistically equivalent on their observables?

```
a. More specific, compare the averages of the income, gender and gamer variables in the test and con
trol groups. You should also report the % difference in the averages. Compute its statistical signif
icance.

b. Briefly comment on what these metrics tell you about probabilistic equivalence for this experimen
t.

c. If you had run this type of analysis BEFORE executing an experiment and found a large difference
 between test and control groups, what you should do?

d. (Open/Ended Question) If you had millions of consumers, your “classic” statistical significance t
ests would not work (this is because the number of samples is used to compute those classic statisti
cal tests). Do some research online and propose what significance test would you do in case you had
 “big data”?https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.201800195#bimj1953-bib-0006
```

In [22]:

```
# df_equi = df.groupby(by='test').mean().iloc[:,[3,4,5]]
# df_percentDiff = pd.DataFrame({'%Diff':(df_equi.iloc[0,:]-df_equi.iloc[1,:])/((df_equi.iloc[0,:]+df_equi.il
# print(df_equi)
# print(df_percentDiff)

df_equi<-df %>%
    select(test,income,gender,gamer)%>%
    group_by(test) %>%
    summarise_all(mean)
col_mean<-(df_equi[1,]+df_equi[2,])/2
col_diff<-abs(df_equi[1,]-df_equi[2,])
perc_diff<-(col_diff/col_mean)*100
print('Average: ')
print(df_equi)

print('%Diff:')
print(perc_diff[,-1])
print('According to the numbers above, it seems that test group and control group are probablistically equiva

df_test<-df%>%
    filter(test=='1')
df_control<-df%>%
    filter(test=='0')
```

executed in 152ms, finished 20:32:32 2020-04-08

```
[1] "Average: "
# A tibble: 2 x 4
   test income gender gamer
  <dbl>  <dbl>  <dbl> <dbl>
1     0   55.2  0.648 0.602
2     1   54.9  0.647 0.601
[1] "%Diff:"
     income     gender      gamer
1 0.4137444 0.09509393 0.08175384
[1] "According to the numbers above, it seems that test group and control group are probablistically equival
ent."
```

In [4]:
```
print('T-test for income: ')
print('Null Hypothesis: There is no significant difference between test group income and control group income
print('Alternative Hypothesis: There is significant difference between test group income and control group in
print('Significance Level: 5%')
var.test(df_test$income, df_control$income, ratio = 1, alternative = "two.sided") ### Variances are equal
print('Sample variances are equal in two goups.')
t.test(df_test$income, df_control$income, alternative = "two.sided", mu = 0, paired = FALSE, var.equal = TRUE
print('Conclusion: At 5% significance level, there is no significant difference between test group income and
```

executed in 87ms, finished 20:20:20 2020-04-08

```
[1] "T-test for income: "
[1] "Null Hypothesis: There is no significant difference between test group income and control group incom
e."
[1] "Alternative Hypothesis: There is significant difference between test group income and control group inc
ome."
[1] "Significance Level: 5%"


        F test to compare two variances

data:  df_test$income and df_control$income
F = 1.0104, num df = 28090, denom df = 11956, p-value = 0.5055
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9801602 1.0413289
sample estimates:
ratio of variances
          1.010376

[1] "Sample variances are equal in two goups."


        Two Sample t-test

data:  df_test$income and df_control$income
t = -1.5206, df = 40046, p-value = 0.1284
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.52136490  0.06581471
sample estimates:
mean of x mean of y
 54.93824  55.16601

[1] "Conclusion: At 5% significance level, there is no significant difference between test group income and
control group income."
```

In [25]:
```
n_test<-length(df_test$gender)
n_control<-length(df_control$gender)
male_test<-sum(df_test$gender)
male_control<-sum(df_control$gender)
```

executed in 27ms, finished 20:33:04 2020-04-08

In [26]:
```
print('T-test for gender: ')
print('Null Hypothesis: There is no significant difference between test group gender and control group gender
print('Alternative Hypothesis: There is significant difference between test group gender and control group ge
print('Significance Level: 5%')
prop.test(c(male_test,male_control), c(n_test,n_control),alternative = c("two.sided"), conf.level = 0.95, cor
print('Conclusion: At 5% significance level, there is no significant difference between test group gender and
```

executed in 40ms, finished 20:33:05 2020-04-08

```
[1] "T-test for gender: "
[1] "Null Hypothesis: There is no significant difference between test group gender and control group gender"
[1] "Alternative Hypothesis: There is significant difference between test group gender and control group gen
der"
[1] "Significance Level: 5%"


        2-sample test for equality of proportions with continuity correction

data:  c(male_test, male_control) out of c(n_test, n_control)
X-squared = 0.011368, df = 1, p-value = 0.9151
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.010898505  0.009666854
sample estimates:
   prop 1    prop 2
0.6472892 0.6479050

[1] "Conclusion: At 5% significance level, there is no significant difference between test group gender and
control group gender"
```

In [27]:
```
n_test<-length(df_test$gamer)
n_control<-length(df_control$gamer)
gamer_test<-sum(df_test$gamer)
gamer_control<-sum(df_control$gamer)
```

executed in 26ms, finished 20:33:08 2020-04-08

In [28]:
```
print('T-test for gamer: ')
print('Null Hypothesis: There is no significant difference between test group gamer and control group gamer')
print('Alternative Hypothesis: There is significant difference between test group gamer and control group gam
print('Significance Level: 5%')
prop.test(c(gamer_test,gamer_control), c(n_test,n_control),alternative = c("two.sided"), conf.level = 0.95, c
print('Conclusion: At 5% significance level, there is no significant difference between test group gamer and
```

executed in 43ms, finished 20:33:09 2020-04-08

```
[1] "T-test for gamer: "
[1] "Null Hypothesis: There is no significant difference between test group gamer and control group gamer"
[1] "Alternative Hypothesis: There is significant difference between test group gamer and control group game
r"
[1] "Significance Level: 5%"


        2-sample test for equality of proportions with continuity correction

data:  c(gamer_test, gamer_control) out of c(n_test, n_control)
X-squared = 0.0065358, df = 1, p-value = 0.9356
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.01102858  0.01004496
sample estimates:
   prop 1    prop 2
0.6013314 0.6018232

[1] "Conclusion: At 5% significance level, there is no significant difference between test group gamer and c
ontrol group gamer"
```

In [1]:
```
print('Overall, for the current dataset, the test group and control group are probablistically equivalent.')
print('If I can run this test before the experiment and find the two groups are not probablistally equivalent
I will re-design the test group and control group and make sure that they are matched.')


print('Question d: If you had millions of consumers, your “classic” statistical significance tests would not
print('Answer: The article "Too Big to Fail: Large Samples and the p-Value Problem" (Lin, 2013) reviewed seve
the "big data"situation. Usually, people would decrease the significant threshold to ajust the test and repor
interval.Some people also notice the audience about the "big data"issue. "A critical evaluation of the curren
(Stefan, 2017) suggests people to calculate the effect size rather than use p-value')
```

executed in 58ms, finished 15:28:34 2020-04-09

```
[1] "Overall, for the current dataset, the test group and control group are probablistically equivalent."
[1] "If I can run this test before the experiment and find the two groups are not probablistally equivalen
t,\nI will re-design the test group and control group and make sure that they are matched."
[1] "Question d: If you had millions of consumers, your “classic” statistical significance tests would not w
ork (this is because the number of samples is used to compute those classic statistical tests). Do some rese
arch online and propose what significance test would you do in case you had “big data”?"
[1] "Answer: The article \"Too Big to Fail: Large Samples and the p-Value Problem\" (Lin, 2013) reviewed sev
eral ways to dealing with\nthe \"big data\"situation. Usually, people would decrease the significant thresho
ld to ajust the test and report the confidence \ninterval.Some people also notice the audience about the \"b
ig data\"issue. \"A critical evaluation of the current “p-value controversy”\"\n(Stefan, 2017) suggests peop
le to calculate the effect size rather than use p-value"
```

5 Question 2: Evaluate the Purchase Rates for Groups

Evaluate the average purchase rates in the test and control for the following groups. For each comparison, report the average purchase rate for the test, average purchase rate for the control and the absolute difference (not the % difference) between the test and control.

```
a. Comparison 1: All customers
b. Comparison 2: Male vs Female customers
c. Comparison 3: Gamers vs Non-Gamers Customers
d. Comparison 4: Female Gamers vs Male Gamers
```

In [41]:
```
print('Comparison 1: All customers')
df_equi<-df %>%
    select(test,purchase)%>%
    group_by(test) %>%
    summarise_all(mean)

col_diff<-abs(df_equi[1,]-df_equi[2,])

print('Average Purchase Rate: ')
print(df_equi)

print('Absolut Difference:')
print(col_diff[,-1])
print('From the result, test group has higher purchase rate than the control group.')
```

executed in 60ms, finished 20:37:54 2020-04-08

```
[1] "Comparison 1: All customers"
[1] "Average Purchase Rate: "
# A tibble: 2 x 2
   test purchase
  <dbl>    <dbl>
1     0   0.0362
2     1   0.0768
[1] "Absolut Difference:"
[1] 0.04060866
[1] "From the result, test group has higher purchase rate than the control group."
```

In [42]:
```
print('Comparison 2: Male vs Female customers')
df_equi<-df %>%
    select(test,purchase,gender)%>%
    group_by(test,gender) %>%
    summarise_all(mean)

male_diff<-abs(df_equi[2,]-df_equi[4,])
female_diff<-abs(df_equi[1,]-df_equi[3,])

print('Average Purchase Rate: ')
print(df_equi)

print('AbsDiff_male:')
print(male_diff[,-c(1,2)])
print('AbsDiff_female:')
print(female_diff[,-c(1,2)])
print('From the result, gender does not affect the purchase rate a lot.')
```
executed in 85ms, finished 20:38:11 2020-04-08

```
[1] "Comparison 2: Male vs Female customers"
[1] "Average Purchase Rate: "
# A tibble: 4 x 3
# Groups:   test [2]
   test gender purchase
  <dbl>  <dbl>    <dbl>
1     0      0   0.0344
2     0      1   0.0372
3     1      0   0.0809
4     1      1   0.0746
[1] "AbsDiff_male:"
[1] 0.03739947
[1] "AbsDiff_female:"
[1] 0.04650289
[1] "From the result, gender does not affect the purchase rate a lot."
```

In [43]:
```
print('Comparison 3: Gamers vs Non-Gamers Customers')
df_equi<-df %>%
    select(test,purchase,gamer)%>%
    group_by(test,gamer) %>%
    summarise_all(mean)

gamer_diff<-abs(df_equi[2,]-df_equi[4,])
nongamer_diff<-abs(df_equi[1,]-df_equi[3,])

print('Average Purchase Rate: ')
print(df_equi)

print('AbsDiff_gamer:')
print(gamer_diff[,-c(1,2)])
print('AbsDiff_nongamer:')
print(nongamer_diff[,-c(1,2)])
print('From the result, the test improve the purchase rate of gamers.')
```
executed in 79ms, finished 20:38:19 2020-04-08

```
[1] "Comparison 3: Gamers vs Non-Gamers Customers"
[1] "Average Purchase Rate: "
# A tibble: 4 x 3
# Groups:   test [2]
   test gamer purchase
  <dbl> <dbl>    <dbl>
1     0     0   0.0374
2     0     1   0.0354
3     1     0   0.0351
4     1     1   0.104
[1] "AbsDiff_gamer:"
[1] 0.06905098
[1] "AbsDiff_nongamer:"
[1] 0.002294685
[1] "From the result, the test improve the purchase rate of gamers."
```

In [44]:
```
print('Comparison 4: Female Gamers vs Male Gamers')
df_equi<-df %>%
    select(test,purchase,gender,gamer)%>%
    filter(gamer==1)%>%
    group_by(test,gender) %>%
    summarise_all(mean)

malegamer_diff<-abs(df_equi[2,]-df_equi[4,])
femalegamer_diff<-abs(df_equi[1,]-df_equi[3,])

print('Average Purchase Rate: ')
print(df_equi[,-4])

print('AbsDiff_malegamer:')
print(malegamer_diff[,-c(1,2,4)])
print('AbsDiff_femalegamer:')
print(femalegamer_diff[,-c(1,2,4)])
print('From the result, the test has effect on both male gamers and female gamers.')
```

executed in 96ms, finished 20:38:31 2020-04-08

```
[1] "Comparison 4: Female Gamers vs Male Gamers"
[1] "Average Purchase Rate: "
# A tibble: 4 x 3
# Groups:   test [2]
   test gender purchase
  <dbl>  <dbl>    <dbl>
1     0      0   0.0320
2     0      1   0.0373
3     1      0   0.110
4     1      1   0.101
[1] "AbsDiff_malegamer:"
[1] 0.06412899
[1] "AbsDiff_femalegamer:"
[1] 0.0780506
[1] "From the result, the test has effect on both male gamers and female gamers."
```

5.1 Question 3: Assess the Revenue

Assess the expected revenue in the test vs. control for the following comparisons:

```
a. Comparison 1: All customers
b. Comparison 4: Female Gamers vs Male Gamers
```

In [34]:
```
print('Comparison 1: All customers')
cat('Test Group Expected Revenue: $',sum(df_test$purchase*37.5),'\n')
cat('Control Group Expected Revenue: $',sum(df_control$purchase*37.5))
print('The promotion increased the expected revenue.')
```

executed in 35ms, finished 20:33:26 2020-04-08

```
[1] "Comparison 1: All customers"
Test Group Expected Revenue: $ 80925
Control Group Expected Revenue: $ 16237.5[1] "The promotion increased the expected revenue."
```

In [35]:
```
print('Comparison 2: Female Gamers vs Male Gamers')
df_gamer<-df %>%
    select(test,purchase,gender,gamer)%>%
    filter(gamer==1)

df_gamer$revenue<-df_gamer$purchase*37.5
df_gamerR<-df_gamer%>%
    select(test,gender,revenue)%>%
    group_by(test,gender)%>%
    summarise_all(sum)
print(df_gamerR)
print('This promotion increased revenue no matter the gamer is female or male.')
```

executed in 71ms, finished 20:33:29 2020-04-08

```
[1] "Comparison 2: Female Gamers vs Male Gamers"
# A tibble: 4 x 3
# Groups:   test [2]
   test gender revenue
  <dbl>  <dbl>   <dbl>
1     0      0   3038.
2     0      1   6525
3     1      0  24750
4     1      1  41438.
[1] "This promotion increased revenue no matter the gamer is female or male."
```

5.2 Question 4: Recommendation

Based on your previous answers, provide a brief recommendation to your management team summarizing the expected financial outcome for Game-Fun.

```
a. Should Game-Fun run this promotion again in the future? If no, explain why. If yes, should Game-F
un offer it to all customers or a targeted segment.
```

In [8]:
```
print('According to our findings, Game-Fun should run this promotion in the future. Since we found that this
is work on gamers rather than non-gamers, Game-Fun should offer it to gamers.')
```

executed in 23ms, finished 18:54:51 2020-04-06

```
[1] "According to our findings, Game-Fun should run this promotion in the future. Since we found that this p
romotion \nis work on gamers rather than non-gamers, Game-Fun should offer it to gamers."
```

6 Appendix

6.1 K-Factor Explained

Assume you have a market (e.g., a US state or a country) where you have almost zero daily organic downloads. You agree to buy users though targeted advertisement. Let’s say that you buy 10,000 users. If after a specific period of time, you suddenly have 12,000 users, you know that 2000 of them came as an indirect result of your paid user acquisition campaign. So, your KFactor is 1.2. A K-Factor higher than 1 is considered viral, and the higher the K factor, the better.

6.2 Statistical Power

In order to determine the fraction of users that need to be assigned to the control group, a data scientists needs to consider three factors: 1) baseline conversion rate, 2) campaign reach and 3) expected minimum advertising lift. See here for more info and a calculator: [https://www.optimizely.com/sample-size-calculator/ (https://www.optimizely.com/sample-size-calculator/)](https://www.optimizely.com/sample-size-calculator/)