Tuango Case write-up

Yuhan Xu 474154

2/5/2020

Load the data

```
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.0 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     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()

library(gmodels)
library(statar)
tuango = read.csv("tuango.csv")
attach(tuango)
```

Q1

Q: What percentage of customers responded (i.e. bought anything) after the push message?

```
CrossTable(buyer, digits=4)

##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  13939
##
##
##           |         0 |         1 |
##           |-----------|-----------|
##           |     13507 |       432 |
```

```
##            |     0.9690 |     0.0310 |
##            |-----------|-----------|
##
##
##
##
```

A: About 3.1% customers responded.

Q2

Q: Of those who bought, what was the average spending? (Hint: constrain the summary command with the buyer==1")

```
summarise(tuango[which(buyer == 1),], avg_spending = mean(ordersize))

##   avg_spending
## 1     202.3565
```

A: The average spending was 202.3565RMB.

Q3

Q: Create quintile variables for recency, frequency and monetary. (Hint: review the file "RFM_BBB.R" which goes through the calculations for the Bookbinders RFM analysis)
A:

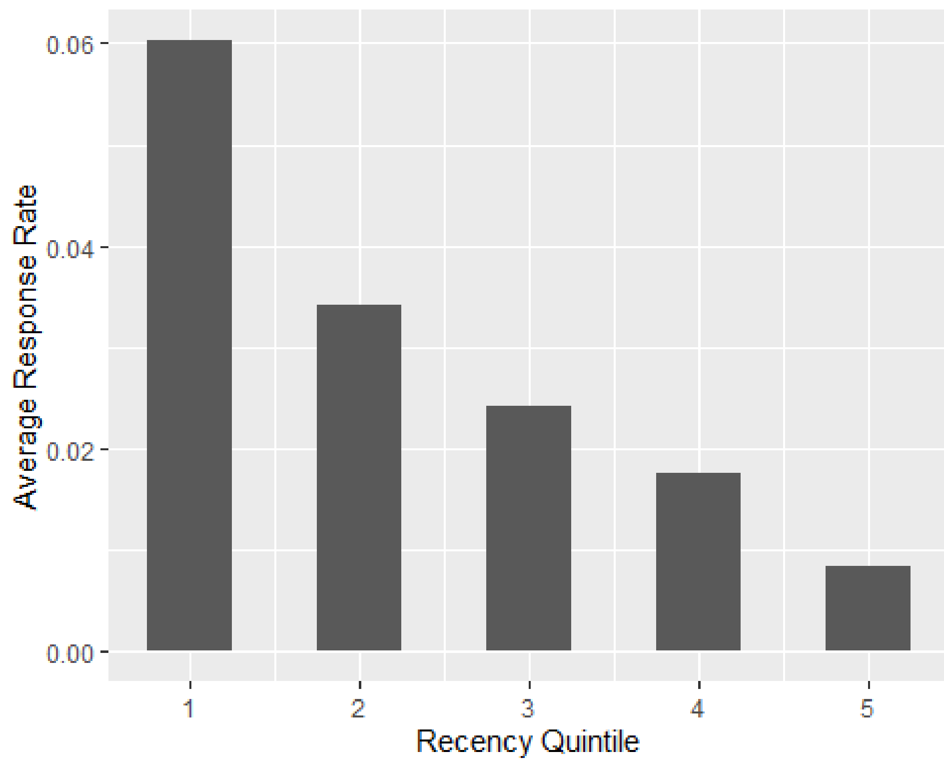
```
tuango = tuango %>%
  mutate(rec_quin = xtile(recency, 5),
         fre_quin = xtile(frequency, 5),
         mon_quin = xtile(monetary, 5))

tuango$fre_quin <- max(tuango$fre_quin) + 1 - tuango$fre_quin
tuango$mon_quin <- max(tuango$mon_quin) + 1 - tuango$mon_quin
attach(tuango)

## The following objects are masked from tuango (pos = 3):
##
##     buyer, category, frequency, mobile_os, monetary, ordersize,
##     platform, recency, rfm1, rfm2, userid

tuango %>% group_by(rec_quin) %>% summarise(mean(recency))

## # A tibble: 5 x 2
##   rec_quin `mean(recency)`
##      <int>           <dbl>
## 1        1            9.37
## 2        2           12.6
## 3        3           22.1
## 4        4           51.0
## 5        5          184.
```

```r
tuango %>% group_by(fre_quin) %>% summarise(mean(frequency))

## # A tibble: 4 x 2
##   fre_quin `mean(frequency)`
##      <dbl>             <dbl>
## 1        1              6.59
## 2        2              3.42
## 3        3              2
## 4        5              1

tuango %>% group_by(mon_quin) %>% summarise(mean(monetary))

## # A tibble: 5 x 2
##   mon_quin `mean(monetary)`
##      <dbl>            <dbl>
## 1        1           220.
## 2        2            85.3
## 3        3            48.1
## 4        4            26.9
## 5        5             9.79
```

Q4

Q: Create a bar chart showing the response rate (i.e., the proportion of customers who bought something) to this deal by recency quintile.
A:

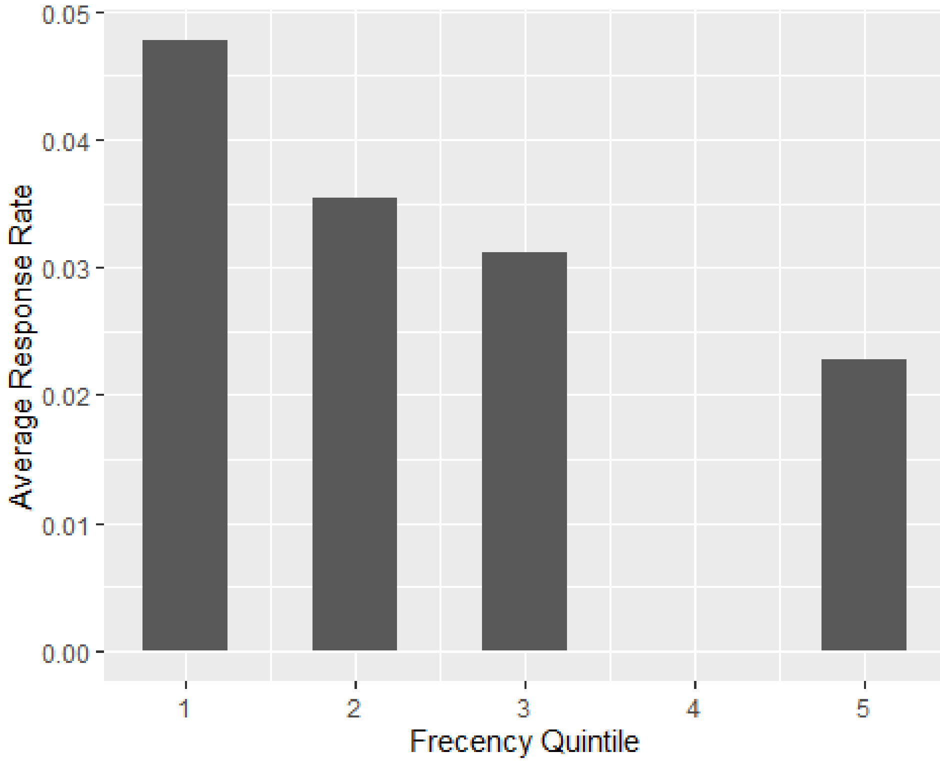
```r
avg_resp_rate_rec <- tuango %>% group_by(rec_quin) %>%
  summarise(avg_resp_rate_rec = mean(buyer))
bar_avg_resp_rate_rec <-
  ggplot(data=avg_resp_rate_rec,
         aes(x = rec_quin, y = avg_resp_rate_rec)) +
  labs(x="Recency Quintile", y="Average Response Rate") +
  geom_bar(stat="identity", width=0.5)
bar_avg_resp_rate_rec
```



Q5

Q: Create a bar chart showing the response rate to this deal by frequency quintile.
A:

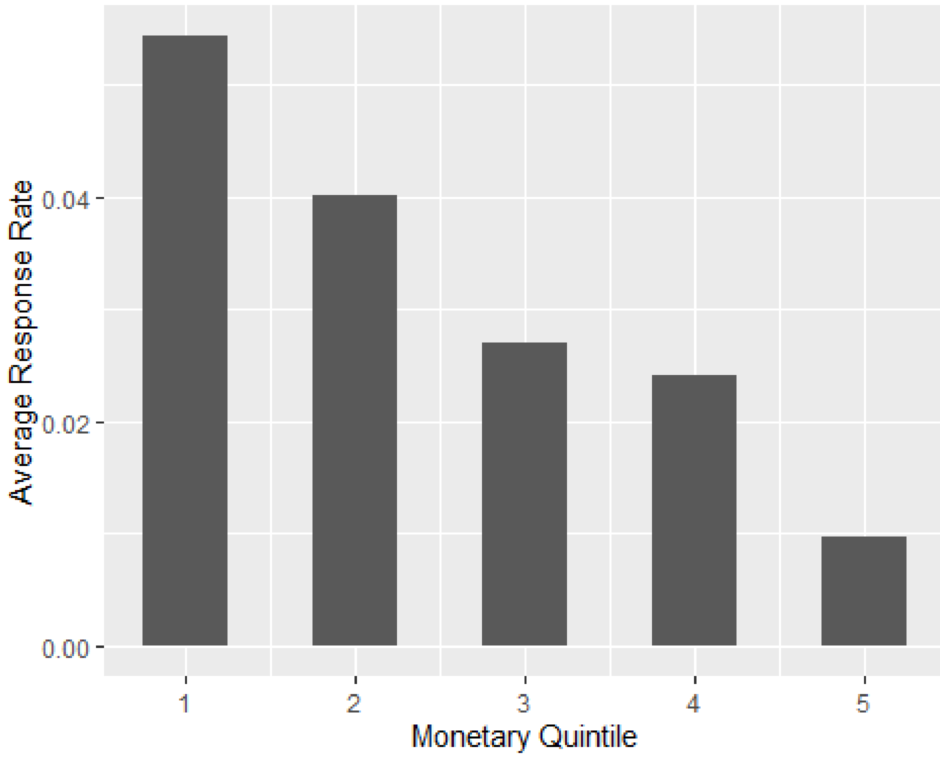
```
avg_resp_rate_fre <- tuango %>% group_by(fre_quin) %>%
  summarise(avg_resp_rate_fre = mean(buyer))
bar_avg_resp_rate_fre <-
  ggplot(data = avg_resp_rate_fre,
         aes(x = fre_quin, y = avg_resp_rate_fre)) +
  labs(x="Frecency Quintile", y="Average Response Rate") +
  geom_bar(stat="identity", width=0.5)
bar_avg_resp_rate_fre
```



Q6

Q: Create a bar chart showing the response rate to this deal by monetary quintile.
A:

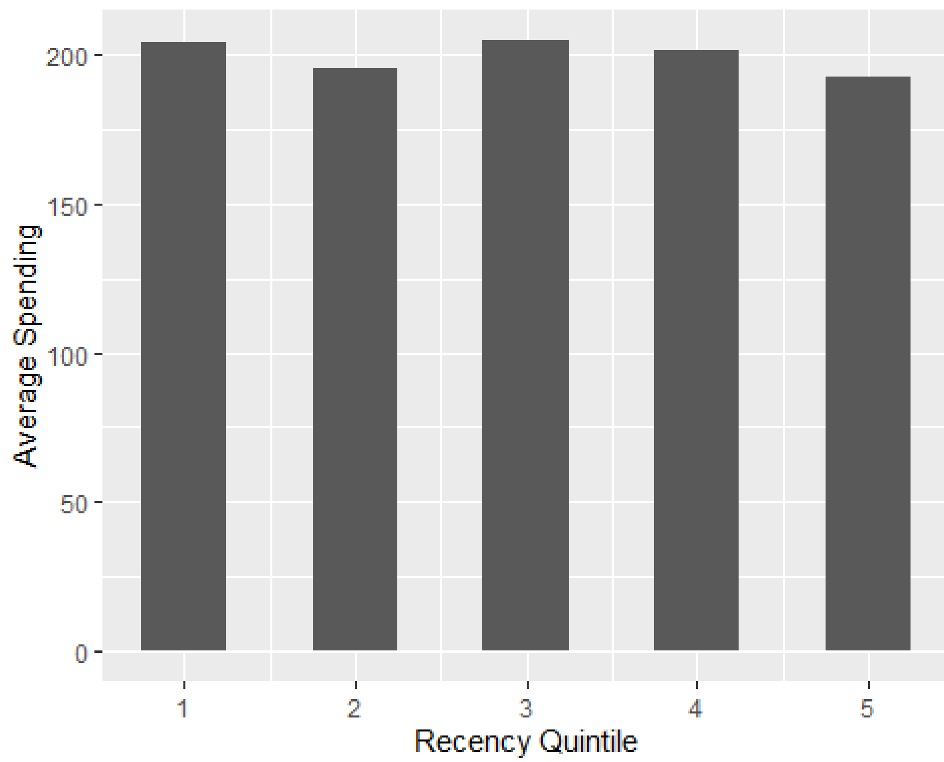
```
avg_resp_rate_mon <- tuango %>% group_by(mon_quin) %>%
  summarise(avg_resp_rate_mon = mean(buyer))
bar_avg_resp_rate_mon <-
  ggplot(data=avg_resp_rate_mon,
         aes(x = mon_quin, y = avg_resp_rate_mon)) +
  labs(x="Monetary Quintile", y="Average Response Rate") +
  geom_bar(stat="identity", width=0.5)
bar_avg_resp_rate_mon
```



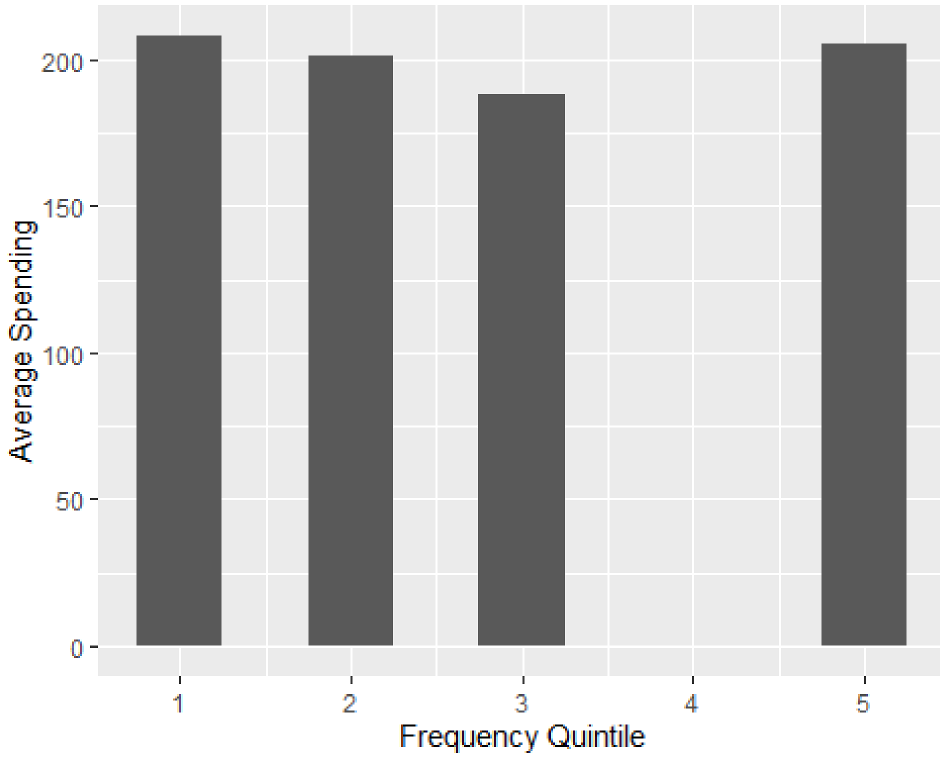
Q7

Q: Repeat questions 4-6 using only those customers who placed an order after the push message, i.e. create bar charts showing the average spending (in RMB) spent by recency, frequency and monetary quintile.
A:

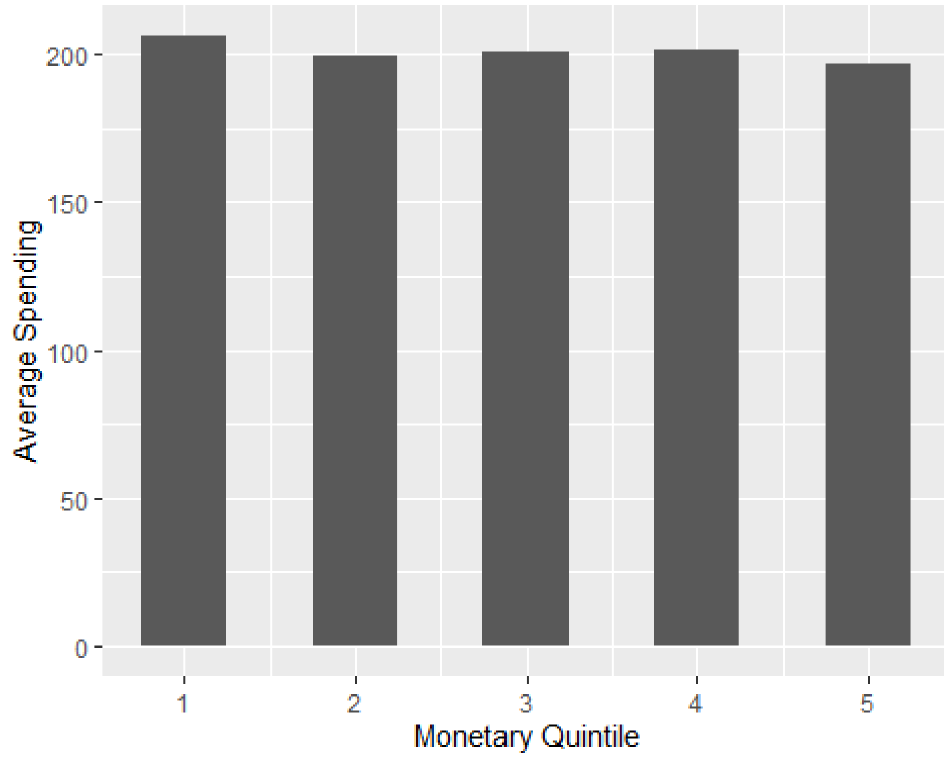
```
avg_spending_rec <- tuango[which(buyer == 1),] %>% group_by(rec_quin) %>%
  summarise(avg_spending_rec = mean(ordersize))
avg_spending_rec <-
  ggplot(data=avg_spending_rec,
         aes(x = rec_quin, y = avg_spending_rec)) +
  labs(x="Recency Quintile", y="Average Spending") +
  geom_bar(stat="identity", width=0.5)
avg_spending_rec
```




```
avg_spending_fre <- tuango[which(buyer == 1),] %>% group_by(fre_quin) %>%
  summarise(avg_spending_fre = mean(ordersize))
avg_spending_fre <-
  ggplot(data=avg_spending_fre,
         aes(x = fre_quin, y = avg_spending_fre)) +
  labs(x="Frequency Quintile", y="Average Spending") +
  geom_bar(stat="identity", width=0.5)
avg_spending_fre
```




```
avg_spending_mon <- tuango[which(buyer == 1),] %>% group_by(mon_quin) %>%
  summarise(avg_spending_mon = mean(ordersize))
avg_spending_mon <-
  ggplot(data=avg_spending_mon,
         aes(x = mon_quin, y = avg_spending_mon)) +
  labs(x="Monetary Quintile", y="Average Spending") +
  geom_bar(stat="identity", width=0.5)
avg_spending_mon
```



Q8

Q: What do the above bar charts reveal about the likelihood of response and the size of the order across the different recency, frequency, and monetary quintiles?
A: Customers who bought more frequently and recently as well as spent more are more likely to response to the deals. But they may not spend more on the Karaoke deal.

Q9

Q: What is the breakeven response rate?

```
break_even = 1.6/(mean(ordersize[which(buyer == 1)])*0.5)
break_even

## [1] 0.01581368
```

A: The breakeven response rate is 1.58%.

Q10

Q: What is the projected
(a) profit in RMB (b) return on marketing expenditures
if you offer the deal to all remaining 264,841 customers.

```
profit1 = sum(ordersize)*264841/13939*0.5-1.6*264841
ROM1 = profit1/(1.6*264841)
profit1

## [1] 406725.4

ROM1

## [1] 0.9598339
```

A: The projected profit is 406,725.4RMB, the return on marketing expenditures is 96.0%.

Q11

Q: Determine which RFM cells (using the sequential n-tiles approach) have response rates exceeding the breakeven rate (no need to report them).
Determine the number of customers belonging to these profitable cells.
Determine the number of buyers belonging to these profitable cells.
What is the projected
(a) profit in RMB (b) return on marketing expenditures

```
tuango <- tuango %>%
  group_by(rfm1) %>%
  mutate(resp_rate_rfm_sq = mean(buyer))
tuango <- tuango %>%
  mutate(mailto_sq = resp_rate_rfm_sq > break_even)

CrossTable(tuango$buyer, tuango$mailto_sq, prop.r = FALSE, prop.t = FALSE,
prop.chisq = FALSE)

##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Col Total |
## |-------------------------|
##
##
## Total Observations in Table:  13939
##
##
##              | tuango$mailto_sq
## tuango$buyer |     FALSE |      TRUE | Row Total |
## -------------|-----------|-----------|-----------|
##            0 |      5904 |      7603 |     13507 |
##              |     0.994 |     0.950 |           |
## -------------|-----------|-----------|-----------|
##            1 |        34 |       398 |       432 |
##              |     0.006 |     0.050 |           |
## -------------|-----------|-----------|-----------|
```

```
## Column Total |      5938 |      8001 |     13939 |
##              |     0.426 |     0.574 |           |
## -------------|-----------|-----------|-----------|
##
##
```

```
m_cost = 264841*mean(tuango$mailto_sq)*1.6
revenue = sum(tuango$ordersize[which(tuango$buyer == 1 & tuango$mailto_sq ==
TRUE)])
profit2 = revenue*0.5*264841/13939 - m_cost
ROM2 = profit2/m_cost
profit2
```

```
## [1] 523818.6
```

```
ROM2
```

```
## [1] 2.15359
```

A: The projected profit is 523,818.6RMB, the return on marketing expenditures is 215.4%.

Q12

Q: Examine the first 20 or so observations in the database. What do you notice about the rfm1 and rfm2 values? That is – do the two approaches generally yield the same RFM index for any given customer? What do you see as the pros and cons of the two approaches (from a statistical as well as logical perspective) and why?
A: The recency and frequency index are the same for these two approaches. The monetary index may be a little bit different.

In the independent approach, the customers are ranked and grouped based on three indepedent factors. While in sequential approach, the recency and monetary rank of customers are based on their frequency rank. So their recency and monetary rank are relative ranks. We cannot compare the recency and monetary rank of customers who have different frequency ranks. This is the cons of sequential approach. On the contrary, in the independent approach, we can compare the three ranks directly.

On the other hand, using independent approach may result in uneven distribution. When the sample size is small and the data are highly skewed, there may be empty groups.