Kellogg School of Management
Northwestern University

Tuango: RFM Analysis for Mobile App Push Messaging

Tuango is one of the major “deal-of-the-day” websites in China. The website’s business model is similar to that of Groupon, promoting discounted gift certificates that can be used at local or national retailers. The pronunciation of “Tuango” in Chinese sounds similar to “group buying,” which refers to the fact that customers are buying as a big group for each “deal.”

Susan Liu had been working as a data analyst in Tuango’s Internet Marketing group for over three years and had recently been appointed as Chief Data Scientist to the newly founded Mobile Marketing group. As Chief Data Scientist, Liu managed a small but highly competent analytics team. The Mobile Marketing group managed Tuango’s marketing campaigns on mobile apps installed on customers’ Android and iOS devices. A conservative estimate by Tuango in 2013 was that the company had about 14 million active mobile customers. More importantly, the smartphone user base in China surpassed 500 million in 2013, which meant that there was a great potential for growth in Tuango’s business.

Tuango had been experimenting with promotional push message campaigns through mobile apps for several months. These campaigns had followed a common pattern. It always started with a deal that the company wanted to offer. Tuango then selected customers who had expressed an interest in the product category where the deal fell, either when they signed up with Tuango, or because they had already purchased a deal in the same category during the last 12 months. Finally, if the deal was tied to a physical store, Tuango made sure only to target customers that lived sufficiently close to the promoted store. Once customers were identified, the offer was pushed out using Tuango’s app on customers’ mobile devices.

When Liu became Chief Data Scientist she decided to reevaluate how mobile campaigns were executed. In particular, she had been bothered for quite some time by the view in the company that the cost of pushing deals onto consumers’ phones was essentially zero. However, Liu knew that the true marginal cost of each message was much higher. If consumers received too many deal offers that were not relevant to them, customer could block future messages in the app, thereby preventing Tuango from contacting them at all.

Liu’s first task for her analytics team was to determine the true marginal cost of sending a push message. The team needed two key metrics to determine marginal cost. First, what was the loss in lifetime value associated with a customer blocking deal messages

---

Professors Song Yao and Florian Zettelmeyer prepared this case to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. The names and the data used in this case have been disguised to assure confidentiality and some events are fictionalized. The case is partially based on the Tuscan RFM case by Professor Charlotte Mason.

from Tuango? Second, by how much did an incremental pushed deal increase the probability that a customer would block future deals from Tuango?

Getting at the first metric was easy. Tuango currently had two types of customers. Many consumers used the mobile app but there were also many web-only customers. Liu decided to approximate the loss in value of a customer who refused deals on their mobile app by assuming that they would subsequently behave like web-only customers.

Getting at the second metric was harder. Luckily, the analytics team found that there was a lot of variation in the number of messages customers had received from Tuango in the past. Since this variation seemed to be largely random, the team could approximate the probability of blocking deal messages using the average fraction of customers who blocked deals on their mobile devices across groups of customers who received fewer or more deals from Tuango.

By multiplying the change in customer value from blocking deal messages with the probability of blocking deal messages, Liu's team determined that 1.6RMB was a good approximation for the true marginal cost of sending an additional deal. Liu knew that Tuango should send deal offers only to those customers for whom Tuango's expected value of the deal exceeded this marginal cost, i.e. 1.6RMB.

Liu did not have much first-hand experience in mobile marketing. However, for years she had been applying a variety of targeting techniques in direct marketing campaigns, including for direct mail, emails, banner ads, online video ads, and so on. She believed that some of the techniques she had used so far should be applicable for mobile marketing as well. To test this, she planned to try some techniques for deal targeting. A good place to start would be RFM analysis, a simple and very popular targeting technique.

Marketers had a long and successful history of using RFM in database marketing to target customers. RFM stands for Recency, Frequency, and Monetary. The fundamental premise underlying RFM analysis is that those customers of yours who have purchased more recently, have made more purchases, and have made larger purchases are more likely to respond to your offering than other customers who have purchased less recently, less often, and in smaller amounts. Although originally developed before the Internet, RFM and its derivatives are still popular in the digital age, particularly because they are very easy to implement even with the huge customer databases that many Internet companies typically have.

RFM Classification

Typically firms group consumers into quintiles – or five groups – when performing a RFM analysis. This makes the number of possible RFM groups or segments at most 125 (5 x 5 x 5).

There are at three typical variations of how RFM segments are determined:

- 'Independent n-tiles' Approach

  This approach computes quintiles independently for recency, frequency, and monetary. That is, the entire customer list is sorted based on recency and divided into recency quintiles. Then, the entire customer list is resorted based on frequency and divided into frequency quintiles. Finally, the entire customer list is sorted one more time based on monetary and divided into monetary quintiles. Then these three quintile variables are combined to form the RFM index. For example, a customer in the 1st recency quintile, 2nd frequency quintile, and 4th monetary quintile – is assigned an RFM index of 124.

- ‘Sequential n-tiles’ Approach

  This approach computes quintiles first for recency. Then, *within* each of the five quintiles for recency, quintiles for frequency are computed – resulting in a total of 25 recency/frequency combinations. Finally, *within* each of these 25 groups, quintiles for monetary are computed.

- ‘Intuitive Groupings’ Approach

  This approach uses intuitive splits rather than quintiles to form the groups. For example, customers may be grouped into recency groups determined by: (1) purchase in last 6 months, (2) last purchase 6-12 months ago, (3) last purchase 12-24 months ago, (4) last purchase 24-36 months ago, and (5) last purchase more than 36 months ago. This approach is ‘intuitive’ in that it is easy to know what is meant if a customer is in the second recency group. But it relies on the analyst’s judgment to know where to ‘draw the lines.’

Testing the performance of RFM for mobile deal offers

Liu decided to test the performance of RFM on a deal for a 1, 2, or 3-hour Karaoke session at one of Hangzhou’s leading Karaoke chains. The deals were priced at 129, 209, and 259RMB, respectively. Tuango’s commission was 50% of the deal price when a deal was sold to consumers (similar to Groupon in the US).

To guide her team on how to perform a RFM analysis for targeting deals to Tuango’s mobile customers, Liu wrote down the key steps:

1. Select all mobile customers who have expressed interest in the category into which the deal falls. This meant 278,780 customers in Hangzhou.
2. Categorize the customers into RFM cells.
3. Randomly select a 5% sample of mobile consumers and offer the deal. This meant that at total of 13,939 received the Karaoke deal. This would be the data used for all subsequent analysis.
4. Track the response (% who respond, order sizes) by RFM cell.
5. Assess response across recency, frequency and monetary quintiles to get a feel for the data before performing a RFM analysis.
6. Assess response as well as profitability and return on marketing expenditure by RFM cell.
7. Use results of the test to determine whether RFM analysis is effective in improving profits and return on marketing expenditures in mobile deal targeting campaigns.
8. If so, use the profitable RFM cells to target the remaining 264,841 customers.

The data

After the 5% random sample had been offered the deal, Liu's analytics team received a dataset with the results. The dataset contained all information needed for the RFM analysis (see Exhibit 1 for a summary and definitions of key variables found in the dataset *Tuango_RFM.dta*).

Exhibit 1
Variable Names and Descriptions
(Tuango_RFM.dta dataset)

Name	**Description**
userid	Unique user ID

Values of Recency, Frequency, and Monetary variables

recency	Days since last purchase of a deal (before the Karaoke deal offer)
frequency	Number of deals purchased during the one year period before the Karaoke deal offer
monetary	Average spending per order (in RMB) during the one year period before the Karaoke deal offer

RFM indices:

rfm1	Sequential RFM indices
rfm2	Independent RFM indices

Response to the customized push message

buyer	Bought one of the three Karaoke deals? 1=yes/0=no
ordersize	The spending on the Karaoke deal (in RMB)

Other observed variables in the dataset

platform	The platform used when consumer registered with Tuango
category	The category of last purchase before the current targeting
mobile_os	The mobile OS of the customer

Assignment

Everything you need to know to prepare this report is contained in the lecture notes from the RFM lecture, the file "RFM_BBB_stata.do" which goes through the calculations for the Bookbinders RFM analysis, and the "Stata Tutorial."

Part I: Preliminary and Quintile Analysis (Q1-Q8 36 points)

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

2. Of those who bought, what was the average spending?
   (Hint: constrain the *summarize* command with the *if* command, see the "Stata Tutorial")

(For Questions 3-8, use the independent n-tiles approach.)

3. Create *quintile* variables for recency, frequency and monetary.
   (Hint: review the file "RFM_BBB_stata.do" which goes through the calculations for the Bookbinders RFM analysis)

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

5. Create a bar chart showing the *response rate* to this deal by frequency quintile.

6. Create a bar chart showing the *response rate* to this deal by monetary quintile.

7. 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.
   (Hint: constrain the *graph bar* command with the *if* command, see the "Stata Tutorial")

8. 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?

Part II: Profitability Analysis

The RFM indices using both independent and sequential n-tiles approaches **are already in the database**. (For those interested in re-computing them, the specific steps for how these values can be computed for the Bookbinders case are in the file "RFM_BBB_stata.do")

The following questions will ask you to use your data to forecast the profit and the return on marketing expenditures of offering the deal to the remaining 264,841 potential customers (i.e. 278,780-13,939).

To calculate profit and return on marketing expenditures assume the following:

> Marginal cost to offer a deal is = 1.6 RMB
> Commission on each deal sold = 50% (of sales revenues)

9. (10 points) What is the breakeven response rate?

10. (12 points) What is the projected
    (a) profit in RMB
    (b) return on marketing expenditures

    if you offer the deal to ***all*** remaining 264,841 customers.

11. (12 points) Consider offering the deal only to those of the 264,841 customers in RFM cells (using the sequential n-tiles approach, coded in the *rfm1* variable) with a response rate that is equal to or greater than the breakeven response rate. Specifically, follow these steps (using Stata and a calculator):

    - 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

12. (15 points) 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?

13. (15 points) So far, you have calculated expected profits or return on marketing expenditures for the Karaoke deal when you offer the Karaoke deal to the remaining 264,841 customers. However, if you consider all orders (including the Karaoke orders, but not just the Karaoke orders) that these customers can make on Tuango, can you say that the profit generated from all the orders always (100%) increases when you offer the Karaoke deal to these customers? How can you measure the impact of the Karaoke deal on the overall profit across all orders? You do not need to calculate anything to answer these questions.