Tuango – Targeting Mobile App Messages

Tuango is a “deal-of-the-day” website in China. Their business model is similar to 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”.[1]

Susan Liu had been working as a data analyst in Tuango’s Internet Marketing group for three years and had recently been appointed as Chief Data Scientist to the 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. In 2020 Tuango had about 20 million active mobile customers. More importantly, the smartphone user base in China was projected to surpass 1.3 million in 2030, which suggested potential for growth in Tuango’s business.

Tuango had been experimenting with promotional campaigns through mobile apps for several months. These campaigns 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 the deal was aimed at, either when they signed up with Tuango, or because they had already purchased a deal in the same category during the past 12 months. Finally, if the deal was tied to a physical store, Tuango made sure to only target customers that lived

[1] Professor Blake McShane 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 a case created by Professors Song Yao and Florian Zettelmeyer.

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 was bothered by the view in the company that the cost of pushing offer messages to customers' phones was essentially zero. However, Liu knew that the true marginal cost of each message was much higher. If customers received too many messages for deals that were not relevant to them, customers could block future messages in the app, thereby closing an important marketing channel.

Liu's first task for her analytics team was to determine the true marginal cost of sending an offer message in the app. The team needed two key metrics to determine marginal cost. First, what was the loss in customer lifetime value associated with a customer blocking deal messages from Tuango? Second, by how much did an incremental message increase the probability that a customer would block future messages from Tuango?

Getting at the first metric was easy. Tuango currently had two types of customers. Many customers 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 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 9 RMB was a good approximation for the true marginal cost of sending an additional deal message. Liu knew that Tuango should send deal messages only to those customers for whom Tuango's expected profit exceeded this marginal cost.

Liu did not have much first-hand experience in mobile marketing. However, for years she had been applying a variety of targeting methods for email, banner ads, online video ads, and so on. She believed that some of these techniques should be applicable to mobile marketing and deal targeting as well.

Testing the performance of targeting for mobile deal offers

To guide her team on how to use models to target deals to Tuango's mobile customers, Liu wrote down the key steps:

1. Select all mobile customers that expressed interest in the karaoke category (i.e., 418,160 customers in Hangzhou)

2. Randomly sample 5% of these customers (i.e., 20,908) and offer all of them the karaoke deal
3. Track if a customer responds (i.e., purchases) and determine the order sizes (i.e., number of 30-minute sessions purchased). This would generate the data used in analyses
4. Build targeting models that predict customer response and order size
5. Use these models to determine if targeting can improve the profitability of the karaoke mobile deal campaign
6. If so, target only the profitable customers among the remaining 397,252 customers

The data

After the sample of customers had been offered the deal, Liu's analytics team received a dataset with the results. The dataset contains the information needed for the analysis.

Exhibit 1: Variable Names and Descriptions

- userid {string}: Unique user ID

Response to the karaoke deal offer message

- buyer {category}: Did the customer buy the karaoke deal? ("yes" or "no")
- ordersize {integer} Number of 30-minute karaoke sessions the customer purchased

Recency, Frequency, and Monetary variables (RFM)

- recency {integer}: Days since last purchase (excluding the karaoke deal offer)
- frequency {integer}: Number of deals purchased during the one-year period prior to the karaoke deal offer
- monetary {float}: Average amount spent per order (in RMB) during the one-year period prior to the karaoke deal offer

Other variables in the dataset

- age {integer}: Age of the customer
- gender {category}: Gender identity of the customer (M = male, F = female, O = Other or unknown)
- music {category}: Purchased a deal in the music category during the one-year period before the karaoke deal offer ("yes" or "no")
- test {integer}: Splits the dataset into test (1) and rollout (0) data
- training {integer}: Splits the dataset into training (1) and test (0) data. This variable will not be used in this assignment

Assignment

To help you complete this assignment, review the applied-logistics regression document and the Python code for the Bookbinders test-and-rollout analysis. You can get the files for this worked example in .zip format here.

To run the analysis run the commands below in a terminal set to the project directory for the tuango-case (e.g,. ~/git/mgta455-tuango)

```
uv init
uv venv
source .venv/bin/activate
uv add pyrsm==2.2.0
```

Then make sure to select the jupyter kernel for `mgta455-tuango .venv/bin/python`

Data for the Tuango case are included in the GitHub repo https://github.com/orgs/rsm-msba-25-26/repositories. For questions 1-15 use `data/tuango_pre.parquet`. For question 16 use the `data/tuango_post.parquet` dataset. The dataset descriptions are in `data/tuango_pre_description.md` and data/`tuango_post_description.md`.

IMPORTANT: Just showing results without discussion and explanation is NOT sufficient. Explain what you did and why and what the results tell us.

Part I: Preliminary Analysis (4 points)

Use the "tuango.ipynb" notebook in the repo you cloned from GitHub to answer all questions.

1. What proportion of customers responded to the deal offer message (i.e., bought the deal)? (2 points)
2. What was the average number of karaoke sessions purchased by customers that bought one or more 30-minute sessions? Use the "ordersize" variable for your calculation (2 points)

Part II: Build Targeting Models (21 points)

3. Estimate a logistic regression model using "buyer" as the response variable (aka target or dependent variable) and, recency, frequency, monetary, age, gender, and music as the explanatory variables (aka features or independent variables). Use either pyrsm or statsmodels (formula interface). Do **not** use `sklearn` for logistic regression in this class (2 points)

4. Create Partials-Dependence-Plots (PDP) for all explanatory variables using `pyrsm`. Describe the effect of a change in each explanatory variable on the probability that a customer will purchase the karaoke deal (3 points)
5. Use `pyrsm` to generate a Permutation-Importance-Plot (PIP) to assess variable importance. Which variables seem to be most and least important in the model? Explain how Permutation Importance allows you to evaluate variable importance (3 points)
6. Add the predicted values from the logistic regression to the "tuango" DataFrame. Use `pred_logit` as the variable name. Compare the average of the predicted values in `pred_logit` to the overall response rate (i.e., proportion of buyers). What do you notice when you do this calculation on only the data used to estimate the model? Explain. (4 points)
7. Estimate a linear regression model using "ordersize" as the response variable and recency, frequency, monetary, age, gender, and music as the explanatory variables. Estimate this regression using only those customers who placed an order after the deal offer message. Describe why you think it does, or does not, make sense to focus on this group of customers. Use either `pyrsm` or `statsmodels` (formula interface). Do not use `sklearn` for linear regression in this class (3 points)
8. Use `pyrsm` to create a Permutation Importance Plot and to assess variable importance in the estimation data of the linear regression model. Which variables seem to be most important in the model? (2 points)
9. What do the linear regression model results suggest about our ability to predict ordersize for customers who responded to the deal offer message? (2 points)
10. Add the predicted values from the linear regression to the "tuango" DataFrame. Compare the average of the predicted values to the average value of ordersize. Make sure to focus only on buyers. What do you notice? (2 points)

Part III: Profitability Analysis

The following questions focus on the profit and return on marketing expenditures from offering the deal to (some of) the remaining 397,252 potential customers in Hangzhou (i.e., 418,160 – 20,908).

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

- Price per 30-minute session is 49 RMB
- Marginal cost of sending a deal offer message is 9 RMB
- Tuango's fee on each deal sold is 50% of sales revenues

11. What is the breakeven response rate? Use the average ordersize from question 2 in your revenue calculations. (2 points)

12. What is the projected profit in RMB and the return on marketing expenditures if you offer the deal to all 397,252 remaining customers (i.e., target everyone)? (6 points)

13. Evaluate the performance implications of offering the deal to only those customers (out of 397,252) with a predicted probability of purchase greater than the breakeven response rate. Determine the projected profit in RMB and the return on marketing expenditures and compare the results to the values you calculated in question 12. (6 points)

    Note: Fine tune your estimate from Q2 above by determining the average amount spent among the people that (1) will receive a message and (2) bought a karaoke deal. Also, use the actual number of messages you plan to send out to the group of customers in the rollout sample (i.e., "test == 0")

14. Create a bar chart with profit information for the analyses conducted in questions 12 and 13 (2 points)

15. Create a bar chart with ROME numbers for the analyses conducted in questions 12 and 13 (2 points)

16. You also have access to a dataset with the results from the deal offer roll-out (tuango_post.parquet). Tuango decided to contact all remaining 397,252 customers because this would provide data that could be used to evaluate different targeting approaches. The data has a "test" variable (`test = 1` for the data used in the test, `test = 0` for the remaining customers). Use this variable to evaluate the actual performance on the 'roll out' sample for the targeting approaches from questions 12 and 13. Also re-create the plots from questions 14 and 15 based on this new dataset.

Specifically, redo questions 12-15 using the `tuango-post.parquet` data and make the required code adjustments to calculate the actual performance on the 'roll out' sample correctly.

- Hint 1: It is important that you do NOT use any information about buyers that were in the 'roll out' sample (i.e., test == 0) when calculating the break-even response rate etc. for targeting.
- Hint 2: You have the actual data on what happened in the "post" data. Use that information to calculate performance (i.e., do not "project" the performance like you had to do for questions 12 and 13).