Tuango - RFM Analysis for Mobile App Push Messaging

Professor Song Yao
Olin Business School

Customer Analytics

How many customers bought and how much?

PRELIMINARY CALCULATIONS

```
# Q1. What percentage of customers responded (i.e. bought anything) after the push message?
# Group by 'buyer' and count the occurrences
buyer_summary = Tuango.groupby('buyer').size().reset_index(name='count')
# Calculate the percentage
buyer_summary['percentage'] = buyer_summary['count'] / buyer_summary['count'].sum() * 100

# Print the table with formatting
print(tabulate(buyer_summary, headers='keys', tablefmt='grid',
               floatfmt=".2f", showindex=False))
```

```
+---------+----------+--------------+
|   buyer |    count |   percentage |
+=========+==========+==============+
|    0.00 | 13507.00 |        96.90 |
+---------+----------+--------------+
|    1.00 |   432.00 |         3.10 |
+---------+----------+--------------+
```

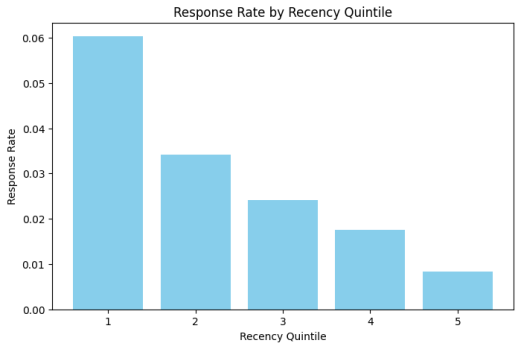
```
# 2.    Of those who bought, what was the average spending?
# Filter using 'buyer==1' and calculate the average of 'ordersize'
average_spending = Tuango[Tuango['buyer'] == 1]['ordersize'].mean()
print(f"Average spending of buyers: {average_spending:.2f}")
```

```
Average spending of buyers: 202.36
```

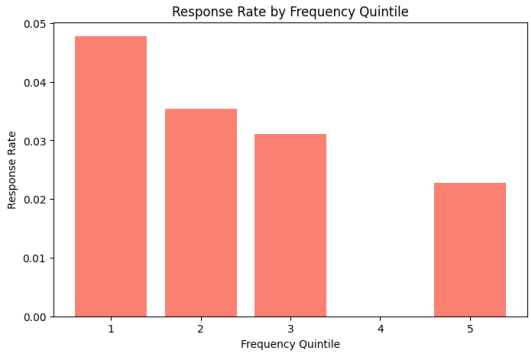
What does the quintile analysis reveal about the premise of RFM?

RESPONSE RATES BY RFM

Recency



Frequency

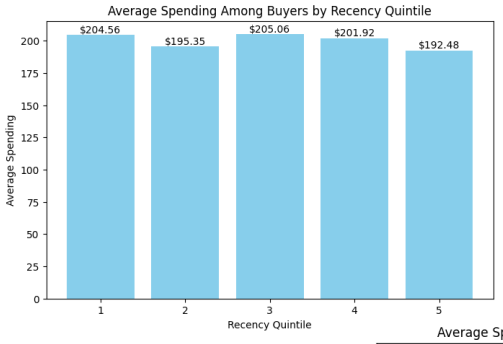


Monetary

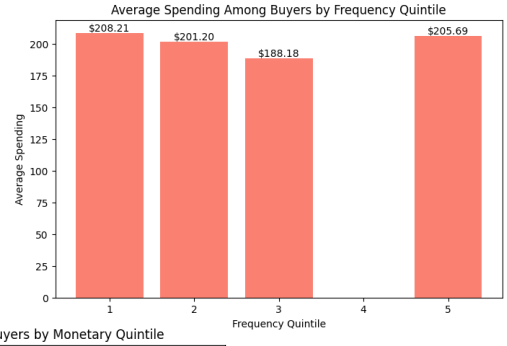
How predictive are RFM deciles for the average $ order from purchasing customers?

ORDER SIZE BY RFM QUINTILES

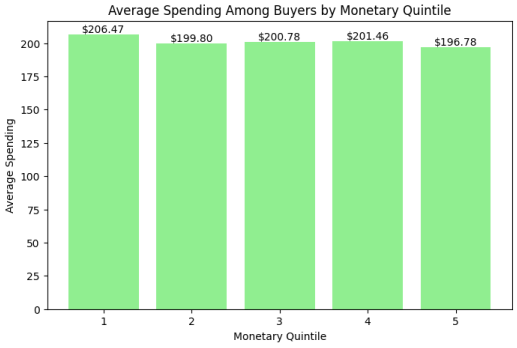
Recency



Frequency



Monetary



As a benchmark we calculate the profitability of messaging the full 264,841 customers

PROFITABILITY (FULL SAMPLE)

- Push messages to 264,841 (278,780-13,939)
- Average response rate

buyer	count	percentage
0.00	13507.00	96.90
1.00	432.00	3.10

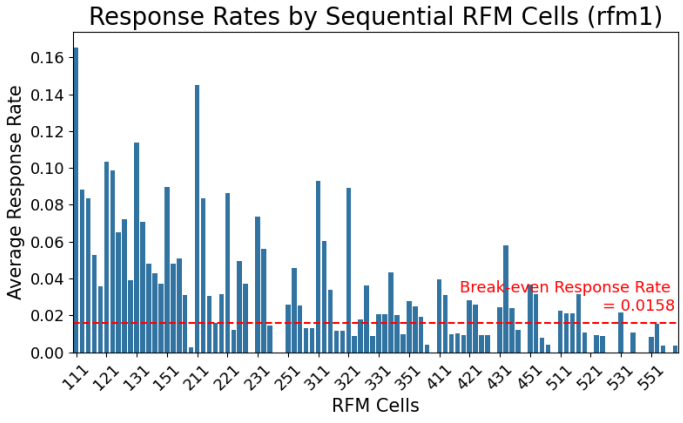
Aver. response rate: 3.10%
Expected number of buyers:
3.10% * 264,841 **= 8,210**

- Profit = ($202.36)(.5) * **8,210** -264,841*($1.6) = **$406,942**
- Return on marketing expenditure = $**406,942**/$423,745.6 = **96%**

The break-even response rate is 1.58%

BREAK EVEN RESPONSE RATE

- Cost of a push message = $1.6
- Average order size = $202.36
- COGS = 50%
- Break-even = Cost to push / profit per sale
  = 1.6/ (202.36*0.5) = **1.58%**



Next, we calculate profitability using a sequential N-Tile RFM approach

PROFITABILITY (RFM-INDEX BASED ON SEQUENTIAL N-TILE)

```
### Percentage of customers being targeted
Tuango %>% group_by(push_rfm1) %>%
  summarise(n = n(), .groups = "drop") %>%
  mutate(Percentage=n/sum(n) * 100) %>%
  kable(digits = 2)
```

push_rfm1	n	Percentage
FALSE	5938	42.6
TRUE	8001	57.4

Message to 57.40% of sample: 264,841 * **57.40%=152,019**

```
### Percentage of targeted customers buying the book
Tuango %>% filter(push_rfm1 == TRUE) %>%
  group_by(buyer) %>%
  summarise(n = n(), .groups = "drop") %>%
  mutate(Percentage=n/sum(n) * 100) %>%
  kable(digits = 2)
```

buyer	n	Percentage
0	7603	95.03
1	398	4.97

Expected response rate: 5% (using 4.97% is fine)
Expected number of buyers: 5% * 152,019 = 7,601

- Profit = ($202.36)(.5)* **7,601** -**152,019** ($1.6) = **$525,839**
- Return on marketing expenditure = $**525,839**/ (152,019 * $1.6) = **216%**

RFM yields large profitability gains

	Mass Messaging	Targeted: Sequential RFM	Targeted: Independent RFM
Number of Customers	264,841	264,841	264,841
Number Messaged	264,841	152,019	161,421
% Messaged	100%	57%	61%
Number of Orders	8,210	7,601	7,716
Response Rate	3.1%	5.00%	4.78%
Revenues	$1,661,375.60	$1,538,138.36	$1,561,409.76
COGS	$830,687.80	$769,069.18	$780,704.88
Messaging Cost	$423,745.60	$243,230.40	$258,273.60
Profit	$406,942.20	$525,838.78	$522,431.28
Return on Mark.	96.0%	216.2%	202.3%

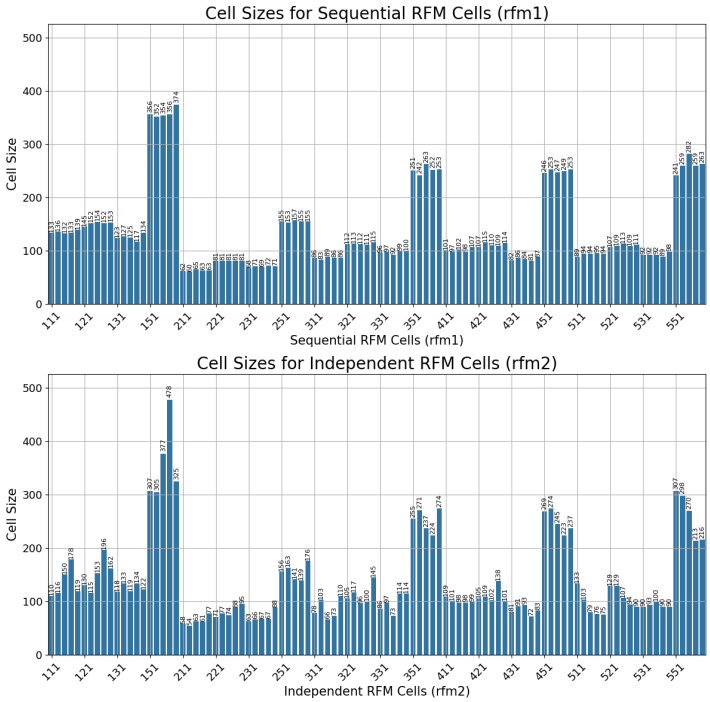
How do N-Tile approaches differ?

COMPARISON BETWEEN INDEPENDENT AND SEQUENTIAL N-TILE APPROACHES

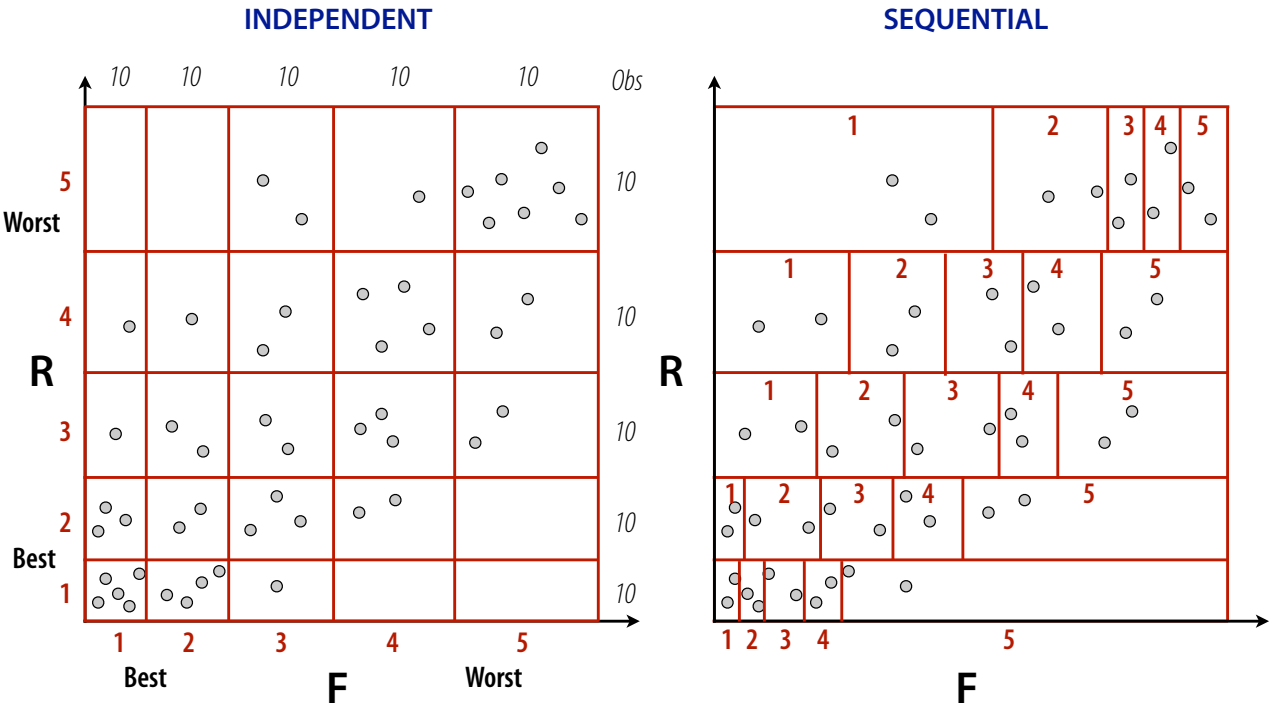
rfm1	rfm2
311	312
253	252
121	122
111	112
523	522
512	511
221	222
554	553
151	152
122	123
553	552
453	452
552	551
252	251
554	553
423	424
454	453
254	255
234	235
123	124

How do cell sizes compare across N-Tile approaches?

DISTRIBUTION OF CELL SIZES BY N-TILE APPROACH



An example with 50 observations can explain the difference in the approaches



How do N-Tile approaches differ?

COMPARISON BETWEEN INDEPENDENT AND SEQUENTIAL N-TILE APPROACHES

rfm1	rfm2
311	312
253	252
121	122
111	112
523	522
512	511
221	222
554	553
151	152
122	123
553	552
453	452
552	551
252	251
554	553
423	424
454	453
254	255
234	235
123	124

- R always same, F & M may differ
- Sequential approach has advantage of looking at more homogeneous groups b/c M value computed within R and F groups
- Hence, being a frequent buyer or "big spender" is only relative (e.g., 55**1** is a big spender in the RF=55 group, but actually spend less than group 11**5**).
  - Only independent approach allows comparisons of F and M across groups
- Sequential approach implicitly weights R > F > M
- From targeting perspective, often little difference in performance (in terms of return or marketing)
- Sequential approach has an advantage for testing campaigns: Since response rates will be computed for each cell it is important to have a sufficient and stable number of customers per cell (sufficient stat power and stable standard errors)