

Customer Segmentation & Churn Prediction

Cheng-Pang (**Tego**) Chang

Master in Interdisciplinary Data Science

Duke University

Duke

Key Takeaways

- Deliver a **dynamic** framework to categorize our merchants
- Proposed models to help **focus** resources to those potential churns

Assignments for Each Merchant

- RFM Framework
- Categorize Considering Product User Growth

Applying RFM Framework

- Treat the first question as a **customer segmentation** problem.
- RFM measures a customer's (merchant's) value in three perspectives:
 - **R** – Recency: how long it has been since **the last time** the merchant has a transaction till **today** (2034/12/31).
 - **Assumption**: the larger the gap, the more likely the merchant could have churned.
 - **F** – Frequency: how many **transaction times** does the merchant have during the 2-year timeframe.
 - **Assumption**: the fewer transactions the merchant has conducted, the more likely it churns.
 - **M** – Monetary: what is the **sum of the transaction amounts** in USD from the merchant throughout the 2 years, 2033 and 2034.
 - **Assumption**: the less the sum of transaction amount is, the more likely it churns

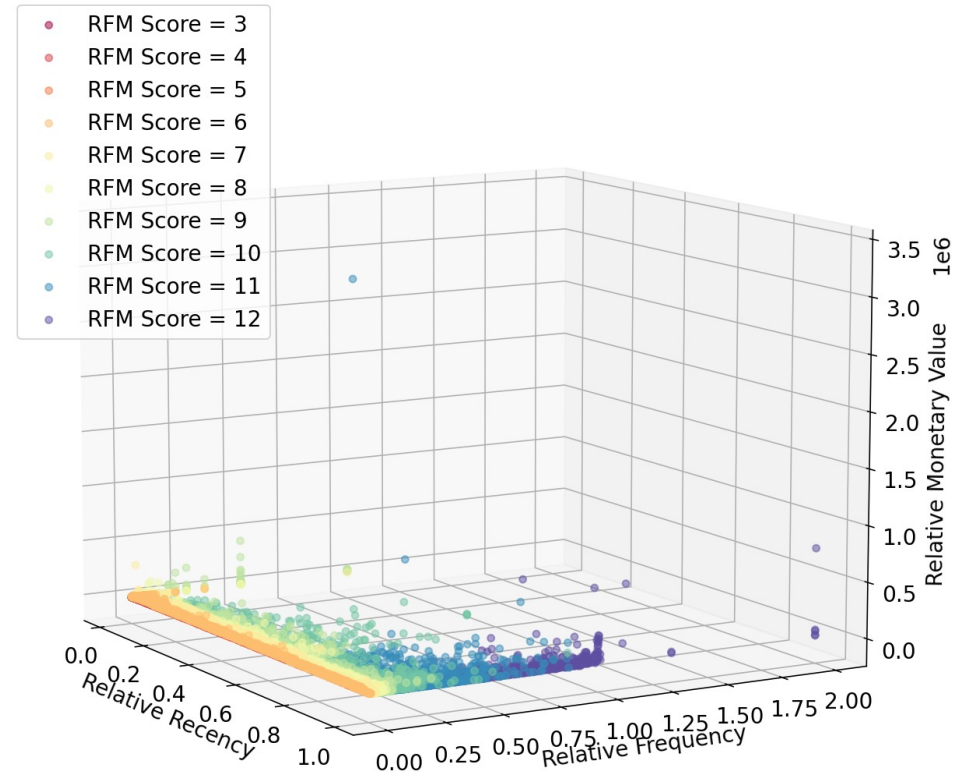
Relative RFM

- How we calculate the 3 metrics in RFM:
 - Relative recency (**R**) = $1 - \text{Recency} / \text{Lifetime of the merchant}$
 - Relative frequency (**F**) = $\text{Frequency} / \text{Lifetime of the merchant}$
 - Relative monetary (**M**) = $\text{Monetary} / \text{Lifetime of the merchant}$
- Why relative?
 - Values above have to be **normalized** to avoid potential bias, e.g.,
 - Comparing with a merchant who had its first transaction **2 week ago** disappearing in our record for 10 days, the merchant who disappears for the same length of time but with the first transaction **1 year ago** is considered a **higher score** merchant.

Note: **Lifetime of a merchant** = Today (2034/12/31) – The day of the first transaction of the merchant

RFM Assignment

- Applied 25%, 50%, 75% percentiles to classify each metric (R, F, M) into 4 leveled-score (1, 2, 3, 4) **Rs, Fs, Ms** corresponding..
- After **summing** the scores for each metric, we will have each merchant being leveled from 3 to 12, totally **10 groups**.



Updated Merchant Assignment

- Applying the concept of **product user growth** to simplify the segmentation from 10 groups to 4 groups.
 - Engagement (F) >= Revenue (M)**

Groups	Conditions	Percent of Merchant	Identification	Follow Up Business Strategy
01. high engagement and high value	$F_s \geq 3$ AND $M_s \geq 3$	39% (5615)	Most valuable customers (MVC)	Barrier in between not obvious: <ul style="list-style-type: none"> Retention approaches Incentive program to increase their transaction amounts. Referral program to attract more merchants
02. high engagement and low value	$F_s \geq 3$ AND $M_s \leq 2$	11% (1552)	Active customers	
03. low engagement but high value or recent usage	$F_s \leq 2$ AND ($M_s \geq 3$ OR $R_s \geq 3$)	20% (2866)	Potential MVC or active customer	Conduct interview to build merchant persona to investigate how to upward to group 01 or 02
04. potential churn	$F_s, M_s, R_s \leq 2$	30% (4318)	Potential churn	Provide large scale of promotions or discounts

Dynamically Adjustable
based on Business Context

Decision boundary in values:
- Relative frequency (R) equals to **0.035**.

- Relative monetary (M) equals to **700**.
- Relative recency (F) equals to **0.67**.

Identifying and Predicting the Churn Merchant

- Defining Churn through RFM
- Models based on budgets

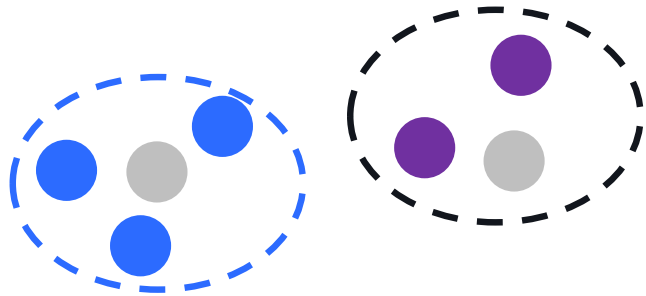
Churn

- Definition of **Churn**
 - Score of 1 in all Rs, Fs, and Ms (< 25% percentile of R, F, and M)
 - Alternative, e.g.,
 - Consecutive no transactions/monetary below a threshold for N months/seasons.
 - Not preferred:
 - As we want you to understand **how the results of model are formed**.
- Identification
 - 11% (1517) merchants are considered to have already churned in the 2-year timeframe.
- Definition of **Loyal**
 - Score of 4 in all Rs, Fs, and Ms (> 75% percentile of R, F, and M)

Locating the Most Likely Churn

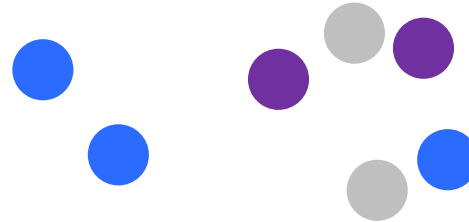
Label-propagation

- Assumption
 - Churn or loyal is a characteristic that can be considered to have a behavior boundary like a group.



Self-training

- Assumption
 - Merchants having the similar transaction patterns might share the same characteristic of being churn or loyal.



Criteria for Identifying the Mostly Likely Churn

Groups	Conditions	Percent of Merchant	Identification	Follow Up Business Strategy
04. potential churn	Fs, Ms, Rs <= 2	30% (4318 = 1517 + 2801)	Potential churn	Provide large scale of promotions or discounts

Criteria	Label Propagation	Criteria	Self Learning
Min. RelRecency	0	Avg. RelRecency	0.14
Min. RelFrequency	0.01	Avg. RelFrequency	0.01
Min. RelMonetary	0.03	Avg. RelMonetary	1.66
Mostly Likely Churned Merchants	42	Dynamic Adjustable based on Business Context	2801

Thank you and Comments
Appreciated!