

Analyzing characteristics of users from MoMo Loyalty Program



Tên: Phạm Quốc Huy

Mentor: Nguyễn Thị Thắm

Lớp/Nhóm: DA18/Nhóm 2

| **TABLE OF CONTENTS**

01 Intro

02 Data

**03 Customer
Segmentation**

04 Scale Data

05 Find K

**06 Modeling
KMeans**

**07 Cluster
Analysis**

08 Conclusison

1. Introduction

1.1. Problem

1.2. My approaches

1. Introduction

1.1. Problem

- Users' retention has always been one of the key targets that **MoMo** is striving to be better. A **Loyalty program** called "**MoMo Hoàn Tiền**" - one of the projects aiming to achieve such goal was launched from January 1st 2021 to March 2022.
- What ideas do you have for **MoMo** in loyalty program development strategy?
- Do you have any advice for the Marketing department in designing promotion campaigns to increase user retention's performance?

1. Introduction

1.2. Approaches

- Using Machine Learning model KMeans to find clusters in by analyzing RFM.
- Find characteristics and true needs of users to offer the right service.
 - **Recency (R)**: How recently a customer has made a transaction
 - **Frequency (F)**: How often a customer makes a transaction
 - **Monetary Value (M)**: Money value of the transaction

2. Data

2.1. Data

2.2. Data Processing

2.3. EDA

2. Data

2.1. Data

Transaction

User_id	Order_id	Date	GMV	Service Group	Merchant_id
123	234	01-17-2020	100000	supermarket	12
...

User_id: each user in MoMo will be given an unique id
Order_id: each transaction will be given an unique id
Date: date on which the transaction takes place
GMV (Gross Merchandise Value): total amount of money that user spends (VND)
Service Group: group services that users spend on
Merchant_id: each merchant will be given a unique id

Merchant

ServiceID	Service Group	Merchant	MerchantID
7eleven	CVS	7eleven	1
MOMOIOQC20191008	CVS	7eleven	1

Service_id: each service group will be given an unique id
Order_id: each transaction will be given an unique id
Merchant: merchant services that users spend on
Merchant_id: each merchant will be given a unique id

2. Data

2.2. Data Processing

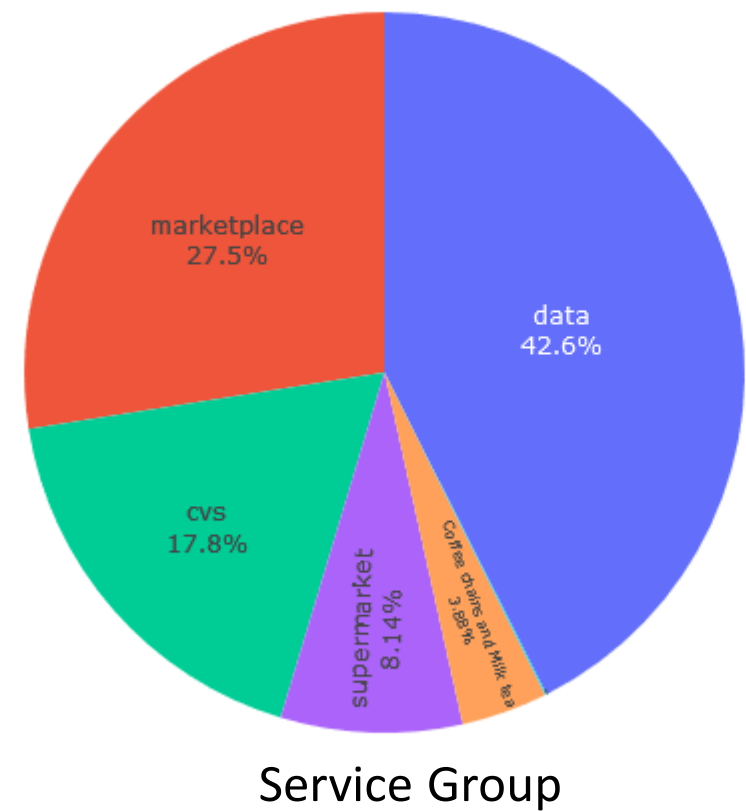
#	Column	Non-null	Count	Dtype
0	DATE	58158	non-null	datetime64
1	Order_id	58158	non-null	int64
2	NEWVERTICAL_Merchant	58158	non-null	object
3	MerchantID	58158	non-null	int64
4	User_id	58158	non-null	int64
5	GMV	58158	non-null	int64
6	Service Group	58158	non-null	object
7	ServiceID	58158	non-null	object
8	Merchant	58158	non-null	object

Steps:

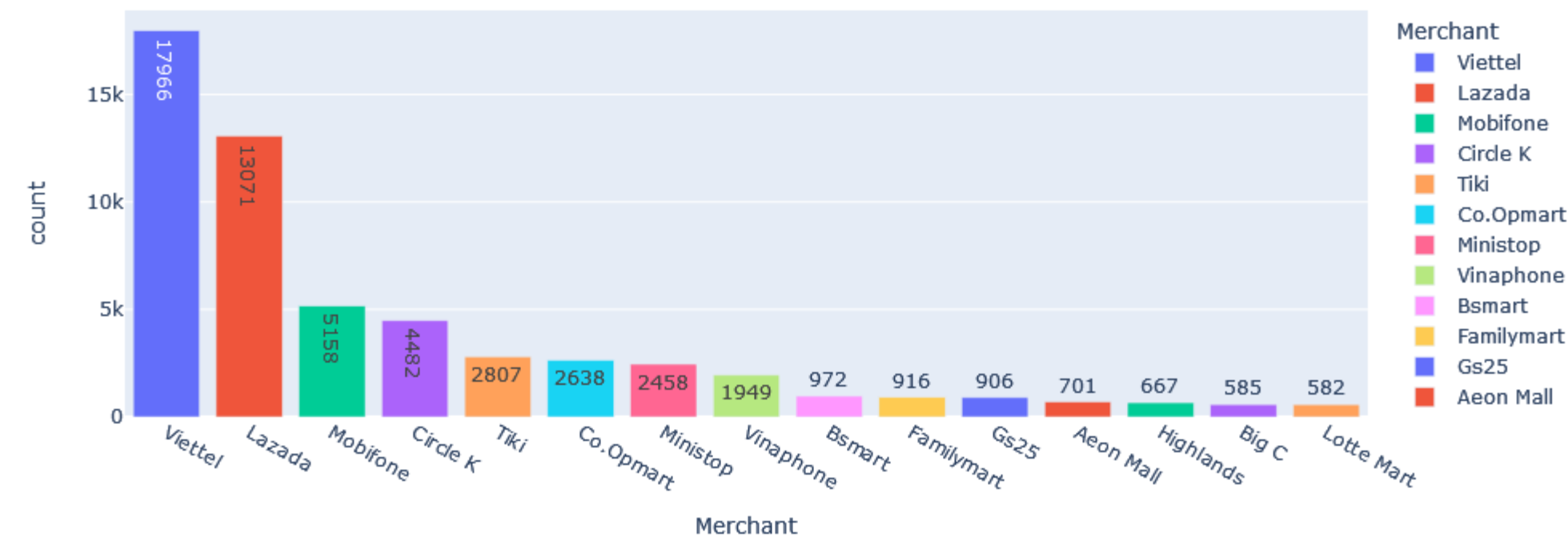
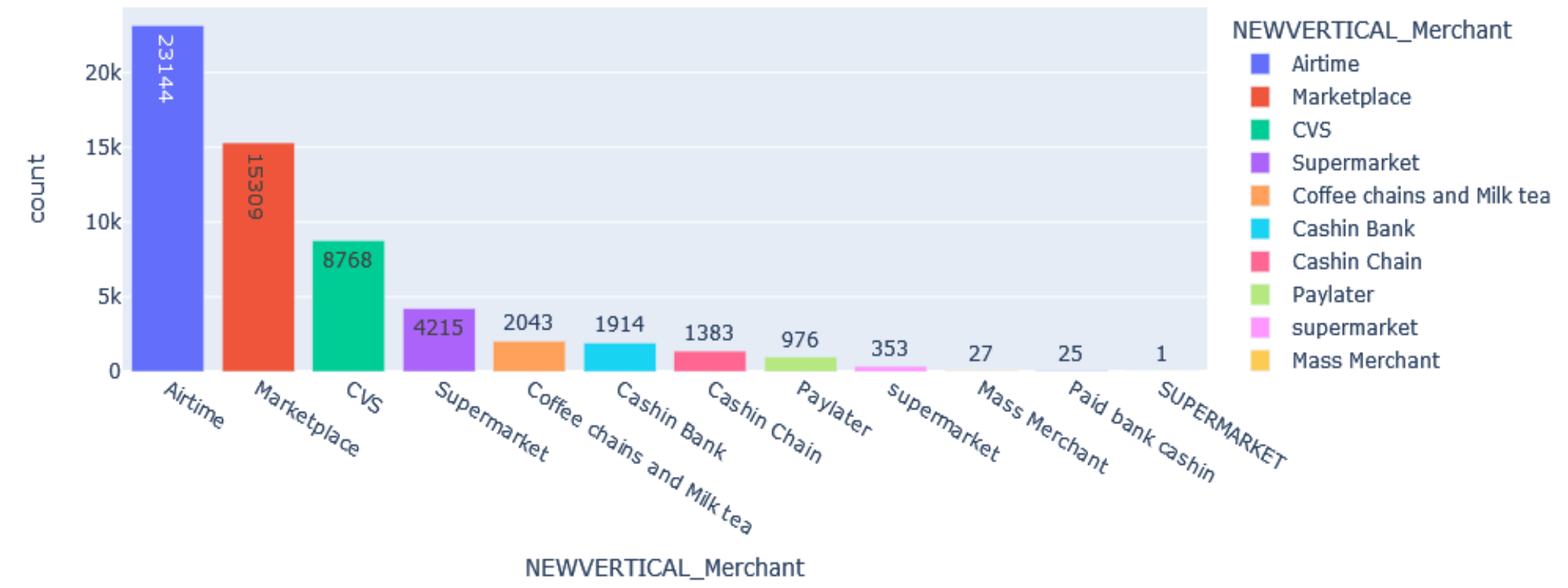
- Deleted nulls
- Dropped duplicates
- Lowercased
- Merged two dataframes

2. Data

2.3. EDA



- data
- marketplace
- cvs
- supermarket
- Coffee chains and Milk tea
- Offline Beverage

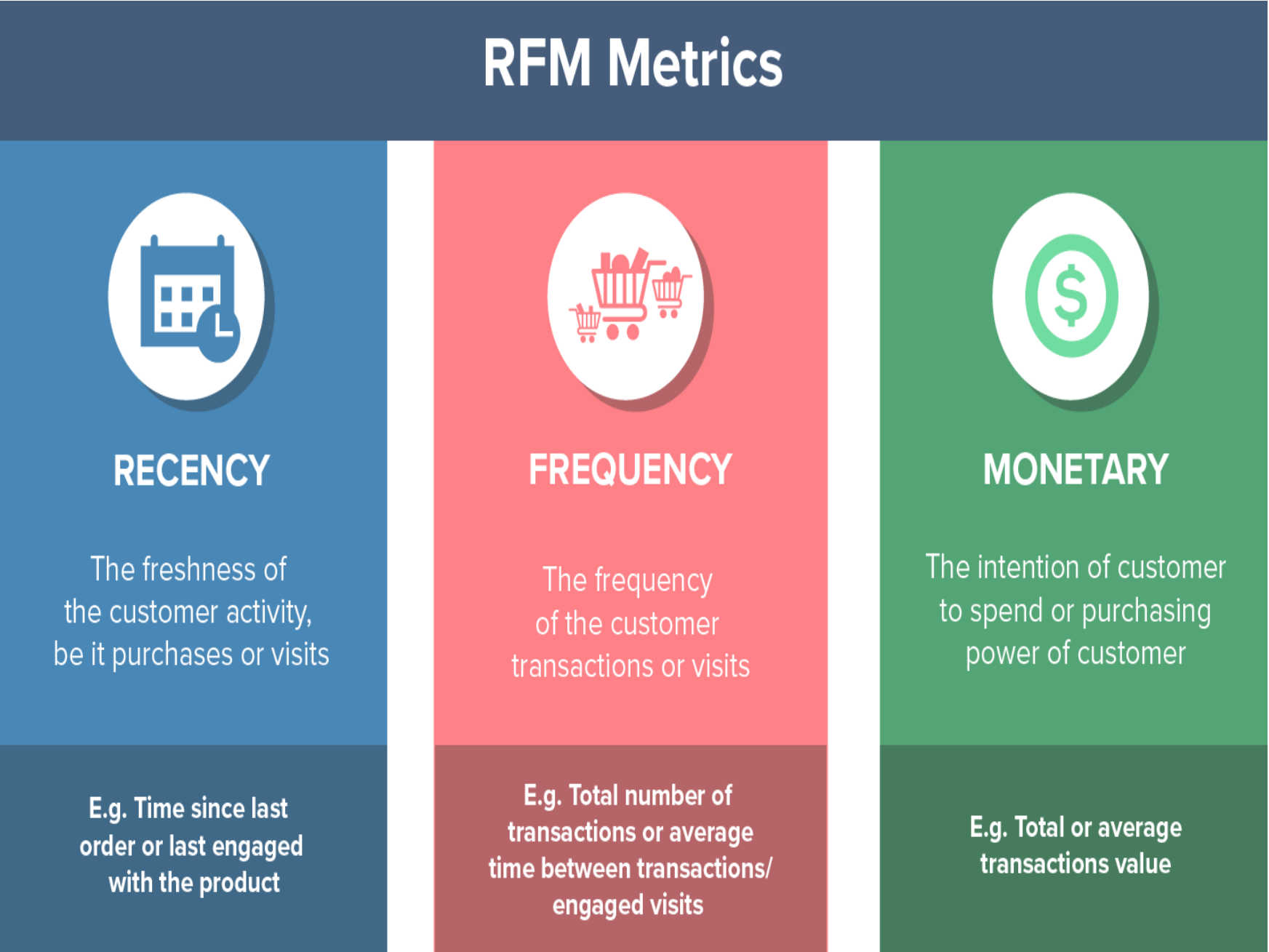


3. Customer Segmentation

3.1. RFM

3. Customer Segmentation

3.1. RFM



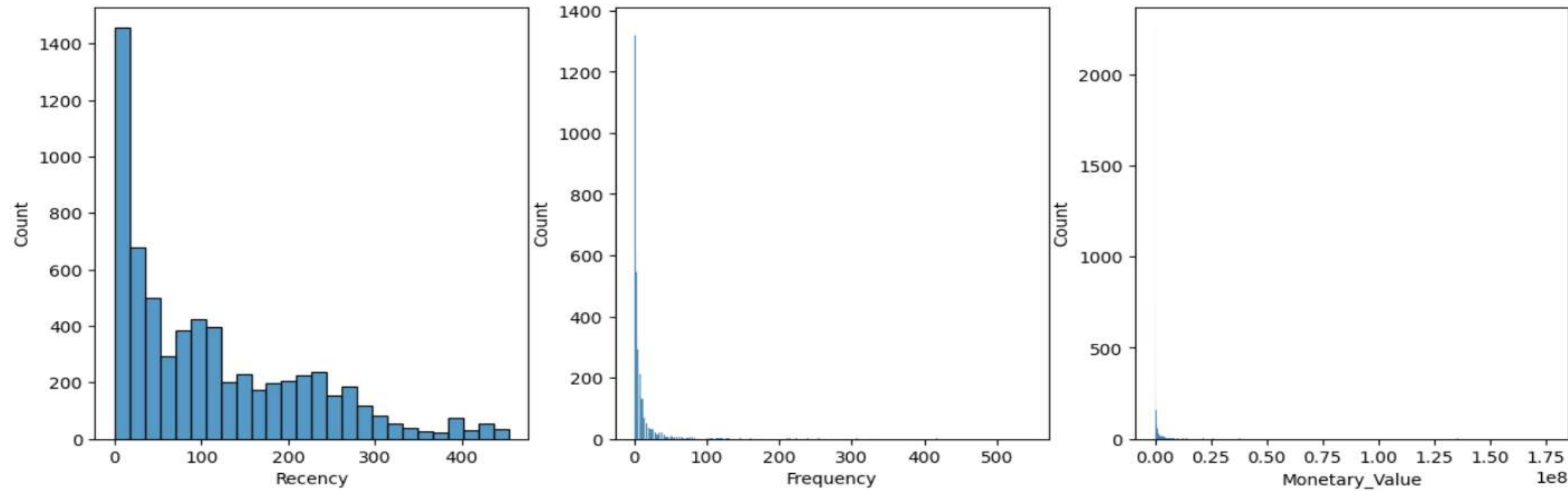
4. Feature Engineering

4.1. Transform Data

4.2. Scale Data

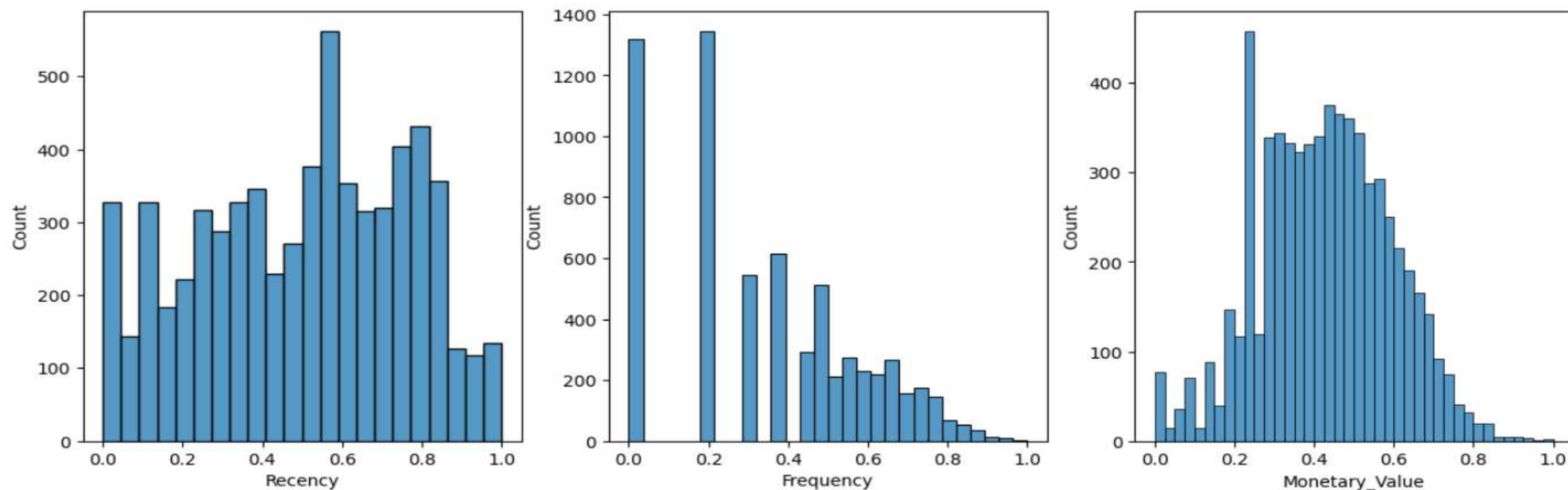
4. Scale Data

4.1. Transform Data



- From this subplots, our data is skewed to the left, meaning it is **not distributed well**.
- So we need to **transform** and **scale** the data

4.2 Scale Data



This is the distribution of RFM after using:

- **Box-Cox & Log Transformation**
- **Min-max Scaler**

5. Find K

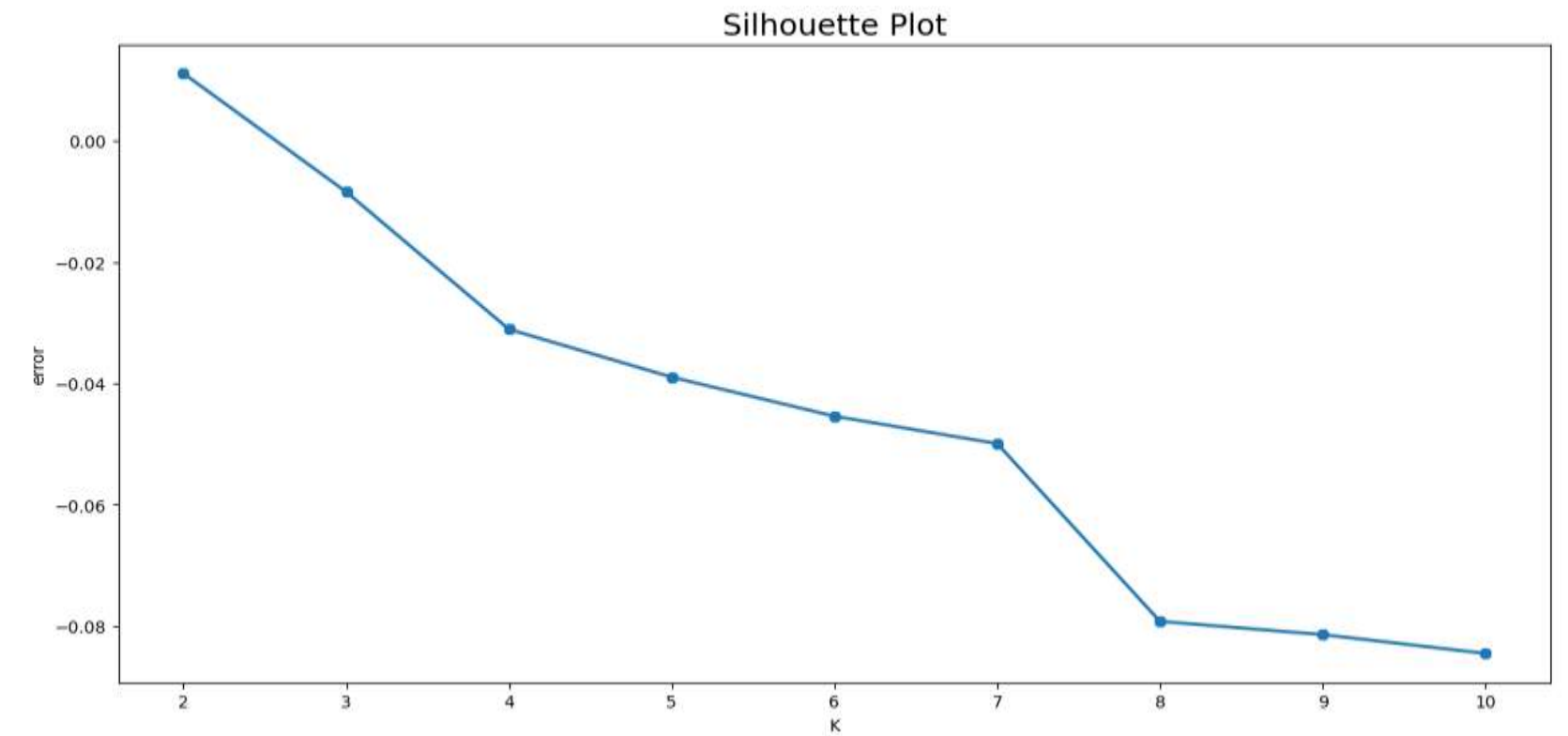
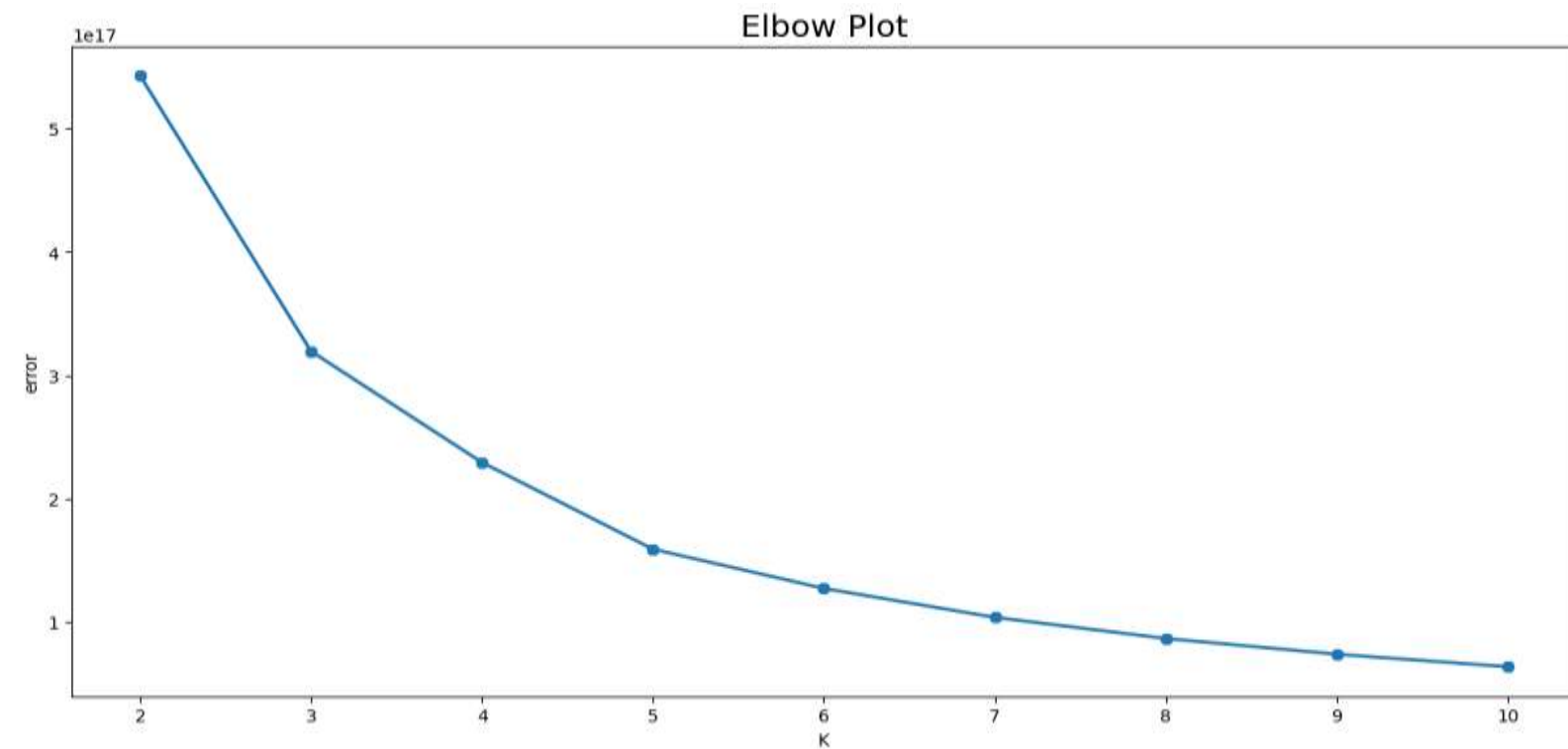
5.1 Find K with Elbow, Silhouette

5.2 Find K using KneeLocator

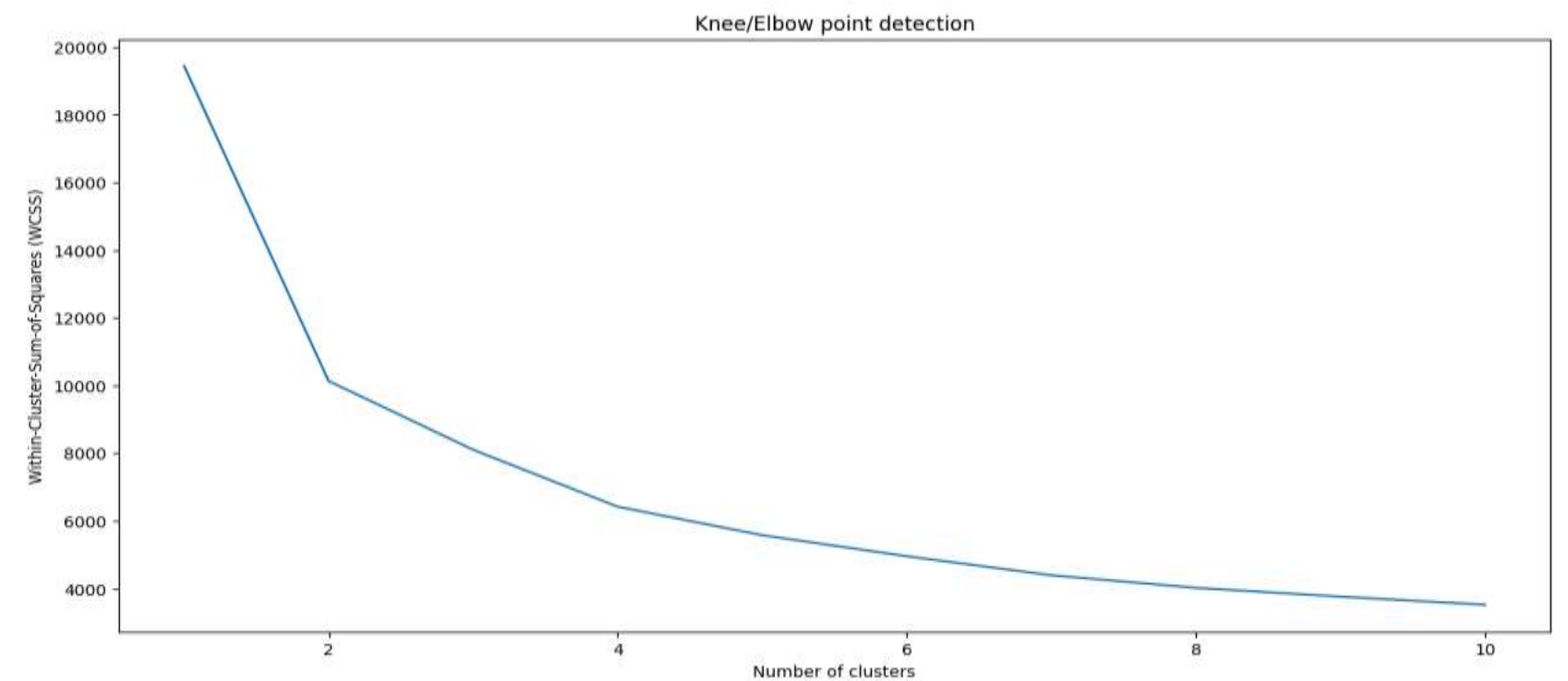
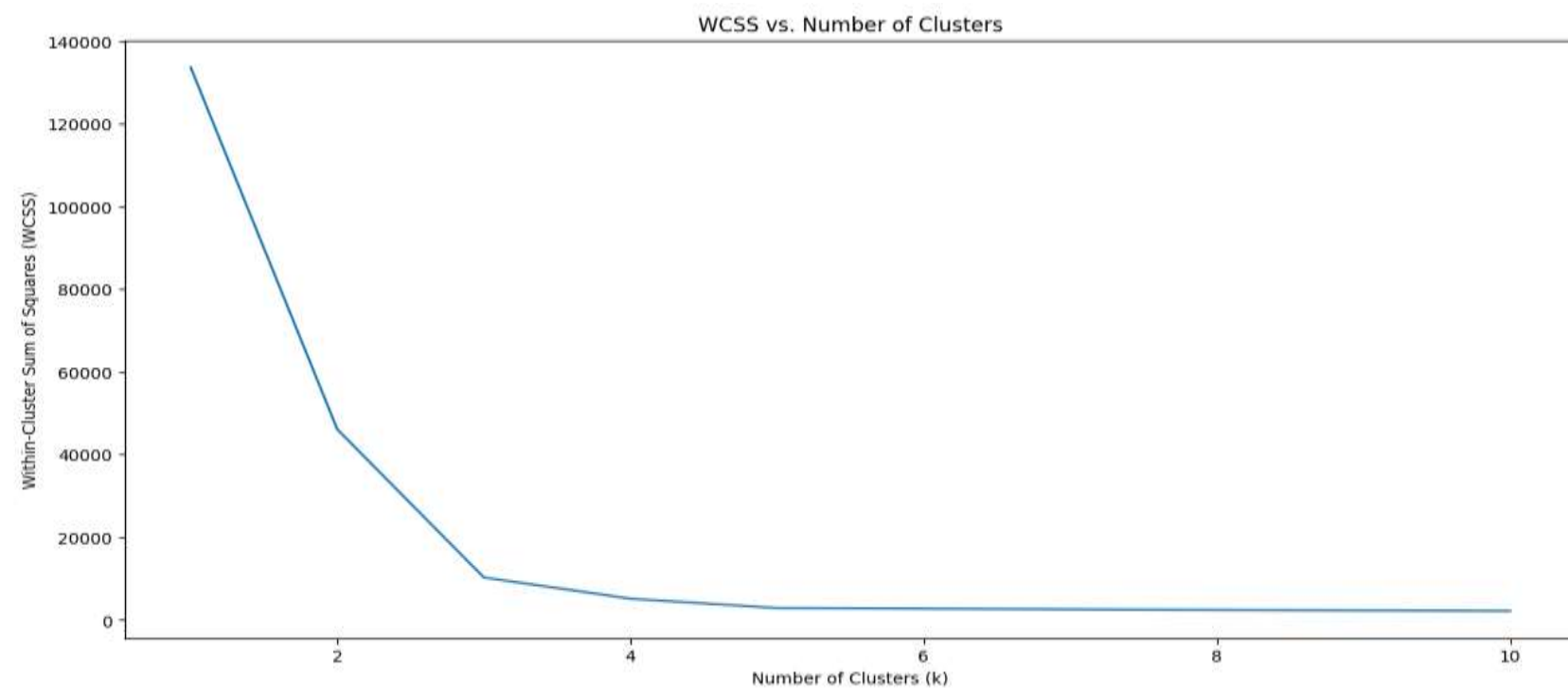
5.3 Find K using KElbowVisualizer

5. Find K

5.1 Find K with Elbow, Silhouette



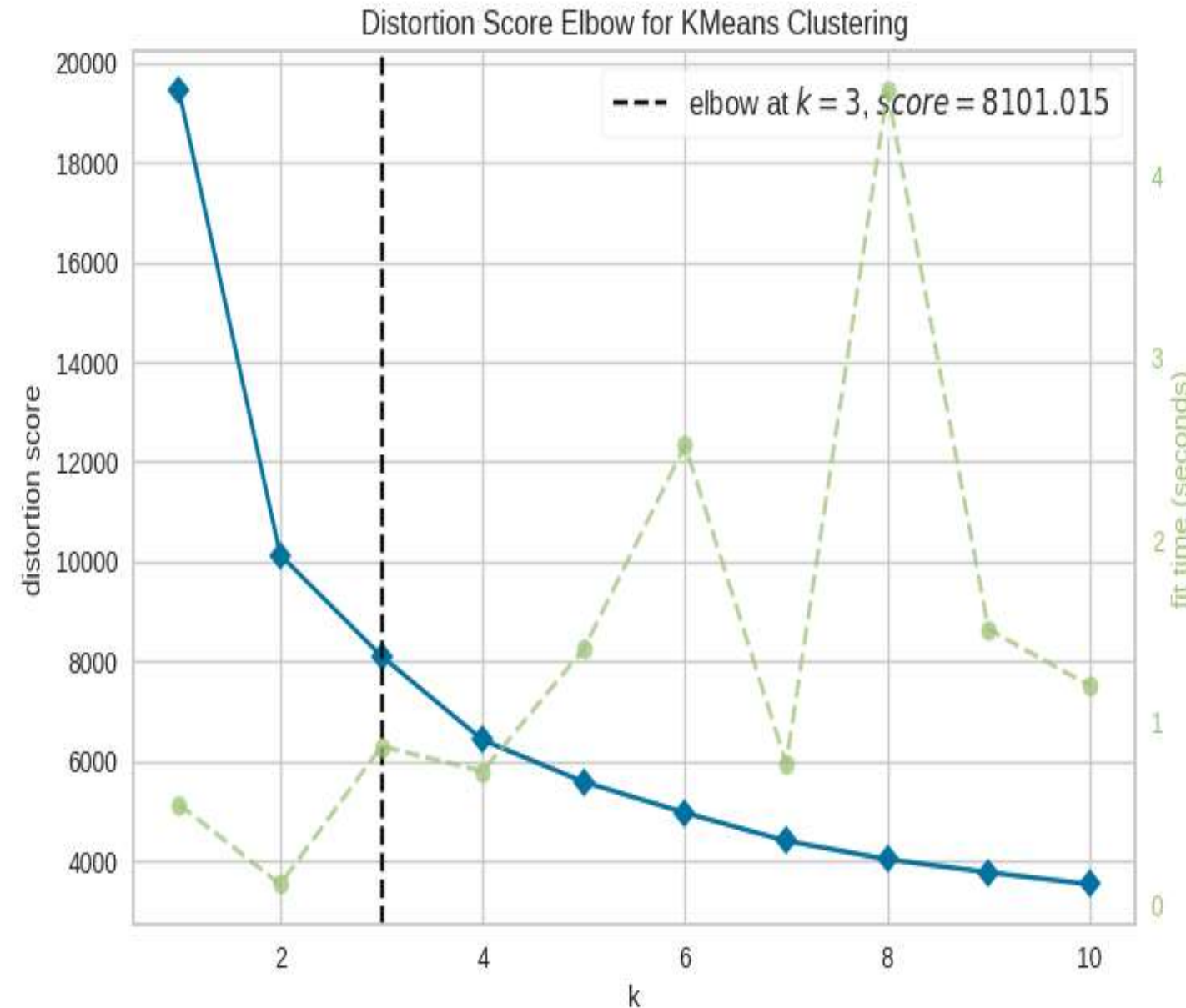
5.2 Find K with WCSS KneeLocator



Optimal number of clusters (k): 3

5. Find K

5.3 Find K with KElbow Visualizer



→ Choose K = 3

- Uses the Elbow method.
- Plots the within-cluster sum of squares (WCSS).
- Identifies the "elbow" point where the WCSS curve starts to level off.

6. Modeling

KMeans = 3

6.1. Applying clusters

6. Modeling KMeans = 3

6.1. Applying clusters

DATE	Order_id	NEWVERTICAL_Merchant	MerchantID	User_id	GMV	Service Group	Merchant	Recency	Frequency	Monetary_Value	Cluster
2021-01-01	8733622706	Marketplace	37	61386143	100000	marketplace	Tiki	36	49	2904300	0
2021-01-03	8759351785	Airtime	41	61386143	10000	data	Viettel	36	49	2904300	0
2022-02-08	20747004308	Coffee Chains And Milk Tea	35	44018659	69000	Coffee chains and Milk tea	The Coffee House	51	3	207000	1
2022-03-06	21673509622	Coffee Chains And Milk Tea	35	48918259	45000	Coffee chains and Milk tea	The Coffee House	25	3	135000	2
...

Cluster	Count
0	2297
1	2473
2	1709

7. Cluster Analysis

7.1. Average GMV of users in Clusters

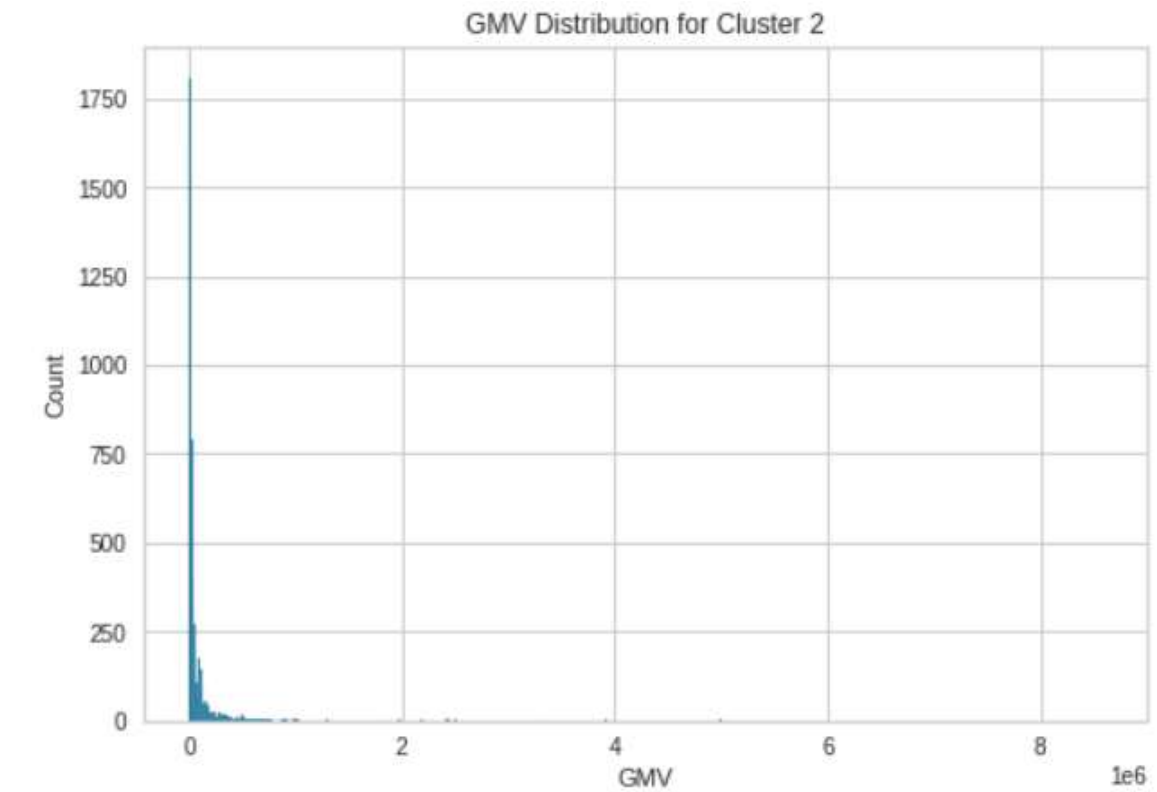
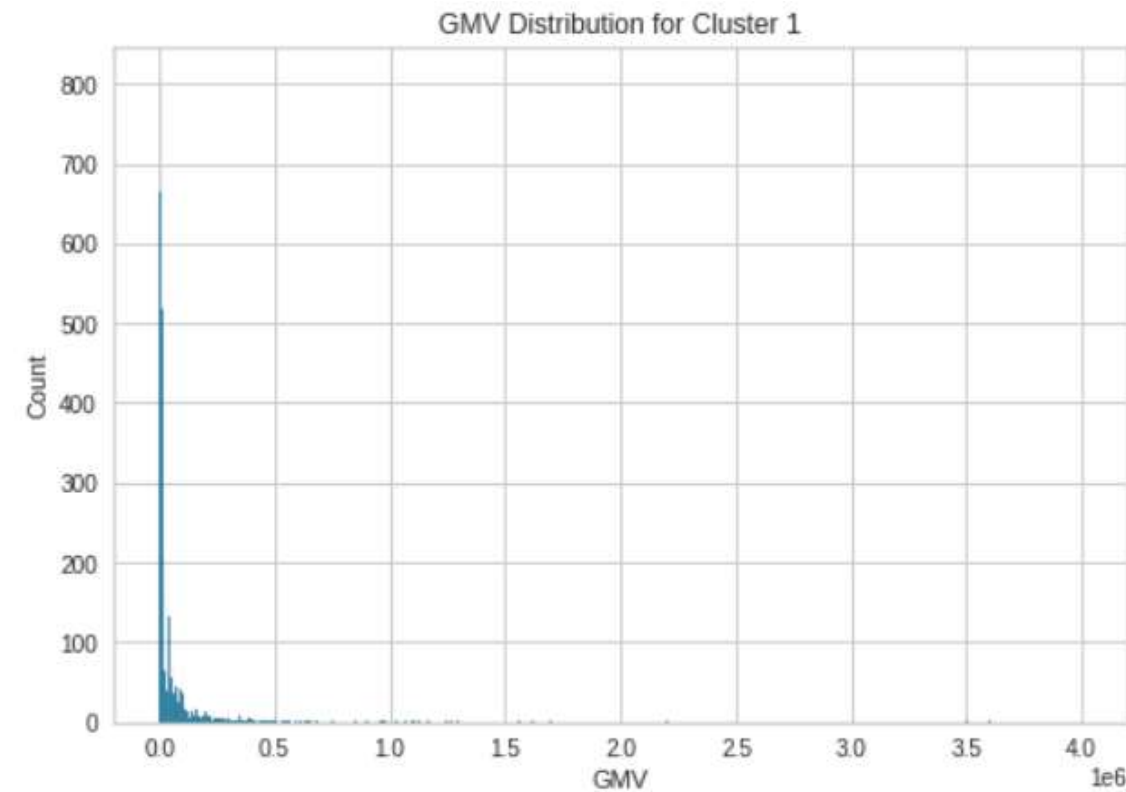
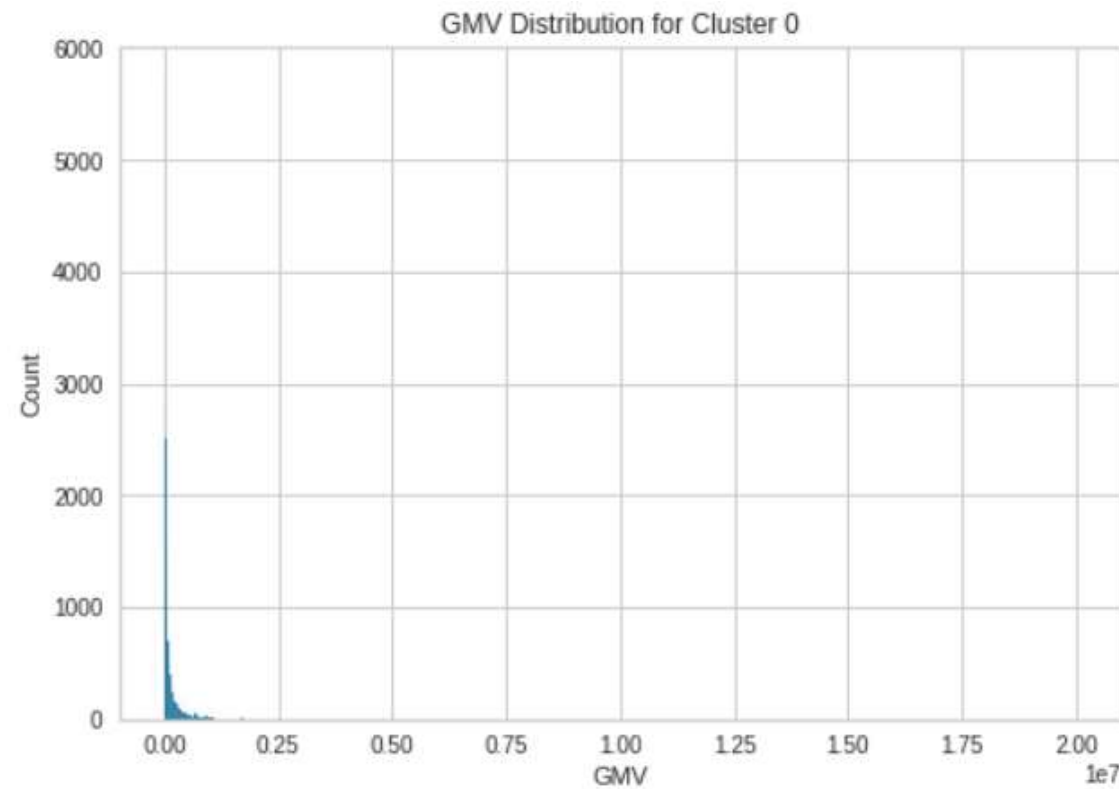
7.2. Most common service groups & merchants of users

7.3. Average recency and frequency of purchases from users in Clusters

7.4. Average monetary value of purchases from users in Clusters

7. Cluster Analysis

7.1. Average GMV of users in Clusters



Observations:

- Average GMV for Cluster 0: 122.457 VND
- Average GMV for Cluster 1: 52.087 VND
- Average GMV for Cluster 2: 157.749 VND

7. Cluster Analysis

7.2. Most common service groups & merchants of users

Cluster 0

Top Service Groups for Cluster 0:

data	4630
marketplace	2984
cvs	1490
supermarket	791
Coffee chains and Milk tea	330
Offline Beverage	2

Name: Service Group, dtype: int64

Top Merchants for Cluster 0:

Viettel	2624
Lazada	2256
Mobifone	1435
Circle K	823
Tiki	728
Vinaphone	571
Co.Opmart	548
Ministop	407
Highlands	121
Lotte Mart	101

Name: Merchant, dtype: int64

Cluster 1

Top Service Groups for Cluster 1:

data	1890
marketplace	793
cvs	312
supermarket	126
Coffee chains and Milk tea	93
Offline Beverage	1

Name: Service Group, dtype: int64

Top Merchants for Cluster 1:

Viettel	846
Lazada	644
Mobifone	618
Vinaphone	426
Circle K	204
Tiki	149
Co.Opmart	79
Ministop	52
Highlands	44
Lotte Mart	23

Name: Merchant, dtype: int64

Cluster 2

Top Service Groups for Cluster 2:

marketplace	12101
data	9570
cvs	8533
supermarket	3810
Coffee chains and Milk tea	1689
Offline Beverage	30

Name: Service Group, dtype: int64

Top Merchants for Cluster 2:

Lazada	10171
Viettel	5513
Circle K	3455
Mobifone	3105
Co.Opmart	2011
Ministop	1999
Tiki	1930
Vinaphone	952
Bsmart	935
Gs25	825

Name: Merchant, dtype: int64

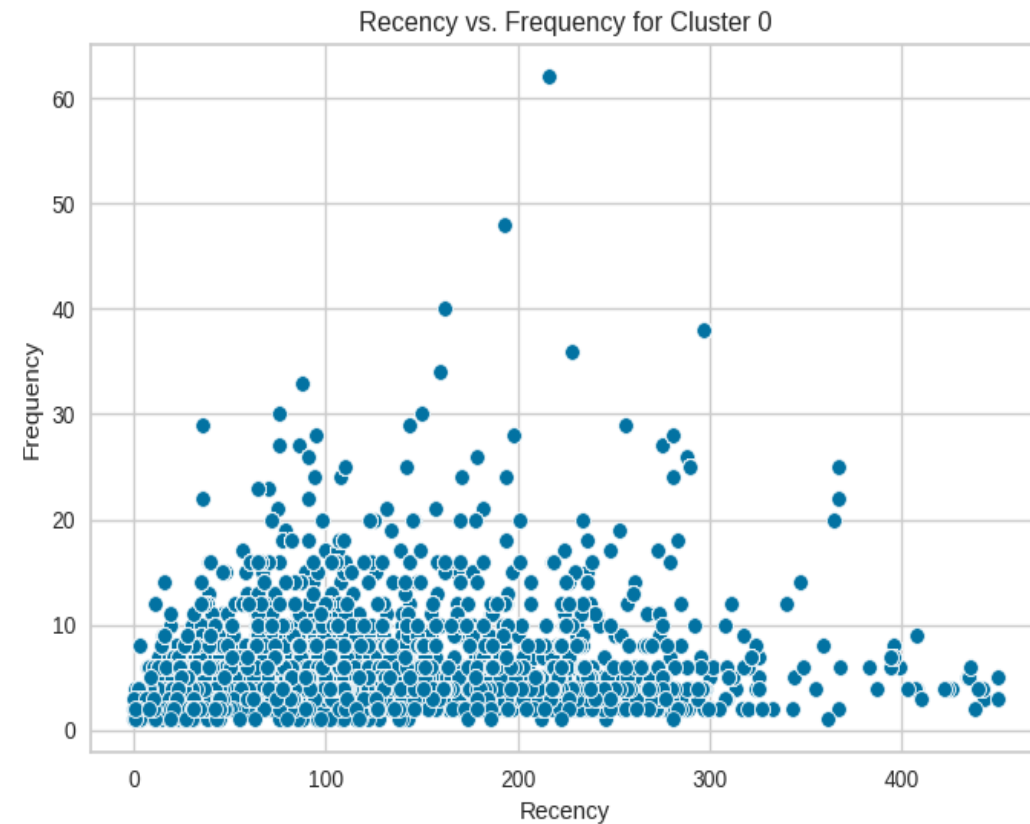
Observations:

- Top Service Groups for Cluster 0: Data, Marketplace, CVS, Supermarket, Coffee & Milk tea
- Top Service Groups for Cluster 1: Data, Marketplace, CVS, Supermarket, Coffee & Milk tea
- Top Service Groups for Cluster 2: Marketplace, Data, CVS, Supermarket ,Coffee & Milk tea

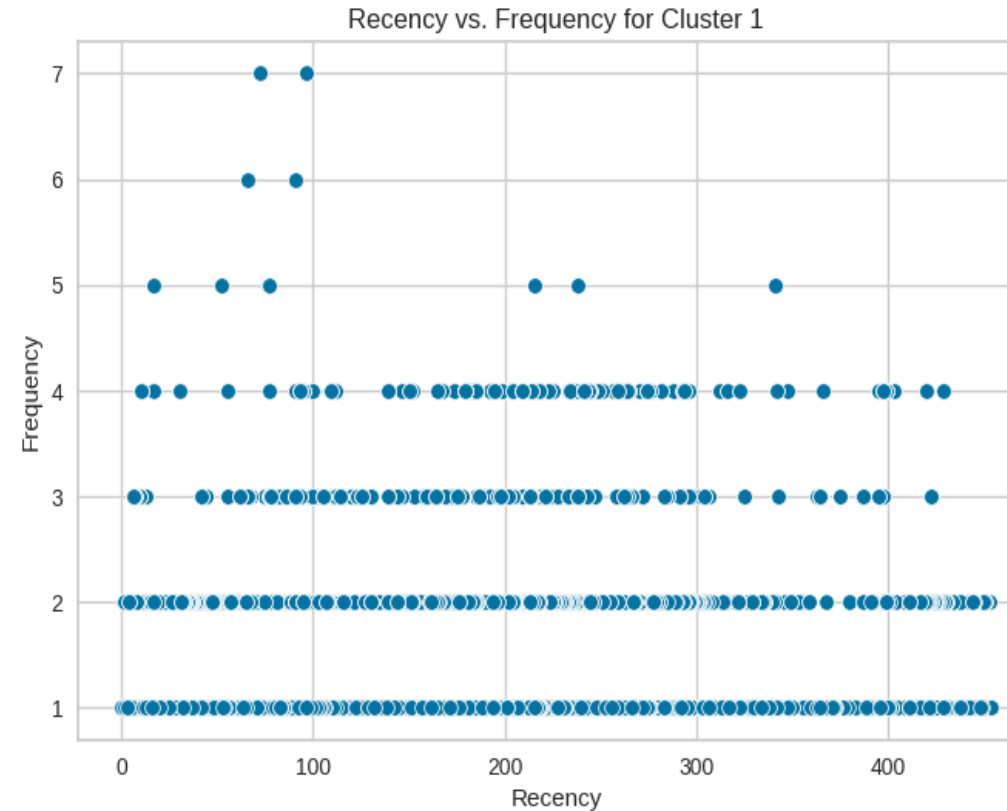
7. Cluster Analysis

7.3. Average recency and frequency of purchases from users in Clusters

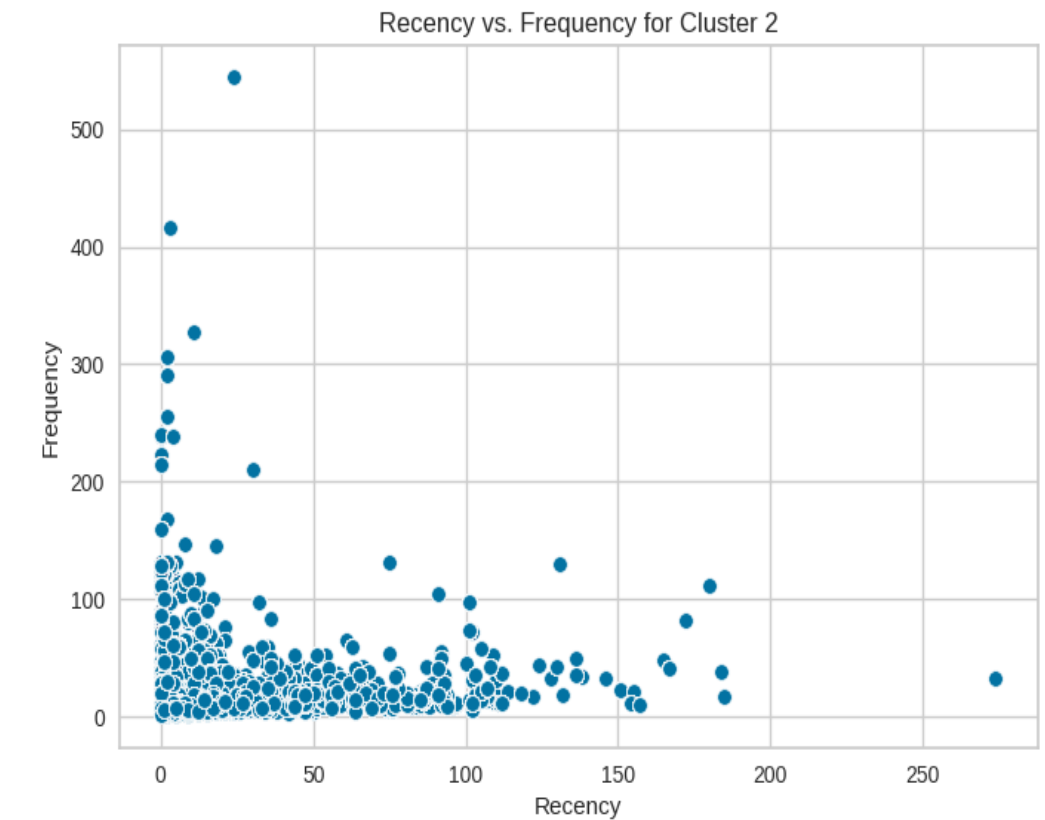
Average Recency for Cluster 0: 119.20983670675662
Average Frequency for Cluster 0: 8.83377334506698



Average Recency for Cluster 1: 178.50979782270608
Average Frequency for Cluster 1: 1.9399688958009331



Average Recency for Cluster 2: 19.588111829401395
Average Frequency for Cluster 2: 69.80396272353286



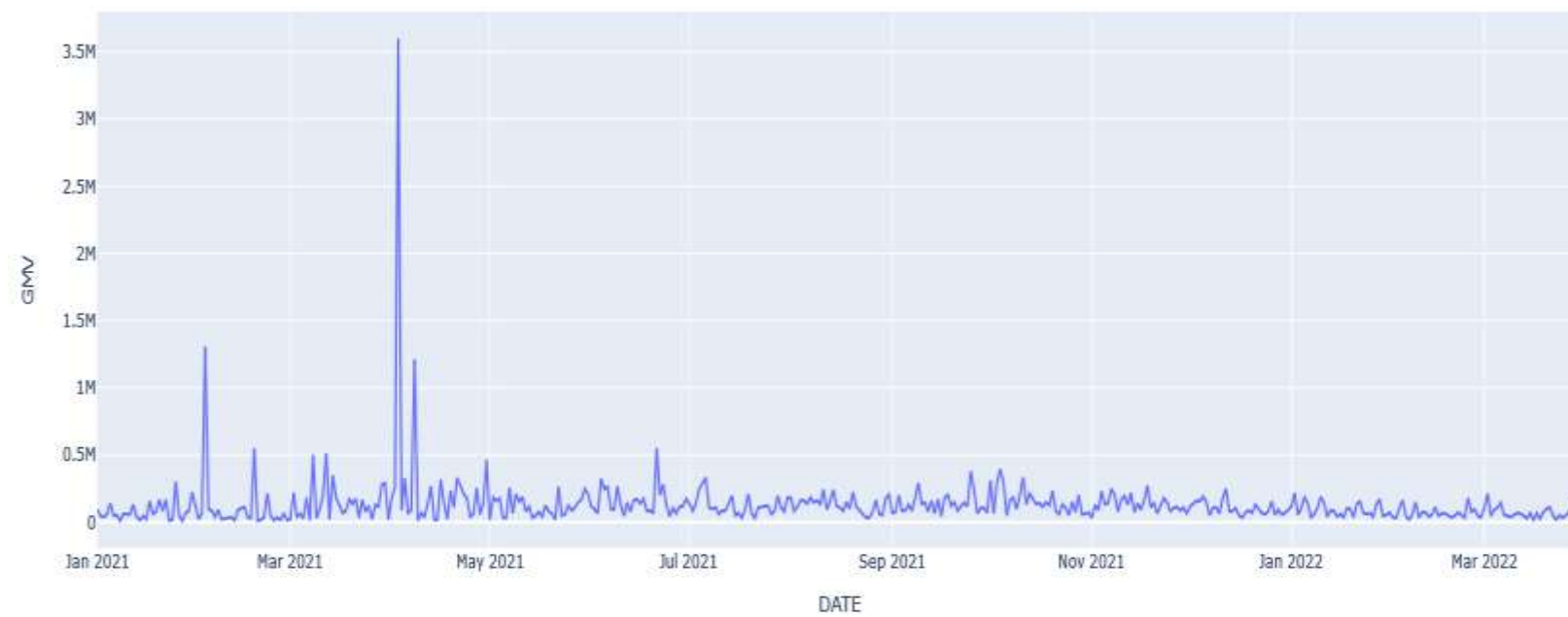
Observations:

- Average Recency, Frequency for Cluster 0: 119 & 9
- Average Recency, Frequency for Cluster 1: 178 & 2
- Average Recency, Frequency for Cluster 2: 19 & 69

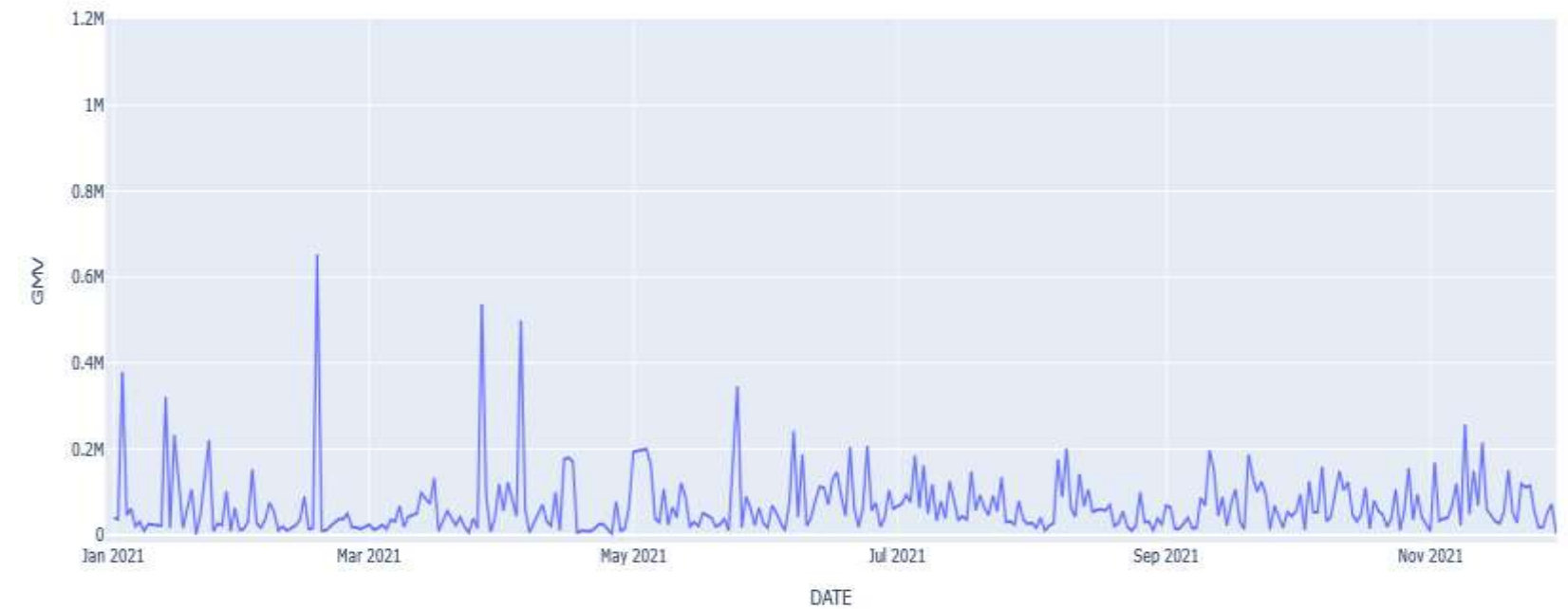
7. Cluster Analysis

7.4. Average monetary value of purchases from users in Clusters

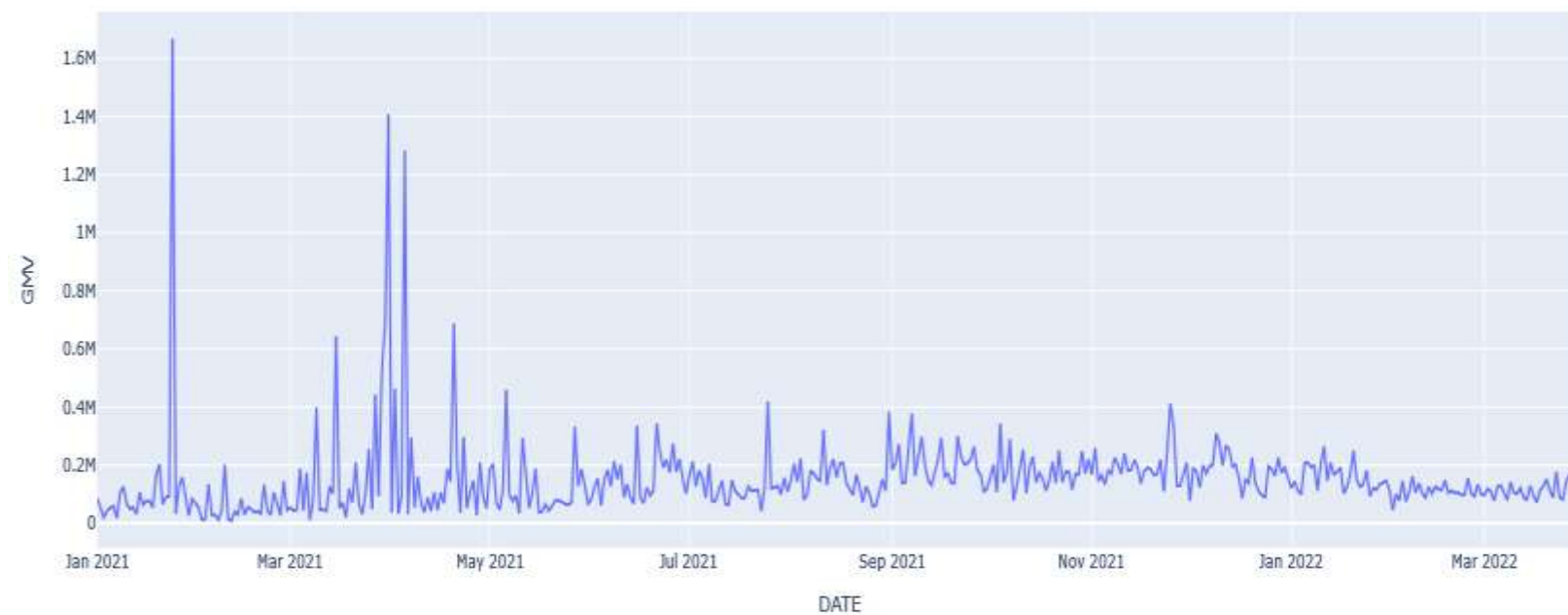
GMV over Time for Cluster 0 Customers



GMV over Time for Cluster 1 Customers



GMV over Time for Cluster 2 Customers



Observations:

- Average Monetary Value for Cluster 0: 677.205
- Average Monetary Value for Cluster 1: 68.022
- Average Monetary Value for Cluster 2: 10.300.165

8. Conclusion

8.1. Summary and Recommendations

Cluster 0

- Has the highest **Average Monetary Value** which is ~ **677.205 VND**
- Users belong in this cluster usually have transactions at **Data, Marketplace, Convenient Stores, Supermarket.**
- **Favorite merchants:** Viettel, Lazada, Mobifone, Circle K, Tiki.
- **Average Frequency** is **8** & **Average Recency** is **119**.

Recommendations:

- + Offer discounts from Data Mobile channels for 4G, Wi-fi.
- + Offer Tiki e-code for Tiki coin exchange.

Cluster 1

- The least **Average Monetary Value** which is ~ **52.022**.
- Users belong in this cluster usually have transactions at **Data, Marketplace, Convenient Stores, Coffee chains and Milk tea.**
- **Favorite merchants:** Viettel, Lazada, Mobifone, Vinaphone.
- **Average Frequency** is **2** and **Average Recency** is **178**.
- They likely to purchase for mobile service

Recommendations:

- Promote vouchers to pay bills monthly:
- + Discount 5.000 VND from 50.000-bill for Internet
 - + Discount 5.000 VND from 200.000-bill for Mobile Service

Cluster 2

- The second highest **Average Monetary Value** which is ~ **157.749**.
- Users belong in this cluster usually have transactions at: **Marketplace, Data, CVS, Supermarket.**
- **Favorite merchants:** Lazada, Viettel, Circle K, Mobiphone, Co.op Mart.
- **Average Frequency** is **69** and **Average Recency** is **19**.
- GMV over time significantly **increased in Jan, Mar, April**

Recommendations:

- + Discount 10.000 VND from 300.000-bill for Electricity
- + Discount Buy 2 get 1 for Highlands
- + Convert Tiki Coin from eGift code

Reference:

- An article by Jan Roelf Bult and Tom Wansbeek, titled “Optimal Selection for Direct Mail,” published in a 1995 issue of *Marketing Science*.
- [KElbow Visualize](#)
- [Knee-elbow-point-detection](#)

THANKS !!!

Tên: Phạm Quốc Huy

Mentor: Nguyễn Thị Thắm

Lớp/Nhóm: DA18/Nhóm 2

