IMPROVING CUSTOMER LOYALTY OF A VIETNAMESE FASHION BRAND BASED ON CLV

Author: Nguyen Thanh Luan

1. Abstract

Over the past decade, Vietnam has witnessed rapid growth in internet users, driving a parallel rise in digital buyers. Fashion e-commerce has emerged as a dominant segment among these buyers, with a significant portion of transactions involving apparel purchases. This growth highlights the strategic importance of customer loyalty to maintaining competitiveness in Vietnam's rapidly growing e-commerce market.

This study aims to identify the loyalty level of an electronic customer based on Customer Lifetime Value (CLV) of the customer segmentation and design the CLV improvement. Customer segmentation performed using the K-means algorithm and RFM analysis. This paper used 2 transaction datasets from 2 different time points (2023 and 2024) of local brand fashion e-commerce in Viet Nam.

2. Literature Review

CRM: is an analytical tool that can support companies to improve their business units performance such as increase their profit, improve their customers satisfaction, and also as the effective marketing and communication strategies with customers loyalty, potency and price reduction.

CLV: is the current worth of all profits obtained from customer and it helps the owner of the business unit target the acceptable markets additional effectively. By using CLV, company can targeting only the profitable customers who can give more profits to the company in the future. CLV can be counted in a certain time period or from the first time doing transaction until now.

Customer segmentation: has a significant effect on customer management through dividing customers into several groups according to their character, where companies can utilize the information contained in this segmentation of customers to market products differently by focusing on the needs of each type of customer group.

Clustering: is the most common form of unsupervised learning. It is the process that involves the grouping of data points into classes of similar objects.

RFM: RFM variables consist of recency (R), frequency (F), and monetary (M). R measured by analyzing the time gap between the last transaction date performed by any customer and the last transaction date, F measured by counting the number of transaction(s) which happen throughout the year for each customer and M measured by summing all the total cost of every customer.

AHP: Analytical Hierarchy Process is a decision-making method used by individuals and organizations to rank alternatives they are considering based on pairwise comparisons. It is widely used for decision-making in many fields, including business, government, engineering, ...

3. Experiments

3.1. Data situation

Columns	Description
Phone	unique identifier for each customer
Cus name	name of the customer
Recency	number of days since the last purchase
Frequency	number of purchases made in the year
Monetary	amount due for the entire purchase
diversity	average number of unique products purchased
first_ord	number of days since from start date to first purchase
aov	Average order value

Table 1. Data infomation

	Recency	**Frequency**	**Monetary**	**Diversity**	**AOV**
2023	175	1.223	5.75e+5	1.69	4.95e+5
2024	182	1.288	5.87e+5	1.59	4.90e+5

Table 2. Overview table of indicators

From Table 2, it can be observed that although customers in 2024 tend to place more orders and generate higher revenue compared to 2023, the average number of product categories per order (diversity) and the average order value (AOV) have slightly decreased. A preliminary explanation for this could be as follows: "The brand has successfully attracted returning customers and increased their spending through advertising, marketing, and customer appreciation campaigns. However, the product catalog and designs lack significant innovation or appeal, resulting in customers mainly returning to purchase seasonal or essential products."

Additionally, it was assumed that as customers make repeat purchases more frequently, the gap between their last purchase and the end of the data collection period would decrease. However, the statistical data shows results contrary to this assumption. A possible explanation could be that the brand launched several customer acquisition campaigns earlier in the year, but recent campaigns were less effective in sustaining customer engagement.

Based on the statistical results and hypotheses, further analysis can be conducted in two directions:

+ Analysis based on business and marketing activities.

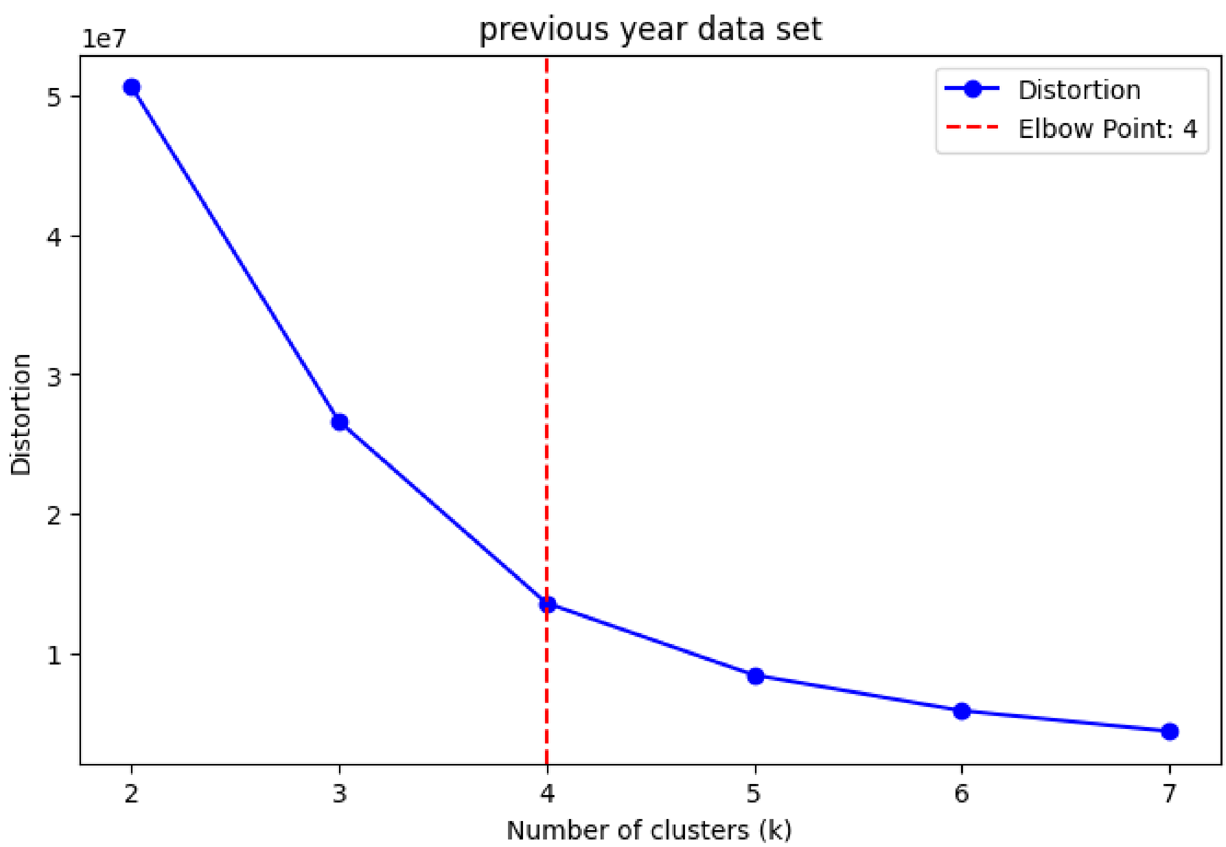
+ Analysis based on customer group behavior (*), this analysis will focus on the latter direction.

3.2. Customer segmentation

In this process, I chose to use the K-means algorithm with three features, R, F, and M, as input, normalized using the min-max method (eq.1). The optimal number of clusters was determined by combining the elbow and silhouette methods. Finally, the clusters were ranked based on CLV, calculated using eq.2, where N represents normalized values and W denotes the weights of R, F, and M, determined using the AHP method.

$$NV = \frac{V-min}{\max - min} new_{max} - new_{min} + new_{min} \quad (Eq.1)$$

$$CLV = (NR * WR) + (NF * WF) + (NM * WM) \quad (Eq.2)$$



After identifying the optimal number of clusters as 4 in 2023, I proceeded to infer insights from the 2024 dataset and calculated the CLV. First, my team and I assessed the data, considered industry-specific characteristics, and aligned with the business objectives. Based on this evaluation, we prioritized the factors R, F, and M in the following order: M > R > F. The initial weights were set as follows: weightR = 1/3, weightF = 1/6, weightM = 1/2.

```python
import numpy as np

priority_matrix = np.array(
                        [[1,2,2/3],
                        [1/2,1,1/3],
                        [3/2,3,1]]
                    )
print('Original matrix')
print(priority_matrix)
print('===============================')

col_sum = np.sum(priority_matrix, axis=0)
priority_matrix = priority_matrix/col_sum[np.newaxis, :]

print('Priority matrix')
print(priority_matrix)
print('===============================')

weight = np.average(priority_matrix, axis=1)
print('Weight')
print(weight)
```

✓ 0.0s

```
Original matrix
[[1.         2.         0.66666667]
 [0.5        1.         0.33333333]
 [1.5        3.         1.        ]]
===============================
Priority matrix
[[0.33333333 0.33333333 0.33333333]
 [0.16666667 0.16666667 0.16666667]
 [0.5        0.5        0.5       ]]
===============================
Weight
[0.33333333 0.16666667 0.5       ]
```

After calculating the basic weights, I proceeded to validate the CR (Consistency Ratio), CI (Consistency Index), and RI (Random Index) to determine whether the evaluations made by my team and I achieved consistency in prioritization, as shown in (Eq.3). The results indicated that CR < 10%, meaning it is feasible to continue using this approach for the research.

$$CR = \frac{CI}{RI} \; (eq.3)$$

Using these weights in combination with the centers of each cluster, I calculated the representative CLV for each cluster, as shown in Table 3.

	2023	2024
Best	0.597308	0.567622
Valuable	0.441471	0.439823
Potentially valuable	0.336614	0.360096
Likedly churn	0.277341	0.295365

Table 3. CLV table

Table 3 indicates that customers in the *Best* and *Valuable* groups experienced a decline in value, whereas those in the *Potentially Valuable* and *Likely Churn* groups showed an increase in value in 2024. This may potentially explain the observed decrease in the statistical results in Table 2 for 2024. To further investigate this, I will analyze the transition of customer segments across the two periods.

	label_1	label_2	count	%_count	total_monetary_1	%m_last_year	total_monetary_2	%m_current_year
24	NaN	NaN	93217	1.000000	2.656033e+10	1.000000	2.908615e+10	1.000000
16	Unknown	Likedly churn	19433	0.208471	0.000000e+00	0.000000	1.156598e+10	0.397646
11	Potentially valuable	Churn customer	16088	0.172587	8.918922e+09	0.335799	0.000000e+00	0.000000
17	Unknown	Potentially valuable	15292	0.164047	0.000000e+00	0.000000	8.881452e+09	0.305350
6	Likedly churn	Churn customer	14679	0.157471	8.655001e+09	0.325862	0.000000e+00	0.000000
20	Valuable	Churn customer	11263	0.120826	6.251425e+09	0.235367	0.000000e+00	0.000000
18	Unknown	Valuable	10219	0.109626	0.000000e+00	0.000000	5.662965e+09	0.194696
15	Unknown	Best	2088	0.022399	0.000000e+00	0.000000	1.482904e+09	0.050983
1	Best	Churn customer	1691	0.018140	1.253275e+09	0.047186	0.000000e+00	0.000000
12	Potentially valuable	Likedly churn	382	0.004098	2.177404e+08	0.008198	2.319730e+08	0.007975
7	Likedly churn	Likedly churn	309	0.003315	1.923511e+08	0.007242	1.865736e+08	0.006415
13	Potentially valuable	Potentially valuable	308	0.003304	1.787348e+08	0.006729	1.866643e+08	0.006418
21	Valuable	Likedly churn	251	0.002693	1.453632e+08	0.005473	1.570361e+08	0.005399
8	Likedly churn	Potentially valuable	245	0.002628	1.522508e+08	0.005732	1.429220e+08	0.004914
22	Valuable	Potentially valuable	204	0.002188	1.162900e+08	0.004378	1.226003e+08	0.004215
14	Potentially valuable	Valuable	176	0.001888	1.000452e+08	0.003767	1.036298e+08	0.003563
9	Likedly churn	Valuable	162	0.001738	1.011018e+08	0.003806	9.425503e+07	0.003241
23	Valuable	Valuable	128	0.001373	7.903955e+07	0.002976	7.313914e+07	0.002515
5	Likedly churn	Best	62	0.000665	3.990484e+07	0.001502	4.132176e+07	0.001421
10	Potentially valuable	Best	61	0.000654	3.537095e+07	0.001332	4.126310e+07	0.001419
2	Best	Likedly churn	51	0.000547	3.783708e+07	0.001425	2.956935e+07	0.001017
3	Best	Potentially valuable	45	0.000483	3.347753e+07	0.001260	2.713833e+07	0.000933
19	Valuable	Best	42	0.000451	2.308855e+07	0.000869	3.127023e+07	0.001075
4	Best	Valuable	25	0.000268	1.921946e+07	0.000724	1.502983e+07	0.000517
0	Best	Best	13	0.000139	9.893330e+06	0.000372	8.458020e+06	0.000291

Table 4. Conversion statistics table between customer groups in two periods

The statistical results on segment transitions in Table 4 reveal that the company has excelled in attracting new customers. Specifically, in 2024, the company acquired 47,032 new customers, contributing revenue of 25.6 billion VND, accounting for 94.46% of the company's total annual revenue. This indicates a positive business performance. However, the company needs to address the question: “Why was the company able to attract such a high number of new customers? Was it due to high-quality products, attractive designs, competitive pricing, or marketing strategies that effectively targeted customer needs?”

However, in 2024, as many as 43,721 customers left the company, and this group contributed 25.08 billion VND, accounting for 94.42% of the company’s total revenue in 2023. This reflects a negative aspect of the company’s performance. The question arises: *“Did customers leave the brand due to dissatisfaction with customer service, products no longer meeting their needs, or negative experiences from their last purchases? Or, in a worse scenario, were they attracted to a competing brand?”*

Quoting a study from Harvard Business School: *“Increasing customer retention by 5% can boost profits by 25% to 95%. Many businesses affirm that their survival depends on loyal customers.”* Or another conclusion from researchers: *“It has been proven that acquiring a new customer costs a business 5–7 times more than retaining an existing one. Additionally, for new customers to spend as much as existing ones, businesses need to invest 16 times more. This figure continues to rise amidst intensifying competition and soaring advertising expenses.”*

Therefore, I do not conclude whether the company's business performance is positive or negative. However, from my perspective, the company seems to be missing out on potential customers and wasting resources to pursue new ones, even though they could optimize existing resources at a lower cost.

Finally, I will address the question of why the "Best" and "Valuable" customer segments—considered the most desirable—experienced a decline in CLV in 2024, particularly the "Best" segment. Conversely, I will explain why the opposite trend was observed in the "Likely to Churn" and "Potentially Valuable" segments, as concluded in Table 3.

	label_1	label_2	count	aov1	aov2	monetary1	monetary2
0	Best	Best	13	380512.692308	325308.461538	761025.384615	650616.923077
1	Best	Likedly churn	51	370951.754902	443777.980392	741903.509804	579791.254902
2	Best	Potentially valuable	45	371972.511111	456349.533333	743945.022222	603073.977778
3	Best	Valuable	25	384389.200000	492131.200000	768778.400000	601193.200000
4	Likedly churn	Best	62	472791.048387	333240.040323	643626.451613	666480.080645
8	Potentially valuable	Best	61	467170.491803	338222.163934	579851.639344	676444.327869
12	Valuable	Best	42	468194.285714	372264.619048	549727.380952	744529.238095

Table 5. Revenue Metrics Summary Table for the Best Group Over Two Years

From the statistical results in Table 5, it is evident that customers who maintained the "Best" status continuously for two years experienced a decline in AOV and monetary value. Additionally, customers transitioning from three other segments in 2023 to the "Best" segment in 2024 had lower AOV and monetary value compared to customers who belonged to the "Best" segment in 2023 but transitioned in 2024. Similar analyses for other segments reveal that AOV and monetary value are the primary factors driving the increase or decrease in CLV across segments.

4. Conclusion

In this report, I have outlined a method for customer segmentation using unsupervised learning algorithms, combined with hierarchical analysis, to identify customer lifetime value (CLV) and assign descriptive names to each cluster. I also highlighted the challenges and opportunities the business is facing through a flowchart illustrating customer segment transitions. However, due to the absence of data on campaigns, detailed orders, or customer feedback after purchases, the analysis could not delve deeper into the underlying causes.

5. Repo: https://github.com/thanhluan7702/Improve-customer-loyalty-Vfashion