International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

RFM-T Model Clustering Analysis in Improving Customer Segmentation

Astrid Dewi Rana[1], Quezvanya Chloe Milano Hadisantoso[1] and Abba Suganda Girsang[1]

[1]Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University Jakarta, Indonesia, 11480.

E-mail address: astrid.rana@binus.ac.id, quezvanya.hadisantoso@binus.ac.id, abba.girsang@binus.ac.id

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: In the dynamic landscape of business, understanding and identifying customers are paramount for effective marketing strategies. This study delves into the realm of customer segmentation, a crucial component of robust marketing strategies, particularly focusing on the widely adopted RFM (Recency, Frequency, and Monetary) model. Various new models of RFM have been explored, with a notable extension being the RFM-T model, introducing the "T" variable to represent Time. This study aims to compare the performance of the traditional RFM model with the innovative RFM-T model, assessing their efficacy in customer segmentation. Utilizing a dataset sourced from a US-based online retail platform, the study employs the K-Means algorithm for segmentation, a method commonly utilized for partitioning data points into distinct clusters. To ascertain the optimal number of clusters, the Elbow Curve approach is employed, offering insight into the granularity of segmentation. Subsequently, the Silhouette Score, a metric used to assess the cohesion and separation of clusters, is leveraged to evaluate the quality and effectiveness of both models. By conducting a comparative analysis of the traditional RFM model and its enhanced RFM-T counterpart, the study endeavors to shed light on their respective contributions to the refinement of customer profiling and segmentation strategies within the online retail industry. Through this exploration, businesses can glean valuable insights into the evolving landscape of customer segmentation, thereby enabling them to tailor their marketing efforts more precisely and effectively to meet the dynamic needs and preferences of their target audience.

Keywords: RFM, RFM-T, Time, K-Means algorithm, Customer segmentation

1. Introduction

This In the business domain, a pivotal emphasis lies in identifying and understanding customers to implement effective marketing strategies and optimize their lifetime value. A robust marketing strategy involves deploying diverse and impactful tactics tailored to individual customer needs. Furthermore, database marketing is a common approach employed in customer segmentation for direct marketing endeavors. With the rapid expansion of collected data, marketers encounter the challenge of allocating their marketing communication budget judiciously, focusing on the most promising customers [1]. Leveraging advanced data insights allows us to know who our customers are and their behavior. This identification of customer profiling serves as a valuable tool in enhancing marketing strategies, facilitating more informed and precise decision-making processes.

Segmenting customers according to the RFM (Recency, Frequency and Monetary) variables has become widely adopted in contemporary business practices. This method categorizes customers based on their distinct characteristics and behavior. Various clustering algorithms have been applied to effectively group customers, aiming to enhance the precision and efficacy of clustering outcomes [2]. This model evaluates customer behavior based on three key metrics: recency, which assesses how recently a customer has made a purchase; frequency, which measures the rate of customer transactions; and monetary, which gauges the total monetary value of a customer's transactions. RFM facilitates the identification of buyer characteristics that influence responses [3].

The variable "T" has been added to the RFM model, which goes by the name RFM-T, to indicate inter-purchase time. T determines the typical amount of time that passes between a customer's subsequent transactions [4]. The purpose of this T addition is to examine the connection between the ocurrency and tendency of internet purchasing [5]. As a result, there is a strong likelihood that RFM-T

E-mail:author's email

could surpass traditional RFM models, especially in the online retail industry.

The primary objective of this study is to conduct a comparative analysis between the widely recognized RFM model and the RFM-T model, aiming to discern if the latter demonstrates significant enhancements in the customer segmentation process. To achieve this objective, we utilize a comprehensive dataset comprising all transactions conducted within a US-based online retail platform. Customer segmentation will be performed using the K-MEANS algorithm, employing both the extended RFM-T model and the traditional RFM model. Subsequently, the optimal number of clusters for the study will be determined precisely through the application of the elbow curve method. Furthermore, an evaluation of both models' quality will be conducted based on the silhouette score.

2. LITERATURE REVIEW

The RFM model's first concepts were presented. The RFM model's definition was initially put out by Hughes [6]. Only transactional factors like recency, frequency and monetary are taken into account by this classic RFM model, which excludes other customer attributes [7]. Consequently, a great deal of research has been done to enhance customer segmentation effectiveness by applying machine learning and including additional variables into the conventional RFM model.

In 2021, Time or Interpurchase Time was applied to the RFM model [4]. An algorithm that can improve the segmentation method is used by RFM-T. Inter-purchase time, designated by the variable T in the dataset, is the interval of time between two consecutive transactions by the same customer. According to Donald, G.M., this approach has been used for behaviour analysis in business since the 1960s [8]. Then, in 2023, RFM-T was analyzed using K-Means, Gaussian and DBSCAN clustering algorithms [9]. However, it hasn't proven the effectiveness of adding T variables compared to the classic model.

Acknowledging the significance of assessing the influence of the T variable in the RFM model, a comparative analysis is conducted to evaluate the clustering performance of RFM and RFM-T. Moreover, it extends this investigation to include RFM-D, where D signifies the inclusion of product diversity as an additional variable in the RFM framework [10]. Through empirical analysis, particularly employing the K-Means clustering algorithm, the study observes that RFM-D outperforms RFM in terms of clustering quality, as evidenced by its higher Silhouette Score. This finding underscores the significance of incorporating product diversity into customer segmentation models for more accurate and insightful clustering outcomes.

The most popular clustering techniques include K-Means [11]. RFM models and other customer segmentation models could potentially be used with the K-Means algorithm. K-Means is a non-hierarchical clustering method that partitions data into clusters, emphasizing high intra-cluster similarity and low inter-cluster similarity [12]. While K-Means is popular due to its algorithmic simplicity and speed in selecting the cluster center (centroid), however, a challenge still exists in figuring out the ideal number of clusters (k) [13]. According to research by Subbalakshmi et al., choosing the right initial value and cluster selection can increase the K-Means method's accuracy [14].

The Elbow Method (EM) plays a crucial role in identifying the optimal number of clusters by assessing the curvature of the curve formed when plotting the number of clusters against a certain metric, typically the within-cluster sum of squares (WCSS) or inertia [15]. This method aids in determining the point where the curve exhibits a significant change in slope, often resembling an elbow shape. This pivotal point indicates the optimal number of clusters, denoted by 'k'. Typically, the elbow point signifies the balance between minimizing within-cluster variance and avoiding overfitting, making it a vital step in cluster analysis.

The Silhouette Score stands out among various methods used to evaluate clustering results. Unlike most other performance assessment techniques, the Silhouette Score does not necessitate a training set for evaluating clustering outcomes. This characteristic makes it particularly well-suited for assessing RFM clustering [16]. Since RFM clustering aims to categorize customers based on their transaction behavior, the Silhouette Score's ability to evaluate clustering quality without requiring labeled data makes it a valuable tool in this context. This approach enables businesses to effectively measure the coherence and separation of clusters generated by the RFM clustering algorithm, providing insights into the effectiveness of the segmentation process.

3. RESEARCH METHODOLOGY

This section describes the suggested approach for identifying customer segments utilizing the RFM and RFM-T model then using clustering algorithms (Elbow Curve and K-means) in order to maximize benefits and compare both of the models.

The proposed systematic framework in Figure 1, the dataset undergoes pre-processing and normalization to handle missing values and standardize scales. Feature extraction follows, generating Recency (R), Frequency (F), Monetary (M) and Time (T) attributes that encapsulate essential customer behavior patterns. The K-Means clustering algorithm is then applied to both RFM (Recency, Frequency, Monetary) and RFM-T (including Time) models for segmentation. Silhouette analysis evaluates the quality of clusters, facilitating a comparative study between RFM and RFM-T models. The final stage involves detailed customer analysis within each segment, providing insights into customer characteristics and preferences. This comprehensive approach, spanning pre-processing, feature extraction, clustering and customer analysis, forms a structured framework for effective customer segmentation using the RFM and RFM-T models.

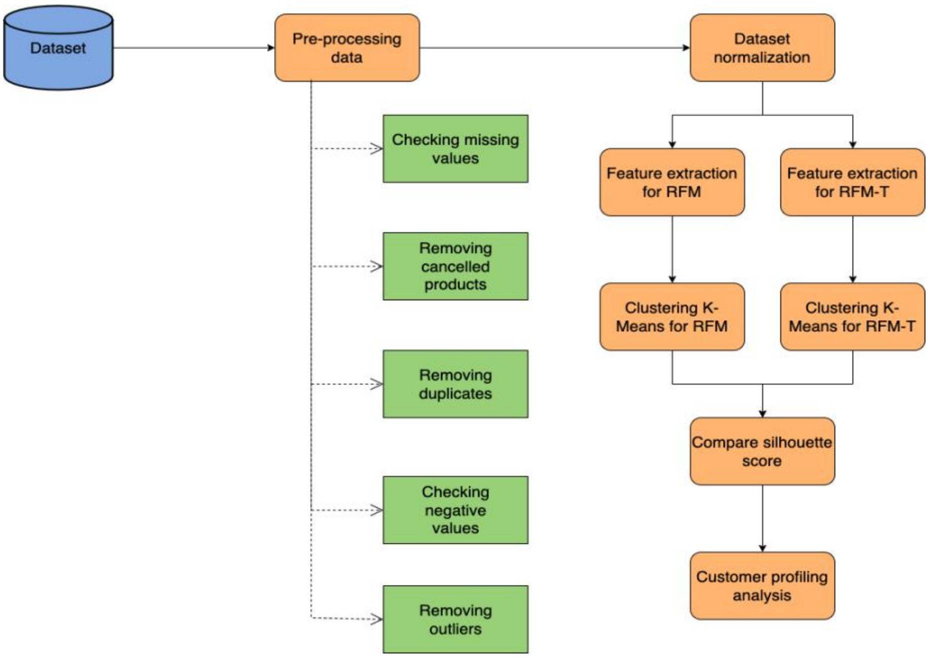

Figure 1. Proposed customer segmentation framework.

A. Dataset

This research utilized the US based online retail dataset encompassing one-year transactions recorded from 01/01/2019 to 31/12/2019 contained 52,555 entries across 12 attributes. The company specializes in selling electronic, office and apparel products. They also offer coupons for the customers to use on each product they bought. On the dataset, customer age and location are provided for better customer profiling.

TABLE I. DATASET ATTRIBUTES DESCRIPTION

No	Attribute	Description
1	Customer ID	Unique identifier for each customer
2	Gender	Customer's gender
3	Location	Customer's location
4	Invoice ID	Unique identifier for each invoice
5	Date	The date of the transaction
6	Product ID	Unique identifier for each product
7	Product Description	A brief description of the product
8	Product Category	Categorization of the product
9	Quantity	The quantity of the product purchased in a transaction
10	Price	The price of the product per unit
11	Coupon Status	Indicates whether a coupon was applied

B. Data Pre-processing

From the dataset, it is evident that the data contained in 53,555 entries requires data processing. The main goal of this step is data pre-processing, which removes erroneous data that may affect the analysis and final segmentation. Data with null stock codes, transactions with negative values, and lines without customer numbers are impacted by this step.

Upon analysis, the dataset exhibits missing values for certain Customer IDs which need to be removed. After removal, the data undergoes a removal of outliers in quantity and price, as visualized in Figure 2 and Figure 3.

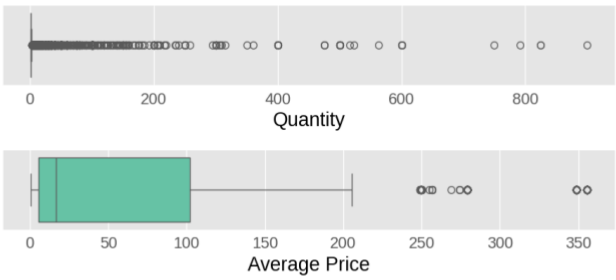

Figure 2. Before the outliers were removed.

Figure 3. After the outliers were removed.

C. RFM and RFM-T Model Score

When it comes to segmenting and analyzing potential customers, the RFM model is highly used. It is a model that primarily analyzes customer behavior with regard to transactions and purchases before making a database forecast. This model consists of three measures: monetary, frequency and recency, which are combined into the concept of RFM [17].

In this study, Time variable is added to analyze its effect on improving the performance of RFM in customer segmentation. According to the suggested RFM-T model, Time is calculated by adding up and averaging the number of days between consecutive transactions. This means it takes into account not just how often and how much a customer spends, but also how quickly they repeat purchases. The "Time" component focuses on measuring the time interval between successive transactions for each customer. It captures their purchasing rhythm and allows for identifying customers who tend to buy frequently within a short timeframe or those with longer intervals between purchases [17].

To build the model, each variable's score needs to be calculated first. Here, let's examine the meanings of R, F, M and T:

- Recency Score: The number of days that have passed between the customer's most recent transaction and their last purchase made within the analysis period. The defined last purchase date in this study is 01/06/2011. The recurring visits from satisfied customers are indicative from a modest value of recency.
- Frequency Score: The frequency throughout the study time indicates how many visits the consumer made. As long as frequency has a high value, he is regarded as loyal.
- Monetary Score: This metric indicates the total expenditure by a customer over the test dataset period. A higher monetary score suggests that a customer is likely more satisfied with the store, as they are willing to spend more. Equation (1) illustrates the calculation of the Monetary Score, where Q represents quantities multiplied by the price per unit (P).

$$M = \sum_{i=1}^{n} \left(Q_i \, x \, P_i \right) \quad (1)$$

- Monetary Time Score: The sum of days between all consecutive transactions made by a customer is calculated. This sum is then divided by the total number of transactions to obtain the average time interval between purchases. Equation (2) details how the Time Score is calculated, where L represents the shopping cycle, obtained by summing transaction date gaps (Ti). Then, L is divided by the number of frequencies (F) minus one, considering only transactions with a frequency greater than 1.

$$T = \frac{L}{F-1} = \frac{\sum_{i=2}^{n} \left(T_i + T_{i-1} \right)}{F-1} \quad (2)$$

Upon determining the values of R, F, M and T as indicated in Table 2. The analysis reveals significant variance in the range of variables. Since K-Means depend on distance, it is necessary to modify the common range in order to prevent creating biased models.

TABLE II. R, F, M AND T VARIABLE STATISTICS

	Recency	Frequency	Monetary	Time
count	1468.00000	1468.00000	1468.00000	1468.00000
mean	145.292234	2.185286	2546.60983	19.502725
std	101.936959	2.235245	3641.88249	30.139134
min	1.000000	1.000000	1.000000	0.000000
25%	56.000000	1.000000	561.403750	0.000000
50%	132.000000	1.500000	1468.15500	0.000000
75%	221.000000	3.000000	3311.85750	36.000000
max	365.000000	34.000000	42433.2500	175.000000

D. *Normalization*

Table 2 shows that the variables have different scales that can lead to issues in the algorithm later. Especially, K-Means involves distance calculations, hence features with larger scales could disproportionately influence the results. Furthermore, when the skewness of the data is checked, Figure 4 proves that the variables are right-skewed. This distribution is identified from the long tail that extends to the right or positive side of the x-axis.

For normalization, this study used the Quantile Transformer method. It is a technique used to transform the probability distribution of a dataset into a specific known distribution or a uniform distribution. It is particularly useful when the data has a skewed distribution. After normalization, the variables are distributed normally as shown in Figure 5.

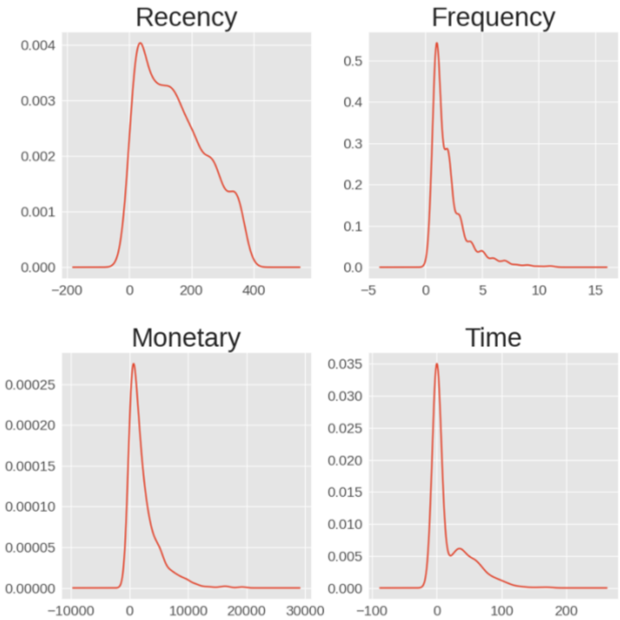

Figure 4. R, F, M and T skewness.

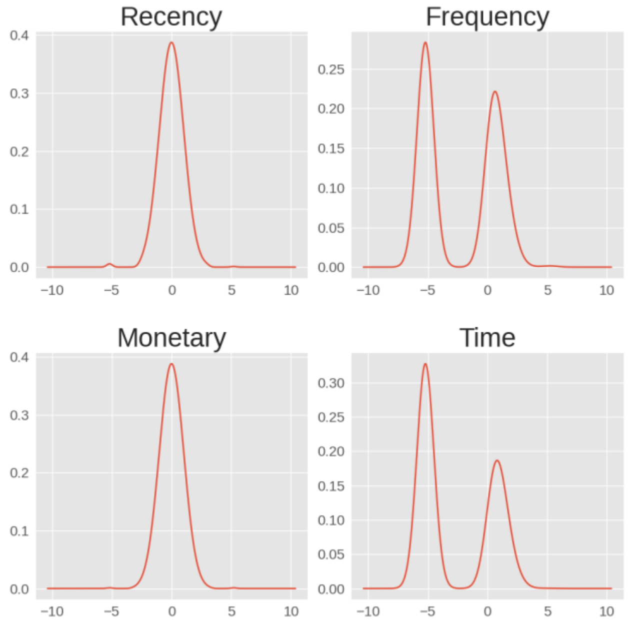

Figure 5. R, F, M and T skewness after normalization.

E. Clustering K-Means

K-Means used as the algorithm for clustering in this study. A few advantages K-Means assures convergence, scales huge data sets, is quite easy to construct, and is easily adaptable to new instances [17].

K-Means determines how many clusters to form by calculating the Euclidean Distance. Euclidean Distance as a method to calculate the distance between data points and centroids. The formula for Euclidean Distance shown as on Equation (3). In this case, xi denotes the ith point in the dataset, k stands for K cluster center, and µk for the kth center. This calculation will be assisted by implementing the Scikit-Learn library.

$$d = \sum_{k=1}^{K} \sum_{i=1}^{n} (x_i - \mu_k)^2 \quad (3)$$

The computer calculates the centroid value before each iteration step in the K-Means algorithm. The Euclidean Distance metric is employed to determine the cluster with the closest centroid for each data point. The iterative process continues until the results of the cluster obtained are comparable to the results of the previous iteration, the procedure as mentioned earlier will be repeated.

The potential of the algorithm to recognize inherent patterns in the dataset is demonstrated by the examination of the K-means clustering findings in RFM and RFM-T. The resulting clusters show distinct separation, suggesting that the data contains significant patterns of groups of customers. It is crucial to recognize that the selection of K classifications has carefully assessed how many clusters are acceptable given the specifics of their study.

F. Elbow Curve

The "Elbow Method" is used to calculate the number of clusters. In order to determine the ideal number of clusters for RFM and RFM-T segmentation, the Elbow Curve technique is thought to be the most accurate and efficient approach. The graph's slope is used to determine how many clusters to create [14]. Using the square of the length between each cluster's centroid and data points, the Elbow Curve rule generates a range of potential values for K. The sum of squared errors (SSE) is used as an indicator to gauge performance, known as distortion or inertia score (Equation 4). Clusters converge as long as the SSE values are low. As the number of clusters approaches the optimal amount, the SSE exhibits a sharp decrease. The SSE declines, although very slowly, if the ideal number of clusters is surpassed [10].

$$inertia = \sum_{i=0}^{n} \min(\|x_i - \mu_j\|^2) \quad (4)$$

The inertia of a cluster is a measure of how far the data points within the cluster are from the centroid (center) of that cluster. The formula for inertia in the context of K-means clustering is the sum of squared distances between each data point in a cluster and its centroid, summed over all clusters. Mathematically, the inertia is calculated as in Equation (4), the term μj represents the mean of a cluster's sample, denoting the average value of the data points within a specific cluster. The symbol xi represents each unique data point inside a cluster, while the number n denotes the total number of clusters in the dataset. Therefore, to provide a representative measure for that specific set of data points, μj is computed as the mean of all xi values within the corresponding cluster.

G. Silhouette Score

The optimal cluster number discovered using the Elbow technique was verified using the silhouette methodology. It quantifies an object's cohesion—how similar it is to its own cluster—as opposed to separation—how similar it is to other clusters. A high silhouette score means that the item is well matched to its own cluster and poorly matched to neighboring clusters. The silhouette score goes from -1 to 1. For a given number of clusters, the silhouette score is used to evaluate the quality of clustering.

4. Results and Discussion

The retail business may implement more targeted marketing tactics to particular customer groups for improved retention by analyzing the features of each grouped cluster and comparing the quality and performance of the RFM and RFM-T models to determine which is the best. In the sections that follow, those results are addressed.

A. Time as Extended Variable

The variables R, F, M and T of the model underwent normalization using Quantile Transformers to address any asymmetry in their values. Analysis based on RFM-T variables revealed a significant correlation. In Figure 6, the heatmap of R, F, M and T variables exhibited a strong positive correlation between the Time and Frequency variables (0.86). Additionally, a moderate positive correlation (0.40) was observed between the Time and Frequency variables. Thus, these findings underscore the importance of considering variable Time when analyzing customer behavior in the context of RFM-based segmentation.

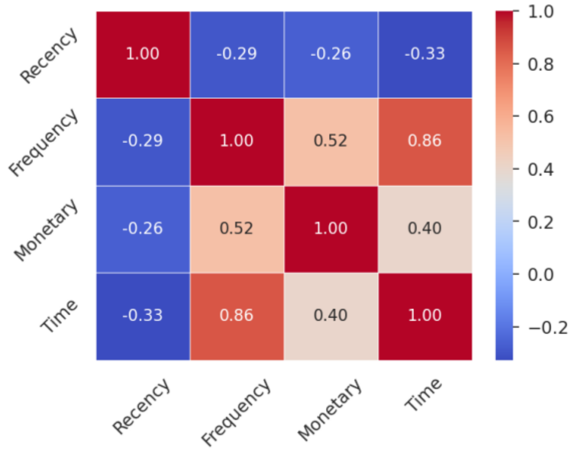

Figure 6. R, F, M and T correlations.

B. *RFM vs RFM-T*

The purpose is to compare the segmentation based on the two models and provide evidence that enhancing the quality of the clustering requires using the variable "Time" in the customer segmentation.

Both the RFM and RFM-T models are used in the application of the K-Means method to carry out the segmentation. The Elbow Curve method is used to segment data according to the two models in order to determine the ideal number of clusters. As Figure 7 and Figure 8 illustrate, the segmentation using the RFM model yields five clusters, while the segmentation using the RFM-T model creates three distinct clusters.

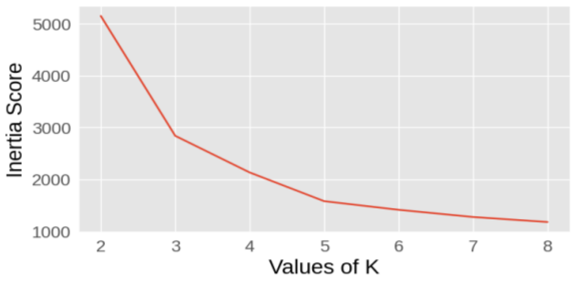

Figure 7. RFM Elbow Curve

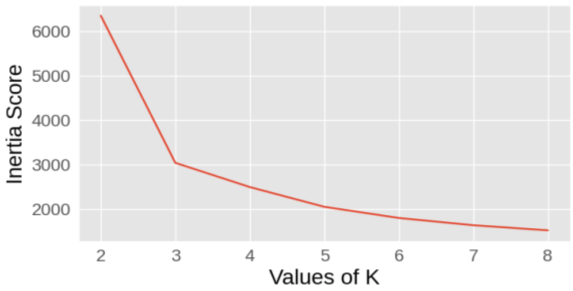

Figure 8. RFM-T Elbow Curve

When analyzing the score on Table 3, RFM is leading by its lower inertia value of 1.584 and faster fit time by only 0.486. While inertia value and fit time of RFM-T are 3.043 and 1.556 respectively. However, upon analyzing the Silhouette Score, RFM achieves a score of 0.3347, while RFM-T attains a score of 0.7096. This suggests that RFM-T has better quality and high performance in clustering than the RFM model alone without T.

TABLE III. K-MEANS AND SILHOUETTE SCORE ANALYSIS

Model	Clusters	Inertia	Fit Time(s)	Silhouette Score
RFM	5	1.584	0.486	0.3347
RFM-T	3	3.042	1.556	**0.7096**

Furthermore, Figure 9 of RFM Monetary-Frequency-Recency and Figure 10 RFM-T Monetary-Time-Frequency demonstrate that RFM-T clusters exhibit a much clearer separation between clusters compared to RFM.

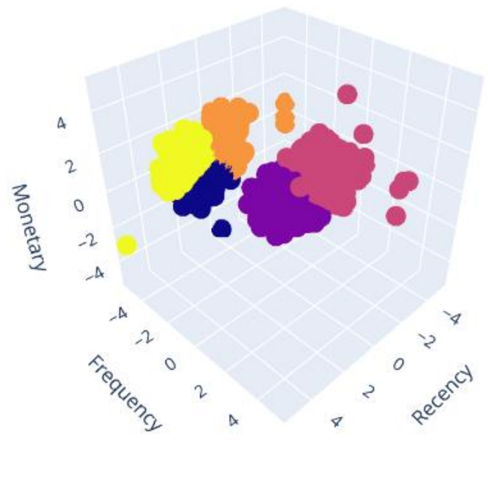

Figure 9. RFM cluster visualization

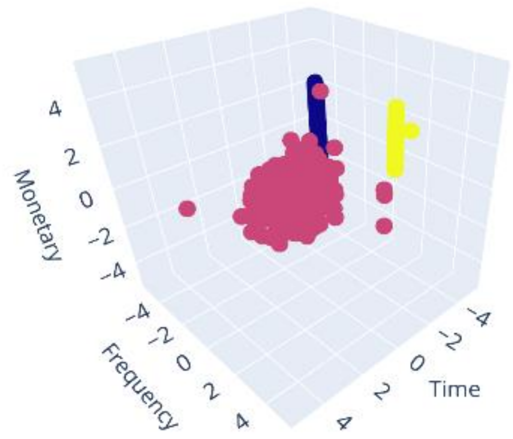

Figure 10. RFM-T cluster visualization

C. *RFM-T Customer Profiling Analysis*

From the previous section, it's evident that the RFM-T clustering model outperforms the RFM model. So, this study will continue the customer profiling on RFM-T model, as a result there will be 3 customer segments across the data. When counted on Table 4, Cluster 0 has the highest number of customers, then Cluster 2 and Cluster 3 came second and third respectively.

TABLE IV. NUMBER OF CUSTOMERS IN EACH CLUSTER

	Count	Percentage (%)
Cluster 0	734	50.3 %
Cluster 1	608	41.67 %
Cluster 2	117	8.02 %
Total	1.459	100%

The distribution of R, F, M and T variables for each cluster is clearly visualized in Figure 11. Customers belonging to Cluster 0 exhibit the highest frequency scores ranging from 0 to 4, visit the store frequently based on their inter-purchase time scores and spend monetarily at a mid to high level. Similarly, customers in Cluster 2 have moderate frequency and monetary scores, but their visits to the store are less frequent due to a low inter-purchase time score. In contrast, Cluster 1 has low scores for all variables, indicating that customers in Cluster 1 visit the store only once.

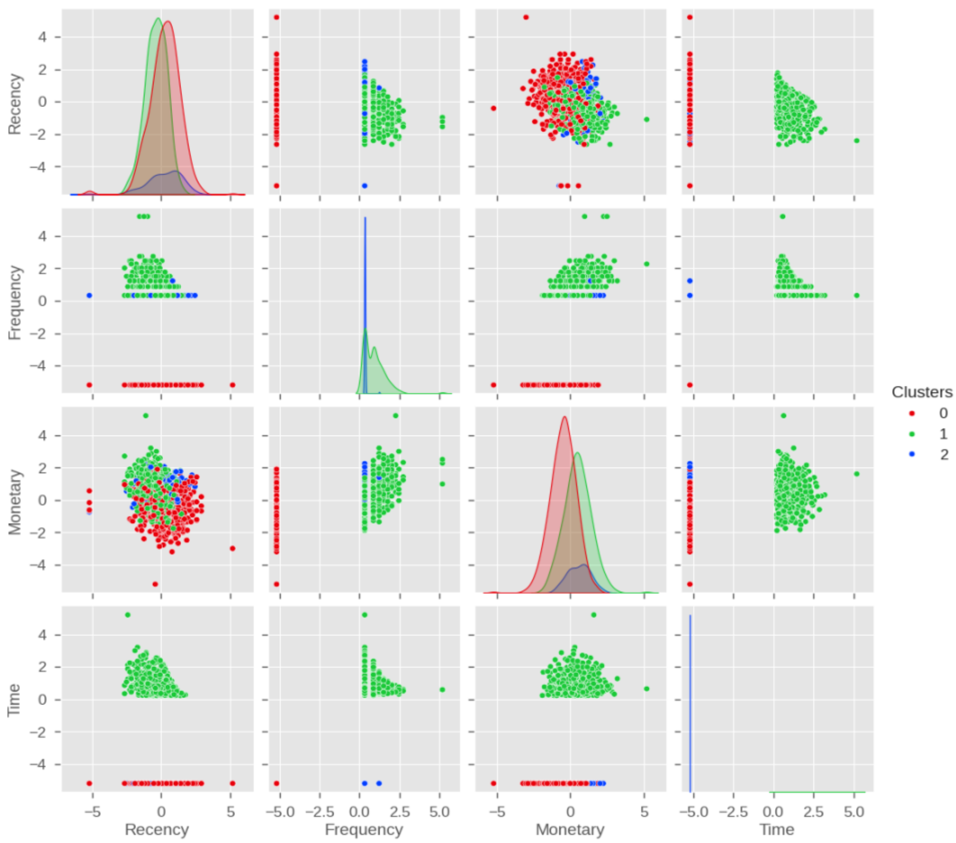

Figure 11. Cluster R, F, M and T distribution.

Customer identities within each cluster exhibit similar distributions, as illustrated in Table 5. The majority of each cluster is composed of female customers, primarily originating from Chicago or New York. Examining Figure 12 reveals distinct purchase behaviors for each cluster. In Cluster 0, customers tend to buy a large quantity and variety of products but with low value. They frequently use coupons and predominantly visit the store on weekends. Cluster 1, on the other hand, consists of customers who typically purchase a small quantity of products but with high value. These customers are high spenders, as indicated by their average monthly spending and rarely use coupons, often visiting the store on weekdays. Lastly, Cluster 2 customers exhibit purchase behaviors similar to those in Cluster 0. They prefer using coupons, resulting in lower average transaction values and monthly spending compared to Cluster 1. However, their product diversity and quantity are considered low, and they tend to shop on weekdays.

Here are some suggestions based on valuable segments profiles:

- **Cluster 0 is defined as Loyal Customers** by Author because they are regular patrons of the store and have lower spending habits. This group requires maintenance through loyalty promotions and is an ideal target for bulk coupons, given their preference for purchasing high quantities of products at low prices.
- **Cluster 1 is defined as Thrifters** by Author because they are infrequent visitors who only come to the store occasionally for specific high-value items. This group is interested in niche luxury items, making them suitable targets for new trending products in the market.
- **Cluster 2 is defined as Attention-Needed Customers** by Author because this group has high

potential to become regular customers like those in Cluster 0, but additional effort is required to increase their frequency, quantities and diversity of purchased products. Similar to Cluster 0, they appreciate coupons, so offering them flash sale promotions on a variety of products on weekdays might be more effective.

TABLE V. RFM-T CUSTOMER IDENTITIES ON EACH CLUSTER

Cluster	Gender		Location			
	Female	*Male*	*New York*	*Chicago*	*New Jersey*	*Washington DC*
Cluster 0	61.5%	38.5%	32.3%	44.2%	14.4%	9.1%
Cluster 1	65.3%	34.7%	31.9%	45.7%	16.3%	6.2%
Cluster 2	63.2%	36.8%	36.6%	48.8%	8.5%	6.1%

Figure 12. RFM-T cluster behavior.

5. CONCLUSION

This study enhanced the traditional RFM (Recency, Frequency and Monetary) model by introducing the Time (T) variable, resulting in the RFM-T model for more precise customer segmentation. Utilizing a US-based online retail dataset, we conducted pre-processing techniques to ensure data balance and extracted RFM-T values. Employing K-Means clustering and the Elbow Curve method, to determine the optimal value of K. Subsequently, evaluate the Silhouette Scores of RFM and RFM-T, revealing a significant improvement in the new model, with scores of 0.3347 and 0.7096 respectively (as shown in Table 3). For customer profiling, demographic information was applied to each RFM-T cluster, allowing for insights into marketing strategies to enhance customer relationships. Specifically, Cluster 0: Lower spenders requiring maintenance and bulk coupons for high quantity, low-price purchases. Cluster 1: Infrequent visitors interested in niche luxury items, ideal for new products. Cluster 2: Potential regulars needing increased purchases, respond to coupons and flash sales.

ACKNOWLEDGMENT

The authors would like to extend their appreciation to the individuals who contributed to the online-shopping-dataset, as well as to Jackson Divakar R for his valuable contributions. They also acknowledge the reviewers and peers for their insightful feedback, and express gratitude to the research participants who indirectly supported this study through Kaggle comments.

REFERENCES

[1] D. Bratina and A. Faganel, “Using Supervised Machine Learning Methods for Rfm Segmentation: a Casino Direct Marketing Communication Case,” *Market-Trziste*, vol. 35, no. 1, pp. 7–22, Jun. 2023, doi: 10.22598/mt/2023.35.1.7.

[2] S. H. Shihab, S. Afroge, and S. Z. Mishu, *RFM based market segmentation approach using advanced k-means and agglomerative clustering: a comparative study*. IEEE, 2019.

[3] J. R. Bult and T. Wansbeek, “Optimal Selection for Direct Mail. Marketing Science,” vol. 14, no. 4, pp. 378–394, 1995.

[4] J. Zhou, J. Wei, and B. Xu, “Customer segmentation by web content mining,” 2021, doi: 10.1016/j.jretconser.2021.102588.

[5] D. Vakratsas and F. M. Bass, “The relationship between purchase regularity and propensity to accelerate,” 2002.

[6] X. He and C. Li, “The Research and Application of Customer Segmentation on E-commerce Websites,” 2016, doi: 10.1109/ICDH.2016.47.

[7] A. M. Hughes, “Boosting response with RFM. Mark,” vol. 5, pp. 4–10, 1994.

[8] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, “RFM ranking – An effective approach to customer segmentation,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, Dec. 2021, doi: 10.1016/j.jksuci.2018.09.004.

[9] A. Ullah *et al.*, “Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation Using Recency, Frequency, Monetary, and Time,” *Sensors*, vol. 23, no. 6, Mar. 2023, doi: 10.3390/s23063180.

[10] M. Y. Smaili and H. Hachimi, “New RFM-D classification model for improving customer analysis and response prediction,” *Ain Shams Engineering Journal*, vol. 14, no. 12, Dec. 2023, doi: 10.1016/j.asej.2023.102254.

[11] S. Idowu and O. Idowu, “Customer Segmentation Based on RFM Model Using K-Means, Hierarchical and Fuzzy C-Means Clustering Algorithms”, doi: 10.13140/RG.2.2.15379.71201.

[12] C. Subbalakshmi, G. Rama Krishna, S. Krishna Mohan Rao, and P. Venketeswa Rao, “A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set,” in *Procedia Computer Science*, Elsevier B.V., 2015, pp. 346–353. doi: 10.1016/j.procs.2015.02.030.

[13] J. T. Wei, S. Y. Lin, Y. Z. Yang, and H. H. Wu, “Applying data mining and RFM model to analyze customers’ values of a veterinary hospital,” in *Proceedings - 2016 IEEE International Symposium on Computer, Consumer and Control, IS3C 2016*, Institute of Electrical and Electronics Engineers Inc., Aug. 2016, pp. 481–484. doi: 10.1109/IS3C.2016.126.

[14] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1757-899X/336/1/012017.

[15] M. Aliyev, E. Ahmadov, H. Gadirli, A. Mammadova, and E. Alasgarov, “Segmenting

E-mail:author’s email

Bank Customers via RFM Model and Unsupervised Machine Learning.”

[16] M. Shutaywi and N. N. Kachouie, “Silhouette analysis for performance evaluation in machine learning with applications to clustering,” *Entropy*, vol. 23, no. 6, Jun. 2021, doi: 10.3390/e23060759.

[17] A. Amine, B. Bouikhalene, R. J. I. J. o. C. Lbibb, and I. Engineering, “Customer segmentation model in e-commerce using clustering techniques and LRFM model: The case of online stores in Morocco,” vol. 9, no. 8, pp. 1993–2003, 2015.

Astrid Dewi Rana is engaged in her undergraduate and Master's studies in Computer Science at Bina Nusantara University since 2020 via the Master Track programme. She has been an IT Business Analyst at a national company since 2023, and has been working in the technology industry since 2021. She aims to focus her studies on developing innovative machine learning and deep learning models for customer segmentation. She can be contacted at email: astrid.rana@binus.ac.id.

Quezvanya Chloe Milano Hadisantoso is currently pursuing her undergraduate and Master's degrees in Computer Science at Bina Nusantara University since 2020 through the Master Track programme. In 2023, she worked as an IT Consultant at a national company. She aims to focus her studies on developing innovative machine learning and deep learning models for customer segmentation. She can be reached at email: quezvanya.hadisantoso@binus.ac.id.

Abba Suganda Girsang is currently lecturer at master information technology at Bina Nusantara University Jakarta. He obtained Ph.D. degree in the Institute of Computer and Communication Engineering, Department of Electrical Engineering and National Cheng Kung University, Tainan, Taiwan, in 2014. He graduated bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta Indonesia, in 2000. He then continued his master’s degree in the Department of Computer Science in the same university in 2006–2008. He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and also worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003-2015. He also taught some subjects at some universities in 2006–2008. His research interests include swarm intelligence, combinatorial optimization, and decision support system. He can be contacted at email: agirsang@binus.edu.