Volume 9, Issue 10, October– 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

Cluster Analysis of Furniture for Export Customers Position at UD HK Jepara

Achmad Fawwaz Bahaudin
Teknik Manajemen Industri, Universitas Diponegoro
Kota Semarang, Jawa Tengah 50275

Abstract:- A furniture production company needs to categorize its customers, especially those who are involved in export. The company has analyzed export shipment records from June 6, 2015, to April 23, 2022. The analysis used the Recency, Frequency, and Monetary (RFM) variables, which have been widely used in the field of marketing. The k-means clustering algorithm was employed for the analysis, resulting in the division of customers into three clusters. Cluster 1 comprises customers with the highest and most valuable purchases, including those with the codes CL, FR, NC, and RE. Cluster 2 includes customers who excel in one of the RFM variables, with codes AN, AR, BN, IN, IT, KE, KR, LK, MU, MY, SA, SC, SM, TW, and UN. Customers in Cluster 3 have the fewest and least valuable purchases, including those with the codes GP, KN, NL, OM, PT, and TZ.

Keywords:- *Analysis, K-Means Clustering, RFM Model, Recency, Frequency, Monetary.*

I. INTRODUCTION

Furniture encompasses household items like tables, chairs, cabinets, and more. It is a crucial element that enhances a home, making it a comfortable place to live. Furniture transforms a house into a cozy space for both relaxation and work. Nowadays, there are various types of furniture available, made from metal, plastic, or wood. Wooden furniture is a prominent product of Jepara Regency, Central Java Province.

UD HK Jepara is a furniture production company located in Jepara Regency. The company caters to both local and international customers. Since 2015, UD HK Jepara has exported over 400 containers to various countries. Their primary customers are retail companies in different countries who order furniture as needed for restocking. As a result, furniture orders are not placed on a regular schedule. Additionally, UD HK Jepara is not the exclusive supplier for these companies. Therefore, it is crucial for the company to retain its customers, especially those who have remained loyal in purchasing their products. Customer satisfaction is crucial for building customer loyalty. The company deems it necessary to provide extra services to VIP customers to prevent them from switching to other companies.

When categorizing customers, it is essential to segment them based on their frequency of furniture purchases from the company. Numerous studies have been conducted to identify the most effective models and approaches for customer categorization. Previous research has shown that the Recency, Frequency, Monetary (RFM) model is one of the most widely utilized models, offering a high level of accuracy. This model assesses customer behavior by considering the time of the customer's last transaction (Recency), the number of transactions (Frequency), and the amount of money spent (Monetary) [1].

In a study conducted by Syakur et al. [2],the profiles of the top customers in SMEs were identified using K-Means Clustering. The K-Means algorithm, known for its simplicity compared to other clustering algorithms, offers the advantages of easy implementation and execution, speed, and adaptability [3]. This study successfully grouped customers into 8 clusters. Additionally, Khajvand & Tarokh [4] conducted research on estimating customer segment value by applying the RFM Model with the K-Means Clustering algorithm, resulting in the division of customers into 4 clusters.

In this paper, the data grouping involves clustering customer order data using the variables of Recency, Frequency, and Monetary (RFM). The method used is Cluster Analysis with the K-Means algorithm. Customers will be grouped into 3 clusters.

II. LITERATURE REVIEW

This chapter contains the theoretical foundation or references that are briefly outlined and will later be used as the supporting basis for addressing the problem to be researched.

A. Multivariate Analysis

Multivariate analysis is a statistical method used to analyze data that involves multiple independent and dependent variables. It is generally divided into two methods: the dependent method and the independent method. The dependent method is utilized to analyze the relationship between variables and to explain or predict the dependent variable based on two or more independent variables. On the other hand, the independent method is used to group variables based on specific criteria and is categorized into factor analysis, cluster analysis, and multidimensional scaling.

B. Factor Analysis

Factor analysis is a statistical technique used to explore and understand the relationships among a large set of variables. Its main goal is to simplify the information within many variables by reducing them into a smaller set of latent factors, which represent the core structure of the data. This method

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

helps identify key variables and summarize complex data, making it easier to understand the relationships between variables without losing important information.

C. Multicollinearity Test

The multicollinearity test is used to check for any correlation between independent variables in a regression model. Multicollinearity suggests that there is a perfect linear relationship between some or all of the explanatory variables in the regression model [5].

- *The Decision-making Criteria for the Multicollinearity Test are Outlined by Ghozali [6].*

- If the VIF value is < 10 or the Tolerance value is > 0.01, it is concluded that multicollinearity does not occur.
- If the VIF value is > 10 or the Tolerance value is < 0.01, it is concluded that multicollinearity occurs.
- If the correlation coefficient between each independent variable is > 0.8, multicollinearity occurs. However, if the correlation coefficient between each independent variable is < 0.8, multicollinearity does not occur.

D. KMO & Bartlett's Test Analysis

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy is a statistic used to determine if a dataset is suitable for factor analysis. It compares the strength of the observed correlations with the strength of the partial correlations. For factor analysis to be valid, the KMO measure must be greater than 0.50.

Bartlett's Test of Sphericity is a statistical test used to assess the interdependence between variables that are indicators of a factor. The purpose of this analysis is to determine if the variables being examined are uncorrelated in the population. In order to conduct factor analysis, the significance in Bartlett's test should be less than 0.05 [7].

E. Cluster Analysis

Clustering involves creating groupings where all members of each group share similarities based on a specific criteria. Cluster analysis is a data analysis technique used to group individuals or objects into clusters with distinct characteristics. This ensures that individuals or objects within the same cluster have relatively homogeneous traits.

F. K-Means Clustering

The K-Means method was first introduced by MacQueen JB in 1976. K-Means is a non-hierarchical data clustering method that aims to partition existing data into two or more groups. This method partitions data into groups so that data with similar characteristics are placed into the same group, while data with different characteristics are grouped into another.

III. METODHOLOGY

The research method used in this study is quantitative research, where the analysis is conducted based on the results of methods that produce numerical output.

The research stages refer to a series of activities conducted in an orderly and systematic manner. The stages of this research are as follows:

A. Field Study

This stage involves conducting field observations to assess current conditions and understand the problems encountered. The focus is on observing the export furniture ordering process at the company.

B. Literature Review

In this stage, theories and knowledge related to the existing problems are searched for and studied to support finding solutions. It includes studying relevant literature, journals, and information related to the research problem.

C. Problem Identification

At this stage, the issues occurring in the company, particularly UD HK Jepara, are identified to discover the underlying problems within the company.

D. Problem Formulation

This stage results from the problem identification process. The research topic and identified problems are used as a basis for formulating the research problem, which becomes the focus of the study.

E. Determining Research Objectives

The next step involves defining the research objectives, crucial for establishing benchmarks to measure the success of the research.

F. Data Collection

Data collection is carried out to gather information that supports the research. The type of data used is secondary data, including documentation submissions related to V-Legal for exporters' needs.

G. Data Standardization

The collected data needs to be standardized to ensure uniformity, especially when the input format is inconsistent, so that all data adheres to a specific standard.

H. Non-Multicollinearity and Representativeness of Sample Assumption Tests

Before clustering, the data must pass cluster analysis assumption tests. The first required test is the Non-Multicollinearity Assumption Test by examining VIF values and tolerance. The second test is the sample representativeness assumption test by reviewing the KMO and .Sig values.

I. K-Means Clustering Implementation

Clustering will be set with a maximum of 10 iterations, and the data will be grouped into 3 clusters.

J. Conclusion and Recommendations

This final stage contains conclusions derived from the analysis and discussion, aimed at answering the research objectives set earlier. Recommendations are also provided based on the findings.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

IV. RESULT AND DISCUSSION

This chapter explains the general overview of the company, the data collection required for the research, data processing using the methods described in the previous chapter, as well as the analysis and discussion of the research conducted.

A. Data Collection

The research utilized V-Legal document submission records from UD HK Jepara spanning from June 6, 2015, to April 23, 2022. The collected data includes the frequency of V-Legal document submissions, representing customer purchase frequency, the latest submission date, and the export value. This Data can be seen in Table 1.

This data was then used to calculate Recency, Frequency, and Monetary (RFM) values. Recency was computed by subtracting the last submission date from the research start date (May 1, 2022) in months. Frequency was determined by tallying the number of shipments, while Monetary was based on the total export value.

Table 1 RFM Data Export

No	Code	Recency	Frequency	Monetary
1	AN	4	24	Rp 97,700,000
2	AR	5	2	Rp 98,550,000
3	BN	44	1	Rp 82,450,000
4	CL	1	56	Rp 96,978,000
5	FR	1	76	Rp 105,330,000
6	GP	51	1	Rp 15,000,000
7	IN	69	3	Rp 107,000,000
8	IT	65	1	Rp 70,975,000
9	KE	50	2	Rp 116,720,000
10	KN	50	1	Rp 20,000,000
11	KR	5	4	Rp 102,750,000
12	LK	5	2	Rp 100,550,000
13	MU	6	21	Rp 99,300,000
14	MY	13	3	Rp 98,822,000
15	NC	1	100	Rp 102,690,000
16	NL	4	3	Rp 47,000,000
17	OM	10	16	Rp 50,250,000
18	PT	8	2	Rp 46,000,000
19	RE	1	120	Rp 90,650,000
20	SA	29	22	Rp 100,572,500
21	SC	21	23	Rp 100,772,500
22	SM	5	18	Rp 99,200,000
23	TW	30	2	Rp 100,930,000
24	TZ	50	1	Rp 17,350,000
25	UN	14	16	Rp 97,500,000

B. Standardization Cluster Analysis

The information about the standardization process of Recency, Frequency, and Monetary (RFM) data obtained from the company's shipment records. The process was carried out using IBM SPSS software. The first step involved defining the variables to be used, which include customer code, recency, frequency, and monetary value.

The results from the data processing in IBM SPSS indicate that the recency variable has a minimum value of 1 and a maximum value of 69, with a mean of 21.68 and a standard deviation of 22.50504. The frequency variable has a minimum value of 1, a maximum value of 120, and a mean of 20.80, with a standard deviation of 32.55124.

Regarding the monetary variable, the minimum and maximum values are Rp 15,000,000.00 and Rp 116,720,000.00 respectively. The mean and standard deviation for the monetary variable are Rp 82,601,600.00 and Rp 30,771,674.64 respectively. Z-scores for each attribute were also calculated and can be found in Table 2, in addition to the minimum, maximum, mean, and standard deviation values.

Table 2 Z-Score Value

No	Code	Zscore Recency	Zscore Frequency	Zscore Monetary
1	AN	-,78560	,09831	,49066
2	AR	-,74117	-,57755	,51828
3	BN	,99178	-,60827	-,00493
4	CL	-,91891	1,08137	,46720
5	FR	-,91891	1,69579	,73861

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

No	Code	Zscore Recency	Zscore Frequency	Zscore Monetary
6	GP	1,30282	-,60827	-2,19688
7	IN	2,10264	-,54683	,79289
8	IT	1,92490	-,60827	-,37783
9	KE	1,25838	-,57755	1,10876
10	KN	1,25838	-,60827	-2,03439
11	KR	-,74117	-,51611	,65477
12	LK	-,74117	-,57755	,58328
13	MU	-,69673	,00614	,54265
14	MY	-,38569	-,54683	,52712
15	NC	-,91891	2,43309	,65282
16	NL	-,78560	-,54683	-1,15696
17	OM	-,51899	-,14746	-1,05134
18	PT	-,60786	-,57755	-1,18946
19	RE	-,91891	3,04750	,26155
20	SA	,32526	,03686	,58401
21	SC	-,03022	,06759	,59051
22	SM	-,74117	-,08602	,53941
23	TW	,36970	-,57755	,59563
24	TZ	1,25838	-,60827	-2,12051
25	UN	-,34126	-,14746	,48416

C. Non-Multicollinearity Tests

When examining a regression model for multicollinearity, it's important to analyze the Tolerance value and the Variance Inflation Factor (VIF). The Tolerance value indicates the amount of variability in a particular independent variable that is not explained by the other independent variables. This analysis was performed using IBM SPSS software, and the findings are detailed in Table 3.

Table 3 Multicollinearity Test Result

	Colineary Statistics	
Mode	***Tolerance***	**VIF**
Dependent Variable : Z-score recency		
Zscore(Frequency)	,914	1,094
Zscore(Monetary)	,914	1,094
Dependent Variable : Z-score frequency		
Zscore(Frequency)	,890	1,124
Zscore(Monetary)	,890	1,124
Dependent Variable : Z-score monetary		
Zscore(Frequency)	,773	1,293
Zscore(Monetary)	,773	1,293

Since all three VIF values are < 10 and the Tolerance values are > 0.01, it can be concluded that multicollinearity does not occur.

D. Representativeness of Sample Assumption Tests

To determine whether a variable is correlated with other variables, Bartlett's Test of Sphericity is conducted. The results of the test can be seen in Table 4.

Table 4 KMO and Bartlett's Test Result

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,623
Bartlett's Test of Sphericity	*Approx. Chi-Square*	8,882
	df	3
	Sig.	,031

In Table 4, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy is 0.623. Since the KMO value is greater than 0.5, it indicates that factor analysis is appropriate. Additionally, the Sig. value is 0.31, which is smaller than the alpha value of 0.05. This means the correlation between variables is high, and the factor analysis process can proceed.

V. CLUSTER ANALYSIS USING K-MEANS

After ensuring that there is no multicollinearity and that the sample is representative, you can use IBM SPSS software to perform a Cluster Analysis using the K-Means method. The iteration was limited to a maximum of 10 cycles. The results indicate that customers can be divided into three clusters:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

- Cluster 1 includes customers who have made the highest number and value of purchases.
- Cluster 2 includes customers who excel in one specific variable.
- Cluster 3 includes customers who have made the fewest and lowest-value purchases.

Customers with the following codes CL, FR, NC, and RE fall into Cluster 1. Customers with codes AN, AR, BN, IN, IT, KE, KR, LK, MU, MY, SA, SC, SM, TW, and UN are placed in Cluster 2. Meanwhile, customers with codes GP, KN, NL, OM, PT, and TZ are categorized in Cluster 3. The results of the clustering process obtained using IBM SPSS can be seen in Table 5.

VI. CONCLUSION

Based on the research conducted, the conclusions that can be drawn from this study are as follows.

The study utilized Recency, Frequency, and Monetary (RFM) variables derived from V-Legal document submission data for container shipments. Recency was calculated by subtracting the last shipping date from the research start date, frequency represents the number of shipments made, and monetary signifies the average value of the shipments.

Before applying K-means clustering to categorize UD HK Jepara's customers, the recency, frequency, and monetary data from the company needed to be standardized. After standardization, non-multicollinearity assumption tests and KMO and Bartlett's Test were performed. Once these assumptions were met, the clustering process was conducted, resulting in customers being grouped into 3 clusters.

It was identified that customers labeled CL, FR, NC, and RE require more attention from the company.

Table 5 K-means Cluster Result

Cluster	**CODE_BUYER**	***Distance***
	CL	,985
1	FR	,424
	NC	,389
	RE	1,019
	AN	1,006
	AR	,890
	BN	1,047
	IN	2,015
	IT	2,030
	KE	1,310
	KR	,888
2	LK	,893
	MU	,887
	MY	,543
	SA	,440
	SC	,445
	SM	,898
	TW	,354
	UN	,500
	GP	1,143
	KN	1,030
	NL	1,199
3	OM	1,079
	PT	1,025
	TZ	1,067

RECOMMENDATIONS

Based on the previous conclusions, the following recommendations can be made from this research:

Maintain Loyalty of Cluster 1 Customers: Customers in Cluster 1 should be carefully nurtured to ensure their continued loyalty. These customers contribute the most in terms of purchase frequency and value, so maintaining strong relationships with them is crucial for sustaining the company's revenue.

Follow-up with Cluster 3 Customers: Efforts should be made to follow up with customers in Cluster 3, who have made fewer and lower-value purchases. By engaging them more actively, there may be opportunities to increase their purchasing frequency and value.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24OCT731

REFERENCES

[1]. A. T. Widyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Global)," KONSTELASI: Konvergensi Teknologi dan Sistem Informasi, pp. 204-215, 2021.

[2]. M. A. Syakur, K. B. K., E. M. S. Rochman and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of the Best Customer Profile Cluster," IOP Conf. Series: Materials Science and Engineering 336, 2017.

[3]. D. Grossman and O. Frieder, Information Retrieval Algorithms and Heuristics Second Edition, The Netherlands: Springer, 2004.

[4]. M. Khajvand and M. J. Tarokh, "Estimating Customer Future Value of Different Customer Segments Base on Adapted RFM Model in Retail Banking Context," Procedia Computer Science, pp. 1327-1332, 2011.

[5]. S. R. d. Ajija, Cara Cerdas Menguasai Eviews, Jakarta: Salemba Empat, 2011.

[6]. I. Ghozali, Aplikasi Analisis Multivariate dengan Program IBM SPSS 23 Edisi 8, Semarang: Penerbit Universitas Diponegoro, 2016.

[7]. S. Santoso, Analisis SPSS pada Statistik Parametrik, Jakarta: PT Elex Media Komputindo, 2012.

[8]. S. Santoso, Mahir Statistik Multivariat Dengan SPSS, Jakarta: PT Elek Media Komputindo, 2014.