

Implementation of Data Mining for Analyzing Consumer Purchasing Patterns at TeTa Ino Café

Theresia Elvita Tjia^{1*}, Fajar Novriansyah Yasir², Shindy Ekawati³

^{1,2,3}Cokroaminoto Palopo University, Indonesia

¹theresiaelvita1004@gmail.com, ²fajarnovriansyah@uncp.ac.id, ³shindy.ekawati@uncp.ac.id



*Corresponding Author

Article History:

Submitted: 29-07-2025

Accepted: 07-08-2025

Published: 14-08-2025

Keywords:

Clustering; CRISP-DM; K-Means Algorithm; Marketing Strategy; Purchase Pattern.

Brilliance: Research of

Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

TeTa Ino is a micro, small, and medium enterprise by university students focusing on the production and marketing of innovative products based on butterfly pea tea. The significant decrease in sales volume from approximately 50–80 units per promotional event to only 20–40 units indicates a potential issue in the current marketing strategy. Therefore, this study aims to identify consumer purchasing patterns that can serve as the foundation for developing more targeted marketing strategies to enhance the competitiveness of TeTa Ino. This research employs the Cross Industry Standard Process for Data Mining (CRISP-DM) approach by applying the K-Means Clustering algorithm to four months of transaction data, including variables such as number of transactions, total transaction value, and discounts offered. The analysis resulted in four distinct consumer clusters: passive consumers, loyal consumers, non-loyal consumers, and consumers with moderate purchasing frequency. Each cluster is recommended to be approached with tailored marketing strategies, such as loyalty programs, product benefit education, and bundling promotions. The clustering evaluation achieved a Silhouette Score of 0.9008 and a Calinski Harabasz Score of 7630.34, indicating good segmentation quality and clear separation among clusters. This study concludes that applying the K-Means Clustering algorithm is effective in mapping consumer purchasing behavior as a basis for data-driven marketing strategy formulation. Future research is recommended to incorporate time-related variables and explore other clustering methods to further strengthen the analysis.

INTRODUCTION

The rapid advancement of information technology and the intensifying competition among micro, small, and medium enterprises have driven business actors to continuously innovate to sustain their existence and increase sales performance. TeTa Ino is by university students focusing on the production and marketing of innovative beverage products based on butterfly pea flowers. Based on historical sales data, in the first two months, the sales volume of butterfly pea tea at TeTa Ino reached between 50 and 80 units per promotional event. However, a significant decline was recorded in the subsequent two months, with sales dropping to only 20–40 units, which directly affected revenue stability and business continuity.

In line with the increasingly complex consumer behavior patterns, several previous studies emphasize the importance of utilizing data mining methods, particularly the K-Means

Clustering algorithm, to identify and analyze consumer purchasing patterns. For instance, the research conducted by Mulyadi (Mulyadi, 2022) demonstrated the effectiveness of K-Means Clustering in improving marketing strategies for Tupperware products. Similarly, the study by Astuti (Astuti, 2019) which adopted the Cross Industry Standard Process for Data Mining (CRISP-DM) framework, successfully identified consumer segmentation for micro, small, and medium enterprises more accurately. Moreover, there is also research that highlights the important role of understanding consumer behavior patterns as a basis for developing data driven marketing strategies (Noperahila, 2024).

Based on these considerations, this study is grounded in the hypothesis that applying the K-Means Clustering algorithm to TeTa Ino's transaction data can facilitate more accurate consumer segmentation, thereby supporting the formulation of more effective marketing strategies. The research questions addressed in this study are as follows: (1) What are the consumer purchasing patterns at TeTa Ino based on the collected transaction data? and (2) What marketing strategies are suitable to increase the sales of butterfly pea tea products amid the competitive micro, small, and medium enterprise food and beverage sector?.

Accordingly, the primary objective of this research is to identify consumer purchasing patterns through the application of the K-Means Clustering algorithm and to formulate recommendations for marketing strategies based on the clustering results to enhance the competitiveness and sales stability of TeTa Ino.



LITERATURE REVIEW

The rapid growth of data-driven business practices has led to the increasing adoption of data mining techniques to identify consumer behavior patterns, particularly in the micro, small, and medium enterprise sector. Data mining is defined as a systematic process of discovering meaningful patterns from large datasets, serving as a foundation for strategic decision-making (Wahono, 2023). Among various algorithms, *K-Means Clustering* has gained considerable attention due to its simplicity and effectiveness in grouping data based on similarity, thus enabling businesses to develop targeted marketing strategies.

Several studies have demonstrated the applicability of K-Means Clustering in improving business outcomes. For instance, Mulyadi (Mulyadi, 2022) utilized K-Means to analyze consumer purchase behavior, resulting in more precise promotional strategies that boosted product sales. Similarly, Astuti (Astuti, 2019), offering actionable insights for market positioning. In another study, Noperahila and Sulianta (Noperahila & Sulianta, 2024) highlighted the importance of understanding consumer shopping patterns as the basis for crafting data-driven marketing strategies, reinforcing the value of segmentation analysis in dynamic market contexts. Furthermore, Safitri (Safitri, 2024) emphasized that accurate segmentation could help businesses design tailored promotions and loyalty programs that better match consumer needs and preferences.

While these studies provide compelling evidence of the usefulness of K-Means Clustering, they also present several limitations that create opportunities for further research. Most existing studies have primarily focused on a single clustering algorithm without exploring alternative methods such as hierarchical clustering or DBSCAN, which could potentially offer more nuanced insights when dealing with non-linear data distributions. Additionally, prior research has rarely incorporated temporal variables such as purchase timing or seasonal trends—that could reveal deeper insights into consumer buying cycles (Ahmed et al., 2020). These gaps underscore the need for comprehensive studies that combine systematic data mining methodologies with broader data dimensions.

To address these limitations, the present research applies the CRISP-DM framework holistically and uses K-Means Clustering to analyze consumer purchase patterns at TeTa Ino, an micro, small, and medium enterprise initiated by university students. By focusing on variables including transaction count, total purchase value, and discount usage, this study aims to produce practical consumer segmentation that can directly inform the development of data-driven marketing strategies. In doing so, the research contributes to bridging the gap in micro, small, and medium enterprise focused segmentation studies, while also offering practical recommendations for improving sales performance in competitive markets.

Theoretical Foundation

1. Purchase Pattern Analysis

Purchase pattern analysis is the study of consumer behavior that focuses on identifying habits, preferences, and tendencies during the process of purchasing goods or services. This analysis provides a deeper understanding of how consumers interact with products, how frequently they purchase, which products are often bought together, and specific preferences associated with certain demographic segments (Valaskova et al., 2021). Beyond simply understanding consumer behavior, purchase pattern analysis also facilitates the prediction of future trends through various analytical techniques. It is particularly important in the industry for enhancing customer experience and supporting strategic decision-making by identifying recurring trends. By recognizing purchase patterns, businesses can design more effective sales strategies, target specific customer segments, and develop products that better meet customer needs (Ramanto et al., 2024).

In practice, companies implement various strategies and plans to increase the sales of their products or services, aiming to improve their market position and revenue. These are commonly referred to as sales enhancement strategies. Such strategies may include customer service improvements, product development, marketing initiatives, and price optimization, among other approaches (Permata Sari, 2020).

In parallel, data mining plays a critical role in supporting sales enhancement strategies by enabling businesses to systematically analyze consumer purchasing behaviors and market trends. Through uncovering hidden patterns in transaction data, businesses can design targeted promotions, optimize pricing, and tailor product offerings to better meet customer needs. Furthermore, Ramanto (Ramanto et al., 2024) highlights that incorporating predictive analytics based on historical data allows companies to anticipate future purchasing trends, plan inventory more effectively, and implement timely discount strategies. Beyond merely boosting sales figures, these data-driven approaches aim to build long-term customer relationships, foster product loyalty, and create a sustainable competitive advantage by aligning business decisions with actual consumer behavior patterns.

2. CRISP-DM (Cross Industry Standard Process of Data Mining)

CRISP-DM (Cross Industry Standard Process for Data Mining) is a widely adopted standard methodology for data mining projects, consisting of six main phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The Business Understanding phase plays a crucial role in translating business needs into well-defined data mining problems. This phase often involves iterative processes between business objectives and data analysis techniques to ensure that the final outcome aligns with the project's goals. Although time-



consuming, this iterative approach facilitates better management of business requirements (Schroer et al., 2021).

Similarly, the Data Understanding and Data Preparation phases frequently require iterations, especially in complex projects. These phases aim to identify data quality issues, construct relevant datasets, and conduct comprehensive exploratory data analysis to gain deeper insights. In practice, the CRISP-DM methodology has been adapted across diverse fields, ranging from marketing to industrial engineering, due to its flexibility and systematic structure (Plotnikova et al., 2021)

CRISP-DM enables comprehensive data exploration and preprocessing, supports the selection of appropriate models, and includes evaluation steps to ensure that the developed model meets the defined objectives before deployment. Its structured and iterative nature ensures that data mining projects are not only methodologically sound but also aligned with business goals. This flexibility and systematic framework have contributed to CRISP-DM becoming a standard approach for ensuring the success of data mining initiatives in various industries.

3. Google Colab

Google Colab, or Google Colaboratory, is a cloud-based platform that provides an interactive Python programming environment, primarily designed for scientific computing and data analysis, including the implementation of machine learning algorithms such as K-Means Clustering. Built on top of Jupyter Notebook, Google Colab enables users to write and execute Python code directly within a web browser without the need to install additional software (Nur Rohim & Zuliarso, 2022).

One of its key advantages is the ability to access Google's computational resources, including free GPU and TPU support, which facilitates the handling of large datasets and accelerates the training and experimentation of machine learning models (Gelar Guntara, 2023). In addition, Google Colab supports seamless integration with popular Python libraries, such as Scikit-learn for implementing K-Means Clustering, Pandas for data manipulation, and Matplotlib for visualizing analytical results.

Research also shows that Google Colab enables complex data processing and analysis by offering a flexible and cost-effective environment, especially for cloud-based machine learning applications. With these features, Google Colab serves as a valuable tool in data-driven research and machine learning projects by providing an interactive cloud-based environment, real-time collaboration capabilities, free GPU access, and strong integration with various data libraries and storage services.

4. Data Mining

Data mining is a systematic process of collecting, processing, and analyzing large volumes of data to uncover hidden patterns, correlations, or valuable insights that support decision-making. As a key component of the broader discipline known as Knowledge Discovery in Databases (KDD), data mining employs a variety of analytical techniques, including classification, clustering, regression, and association rule analysis (Wahono, 2023). These methods enable researchers and practitioners to transform raw data into meaningful knowledge that can guide strategic and operational decisions.

Over time, data mining has evolved from a purely technical process into a multidisciplinary tool with significant applications across diverse sectors such as business, healthcare, finance, and education. In the business context, for example, data mining helps organizations identify customer segments, predict purchasing trends, optimize marketing campaigns, and improve customer retention strategies. In healthcare, it assists in detecting disease patterns and supporting diagnostic processes, while in education, it can help analyze student performance and identify factors affecting learning outcomes (Sekar Setyaningtyas et al., 2022).

The strength of data mining lies not only in its ability to analyze complex datasets but also in its capacity to produce actionable insights that enhance innovation and efficiency. By uncovering hidden patterns and trends, data mining enables organizations to make data-driven decisions, anticipate market changes, and maintain competitive advantage. In this research, data mining specifically through K-Means clustering serves as the foundation for segmenting consumer purchasing behavior, supporting the development of targeted marketing strategies tailored to diverse customer segments. Moreover, data mining helps businesses understand customer preferences and buying habits, which can inform product development and promotional planning. It also contributes to optimizing inventory management by predicting demand fluctuations based on historical trends. Beyond improving sales performance, these insights foster long-term customer loyalty and brand trust. Ultimately, the integration of data mining into business strategy transforms raw data into strategic knowledge that drives sustainable growth.

5. K-Means Clustering

K-Means Clustering is one of the most popular clustering algorithms in data mining, widely recognized for its simplicity and efficiency in partitioning data into clusters based on shared characteristics. The algorithm divides data into K clusters by iteratively updating centroids to minimize the total intra-cluster distance, resulting in compact and well-defined groups. Despite its strengths, K-Means has several limitations, including sensitivity to the initial placement of centroids, which can influence the final clustering outcome, and the requirement to predefine the number of clusters before analysis. To address these challenges, numerous studies have proposed adaptive variants of the K-Means algorithm tailored to different data types and distributions (Ahmed et al., 2020).

K-Means has been successfully applied across a range of domains, such as market analysis, customer



segmentation, and image clustering, due to its ability to reveal hidden patterns in unlabeled datasets through an unsupervised learning approach (Oti et al., 2021). Complementing this, the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology provides a systematic framework for conducting comprehensive data exploration and preparation, selecting appropriate models, and evaluating outcomes to ensure alignment with project objectives before deployment. This structured yet flexible methodology is applicable across various industries, supporting the success of data mining projects by bridging technical analysis with practical business needs (Ahmed et al., 2020).

METHOD

Bussines Understanding

The researcher recognizes that TeTa Ino's product sales face significant challenges in understanding and adapting to the dynamic nature of consumer behavior. These changes include shifts in price sensitivity, purchasing location preferences, payment methods, and consumer responses to promotional discounts. Consequently, it becomes essential to conduct a more in-depth analysis of transaction data to design sales and marketing strategies that are better aligned with actual consumer needs.

This analysis specifically aims to extract valuable insights that can support data-driven business decision-making. Key areas of focus include identifying which TeTa Ino product variants are most popular among consumers, determining the sales locations with the highest transaction volumes, examining the most frequently used payment methods, and analyzing the impact of discounts on the total transaction value. To achieve this, the study employs the K-Means Clustering algorithm to segment consumers based on their transaction characteristics. The insights derived from this segmentation are expected to inform the development of targeted marketing strategies that ultimately enhance sales performance at TeTa Ino Cafe.

Data Understanding

This study utilizes detailed sales transaction data from TeTa Ino Cafe, comprising a total of 685 individual transactions collected over a four-month observation period. Each record in the dataset represents a single sales transaction and includes critical information such as the quantity of products purchased, the total transaction value, and the discounts applied. This comprehensive dataset serves as the empirical foundation for conducting an in-depth analysis to better understand consumer purchasing behavior and support the design of targeted marketing strategies.

The dataset is structured around several key attributes: transaction code, transaction date, sales location, payment method, product variant, unit price, quantity, subtotal, discount, and final transaction amount. These attributes collectively allow the researcher to conduct exploratory data analysis and apply clustering techniques to identify distinct consumer segments, purchasing habits, and emerging trends that may not be immediately visible through simple descriptive analysis.

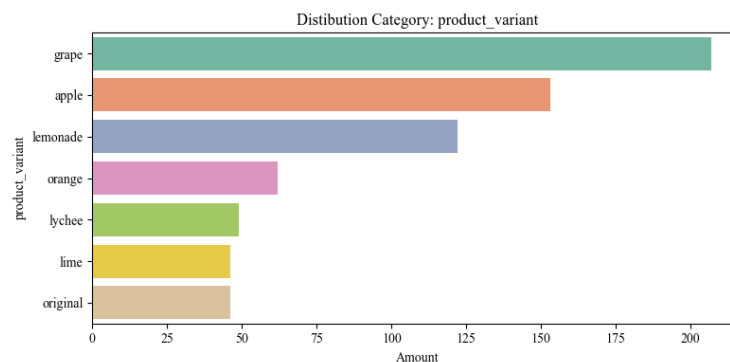


Figure 1. Product Variant Distribution Category

Preliminary analysis of the dataset revealed several noteworthy insights. The grape variant emerged as the most popular product, recording 207 transactions during the observation period, whereas the original and lime variants were less preferred, with only 46 transactions each. Furthermore, cash was identified as the overwhelmingly dominant payment method, accounting for approximately 95.3% of all transactions. From the perspective of sales location, the courtyard of UNCP Campus was the most active point of sale, generating 257 transactions, highlighting its strategic importance as a sales hub. These initial findings not only illustrate consumer preferences and behavioral patterns but also establish a critical starting point for applying the K-Means clustering algorithm. Through further analysis, the study aims to transform these preliminary observations into actionable insights that can inform data-driven marketing strategies, improve sales performance, and strengthen customer engagement at TeTa Ino Cafe.



Figure 2. Sales Location Distribution Category

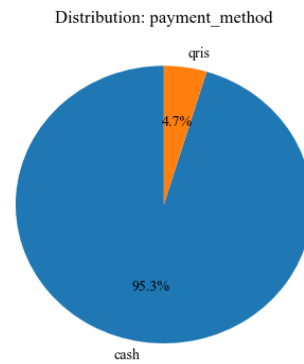


Figure 3. Payment Method Distribution

In the data understanding phase, the researcher also conducted an analysis of the correlations between numerical features. Understanding these correlations is important as it can support subsequent modeling and evaluation processes by highlighting relationships that may affect clustering outcomes. Based on the correlation matrix visualization, it was observed that the features quantity, subtotal, and total transaction value exhibit a strong positive correlation. This indicates that as the number of items purchased increases, the total transaction value also tends to rise proportionally.

Conversely, the price feature shows a relatively weak correlation with other variables, suggesting that price variations within the dataset do not significantly influence the total transaction value. These findings imply that purchase volume plays a more substantial role in determining transaction totals than product price differences. This insight is valuable in guiding the feature selection process for clustering, as it highlights which attributes most effectively capture consumer purchasing behavior.

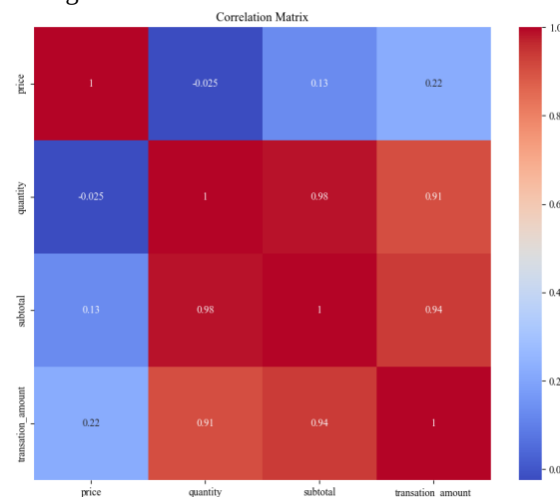


Figure 4. Correlation Matrix

Data Preparation

1. Feature Selection

The first step in the data preparation process involves selecting the features to be used for clustering. In this study, the researcher chose three primary features: subtotal, quantity, and discount. By combining these three features, the clustering process is expected to generate more meaningful and representative consumer segments that reflect actual purchasing behavior.

The resulting clusters are anticipated to serve as a strong foundation for formulating more targeted and effective marketing and sales strategies, ultimately aimed at improving the sales performance and competitiveness of TeTa Ino Café.

discount	quantity	subtotal
0	1	10000
0	1	10000
0,2	2	20000
0	1	10000
0,2	2	20000

Figure 5. Selected Columns: Discount, Quantity and Subtotal

2. Transformation Data

Data transformation was performed because the discount feature was originally stored as an object or categorical data type. Since clustering algorithms can only process numerical data, it was necessary to convert the values in the discount feature into a numerical format, specifically floating point numbers. This transformation step ensures that the discount feature can be included in the clustering process. Consequently, transforming the data type of this feature is a critical step to ensure that the resulting clusters are valid, reliable, and accurately reflect consumer purchasing patterns.

discount	quantity	subtotal
0.0	1	10000
0.0	1	10000
2.0	2	20000
0.0	1	10000
2.0	2	20000

Figure 6. Transformation Data: Selected Columns

3. Data Standardization

Data standardization was performed to prevent features with larger numeric ranges from disproportionately influencing the determination of cluster centroids. Standardizing the data transforms all feature values into a uniform scale, ensuring that each feature contributes equally to the clustering process. By applying this step, the clustering algorithm can evaluate all features objectively, which is expected to produce more accurate and unbiased consumer segments.

	discount	quantity	subtotal
0	-1.666023	-1.400590	-1.400590
1	-1.666023	-1.400590	-1.400590
2	-0.498258	-0.804456	-0.804456
3	-1.666023	-1.400590	-1.400590
4	-0.498258	-0.804456	-0.804456
..	...	...	...
636	-0.498258	-0.804456	-0.804456
637	1.253388	1.580082	1.580082
638	1.253388	0.983947	0.983947
639	0.085624	0.387813	0.387813
640	1.253388	1.580082	1.580082

Figure 7. Standardization Data: Transformation Data

4. Outlier Removal

In this study, the researcher identified 44 outlier data points, representing approximately 6.3% of the total dataset. Since this proportion is below the commonly accepted threshold of 10%, the removal of these 44 outliers was considered reasonable to improve data quality. After outlier removal, the dataset size was reduced from 685 records to 641 records.

The detection of outliers was performed using the Interquartile Range (IQR) method, which is widely used to identify extreme values in a dataset. Applying the IQR technique helps stabilize the data distribution, which is particularly beneficial for machine learning algorithms like K-Means Clustering used in this research. Removing outliers ensures that the clustering process produces more reliable and meaningful segments that accurately reflect typical consumer purchasing behavior.

Modeling

An examination of the Elbow Method graph reveals a pronounced decrease in the Within-Cluster Sum of Squares (WCSS) values from cluster counts 1 to 4. Beyond this point, the decline in WCSS becomes noticeably more gradual and less substantial. This trend suggests that adding more clusters beyond four does not significantly reduce intra-cluster variation. Accordingly, it can be concluded that the “elbow point” occurs at cluster number 4, indicating that the optimal number of clusters for this study is four.

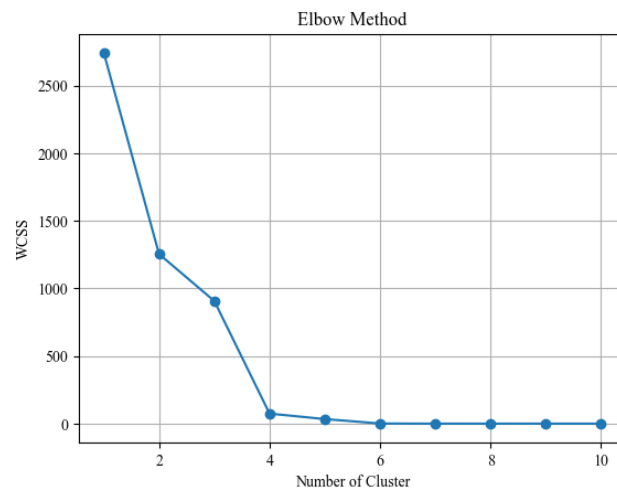


Figure 8. Elbow Method

In the K-Means algorithm, the distance between each data point in the dataset and all cluster centroids is calculated. Each data point is then assigned to the cluster whose centroid is closest to it. This process aims to minimize the Within-Cluster Sum of Squares (WCSS), thereby ensuring that each cluster becomes as compact and internally homogeneous as possible. In this study, the distance between data points and centroids was computed using the Euclidean Distance method, which is one of the most widely used distance metrics and is particularly well-suited for numerical data.

```
1.38270624 4.77915303 0.37537801
1.38270624 4.77915303 0.37537801
0.43302887 3.41983657 1.38270624
..
3.10323088 0.39639397 4.44431797
1.73172429 1.71657051 3.10249237
3.77362411 0.44817744 5.14443395
```

Figure 9. Eucliden Distance

The visualization below illustrates the results of the K-Means clustering process, which segmented the dataset into four distinct clusters using three primary features: the X-axis represents quantity, the Y-axis represents subtotal, and the Z-axis represents discount. The colored dots in the plot correspond to individual data points, with each color indicating a different cluster: purple for cluster 0, green for cluster 1, blue for cluster 2, and yellow for cluster 3. The large black “X” symbols mark the centroids of each cluster.

This three-dimensional visualization provides an intuitive view of how the data points are distributed and grouped based on the selected features. The following section explains the characteristics and business interpretation of each identified cluster in more detail.

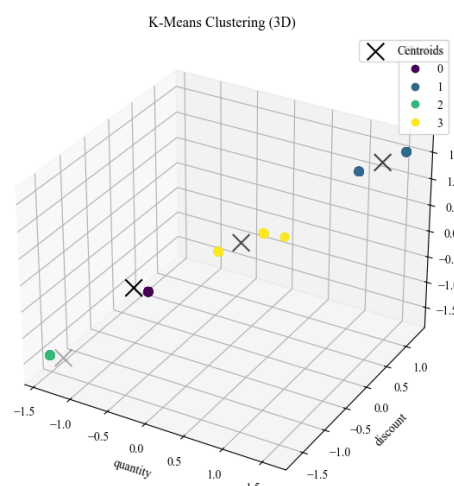


Figure 10. K-Means Clustering Visualization

Evaluation

As part of the evaluation process, the Silhouette Score was employed as a metric to assess the performance of the K-Means clustering algorithm. The analysis yielded a Silhouette Score of 0.9008, which signifies that data points are positioned very close to the centroid of their respective clusters while remaining relatively distant from the centroids of other clusters. This result indicates that the clusters exhibit high internal cohesion and clear separation from one another. A Silhouette Score exceeding 0.70 is generally regarded as excellent in the context of K-Means clustering, thereby confirming that the clustering solution produced in this research demonstrates strong internal consistency and distinct inter-cluster boundaries.

Silhouette Score: 0.9008

Figure 11. Silhouette Score

RESULT

Deployment

In this stage, the researcher presents the findings obtained after applying the K-Means Clustering algorithm to the sales transaction dataset collected from TeTa Ino Cafe. This analysis aims to explore and categorize consumer purchasing patterns by identifying meaningful segments within the data. The clustering process was conducted using three primary features: quantity of products purchased, subtotal transaction value, and discount received. By segmenting consumers based on these features, the study seeks to reveal differences in purchasing frequency, spending behavior, and responsiveness to discounts. As a result, four distinct consumer clusters were identified, each reflecting unique behavioral characteristics and transaction patterns within the observed period. The descriptive results for each cluster are outlined below to provide a clearer understanding of the consumer segments uncovered through this analysis.

1. Cluster 0 represents consumers with relatively high purchasing activity, where the standardized values for quantity and total transaction features range approximately from 0.6 to 1.0. However, consumers in this cluster tend to receive few or only small discounts, with standardized discount feature values between -0.8 and -0.2.
2. Cluster 1 shows the most active consumers who also receive discounts most frequently. The standardized values for data points in this cluster are generally higher, ranging from above 1.1 to 1.5 for all three features: quantity, total transaction, and discount.
3. Cluster 1 shows the most active consumers who also receive discounts most frequently. The standardized values for data points in this cluster are generally higher, ranging from above 1.1 to 1.5 for all three features: quantity, total transaction, and discount.
4. Cluster 3 consists of active consumers, although not as active as those in Cluster 1, who are still fairly responsive to discounts. Data points in this cluster have standardized values for quantity and total transactions around 0.7 to 1.0, and discount values ranging from approximately 0.7 to 0.9.

transaction_code	transaction_date	sales_location	payment_method	product_variant	price	discount	quantity	subtotal	Cluster
A001	2024-07-15	uncp campus	cash	lemonade	10000	-1.666023	-1.400590	-1.400590	2
A002	2024-07-15	uncp campus	cash	grape	10000	-1.666023	-1.400590	-1.400590	2
A003	2024-07-15	uncp campus	cash	grape	10000	-0.498258	-0.804456	-0.804456	0
A004	2024-07-15	uncp campus	cash	lime	10000	-1.666023	-1.400590	-1.400590	2
A005	2024-07-15	uncp campus	cash	grape	10000	-0.498258	-0.804456	-0.804456	0
A005	2024-07-15	uncp campus	cash	lemonade	10000	1.253388	1.580082	1.580082	1
A006	2024-07-15	uncp campus	cash	apple	10000	0.085624	-0.208322	-0.208322	3
A007	2024-07-15	uncp campus	cash	apple	10000	-0.498258	-0.804456	-0.804456	0
A008	2024-07-15	uncp campus	cash	grape	10000	-0.498258	-0.804456	-0.804456	0
A009	2024-07-15	uncp campus	cash	lemonade	10000	1.253388	0.983947	0.983947	1
A010	2024-07-15	uncp campus	cash	original	7000	-1.666023	-1.400590	-1.400590	2
A010	2024-07-15	uncp campus	cash	orange	10000	0.085624	-0.208322	-0.208322	3
A011	2024-07-15	uncp campus	cash	orange	10000	1.253388	1.580082	1.580082	1
A012	2024-07-15	uncp campus	cash	grape	10000	-1.666023	-1.400590	-1.400590	2
A012	2024-07-15	uncp campus	cash	apple	10000	0.085624	0.387813	0.387813	3
A012	2024-07-15	uncp campus	cash	orange	10000	-0.498258	-0.804456	-0.804456	0
A012	2024-07-15	uncp campus	cash	lime	10000	0.085624	0.387813	0.387813	3
A013	2024-07-15	uncp campus	cash	lemonade	10000	-1.666023	-1.400590	-1.400590	2
A014	2024-07-15	uncp campus	cash	grape	10000	-1.666023	-1.400590	-1.400590	2
A015	2024-07-15	uncp campus	cash	original	7000	-0.498258	-0.804456	-0.804456	0

Figure 12. Clustering Result

In addition to mapping these cluster characteristics, the modeling results also provide several insights that can support targeted sales improvement strategies for TeTa Ino Cafe:

1. Consumers in Cluster 0, who tend to be passive with relatively low transaction amounts, may benefit from educational content highlighting the health benefits of butterfly pea flower to stimulate greater interest and increase purchases.

2. Consumers in Cluster 1, who show high transaction volumes and frequent engagement, could be retained through loyalty programs, such as offering reward points for transactions above a certain amount. This program may be particularly effective when focused on lower cost payment methods like QRIS to avoid over transactions.
3. Consumers in Cluster 2, who appear less loyal and make only occasional or even single purchases, could be targeted with reactive promotional offers using the best-selling variant, grape. These promotions could be strategically launched at sales locations with lower transaction volumes, such as graduation events, to attract new buyers.
4. Consumers in Cluster 3, who have average transaction volumes, could be encouraged with attractive bundle promotions, such as buy-one-get-one-free offers. For example, consumers could purchase the least popular variant (original) and receive the most popular variant (grape) for free.

DISCUSSION

This study applied the K-Means Clustering algorithm to analyze consumer purchasing behavior at TeTa Ino Café, resulting in the identification of four distinct consumer clusters. The clustering process, which used three key features quantity, subtotal, and discount successfully revealed meaningful patterns that offer practical insights for developing targeted marketing strategies.

A closer examination of the results shows that Cluster 1 represents the most active consumers who frequently benefit from discounts, suggesting a strong responsiveness to promotional offers and a potential for further engagement through loyalty programs. In contrast, Cluster 2 includes passive or non-loyal consumers who rarely purchase and seldom receive discounts. This highlights an opportunity for targeted reactivation campaigns, for example by introducing promotional incentives or bundling offers focused on best-selling variants like grape. Meanwhile, Cluster 0 and Cluster 3 show moderate to high purchasing activity with varying sensitivity to discounts, indicating that tailored messaging and educational content emphasizing product benefits might encourage higher purchase frequency.

Compared to the raw transactional data, the clustering results provide a clearer understanding of underlying consumer behaviors that are not immediately apparent from descriptive statistics alone. The use of K-Means also proves effective in simplifying complex transactional data into actionable consumer segments, offering practical value for decision-making despite the known limitations of the algorithm, such as sensitivity to initial centroid selection.

Furthermore, the high Silhouette Score obtained (0.9008) demonstrates strong cohesion within clusters and distinct separation between clusters, confirming the robustness of the model in capturing real consumer differences. These findings align with previous studies suggesting that data-driven segmentation can enhance the effectiveness of marketing strategies, and extend this insight by providing a practical case in the context of small and medium-sized enterprises.

CONCLUSION

This study applied the K-Means Clustering algorithm to analyze consumer purchasing patterns at TeTa Ino Café, resulting in the identification of four distinct consumer clusters based on quantity, subtotal transaction value, and discount received. The clustering process successfully revealed meaningful differences among consumer groups, which can be used to design targeted marketing strategies, such as loyalty programs for highly active consumers, educational content for moderately active consumers, and promotional campaigns to reactivate passive consumers.

The findings highlight the practical benefits of using data-driven segmentation, even for small and medium sized enterprises with relatively limited datasets. By transforming raw transactional data into actionable insights, businesses can better understand consumer behavior and optimize their marketing approaches accordingly.

However, this research also has limitations. The dataset used was relatively small and limited to a four-month observation period, which may not fully capture seasonal or long-term purchasing trends. Additionally, the use of only three features (quantity, subtotal, and discount) may not account for other factors influencing consumer behavior, such as demographic data or product preferences.

For future studies, it is recommended to expand the dataset over a longer time period, incorporate additional features (such as consumer demographics, location-based data, or time of purchase), and explore the use of alternative clustering algorithms or ensemble methods to validate and enhance segmentation accuracy. Overall, this research demonstrates the potential of simple clustering techniques to provide valuable insights for improving sales performance and supporting data.

REFERENCES

- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics (Switzerland)*, 9(8), 1–12. <https://doi.org/10.3390/electronics9081295>
- Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 60–72. <https://doi.org/10.20895/inista.v1i2.71>



- Gelar Guntara, R. (2023). Deteksi Atap Bangunan Berbasis Citra Udara Menggunakan Google Colab dan Algoritma Deep Learning YOLOv7. *Jurnal Manajemen Sistem Informasi (JMASIF)*, 2(1), 9–18. <https://doi.org/10.59431/jmasif.v2i1.156>
- Mulyadi, I. P. (2022). Klasterisasi Menggunakan Metode Algoritma K-Means dalam Meningkatkan Penjualan Tupperware. *Jurnal Informatika Ekonomi Bisnis*, 4, 172–179. <https://doi.org/10.37034/infec.v4i4.164>
- Noperahila, N. R., & Sulianta, F. (2024). Analisis Pola Perilaku Konsumen dan Kebiasaan Belanja menggunakan Metode K-Means Clustering. *July*. <https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and->
- Nur Rohim, & Zuliarso, E. (2022). Penerapan Algoritma Deep Learning Untuk Pengembangan Chatbot Yang Digunakan Untuk Konsultasi Dan Pengenalan Tentang Virus Covid-19. *Pixel : Jurnal Ilmiah Komputer Grafis*, 15(2), 267–278. <https://doi.org/10.51903/pixel.v15i2.777>
- Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>
- Permata Sari, S. (2020). Strategi Meningkatkan Penjualan Di Era Digital. *SCIENTIFIC JOURNAL OF REFLECTION: Economic, Accounting, Management and Business*, 3(3), 291–300. <https://doi.org/10.5281/zenodo.3930698>
- Plotnikova, V., Dumas, M., & Milani, F. (2021). Adapting the CRISP-DM Data Mining Process: A Case Study in the Financial Services Domain. *Lecture Notes in Business Information Processing*, 415 LNBIP, 55–71. https://doi.org/10.1007/978-3-030-75018-3_4
- Ramanto, A., Irma Purnamasari, A., & Ali, I. (2024). Analisis Pola Pembelian Produk Pada Toko Online Daun Indah Menggunakan Association Rule Dan Algoritma Fp-Growth Di Platform E-Commerce Shopee. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 1215–1220. <https://doi.org/10.36040/jati.v8i1.8341>
- Safitri, E. (2024). Implementasi Algoritma K-Means Clustering Dalam Menentukan Strategi Marketing Dalam Penjualan Ikan (Studi Kasus: Grosir Ikan Tani Mas Tanjung Morawa). *JIKTEKS : Jurnal Ilmu Komputer Dan Teknologi Informasi*, 02(02), 23–33.
- Schroer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Valaskova, K., Durana, P., & Adamko, P. (2021). Changes in consumers' purchase patterns as a consequence of the COVID-19 pandemic. *Mathematics*, 9(15), 1–17. <https://doi.org/10.3390/math9151788>
- Wahono, R. S. (2023). Data Mining Data mining. In *Mining of Massive Datasets* (Vol. 2, Issue January 2013). https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part