



Utilizing Data Mining Techniques to Analysis Changes in Purchase Behavior of Batik's Customers

Danang Setiawan^{1*}, Lisa Alfiyani², Joko Sulistio³, Qurtubi⁴

¹Industrial Engineering Department, Faculty of Industrial Technology, Universitas Islam Indonesia, Jl. Kaliurang km 14.5, Sleman, Yogyakarta 55584, Indonesia

*danang.setiawan@uui.ac.id

Abstract. Sales transaction data contains rich information potentially used to support company competitiveness. However, interpreting and utilizing transaction data in developing marketing strategies remains a challenge, even for big companies. Therefore, this research aims to develop marketing strategies using data mining techniques. A medium-sized company focusing on producing and selling traditional motif clothes (batik) will be used as a case study. The negative sales trend is the biggest issue currently faced by the company. Hypothetically, this problem is caused by imported products sold at lower prices or changing consumer behavior after pandemic covid. Currently, the company only implements simple analysis of its transaction data. The analysis of transaction data, conducted through five data mining stages, yielded a shift from purchasing small quantities to larger quantities, increased purchases during the final week of each month, and increased purchases on religious occasions. Furthermore, the analysis revealed that 31.29% of all transactions were attributed to loyal consumers, and 192 customers exhibited in Cluster 1 (high transaction quantities and high transaction values). Further investigation also revealed that customers categorized as loyal customers and Cluster 1 have different behaviors that can be used to develop further customer relationship programs. Future research can be conducted by employing data mining techniques to study the organization's assortment of products. Management discussions reveal that changes in consumer buying behavior extend to the selection of items and batik themes.

Keywords: author guidelines, ASSET, article template

(Received 2024-02-27, Accepted 2024-03-09, Available Online by 2024-04-19)

1. Introduction

The retail industry contains abundant transaction data that provide both opportunities and challenges. Transaction data includes essential information that can provide valuable insights into customer preferences and behavior to support the company's competitiveness and customer satisfaction [1]. However, getting useful information from the transaction data is a challenge even for big companies because the previously unstructured data must be processed first [2]. In addition, organizational, cultural, and technological barriers hinder data mining and business intelligence tools in the retail industry [3].

For this reason, up until now, not all industries have successfully managed their transaction data to improve the company's competitiveness or increase customer satisfaction.

Previous research implements different data mining techniques in other cases of utilizing transaction data. Transaction data representing consumer shopping behavior can be used to formulate promotional strategies [1] [4]. Samuel and Gwendolyn [5] studied by integrating Market Basket Analysis (MBA) and 4P strategy (product, price, place, and promotion) in the retail industry. Panjaitan et al. [6] utilized the MBA to develop promotional strategies in product bundling. In this study, data mining techniques will be implemented to analyze transaction data, where the output of analyzing data will then be used to determine the marketing strategy.

This study used a medium-sized company that produces and sells Indonesian traditional motif clothes called *batik* as a case study. The company faces tremendous challenges due to changing consumer purchase behavior and abundant imported products that offer clothes with almost the same motif but far lower prices. Previous research also stated that after the COVID-19 pandemic, the way consumers fulfilled their needs changed significantly, where the clothing industry was the industry that received the most significant effect of the COVID-19 [7]. The pandemic has accelerated the shift towards e-commerce, with customers increasingly relying on online shopping and e-payment methods [8]. Eger et al. [9] stated that retailers and suppliers must successfully consider the new knowledge of consumer behavior and apply it to their selling strategy.

Nevertheless, it is hypothetic that the observed company has loyal customers because the company has been established for over ten years. Therefore, one of the goals of this study is to identify loyal customers by customer segmentation and finding recurring buying behavior. The result of this study, combined with a discussion with the management team, will then be used to determine the marketing strategy.

Several research studies focused on analyzing the behavior of batik consumers in purchasing. Rahadi et al. [10] researched the purchasing behavior of batik customers using a questionnaire. This study found that consumers prefer online and word-of-mouth promotion and select batik designs that combine traditional and modern motifs [10]. Ardiansyah and Febrianti [11] researched customer purchase decisions for batik products using path analysis or regression analysis. The study found that customer engagement influences brand attachment, while brand attachment also influences the purchase decision [11]. The research mentioned above uses direct observation techniques to gain insight into consumer shopping behavior.

In addition to research, research is also conducted using data mining techniques to evaluate consumer shopping patterns. Research by Mulyawan et al. [12] uses clustering techniques using k-means clustering to analyze the motives that consumers are interested in. Clustering techniques are also used by research by Salbinda et al. [13] to group the clothes that customers most demand. Meanwhile, Mardalius and Christy's [14] used clustering techniques to group consumer areas based on sales transaction value.

Based on the abovementioned research, no research comprehensively uses data mining techniques to analyze the sales of batik companies. Therefore, this study aims to identify loyal customers using Pareto, churn analysis, and clustering. The loyal customer identified will be the focus of the company's customer engagement program. This research aims to contribute to the knowledge of the batik business in Indonesia by specifically examining sales transaction data through data mining techniques. The objective is to enhance the company's competitiveness and improve customer satisfaction.

2. Methods

2.1. Data mining technique

Data mining is a technique used to transform unstructured data into valuable knowledge. Data mining can be classified into five techniques: (1) classification, (2) association, (3) prediction, (4) clustering, and (5) outlier analysis [2]. These techniques are advantageous for analyzing large datasets and have been applied across healthcare, finance, and telecommunications. Classification is a method of

classifying data into predefined classes that will become a model of a training set with known class labels. The association is used to identify a frequently used product from data. Prediction is a method to predict how an attribute in data will behave. It is mainly used to identify a pattern within the data. Clustering is a method to group data objects into clusters or groups. Outlier analysis is a phase in which data has a pattern that does not conform to a well-defined notion of normal behavior.

Data mining consists of a series of processes, starting from business understanding, data understanding, data preparation, modeling, evaluation, and deployment [15] [16]. Business understanding is the phase of determining the objectives of the data mining project from a business point of view. Data understanding is the data collection phase and continues with data preparation, consisting of data selection, cleaning, integration, and transformation. After the data is transformed, the modeling phase occurs, where the data is modeled based on the chosen method and parameters. Evaluation is the phase where the researcher evaluates the model results from the modeling phase by validating the model results. Deployment is the phase where the knowledge is obtained through a pattern deployed for the objectives [15]. The following process is selecting a model or analysis technique and evaluating and deploying the results.

2.2. Research flow

This research will be conducted by following the general steps of the data mining technique. Anshu [15] states that data mining is a process that cannot be completed in a single step. Figure 1 illustrates the data mining process used in this study.

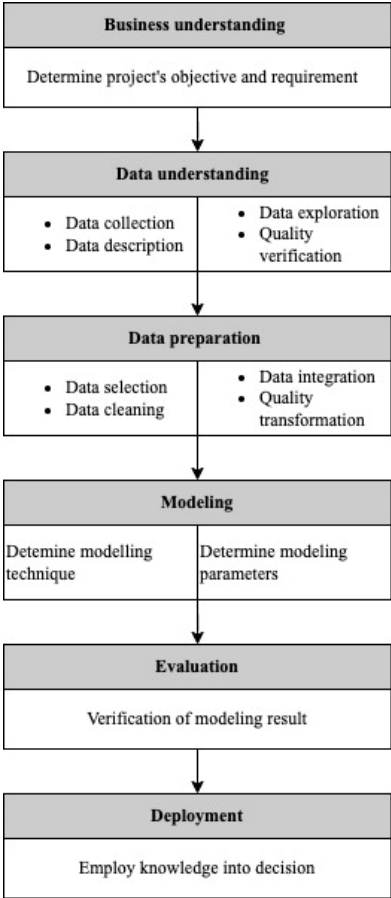


Figure 1. Research step

The description of each step is detailed as follows:

1. Business understanding

Business understanding aims at determining the objectives and requirements of the project. In this study, discussion with the management teams will be conducted to gather the project's goals and needs. Initial meeting obtained that the company's owner wanted to get input for marketing strategies by analyzing current transaction data.

2. Data understanding

Data understanding consists of several activities: data collection, description, exploration, and quality verification. In this study, data collection will be carried out by downloading sales transaction data through web-based information systems owned by the company. Downloaded data will be verified first with the sales team and the company's owner. Because the pandemic happened from the beginning of 2020 until 2022, the focus group discussion with the management team will be conducted to determine whether data during the COVID-19 pandemic will be used in the next step.

3. Data preparation

Data preparation contains activities for constructing the final data set to be modeled in the next step. This phase consists of selecting data, cleaning data, data integration, and data transformation.

4. Modeling

In the modeling phase, the modeling technique and parameters are determined based on the project's objective. The data mining techniques can be classified into classification, clustering, prediction, association rule, neural networks, time series analysis, summarization, and sequence discovery [15].

5. Evaluation

In the evaluation step, the result of the modeling phase is verified in the context of achieving business objective. The decision in this phase will be to continue to the next stage or review the previous step.

6. Deployment

In the deployment phase, knowledge obtained is used to generate decisions. In this study, the decision will be the marketing strategy.

3. Results and Discussion

3.1. Data Mining Step

This chapter is presented based on 6 (six) steps of data mining as mentioned in the chapter methodology. However, due to time limitations, this study will cover the 5 (five) steps as follows:

1. Business understanding

Business understanding is a step to determine the objective of a project. In this study, as mentioned in the introduction chapter, the company currently faces tremendous challenges due to changes in customer behavior after conceived-19 pandemic. However, because the company has already been established for more than ten years, it is hypothetic that the observed company has loyal customers. Therefore, this study aims to understand the customers' purchase behavior and identify the company's loyal customers.

2. Data understanding

Data was collected by downloading the point-of-sales data from the company's information system. The company currently has a web-based platform to manage the transaction data. However, the company has not recorded the customer ID of their customer. Therefore, as seen in Figure 1, there was no standard for inputting the phone number. The data downloaded from the company's information system contain information on each customer's transactions across months in one year. Therefore, this study aims to

understand the purchase pattern and identify the customer with the priority on customer engagement program.

3. Data preparation

The data downloaded consists of 12 sheets that represent monthly transactions. In total, there were 3900 transactions in 12 months. The initial dataset can be seen in Figure 2, whereas the dataset after the preparation step is shown in Figure 3.

The table shows a grid with columns labeled A through AH. Row 1 contains headers for months: T, U, V, W, X, Y, Z, AA, AB, AC, AD, AE, AF, AG, AH. Row 2 contains headers for months: 17/01, 18/01, 19/01, 20/01, 21/01, 22/01, 23/01, 24/01, 25/01, 26/01, 27/01, 28/01, 29/01, 30/01, and Rata-rata Harga. The data rows contain phone numbers in column A, names in column B, and cities in column C. The bottom row shows month abbreviations from Januari to Desember.

Figure 2. Initial dataset

The table shows a grid with columns labeled A through Y. Row 1 contains headers for months: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24. The data rows contain phone numbers in column A, names in column B, and cities in column C. The bottom row shows month abbreviations from January to Desember.

Figure 3. Dataset after preparation step

The data preparation consists of the following steps:

- a. Phone number standardization. The phone number is standardized using the following action:
 - Delete the + (plus), - (minus), and . (dot).
 - Change 62 at the beginning of with 0, where 62 is the country code for Indonesia.
 - Delete leading zero; therefore, all phone numbers start with non-zero values.
- b. City standardization. In the initial database, customer addresses were stated in an unstandardized format. Therefore, the address is standardized by only considering the city's name.
- c. Removing the name column because, in this research, the phone number is the unique identifier of the consumer.
- d. Standardize the month from Indonesian to English.
- e. Delete rows that have a phone number value of 0 (NaN).

4. Modeling

This research focuses on analyzing loyal customers and customer segmentation. As mentioned in the previous step, the phone number will be used as a unique identifier because there was no unique identifier for each customer in the current database. In the modeling step, several techniques were used: descriptive statistics, retention analysis, Pareto charts, and customer segmentation.

a) Descriptive statistic

The descriptive statistics aimed to recognize the transaction pattern based on month, day, and city. Figure 4 depicts the total monthly transactions (Figure 4.a) and average transaction value (Figure 4.b). Total monthly transactions experienced a significant decline from May until November. Although total transactions decreased, average transactions increased significantly after April. This finding indicates that consumer spending has changed from buying in small quantities to buying in large quantities.

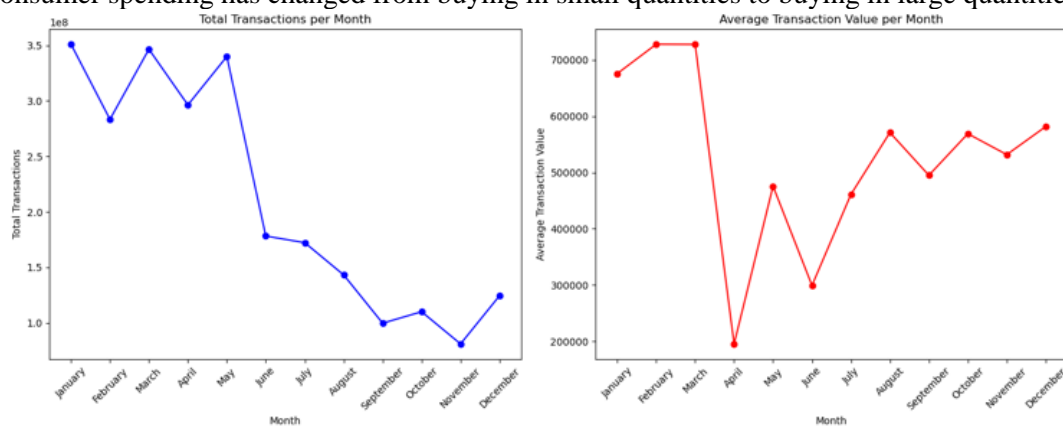


Figure 4. (a) Total monthly transaction, and (b) Average monthly transaction

Figure 5 indicates the daily transactions across all months. Most transactions occurred in the last week (fourth week in Figure 1), while the lowest occurred in the third week. Transactions increase significantly from the 23rd to the 27th. This period is the payroll period for most workers in Indonesia. Thus, companies can implement promotional policies to capture the increase in customer orders in the last week of each month.

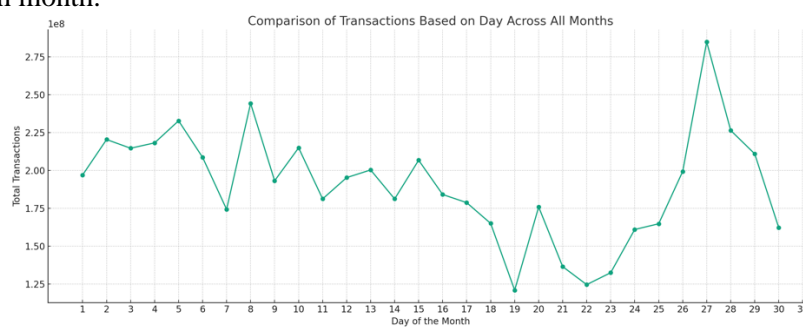


Figure 5. Comparison of Transaction Based on Day Across All Months

Figure 6.a indicates the number of transactions based on city, whereas Figure 6.b depicts the number of unique customers. Sleman, Yogyakarta, Kendal, Bekasi, and South Jakarta are the cities with the most transactions. The production location of the company used in this case study is in the Sleman area, so Sleman and Yogyakarta, the cities with the highest transactions and number of unique consumers, are reasonable. However, the high number of transactions and unique consumers in Kendal, Bekasi, and South Jakarta are interesting things to pay attention to. The company can take the free shipping policy to increase the number of transactions in cities other than DIY Province.

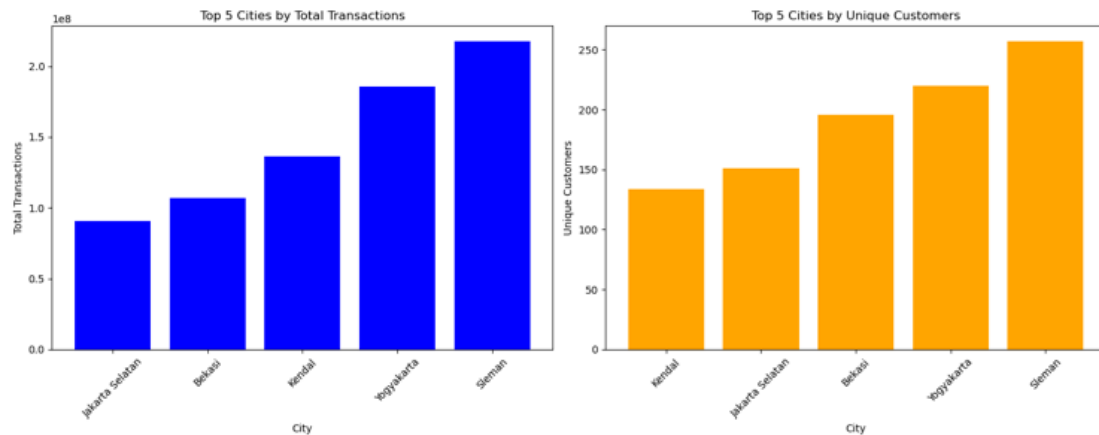


Figure 6. Geographical analysis: (a) top 5 cities by total transactions, and (b) top 5 cities by unique customers

b) Retention analysis

Figure 6 illustrates the customer retention rates. The retention rate is calculated as the percentage of customers from the previous month who remain customers in the current month. The highest retention occurs in April, July, and December. This finding indicates that of customers who purchased in March, 14% will buy in April. If reviewed more deeply, April is the month of fasting for Muslims. In addition, the increase also occurred in December, when this month is also a big day for Christians. Therefore, the recommendation that can be given to companies is to prepare attractive promotions during the month of religious holidays.



Figure 7. Customer retention rates and customer churn rates

c) Pareto analysis

The dataset consists of 3,271 unique customers. The Pareto analysis obtained 1,366 customers, contributing to 80% of the total transaction value (Figure 7). These 1,366 customers represent 41,76% of the total customer. This result is slightly different from the 80:20 classic Pareto Principle, which states that 80% of total value is contributed by 20% of customers.

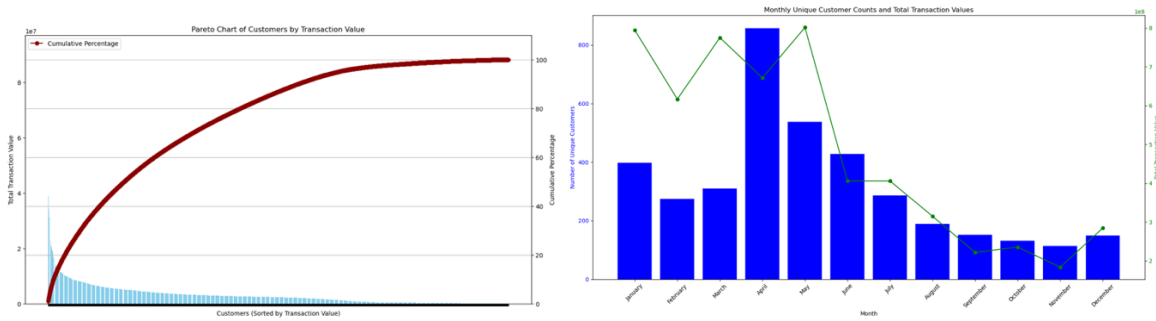


Figure 8. (a) Pareto chart and (b) Monthly unique customer

Approximately 31.29% of the total transaction value is attributed to loyal customers. This study defines a loyal customer as a customer who has made more than one transaction across the dataset. On the other hand, around 68.71% of the total transaction value comes from new customers, defined as those who have made only one transaction. Figure 8 further displayed monthly unique customers. The number of unique customers increased in April and December, whereas in the previous section, this increase was caused by religious events. This analysis indicates that new customers generate a significant portion of transaction value. However, loyal customers, though fewer in number, also contribute a substantial part of the total transaction value. Therefore, the recommendation given to the observed company is (1) increasing the number of new customers, (2) converting new customers to loyal customers, and maintaining the number of loyal customers.

d) Customer Segmentation

Figure 8 shows that the distribution of most consumers made 1 (one) transaction, with a value ranging below IDR 1 million. This finding indicates that most consumers conduct transactions in low quantities and values.

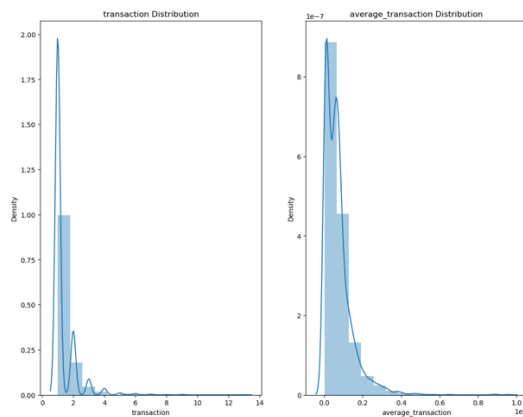


Figure 9. Density analysis on “transaction” and “average-transaction”

As a reference for determining the number of clusters, the elbow method is used in this study to determine the best number of clusters. The results of the elbow method obtained the best number of clusters, which is 3 clusters (Figure 9.a).

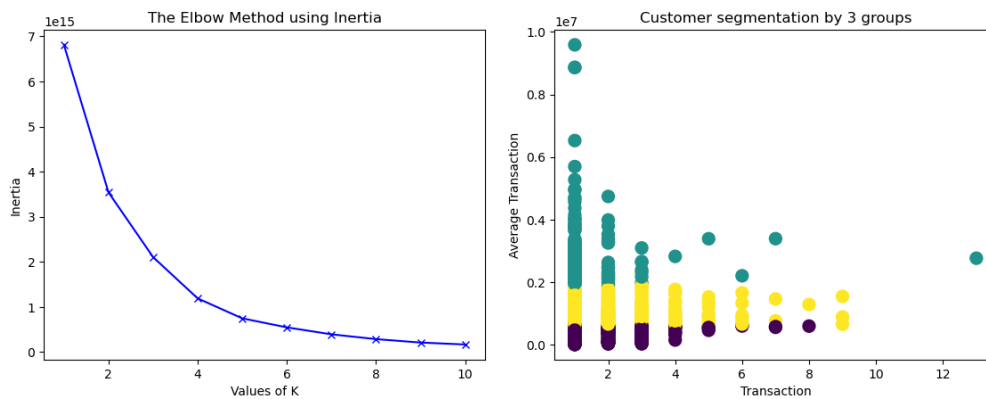


Figure 10. (a) Elbow method and (b) customer segmentation

The K means clustering results obtained 3 clusters illustrated in the scatter plot as shown in Figure 9b.

Cluster 1. High number of transactions, high average value of transactions: 192 customers.

Cluster 2. Medium number of transactions, medium average value of transaction: 1264 customers.

Cluster 3. Low number of transactions, low average value of transaction: 1815 customers.

Further analysis is carried out on cluster 1 to see customers' behavior included in the high transaction category and transaction value. The buying behavior of two customers from cluster 1 is depicted in Figure 10. Consumer number 82136370755 tends to make large purchases in May, where the results of the previous analysis of the month coincided with the religious day of Eid al-Fitr. Meanwhile, consumer number 85290006888 tends to make large purchases in December, coinciding with the religious day of Christmas.

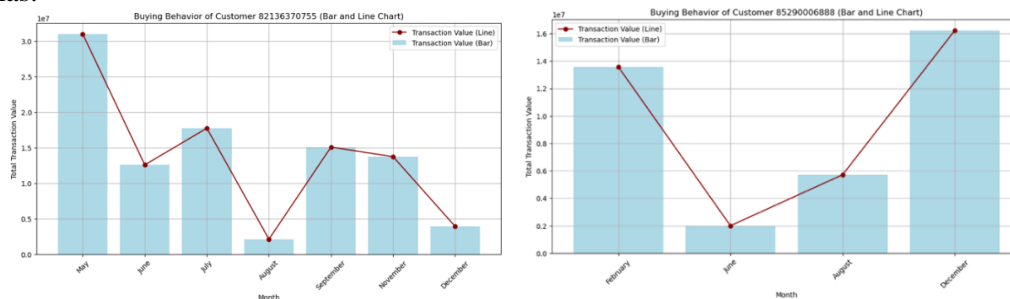


Figure 11. Buying pattern of two customers in Cluster 1: (a) customer 82136370755, and (b) customer 85290006888

5. Evaluation

Based on the analysis carried out in the previous stage, the evaluation stage recapitulates the findings from the analysis. These findings are then used to provide recommendations for companies to improve their competitiveness.

- Consumer spending has changed from buying in small quantities to buying in large quantities.
- Most transactions occurred in the last week, while the lowest occurred in the third week.
- Apart from Yogyakarta, which is the location of production and display stores, Jakarta and Bekasi are the cities with the largest unique consumers and sales.
- Sales increase during religious events (Idul Fitri and Christmas).
- 41.76% of the total customers contribute to 80% of sales.
- 31.29% of the total transaction value is attributed to loyal customers.
- 192 customers are in cluster 1 (high number of transactions and high transaction value).

Recommendations are given based on the findings of the above analysis and discussions with the company's management team. The recommendations given are as follows:

- a. Provide discounted prices for customers who buy in large quantities.
- b. Provide promotions and increase production capacity in the 4th week of each month and on religious days (Eid al-Fitr and Christmas).
- c. Provide discounted shipping promotions for areas outside Jogja.
- d. Conduct programs to reach loyal customers and turn new customers into loyal customers.

3.2. Discussion

This research focused on data mining to analyze sales data in a company that produces traditional motif clothing- batik. Data mining techniques in this study were implemented through 5 steps: business understanding, data understanding, data preparation, modeling, and evaluation. In the modeling stage, several analysis techniques are used to evaluate consumer behavior: descriptive statistics, retention analysis, Pareto charts, and customer segmentation. The evaluation step obtained several findings related to customer behavior, as presented in the evaluation stage.

Compared to previous research, this finding has similarities with previous research. Previous research found that the pandemic has led to a change in consumer preference, with a preference for online shopping, including for clothing [17]. Previous research has also found that after the pandemic, consumers tend to refrain from buying clothes and prioritize basic needs, which impacts the decline in clothing sales [18]. Using observation and interview methods, Khan and Sharma [19] found a positive relationship between consumer purchases and religious holidays. This is relevant to the findings on batik sales that increase in the month coincides with religious holidays.

This research can contribute to both managerial implications and theoretical implications. In managerial implications, the findings in this study can be the basis for the observed companies in particular and all batik producers to formulate marketing strategies. Regarding data mining procedures, this research uses a phone number as a unique identifier for each customer. Therefore, each customer should use a unique identifier, such as a customer number, to facilitate customer identification. On the other hand, this research can also provide theoretical implications. In the results of previous reviews, no research comprehensively uses data mining to analyze sales in batik producers. Research by Mulyawan et al. [12] studied batik sales in terms of motifs and sales using clustering techniques. Shop owners then used the results to formulate sales strategies. On the other hand, research by Bhaskara et al. [20] focused on creating a data warehouse system to record and monitor sales transaction data for batik companies. The researcher stated that real-time access to information could help companies develop their strategies. In addition, the data warehouse is the first step in the data mining process in this study.

4. Conclusion

The analysis using five data mining stages obtained insights that can be used as a basis for formulating batik sales strategies. This research focuses on analyzing customer purchasing behavior and evaluating loyal consumers. The analysis obtained insights into consumer behavior: changes from purchases in small quantities to large quantities, increased purchases in the final week of each month, and increased purchases on religious days. In addition, the analysis obtained that 31.29% of total transactions were contributed by loyal consumers, and 192 consumers were in the cluster of high transaction amounts and high transaction values. Future research can be performed by applying data mining to analyze the types of products the company sells. The results of discussions with management indicate that changes in consumer shopping behavior also occur in the types of products and batik motifs chosen by consumers.

References

- [1] J. Bermudez, K. Apolinario, and A. G. Abad, "Layout Optimization and Promotional Strategies Design in a Retail Store based on a Market Basket Analysis," 2016.
- [2] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.
- [3] J. Kenyon and M. Vakola, "Customer Relationship Management: A Viable Strategy for the Retail Industry?," *Int. J. Organ. Theory Behav.*, vol. 6, no. 3, pp. 329–353, 2003, doi:

- <https://doi.org/10.1108/IJOTB-06-03-2003-B001>.
- [4] R. Karthiyayini and R. Balasubramanian, "Affinity and Association Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 6, no. 10, pp. 241–246, 2016, [Online]. Available: www.ijarcsse.com.
- [5] M. Samuel and Z. Gwendolyn, "An Investigation of the Role of Product , Place , Promotion and Price (4Ps) in Market Basket Analysis : A Case Study of Retail Shops in Gweru Zimbabwe," *Int. J. Sci. Res.*, vol. 3, no. 2, pp. 132–135, 2014.
- [6] S. Panjaitan *et al.*, "Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012057.
- [7] A. K. S. Ong *et al.*, "Consumer behavior in clothing industry and its relationship with open innovation dynamics during the covid-19 pandemic," *J. Open Innov. Technol. Mark. Complex.*, vol. 7, no. 4, p. 211, 2021, doi: 10.3390/joitmc7040211.
- [8] T. N. Hashem, "Examining the Influence of COVID 19 Pandemic in Changing Customers' Orientation towards E-Shopping," *Mod. Appl. Sci.*, vol. 14, no. 8, p. 59, 2020, doi: 10.5539/mas.v14n8p59.
- [9] L. Eger, L. Komárková, D. Egerová, and M. Mičík, "The effect of COVID-19 on consumer shopping behaviour: Generational cohort perspective," *J. Retail. Consum. Serv.*, vol. 61, no. February, 2021, doi: 10.1016/j.jretconser.2021.102542.
- [10] R. A. Rahadi, D. Rahmawati, N. A. Windasari, and P. F. Belgiawan, "The Analysis of Consumers ' Preferences for Batik Products in Indonesia," *Rev. Integr. Bus. Econ. Res.*, vol. 9, no. 1, pp. 278–287, 2020.
- [11] R. Ardiyansyah and R. A. M. Febrianti, "Understanding the driver of customer purchase decision : The role of customer engagement and brand attachment in Batik Products," *Fair Value J. Ilm. Akunt. dan Keuang.*, vol. 5, no. 4, pp. 1979–1985, 2022.
- [12] Mulyawan, A. Bahtiar, and I. Ali, "Transformasi Strategi Penjualan Batik Cirebon Dengan Pendekatan Analisis Pengelompokan K-Means," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 7, no. 1, pp. 1–7, 2023, doi: 10.32485/kopertip.v7i1.309.
- [13] V. Salbinda, R. I. Handayani, and N. Normah, "Clustering of Clothing Sales Data At Top Store Using K-Means Method," *Sinkron*, vol. 7, no. 2, pp. 314–320, 2022, doi: 10.33395/sinkron.v7i2.11244.
- [14] M. Mardalius and T. Christy, "Mapping of Potential Customers As a Clothing Promotion Strategy Using K-Means Clustering Algorithm," ... *J. Ilmu Pengetah. Dan ...*, vol. 6, no. 1, pp. 67–72, 2020, doi: 10.33480/jitk.v6i1.1414.
- [15] A. Anshu, "Review Paper on Data Mining Techniques and Applications," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 7, no. 2, pp. 22–26, 2019, doi: 10.21276/ijrcst.2019.7.2.4.
- [16] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1090, no. 1, p. 012053, 2021, doi: 10.1088/1757-899x/1090/1/012053.
- [17] H. Shen, F. Namdarpour, and J. Lin, "Investigation of online grocery shopping and delivery preference before, during, and after COVID-19," *Transp. Res. Interdiscip. Perspect.*, vol. 14, no. March, p. 100580, 2022, doi: 10.1016/j.trip.2022.100580.
- [18] C. Liu, S. Xia, and C. Lang, "Clothing Consumption During the COVID-19 Pandemic: Evidence From Mining Tweets," *Cloth. Text. Res. J.*, vol. 39, no. 4, pp. 314–330, 2021, doi: 10.1177/0887302X211014973.
- [19] M. R. Khan and K. Sharma, "Purchase Preferences and Buying Influences on Religious Occasions," *FIIIB Bus. Rev.*, vol. 9, no. 3, pp. 216–227, 2020, doi: 10.1177/2319714520942624.
- [20] I. M. A. Bhaskara, L. G. P. Suardani, and M. Sudarma, "Data Warehouse Implementation To Support Batik Sales Information Using MOLAP," *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 45–51, 2018.