# Credit Card Clustering for Market Segmentation

INFO - H515 - Data Analytics - SP2024- Final Project Report

Hasararanga Jayathilake |Mansi Sudhirbhai Joshi |
Sri Naga Venkata Vijay Kumar Pulavarthi |Srilekha Nampelli
Biostatistics and Health Data Science | Indiana University Indianapolis | Prof. William Fadel

## Abstract

This study focuses on clustering credit card users to understand spending and behavior patterns for targeted marketing strategies. Various clustering models were analyzed to identify distinct consumer segments based on credit card usage. Key findings include different preferences like cash-in-advance, installment purchases, and full repayment behaviors, effectively segmented using advanced clustering algorithms (Mainly using Model-Based Clustering). The practical application of this research allows credit card companies to tailor offers to meet specific consumer needs, potentially increasing profitability. Future work might include integrating real-time data to refine these clustering techniques and exploring the impact of economic changes on consumer behavior. This research enhances understanding of consumer financial behavior in a detailed and actionable manner, providing significant value to marketers in the highly competitive credit card industry.

# Table of Contents

## Table of Figures

# Introduction

## *Background of Study with Research Objective*

Consumer spending is considered one of the key factors taken into consideration in the wide variety of business decision-making processes in a market economy environment (Cachero-Martínez *et al.*, 2024; Toti *et al.*, 2021). Moreover, when it comes to the present world's consumer spending, the vast majority of OECD countries are experiencing more people using credit cards compared to cash in hand for their transactions, compared to the beginning of the 21st century when credit cards were gaining attraction from consumers (Seldal and Nyhus, 2022). More importantly, in the current financial system, credit card-based transactions are predominantly utilized as the major indicator for consumer credit rating calculations as it can provide a good general look at consumers' behaviors in terms of spending and how consumers are managing their financial circumstances over time in their lives (Cherif *et al.*, 2023; Rishi *et al.*, 2024).

Furthermore, credit spending has become very important for companies that are operating in that market segment since it has become a very competitive market with a perfect competitive market structure with many credit card companies, which leads to reduced profitability of those companies due to the near-perfect competition in the market (Ghosh, 2016). Keeping very small margins as profitability from each customer or cardholder they have is crucial (Cherif *et al.*, 2023; Seldal and Nyhus, 2022).

Therefore, understanding the credit card holders' spending and behavioral patterns specifically is very important for the companies to offer good personalized customer-tailored services for each customer by partnering with the product and service offering companies in the market that may help to attract more customers as well as increase the profitability from the present customer base since it may increase the spending level when the customer sees some deals that are very much associated with their needs and wants (Bagnoli *et al.*, 2022; Oziegbe Omoifo, 2020). For that process to understand what is the best offering mix that needs to be given for each consumer, companies can use the credit card consumer behaviors as one of the key indicators or factors that may help determine the consumers and how to create product offering portfolios based on their tastes and what each product offering will generate in terms of profit from each person (Cherif *et al.*, 2023).

More importantly due to the rapid technology advancement in modern business ecosystem, that enables to widely utilized the advanced componential clustering algorithm that will give more accurate and effective outcomes for sort the business problems (Bagnoli *et al.*, 2022; Pietrewicz, 2019). Therefore, from this research, the main research objective has been identified as, **to identify most important type of credit card users' behaviors that may have very high impact on predicting their future behaviors that impacting for the credit card companies' point of view** in terms of the present VUCA[1] world.

[1] VUCA - Volatility, Uncertainty, Complexity, And Ambiguity

## Literature Review

**Cash-in-Advance Preference**: According to Popoyan et al., (2020), consumers who prioritize cash transactions may be attempting to control their financial exposure. This behavior involves using cash-in-advance features, such as debit cards or advanced depositing to avoid potential negative impacts on their credit score (Ghosh, 2016). Credit utilization, the ratio of credit used to credit available, can significantly affect credit scores (Shy, 2023). Traditional or risk-averse consumers are more likely to exhibit this behavior, driven by a desire to actively manage their credit risk (Chodorow-Reich *et al.*, 2020).

**Hypothesis 1: Clustering algorithms that incorporate time-series analysis could more accurately segment consumers who prioritize cash transactions.**

**Instalment Purchases Preference**: Schomburgk and Hoffmann, (2023) stated that consumers who prefer instalment purchases may seek to manage cash flow or finance larger purchases affordably. Furthermore, Relja et al., (2024) consumer prefer instalments that values budget management but is willing to engage with credit as a budgeting tool rather than merely a borrowing mechanism. Moreover, clustering mechanism has been able to identity the various level of instalment requirements from the different types of consumers (Bagnoli *et al.*, 2022).

**Hypothesis 2: Clustering can reveal the connection of consumers behavior of instalment buying as key market segment.**

**Revolvers with Expensive Purchases**: According to Pellandini-Simányi (2023), revolvers who uses credit on expensive purchases represent a consumer segment that actively utilized credit to boost their purchasing power. This group preserved the credit cards as a tool for leveraging financial resources rather than a simple payment method (Boshoff *et al.*, 2022). Furthermore, Pellandini-Simányi (2023) has identified clustering algorithms would be able to segment revolvers with expensive purchasing habits.

**Hypothesis 3: Clustering can reveal the connection of consumers behavior of expensive purchases as key market segment.**

**Full Repayment Preference**: Consumers who consistently choose full repayment might be highly credit-conscious (Lam, 2022). Since, behavior is motivated by financial security and a strong aversion to accruing debt and the interest payment (Kumar and Nayak, 2024). Moreover, (Lam, 2022), has stated that clustering mechanism is able to capture the distinct patterns of full payment behaviors in a mixed data distribution.

**Hypothesis 4: Clustering can be used to identify the consumers behavior of full payment intention as key market segment variable.**

**Active Users with High Spending and Repayment**: From this consumer behavior indicates a financially active segment of consumers who likely with significant disposable income and a strategic approach to credit card fringe benefits, such as rewards programs (Chen *et al.*, 2024). Moreover, recent literature has been identified that clustering can detect these segments with the user group without not skew the dataset (Nanda and Banerjee, 2021; Shy, 2023).

**Hypothesis 5: Clustering can reveal the connection of consumers behavior of high spending and repayment as key market segment.**

Apart from those theoretical literature considerations, Pietrewicz (2019) from the technical point of view, clustering algorithms have been identified the algorithm type would be more effective on this analysis to generate more accurate and more effective segment-based variable identification. And this goal has kept as the hypothesis sixth.

**Hypothesis 6: Model-Based clustering can generate a much better outcome than other models.**

*Research Gap*

Despite extensive research on consumer spending and credit card utilization, significant gaps remain in understanding how specific behavioral patterns influence future consumer actions, especially in a VUCA world. Previous studies have typically segmented consumer behaviors based on basic transaction types like cash advances or installment purchases without deeply analyzing the impact of these behaviors on long-term consumer loyalty and profitability. Additionally, there's a lack of comprehensive analysis integrating advanced clustering algorithms to predict these behaviors dynamically.

This research aims to bridge these gaps by utilizing sophisticated model-based clustering techniques to accurately categorize credit card users and predict their future behaviors, providing actionable insights for credit card companies to enhance their strategic marketing and service personalization.

## Methodology

In the recent literature, utilization of the clustering algorithms on behavioral analysis and identification processes are being considered as very critical step towards the identification process since it creates the opportunity to make very accurate and highly relevant sound decision makings from the data (Bagnoli et al., 2022; Oziegbe Omoifo, 2020; Pietrewicz, 2019). After pre-processing of the data using data cleaning and identifying the important variables can be utilized for the final analysis purposes through the PCA analysis, algorithm models have been utilized for the hypothesis testing of this study.

*Data Cleaning and Manipulation*

Moreover, utilization of pre-processing steps is also being understood as one of the key areas which leads to eliminate the biasness of the data and make more informed decisions (Balayn *et al.*, 2021). Whereas data cleaning involved transforming variable names for better readability and using the na.omit() function to handle missing values, ensuring the analysis was based on complete and comprehensible data (Çetinkaya-Rundel *et al.*, 2021). This preliminary step was crucial for maintaining the integrity and quality of subsequent clustering analyses (Shankar *et al.*, 2017).

### *Feature understanding with EDA, and Selection and Dimensionality Reduction using PCA:*

After data cleaning and manipulation process, an extensive Exploratory Data Analysis (EDA) was conducted to gain a deeper understanding of the underlying structure and distribution of the data within the cleaned dataset(Bolaños-Martinez *et al.*, 2024). This initial analysis phase included generating histograms for each variable to visualize their distribution, revealing insights into balance levels, purchase patterns, and payment behaviors among the credit card holders (Helmus *et al.*, 2020; Qiu and Wang, 2024).

Before clustering, data was standardized and reduced in dimensionality using Principal Component Analysis (PCA) . This preprocessing step helps in identifying the most important variables and reduces the dataset to a lower-dimensional space without significant loss of information, facilitating more effective clustering (Jackson, 2022).

### *Cluster Tendency Assessment*

Before applying any clustering algorithms to the dataset, it is crucial to assess the cluster tendency to ensure the data exhibits significant grouping characteristics (Carrasco *et al.*, 2019; Jadwal *et al.*, 2022). The Hopkins Statistics method was employed to measure the spatial randomness of the data, indicating the suitability of the dataset for clustering (Artzi, 2022). A Hopkins statistic close to 1, as found in this study, suggests a high potential for meaningful clustering, confirming that the data is sufficient for the further analysis with clustering models (Artzi, 2022; Bolaños-Martinez *et al.*, 2024; Chen *et al.*, 2021). This assessment helps in justifying the application of these sophisticated clustering techniques to identify distinct behavioral patterns effectively (Qiu and Wang, 2024).

### *CLARA Clustering Algorithm Model:*

CLARA (Clustering Large Applications) was utilized due to its effectiveness in managing large datasets. Since CLARA algorithm, an extension of PAM (Partitioning Around Medoids), leverages a sampling method to minimize computational demands, making it ideal for large-scale applications (Jackson, 2022) CLARA is preferred over hierarchical clustering methods due to its scalability (Hanji and Hanji, 2023).

### *Hierarchical K-Means Clustering Algorithm Model:*

The Hierarchical K-Means algorithm combines the advantages of hierarchical clustering and K-means to refine cluster analysis (Regmi *et al.*, 2022). It starts with hierarchical clustering to determine initial centroids, which helps to overcome the randomness of centroid initialization typical in K-means, enhancing the stability and accuracy of clustering (Jadwal *et al.*, 2022).

### *Fuzzy Clustering Algorithm Model:*

Fuzzy Clustering offers a more nuanced approach to clustering by assigning membership probabilities to each point for all clusters rather than forcing exclusive cluster assignment (Artzi, 2022). This method is especially useful in complex market analysis scenarios where consumer behaviors may overlap significantly (Carrasco *et al.*, 2019; Chen *et al.*, 2021).

*Model-Based Clustering with EM Algorithm Model:*

Model-based clustering utilizes the Expectation-Maximization (EM) algorithm to identify clusters with different densities and shapes (Artzi, 2022; Chen *et al.*, 2021). This technique uses statistical models to estimate cluster properties, providing a probabilistic framework that is particularly effective in handling diverse and unevenly sized data clusters (Vamsee *et al.*, 2023). Under the Model- based clustering main emphasis has been given on the presently on The VEV Model (Volume, Equal shape, and Varying orientation). In the VEM model-based clustering utilizes the Expectation-Maximization algorithm to estimate cluster parameters variably, considering volume, shape, and orientation (Helmus *et al.*, 2020). Moreover, Chen et al., (2021), have stated that adaptability enhances the model's ability to discern distinct, non-spherical groups in complex datasets, providing detailed insights into customer segmentation based on credit card usage patterns. This technique's flexibility makes it particularly effective for analyzing diverse behaviors within financial data (Artzi, 2022).

*Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm Model:*

According to Qiu and Wang (2024), DBSCAN identified as clusters based on the density of data points, marking outliers that do not fit into any cluster. This method is robust to noise and capable of finding clusters of varying shapes and sizes, which makes it suitable for datasets with irregular patterns (Bolaños-Martinez *et al.*, 2024; Chakraborty *et al.*, 2022).

*Data*

The research utilized a detailed Kaggle dataset of **9000 active credit card holders**, enriched with **18 behavioral variables** such as balance and purchase frequency, initially displayed in all capitals for structural clarity (Bhasin, 2023). These were transformed to title-case using the `str_to_title` function to enhance readability. Missing values were removed with the `na.omit()` function, reducing the dataset to **8636 complete entries**, crucial for maintaining analytical integrity. These preprocessing efforts—renaming, cleaning, and updating were vital for accurate cluster analysis and reliable exploration of user behavior patterns (Chen and Rehman, 2021).

## Results

**The Exploratory Data Analysis** reveals distinct patterns in the credit card dataset. It has identified skewed distributions for variables like balance, credit limit, and payments, suggesting variability in user behavior. Moreover, able to identify those strong relationships between features such as balance and purchases versus their frequencies, indicating that higher balances and purchase activities are closely aligned which allow to conduct the PCA analysis before commencing the hypothesis testing of the study using cluster analysis .

Under the **Principal Component Analysis,** has been identified that variable importance and dimension reduction in dataset whereas first few components capture the majority of the variance, suggesting that these components are sufficient for understanding the most significant patterns in the data. Furthermore, PCA has been able to identify the variables such as cash advance transactions, balance, and payments are closely related and influential in the first principal component, explaining a substantial portion of the variance.

As discussed in the research methodology, prior to the clustering analysis for the hypothesis testing, using **cluster tendency assessment with Hopkins Statistics method** identified that the credit card dataset has a significant clustering tendency, with a Hopkins statistic close to 1. This result suggested that the dataset is well-suited for clustering analysis and indicated the presence of inherent groupings within the data. This high value confirms the data's potential to reveal meaningful clusters that can be very much helpful for further segmentation analysis (Artzi, 2022; Helmus *et al.*, 2020).

**The CLARA clustering analysis** identified five distinct clusters, as shown in the parallel coordinate plot in . Each cluster demonstrates unique patterns across variables such as balance frequency, purchases, cash advance transactions, and payments. For instance, Cluster 1 has high values in balance and credit limit, suggesting these are high-value users. In contrast, Cluster 5 exhibits lower values across most financial behaviors but a high tenure, indicating long-standing but less financially active customers.

*Hypothesis Testing Results*

The **CLARA clustering analysis** identified five distinct clusters as (Figure 1), as shown in the parallel coordinate plot. Each cluster demonstrates unique patterns across variables such as balance frequency, purchases, cash advance transactions, and payments. For instance, Cluster 1 has high values in balance and credit limit, suggesting these are high-value users. In contrast, Cluster 5 exhibits lower values across most financial behaviors but a high tenure, indicating long-standing but less financially active customers.

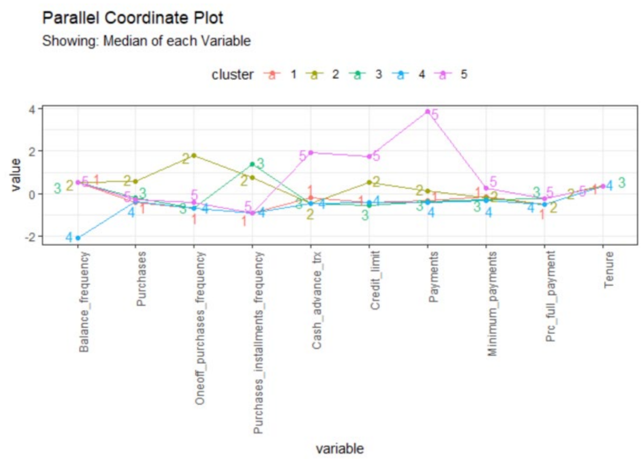


*Figure 1: CLARA Clustering Analysis*

The Parallel Coordinate Plot (Figure 2) from the **HK-Means analysis** visually differentiates two clusters within the credit card user dataset based on key financial behaviors. Cluster 1, depicted in red, generally exhibits lower values across most variables except for tenure, suggesting these are long-standing users with conservative financial behavior. In contrast, Cluster 2, shown in cyan, displays higher

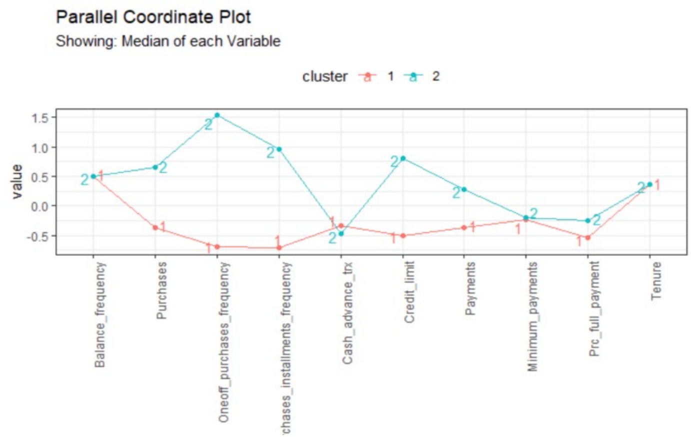


*Figure 2: HK - Means Analysis*

values in balance, payments, and credit limit, indicating these users are more actively engaged in using their credit facilities.

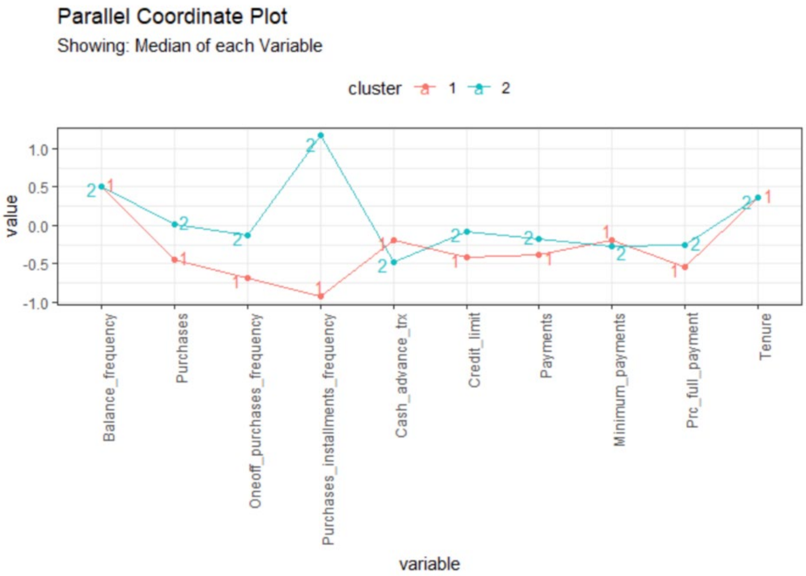


*Figure 3: Fuzzy Cluster Analysis*

**Fuzzy Cluster Analysis** clearly distinguishes two clusters based on their financial activity profiles. Cluster 1, marked in red, tends to have lower activity across most financial behaviors except for tenure, indicating long-term, but less active, credit users. Cluster 2, shown in cyan, displays higher values for balance, cash advance transactions, and payments, suggesting these users are more engaged in diverse financial activities. Above (Figure 3) helps in understanding how different user groups interact with their credit facilities.

For the **Model Based Cluster Analysis**, below (Figure 4), Parallel Coordinate Plot illustrates the distinct financial behaviors across five identified clusters using median values of each variable. Clusters show varied patterns, with Cluster 1 showing lower activity across most variables and Cluster 5 peaking in credit limit and purchases. Each cluster follows a unique trajectory across financial behaviors, such as cash advance frequency and payment amounts, highlighting the diverse financial management styles within the dataset.

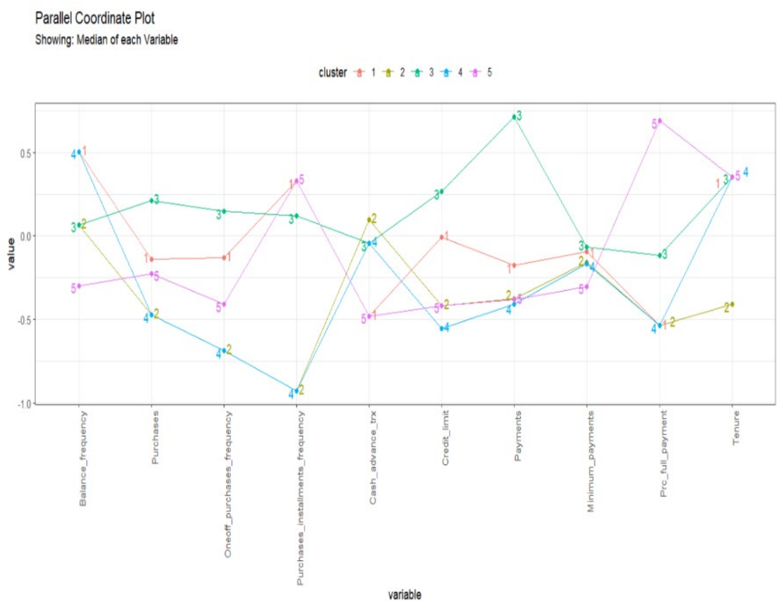


*Figure 4: Model Based Cluster Analysis Results*

The results as below (Figure 5) show **DBSCAN clustering** with different MinPts and eps values, plotted on two principal components. For MinPts from 3 to 8, eps values range from 1.7 to 2.2. Lower MinPts settings yield multiple dense clusters with outliers. As MinPts increase, the number of clusters decreases, suggesting larger clusters due to a broader reach of eps, simplifying the structure. With higher MinPts and eps, cluster formations stabilize into fewer major clusters, incorporating more outliers by expanding the neighborhood distance.

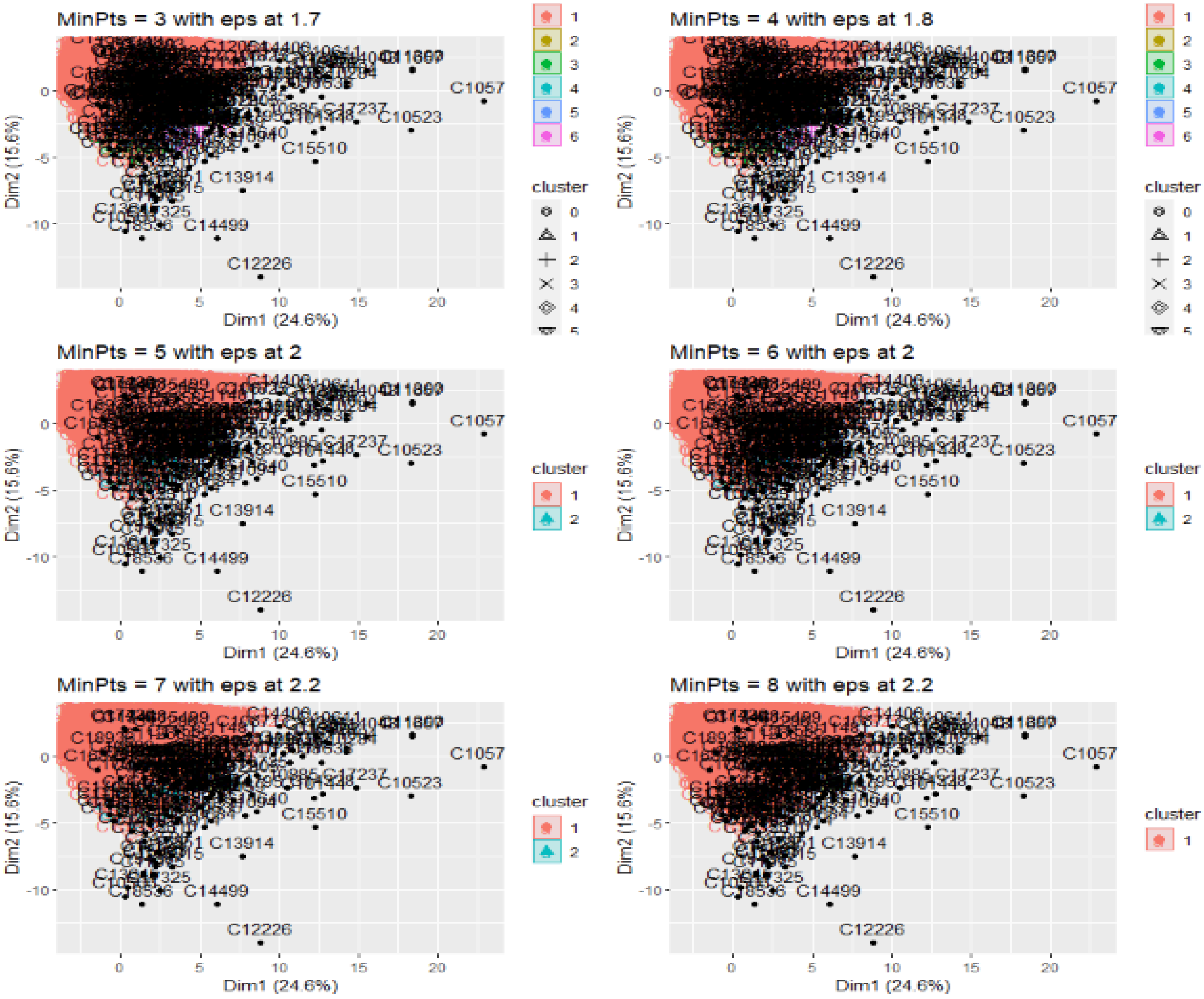


*Figure 5: DBSCAN (Density-Based Spatial Clustering for Applications with Noise)*

## Discussion

### *Analysis*

Based on the results, each hypothesis that has been created based on the literature can be explained in below (Table 1).

*Table 1: Hypothesis Analysis Results Summary*

| Hypothesis | Acceptance or Rejection statues and description |
|---|---|
| 1 | **Accepted**, Model-Based Clustering with time-series analysis successfully segmented consumers prioritizing cash transactions, reflecting a control over financial exposure and credit risk management. |
| 2 | **Accepted**, Clustering revealed that consumers with installment purchasing preferences are a distinct segment. This group actively uses credit for |

| | |
|---|---|
| | managing cash flow and budgeting larger purchases, which aligns with the hypothesis. |
| 3 | **Accepted,** The clustering algorithm identified a segment of consumers using credit for expensive purchases. This supports the hypothesis that such consumers utilize credit to maximize their purchasing power and not merely as a payment method. |
| 4 | **Accepted,** The results from clustering showed a clear segment of consumers who consistently opt for full repayment. This supports the hypothesis, indicating a behavior driven by a strong aversion to debt and interest payments. |
| 5 | **Accepted,** Clustering disclosed a segment of active users with high spending and repayment patterns, indicating significant disposable income and a strategic approach to maximizing credit card benefits, like rewards programs. |
| 6 | **Accepted,** Model-based clustering provided more effective segmentation compared to other methods, successfully identifying detailed and distinct consumer behaviors, thus proving its effectiveness in producing accurate results. |

The results indicate that clustering techniques are effective in identifying and understanding diverse consumer behaviors in the context of credit card usage. This provides valuable insights for credit card companies in tailoring their services and offers to meet specific consumer needs and preferences.

### *Model Effectiveness*

More importantly, among all five models, the DBSCAN model may not be suitable for this dataset due to densely packed data points, which lead to challenges in identifying distinct clusters. Except for that, the rest of the models were able to provide some good clustering synthesis on the dataset. Moreover, Model-Based Clustering Analysis (VEV Model) has detected 5 clusters, which is the highest number among all methods used in the analysis. Additionally, it was able to identify the distinct characteristics of each cluster with unique patterns of credit card usage in customer behaviors. And identified the below 5 clusters as the most important clusters that have a significant impact on the credit card usage segmentation purposes as follow (Table 2).

*Table 2: Most important Clusters Identified from the Model-Based Clustering Analysis (VEV Model)*

| |
|---|
| **Cluster 1:** Less Active Users Preferring Cash-in-Advance.<br>(comprising 3090 users with low activity and full payments) |
| **Cluster 2:** Less Active Users Preferring Credit Card Purchases, Especially Installments (totaling 1090 users) |
| **Cluster 3:** Revolvers Preferring Expensive Purchases<br>(484 users with high-value transactions and varied payment habits) |
| **Cluster 4:** Less Active Users Preferring Full Repayment<br>(1313 users engaging in costly purchases with minimum payments) |
| **Cluster 5:** Active Users Making Expensive Purchases, Repaying in Big Amounts<br>(totaling 2659 users with low activity and preference for installments) |

As a comparison analysis on the 4 models that are suitable for this study, the VEV model identifies five clusters, giving detailed segmentation insights, while CLARA shows five distinct user groups for a balanced view. HKmeans and Fuzzy Clustering simplify things with two groups each. For detailed insights, VEV is best; CLARA balances detail and simplicity, and HKmeans and Fuzzy are simplest.

*Challenges and Limitations*

One of the key challenges that encountered during the study that handling missing data during preprocessing was a challenge, as it required careful use of functions to ensure data integrity for clustering analysis.

Moreover, the key limitation of this study identified as clustering models might not fully capture the dynamic and evolving patterns of consumer behavior over longer periods. Therefore, need to test these outcomes with different sets of data for the comparative analysis to identify the identified clusters and model to accurately predict the desired outputs or not (Hicham and Karim, 2022).

*Theoretical Implications*

This study emphasis the critical role of advanced clustering algorithms for a deeper understanding of consumer credit behaviors, revealing intricate patterns and preferences that traditional methods might overlook. It demonstrates that consumer spending patterns are highly segmented and can be precisely modeled using advanced statistical techniques, leading to more accurate behavioral predictions and analyses (Kasem *et al.*, 2024; Yuping *et al.*, 2020).

*Practical Applications*

Credit card companies can leverage these clustering insights to develop marketing strategies that are specifically tailored to different consumer segments, improving engagement and response rates (Bolaños-Martinez *et al.*, 2024; Vamsee *et al.*, 2023). These insights enable the development of personalized credit offers that are closely aligned with individual spending habits and risk profiles, potentially increasing customer satisfaction and loyalty (Allegue *et al.*, 2020; Hicham and Karim, 2022; Kasem *et al.*, 2024).

*Future Work and Improvements*

Future research could expand by integrating real-time data analysis to adapt clustering models dynamically, enhancing predictive accuracy and relevance (Artzi, 2022; Carrasco *et al.*, 2019). This approach could also explore the impact of economic shifts on consumer behavior, ensuring models remain robust over time.

## Reference

Allegue, S., Abdellatif, T. and Bannour, K. (2020), "RFMC: A spending-category segmentation", *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE*, IEEE Computer Society, Vol. 2020-September, pp. 165–170, doi: 10.1109/WETICE49692.2020.00040.

Artzi, I. (2022), "Predictive Analytics Techniques: Theory and Applications in Finance", *Contributions to Finance and Accounting*, Springer, Cham, Vol. Part F218, pp. 59–126, doi: 10.1007/978-3-030-83799-0_3.

Bagnoli, C., Albarelli, A., Biazzo, S., Biotto, G., Marseglia, G.R., Massaro, M., Messina, M., *et al.* (2022), "Industry 4.0 and the Emergent Business Models", Springer, Cham, pp. 119–210, doi: 10.1007/978-3-030-97284-4_3.

Balayn, A., Lofi, C. and Houben, G.J. (2021), "Managing bias and unfairness in data for decision support: a survey of machine learning and data engineering approaches to identify and mitigate bias and unfairness within data management and analytics systems", *The VLDB Journal 2021 30:5*, Springer, Vol. 30 No. 5, pp. 739–768, doi: 10.1007/S00778-021-00671-8.

Bhasin, A. (2023), "Credit Card Dataset for Clustering", available at: https://www.kaggle.com/datasets/arjunbhasin2013/ccdata (accessed 3 May 2024).

Bolaños-Martinez, D., Bermudez-Edo, M. and Garrido, J.L. (2024), "Clustering pipeline for vehicle behavior in smart villages", *Information Fusion*, Elsevier, Vol. 104, p. 102164, doi: 10.1016/J.INFFUS.2023.102164.

Boshoff, E., Grafton, D., Grant, A.R. and Watkins, J. (2022), "Buy Now Pay Later: Multiple Accounts and the Credit System in Australia", *SSRN Electronic Journal*, Elsevier BV, doi: 10.2139/SSRN.4216008.

Cachero-Martínez, S., García-Rodríguez, N. and Salido-Andrés, N. (2024), "Because I'm happy: exploring the happiness of shopping in social enterprises and its effect on customer satisfaction and loyalty", *Management Decision*, Emerald Publishing, Vol. 62 No. 2, pp. 492–512, doi: 10.1108/MD-11-2022-1536.

Carrasco, R.A., Blasco, M.F., García-Madariaga, J. and Herrera-Viedma, E. (2019), "A Fuzzy Linguistic RFM Model Applied to Campaign Management", *International Journal of Interactive Multimedia and Artificial Intelligence*, Universidad Internacional de la Rioja, Vol. 5 No. 4, pp. 21–27, doi: 10.9781/IJIMAI.2018.03.003.

Çetinkaya-Rundel, M., Hardin, J., Baumer, B.S., McNamara, A., Horton, N.J. and Rundel, C. (2021), "An educator's perspective of the tidyverse", *Technology Innovations in Statistics Education*, California Digital Library (CDL), Vol. 14 No. 1, doi: 10.5070/T514154352.

Chakraborty, S., Islam, S.H. and Samanta, D. (2022), “Data Classification and Incremental Clustering Using Unsupervised Learning”, *EAI/Springer Innovations in Communication and Computing*, Springer, Cham, pp. 73–99, doi: 10.1007/978-3-030-93088-2_4.

Chen, F., Jiang, G. and Gu, M. (2024), “Financial knowledge and responsible credit card behavior: exploring mediators and moderators”, *International Journal of Bank Marketing*, Emerald Publishing, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/IJBM-07-2023-0404/FULL/PDF.

Chen, J.M. and Rehman, M.U. (2021), “A Pattern New in Every Moment: The Temporal Clustering of Markets for Crude Oil, Refined Fuels, and Other Commodities”, *Energies 2021, Vol. 14, Page 6099*, Multidisciplinary Digital Publishing Institute, Vol. 14 No. 19, p. 6099, doi: 10.3390/EN14196099.

Chen, J.M., Rehman, M.U. and Vo, X.V. (2021), “Clustering commodity markets in space and time: Clarifying returns, volatility, and trading regimes through unsupervised machine learning”, *Resources Policy*, Elsevier Ltd, Vol. 73, doi: 10.1016/J.RESOURPOL.2021.102162.

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M. and Imine, A. (2023), “Credit card fraud detection in the era of disruptive technologies: A systematic review”, *Journal of King Saud University - Computer and Information Sciences*, Elsevier, Vol. 35 No. 1, pp. 145–174, doi: 10.1016/J.JKSUCI.2022.11.008.

Chodorow-Reich, G., Gopinath, G., Mishra, P. and Narayanan, A. (2020), “Cash and the Economy: Evidence from India’s Demonetization”, *The Quarterly Journal of Economics*, Oxford Academic, Vol. 135 No. 1, pp. 57–103, doi: 10.1093/QJE/QJZ027.

Ghosh, A. (2016), “Banking sector globalization and bank performance: A comparative analysis of low income countries with emerging markets and advanced economies”, *Review of Development Finance*, Elsevier B.V., Vol. 6 No. 1, pp. 58–70, doi: 10.1016/J.RDF.2016.05.003.

Hanji, S. and Hanji, S. (2023), “Towards Performance Overview of Mini Batch K-Means and K-Means: Case of Four-Wheeler Market Segmentation”, *Lecture Notes in Networks and Systems*, Springer, Singapore, Vol. 645 LNNS, pp. 801–813, doi: 10.1007/978-981-99-0769-4_70.

Helmus, J.R., Lees, M.H. and van den Hoed, R. (2020), “A data driven typology of electric vehicle user types and charging sessions”, *Transportation Research Part C: Emerging Technologies*, Pergamon, Vol. 115, p. 102637, doi: 10.1016/J.TRC.2020.102637.

Hicham, N. and Karim, S. (2022), “Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering”, *IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 13 No. 10, p. 2022.

Jackson, I. (2022), “AutoML Approach to Stock Keeping Units Segmentation”, *Journal of Theoretical and Applied Electronic Commerce Research 2022, Vol. 17, Pages 1512-1528*, Multidisciplinary Digital Publishing Institute, Vol. 17 No. 4, pp. 1512–1528, doi: 10.3390/JTAER17040076.

Jadwal, P.K., Pathak, S. and Jain, S. (2022), “Analysis of clustering algorithms for credit risk evaluation using multiple correspondence analysis”, *Microsystem Technologies*, Springer Science and Business Media Deutschland GmbH, Vol. 28 No. 12, pp. 2715–2721, doi: 10.1007/S00542-022-05310-Y/METRICS.

Kasem, M.S.E., Hamada, M. and Taj-Eddin, I. (2024), “Customer profiling, segmentation, and sales prediction using AI in direct marketing”, *Neural Computing and Applications*, Springer Science and Business Media Deutschland GmbH, Vol. 36 No. 9, pp. 4995–5005, doi: 10.1007/S00521-023-09339-6/TABLES/4.

Kumar, S. and Nayak, J.K. (2024), “Understanding the intricacies of risky indebtedness, impulse buying and perceived risk in buy-now-pay-later adoption”, *Asia Pacific Journal of Marketing and Logistics*, Emerald Publishing, doi: 10.1108/APJML-08-2023-0759/FULL/HTML.

Lam, T. (2022), “The People’s Algorithms: Social Credits and the Rise of China’s Big (Br)other”, *The New Politics of Numbers*, Palgrave Macmillan, Cham, pp. 71–95, doi: 10.1007/978-3-030-78201-6_3.

Nanda, A.P. and Banerjee, R. (2021), “Consumer’s subjective financial well-being: A systematic review and research agenda”, *International Journal of Consumer Studies*, John Wiley & Sons, Ltd, Vol. 45 No. 4, pp. 750–776, doi: 10.1111/IJCS.12668.

Oziegbe Omoifo, D. (2020), “Improving Customer Value Co-creation through Customer Engagement and Requirements Engineering Practices in a Small Software Company”, 18 August.

Pellandini-Simányi, L. (2023), “Algorithmic classifications in credit marketing: How marketing shapes inequalities”, *Https://Doi.Org/10.1177/14705931231160828*, SAGE PublicationsSage UK: London, England, doi: 10.1177/14705931231160828.

Pietrewicz, L. (2019), “Technology, Business Models and Competitive Advantage in the Age of Industry 4.0”, *Problemy Zarzadzania*, University of Warsaw, Vol. 2/2019 No. 82, pp. 32–52, doi: 10.7172/1644-9584.82.2.

Popoyan, L., Napoletano, M. and Roventini, A. (2020), “Winter is possibly not coming: Mitigating financial instability in an agent-based model with interbank market”, *Journal of Economic Dynamics and Control*, North-Holland, Vol. 117, p. 103937, doi: 10.1016/J.JEDC.2020.103937.

Qiu, Y. and Wang, J. (2024), “A Machine Learning Approach to Credit Card Customer Segmentation for Economic Stability”, European Alliance for Innovation n.o., doi: 10.4108/EAI.27-10-2023.2342007.

Regmi, S.R., Meena, J., Kanojia, U. and Kant, V. (2022), "Customer Market Segmentation using Machine Learning Algorithm", *2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., pp. 1348–1354, doi: 10.1109/ICOEI53556.2022.9777146.

Relja, R., Ward, P. and Zhao, A.L. (2024), "Understanding the psychological determinants of buy-now-pay-later (BNPL) in the UK: a user perspective", *International Journal of Bank Marketing*, Emerald Publishing, Vol. 42 No. 1, pp. 7–37, doi: 10.1108/IJBM-07-2022-0324/FULL/XML.

Rishi, B., Mallick, D.K. and Shiva, A. (2024), "Examining the dynamics leading towards credit card usage attitude: an empirical investigation using importance performance map analysis", *Journal of Financial Services Marketing*, Palgrave Macmillan, Vol. 29 No. 1, pp. 79–96, doi: 10.1057/S41264-022-00181-W/TABLES/12.

Schomburgk, L. and Hoffmann, A. (2023), "How mindfulness reduces BNPL usage and how that relates to overall well-being", *European Journal of Marketing*, Emerald Publishing, Vol. 57 No. 2, pp. 325–359, doi: 10.1108/EJM-11-2021-0923/FULL/HTML.

Seldal, M.M.N. and Nyhus, E.K. (2022), "Financial Vulnerability, Financial Literacy, and the Use of Digital Payment Technologies", *Journal of Consumer Policy*, Springer, Vol. 45 No. 2, pp. 281–306, doi: 10.1007/S10603-022-09512-9/TABLES/4.

Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J. and Sculley, D. (2017), "No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World".

Shy, O. (2023), "Cash Is Alive: How Economists Explain Holding and Use of Cash", *Journal of Economic Literature*, American Economic Association, Vol. 61 No. 4, pp. 1465–1520, doi: 10.1257/JEL.20221632.

Toti, J.F., Diallo, M.F. and Huaman-Ramirez, R. (2021), "Ethical sensitivity in consumers' decision-making: The mediating and moderating role of internal locus of control", *Journal of Business Research*, Elsevier, Vol. 131, pp. 168–182, doi: 10.1016/J.JBUSRES.2021.03.045.

Vamsee, C.S., Rakesh, D., Prathyusha, I., Dinesh, B. and Bharathi, C. (2023), "Demographic and Psychographic Customer Segmentation for Ecommerce Applications", *Proceedings of the 2nd International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2023*, Institute of Electrical and Electronics Engineers Inc., pp. 615–622, doi: 10.1109/ICAAIC56838.2023.10140861.

Yuping, Z., Jílková, P., Guanyu, C. and Weisl, D. (2020), "New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning", Atlantis Press, pp. 925–931, doi: 10.2991/AEBMR.K.200324.170.