Spending pattern visualization using unsupervised machine learning

Gabriel Porto Oliveira[1], **Jadson Castro Gertrudes**[2], **Roberta B. Oliveira**[1]

[1]Departamento de Ciência da Computação, Universidade de Brasília

[2]Departamento de Computação, Universidade de Ouro Preto

gportooliveira@icloud.com, jadson.castro@ufop.edu.br,

roberta.oliveira@unb.br

Abstract. *As the amount of financial data generated grows yearly, there is a growing need to leverage this data to develop customized financial products to meet individual users' unique needs and preferences. This study proposes a method for identifying potential spending patterns based on categorized financial transactions. Different clustering and outlier detection algorithms are compared using various internal validation metrics and empirical analysis of cluster balancing. A visualization of the spending patterns is created from the proposed method and validated by an expert in the domain in order to extract more insights based on user behavior. The visualization was found to be helpful when analyzing for insights into spending pattern.*

1. Introduction

The literature has extensively addressed the task of categorizing customers based on their financial transaction history [Jo-Ting et al. 2010][Li and Wu 2021]. The common Recency, Frequency and Monetary value (RFM) model is a common approach for segmentation, dividing customers into different groups. The RFM model uses the date of the last transaction (Recency), the number of transactions (Frequency) and the amount of money spent (Monetary value) to be used in customer segmentation. This data can be valuable for developing targeted marketing strategies or offering customized financial services to particular groups of users [Umuhoza et al. 2020]. Moreover, defining spending pattern profiles can optimize the development of marketing campaigns and potentially save resources for financial institutions [Shaw et al. 2001].

The introduction of OpenFinance[1], an initiative to share financial information between financial institutions and Pix[2], a fast payment mean that generates transaction data when compared to cash transactions has made Brazil resourceful place for gathering financial information. Furthermore, the excess of information also makes Brazil a good place to deploy new financial products.

Machine Learning (ML) refers to a field in which computers are able to learn without the need for explicit programming. This field encompasses several areas, one of them being unsupervised learning. In unsupervised learning, algorithms are designed to learn and present information based on the underlying structure of data [Mahesh 2020].

[1]https://www.bcb.gov.br/en/financialstability/open_finance.

[2]https://www.bcb.gov.br/estabilidadefinanceira/pix.

Clustering is one area within unsupervised ML. Clustering algorithms are used to identify groups of similar data points and group them together while separating dissimilar points into different clusters. To evaluate the quality of clustering results, internal validation metrics are used, which only consider the characteristics of clusters [Gan et al. 2007].

The majority of clustering algorithms studies only utilize information regarding total spending value and frequency of purchases or RFM model-based input [Hu et al. 2020][Umuhoza et al. 2020][Zakrzewska and Murlewski 2005]. This model has a limitation in that it overlooks the nature of spending for each transaction, which could potentially provide valuable information about customer spending behavior. Here, only categorized transaction data is used and clustering algorithms have been employed to identify spending patterns.

Clustering algorithms can present major issues due to the presence of outliers, which can negatively affect the results obtained from clusters by grouping together data that should not be with each other. Outlier points are typically data patterns that deviate from the characteristics found within normal data points [Liu et al. 2008]. To address such issues, outlier detection algorithms are used to identify such data points. In this particular domain, outliers could be linked to users with abnormal spending habits or data that was mistakenly added to databases.

Simply generating spending pattern profiles may not be sufficient to provide meaningful insights to experts. To solve this challenge, the use of graphical visualizations can be beneficial in facilitating the analysis of the generated profiles and extracting valuable insights. The main objective of this paper is to validate a spending pattern profile visualization created using clustering algorithms and categorized user transaction data. Moreover, an analysis of outlier detection and clustering algorithms is used according to [Oliveira 2023] to generate user segmentation.

This paper is organized as follows: Section 2 presents an overview of studies found in the domain; Section 3 details the spending pattern creation method and explains each step within it; Section 4 describes the experiments executed along with results found from the comparison of clustering and outlier detection algorithms, and Section 5 gives a conclusion based on the results found along with a proposal of future work to take place.

2. Literature Overview

This section provides an overview of the literature in the field of customer clustering/segmentation. Some works utilize the RFM model, while others use different types of input data, called non-RFM based. Both RFM and non-RFM based works have employed clustering algorithms to generate potential spending patterns. Some studies have found RFM to be insufficient as it only utilizes three values, thus limiting the amount of data used, and have expanded the model in some way [Allegue et al. 2020][Hu et al. 2020][Huang et al. 2020]. Others have adopted entirely different approaches due to the limitations exhibited by the RFM model [Wu and Lin 2005].

The RFM model has been widely used in the literature for the task of customer segmentation [Ernawati et al. 2021]. Usually, the measurements are taken within periods of time. The studies addressed in this section utilize the RFM model or use data

heavily influenced by it along with some clustering algorithms to better extract information. [Lefait and Kechadi 2010] proposed an architecture that uses the RFM model and k-Means algorithm to segment customers. Their work suggests a way of using RFM data to create several clusters and gain knowledge from a visual representation of selected clusters. Overall, their architecture supports the definition of the best clustering results for expert analysis and creates a visual representation to show customer segmentation.

On the other hand, [Umuhoza et al. 2020] used a dataset of credit card transactions from an Egyptian financial institution. The k-Means algorithm was applied along with validation metrics Silhouette Score and Calinski-Harabasz Score. The number of four clusters was found to be the best. Different types of customer behavior were determined based on the results. Using the generated clusters information, marketing strategies were defined for each of the spending profiles. [Holm 2017] utilized categorized transactions to create clusters in a dataset consisting of banking transactions with features divided into two categories: category statistics and general statistics. The former contained information on the nature of the transactions while the latter had general information such as the amount spent and the day of the week for the expense. The authors employed the Variance Threshold to eliminate features from the dataset. The k-Means and Hierarchical Clustering algorithms with metrics such as Calinski-Harabasz, Silhouette Score, and Dunn index were used for clustering. Additionally, they used credit scores to determine the best number of clusters. The study revealed that k-Means and Hierarchical Clustering algorithms using Ward linkage generated similar clusters when a small number of clusters were used.

To the best of our knowledge, techniques that utilize categorized transaction data have not been widely adopted and have only been observed in a few studies [Allegue et al. 2020][Holm 2017]. Categorized data has been demonstrated to provide more informative insights compared to the traditional RFM model approach [Allegue et al. 2020]. In addition, outlier detection techniques were not widely explored in previous studies, and the topic of outliers was only taken into consideration by [Zakrzewska and Murlewski 2005] when using clustering algorithms, but not with specific techniques to remove such data points. Therefore, this paper aims to investigate the combination of outlier detection algorithms with clustering algorithms. Moreover, a profile visualization method is introduced and tested by an expert in the area, the expert is the current CEO of the company the dataset originated from, possesses a Bachelor in Economics from Universidad San Ignacio de Loyola and has vast experience in the personal economics field.

3. Spending Pattern Generation Method

This paper aims to examine a spending profile visualization generated through a categorized transaction data approach. The data we have utilized is sourced from a private Peruvian financial technology company that operates in Brazil. The company's main objective is to serve the underbanked or unbanked population. The database is from September 2022 and consists of users' financial information containing incomes, expenses, investments, and more information related to financial habits. The database originates from users in Brazil and contains 433 users.

Here, creating the spending patterns involves four steps, which are clearly illus-

Figure 1. Illustration of the proposed method.

trated in Figure 1. These steps include data input, pre-processing, outlier detection, and cluster generation. The raw data is analyzed in the first step, and non-important data, such as email or phone number, is removed. The second step involves formatting the input data into a training-friendly format and mapping transaction categories. The third step aims to remove anomalies from the dataset to avoid adverse effects on the final results. In the last step, clustering algorithms are applied to the standardized data to generate clusters.

3.1. Data Pre-Processing

The original database is structured in a way that could not be used for clustering tasks. In the Data-Preprocessing step, original categories are mapped to new ones to remove redundant information and thus decreasing the number of features. Table 1 shows the original and new categories used, features are created using transaction data available, and users with less than five transactions (bills) are filtered out.

Table 1. Bill category mapping and your features.

Original Category	New Category	Feature
Household, Rent	Household	$n\_household$, $household\_total$
Entertainment	Entertainment	$n\_entertainment$, $entertainment\_total$
Transport, Car, Fuel, Parking, Transportation	Transport	$n\_transport$, $transport\_total$
Telephone	Telephone	$n\_telephone$, $telephone\_total$
Personal Care, Beauty, Fitness	Personal Care	$n\_personalCare$, $personalCare\_total$
Education	Education	$n\_education$, $education\_total$
Feeding, Dining, Groceries, Market	Feeding	$n\_feeding$, $feeding\_total$
Taxes	Taxes	n_taxes, $taxes\_total$
Health, Health Care	Health	n_health, $health\_total$
Bonuses, Transfers, Loans, Financing	Finances	$n\_finances$, $finances\_total$
Travel	Travel	n_travel, $travel\_total$
Shopping, Clothing	Shopping	$n\_shopping$, $shopping\_total$
Others	Others	n_others, $others\_total$
-	-	n_bills, $bills\_total$

Note: $n\_$ is number of transactions bills of a given category, $\_total$ is total spent in a given category.

Motivated by the many RFM-based studies discussed, the parameters Frequency (number of transactions) and Monetary Value (sum value of transactions) inspired the features used in this paper. The final dataset features consist of the number of expenses and the sum of all expenses of each newly mapped category, along with the total amount and the sum of expenses of each user. Here, the Recency value is not employed since several users stopped inserting data into the used mobile application in very different

intervals. Some users stopped using the application a few days after installation others used it for more time.

Based on the excessive number of redundant features, a category mapping process was needed to reduce the number of features. This process reduces the number of features in the dataset and removes redundant ones while keeping relevant categories. The final pre-processed dataset contains 109 users and 28 features. Each feature is explained in Table 1.

3.2. Optimization and Evaluation Process

We use clustering validation metrics to determine the best combinations of outlier detection and clustering algorithms, metrics that take into consideration the structure of the created clusters. The Silhouette Index (SI) was first introduced by [Rousseeuw 1987] and is scored in the interval $[-1, 1]$. The closer the score is to 1, the better the clustering result. Metric Calinski-Harabasz (CH) was presented by [Caliński and Harabasz 1974] and gave a score that the higher the clustering process is, the better. The final metric Davies-Bouldin (DB), created by [Davies and Bouldin 1979], determines that a score closer to 0 indicates a better clustering result.

Table 2 shows the parameters used for the clustering and outlier detection algorithms, the initial and final values used for testing, and the interval. Parameter b is the Bandwith parameter for the Mean-Shift algorithm. The parameter t corresponds to number of trees for the Isolation Forest algorithm and parameter n to number of neighbors for the Local Outlier Factor algorithm.

Table 2. Parameters used for clustering and outlier detection algorithms.

Algorithm	Parameter	Start Value	End Value	Step
Mean-Shift	b	0.15	2.2	0.05
k-Means	k	2	10	1
Bk-Means	k	2	10	1
Isolation Forest	t	50	100	2
Local Outlier Factor	n	10	50	1

3.3. Cluster Visualization

To assist experts in personal finance with identifying spending patterns, a cluster visualization method introduced by [Oliveira 2023] is adopted. To test the potential usefulness of the visualization method, we used a dataset as input to generate a graph following the method steps. A domain expert then analyzed the resulting cluster visualization to understand better and classify the significance of the spending pattern profiles created. The expert Julio Lavalle owns the Peruvia financial techonology company that provided the database and holds a bachelor's degree in Applied Sciences focusing on Economic and International Development. The cluster visualization comprises multiple graphs, one for each feature.

4. Results

This paper presents the results obtained with the best combinations of outlier detection and clustering algorithms. Furthermore, the spending profile visualization is presented. The outlier detection algorithms Isolation Forest and Local Outlier Factor

are used to help determine which data points within the data do not follow the same pattern within the rest of the dataset [Liu et al. 2008][Breunig et al. 2000]. The outlier detection algorithms along with three clustering algorithms, k-Means, Mean-Shift, and Bisecting k-Means, are compared and evaluated using internal validation metrics [Bock 2008][Di and Gou 2018][Fukunaga and Hostetler 1975]. It is important to note that the k-Means algorithm implemented in the SKLearn library uses the k-Means++ variant [Arthur and Vassilvitskii 2007].

4.1. Combination of Outlier Detectors and Clustering algorithms

In this section, the outlier detector and clustering algorithms are combined and an analysis of the effect the detection of outliers has on the clustering validation metrics is presented. Furthermore, the results found are then used to select the combination to generate the results in the profile visualization. Results with the Local Outlier Factor algorithm used a threshold value of 10. Table 3 demonstrates the best results achieved by using the outlier detector with each clustering algorithm, overall the best scores with this outlier detector and clustering algorithms resulted in very similar results. Instances with the best scores of SI and DB, are situations where there are only two clusters, one cluster containing several data points and the other only one. All clustering algorithms scored the same exact best values for all validation metrics. Furthermore, clusters created are more balanced with the k-Means and Bk-Means algorithms.

Table 3. Results with the LOF algorithm and each clustering algorithm.

Algorithm	No. Clusters	Bandwith	SI	CH	DB	No. Neighbors
k-Means	2	-	**0.7932**	37**.0483**	**0.1362**	33 - 50
Bk-Means	2	-	**0.7932**	**37.0483**	**0.1362**	33 - 50
Mean-Shift	2	1.05 - 2.10	**0.7932**	**37.0483**	**0.1362**	36 - 50

Note: best score in **bold**.

Table 4 displays the results of using the Isolation Forest anomaly detector in combination with clustering algorithms k-Means, Bk-Means, and Mean-Shift, respectively. Analyzing how balanced the resulting clusters, when k-Means was utilized with a value of $k = 2$, one cluster was formed with 23 data points, while the other cluster contained 79 data points.

Table 4. Best results with Isolation Forest and each clustering algorithm.

Algorithm	No. Clusters	Bandwith	SI	CH	DB	No. Trees
k-Means	2	-	**0.5608**	14.0976	**0.4520**	90
k-Means	4	-	0.3431	**17.6677**	1.3458	86 - 100
Bk-Means	2	-	0.4440	4.1475	**0.4187**	50 - 72
Bk-Means	3	-	**0.4638**	10.7888	1.2955	50 - 72
Bk-Means	5	-	0.1779	**13.8029**	1.6733	60 - 62
Mean-Shift	2	0.65 - 0.70	**0.5608**	7.5188	**0.3151**	90
Mean-Shift	4	0.50 - 0.55	0.5087	**14.0716**	0.6111	90

Note best score in **bold**.

4.2. Profile Visualization

When generating customer behavior clustering results, clusters with high SI and DB scores are not necessarily balanced, resulting in customer behavior graphs with limited meaningful information. To generate more informative graphs, a combination of highly balanced clusters with good clustering validation metrics can be used, which may result in more useful visualizations. Using Isolation Forest with the number of trees parameter 86 and clustering algorithm k-Means with $k = 4$. This combination generated metrics of CH of 17.6677, DB of 1.3458 and SI of 0.3431. Analyzing the number of data points in each cluster, cluster 0 contained 20 data points, cluster 1 with 20 data points and clusters 2 and 3 contained 75 and 3 data points respectively. The x axis of each graph represents the cluster number, while the y axis displays a box and whisker chart to illustrate the behavior of clusters for each feature. Figure 2 shows a visualization created using the parameters aforementioned.

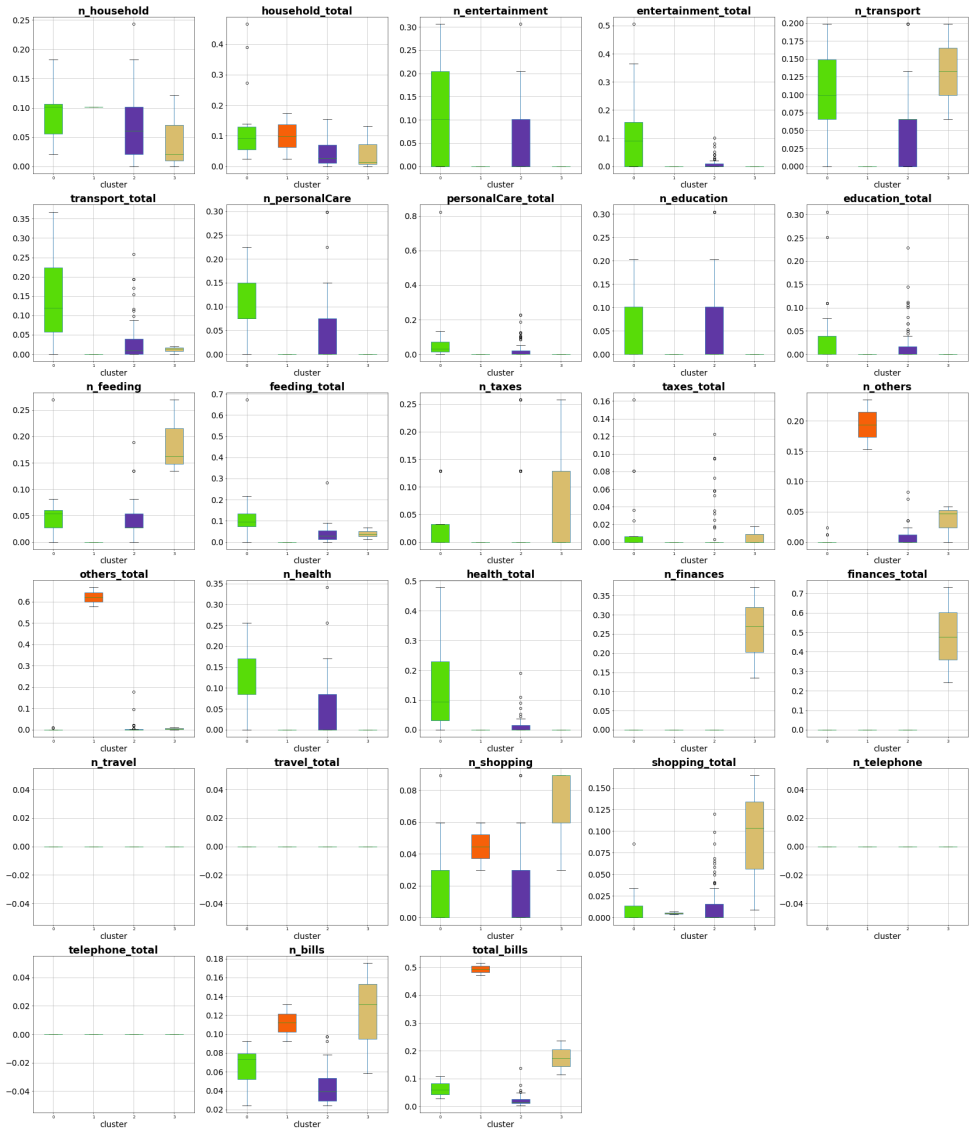

Figure 2. Graphs that show cluster behavior for each feature.

Every graph in the visualization present in Figure 2 contains a title, those beginning with the prefix $n\_$ relate to the number of transactions for a given feature, while those ending with the suffix $\_total$ relate to the total amount spent on that feature. The features n_bills and $bills\_total$ correspond to the total number of transactions and the total amount spent, respectively. The expert reviewed Figure 2 and provided an overview of each generated graph, as well as how individuals in different clusters behave in relation to one another, taking into account the available background information on the financial landscape of Brazil for under and unbanked individuals. The subsequent subsections detail the expert's analysis of some of the graphs in the visualization, initially the analysis would take into consideration behavior in all graphs but the expert decided to do a by graph analysis.

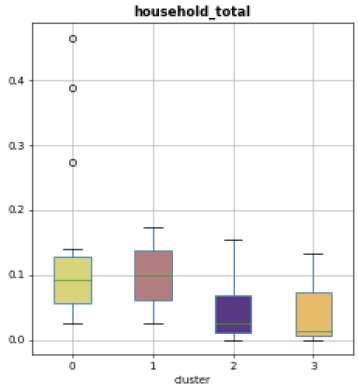

Figure 3. Cluster behavior related to $household\_total$ feature.

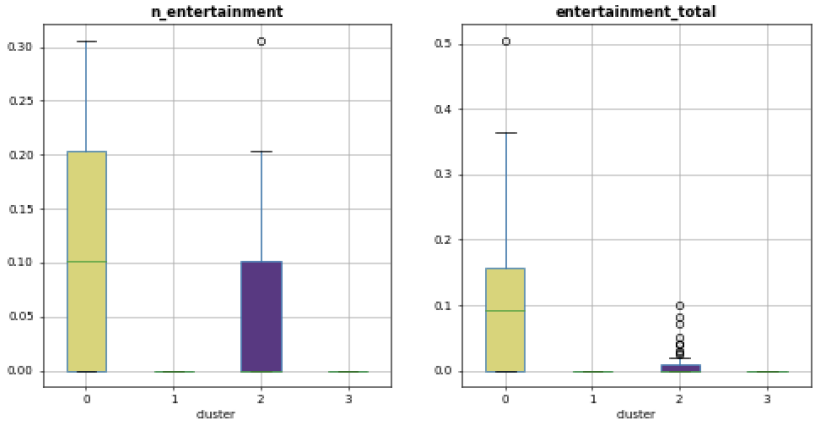

Figure 4. Cluster behavior for features $n\_entertainment$ and $entertainment\_total$.

According to the expert, analyzing the Graph $household\_total$ (Figure 3), users' spending behavior in all clusters is similar despite some outliers. The reported behavior is expected as household expenses could be similar in Brazil's unbanked/underbanked population. Customer spending behavior for Features $n\_entertainment$ and $entertainment\_total$, visible in Figure 4, can be interpreted as users in clusters 1 and 3 tend to have more stable expenses when analyzing household and entertainment-related features. This behavior could be related to the low income available for one or two cat-

egories. Alternatively, one could interpret clusters 1 and 3 as being more financially organized, with a more normalized spending pattern every month.

The total education spending is visible in Figure 5 with $education\_total$ feature behavior. Cluster 0 shows high variation in this feature. This behavior could be related to different school-aged children. On the other hand, Cluster 2 could be spending their money on personal or professional development other than school-aged children.

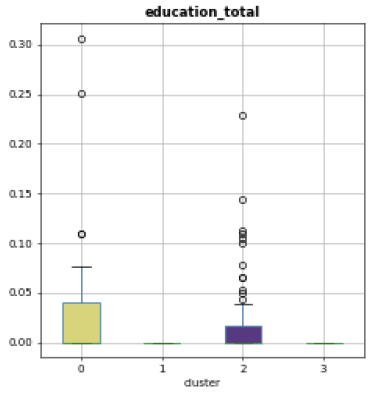

Figure 5. Cluster behavior related to feature $education\_total$.

Graphs n_taxes, $n\_finance$ and $finance\_total$ were analyzed, and the behavior of each was considered a graph when understanding customer behavior. Visualization of each graph is visible in Figure 6. According to the expert, cluster 3 shows higher spending with taxes. The higher spending taxes could indicate an increased need to pay financial fees due to unstable financial health. The identified behavior could be related to interest rates, particularly in Brazil, which has higher interest rates when compared to other countries.

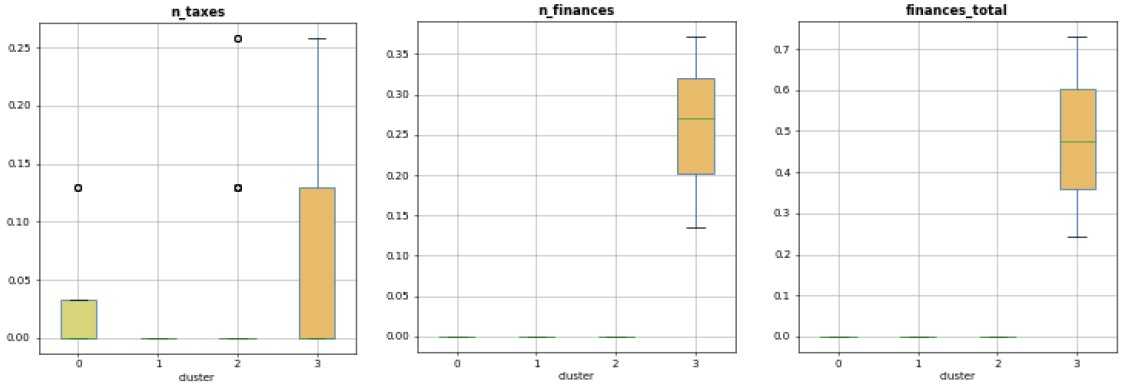

Figure 6. Cluster behavior related to features n_taxes, $n\_finance$ and $finance\_total$.

Graphs $n\_shopping$ and $shopping\_total$, seen in Figure 7, demonstrate spending behavior related to the shopping category. Users in clusters one and three have a higher number of transactions. The higher number of transactions could be interpreted as disorganized purchasing behavior due to a lack of up-front financial resources. Moreover, clusters zero and two show reduced purchasing power.

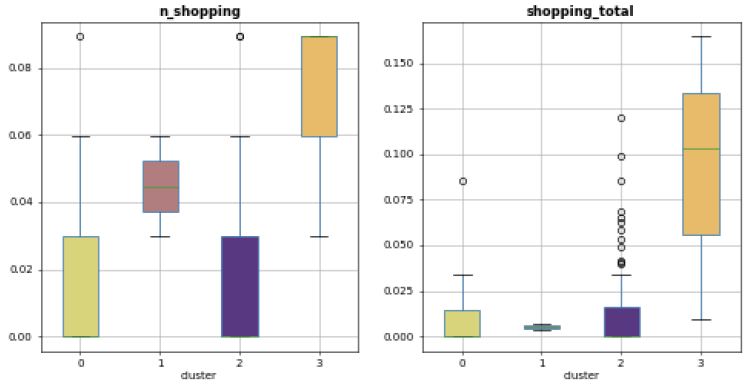

Figure 7. Cluster behavior related to features $n\_shopping$ and $shopping\_total$.

4.3. Discussion

The algorithms Isolation Forest and LOF were effective in identifying outliers. The Local Outlier Factor algorithm achieved higher validation metric results, but the results were generally less balanced. When considering the differences in the validation metrics, the Calinski-Harabasz index was better at finding combinations with more balanced clusters than the other metrics. Using the mentioned information, utilizing the CH index and the Isolation Forest outlier detection algorithm may be the best combination to find more balanced results.

Regarding the results of the clustering algorithms, the Bk-Means clustering algorithm generated a more balanced cluster than the standard k-Means but scored lower validation metrics results with both outlier detectors. The Mean-Shift algorithm, on the other hand, generally scored better validation metrics when utilizing the LOF outlier detector and similar results of the k-Means + Isolation Forest combination when using the same outlier detector. The spending profile visualization helped determine the behavior of users in different clusters. Regarding the visualization quality generated, the expert found the y-axis scale changing from one graph to another a problem. Furthermore, the initial idea was to analyze a cluster-by-cluster basis. The expert found that analyzing graph by graph would be a better approach.

5. Conclusion and Future Work

In this study, an already proposed method using different outlier detection and clustering algorithms has shown promising results when using categorized transaction data. Furthermore, the spending pattern profile visualization was tested by an expert in the area and is able to provide important information.

Internal validation metrics and cluster balancing analysis were used to identify the best combinations of outlier detection and clustering algorithms. Based on the final results, it was identified that better validation metrics alone did not result in the finest clustering results. However, their use with an empirical cluster balancing analysis were a positive way of generating clusters. Overall the more critical the validation metrics and the more balanced the clusters, the more information could be extracted from each cluster to be used in the spending pattern profile visualization. The proposed method was tested using only one dataset, further analysis with other categorized transaction data could help

improve the method steps and results.

A spending profile visualization was further verified with an expert in the area. The visualization was used to identify spending behavior generated with the proposed method and give more insightful information. The analysis done by the expert mainly focused on a small number of behavior graphs at a time. A further study analyzing spending behavior taking into consideration all categories could help find more insightful information. There is a need to improve further the profile visualization based on expert feedback and possible test variations, further analyses from other experts could help improve the method. Studying possible measurements or approaches of balanced clusters could help during the optimization step for future work. The addition of some performance metric-related balancing clusters is something to investigate.

6. Acknowledgments

The authors thank expert Julio Lavalle for helping interpret the behavior found in the spending pattern graphs. The authors thank FAP-DF for their aid through the project F2Dsys.

References

Allegue, S., Abdellatif, T., and Bannour, K. (2020). RFMC: a spending-category segmentation. In *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE.

Arthur, D. and Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA. Society for Industrial and Applied Mathematics.

Bock, H.-H. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Electronic Journal for History of Probability and Statistics*, 4(2):1–18.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227.

Di, J. and Gou, X. (2018). Bisecting k-means algorithm based on k-valued selfdetermining and clustering center optimization. *J. Comput.*, 13(6):588–595.

Ernawati, E., Baharin, S. S. K., and Kasmin, F. (2021). A review of data mining methods in RFM-based customer segmentation. *Journal of Physics: Conference Series*, 1869(1):012085.

Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40.

Gan, G., Ma, C., and Wu, J. (2007). In *Data Clustering: Theory, Algorithms, and Applications*, pages 299–320. Society for Industrial and Applied Mathematics.

Holm, M. (2017). Machine learning and spending patterns: A study on the possibility of identifying riskily spending behaviour. Master's thesis, KTH Royal Institute of Technology.

Hu, X., Shi, Z., Yang, Y., and Chen, L. (2020). Classification method of internet catering customer based on improved rfm model and cluster analysis. In *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pages 28–31. IEEE.

Huang, Y., Zhang, M., and He, Y. (2020). Research on improved RFM customer segmentation model based on k-means algorithm. In *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*. IEEE.

Jo-Ting, W., Shih-Yen, L., and Hsin-Hung, W. (2010). A review of the application of rfm model. *African Journal of Business Management*, 4(19):4199–4206.

Lefait, G. and Kechadi, T. (2010). Customer segmentation architecture based on clustering techniques. In *2010 Fourth International Conference on Digital Society*. IEEE.

Li, H. and Wu, W. (2021). Construction of chinese national geography APP user operation strategy based on RFM model. In *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*. IEEE.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386.

Oliveira, G. P. (2023). A method for defining customer spending behavior based on unsupervised machine learning. Bachelor's thesis, Universidade de Brasília.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Shaw, M. J., Subramaniam, C., Tan, G. W., and Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1):127–137. Knowledge Management Support of Decision Making.

Umuhoza, E., Ntirushwamaboko, D., Awuah, J., and Birir, B. (2020). Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in africa. *SAIEE Africa Research Journal*, 111(3):95–101.

Wu, J. and Lin, Z. (2005). Research on customer segmentation model by clustering. In *Proceedings of the 7th international conference on Electronic commerce - ICEC '05*. ACM Press.

Zakrzewska, D. and Murlewski, J. (2005). Clustering algorithms for bank customer segmentation. In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*. IEEE.