REVISTA DE GESTÃO E SECRETARIADO

ISSN: 2178-9010
DOI: http://doi.org/10.7769/gesec.v15i11.4142

An approach for K-modes cluster analysis to examine direct resource transfers in brazilian public agreements for innovation and rural development

Uma abordagem para análise de *K-modes clusters* para examinar transferências diretas de recursos em acordos públicos brasileiros para inovação e desenvolvimento rural

Un enfoque de análisis de *K-modes clusters* para examinar las transferencias directas de recursos en los acuerdos públicos brasileños para la innovación y el desarrollo rural

Alan Keller Gomes [1]

Márcio Dias de Lima [2]

Paulo Henrique dos Santos [3]

Cassiomar Rodrigues Lopes [4]

Lucas Santos de Oliveira [5]

Daniel Soares de Souza [6]

José Carlos Barros Silva [7]

Karla de Aleluia Batista [8]

Abstract

This article presents an innovative approach to K-modes cluster analysis applied to public

[1] PhD in Computer Sciences. Instituto Federal de Goiás. Inhumas, Goiás, Brazil.
E-mail: alan.gomes@ifg.edu.br Orcid: https://orcid.org/0000-0002-4073-0388
[2] PhD in Computer Sciences. Instituto Federal de Goiás. Goiânia, Goiás, Brazil.
E-mail: marcio.lima@ifg.edu.br Orcid: https://orcid.org/0000-0003-2782-386X
[3] MsC in Production Engineering and Systems. Instituto Federal de Goiás. Senador Canedo, Goiás, Brazil.
E-mail: paulo.santos1@ifg.edu.br Orcid: https://orcid.org/0000-0002-6027-3022
[4] PhD in Agribusiness. Instituto Federal de Goiás. Anápolis, Goiás, Brazil.
E-mail: cassiomar.lopes@ifg.edu.br Orcid: https://orcid.org/0009-0008-9075-4127
[5] Specialist in Applied Statistics. Universidade Pitagóras Unopar Anhanguera. Brasília, Distrito Federal, Brazil.
E-mail: lucas.oliveira4@estudante.ifb.edu.br Orcid: https://orcid.org/0000-0002-7128-8545
[6] MsC in Public Management. Instituto Federal de Brasília. Brasília, Distrito Federal, Brazil.
E-mail: daniel.souza@ifb.edu.br Orcid: https://orcid.org/0000-0003-2210-5412
[7] MsC in High Education Sciences. Instituto Federal de Goiás. Luziânia, Goiás, Brazil.
E-mail: josecarlos.silva@ifg.edu.br Orcid: https://orcid.org/0009-0003-1476-8352
[8] PhD in Biological Sciences. Instituto Federal de Goiás. Goiânia, Goiás, Brazil.
E-mail: karla.batista@ifg.edu.br Orcid: https://orcid.org/0000-0003-4396-032X

agreements signed by the Secretariat of Innovation, Rural Development, and Irrigation (SDI) of the Ministry of Agriculture, Livestock, and Supply (MAPA) of Brazil. The agreements were signed between 2019 and early 2023. The main goal is to identify patterns and trends in resource transfers for innovation and sustainable rural development in Brazil. This study analyzed 10,098 agreements using the Design Science Research Methodology, refined over three cycles. The K-modes method is particularly suitable for handling categorical data. It enables the clustering of agreements based on similar characteristics, such as geographic region served, purpose, the year the agreement was signed, the total value of expense items, and the status of the agreement. The results demonstrate that the K-modes approach overcomes typical limitations of traditional clustering methods, including sensitivity to outliers, restriction to numerical data, and difficulty handling clusters of varying sizes and densities. Additionally, it addresses issues related to the lack of interpretability of the generated clusters. This study advances the application of the K-modes method in analyzing direct resource transfer mechanisms for innovation and rural development, a still unexplored area. The proposed approach can be generalized to examine direct resource transfer mechanisms in different countries, contributing to enhancing public policies on a global scale. Despite structural differences, resource transfers aimed at innovation and rural development aim to direct public funds to initiatives that improve living conditions and the agricultural sector's competitiveness.

Keywords: K-modes Clustering. Direct Transfers. Public Resources. SDI/MAPA Agreements. DSRM.

Resumo

Este artigo apresenta uma abordagem inovadora para análise de cluster K-modes aplicada a acordos públicos assinados pela Secretaria de Inovação, Desenvolvimento Rural e Irrigação (SDI) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) do Brasil. Os acordos foram assinados entre 2019 e o início de 2023. O objetivo principal é identificar padrões e tendências em transferências de recursos para inovação e desenvolvimento rural sustentável no Brasil. Este estudo analisou 10.098 acordos usando a Metodologia de Pesquisa em Design Science, refinada em três ciclos. O método K-modes é particularmente adequado para lidar com dados categóricos. Ele permite o agrupamento de acordos com base em características semelhantes, como região geográfica atendida, finalidade, ano em que o acordo foi assinado, valor total dos itens de despesa e status do acordo. Os resultados demonstram que a abordagem K-modes supera limitações típicas de métodos tradicionais de agrupamento, incluindo sensibilidade a outliers, restrição a dados numéricos e dificuldade em lidar com clusters de tamanhos e densidades variados. Além disso, ele aborda questões relacionadas à falta de interpretabilidade dos clusters gerados. Este estudo avança a aplicação do método K-modes na análise de mecanismos de transferência direta de recursos para inovação e desenvolvimento rural, uma área ainda inexplorada. A abordagem proposta pode ser generalizada para examinar mecanismos de transferência direta de recursos em diferentes países, contribuindo para aprimorar políticas públicas em escala global. Apesar das diferenças estruturais, as transferências de recursos voltadas para inovação e desenvolvimento rural visam direcionar recursos públicos para iniciativas que melhorem as condições de vida e a competitividade do setor agrícola.

Palavras-chave: *K-modes Clustering*. Transferências Diretas. Recursos Públicos. Acordos SDI/MAPA. DSRM.

Resumen

Este artículo presenta un enfoque innovador para el análisis de conglomerados de K-modes aplicado a los acuerdos públicos firmados por la Secretaría de Innovación, Desarrollo Rural e Irrigación (SDI) del Ministerio de Agricultura, Ganadería y Abastecimiento (MAPA) de Brasil. Los acuerdos se firmaron entre 2019 y principios de 2023. El objetivo principal es identificar patrones y tendencias en las transferencias de recursos para la innovación y el desarrollo rural sostenible en Brasil. Este estudio analizó 10.098 acuerdos utilizando la Metodología de Investigación de la Ciencia del Diseño, perfeccionada en tres ciclos. El método K-modes es particularmente adecuado para el manejo de datos categóricos. Permite la agrupación de acuerdos en función de características similares, como la región geográfica atendida, el propósito, el año en que se firmó el acuerdo, el valor total de los ítems de gastos y el estado del acuerdo. Los resultados demuestran que el enfoque K-modes supera las limitaciones típicas de los métodos de agrupamiento tradicionales, incluida la sensibilidad a los valores atípicos, la restricción a los datos numéricos y la dificultad para manejar conglomerados de diferentes tamaños y densidades. Además, aborda cuestiones relacionadas con la falta de interpretabilidad de los conglomerados generados. Este estudio propone una aplicación del método K-modes para analizar los mecanismos de transferencia directa de recursos para la innovación y el desarrollo rural, un área aún inexplorada. El enfoque propuesto puede generalizarse para examinar los mecanismos de transferencia directa de recursos en diferentes países, contribuyendo así a mejorar las políticas públicas a escala global. A pesar de las diferencias estructurales, las transferencias de recursos destinadas a la innovación y el desarrollo rural tienen como objetivo dirigir fondos públicos a iniciativas que mejoren las condiciones de vida y la competitividad del sector agrícola.

Palabras clave: *K-modes Clustering*. Transferencias Directas. Recursos Públicos. Acuerdos SDI/MAPA. DSRM.

Introduction

This study analyzes and clusters data on public agreements signed by the Secretariat of Innovation, Rural Development, and Irrigation (SDI) of the Ministry of Agriculture, Livestock, and Supply (MAPA) of Brazil between 2019 and early 2023. MAPA plays an essential role in promoting the sustainable development of Brazilian agribusiness, ensuring food security, and enhancing the competitiveness of agricultural products (MAPA, 2023c). SDI is dedicated to fostering an environment conducive to innovation in the agricultural sector by allocating resources for implementing technological and operational innovations (Brasil, 2023).

Many previous studies have examined the mechanisms of direct transfers through public agreements, which are formal contracts established between MAPA and governmental or non-governmental organizations. These mechanisms aim to promote innovation and agricultural development and are typically researched through the analysis of data clusters, primarily generated using the K-means method (Costa et al., 2012), (Borges et al., 2013), (Silvério, 2019), (Alvares and Branco, 2018) and (Tang et al., 2017).

Recent research has highlighted the limitations of the K-means clustering method and proposed various strategies to overcome these limitations (Tang et. al, 2017) (Tian et al., 2019), such as sensitivity to the initialization of centroids, the need to pre-specify the number of clusters, sensitivity to outliers, limitation to numerical data, difficulty in dealing with clusters of different sizes and densities (i.e., the challenge of explaining data dispersion in each group), lack of interpretability of clusters, and limitations in capturing non-linear relationships.

The application of the K-modes clustering method to examine direct resource transfer mechanisms through public agreements, specifically in the context of innovation and rural development, has remained unexplored to date. This article presents an approach for cluster analysis of data generated by the K-modes method to investigate these mechanisms. In our approach, categorical data is used instead of numerical data. Clusters of different sizes and densities are explained based on information such as the number of elements, the number of distinct points in each cluster, and a spread measure that correlates the two previous pieces of information. These treatments facilitate the clusters' interpretability.

This study is essential to understand how direct transfers via public agreements contribute to implementing policies aimed at sustainable development and technological innovation in agriculture. The proposed approach, based on the K-modes method, allows us to overcome some of the limitations of the K-means method, offering a more robust and interpretable analysis of patterns and trends in the partnerships established between SDI/MAPA and various governmental and non-governmental entities. By filling this gap and proposing an innovative approach, the study contributes to improving the examination of mechanisms for promoting innovation in the agricultural sector and advancing public policies in this area.

This research presents an approach for analyzing K-modes (Huang, 1998a) clusters in examining direct transfers via public agreements aimed at innovation and rural development in Brazil. Specifically, it seeks to understand how these transfers contribute to implementing public policies for sustainable development and technological innovation in agriculture, identifying patterns and trends in established partnerships.

The methodology used is Design Science Research Methodology (DSRM) (Peffers et al., 2007), with a refinement of the approach in three interaction cycles. K-modes cluster analysis is applied to deal with categorical data, clustering the agreements (Peffers et al., 2007) based on similar characteristics, such as geographic region served, purpose, the year the agreement was signed, the total value of the expense items, and the agreement status. Data from 10,098 agreements signed between January 2019 and March 2023 were analyzed. Python was utilized to clean, transform, and organize the data, applying the K-modes method and visualizing results in a format accessible to non-experts.

Despite the structural differences between the mechanisms of direct transfer of resources for innovation and rural development in different countries, the approach proposed in this study can be generalized to examine international scenarios where the goal is to direct public resources to promote initiatives that improve living conditions and the competitiveness of the agricultural sector. Applying K-modes cluster analysis allows us to identify common patterns and particularities of each mechanism, contributing to improving public policies on a global scale. Adapting the approach to different international contexts depends on observing each country's specificities, such as the legal framework, institutional structures, and rural development priorities, issues that go beyond the scope of this article.

The article is organized as follows: Section 2 presents previous studies addressing correlated issues, offering a solid basis for understanding the context and applied techniques; Section 3 details the application of DSRM in constructing the proposed approach; Section 4 presents the results obtained and Section 5 discusses the results; finally, Section 6 concludes the article, followed by the references.

Related Work

Several studies explore mechanisms for direct resource transfers and strategic partnerships between different government levels and municipalities within the Brazilian context. Research works by Silvério (Silvério, 2019) and Lima (2005) highlight that Voluntary Transfers (VTs), an important mechanism for transferring resources from the federal government to states and municipalities, can foster innovation and improve health conditions in municipalities. Another relevant mechanism is the intermunicipal consortium, which forms partnerships between municipalities for cooperation and regional development. Alvares and Branco (2018) discuss the role of consortia in attracting federal resources and promoting regional development.

Furthermore, strategic partnerships between different government levels are essential for effectively implementing public policies. The study by Parente (2019) shows that transfers from state governments to municipalities can be influenced by political factors, highlighting the importance of objective and transparent criteria in resource allocation.

The mechanisms of VTs, intermunicipal consortiums, and strategic partnerships are implemented to finance priority projects and actions that contribute to promoting innovation and sustainable agricultural development in Brazil. The potential of direct resource transfers (VTs, intermunicipal consortiums, strategic partnerships) for effectively implementing public policies is addressed in various research. Costa et al. (2012) used hierarchical cluster analysis to characterize municipalities in the Brazilian state of Minas Gerais based on socioeconomic conditions, public finances, and economic activity. Borges et al. (2013)

applied hierarchical cluster analysis to analyze the debt of Brazilian states and the Federal District, verifying the reduction in public debt after the enactment of the Fiscal Responsibility Law (Brasil, 2000).

Silvério (Silvério, 2019) employed hierarchical cluster analysis to examine the adherence of agreements from the Brazilian Ministry of Science, Technology, and Innovation (MCTIC) to the priority actions of the National Science, Technology, and Innovation Strategy (ENCTI) and proposed a typology for VTs. Alvares and Branco (2018) used cluster analysis to investigate the acquisition of federal resources by the municipalities of the Cioeste consortium, noting socioeconomic differences and the consortium's impact on resource acquisition. Parente (2019) used cluster analysis to examine the political cycle and the effect of alliances on transfers from the state of Ceará to municipalities, identifying a higher probability of transfers in election years and a preference for politically aligned local governments.

It is important to note that the cluster analyses in the studies (Costa et al., 2012), (Borges et al., 2013), (Silvério, 2019), (Alvares and Branco, 2018), (Tang et al., 2017), and (Parente, 2019) were conducted using statistical data clustering methods. However, these authors do not explore the impact of these methods on examining the effectiveness of implementing public policies via direct transfer mechanisms.

Some studies have addressed the use of the K-means method for examining the effectiveness of implementing public policies via direct resource transfer mechanisms. However, the advantages and disadvantages of using this method are not explicitly detailed. In Lima's (2005) work, cluster analysis with the K-means method is used to classify municipalities based on health, sociodemographic, and financial indicators, identifying different levels of health expenditure. Crispim et al. (2020) used K-means cluster analysis to investigate municipalities' expenses, investments, and loans during election periods, noting increased spending on investments and loans in election years. Cruz and Silva (2020) employed K-means cluster analysis to evaluate the impact of budget structure on education spending in Brazilian municipalities, finding that the minimum spending rule increased education spending for municipalities below the minimum requirement.

Methodological Procedures

This section outlines the study's methodological procedures. The first subsection discusses the data's origin and the steps taken to ensure its quality and reliability, including data cleaning, outlier treatment, standardization, and adjustments via cross-validation. The second subsection introduces the Design Science Research Methodology (DSRM) (Peffers et al., 2007), an iterative and interactive process consisting of six steps. The third subsection details the application of DSRM in constructing the proposed approach for analyzing data clusters generated by the K-modes method across three cycles.

3.1 Data Origin and Processing

The data used in this study were obtained from reliable secondary sources, specifically the Brazilian government's Public Grant and Transfer Agreements Management System (SICONV) (MAPA, 2023a), maintained by the Ministry of Agriculture, Livestock, and Supply (MAPA). SICONV records and manages information about agreements, transfer contracts, and partnership terms entered by the federal public administration and other public or private non-profit agencies or entities.

The analyzed data covers 10,098 agreements signed between MAPA's Secretariat for Innovation, Rural Development, and Irrigation (SDI) and various governmental and non-governmental entities from January 2019 to March 2023. The dataset includes information on the geographic region served, the purpose of the agreement, the year of the agreement, the total value of the expense item, and the status of the agreement.

The data's reliability and suitability are ensured by the official nature of SICONV, which is a central repository of information on agreements signed by the federal public administration. The system is subject to regular audit and control procedures, ensuring the integrity and consistency of the recorded data (Brasil, 2008).

Before being used in our analysis, the data underwent a processing stage involving cleaning and treatment steps using Python (PSF, 2024). This process involved identifying and handling missing values, identifying and treating outliers, and standardizing the data for categorical format. Additionally, cross-validation techniques were applied to identify and correct possible inconsistencies in the data.

It is also worth mentioning the current technological limitations of Python resources (PSF, 2024) in cleaning, transforming, and organizing data, applying the K-modes method, and plotting results on graphs. These limitations influenced the choice of resources that favor the visualization of results in a more accessible and understandable format for laypeople rather than more advanced visual resources, such as matrix graphics (heatmaps) or three-dimensional plots.

The clusters generated by the K-modes method are formed from the cross-referencing of characteristics (Dong and Pei, 2007) (Ahmed and Pathan, 2018) found in the data from 10,098 agreements. These characteristics include ‘State,’ which are associated with the geographic area served; ‘Nature Expense,’ indicating the purpose of the agreement; ‘Agreement Year,’ denoting the year of the agreement; ‘Total Item Value’, representing the total expense value of the agreement; and ‘Agreement Situation,’ reflecting the status of the agreement within the SDI/MAPA framework.

The characteristics (columns) are analyzed in pairs, with the data already in categorical format. Each attribute-value pair from one column is combined with each attribute-value pair from the other. The frequency of each pair combination is calculated and displayed on a bubble chart, where the size of each bubble corresponds to the frequency of occurrence. The data in each column pertains to a public agreement for transferring resources from MAPA to partner entities. Data from 10,098 agreements signed between January 2019 and March 2023 were analyzed. To manage outliers in columns like ‘Total Item Value’, ‘Agreement Situation,’ and ‘Nature Expense,’ a category was created in each column to group outliers into an attribute-value pair.

This strategy reduces the sensitivity of the K-modes method to outliers. Originally, the ‘Total Item Value’ column contained numeric values. Categorizing this data into value ranges and highlighting outliers helps address the limitations of using numerical data for generating clusters and facilitates using the K-modes method. These strategies also improve the interpretability of the generated clusters and highlight the presence of outliers.

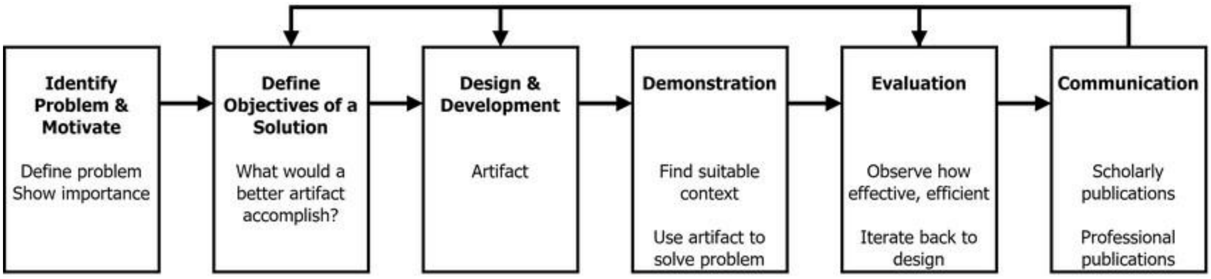
The data groupings that emerge from cross-referencing (Dong and Pei, 2007) (Ahmed and Pathan, 2018) each pair of columns are displayed in a scatter plot, with each cluster represented by a different geometric figure. The key includes the number of elements in each cluster (Element Numbers), the number of distinct points in the cluster plot (Distinct Points), and a measure of the spread of each cluster that correlates the number of elements with the number of distinct points (Spread). Spread measures the relative distribution of points within different clusters on the graph. A higher spread value indicates more dispersed elements within a cluster. A lower spread value indicates more concentrated elements within a cluster.

3.2 Applying DSRM to Build the Proposed Approach

In this article, the DSRM (Peffers et al., 2007) as illustrated in Figure 1, organizes the proposed approach for analyzing data clusters, refining the process through three cycles of interaction and iteration.

Figure 1

DSRM Iterative and Interactive Process Model


Identify Problem & Motivate
Define problem
Show importance
Define Objectives of a Solution
What would a better artifact accomplish?
Design & Development
Artifact
Demonstration
Find suitable context
Use artifact to solve problem
Evaluation
Observe how effective, efficient
Iterate back to design
Communication
Scholarly publications
Professional publications

Source: adapted from Peffers et al., (2007).

Each cycle builds upon the previous one, enhancing the approach's robustness and depth. In the first cycle (Section 3.2.1), the initial mapping of our approach into the stages of the DSRM process is presented. This foundational cycle sets the groundwork for subsequent refinement.

The second cycle (Section 3.2.2) refines the initial approach by incorporating crossing data and analyzing the frequency of occurrences resulting from these intersections. This cycle is mapped into the corresponding stages of the DSRM process, ensuring a more detailed and accurate analysis.

The third cycle (Section 3.2.3) further refines the second cycle by focusing on crossing data and analyzing the data clustering generated by the K-modes method. This cycle is also mapped into the respective stages of the DSRM process, enhancing the comprehensiveness of the approach. Each of these cycles, along with their mapping into the steps of the DSRM process, are described in detail below.

3.2.1 First cycle: mapping the proposed approach in the DSRM stages

1. **Identify the Problem and Motivation:** The research problem is defined as the need to understand how directing transfers through public agreements in Brazil contribute

to implementing policies aimed at the sustainable development of the rural sector and technological innovation in agriculture. The motivation is to gain insights into the effectiveness of these partnerships in promoting rural development and innovation.

2. **Define the Solution's Objective:** The general objective is to perform data crossing (Dong and Pei, 2007) (Ahmed and Pathan, 2018) of the columns 'Agreement Year,' 'Total Item Value,' 'Region,' 'State,' 'Agreement Situation,' and 'Nature Expense.' This data crossing aims to identify patterns and trends in the partnerships established between SDI/MAPA and various governmental and non-governmental entities.
3. **Design and Develop the Artifacts:** In this step, the developed artifact is the dataset with the treatments mentioned in Section 3.1. Data treatment involved identifying and handling missing values, identifying and treating outliers, and standardization for the categorical data format. Additional artifacts were developed from the processed data in subsequent cycles, focusing on the frequency of occurrence and the grouping of data that emerge from data crossing.
4. **Demonstrate:** The artifacts are demonstrated using Python (PSF, 2024). Initially, cross-validation techniques are applied to identify and correct possible inconsistencies in the data. Subsequently, the frequency of occurrence and the clustering of data emerging from data crossing are displayed in graphs, specifically bubble charts and scatter plots.
5. **Evaluate:** The quantitative evaluation of each artifact is conducted by analyzing the results presented in the graphs and their keys. Additionally, a qualitative analysis of the formed clusters is performed to identify patterns and generate new knowledge about direct resource transfers and their impact on rural development and innovation.
6. **Communicate:** The research results and contributions are communicated through this article, sharing the knowledge generated with the academic and professional community.

3.3.2 Second cycle: crossing data for frequency analysis

1. **Identify the Problem and Motivation:** The frequency of each attribute-value pair from one column to another is analyzed through data crossing. The motivation is to

obtain an overview of patterns in the data, identifying the most frequent and potentially relevant combinations of attributes.

2. **Define the Solution's Objectives:** This cycle's objective is to cross-reference information from the agreement columns, calculating the frequency of each combination of pairs. This allows for an initial analysis of patterns in the data.
3. **Design and Develop the Artifact:** In this step, the agreement data columns prepared in the previous cycle are paired two by two. Next, a script is developed to cross-reference the data and calculate the frequency of occurrence of each combination of pairs.
4. **Demonstrate:** The demonstration in this cycle involves executing the developed script using the agreement data. The results are presented in bubble charts, where the size of each bubble represents the frequency of each combination of attribute pairs.
5. **Evaluate:** Evaluation in this cycle focuses on interpreting the generated bubble charts. The most frequent combinations of pairs are analyzed to gain new insights into the distribution of attributes in agreements. Potentiallyrelevant combinations are also identified for further analysis in the next cycle. The interpretability and usefulness of the results are considered.
6. **Communicate:** The results of this cycle are documented to be included in the next cycle of interaction, highlighting the key insights obtained from the bubble chart analysis. These results are communicated internally to the research team to be incorporated into the final scientific article.

3.3.3 Third cycle: data crossing for construction of K-modes clusters

1. **Identify the Problem and Motivation:** The problem identified in this cycle is the need to cluster agreements based on combinations of attributes to uncover patterns and more complex structures in the data. The motivation is to gain a detailed understanding of the similarities and differences between the agreements.
2. **Define the Solution's Objective:** This cycle aims to apply the K-modes clustering method (Hautam"aki et al., 2005) to the agreement data, considering the intersections of the relevant column pairs. This will enable the identification of clusters of agreements with similar characteristics.

3. **Design and Develop the Artifact:** In this step, the data is formatted according to the requirements of the K-modes method. The K-modes algorithm is then executed with an initial number of clusters set to three and other relevant parameters.
4. **Demonstrate:** The demonstration in this cycle involves running the K-modes method on the agreement data. The results are presented in scatter plots, with each cluster represented by a distinct geometric figure. Legends on each scatter plot display the number of elements in each cluster (Element Numbers), the number of times each type of figure appears in the cluster plot (Distinct Points), and a measure of the spread of each cluster that correlates the number of elements with the number of distinct points in each cluster (Spread). The spread measure helps understand the relative distribution of points within different clusters: the higher the value, the more dispersed the elements; the lower the value, the more concentrated the elements. At this stage, the clusters generated by K-modes are visualized and interpreted.
5. **Evaluate:** The evaluation of the artifact in this cycle considers the quality and interpretability of the generated clusters. Cluster evaluation metrics such as silhouette scores and purity can be used. Additionally, feedback from experts is adopted to validate the relevance of the clusters.
6. **Communicate:** The results of this cycle are documented in detail, including the interpretation of the formed clusters and the valuable insights obtained. The research highlights patterns, differences between clusters, and potential implications for policies related to public agreements. These results are incorporated into the final scientific article, along with the findings from the previous cycle.

Results: K-modes Clustering

The results obtained from data crossing are presented in this section in the following order: 1) State vs. Agreement Year, 2) State vs. Nature-Expense, 3) State vs. Total Item Value e 4) State vs. Agreement Situation.

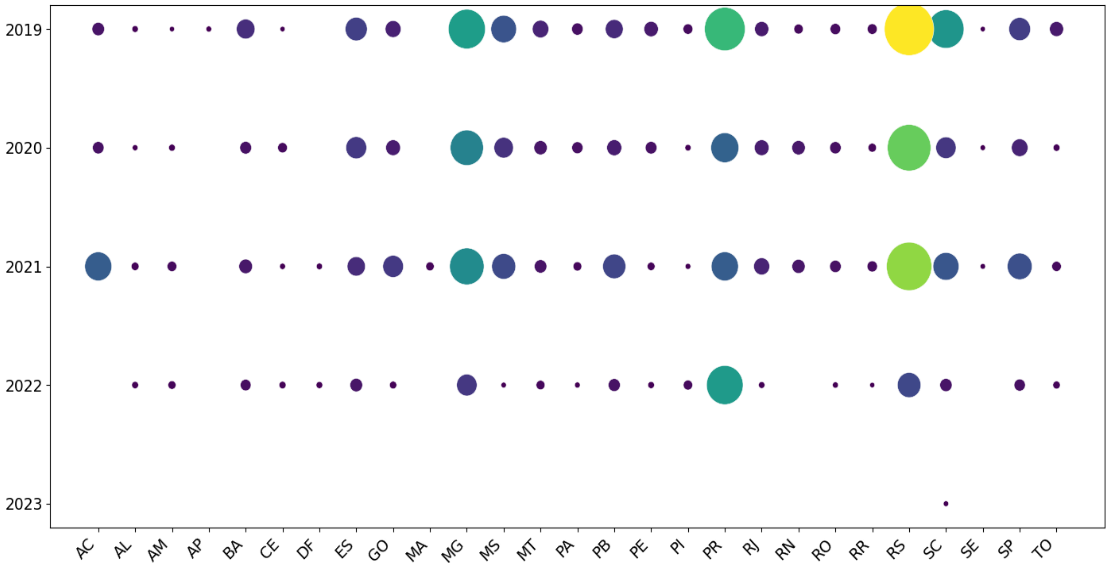
This order allows us to start with an overview of the distribution patterns of agreements by region and state over the years, followed by a more detailed analysis of the areas of activity and the nature of expenses. Next, partnerships and the allocation of resources by value range are discussed. Finally, the situations of the agreements and their relationship with other variables are analyzed. This sequence facilitates a gradual and in-depth understanding of the identified patterns, providing valuable insights for public policies and future actions.

4.1 Frequency and K-modes Clusters: State vs. Agreement Year

Data cross-referencing between the State and Agreement Year columns is presented in Figures 2 and 3. The State data corresponds to each of the 27 states represented on the Y-axis by their respective official acronyms. The Agreement Year data spans from 2019 to 2023.

Figure 2

Bubbles Chart: Frequency - State vs. Agreement Year



Source: Research data

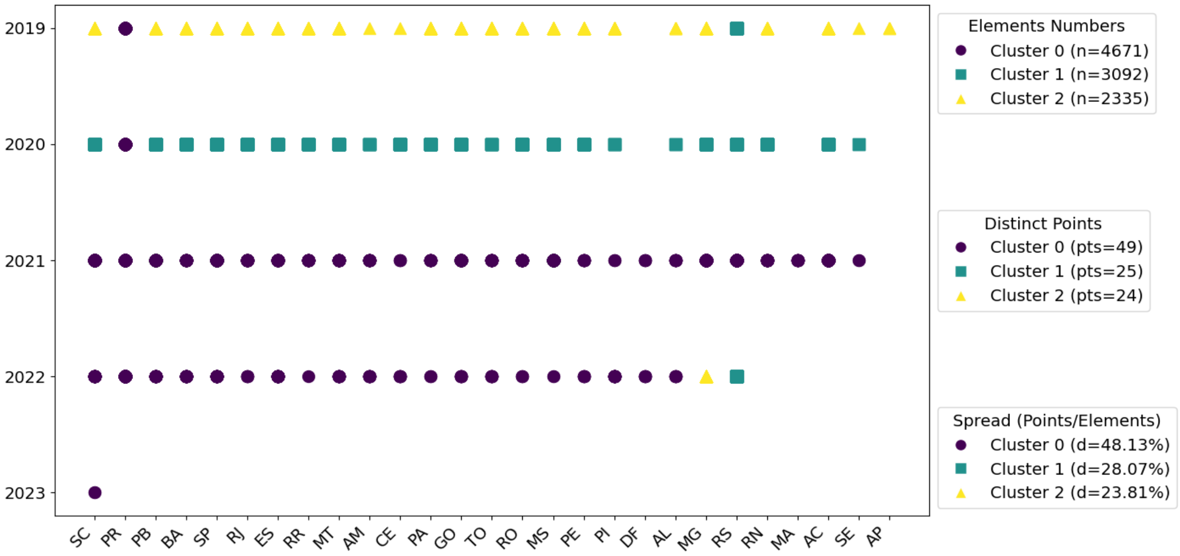
Figure 2 presents the number of agreements signed associated with the frequency of each pair, which is represented by the size of each circle. It can be observed that the states of MG, PR, and RS stand out more frequently than the other states from 2019 to 2023. In 2019, DF and MA did not sign any agreements. In 2020, in addition to DF and MA, the state of AP also did not sign an agreement. In 2021, AP did not sign any agreements, while the number of agreements in AC is notable, especially in the country's northern region. In 2022, the states of AC, AP, MA, RN, and SE did not sign any agreements. In 2023, the only state that signed an agreement was SC.

Figure 3 shows the clusters formed by crossing data from the State vs. Agreement

Year columns. Each cluster corresponds to a geometric form, with each element representing an agreement to transfer resources from MAPA to entities of the associated state.

Figure 3

Scatter Plot: K-modes Clusters - State vs. Agreement Year



Source: Research data

The agreements are grouped into three clusters, indicated by geometric shapes:

- Cluster 0 (circles): This is the largest cluster, with 4,671 agreements. In 2019 and 2020, only agreements with the state of PR are present in this cluster. It is predominant in most states in the years 2021 and 2022. In 2023, only agreements with the state of SC are part of this cluster. It has 49 distinct points on the graph and a spread of 48.13%, making it the most dispersed cluster.
- Cluster 1 (squares): This cluster contains 3,092 agreements, mainly concentrated in 2020, with additional agreements in 2019 and 2022 when crossing data with the state of RS. There are 25 distinct points in this cluster, with a spread of 28.07%, indicating an intermediate concentration of elements.
- Cluster 2 (triangles): This is the smallest cluster, with 2,335 agreements. It is the predominant cluster in 2019 and also includes elements from the data crossing in 2022 with the state of RS. It has 24 distinct points and a spread of 23.81%, making it the most concentrated cluster.

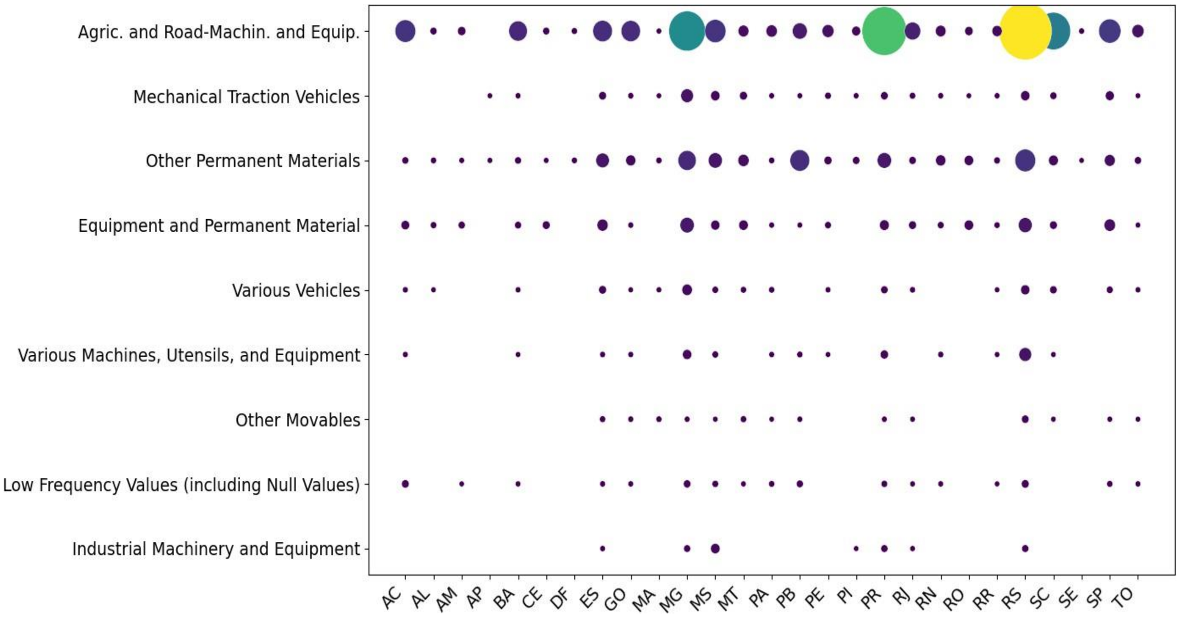
Figures 2 and 3 provide a comprehensive overview of the distribution of agreements by year and state, allowing for an analysis of the allocation patterns of resources from MAPA considering the data crossing between the State and Agreement Year columns.

4.2 Frequency and K-modes Clusters: State vs. Nature Expense

The data crossing from the State vs. Nature Expense columns is presented in Figures 4 and 5. The State data corresponds to each of the 27 states represented on the Y-axis by their respective official acronyms.

Figure 4

Bubbles Chart: Frequency - State vs. Nature-Expense



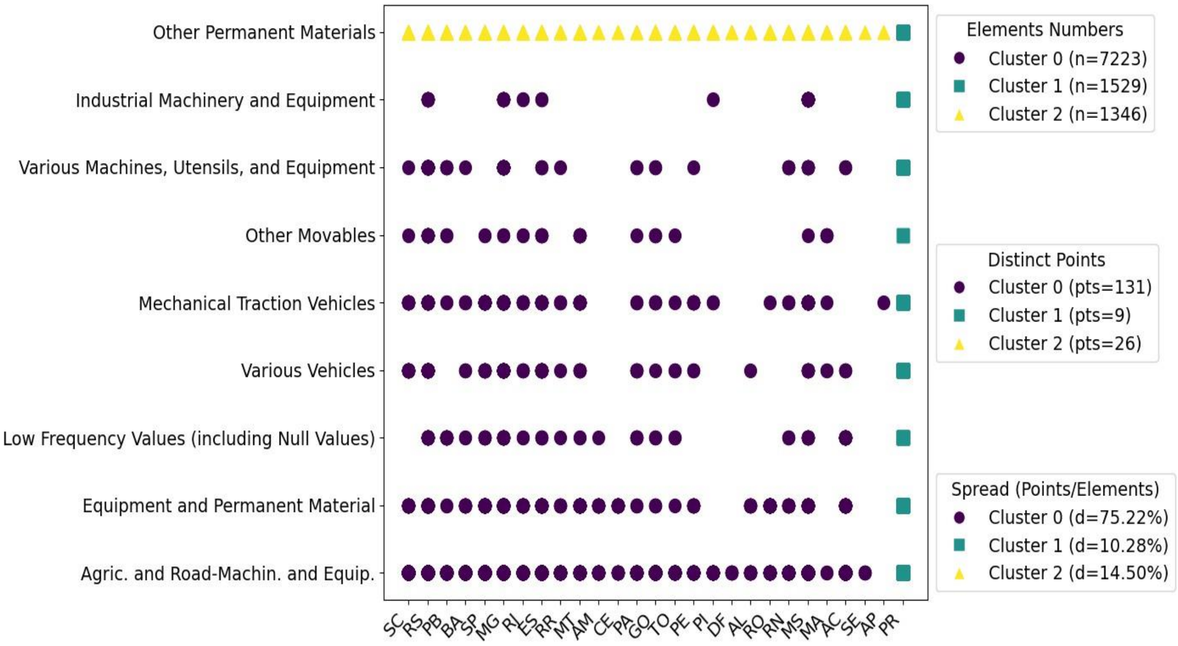
Source: Research data

Nature-Expense data refer to the nature of the expenses described in the signed agreements and are distributed across the following categories: ‘Agricultural and Road Machinery and Equipment,’ ‘Various Vehicles,’ ‘Other Permanent Materials,’ ‘Equipment and Permanent Material,’ ‘Low-Frequency Values (including Null Values),’ ‘Industrial Machinery and Equipment,’ ‘Others Nature-Expenses,’ ‘Mechanical Traction Vehicles,’ ‘Various Machines, Utensils, and Equipment,’ and ‘Other Movables.’

In Figure 4, the number of agreements signed is associated with the frequency of each pair, based on data crossing from the State vs. Nature Expense columns. This frequency is represented by the size of each circle. The data crossing of the nature of expenditure ‘Agricultural and Road Machinery and Equipment’ with the states MG, PR, and RS stands out with the highest frequencies. In this data crossing, AP does not have a signed agreement. Another interesting point is that the highest frequencies, the highest numbers of agreements, involve this nature of expense, even though it is not present in all states. ‘Other Permanent Materials’ is the only type of expense present in all states, yet it has fewer agreements signed than the type ‘Agricultural and Road Machinery and Equipment.’ The type of expense, Industrial Machinery and Equipment’ is present in only 7 out of the 27 states.

Figure 5

Scatter Plot: K-modes Clusters - State vs. Nature-Expense



Source: Research data

Figure 5 shows the clusters formed by crossing data from the State vs. Nature Expense columns. Each cluster corresponds to a geometric figure, and each element of a cluster represents an agreement to transfer resources from MAPA to the associated state entities. The agreements are grouped into three clusters, indicated by geometric shapes as highlighted below:

- Cluster 0 (circles): This is the largest cluster in terms of the number of agreements signed, with 7,223 elements distributed in 131 circles on the graph. With a spread of 75.22%, it is the most dispersed (least concentrated) cluster in the sample. It is present in the data crossing of most types of expenditure with most states, except in the data crossing of ‘Other Permanent Materials’ with all states, and it is not present in the data crossing of other types of expenditure with the state of PR.

- Cluster 1 (squares): This cluster includes 1,529 agreements distributed across 9 different points. It has a spread of 10.28% and is the least dispersed (most concentrated) cluster. It includes the data crossing of all types of expenses only with PR.
- Cluster 2 (triangles): This cluster contains 1,346 agreements distributed across 26 triangles, concentrated at the intersection of the nature of expenses 'Other Permanent Materials' with almost all states except PR. It has a spread of 14.50%, representing an intermediate level of dispersion compared to the other two clusters.

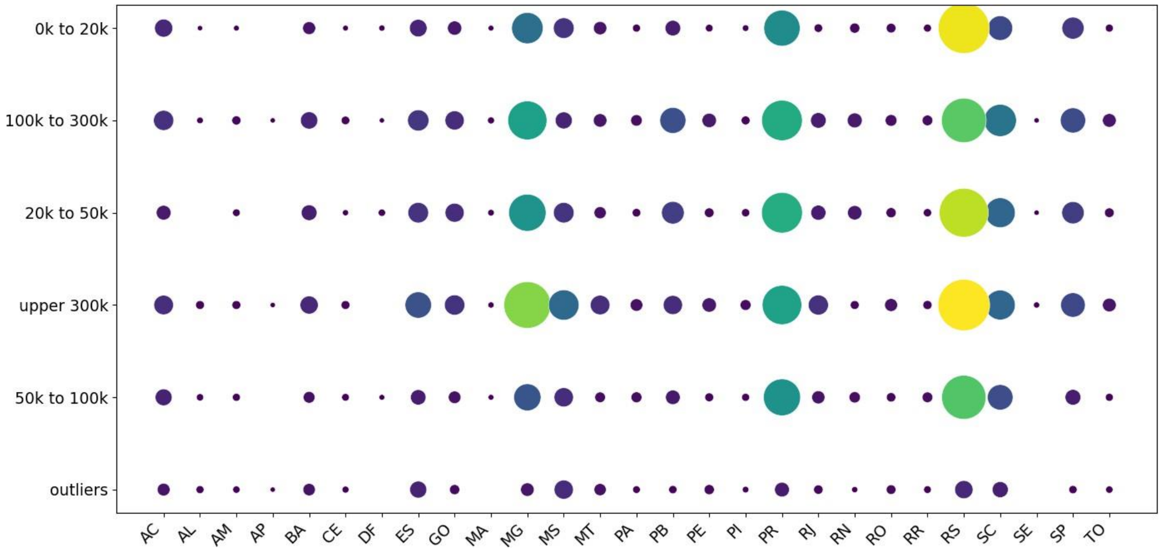
From a broader perspective, Figures 4 and 5 provide a panoramic view of the distribution of agreements, allowing us to analyze the allocation patterns of resources from MAPA based on the data crossing from State vs. Nature Expense

4.3 Frequency and K-modes Clusters: State vs. Total-Item-Value

The data crossing from the State vs. Total Item Value columns is presented in Figures 6 and 7. The State data corresponds to each of the 27 states, represented on the Y-axis by their respective official acronyms. The values in the Total Item Value column are categorized into ranges: ['0k to 20k,' '20k to 50k,' '50k to 100k,' '100k to 300k,' 'upper 300k,' 'outliers'], where k corresponds to the value 1,000. Outliers correspond to atypical values very close to 0k or far above 300k, comprising 505 values.

Figure 6

Bubbles Chart: Frequency - State vs. Total-Item-Value

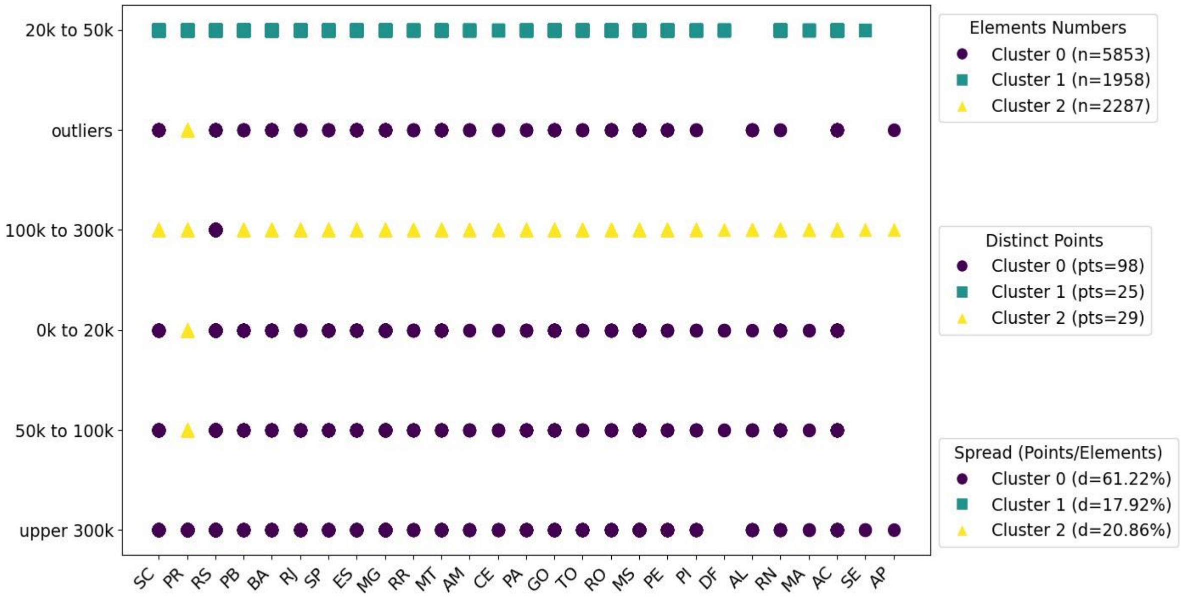


Source: Research data

In Figure 6, the number of agreements is associated with the frequency of each pair from the State vs. Total Item Value columns. This frequency is represented by the size of each circle. MG, PR, SC, and RS have a significant number of agreements across most value ranges, except for outliers.

Figure 7

Scatter Plot: K-modes Clusters - State vs. Total-Item-Value



Source: Research data

The outliers range highlights agreements in ES, MS, RS, and SC. The value ranges ‘0k to 20k,’ ‘20k to 50k,’ and ‘50k to 100k’ are not present in AP. The ‘upper 300k’ range and outliers are not present in DF. The outliers’ range is not present in AM, and the ranges ‘0k to 20k’, ‘50k to 100k,’ and outliers are not present in SE.

Figure 7 shows the clusters formed by crossing data from the State vs. Total Item Value columns. Each cluster corresponds to a geometric figure, and each element of a cluster represents an agreement for the transfer of resources from MAPA to the associated state entities.

The agreements are grouped into three clusters, indicated by geometric shapes as highlighted below:

- Cluster 0 (circles): This is the largest cluster in terms of the number of agreements, with 5,853 elements distributed over 98 different points on the graph. It has a spread of 61.22%, making it the most dispersed (least concentrated) cluster. It is predominant in most value ranges, except for the ‘20k to 50k’ range, where it is absent. In the ‘100k to 300k’ range, the only intersection present is in RS.
- Cluster 1 (squares): This is the smallest cluster, with 1,958 agreements distributed over 25 different points, concentrated in the ‘20k to 50k’ range. It has a spread of 17.92%, making it the least dispersed (most concentrated) cluster. AP and AL are not present in this cluster.
- Cluster 2 (triangles): This cluster has 2,287 agreements distributed across 29 distinct points. The elements of this cluster are mainly concentrated in the ‘100k to 300k’ range, covering almost all states except RS. It has a spread of 20.86%, indicating intermediate dispersion compared to the other clusters. Elements of this cluster are also present at the intersection of PR with the value ranges ‘0k to 20k’, ‘50k to 100k,’ and ‘outliers.’

From a broader perspective, Figures 6 and 7 provide a general overview of the distribution of agreements based on the State vs. Total Item Value data crossing, allowing for an analysis of MAPA’s resource allocation patterns across these dimensions.

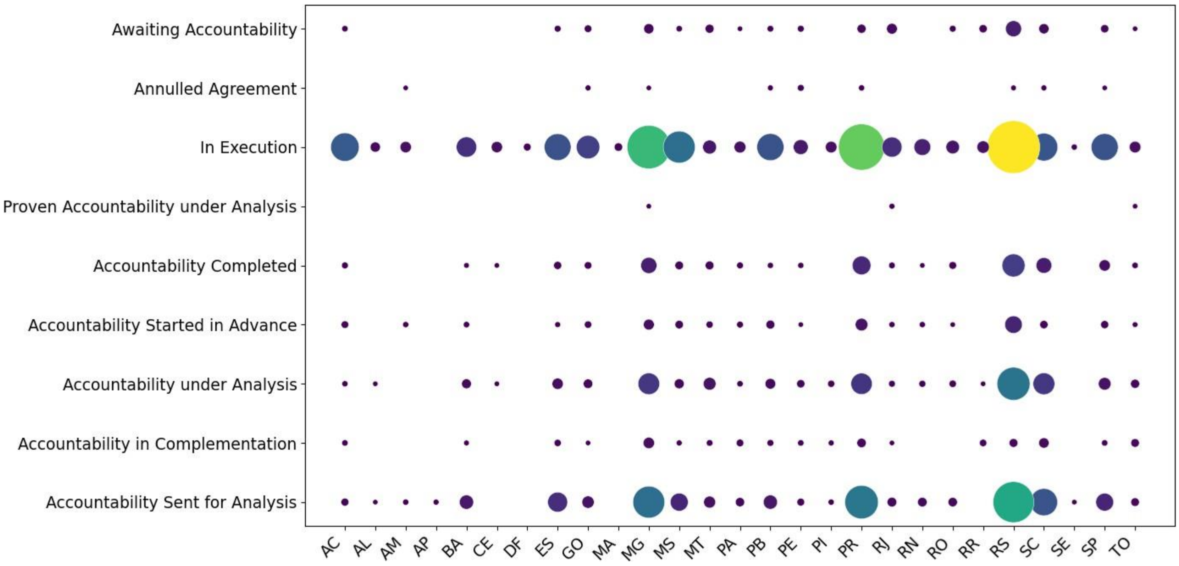
4.4 Frequency and K-modes Clusters: State vs. Agreement Situation

The data crossing from the State vs. Agreement Situation columns is presented in Figures 8 e 9. The Agreement Situation data refers to the status of the contract at the time of building the database: ‘Accountability in Complementation,’ ‘Accountability under Analysis,’ ‘Accountability Sent for Analysis,’ ‘Accountability Completed,’ ‘In Execution,’ ‘Accountability Started in Advance,’ ‘Awaiting Accountability,’ ‘Proven Accountability under Analysis,’ and ‘Annulled Agreement.’ The State data corresponds to each of the 27 states, represented on the Y axis by their respective official acronyms.

In Figure 8, the number of agreements is associated with the frequency of each pair, referring to the data crossing in the State vs. Agreement Situation columns. This frequency is represented by the size of each circle. The ‘In Execution’ situation is present in most states, except AP; the intersection of this situation with the states MG, PR, and RS is highlighted more frequently. In the ‘Accountability Sent for Analysis’ situation, the states MG, PR, and RS have the highest number of agreements signed, while DF stands out for having no signed agreements. The ‘Proven Accountability under Analysis’ situation is notable for having data crossings with only three states: MG, RJ, and TO.

Figure 8

Bubbles Chart: Frequency - State vs. Agreement Situation

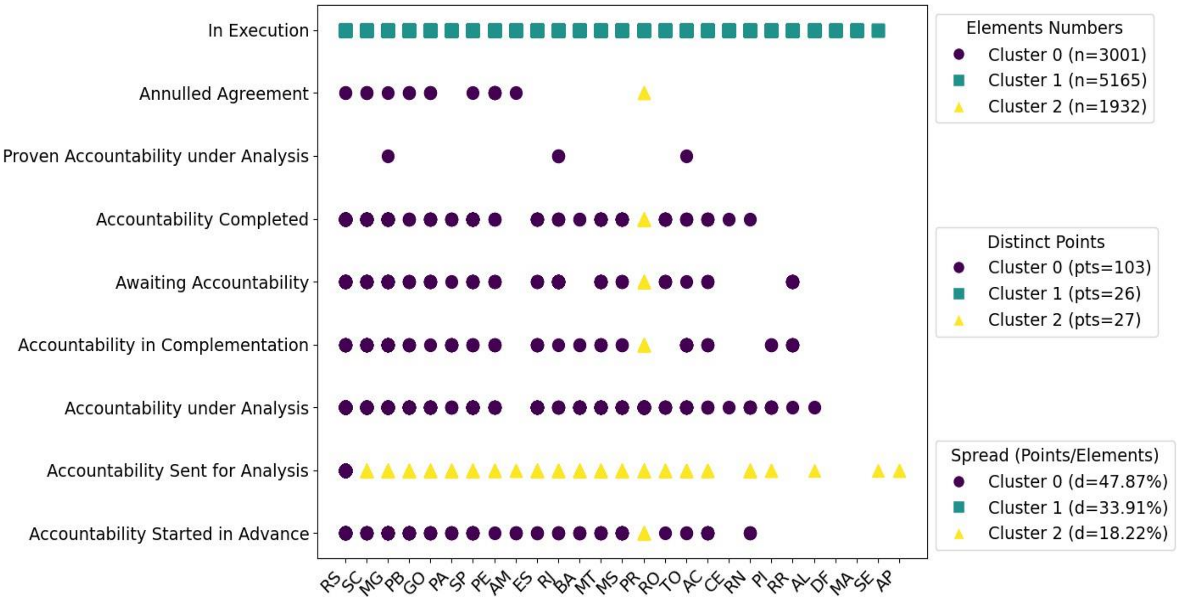


Source: Research data

Figure 9 shows the clusters formed by crossing data from the State vs. Agreement Situation columns. Each cluster corresponds to a geometric form, and each element of a cluster represents an agreement to transfer resources from MAPA to the associated state entities.

Figure 9

Scatter Plot: K-modes Clusters - State vs. Agreement Situation



Source: Research data

Figure 9 shows the clusters formed by crossing data from the State vs. Agreement Situation columns. Each cluster corresponds to a geometric form, and each element of a cluster represents an agreement to transfer resources from MAPA to the associated state entities.

The agreements are grouped into three clusters, indicated by geometric shapes as highlighted below:

- Cluster 0 (circles): This cluster contains 3,001 agreements distributed among 103 circles on the graph. It has a spread of 47.87%, making it the cluster with the most dispersed (or least concentrated) data. There are no elements from the data crossing of the states DF, MA, SE, and AP with any situations. Notably, this cluster includes data crossing from RS with the situation 'Accountability Sent for Analysis'.
- Cluster 1 (squares): This cluster has the largest number of elements, containing 5,165 agreements distributed across 26 squares. It has a spread of 33.91%, indicating intermediate dispersion of elements. The cluster elements are present at the intersection of the 'Accountability Sent for Analysis' situation with most regions. Additionally, elements are present at the intersection of PR with the situations 'Accountability Started in Advance', 'Accountability Sent for Analysis,' 'Accountability in Complementation,' 'Awaiting Accountability,' 'Accountability Completed,' and 'Annulled Agreement'.

- Cluster 2 (triangles): This is the smallest cluster, containing 1,932 agreements distributed across 27 distinct points. It has a spread of 18.22%, making it the cluster with the lowest dispersion. The cluster elements are concentrated at the intersection of the 'In Execution' situation with most states, except AP.

Starting with a broad analysis, Figures 24 and 25 provide an overview of the distribution of agreements, allowing us to analyze the allocation patterns of resources from MAPA by considering the data crossing from the State vs. Agreement Situation columns.

Discussion: K-modes Clustering Analysis

The results analysis presented in Figures 2 and 3 (Agreement Year vs. Region) reveal important patterns in the distribution of MAPA agreements. These patterns highlight the concentration of resources in the South and Southeast regions, especially in the states of Minas Gerais (MG), Paraná (PR), and Rio Grande do Sul (RS). There is a notable predominance of agreements aimed at acquiring agricultural and road machinery and equipment.

The data crossing from State vs. Agreement Year (Figures 4 and 5) shows a trend toward a reduction in the number of agreements over the years, likely influenced by political and economic factors, as well as the COVID-19 pandemic. This decrease may reflect changes in government priorities and the need for adjustments in public policies to address emerging challenges. Cross-checking the data from Region vs. Nature Expense (Figures 6 and 7) and State vs. Nature Expense (Figures 8 and 9) confirms the predominance of the 'Agricultural and Road Machinery and Equipment' expense category in the South and Southeast regions, with an emphasis on the states of MG, PR, and RS. Additionally, 'Other Permanent Materials' is present in all states, although less frequently, indicating a diversification in the types of investments.

The distribution of agreements by the state across the allocated value ranges is detailed in Figures 6 and 7 (State vs. Total Item Value). The results reveal the predominance of the states of MG, PR, RS, and São Paulo (SP) in terms of the number of agreements and coverage of value ranges. These states stand out for their ability to attract resources at different investment levels, suggesting greater maturity in the preparation and execution of innovation and rural development projects.

Data crossing from State vs. Agreement Situation (Figures 8 and 9) reinforces the predominance of the ‘In Execution’ situation in all regions and states, with an emphasis on Minas Gerais (MG), Paraná (PR), and Rio Grande do Sul (RS). Furthermore, the ‘Accountability Sent for Analysis’ situation also stands out in these states, indicating a more advanced accountability process.

These results may indicate a strategic allocation of resources, aiming to strengthen the infrastructure and productive capacity of regions with greater agricultural potential. Furthermore, the temporal analysis shows a trend toward reducing the total value of agreements over the years, possibly influenced by political and economic factors and the COVID-19 pandemic. This decrease may reflect changes in government priorities and the need for adjustments in public policies to address emerging challenges.

Insights generated by K-modes cluster analysis have the potential to inform and improve specific practices in applied agricultural sciences. For example, in agronomy, the identified patterns can guide the selection of crops and the most appropriate management techniques for each region, considering the characteristics of the predominant agreements in each cluster. This can contribute to optimizing the use of natural resources and improving agricultural productivity.

K-modes cluster analysis can also generate valuable insights for improving public rural development policies. By revealing regional patterns in resource allocation, this analysis helps identify priority areas that require greater attention and investment. This allows specific policies and programs to be targeted to meet local needs, promoting more equitable development. Furthermore, understanding the predominant expense categories in agreements guides the prioritization of investments in strategic areas such as infrastructure, training, and research. This resource allocation optimization contributes to a more efficient use of public funds, maximizing positive impacts on the agricultural sector.

The insights generated by K-modes cluster analysis can also support the improvement of mechanisms for monitoring and evaluating agreements. By better understanding the dynamics of execution and accountability, it is possible to establish more precise indicators and develop more effective monitoring methodologies, ensuring transparency and efficiency in the application of resources.

Finally, identifying clusters with a greater concentration of agreements focused on research and technological development can guide policies to promote innovation in the agricultural sector. This includes strengthening innovative ecosystems, disseminating innovative technologies and practices, encouraging the adoption of sustainable solutions, and boosting agriculture's competitiveness and sustainability. By incorporating these insights into formulating and implementing public policies, managers can make more informed decisions aligned with the agricultural sector's real needs. This contributes to the continuous improvement of mechanisms to promote rural development, fostering a more competitive, inclusive, and sustainable agriculture.

Conclusion

The Ministry of Agriculture, Livestock, and Supply (MAPA) plays a crucial role in advancing the sustainable development of Brazilian agribusiness, ensuring food security, and boosting the competitiveness of agricultural products. Through its Secretariat of Innovation, Rural Development, and Irrigation (SDI), MAPA strives to foster an environment that encourages innovation within the agricultural sector. The SDI achieves this by directly transferring resources through public agreements, targeting specific projects and initiatives that can benefit from financial support.

This article examines the application of data clustering techniques to public agreements signed by SDI/MAPA from 2019 to early 2023. Utilizing the K-modes method, this innovative approach addresses challenges typical in data clustering, such as handling non-numeric data, managing outliers, dealing with varying cluster sizes and densities, and improving group interpretability. This methodology provides a more nuanced analysis than traditional clustering techniques.

The analysis highlights significant patterns in the MAPA agreements, revealing a concentration of resources in Brazil's South and Southeast regions, particularly in Minas Gerais, Paraná, and Rio Grande do Sul. These areas show a strong focus on agreements related to acquiring agricultural and road machinery. Insights from these findings are invaluable for shaping public policy and strategic decisions, enabling the development of targeted programs that address local needs and prioritize critical investment areas.

The study underscores the effectiveness of strategic partnerships via these agreements, demonstrating that direct resource transfer is a powerful tool for implementing public policies that promote innovation and sustainable rural development in Brazil. By proposing a new method for analyzing such transfers, the research enhances scientific understanding and contributes to improving mechanisms for promoting both innovation and rural development. It also opens up opportunities for applying this methodology across various contexts, from municipal to national policy assessments.

Adapting the approach to other international scenarios, the strategy requires careful consideration of local differences, such as legal frameworks and institutional setups. However, the methodology holds promise for improving living conditions and strengthening agricultural competitiveness in diverse global contexts. Tailored adaptations can provide insights into strategic resource allocation and rural development priorities, supporting international efforts.

Overall, the article offers valuable contributions by uncovering trends and patterns emerging from SDI/MAPA partnerships with different governmental and non-governmental organizations. The insights have the potential to enhance public policies, supporting more robust innovation and sustainable rural practices. By improving the transparency and accountability of resource management, this study contributes to more effective governance, ensuring that public investments yield tangible benefits for the agricultural sector.

Finally, this research marks the beginning of a journey towards developing better instruments for fostering innovation in agriculture. It invites researchers, policymakers, and industry stakeholders to actively participate in this journey, paving the way for a more resilient, inclusive, and innovative agri-food system that meets present and future challenges.

Acknowledgements

This research was funded by SDI/MAPA, grant number TED nº 939993/2022.

References

Ahmed, M. and Pathan, A.-S. K. (2018). Data analytics: concepts, techniques, and applications. Crc Press.

Alvares, M. A. A. and Branco, M. S. (2018). Captação de recursos via transferências voluntárias: um olhar para os municípios do Consórcio Intermunicipal da Região Oeste Metropolitana de São Paulo (Cioeste). Revista do Serviço Público, 69(3):605–630.

Borges, G. d. F., da Silva, C. M. D., Silva, K. A. T., de Benedicto, G. C., and Antonialli, L. M. (2013). Endividamento Dos Estados Brasileiros Após Uma Década Da Lei De Responsabilidade Fiscal: Uma Análise Com Estatística Multivariada. Revista FSA, 10(4):20–43.

Brasil (2008). Sistema de gestão de convênios e contratos de repasse - SICONV. https://siconv.com.br//. accessed: 2024-04-11.

Brasil (2023). Decreto nº 11.332 de 20 de abril de 2023 – estrutura regimental do ministério da agricultura e pecuária (mapa). https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/D11332.htm.

Brasil. Congresso Nacional (2000). Lei complementar nº 101, de 4 de maio de 2000. Accessed: 2024-04-11.

Costa, C. C. d. M., Ferreira, M. A. M., Braga, M. J., and Abrantes, L. A. (2012). Disparidades Inter-Regionais e Características dos Municípios do Estado de Minas Gerais Disparities and Inter-Regional Characteristics of Municipalities of the State of Minas Gerais. Desenvolvimento em Questão, pages 52–88.

Crispim, G., Alberton, L., Ferreira, C. D., and Lopes, J. E. d. G. (2020). Behavior of budget expenditures during the election period: an analysis in panel data in the Brazilian municipalities. Revista Ambiente Contábil - Universidade Federal do Rio Grande do Norte, 12(2).

Cruz, T. and Silva, T. (2020). Minimum Spending in Education and the Flypaper Effect. Economics of Education Review, 77(May):102012.

Dong, G. and Pei, J. (2007). Data Crossing: Concepts, Techniques, and Applications, pages 115–139. Springer, Boston, MA.

Hautam"aki, V., Cherednichenko, S., K"arkk"ainen, I., Kinnunen, T., and Fr"anti, P. (2005). Improving k-means by outlier removal. In Scandinavian Conference on Image Analysis, pages 978–987. Springer.

Huang, Z. (1998a). Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2(3):283–304.

Huang, Z. (1998b). Extensions to the k-means algorithm for clustering large data sets with categorical values. In Data mining and knowledge discovery, volume 2, pages 283–304. Springer.

Lima, C. R. d. A. (2005). Sistema de informações sobre orçamentos públicos em saúde: confiabilidade e uso das informações na construção de um perfil dos municípios brasileiros. PhD thesis, Fundação Oswaldo Cruz.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento (2023a). Convênios. https://www.gov.br/agricultura/pt-br/acesso-a-informacao/convenios-e-ted/convenios.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento (2023c). Estrutura organizacional. https://www.gov.br/agricultura/pt-br/acesso-a-informacao/institucional/estrutura-organizacional.

Parente, J. I. S. (2019). Ciclo Politico e Efeito Aliança nas Transferências para os Municípios. Dissertação, Universidade Federal do Ceará.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3):45–77.

PSF - Python Software Foundation (2024). Python programming language. https://www.python.org/. Version 3.11.3, accessed: 2024-04-17.

Silvério, M. C. (2019). Ensaios Sobre Políticas Públicas De Ciência, Tecnologia E Inovação No Brasil. Dissertação, Universidade de Brasília.

Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., and Xu, Y. (2017). Weed identification based on k-means feature learning combined with convolutional neural network. Computers and Electronics in Agriculture, 135:63–70.

Tian, K., Li, J., Zeng, J., Evans, A., and Zhang, L. (2019). Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. Computers and Electronics in Agriculture, 165:104962.

Received: 10.25.2024

Accepted: 11.15.2024