# Descriptive analysis of intergovernmental agreements in the brazilian ministry of agriculture and livestock: exploring data clustering of agreements and grantees

## Análise descritiva de contratos intergovernamentais no ministério da agricultura e pecuária brasileiro: explorando a clusterização de dados de convênios e convenentes

## Análisis descriptivo de los contratos intergubernamentales en el ministerio de agricultura y ganadería de brasil: explorando la agrupación de datos de los convenios y beneficiarios

**Márcio Dias de Lima**
PhD in Computer Science
Institution: Instituto Federal de Goiás
Address: Goiânia, Goiás, Brasil
E-mail: marcio.lima@ifg.edu.br

**Alan Keller Gomes**
PhD in Computer Science
Institution: Instituto Federal de Goiás
Address: Goiânia, Goiás, Brasil
E-mail: alan.gomes@ifg.edu.br

**Daniel Soares de Souza**
Master in Public Management
Institution:  Instituto Federal de Brasília
Address: Brasília, Distrito Federal, Brasil
E-mail: daniel.souza@ifb.edu.br

**Lucas Santos de Oliveira**
Specialist in Applied Statistics
Institution:  Instituto Federal de Brasília
Address: Brasília, Distrito Federal, Brasil
E-mail: lucas.oliveira4@estudante.ifb.edu.br

**Paulo Henrique dos Santos**
Master in Production Engineering
Institution: Instituto Federal de Goiás
Address: Senador Canedo, Goiás, Brasil
E-mail: paulo.santos1@ifg.edu.br

**Cassiomar Rodrigues Lopes**
Master in Agrobusiness
Institution: Instituto Federal de Goiás
Address:  Anápolis, Goiás, Brasil
E-mail: cassiomar.lopes@ifg.edu.br

**José Carlos Barros Silva**
Master in High Education Science
Institution:  Instituto Federal de Goiás
Address:  Luziânia, Goiás, Brasil
E-mail: josecarlos.silva@ifg.edu.br

**Karla de Aleluia Batista**
PhD in Biological Science
Institution:  Instituto Federal de Goiás
Address:  Goiânia, Goiás, Brasil
E-mail: karla.batista@ifg.edu.br

**ABSTRACT**
The agricultural and livestock industry is a cornerstone of Brazil's economy, wielding substantial influence over the nation's GDP and food security. Within this framework, the Ministry of Agriculture and Livestock (MAPA) plays an essential role in shaping and implementing policies geared towards fostering the sustainable growth of these critical sectors through public agreements. In this context, the equitable and efficient allocation of resources requires a thorough understanding of these agreements' geographical distribution and temporal dynamics. This study meticulously scrutinizes the geographic spread, temporal progression, and inherent characteristics of the agreements signed by MAPA from 2019 to 2023. It examines the agreements using data mining techniques, encompassing clustering and statistical analyses. The research unveiled a pronounced regional distribution in the agreements, with Brazil's Southern region emerging as the primary destination of grants, closely shadowed by the Southeast. Consequently, the study offers a holistic comprehension of the geographical dispersion, temporal evolution, and intrinsic traits of the agreements brokered by MAPA. Such insights serve as invaluable assets, empowering policymakers to formulate more efficient and targeted strategies for nurturing Brazil's agricultural and livestock industry concerning the allocation of resources and the elaboration of agreements.

**Keywords:** Data Mining. K-means Clustering. Public Agreements. Voluntary Transfers.

## RESUMO

O setor agrícola e pecuário desempenha um papel fundamental na economia do Brasil, contribuindo de forma expressiva para o Produto Interno Bruto (PIB) e a segurança alimentar do país. Nesse contexto, o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) assume um papel central na formulação e execução de políticas destinadas a impulsionar o desenvolvimento sustentável desses setores vitais, por meio de convênios públicos. No entanto, para garantir uma distribuição justa e eficaz dos recursos, é imprescindível compreender a distribuição geográfica e os padrões temporais dos convênios estabelecidos pelo MAPA. Assim, o objetivo desta pesquisa foi analisar detalhadamente a distribuição geográfica, a evolução ao longo do tempo e as características dos convênios firmados pelo MAPA no período de 2019 a 2023. Utilizando técnicas de mineração de dados, incluindo clusterização e análises estatísticas, foram examinados os registros de assinaturas de contratos disponibilizados pelo próprio MAPA. A análise desses dados revelou uma distribuição regional significativa nas assinaturas de contratos pelo MAPA, com a região Sul despontando como líder em quantidade de convênios, seguida de perto pela região Sudeste. Dessa forma, a pesquisa proporcionou uma compreensão abrangente da distribuição geográfica, da evolução temporal e das características dos convênios estabelecidos pelo MAPA, fornecendo assim informações que podem auxiliar em políticas mais eficientes e direcionadas para o desenvolvimento dos setores agrícola e pecuário do Brasil em relação a transferências de recursos e convênios.

**Palavras-chave:** Convênios Públicos. Data Mining. K-means Clustering. Transferências Voluntárias.

## RESUMEN

El sector agrícola y ganadero desempeña un papel fundamental en la economía de Brasil, contribuyendo de manera significativa al Producto Interno Bruto (PIB) y a la seguridad alimentaria del país. En este contexto, el Ministerio de Agricultura, Ganadería y Abastecimiento (MAPA) asume un papel central en la formulación y ejecución de políticas destinadas a impulsar el desarrollo sostenible de estos sectores vitales, mediante convenios públicos. Sin embargo, para garantizar una distribución justa y eficaz de los recursos, es imprescindible comprender la distribución geográfica y los patrones temporales de los convenios establecidos por el MAPA. Así, el objetivo de esta investigación fue analizar detalladamente la distribución geográfica, la evolución a lo largo del tiempo y las características de los convenios firmados por el MAPA en el período de 2019 a 2023. Utilizando técnicas de minería de datos, incluyendo la agrupación y análisis estadísticos, se examinaron los registros de firmas de contratos disponibles por el propio MAPA. El análisis de estos datos reveló una distribución regional significativa en las firmas de contratos por el MAPA, destacando la región Sur como líder en cantidad de convenios, seguida de cerca por la región Sudeste. De esta forma, la investigación proporcionó una comprensión exhaustiva de la distribución geográfica, la evolución temporal y las características de los convenios establecidos por el MAPA, proporcionando así información que puede ayudar en

**Page 3**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

políticas más eficientes y dirigidas para el desarrollo de los sectores agrícola y ganadero de Brasil en relación a las transferencias de recursos y convenios.

**Palabras clave:** Agrupamiento K-means. Convenios Públicos. Minería de Datos. Transferencias Voluntarias.

# 1 INTRODUCTION

The Brazilian Ministry of Agriculture and Livestock (MAPA) plays a crucial role in administering public policies to bolster the country's agricultural sector. Broadly, MAPA facilitates agribusiness development while overseeing and standardizing services related to this industry. Within the Brazilian context, agribusiness encompasses a diverse array of farmers, spanning from small-scale farms to large-scale operations. The sectors' activities include providing agricultural goods and services, production, processing, transformation, and distributing agricultural products until they reach the end consumer (MAPA, 2023; Portal Gov.br, 2024).

Small farmers in Brazil often find themselves marginalized in the market, unable to compete on par with the agricultural sector's epicenters. This disparity is frequently attributed to insufficient production equipment and infrastructure (Braga; Vieira Filho; de Freitas, 2019). Consequently, public grant agreements emerge as a vital tool to propel their production. From this standpoint, one of MAPA's operational modalities involves entering into general agreements with national and subnational governmental agencies or non-governmental organizations (NGOs) to channel financial resources (Braga; Vieira Filho; de Freitas, 2019).

These agreements yield a wealth of data, including details on resource allocation, operational domains, regions of predominant activity, and equipment requested. Within this milieu, data mining emerges as a tool for comprehending this extensive dataset, offering the capability to unearth concealed patterns, trends, and correlations within vast data repositories (Fayyad; Piatetsky-Shapiro; Smyth, 1996). In Fernandes e Chiavegatto Filho (2019), a comprehensive examination of data mining and machine learning (ML) methodologies is

**Page 4**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

presented, catering to the needs of occupational health and safety practitioners. Regarding the subject addressed in this article, these methodologies can effectively scrutinize data from general agreements, adopting cluster analysis to extract pertinent insights.

The ability to forecast future trends based on historical data is indispensable in a refined decision-making process. This predictive prowess facilitates anticipatory scenario analysis, pattern identification, and informed decision-making. Within this framework, statistical modeling and ML algorithms assume paramount importance (Mamede; Fray; Junior, 2023). Among these algorithms, clustering stands out as a prominent tool in data science. This methodology discerns the inherent grouping structures within a dataset, elucidating the highest degree of similarity within clusters and the greatest dissimilarity between distinct groupings (Sinaga; Yang, 2020).

The primary characteristics of a dataset derived from a clustering algorithm encompass heightened homogeneity within clusters and maximal heterogeneity between groups (Duran; Odell, 2013). Moreover, the analysis of data clusters is descriptive, theoretical, and non-inferential. Consequently, knowledge gleaned from data clusters does not constitute a blanket generalization, as it hinges upon the variables utilized for gauging similarity or dissimilarity (Frades; Matthiesen, 2010). Within the machine learning (ML) domain, data clustering stands as a cornerstone of unsupervised learning, wielding significant influence in uncovering latent patterns and structures within unlabeled datasets (Alonso-Betanzos; Bolón-Canedo, 2018).

The current data clustering techniques and algorithms will invariably lead to some sort of group of data, regardless of the actual presence of structured patterns. This task poses considerable challenges owing to the diverse array of shapes and sizes clusters can assume, as well as the myriad of data types and application domains (Jaeger; Banks, 2023). Several categories exist within the realm of clustering algorithms, each boasting strengths and weaknesses. Selecting the most appropriate algorithm often hinges on the data's nature and the analytical objectives. Prominent categories include hierarchical, density-based, grid-based, model-based, constraint-based, and partitional-based

algorithms. Partitional algorithms segment the data into a predetermined number of clusters to optimize criteria such as the sum of quadratic distances. Due to their simplicity and efficacy with large datasets, partitional algorithms are widely favored. However, they often presume clusters of spherical shape and uniform size, which can pose limitations in certain contexts. Notably, the K-means algorithm holds significance in this context.

Cluster analysis finds application across diverse domains, spanning from energy systems (Matenga, 2022), food retail (Hoenink et al., 2023), and social behavior (Jiang; King; Prinyawiwatkul, 2014) to the study of consumer behavior regarding functional foods (Sgroi et al., 2024), identifying investment opportunities (Nemitz et al., 2016), and discerning consumption patterns (Schosler; de Boer; Boersema, 2014).

This article employs clustering methodologies to conduct a descriptive data analysis regarding public agreements concerning resource transfers between Brazilian national and subnational governments from 2019 to 2023.

The study contributes by offering a concise literature review that approaches the topic and conducting a descriptive analysis of the agreements, focusing on the amount of resources transferred and the correlation with Brazil's regions. Additionally, it identifies the resultant clusters across the regions, as well as highlights the participation of the target audience in the execution of the agreements' activities in these territories.

This article is structured into six sections. The introduction is followed by Section 2, which covers related studies. Section 3 explains the materials and methods used in this research. Section 4 presents the results, while Section 5 discusses the findings. Finally, Section 6 provides the conclusion, followed by a list of references used.

## 2 RELATED WORK

### 2.1 VOLUNTARY TRANSFERS AMONG BRAZILIAN SUBNATIONAL ENTITIES

Agriculture stands as one of the primary sources of livelihood and employment globally, yet it confronts the imperative challenge of enhancing productivity while embracing sustainable practices (Laurett; Paco; Mainardes, 2021), particularly amidst the escalating food demands of an expanding global populace.

Brazil is globally renowned for its agricultural prowess, and family farming plays a significant role in the country's production, contributing to both food security and economic stability (Cordeiro et al., 2023). Notwithstanding, small-scale farmers struggle in the market, facing substantial challenges to compete against the dominant hubs within the agricultural industry (Mangiaterra, 2020). This difficulty precipitates a gradual erosion of market share for these farmers, jeopardizing their viability and sustainability. The maximization of these farmers' economic potential often encounters financial or material barriers, underscoring the critical importance of voluntary transfers channeling resources to family farming (Moutinho, 2016).

In this context, the Secretariat of Innovation, Rural Development, and Irrigation (SDI) of the MAPA actively fosters an innovative environment within the agricultural sector. Sustained investment in agribusiness stands as a cornerstone for its growth and success, constituting a pillar of the national economy. However, ensuring parity between large agribusiness corporations and family farmers poses a considerable challenge. To address this disparity, MAPA leverages public agreements to propel sectoral development, channeling financial resources accordingly (Portal da Transparencia, 2024).

The formulation and execution of public policies entail a myriad of instruments that wield direct influence over social realities (Lascoumes; Le Galès, 2007). These instruments, far from being neutral, produce specific effects that may diverge from the policies' stated goals (Capano; Howlett, 2020).

In this industry, funding is fundamental. The allocation of public funds and

Page 7

REVISTA CADERNO PEDAGÓGICO – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

resource distribution mechanisms are linchpins for the implementation of agricultural policies (Borges; Parré, 2021). In addition, recent research has underscored that public policies for agriculture have garnered scholarly attention within the field (Grisa; Schneider, 2014).

An example is a study analyzing voluntary transfers from the Brazilian federal government to subnational entities through MAPA from 2008 to 2021 (Lui; Miquelino, 2023). The research sought to understand the execution of the resources transferred and the impacts on national agriculture, discussing public policy instruments.

Another study in this direction analyzed the Brazilian government's voluntary transfers between 2008 and 2016, using the Public Grant and Transfer Agreements Management System (SICONV) (Chaves; Campelo Filho; Pinto, 2022). The authors report that since the 1960s, these transfers have been essential for decentralizing resources and implementing public policies. However, weaknesses in management were identified, requiring the computerization of transfer procedures. The research sought to fill a gap in the literature, examining the flow of transfers and the effectiveness of SICONV in the equitable distribution of public resources.

Lastly, it is noteworthy to highlight a study that proposed clustering Brazilian municipalities based on their public finances, economic activity, and the socio-economic conditions of their populations (Costa et al., 2012). This approach aimed to streamline the planning and evaluation of public policies tailored to foster local socio-economic development. The research delved into the regional and socio-economic disparities among Brazilian municipalities, focusing on the state of Minas Gerais. These disparities stemmed from political and economic factors, exerting impacts at the state and municipal levels. The decentralization of actions and resources, enshrined in the 1988 Brazilian Constitution, entrusted municipalities with the administration of pivotal policies, such as education and health, in a bid to foster balanced socio-economic development. In this vein, the Municipal Participation Fund (FPM) is a crucial mechanism for leveling municipal capacities to provide essential goods and services. Furthermore, federalism acknowledges the intricacies of each municipality's specific needs, empowering

local administrators to allocate resources strategically to spur development.

## 3 MATERIALS AND METHODS

### 3.1 DATA DESCRIPTION

The MAPA supplied the data utilized in this study, furnishing a sample extracted from its database. These data encapsulate information regarding the public agreements signed by MAPA from 2019 to 2023. The sample encompassed a total of 10098 observations across 40 variables. However, only variables deemed pertinent and consequential for this research endeavor were retained. Table 1 delineates the selected variables of interest that formed the foundation for the study's development.

Table 1. Variables of interest in this study

| Variable | Type | Description |
|---|---|---|
| Year_Signed | Numeric | Year the agreement was signed |
| Type_Expenses | Categorical | Nature of the expense associated with the item |
| Target_Public | Categorical | Target audience benefited from the agreemen |
| Quant_Item | Numeric | Item quantity |
| Region | Categorical | Brazilian region |
| Agreement_Situation | Categorical | Current status of the agreemen |
| State | Categorical | State |
| Total_Value_Item | Numeric | Total value of the item |
| Unit_Value_Item | Numeric | Unit value of the item |

Source: Research data.

The cluster analysis used the numerical variables to categorize regions according to the cumulative values of the public grant agreements. The K-means algorithm, implemented with the assistance of the R programming language, was employed for this endeavor. Further elucidation regarding the specifics of this clustering methodology will be provided in subsequent sections.

### 3.1.1 K-means

The unsupervised algorithm K-means is often used in data sciences for clustering tasks. The algorithm typically partitions the dataset into K groups, each

**Page 9**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

containing instances that exhibit similarity based on a pre-defined similarity measure (Aguiar; Santana Júnior; Bastos Filho, 2018). Specifically, the K-means algorithm thrives when the mean is a meaningful metric, typically requiring data measured on an interval or ratio scale (Macqueen; 1967).

In the clustering process, the dataset undergoes segmentation into groups or clusters determined by the centroids, which represent the average of all points within the cluster. Subsequently, each point is allocated to the cluster whose centroid exhibits the closest proximity. This iterative process persists until the centroids converge and no longer undergo alteration (Meng et al., 2018).

Mathematically, the $K$-means algorithm can be delineated as follows: When $X = \{x_1, x_2, \ldots, x_n\}$ is the dataset and $C = \{c_1, c_2, \ldots, c_k\}$ is the set of initial centroids, where $k$ denotes the desired number of clusters, the $x_i$ is attributed to the cluster whose centroid $c_j$ is the closest for each point $x_i$, utilizing Euclidean distance. The objective is to minimize the ensuing cost function:

$$J(C) = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} \cdot \| x_i - c_j \|^2 \qquad (1),$$

Where

$w_{ij} = 1$ if $j = argmin_j \| x_i - c_j \|^2$ or 0 otherwise.

This process recalculates the centroids repeatedly until the assignments of points to clusters do not change significantly or a stopping criterion is met. To assist in the application of the K-means algorithm, the $kmeans$ function from the $stats$ package was used in the R Studio environment (R Studio Team, 2023).

## 3.2 DETERMINING THE NUMBER OF CLUSTERS

A significant challenge in implementing clustering algorithms, such as K-means and its variants, is the necessity to pre-define the number of clusters as input. This study adopts an exploratory data analysis approach, which proves

relevant when researchers require pre-defined models or hypotheses but aim to comprehend the general characteristics or structure of high-dimensional data (Maione et al., 2017). The primary objective is to comprehensively understand the data, and identifying the number of emerging clusters is one of the principal goals. Given the absence of prior knowledge regarding the number of clusters, a unique methodology was employed to facilitate visualization and estimate the suitable number for our application. Cluster analysis was conducted to categorize regions based on the total values of the agreements. The subsequent steps for clustering were the employment of the $fviz\_nbclust$ function, which is part of the $factoextra$ package within the R Studio environment. Regarded as an extension of the $FactoMineR$ package, the $factoextra$ package furnishes additional graphical and utility functions tailored for factor analysis and related techniques (R Studio Team, 2023).

## 3 RESULTS

Table 2 presents the number of agreements signed by the MAPA in each region.

Table 2. Agreements by region of Brazil

| Region | Frequency | Percentage (%) | Cumulative percentage (%) |
|---|---|---|---|
| North | 754 | 7.5 | 7.5 |
| Northeast | 1022 | 10.1 | 17.6 |
| Central-West | 1043 | 10.3 | 27.8 |
| Southeast | 2528 | 25.0 | 53.0 |
| South | 4751 | 47.0 | 100.0 |
| Total | 10098 | 100.0 | |

Source: Research data.

Table 2 illustrates the prevalence of agreements in South Brazil, totaling 4751 instances and representing nearly 50% of the resources allocated from 2019 to 2023. Conversely, the North region records the lowest number of agreements. The Southeast region accounts for 25% of the overall total, amounting to 2528 agreements. Additionally, the collective sum of agreements between MAPA and subnational entities from the North, Northeast, and Central-West regions amounts to 27.8% of the agreements, a proportion comparable to that of the

Southeast region alone.

Figure 1. Distribution of agreements over the years
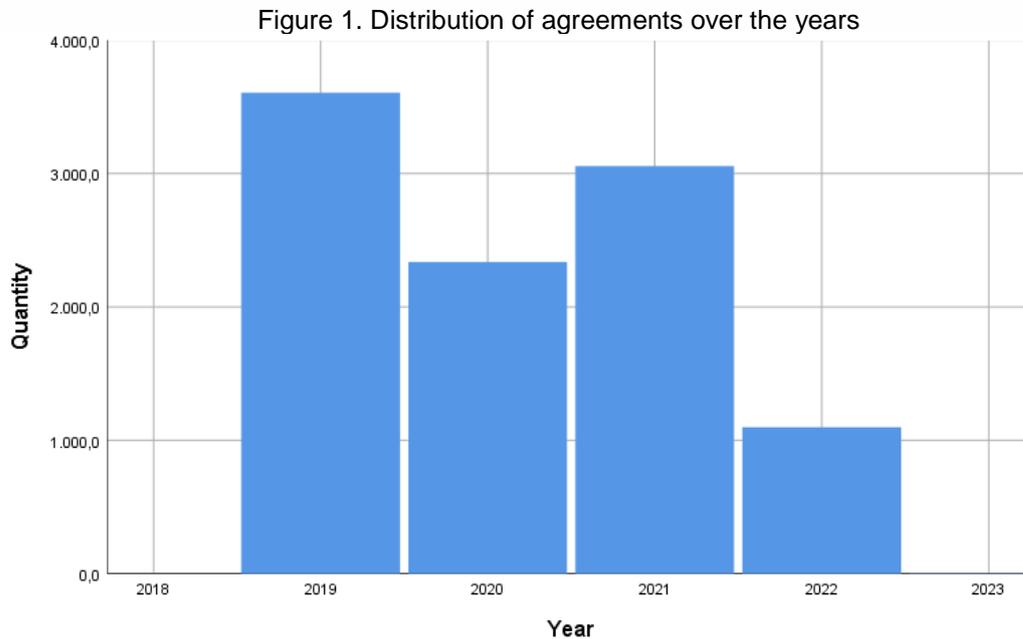


Source: Research data.

Analysis of the number of agreements over the years, as depicted in Figure 1, unveils a fluctuating trend. A decline was noted between 2019 and 2020; however, from 2020 to 2021, there was a resurgence. Notably, a substantial decrease was recorded between 2021 and 2022. These fluctuations can be attributed to various factors, including political influences, the repercussions of the pandemic, and shifts in government, all of which were pivotal events integrated into this timeframe.

Changes in the government's political priorities can directly influence programs and initiatives, impacting both adherence rates and the volume of grant agreements. Alterations in budget allocations may result in diminished resources allocated to certain programs, constraining their efficacy and attractiveness. The onset of the COVID-19 pandemic likely induced economic instability, prompting corporations and individuals to reassess their financial commitments. This was reflected in a decline in the number of agreements and in the processes around requesting and preparing these instruments for resource allocation.

**Page 12**

*REVISTA CADERNO PEDAGÓGICO* – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

Another significant factor was the change in the federal government after the 2022 presidential elections. A new administration tends to adjust policies and programs, disrupting continuity and diminishing interest in existing initiatives.

Figure 2 illustrates the distribution of agreements by state, providing a comprehensive overview from 2019 to 2023. The focus is on state-wise distribution, shedding light on temporal evolution and geographic disparities. The year 2019 marked the peak in the number of agreements, totaling 3606, followed by a decline in 2020 (2336). A resurgence was observed in 2021 (3056), but a significant downturn was apparent in 2022 (1098), with a near absence in 2023 (2). São Paulo (SP) consistently emerged as one of the primary contributors throughout the years, although with a notable decline in 2022. Conversely, other states such as Rio Grande do Sul (RS) and Minas Gerais (MG) showcased significant fluctuations, influencing the overall trend. Additionally, certain states, including Amapá (AP), Roraima (RR), and Maranhão (MA), exhibited limited or negligible participation in specific years.

Figure 1. Relationship between the total value of the items and the agreements' target audience
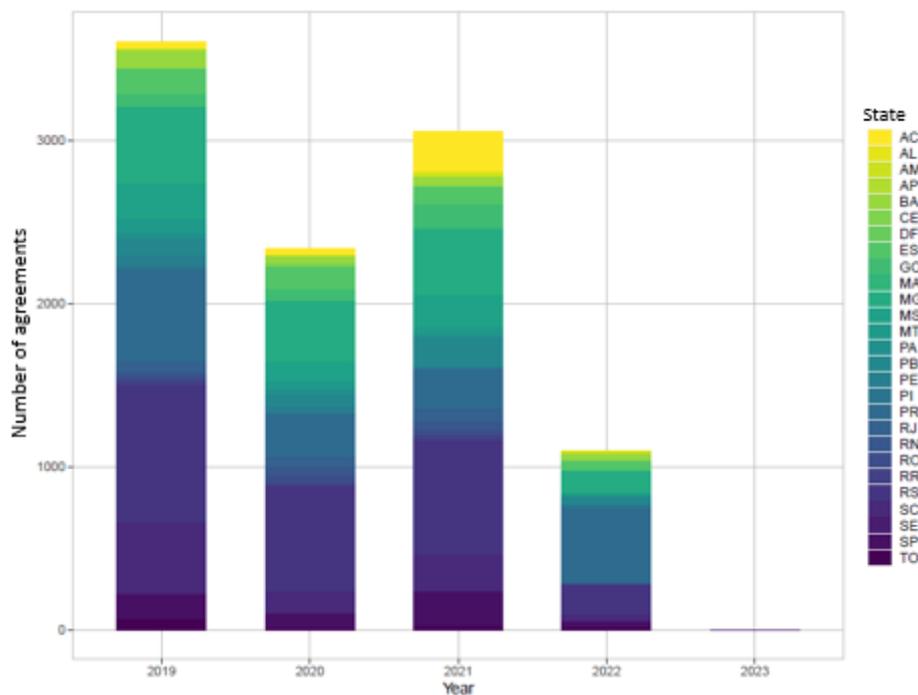


Source: Research data.

Table 3 provides valuable information about the status of agreements. The

most notable category is agreements in progress, encompassing over half of them (51.1%). Approximately 35% of the agreements have already been submitted to auditing and/or are being audited. A spectrum of scenarios is evident, ranging from audited agreements to those under analysis or awaiting supplementary information/documents.

Table 3. Status of agreements signed by MAPA

| Agreement status | Frequency | Percentage (%) |
|---|---|---|
| Awaiting accountability submission | 291 | 2.9 |
| Agreement annulled | 25 | 0.2 |
| In progress | 5165 | 51.1 |
| Accountability documentation verified and under audit | 5 | 0.0 |
| Accountability completed | 611 | 6.1 |
| Audit in progress | 1181 | 11.7 |
| Awaiting supplementary information/documents | 191 | 1.9 |
| Accountability documentation submitted for audit | 2322 | 23.0 |
| Audit initiated in advance | 307 | 3.0 |
| **Total** | **10098** | **100.0** |

Source: Research data.

Figure 3 provides a comprehensive depiction of the relationship between the variables "Target_audience" and "Total_Value_Item." Each point represents an individual observation. Given that the variable "Target_Public" consists of a string of characters, it was recorded using SPSS software to generate its corresponding graph. Analysis of **Erro! Fonte de referência não encontrada.** reveals a dispersion in the values, indicating substantial variability in the data. In essence, the values are distributed in a non-uniform or concentrated manner, implying significant diversity in the target audience.
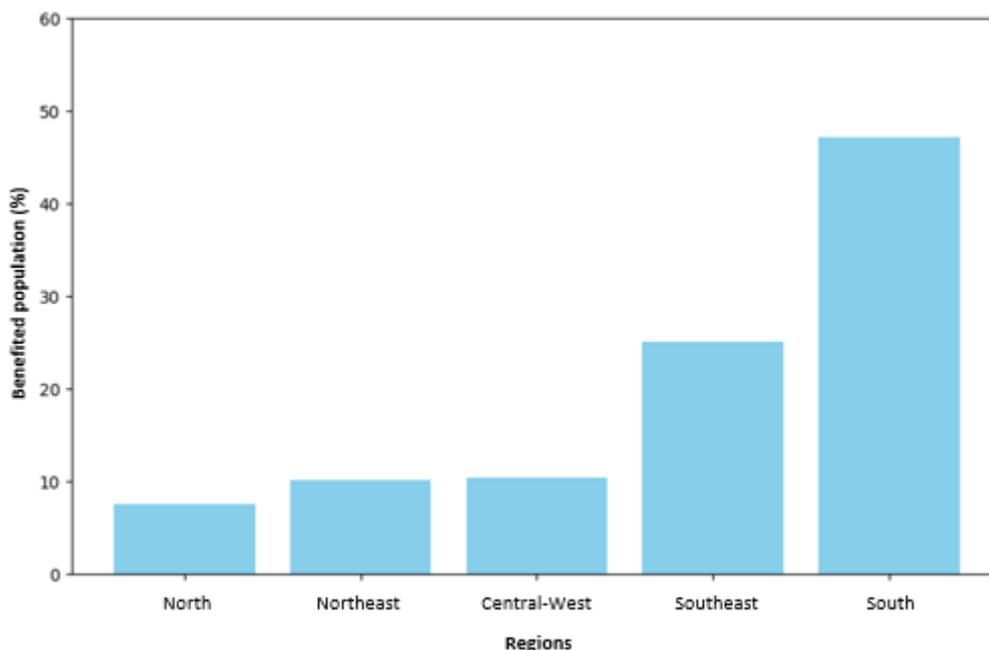
Figure 2. Scatter graph of the distribution of the total values of the items and the target audience
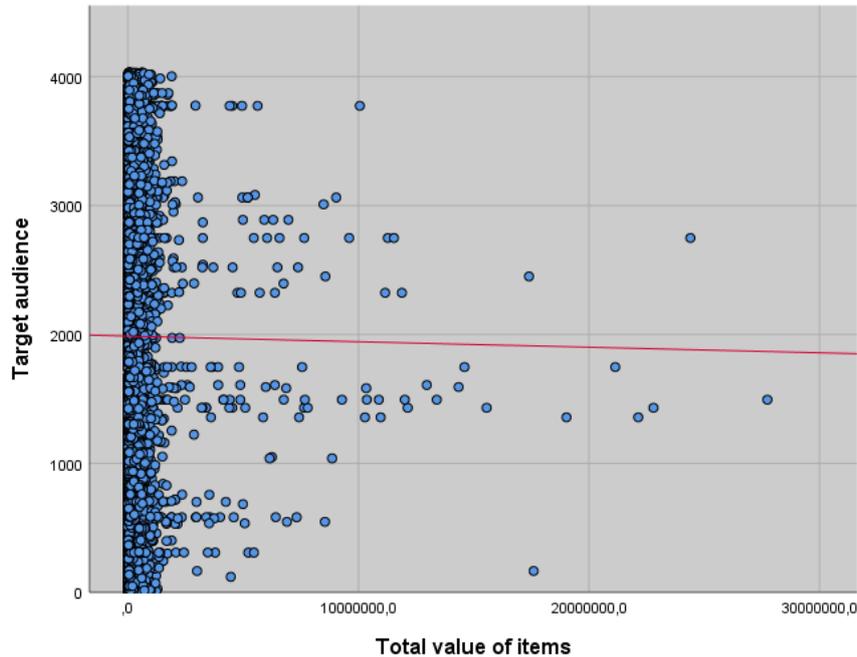


Source: Research data.

The distribution of the benefited population demonstrates a presence across all regions of Brazil, indicating a diverse representation in the execution of the agreements. The South region contributes the most to the total number of people who benefited, closely followed by the Southeast region. This may imply a higher level of interest, involvement, or relevance of the agreement's activities in these areas. Despite the concentration in the South and Southeast regions, the number of people forming the target audience in the North, Northeast, and Cen-tral-West regions is also notable. This balanced distribution may indicate an in-clusive and comprehensive approach to the project, as depicted in Figure 4.

**Page 15**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

Figure 3. Percentage Distribution of Benefited People per Region



Source: Research data.

Regarding the variable "Type_Expenses," movable goods constitute the primary category, encompassing a variety of objects and subcategories, ranging from household utensils to industrial equipment and vehicles. Notably, agricultural and road machinery and equipment comprise almost 70% of these items, followed by permanent materials, which constitute nearly 15%. The municipalities most benefited from movable goods are located within the states of Paraná, Rio Grande do Sul, and Pará.

Table 4 presents statistical information concerning three numerical variables that significantly contributed to this study. These figures reveal that the mean number of items per observation is 5.15, with a high level of dispersion, as indicated by the significantly large standard deviation. A minimum value of 1 suggests at least one observation with just one item, while a maximum value of 3950 indicates a considerable variation in the number of items, with the highest value being quite substantial. Regarding the variable "Total_Value_Item," the values provide insight into the total cost of each observation item. The average is approximately BRL 319,900, yet the dispersion remains high, as evidenced by the considerable standard deviation. The minimum value of BRL 0.01 suggests the presence of observations with very low-cost items (or outliers). In contrast, the

Page 16

REVISTA CADERNO PEDAGÓGICO – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

maximum value of BRL 27.7 million indicates observations with very high costs.

Table 4. Descriptive statistics of some numerical variables in the database

| Variable | Minimum | Maximum | Mean | Standard deviation |
|---|---|---|---|---|
| Quant_Item | 1.00 | 3950.00 | 5.15 | 67.78 |
| Unit_Value_Item | 0.01 | 1980000.00 | 199943.40 | 247053.87 |
| Total_Value_Item | 0.01 | 27738000.00 | 319884.32 | 972267.16 |

Source: Research data.

Figure 5 illustrates the years in which the agreements were signed and the total value of the requested items. A notable growth was observed between 2019 and 2021, with particularly significant values recorded in 2021. However, there was a reduction in comparison to 2022. It is important to note that because the data collection was conducted in 2023, the data for that year was inconclusive.

Figure 5. Relationship between the years the agreements were signed and the total values of the requested items
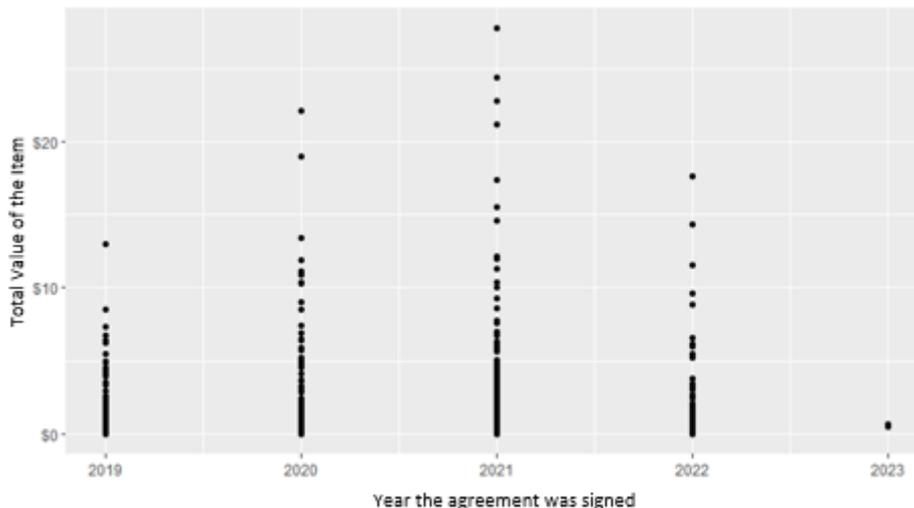


Source: Research data.

Cluster analysis was conducted to categorize regions based on the total values of established agreements. The following steps were applied to perform clustering: utilizing the $fviz\_nbclust$ function from the $factoextra$ package in the R programming language to determine the optimal number of clusters. Additionally, graphical analysis aids in identifying the point at which adding more clusters does not notably reduce intra-cluster variability. From this perspective, **Erro!**

**Fonte de referência não encontrada.** presents the graph that serves as a basis for defining the optimal number of clusters in this research. Through this graph, it is evident that the ideal number of clusters is three.
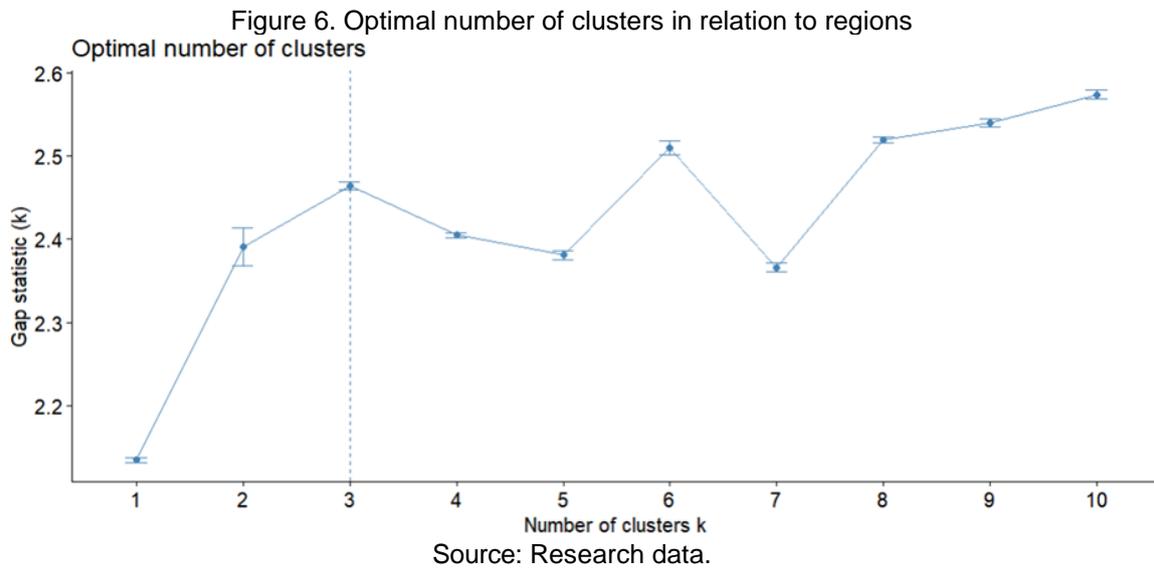
Figure 6. Optimal number of clusters in relation to regions



Source: Research data.

Table 5. Descriptive analysis of the clusters obtained in relation to the Total_Value_Item variable and the regions of Brazil

|  |  | Clusters | | |
|---|---|---|---|---|
|  |  | **1** | **2** | **3** |
| **Total_Value_Item** | Quantity | 65 | 2785 | 7239 |
|  | Minimum | 5196000 | 0 | 5 |
|  | Maximum | 27738000 | 5094000 | 4900000 |
|  | Mean | 10018448 | 336834.7 | 226278.8 |
|  | Standard Deviation | 5166016 | 537111.6 | 342040.6 |
| **State_per_Region** | Minimum | 1 | 1 | 4 |
|  | Maximum | 5 | 3 | 5 |
|  | Mean | 3.738 | 2.099 | 4.653 |
|  | Standard Deviation | 1.278 | 0.792 | 0.476 |

Source: Research data.

Table 5 illustrates that cluster 3 contains the largest number of observations. The total values of the items vary considerably, with an average of approximately BRL 226,300 and a notable standard deviation, indicating a significant dispersion in the values. The distribution of states by region suggests a general consistency in regions 4 and 5, encompassing the South and Southeast regions.

Cluster 2 exhibits significant observations, albeit smaller than cluster 3.

The total values of the items also vary considerably, with a higher average (approximately BRL 336,800) and a high standard deviation, indicating a greater dispersion in the values. The distribution of states by region suggests regional diversity, with a lower average of states per region, encompassing the North, Northeast, and Central-West regions.

Cluster 1 is the smallest in terms of the number of observations. However, it stands out for having much higher total item values, with an average of approximately BRL 10 million and a considerable standard deviation, indicating significant variability in values. The distribution of states by region suggests moderate diversity, with an intermediate average of states per region, covering all regions of the country.

Figure 7 graphically presents the distribution of clusters in relation to Brazilian regions and the total value of the items in each agreement signed by MAPA.

Figure 7. Cluster obtained in relation to regions and the total value of items
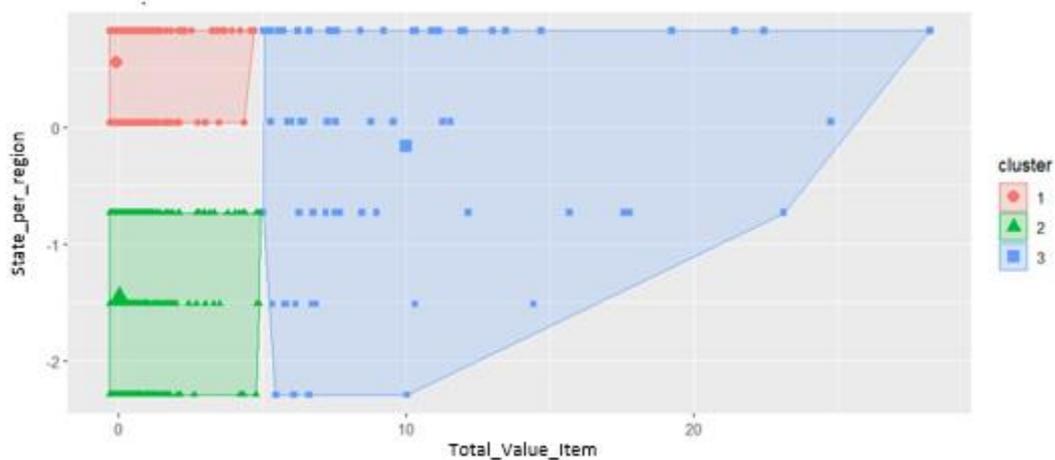


Source: Research data.

# 5 DISCUSSION

## 5.1 K-MEANS CLUSTERS ANALYSIS

The decrease in agreements over the years can be attributed to political factors, the COVID-19 pandemic, and changes in government, reflecting economic instabilities and adjustments in government policies. The regional

distribution indicates varying levels of participation, possibly associated with the projects' relevance in each region. Variations across states and expense categories highlight the diversity and importance of different sectors in implementing agreements.

The analysis of the results reveals significant insights into the dynamics of public agreements signed by MAPA from 2019 to 2023. By employing advanced clustering techniques, notably the k-means algorithm, this study provided an in-depth understanding of the underlying structure of the data, identifying patterns and trends crucial for informing future decisions.

A critical issue in implementing clustering algorithms, especially k-means, lies in pre-specifying the number of clusters. This study utilized the $fviz\_nbclust$ function and the $gap\_stat$ method to determine the optimal number of clusters. This methodology offered a comprehensive perspective of the data, allowing the identification of three relevant clusters to categorize Brazilian regions based on total agreement values. This approach features robustness and adaptability to high-dimensional datasets.

The temporal analysis of agreements, as depicted in Figure 1, offers a fascinating narrative about changes over the years. The reduction in the number of agreements from 2022 to 2023, does not significantly affect the data, as they were collected at the beginning of 2023. This temporal contextualization promotes understanding not only of data variations but also anticipates future scenarios.

The geographic distribution of agreements by state, highlighted in Figure 2, reveals distinct patterns. The state of São Paulo (SP), for example, maintains agreements with regularity. At the same time, Rio Grande do Sul (RS) and Minas Gerais (MG) exhibit significant variations. Including a detailed analysis of the distribution of resources by state, as shown in Figure 3, enriches the discussion by identifying the states that stand out and require greater attention.

This research addresses the decrease in the number of agreements by examining potential influences – such as the pandemic and economic instabilities mentioned before – and underscores the importance of considering them when making future decisions. Understanding the potential impacts of these factors is

decisive for adjusting programs and initiatives and ensuring their continued effectiveness.

The inclusive approach of projects funded by the agreements, as reflected in the distribution of benefited people across Brazil's regions (Figure 4), suggests that MAPA adopts a comprehensive strategy. The detailed analysis of variables such as "Type_Expenses" and the relationship between "Target_Audience" and "Total_Value_Item" adds depth to the discussion, highlighting the diversity in categories and the extensive scope of the projects.

Considering these points, it is evident that applying advanced data analysis techniques significantly contributes to understanding and interpreting complex datasets. Furthermore, contextualizing results within a broader temporal and political landscape provides a solid basis for making informed decisions. This interdisciplinary approach reinforces the relevance and impact of this study in the governmental sphere and the management of public policies.

By conducting a detailed statistical analysis of the values of the agreements, exploring their association with geographic regions, and utilizing advanced clustering techniques, particularly the k-means algorithm, the formation of distinct clusters revealed significant patterns in the data of the agreements. These clusters not only highlight variability in the values of the agreements but also provide valuable information about the geographic distribution of these resources.

## 6 CONCLUSION

Based on the results examined and the analyses conducted, this research provides essential insights into the geographic distribution, temporal evolution, and characteristics of the agreements established by the MAPA from 2019 to 2023. The detailed examination of the agreements reveals a significant regional distribution, with the South emerging as the leader in the number of agreements, closely followed by the Southeast region. These findings underscore the importance of considering Brazil's geographic diversity when planning and implementing government programs and policies related to agriculture and livestock.

**Page 21**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

Moreover, the temporal fluctuations in the number of agreements over the years shed light on the influence of external factors, such as political changes, the COVID-19 pandemic, and shifts in government administration, on the dynamics of agreements. This temporal analysis emphasizes the necessity for a flexible and adaptive approach to managing government programs to address unforeseen events and ensure the continuity of initiatives.

The diverse nature of the agreements, as evidenced by the analysis of the various stages of implementation, underscores the complexity involved in executing these programs and the significance of effective monitoring to ensure compliance with established objectives. Additionally, the correlation between the total value of the items and the target audience of the agreements reflects the diversity and breadth of activities undertaken by MAPA in collaboration with various sectors of society.

Finally, clustering regions based on the total values of established agreements unveiled distinct patterns of geographic distribution and total values of items, providing a deeper understanding of regional heterogeneity within the scope of MAPA agreements. This clustering technique can yield valuable insights for managing and allocating resources, facilitating a more targeted and efficient approach to implementing public programs and policies.

The findings of this study provide a robust foundation for the development of policies and practices that can have a significant positive impact on society. A comprehensive grasp of the geographical distribution and temporal fluctuations of the agreements established by MAPA enables the formulation of more targeted and efficacious public policies, tailored to the specific requirements of each region. By identifying distribution patterns and areas with greater demand or growth potential, managers can allocate resources more strategically, thereby promoting more balanced regional development. Furthermore, the analysis of temporal variations highlights the importance of adopting flexible and resilient approaches to dealing with crises and guaranteeing the continuity of initiatives. This approach not only improves the efficiency of programs, but also contributes to the sustainable development of the agricultural sector, directly benefiting farmers, ranchers, and society in general.

In terms of research limitations, it should be noted that the article only covered the period from 2019 to 2023. Consequently, recent or future changes that could influence the scenario were not covered. From this perspective, the absence of qualitative data on the perception and satisfaction of the beneficiaries of the agreements also constrains a comprehensive understanding of the actual impact of the programs.

It is recommended that future work extend the period of analysis to include previous and future years, as this will provide a more complete picture of trends and their long-term implications. Moreover, comprehensive regional studies can facilitate the identification of particular needs and contexts that may not be discernible in a national analysis. The combination of these methods can facilitate the attainment of a more comprehensive and nuanced perspective, thereby enhancing the efficacy of policies and programs pertaining to agriculture and livestock.

This research contributes to a comprehensive understanding of the agreements signed by MAPA. It highlights the significance of considering geographic, temporal, and contextual factors in formulating and executing government policies and programs related to agriculture and livestock. This knowledge can inform future decisions and enhance the effectiveness and efficiency of MAPA's initiatives in fostering the sustainable development of the Brazilian agricultural sector.

## ACKNOWLEDGEMENTS

# REFERENCES

AGUIAR, P. A. d. A., SANTANA JÚNIOR, C. J., & BASTOS FILHO, C. J. A. (2018). Aplicação de algoritmos de clusterização em uma base de dados de reservas de hotéis. **Revista de Engenharia e Pesquisa Aplicada**, 3.

ALONSO-BETANZOS, A., & BOLÓN-CANEDO, V. (2018). Big-data analysis, cluster analysis, and machine-learning approaches. **Sex-specific analysis of cardiovascular function**, 607–626.

BORGES, M. J., & PARRÉ, J. L. (2021). O impacto do crédito rural no produto agropecuário brasileiro. **Revista de Economia e Sociologia Rural**, 60.

BRAGA, M. J., VIEIRA FILHO, J. E. R., & de FREITAS, C. O. (2019). IMPACTOS DA EXTENSÃO RURAL NA RENDA PRODUTIVA. In **Diagnóstico e desafios da Agricultura Brasileira** (chapter Impactos d, p. 340). IPEA, Rio de Janeiro.

CAPANO, G., & HOWLETT, M. (2020). The knowns and unknowns of policy instrument analysis: Policy tools and the current research agenda on policy mixes. **Sage Open**, 10(1):2158244019900568.

CHAVES, R. S., CAMPELO FILHO, E., & PINTO, R. S. (2022). Panorama das transferências voluntárias brasileiras no período 2008-2016. **Economia & Região**, 10(2):139–159.

CORDEIRO, D. F., da SILVA SOUZA, L. R., LIMIRO, R. M., & da SILVA, N. R. (2023). Convênios públicos no fomento à agricultura familiar: análise exploratória face à pandemia da covid-19. **Revista de Gestão e Secretariado**, 14(2):2211–2234.

COSTA, C. C. d. M., FERREIRA, M. A. M., BRAGA, M. J., & 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**, 52–88.

DURAN, B. S., & ODELL, P. L. (2013). **Cluster analysis: a survey**, volume 100. Springer Science & Business Media.

FAYYAD, U., PIATETSKY-SHAPIRO, G., & SMYTH, P. (1996). From data mining to knowledge discovery in databases. **AI magazine**, 17(3):37–37.

FERNANDES, F. T., & CHIAVEGATTO FILHO, A. D. P. (2019). Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. **Revista Brasileira de Saúde Ocupacional**, 44.

FRADES, I., & MATTHIESEN, R. (2010). Overview on techniques in cluster analysis. **Bioinformatics methods in clinical research**, 81–107.

**Page 24**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

GRISA, C., & SCHNEIDER, S. (2014). Três gerações de políticas públicas para a agricultura familiar e formas de interação entre sociedade e estado no brasil.

HOENINK, J. C., EISINK, M., ADAMS, J., PINHO, M. G., & Mackenbach, J. D. (2023). Who uses what food retailers? a cluster analysis of food retail usage in the netherlands. **Health & place**, 81:103009.

JAEGER, A., & BANKS, D. (2023). Cluster analysis: A modern statistical review. **Wiley Interdisciplinary Reviews: Computational Statistics**, 15(3).

JIANG, Y., KING, J. M., & PRINYAWIWATKUL, W. (2014). A review of measurement and relationships between food, eating behavior and emotion. **Trends in Food Science & Technology**, 36(1):15–28.

LASCOUMES, P., & LE GALÈS, P. (2007). Introduction: Understanding public policy through its instruments—from the nature of instruments to the sociology of public policy instrumentation. **Governance**, 20(1):1–21.

LAURETT, R., PAÇO, A., & MAINARDES, E. W. (2021). Sustainable development in agriculture and its antecedents, barriers and consequences–an exploratory study. **Sustainable Production and Consumption**, 27:298–311.

LUI, L., & MIQUELINO, W. (2023). Evolução dos convênios celebrados pelo ministério da agricultura com os entes subnacionais no brasil. **Revista de Economia e Sociologia Rural**, 61.

MACQUEEN, J. (1967). Some methods for classification and analysis of multivariate observations. **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**, 1(14):281–297.

MAIONE, C., de OLIVEIRA SOUZA, V. C., TOGNI, L. R., da COSTA, J. L., CAMPIGLIA, A. D., BARBOSA, F., & BARBOSA, R. M. (2017). Using Cluster Analysis and ICP-MS to Identify Groups of Ecstasy Tablets in Sao Paulo State, Brazil. **Journal of Forensic Sciences**, 62(6):1479–1486.

MAMEDE, P., FRAY, R., & JUNIOR, I. D. B. (2023). Deep Learning and Statistical Models for Forecasting Transportation Demand: A Case Study of Multiple Distribution Centers. **Logistics**, 86:1–19.

MANGIATERRA, B. (2020). A situação e os desafios do pequeno agricultor no brasil. **Revista Cultivar**. Recuperado de https://revistacultivar.com.br/artigos/a-situacao-e-os-desafios-do-pequeno-agricultor-no-brasil.

MAPA, Ministério da Agricultura e Pecuária (2023). Atribuições. **https://www.gov.br/agricultura/pt-br/acesso-a-informacao/institucional/o-ministerio/atribuicoes**. Acessado em 28 de novembro de 2023.

**Page 25**

**REVISTA CADERNO PEDAGÓGICO** – Studies Publicações e Editora Ltda., Curitiba, v.21, n.9, p. 01-26. 2024.

MATENGA, Z. (2022). Assessment of energy market's progress towards achieving sustainable development goal 7: A clustering approach. **Sustainable Energy Technologies and Assessments**, 52:102224.

MENG, Y., LIANG, J., CAO, F., & HE, Y. (2018). A new distance with derivative information for functional k-means clustering algorithm. **Information Sciences**, 463-464:166–185.

MOUTINHO, J. d. A. (2016). Transferências voluntárias da união para municípios brasileiros: mapeamento do cenário nacional. **Revista de Administração Pública**, 50:151–166.

NEMITZ, M. C., ARGENTA, D. F., KOESTER, L. S., BASSANI, V. L., von POSER, G. L., & TEIXEIRA, H. F. (2016). The international scenario of patents concerning isoflavones. **Trends in food science & technology**, 49:85–95.

PORTAL DA TRANSPARÊNCIA (2024). **Portal da transparência:** Convênios e outros acordos. Disponível em: https://portaldatransparencia.gov.br/entenda-a-gestao-publica/convenios-e-outros-acordos#:~:text=Conv%C3%AA-nio%2C%20contratos%20de%20re-passe%20e,execu%C3%A7%C3%A3o%20de%20um%20objetivo%20comum. Acesso em 04 de abril de 2024.

PORTAL Gov.br (2024). **Ministério da agricultura e pecuária**: Atribuições. Disponível em: https://www.gov.br/agricultura/pt-br/acesso-a-informacao/institu-cional/o-ministerio/atribuicoes. Acesso em 04 de abril de 2024.

R Studio Team. (2023). **R Studio 4.2**. R Studio, PBC. Versão 4.2.

SCHÖSLER, H., de BOER, J., & BOERSEMA, J. J. (2014). Fostering more sustainable food choices: Can self-determination theory help? **Food Quality and Preference**, 35:59–69.

SGROI, F., SCIORTINO, C., BAVIERA-PUIG, A., & MODICA, F. (2024). Analyzing consumer trends in functional foods: A cluster analysis approach. **Journal of Agriculture and Food Research**, 101041.

SINAGA, K. P., & YANG, M. S. (2020). Unsupervised K-means clustering algorithm. **IEEE Access**, 8:80716–80727.