Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.3383

Analysis of Expenses from Brazilian Federal Deputies between 2015 and 2018

Felippe Pires Ferreira [**Universidade de São Paulo** | ***felippe_pires@usp.br***]
Ilan S. G. de Figueiredo [**Universidade de São Paulo** | ***ilan.figueiredo@usp.br***]
Larissa R. Teixeira [**Universidade de São Paulo** | ***rteixeira.larissa@usp.br***]
William Zaniboni Silva [**Universidade de São Paulo** | ***williamzaniboni@usp.br***]
Caetano Traina Junior [**Universidade de São Paulo** | ***caetano@icmc.usp.br***]
Cristina Dutra de Aguiar [**Universidade de São Paulo** | ***cdac@icmc.usp.br***]
Robson L. F. Cordeiro [**Carnegie Mellon University** | ***robsonc@andrew.cmu.edu***]

Computer Science Department, University of São Paulo (USP), Av. Trabalhador São-carlense, 400, Centro, São Carlos, SP, 13566-590, Brazil.

Received: 6 May 2023 • **Published:** 13 January 2025

Abstract The analysis of public expenses is fundamental to foster the correct use of public resources, guaranteeing the application of the principles of publicity and efficiency. Within the scope of the Brazilian parliament, Parliamentary Quotas are also identified as public resources, therefore they need to be subject to the same control criteria. This research aims to carry out analyzes of parliamentary expenses related to Parliamentary Quotas, presenting the distribution of expenses related to the 55th Legislature (2015-2018) of Brazil, in addition to identifying anomalies in such expenses. Through a clustering-based analysis, the expenses were compared with the goal of finding similarities between the spending behavior of the federal deputies. This study, through data mining, presents the results obtained from analyzing different parliamentary expenses under the party or regional aspect of each deputy. The results obtained allowed us to answer questions related to the characteristics of the expenses involving Parliamentary Quotas, anomalous expenses, and similarity between parliamentary expenses, such as, the identification of expenditure patterns, which allow the verification of regional variability, as well as identifying some of the expenditures as possibly anomalous.

Keywords: Analysis of Expenses, Public Expenses, Brazilian Deputies, Outliers

1 Introduction

Public expenses are the subject of recurrent analysis by control and inspection entities, in addition to being of direct interest to the civil society [Paranhos *et al.*, 2015; Bueno, 2017; Thiago Alencar Gomes, 2017] since they must serve the public interest and respect the constitutional principles. Within the constitutional principles, we can mention: legality, impersonality, morality, publicity, and efficiency, according to the Constitution of Brazil in article 37 [Brasil, 1988]. In short, public spending must be used in accordance with legal guidelines, so as not to be directed to favor any interested third party, and not be used in goods or services that do not comply with moral or ethical standards, guaranteeing transparency and availability of its use to any Brazilian citizen, to achieve clear objectives and with as little waste as possible.

In order to ensure primarily the principle of publicity, the governments of the three branches of power make available the data on public expenditures through electronic portals or publications in official journals, aiming to describe the values used during their management, the products and services acquired, and the suppliers contracted in these procedures. Within the scope of the Federal Legislative Power, more specifically in the Chamber of Deputies, in addition to the expenses related to the contracts held by the institution, federal deputies receive monthly allowances to cover their office expenses throughout their legislature, known as *Quota for the Exercise of Parliamentary Activity* [1]. The value of the Quota depends on the state of the federation to which the parliamentarian belongs to, and it is adjusted according to the inflation. It should be noted that the expenses of the Parliamentary Quota include: the maintenance of activities in the parliamentary office or participation in events; transportation tickets; consulting or services necessary for the operation of the office; travel and accommodation expenses; and acquisition of materials.

The focus of this work is to analyze the parliamentary expenses of the Chamber of Deputies, more specifically the expenses of the Parliamentary Quotas of the 55th Legislature, which corresponds to the period from 2015 to 2018. Our study involves data science activities, aiming to create visualizations and interpretations for the data available, besides the search for expenses with strange patterns when compared to other expenses in the dataset. Additionally, with the use of clustering algorithms over the expenditure information of each deputy, we target to group the deputies, either both by party or region, depending on spending patterns over the legislature. At the end of the research, we provide answers to the following questions:

Q1 Exploratory Data Analysis: How do parliamentarians use the values of the Parliamentary Quota, considering

[1] https://www2.camara.leg.br/transparencia/acesso-a-informacao/copy_of_perguntas-frequentes/cota-para-o-exercicio-da-atividade-parlamentar

parties and regions, according to the distribution of expenses and the trends of each expense?

Q2 Outlier Detection: Are there deputies' expense entries that may be considered abnormal in comparison with those of other parliamentarians?

Q3 Cluster Analysis: Can the expenses made during the four years of the legislature be grouped by similarity, thus identifying deputies who have common behaviors?

The rest of our article is organized as follows: Section 2 presents the necessary theoretical foundation to understand the work's methodology; Section 3 discusses the related work; Section 4 details the dataset used and the algorithms employed to answer the presented questions, while Section 5 presents the information found relating the data and data mining techniques. Finally, Section 6 presents conclusions and discusses the findings produced by our research.

2 Background

This section introduces required basic concepts. In particular, we briefly discuss concepts and techniques related to data mining and the analysis of time series.

2.1 Data Mining

Data mining is a technique used to find patterns in a large data set with the aim of supporting decision making in enterprises and other corporations in general. Outlier detection and clustering are among the main tasks of data mining [Han *et al.*, 2022]. They are introduced in the following.

2.1.1 Outlier Detection

Outlier detection (also known as anomaly detection) is the process of finding points (also known as instances or objects) in a dataset that are much different from the majority of other points [Han *et al.*, 2022]. Outliers are interesting because they are suspected of not being generated by the same mechanisms as the rest of the data [Dang *et al.*, 2015], thus indicating potential frauds, invasions and other rare phenomena, or even errors in the data collection. The kNN-outlier detection method [Angiulli and Pizzuti, 2002] detects outliers by exploring the relationship between neighborhoods in data points through the use of a distance function that quantifies the dissimilarity between the points. The further a data point is from its neighbors, the greater its chance to be an outlier. Other well-known methods for outlier detection are LOF [Breunig *et al.*, 2000], iForest [Liu *et al.*, 2008], ABOD [Kriegel *et al.*, 2008] and LOCI [Papadimitriou *et al.*, 2003].

2.1.2 Clustering

Clustering is the process of dividing a data set into groups, such that points belonging to the same group are similar to each other and dissimilar to the points in other groups. Clustering methods can be classified into two main categories: partitioning methods and hierarchical ones.

Hierarchical methods produce a hierarchy with several levels consisting of nested partitions of the dataset according to the similarity between the points. Partitioning methods try to find the best k partitions of a data set by dividing the set into k subsets. Most partitioning methods are based on distance (dissimilarity), like k-means, or on density, such as DBScan [Ester *et al.*, 1996] and OPTICS [Ankerst *et al.*, 1999]. k-means is probably the most popular clustering algorithm due to its simplicity and efficiency [Krishna and Murty, 1999]. It divides the data set into k groups where each group has a center (average value of the points within the group). The groups are generated in such a way that the distance between the points in a group and the corresponding group's center is minimized.

One of the problems in k-means is finding the ideal value of k for the analysis. This problem is commonly tackled through the Elbow Method [Bholowalia and Kumar, 2014], which consists in identifying the ideal k as the one that generates an 'elbow' in the plot of k versus the sum of the squared distances between each point and the corresponding cluster's center.

2.2 Analysis of Time Series

Data mining techniques can be applied to different datasets, such as time series. A time series is defined as a sequence of values collected from a phenomenon or object of interest in distinct time instants. The values are normally measured at equal time intervals, e.g., every minute, hour or day [Han *et al.*, 2022]. In this setting, the distance function Dynamic timewarping (DTW) [Bellman and Kalaba, 1959] is commonly used to measure the dissimilarity between the time series. Essentially, it finds trends between two temporal sequences, which may vary in speed [Mueen and Keogh, 2016].

3 Related Work

This section discusses several studies focused on analyzing public expenses in diverse contexts. We begin by presenting the studies performed with **general public expenses**. Then, in Subsection 3.1, we move on to those studies that **evaluate parliamentary public expenses during the legislature.**

Esen and Celik Kecili [2022] studied public expenses to understand their relationship with a country's economic growth. The study focused on the factors that drive economic growth over time and the analysis of the forces that allow some economies to grow quickly, while others grow slowly, and others do not grow at all. Specifically, the authors aimed to analyze the effects of health expenditure on economic growth in Turkey in a time series from the period 1975-2018. The study showed that there is an unidirectional causality from health expenditure to economic growth in the short term scenario, and also that there exists cointegration between all variables in the long term scenario. While this study analyzes the impact of health expenditure on economic growth, it does not take into account other public expenses in their analysis.

Heiler *et al.* [2016] analyzed the expenses of an electoral campaign. Their study focused specifically on the electoral expenses for the 2010 election in Brazil. As a result, it was observed that the focus on communication and publicity expenses, combined with the participation in more organized and centralized parties, are essential for the electoral success. This study delves into electoral campaign expenses, whereas our work focuses on a broader spectrum of public expenses over multiple years.

Anomaly detection techniques have been used previously in data of public expenditure. For instance, Ravisankar *et al.* [2011] analyzed bidding and stock market data using Feed Foward Neural Networks, Support Vector Machines, Generic Programming, Group Method of Data Handling, and Regression and Probabilistic Neural Network. As a result, the authors were able to identify companies that were later found guilty of fraud in China. Despite identifying anomalies, the work is directed to the Chinese context, which differs from the Brazilian reality.

When considering public spending in general, there is a need to monitor and detect anomalies of legality in bidding values. With that in mind, da. Silva *et al.* [2022] presented a framework that detects irregularities in bidding processes performed in Brazil based on the current local legislation, thus serving as a supporting tool for decision making. The Key Performance Indicators (KPI's) by region were introduced and it was also investigated the association between bidding items with evidence of irregularity using the Apriori algorithm. Additionally, the authors identified potentially suspect companies that won or lost all the bidding processes they participated. Despite containing an analysis of public expenses, the research analyzes problems related to bidding, which has no relation to the object of our research, which is parliamentary spending.

3.1 Analysis of Parliamentarias Public Expenditure

Paranhos *et al.* [2015] performed a statistical study of Brazilian parliamentary expenses related to Parliamentary Quotas. The authors restricted their study to only two Brazilian states (Ceará and Paraíba), selecting the period between 2007 and 2010. The study employed descriptive statistics and principal component analysis, which allowed the authors to conclude that parliamentarians from these regions used the amounts received mainly for transportation tickets, daily expenses and consulting. Unlike this study that focuses on specific states, our research covers all Brazilian states and a more recent period (2015-2018) to provide a comprehensive view of public expenses.

Other possible sources of financial fraud are the expenses of parliamentarians. With that in mind, Thiago Alencar Gomes [2017] analyzed expenditure statements of the Brazilian chamber of deputies, from 2009 to 2017, aimed at identifying outliers using Deep Autoencoders. The expenditures were analyzed individually, without being grouped by candidate or party. Although the model found suspicious entries, such as expenses with energy company and telephone bills, the method demands a large processing time and a parallel architecture to complete the analysis in a feasible time, which prevented the authors from realizing a more complete analysis. In contrast to Thiago Alencar Gomes [2017] that uses Deep Autoencoders to analyze individual expenditure statements, our work employs the k-means algorithm on Parliamentary Quotas data to identify general spending patterns, including anomalous behavior, over a specific legislative period (2015-2018). Unlike our study, the author chose to analyze the deputies individually, not carrying out analyses by party or state.

Bueno [2017] presented an analysis of outliers based on expenditure data of Brazilian federal deputies from 2013 to 2016. A prediction of expenditures was also made for the year 2017. Regarding the detection of outliers, the k-means algorithm was used, with $k = 5$. The authors studied both the entire period (2013 to 2016) and each year individually. The analyses consider the costs of each party's expenditure types per month. As a result, a total of 25 outliers were detected, out of which 23 are related to expenditures for publicizing parliamentary activity. The author chose to carry out separate analyses by year and didn't carry out studies differentiating the types of expenses produced by deputies.

In conclusion, our work aims providing a comprehensive analysis of Parliamentary Quotas of the 55th Legislature (from 2015 to 2018) in Brazil. We utilize the k-means algorithm to uncover general spending patterns within different categories of public expenses. Additionally, we employ time series analysis, utilizing the Dynamic Time Warping (DTW) function to identify anomalies in parliamentary expenses. By combining clustering algorithms and time series analysis, our research contributes to a more nuanced understanding of the behavior of parliamentary expenses, contributing to the existing body of knowledge on public expenditure analysis.

4 Methodology

Here we describe the methodology employed to answer the research questions presented at introduction. In a nutshell, we followed three main steps: preparation, selection and analysis of the data. Section 4.1 presents the preparation, selection and analysis processes related to the **Exploratory Data Analysis**; the corresponding processes regarding the questions of **Outlier Identification** and **Clustering** are described in Sections 4.2 and 4.3, respectively.

4.1 Exploratory Data Analysis

Here we focus on answering the first question **(Q1)**: *How do parliamentarians use the values of the Parliamentary Quota, considering parties and regions, through the analysis of the distribution of expenses and the trends of each expense?*. To this end, we began by performing a data cleaning and preparation. The data related to the use of the Parliamentary Quotas of the federal deputies had been made available publicly at the Transparency Portal of the Federal Chamber of Deputies (in Portuguese, *"Portal da Transparência da Câmara dos Deputados"*) [2]. The data is available in a .csv file, which was obtained from the aforementioned portal. The data was processed using scripts to MinMax normalize the values.

The expenses are presented in a monthly basis: each expense is linked to a deputy (with the corresponding name, unique identifier, federal unity and party) and presented as a receipt (in Portuguese, "*nota fiscal*") with the name and CNPJ/CPF of the provider, the issuing date and the amount

[2] `https://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar`

spent. Each expense is categorized into one of 23 spending classes, including, for example, airfare, postal services, accommodation and among others.

The data contains expense postings from 2009 to 2020. The purpose of our work is to analyze the most recent legistature that is entirely covered; thus, we studied expenses regarding the period of office from 2015 to 2018, aimed at finding patterns in the records. For this period, only the deputies who registered every month in every year were considered, which led to a total of 242 deputies out of the 513 deputies. Therefore, roughly 53% of the original dataset had to be excluded due to the lack of entries of the expenditures of the deputies. The percentage of deputies excluded by region are 65.8% on Midwest, 64.9% on Northeast, 64.6% on North, 54.1% on Southeast, and 53.2% on South. Note that the percentages are relatively well balanced. For a better understanding of the profile of the deputies with missing data, we also separated them by party and grouped according to the number of missing records, as shown later in Section 5.1. We standardized the expenditure values according to the inflation using the IPCA index Khol [1992], which is the official inflation index of the Brazilian Central Bank (in Portuguese, *"Banco Central do Brasil"*). In addition, the expenses were grouped into five large categories:

- **Parliamentary Activity**: 1. Maintenance of an office to support parliamentary activity, 2. Dissemination of parliamentary activity, 3. Participation in a course, lecture or similar event.
- **Transportation Tickets**: 1. Air ticket - RPA, 2. Air ticket - reimbursement, 3. Land, sea or river tickets, 4. Air ticket - SIGEPA.
- **Consulting and Security Services**: 1. Consulting, research and technical work, 2. Security service provided by a specialized company.
- **Daily Expenses**: 1. Locomotion, food and accommodation, 2. Fuel and lubricants, 3. Leasing of motor vehicles or chartering vessels, 4. Provision of food for the parliamentarian, 5. Accommodation, except for the parliamentarian in the federal district , 6. Lease or charter of aircraft, 7. Lease or charter of motor vehicles, 8. Taxi service, tolls and parking, 9. Lease or charter of vessels.
- **Materials and General Services**: 1. Telephony, 2. Acquisition or licensing of software, 3. Acquisition of office supplies, 4. Postal services, 5. Subscription to publications.

Thus, instead of being represented in 23 expense categories, expenses are represented by only 5 categories. Finally, the data were analyzed over the four years in order to illustrate behavioral trends for each expense.

4.2 Outlier Detection

Here we focus on answering the second question **(Q2)**: *Are there any expenditure entries for deputies that may be considered abnormal in comparison with those of other deputies?*. To this end, we separated the problem into two scenarios: (1) analysis of the expenses during the full legislature; and (2) analysis of only during months of parliamentary recess. The emphasis on the parliamentary recess was given because we believe that an improper use of public money would be highlighted in the recess, due to the expected decrease of the legitimate expenses. The dataset employed for the analyses is the one obtained from the data preparation described in Section 4.1. The steps performed for each scenario are: (1) grouping of parliamentary expenses by deputies; (2) identification of the best value of the parameter k for the kNN-outlier algorithm; (3) determination of a threshold to identify expenses as *outliers*; and (4) identification of the expenses considered to be *outliers* according to the threshold. The DTW distance function was employed in this study. The algorithm kNN-outlier was implemented in Python through the package *pyod* Zhao *et al.* [2019].

4.3 Clustering

Here we focus on answering the third and final question **(Q3)**: *Can the expenses made during the four years of the legislature be clustered together, thus identifying deputies who have similar behaviors?*. To this end, we performed the following steps considering initially the whole period of the legislature and then each year of the legislature individually: (1) Deputies considered anomalous in the analysis of Section 4.2 were discarded; (2) For each of the five categories of expenses, the records were normalized in relation to the respective month; (3) The clustering process for each expense category was then performed with the k-Means algorithm, employing the DTW as the distance function; (4) The value of k was chosen using the Elbow Method, as described in Section 2; (5) The implementation and analysis of the clusters were performed in Python, with the algorithms from the package *tslearn* Tavenard *et al.* [2020], which is a package dedicated to clustering time series; (6) Finally, the analysis of the clusters discovered was performed in relation to the deputies' geographic regions and the political parties, taking into account the corresponding year, when applicable.

5 Results

Here we report the results obtained in our study. Specifically, the following three subsections presents each individual results regarding each of the research questions Q1, Q2 and Q3.

5.1 Exploratory Data Analysis

Here we answer the first question **(Q1)**: *How do parliamentarians use the values of the Parliamentary Quota, considering parties and regions, through the analysis of the distribution of expenses and the trends of each expense?* The aim of the exploratory analysis is to describe the distribution of data to complement the other analyses of clustering and detection of outliers.

The dataset comprises the legislature from the years 2015 to 2018, including 29 political parties distributed among the 27 federative units. These states are grouped into five regions, as described in section 4.1. Four analyses were performed on parliamentary expense data: understanding the distribution of expenses by state; understanding the distribution of expenses by party; distribution of expenses in Brazil-

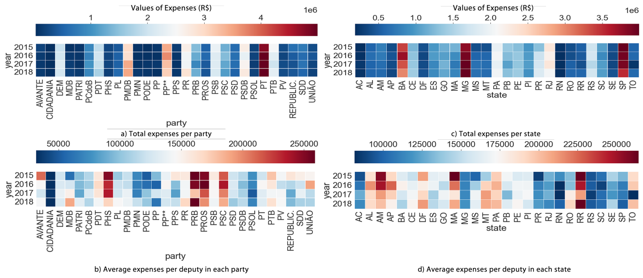

Figure 1. Comparison of deputies' expenses during the legislature (2015-2018) in relation to parties and states. Top: total expenses. Bottom: average expenses per party/deputy.

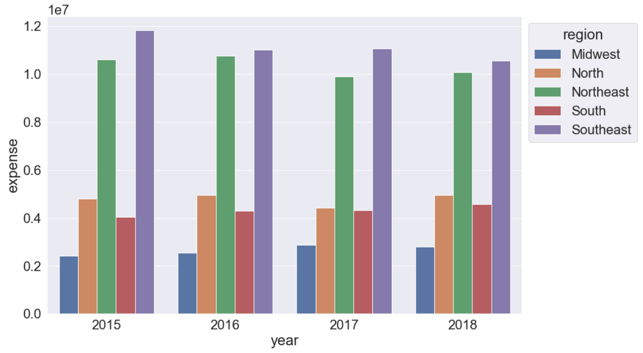

Figure 2. Comparison of deputies' expenses during the legislature (2015-2018) in relation to regions.

ian regions; and how the types of expenses behave during the legislature.

The analysis of expenses in relation to states and parties used a heatmap to illustrate the distribution of values throughout the legislature. Figure 1a illustrates the annual expenses of parties, while Figure 1c shows the expenses of states. We can observe that from the perspective of parties, three parties (PT, PMDB, and PP) stand out and account for more than 33% of the expenses in the legislature. While from the state perspective, the states of BA, MG, and SP also represent more than 33% of the expenses. However, when observing the representation of these parties and states in the Brazilian Congress, we note that they have around one-third of the deputies, which justifies the predominance of these parties and states in the public expenses. Figures 1b and 1d illustrate the average expenses of deputies concerning their parties and states, respectively. In Figure 1b, deputies from smaller parties (PHS, PRB, and PROS) have a higher average expenditure than deputies from traditional parties. Likewise, in Figure 1d, the average spending of deputies from smaller states (RR, AM, and MA) is higher than the average spending of the largest Brazilian states.

Shifting to a regionalized analysis, it is possible to observe that most expenses originate from the Northeast and Southeast regions, as shown in Figure 2. However, when comparing the proportion of deputies from each region to the proportion of expenses produced, we note that the North, Northeast, and Midwest regions have a higher quantity of expenses than the proportion of deputies from their regions, as described in Table 1.

Finally, some analyses were carried out to help understand the distribution of each expense. Figure 3 reports the average and the standard deviation of the monthly expenses made by all deputies during the entire legislature. By analyzing the legislature's spending averages, it is possible to observe spending peaks, seasonal patterns, and, particularly regarding Transportation Tickets, a long-term trend of growth over the months. For example, in Figure 3a, it is possible to observe peaks in the expenses that precede the change from one year to another. During the four years, the expenses of Parliamentary Activities immediately before the month of January present a sharp increase when compared with other months of the same year. It's important to observe that the expenses described in Figures 3a, 3b, and 3c have the largest values in the election period, i.e., October 2018. Particularly, these results may indicate last minute expenses executed to use the year's budget. If this assumption is true, it would be questionable whether the expenses were really necessary. End-of-year spending spikes also occur in other expense categories, the Consulting and Security Services; see Figure 3b. Additionally, when observing those expenses in Figure 3b, it is possible to notice two patterns of similar expenses in the first two years of the legislature. Finally, in Figure 3c, it is possible to verify a clear trend of growth in the expenses regarding Transportation Tickets throughout the legislature. Figure 3d does not have a clear pattern, whereas Figure 3e maintains a standard range, with the exception of a peak in the transition from the first to the second year.

Table 1. Relationship between the number of deputies per region and the proportion of expenses in the legislature.

Regions	Deputies (%)	Expenses (%)					
		2015	2016	2017	2018	Mean	Std
North	10,33	14,23	14,75	13,55	15,04	14,39	0,65
Northeast	26,03	31,54	32,05	30,38	30,56	31,13	0,79
Midwest	6,61	7,16	7,58	8,83	8,5	8,01	0,77
Southeast	39,26	35,1	32,81	33,96	32,03	33,47	1,34
South	17,7	11,97	12,81	13,28	13,87	12,98	0,80

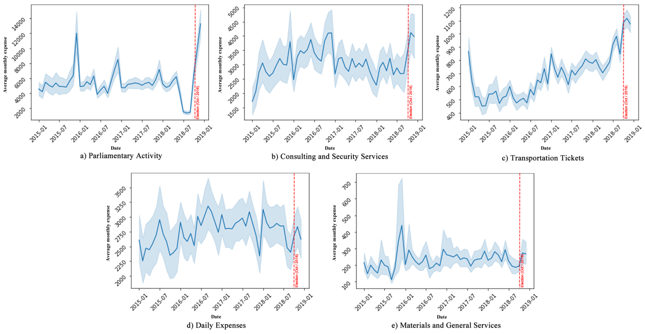

Figure 3. Average and standard deviation of the monthly expenses from each category over the four years of the legislature. Red dashed lines indicate the election period.

5.1.1 Considerations and Restrictions of the Data

Despite existing 513 deputies in the Brazilian Congress, the parliamentary expense data was not reported by every parliamentarian, due either by to omission by the public agent, or by the impossibility of disclosing the data because the deputy didn't complete the legislature. From the 513 existing deputies, only 242 published their expenses for all months of the legislature. Figure 4 illustrates the proportion of deputies who fully or partially reported their expenses. The x-axis shows the proportion of deputies related to the number of months that were disclosed. There are four groups of information related to the number of months of expenses pub-

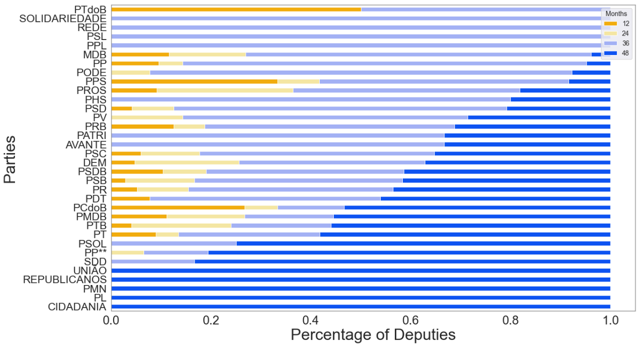

Figure 4. Proportion of deputies per party by number of months with expenses published. Parties are sorted by the number of months with published expenses.

lished by each parliamentarian. The y-axis lists the political parties. As shown, nearly all parties have deputies that did not make their expenses public.

5.2 Outlier Detection

Here we answer the second question **(Q2)**: *Are there deputies whose expenditure entries may be considered abnormal in comparison with those of other deputies?* It is also the focus of our research to make comparisons between parliamentary expenses, looking for situations that are unusual or considered anomalous when compared with other expenses. The analysis of abnormalities was carried out separately for each category of expenses. Our goal is to identify amounts spent that have significant differences when compared with the expenses of other deputies. Two sets of data were created, with 242 instances each, one per deputy. At first, all entries of the deputies are considered, where each deputy is represented by an array of 48 values that correspond to the months of the four years of the legislature. The second analysis considers only the months of the periods of parliamentary recess, that is, the months of January, July, and December. Thus, there are 12 values for each deputy. The analyses were carried out independently for each of the five categories of expenses.

5.2.1 *Outliers* in the entire period of the legislature

To identify outliers in the expenses, the kNN-outlier algorithm was selected, together with the DTW distance function to perform the similarity comparisons over the time series. Five values were considered for the parameter k of the algorithm (3, 5, 7, 10 and 20), and, through a method similar to the *Elbow Method*, the best value for k was selected for each of the categories of expenses evaluated. Specifically, the ideal value for k was selected by how the *outlierness scores* decrease in the ranking of deputies generated by the kNN-outlier algorithm, and identifying the sharp deviations in the scores. Figure 5 is an example of the outlierness scores obtained with each value of k for the Daily Expenses. The scores are presented in decreasing order, so the order of the deputies shown in the X axis of each plot is the corresponding ranking of deputies by their outlierness. As shown, each plot presents a ”knee” (i.e., a sharp variation in the scores) indicating that the first $\sim$ 10 deputies in the ranking have considerably larger scores than the others.

It also provides us with a threshold to distinguish outliers and inliers. Consider the dashed, vertical line in Figure 5: each line identifies the limit that separates data considered outliers from regular data (inliers). Score values equal to or greater than this intersection are considered abnormal. The more abrupt change in the direction of the curve of scores, the better the identification of the threshold that separates the two sets of data (outliers and inliers).

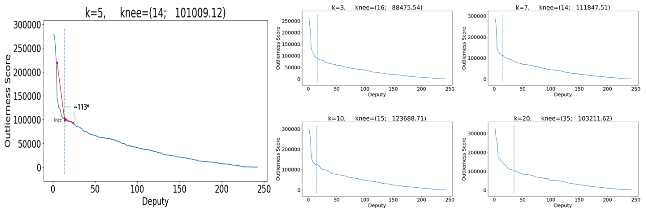

Figure 5. Rankings of deputies by their outlierness according to the Daily Expenses and the value of k. The ”knee”, shown in plot, and identified with the method [Satopaa *et al.*, 2011], reveals that k=5 is the configuration that best distinguishes outliers and inliers.

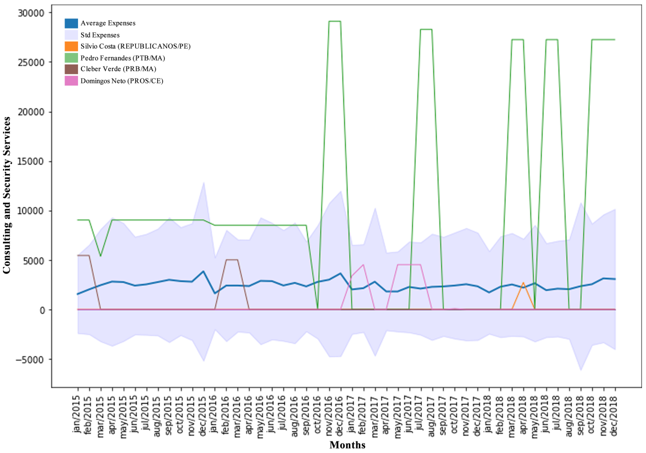

Figure 6. Comparison between the average and the outlying expenses regarding Consulting and Sercurity Services for the entire legislature.

Using the knee identification method [Satopaa *et al.*, 2011], it was possible to determine a ”knee” point for each plot. However, it is still necessary to identify the best value of k for each expense. To identify the best k, we use 10 points before the knee point and 10 points after the knee point to create, through linear regression, two straight lines (red lines). The inclination angles of each straight line were identified, and the variation between them was calculated. The smallest angle formed by the two straight lines represents the best value of k for the expense. This procedure was repeated for all expenses, with the five pre-defined k values. In Figure 5, we could identify that best k is k=5, because there is the smallest angular variation.

After selecting the best values of k for each category of expenses, the deputies selected as outliers are identified, allowing the construction of plots to illustrate the dispersion of anomalies when compared with the average of expenses. For example, Figure 6 presents the expenses in Consulting and Security Services. The blue line represents the average value of this expense category for the four years of the legislature. The shaded region around it represents the standard deviation. The lines in orange, green, brown and pink shows the outliers identified for the respective category of expenses. In addition to expenses that have values much higher than the average, sets of data that have extremely low values, i.e., near zero (Silvo Costa-REPUBLICANOS/PE), are also represented as anomalies.

Table 2. Deputies who have expenses considered to be outliers in the time series of the entire legislature.

Parliamentary Activity	Transportation Tickets	Consulting and Security Services	Daily Expenses	Materials and General Services
Arnaldo Faria De Sá (PTB/SP)	Arlindo Chinaglia (PT/SP)	Silvio Costa (REPUBL./PE)	Wladimir Costa (PMDB/PA)	Hermes Parcianello (PMDB/PR)
Zé Geraldo (PT/PA)	Átila Lins* (PMDB/AM)	Pedro Fernandes* (PTB/MA)	Átila Lins* (PMDB/AM)	Josué Bengtson (PTB/PA)
Carlos Manato (PDT/ES)	Pedro Chaves (PMDB/GO)	Cleber Verde* (PRB/MA)	Paes Landim (PTB/PI)	Jair Bolsonaro (PP/RJ)
Josué Bengtson* (PTB/PA)	Assis Carvalho (PT/PI)	Domingos Neto (PROS/CE)	Silas Câmara* (PSC/AM)	Eduardo Da Fonte (PP/PE)
Silas Câmara (PSC/AM)	Jerônimo Goergen (PP/RS)	-	Nilton Capixaba (PTB/RO)	Lelo Coimbra (PMDB/ES)
Giacobo* (PR/PR)	-	-	Giacobo* (PR/PR)	Eros Biondini (PTB/MG)
Gonzaga Patriota (PSB/PE)	-	-	Pedro Fernandes* (PTB/MA)	-
Cleber Verde* (PRB/MA)	-	-	Átila Lira (PSB/PI)	-
Lázaro Botelho (PP/TO)	-	-	Givaldo Carimbão (PSB/AL)	-
Pr. Marco Feliciano (PSC/SP)	-	-	-	-
Félix Mendonça Júnior (DEM/BA)	-	-	-	-

*Deputies appearing as *outliers* in more than one category of expenses.

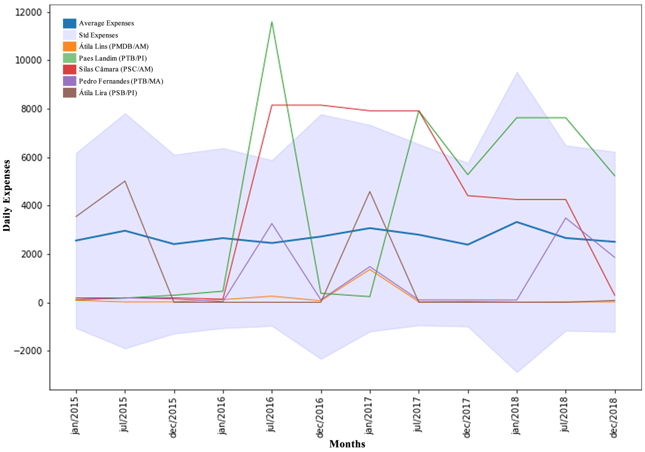

Figure 7. Comparison between the average and the outlying Daily Expenses for the months of parliamentary recess.

Table 2 shows the deputies who presented expenses considered as *outliers* when observing valid entries in the dataset. It is possible to see that some deputies appear in more than one category of expenses. In addition, it was identified that the majority (88%) of outlying deputies regarding Daily Expenses are from the North and Northeast regions of Brazil, while the South and Southeast regions had 2/3 of the anomalies for Materials and General Services. From a party perspective, the parties PMDB, PTB, and PP generated 55% of the outliers identified.

5.2.2 Outliers during the recess

Similarly, the second analysis started with the search for the best value for the parameter k of the algorithm kNN-outlier, and the threshold that distinguishes outliers from inliers. It was performed independently for each category of expenses. Subsequently, we identified the instances that presented scores equal to or greater than the threshold of each category. To exemplify the difference in values between regular and outlying instances, a comparative plot was created between the average amount of Daily Expenses in recess periods and the abnormal ones; see Figure 7. Importantly, note that the outlying deputies (in orange, green, red, pink, and brown) present expense patterns that are indeed remarkably distinct than the average (in blue).

Table 3 reports the deputies who had their expenses tagged as abnormal during the parliamentary recess. Some of them appear as anomalies in more than one category of expenses. It should be noted that, in this scenario, near-zero expense values for all months (Atila Lins-PMDB/AM) of the time series are also considered as outliers. The proportionality of outliers per state and parties presented in the first scenario were maintained in the second scenario with minor variation, expect for the caveat that the state of Rio Grande do Sul had 57% of the anomalies regarding Transportation Tickets, and only states of North and Northeast regions have anomalies in Daily Expenses for the months of recess.

5.3 Cluster analyzing

Here we answer the third and last question (**Q3**): *Can the expenses made during the four years of the legislature be clustered, thus identifying deputies who have similar behaviors?* To this end, we analyzed each one of the five expense categories during the entire legislature (Section 5.3.1) and also during each year of the legislature individually (Section 5.3.2). Specifically, our intent is to analyze both long- and short-term similarities between the expenses of each category.

5.3.1 Analysis of the entire legislature

For this analysis, we removed the deputies considered as outliers in Section 5.2 intending to avoid bias in the generation of the clusters. Therefore, each category of expenses was represented by 213 instances. Each instance consists of 48 records of expenses from the same deputy. As one of our goals is to verify whether the clusters are interpretable, the analysis of interpretability was based on geographic information (region) and on the political party of each deputy.

As described in Section 4.3, the clustering process was performed using the k-Means algorithm and the distance function DTW. The choice of the best value for the parameter k of the k-Means algorithm was made using the Elbow Method. Table 4 reports the best values for k obtained for each category of expenses. The summaries of the clusters identified are shown in Table 5: due to the large number of political parties and the small representation of some acronyms, the political parties that are not among the ten parties with the highest average spending were identified as ”Others” and only the three labels with the highest incidence for each analyzed point in each cluster are presented.

To interpret the identified clusters, one can notice, for example, the impact of the geographic region on the clusters of Daily Expenses. Let us detail this particular case. Figure 8 illustrates the choice of the best k and it also reports the number of instances in each of the clusters of that specific expense category. In Cluster 1, nearly 78% of the deputies belong to the northernmost regions of the country (Northeast has

Table 3. Deputies who have outlying expenditure entries in months of parliamentary recess.

Parliamentary Activity	Transportation Tickets	Consulting and Security Services	Daily Expenses	Materials and General Services
Arnaldo Faria De Sá (PTB/SP)	Luis Carlos Heinze (PP/RS)	Silvio Costa (PTB/PE)	Átila Lins (PMDB/AM)	Henrique Fontana (PT/RS)
Carlos Manato (PDT/ES)	Darcísio Perondi (PMDB/RS)	Josué Bengtson (PTB/PA)	Paes Landim (PTB/PI)	Pedro Chaves (PMDB/GO)
Silas Câmara* (PSC/AM)	Heráclito Fortes (DEM/PI)	Takayama (PSC/PR)	Silas Câmara* (PSC/AM)	Eduardo Da Fonte (PP/PPE)
Givaldo Carimbão* (PSB/AL)	Lúcio Vale (PR/PA)	Pedro Fernandes* (PTB/MA)	Pedro Fernandes* (PTB/MA)	Waldenor Pereira (PT/BA)
Arthur Lira (PP/AL)	Bohn Gass (PT/RS)	Antonio Bulhões (PRB/SP)	Átila Lira (PSB/PI)	Eros Biondini (PTB/MG)
Professora Dorinha (DEM/TO)	Jerônimo Goergen (PP/RS)	Domingos Neto (PROS/CE)	Givaldo Carimbão* (PSB/AL)	-
-	Arnaldo Jordy (PPS/PA)	Flávia Morais (PDT/GO)	-	-

*Deputies appearing as *outliers* in more than one category of expenses.

Table 4. Entire legislature (2015-2018): Best values for the parameter k of the k-Means clustering algorithm.

Category	Best k
Transportation Tickets	6
Materials and General Services	5
Daily Expenses	2
Parliamentary Activity	3
Consulting and Security Services	3

Table 5. Clusters formed considering the entire legislature.

Expense	Cluster	Geographical region (%)	Political party (%)
Transportation Tickets	*Cluster 1*	Midwest (55.56), Southeast (22.22), Northeast (11.11)	Others (77.78), PR (11.11), PDT (11.11)
	Cluster 2	Southeast (40.0), South (25.0), North (17.5)	Others (82.5), PDT (7.5), PSB (5.0)
	Cluster 3	Northeast (40.0), Southeast (34.29), South (14.29)	Others (65.71), PR (14.29), PSB (5.71)
	Cluster 4	Southeast(39.58), Northeast (27.08), South (22.92)	Others (68.75), PSB (8.33), PR (6.25)
	Cluster 5	Southeast (49.35), Northeast (23.38), South (16.88)	Others (72.73), PR (7.79), PSB (5.19)
	Cluster 6	Southeast (25.0), South (25.0), Northeast (25.0)	Others (75.0), PHS (25.0)
Parliamentary Activity	*Cluster 1*	Southeast (38.78), Northeast (26.53), North (14.29)	Others (67.35), PR (10.2), PRB (4.08)
	Cluster 2	Southeast (43.8), South (23.97), Northeast (23.14)	Others (76.86), PR (6.61), PSB (6.61)
	Cluster 3	Southeast (37.21), Northeast (25.58), South (16.28)	Others (67.44), PR (6.98), PSB (6.98)
Consulting and Security Services	*Cluster 1*	Southeast (48.0), Northeast (20.0), South (18.4)	Others (73.6), PSB (7.2), PR (4.8)
	Cluster 2	Southeast (39.39), Northeast (30.3), South (18.18)	Others (69.7), PR (9.09), PRB (6.06)
	Cluster 3	Northeast (30.91), Southeast (27.27), South (20.0)	Others (72.73), PR (12.73), PSB (3.64)
Daily Expenses	*Cluster 1*	Northeast (43.06), North (20.83), Southeast (15.28)	Others (69.44), PR (12.5), PDT (5.56)
	Cluster 2	Southeast (54.61), South (24.82), Northeast(14.89)	Others (74.47), PSB (7.09), PRB (4.96)
Materials and General Services	*Cluster 1*	Southeast (39.13), South (21.74), Northeast (17.38)	Others (69.57), PR (6.52), PTB (6.52)
	Cluster 2	Southeast (41.38), South (24.14), Northeast (17.24)	Others (75.86), PR (10.34), PSB (6.9)
	Cluster 3	Southeast (36.36), North (27.27), Northeast (27.27)	Others (72.73), PDT (18.18), PRB (9.09)
	Cluster 4	Southeast (42.48), Northeast (28.32), South (16.81)	Others (72.57), PR (8.85), PSB (7.08)
	Cluster 5	Southeast (42.86), Northeast (28.57), South (28.57)	Others (78.57), PDT (14.29), PSB (7.14)

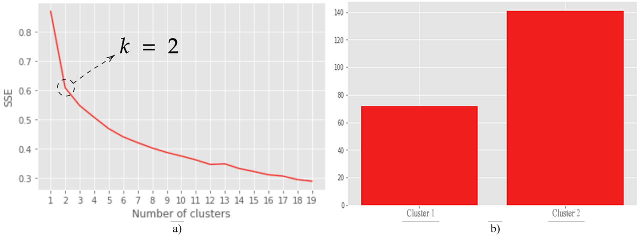

Figure 8. a) Choice of best k; b) Choice of best k and cluster's sizes with reference to Daily Expenses.

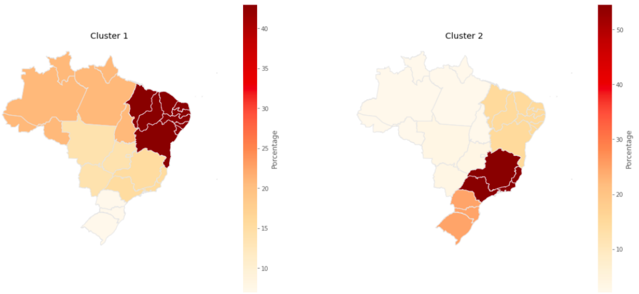

Figure 9. Geographical representation of the clusters of deputies according to their Daily Expenses

43.06%, North has 20.83% and Midwest has 13.89%): for instance, Bahia has 14.47% and Ceará has 7.89%, whereas São Paulo has only 2.63% of the deputies. On the other hand, in Cluster 2, nearly 80% of the deputies belong to the southernmost regions (Southeast has 54.61% and South has 24.82%): for instance, São Paulo has 22.63% and Minas Gerais has 20.44%, whereas Ceará has only 0.73% of the deputies. Figure 9 illustrates the geographical distribution of each cluster. These results lead us to conclude that there exist considerably distinct patterns in the Daily Expenses of deputies according to their regions. The reasons behind such a variation in Daily Expenses' requirements from deputies of distinct regions are not clear; note that Clusters 1 and 2 clearly represent distinct regions, with the former regarding the northern portion of the country, and the latter the southern portion.

5.3.2 Analysis over the years

Using the same methodology of the previous section, all deputies considered outliers in Section 5.2 were discarded from this analysis. To perform an analysis based on years of the legislature, each of the 213 time series instances of each expense category was segmented according to the years; leading to $213 \times 4 = 852$ time series instances, each one consisting of 12 records of expenses made by the same deputy during a single year. All the 852 instances were analyzed together. Our intention is to allow the identification of short-term similarities involving expenses from distinct years. Thus, in addition to verifying the interpretability of the clusters by the geographic region and political party, we also verify the impact of the years in the cluster formation.

As before, we used the k-Means clustering algorithm with the DTW distance function here. Table 6 reports the best values for parameter k, identified with the Elbow Method. The summaries of the clusters identified are shown in Table 7. Only the three labels with the highest incidence for each analyzed point in each cluster are presented and, like in Table 5, political parties with low representativity are identified as ”Others”.

When interpreting the clusters identified, one can notice, for example, the impact of the years on the clusters of Transportation Tickets. Let us detail this particular case. Figure 10

Table 6. Analysis over the years, from 2015 to 2018: Best values for the parameter k of the k-Means clustering algorithm.

Category	Best k
Transportation Tickets	2
Materials and General Services	3
Daily Expenses	5
Parliamentary Activity	6
Consulting and Security Services	4

Table 7. Clusters formed considering each year of the legislature.

Expense	Cluster	Geographical region (%)	Political party (%)	Year (%)
Transportation Tickets	*Cluster 1*	Southeast (42.16), Northeast (25.17), South (18.54)	Others (74.95), PDT (6.35), PSB (6.35)	2015 (33.55), 2016 (30.46), 2017 (21.63)
	Cluster 2	Southeast (40.35), Northeast (23.56), South (19.05)	Others (71.61), PR (8.86), PDT (5.24)	2018 (37.09), 2017 (28.82), 2016 (18.8)
Parliamentary Activity	*Cluster 1*	Southeast (45.99), Northeast (23.36), South (14.6)	Others (73.72), PR (8.03), PSB (5.84)	2018 (28.47), 2017 (26.28), 2015 (25.55)
	Cluster 2	Southeast (27.69), Northeast (24.62), North (21.54)	Others (50.77), PRB (15.38), PSB (9.23)	2016 (32.31), 2015 (24.62), 2018 (23.08)
	Cluster 3	Southeast (46.3), Northeast (24.07), South (11.11)	Others (71.3), PSB (10.19), PR (7.41)	2017 (30.56), 2016 (26.85), 2018 (22.22)
	Cluster 4	Northeast (39.02), Southeast (39.02), North (9.76)	Others (78.05), PR (9.76), PSC (7.32)	2015 (39.02), 2018 (36.59), 2016 (12.2)
	Cluster 5	Northeast (33.96), Southeast (32.08), Midwest (15.09)	Others (75.47), PSC (5.66), PR (3.77)	2015 (32.08), 2018 (28.3), 2016 (20.75)
	Cluster 6	Southeast (41.86), South (24.33), Northeast (22.32)	Others (75.22), PR (7.59), PSC (4.69)	2016 (26.77), 2017 (25.89), 2015 (23.88)
Consulting and Secutiry Services	*Cluster 1*	Southeast (43.85), Northeast (21.84), South (19.58)	Others (74.18), PR (7.28), PSB (5.89)	2017 (25.48), 2015 (25.13), 2018 (25.13)
	Cluster 2	Southeast (29.17), Northeast (27.08), South (27.08)	Others (79.17), PR (9.38), AVANTE (3.12)	2015 (27.08), 2016 (27.08), 2017 (20.83)
	Cluster 3	Southeast (53.49), Northeast (18.6), Midwest (16.28)	Others (46.51), PRB (13.95), PSC (11.63)	2015 (34.88), 2016 (23.26), 2017 (20.93)
	Cluster 4	Southeast (35.29), Northeast (35.29), South (12.5)	Others (70.59), PR (6.62), PSB (6.62)	2016 (27.21), 2017 (27.21), 2018 (25.74)
Daily Expenses	*Cluster 1*	Northeast (51.61), Southeast (14.52), Midwest (14.52)	Others (75.81), PR (9.68), PTB (4.84)	2017 (29.03),2015 (27.42), 2018 (25.81)
	Cluster 2	Southeast (55.35), South (27.68), Northeast (12.53)	Others (74.15), PSB (9.4), PDT (4.4)	2015 (25.59), 2016 (25.33), 2017 (25.07)
	Cluster 3	Northeast (48.65), North (27.03), South (13.51)	Others (62.16), PR (13.51), PRB (10.81)	2015 (29.73), 2016 (24.32), 2017 (25.07)
	Cluster 4	Northeast (34.21), Southeast (27.19), North (21.93)	Others (71.93), PR (14.04), PDT (5.26)	2016 (27.19), 2018 (26.32), 2017 (24.56)
	Cluster 5	Southeast (39.06), Northeast (27.73), South (14.06)	Others (71.88), PR (8.98), PDT (4.69)	2018 (26.17), 2016 (25.39), 2015 (24.22)
Acquisition and General Services	*Cluster 1*	Southeast (42.13), Northeast (25.62), South (18.06)	Others (72.38), PR (7.41), PSB (6.33)	2015 (27.16), 2016 (24.85), 2018 (24.69)
	Cluster 2	Southeast (38.29), South (21.14), Northeast (20.57)	Others (74.29), PR (7.43), PDT (7.43)	2017 (28.57), 2016 (28.57), 2018 (26.26)
	Cluster 3	Southeast (41.38), South (20.69), Northeast (20.69)	Others (74.41), PR (10.34), PSB (6.33)	2017 (41.38), 2015 (27.59), 2018 (24.14)

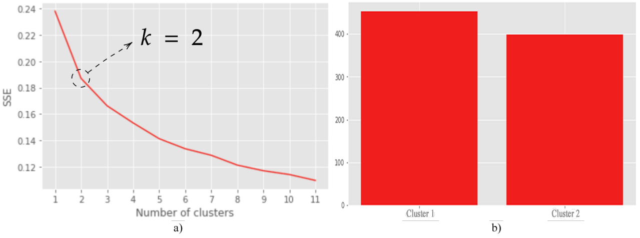

Figure 10. a) Choice of best k; b) Number of instances in each cluster.

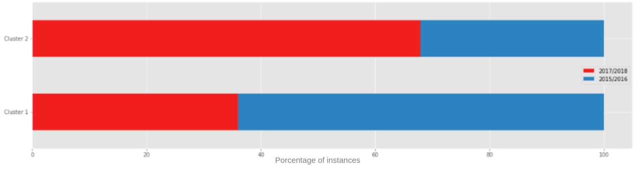

Figure 11. Impact of the years on the clusters of deputies with reference to their expenses with Transportation Tickets.

illustrates the choice of the best k and it also reports the number of instances in each of the clusters of that specific expense category.

By analyzing the two clusters, it is possible to verify the impact of the 2015/2016 and 2017/2018 periods: one of the clusters is formed mostly by the 2017/2018 period while the other is formed mostly by the 2015/2016 period. This behavior is illustrated in Figure 11. These results lead us to conclude that there exist considerably distinct patterns in the expenses with Transportation Tickets of deputies through the years, which may be explained as an impact of inflation in the corresponding periods.

6 Conclusion

This paper presented an analysis of the public expenditures of federal deputies in Brazil considering the period from 2015 to 2018. For this purpose, an exploratory data analysis was performed. We also presented analyses based on outlier detection and clustering to find anomalies and other patterns hidden in the expenses.

The **exploratory data analysis** revealed peaks of expenses at the end of each year, which may make one wonder whether these expenses were made to enforce using the full annual budget, or if they were truly necessary. We could observe that there are some expense patterns (Figure 3), including some expenses reaching high levels during election period (October/2018). Besides, we verified a trend of growth in the expenses with Transportation Tickets, which may be explained by the inflation in the period. Additionally, the exploratory analysis also revealed that roughly 53% of the deputies did not make all their expenditure entries public.

The analysis of **outlier detection** for the entire period of the legislature revealed that expenses on Materials and General Services from the South and Southeast regions of Brazil correspond to roughly 66% of the outliers identified. Also, when considering the parties, PMDB, PTB and PP represented 55% of the outliers found. Analyses for the recess period, showed that some deputies are considered as outliers regarding more than one category of expenses (section 5.2.2). Here, it is worth noting that an "outlying expense" may be either much larger or much smaller than what is considered to be usual when looking at the other expenses.

The results obtained from the **analysis of clustering** have shown that distinct patterns of expenses exist based on the Daily Expenses and the Transportation Tickets. Considering the Daily Expenses for the entire period of the legislature (2015-2018), two clusters were identified and they can be characterized by region: 1. Most of the deputies in Cluster 1 belong to regions further north in Brazil. 2. Most of the deputies in Cluster 2 belong to southernmost regions of Brazil. The reasons behind such a variation could not be understood by us, tough. Considering Transportation Tickets, two clusters were identified and they can be characterized by year: 1. Cluster 1 corresponds mainly to the first two years of the legislature (2015 and 2016). 2. Cluster 2 corresponds mainly to the last two years (2017 and 2018). These two clusters may be explained by the prices of tickets that have increased considerably over the years.

It is important to highlight that parliamentary quotas have no direct relationship with social and political aspects, or with society's public policies. These are amounts used to keep parliamentary activities running. However, the continuity and consistency of expenses information made by deputies is vitally important, ensuring transparency and publicity of information to taxpayers. The absence or inconsistency of information may be considered violations of the principle of publicity. The lack of constant updating of expense information or the publication of incomplete or partial data may influence the conclusions obtained, in addition to compromising the transparency of public spending.

In conclusion, this work demonstrated that data mining and machine learning models are useful in auditing public expenses, focusing on the identification of anomalous expenditures, as well as finding similarities and dissimilarities in expenditures between parties and regions of Brazil. Thus, given the potential that data science has to find patterns and extract useful information from the data, as future work, we expect to perform further analyses in the context of Federal Deputies' expenses in Brazil. A deeper and more detailed investigation would be beneficial to broaden the understanding of deputies' behavior, allowing comparing federal and state expenses, finding similarities among the suppliers related to anomalous expenses, or investigating if the expenses reported match the market price of similar items.

References

Angiulli, F. and Pizzuti, C. (2002). Fast outlier detection in high dimensional spaces. In Elomaa, T., Mannila, H., and Toivonen, H., editors, *Principles of Data Mining and Knowledge Discovery*, pages 15–27, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: Ordering points to identify the clustering structure. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, SIGMOD '99, page 49–60, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/304182.304187.

Bellman, R. and Kalaba, R. (1959). On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9. DOI: 10.1109/TAC.1959.1104847.

Bholowalia, P. and Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).

Brasil (1988). Constituição da república federativa do brasil. *Diário Oficial da República Federativa do Brasil.*

Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 93–104, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/342009.335388.

Bueno, W. F. (2017). *Analysis of Brazilian deputies expenses claims from 2013 to 2016.* PhD thesis, Dublin, National College of Ireland.

da. Silva, E. F., Fragoso, G. A., Rodrigues, L. R., Freitas, N. C. A., and Vinuto, T. (2022). Exposer: Framework para detecção de anomalias em licitações públicas. *Anais Estendidos do XXXVII Simpósio Brasileiro de Banco de Dados (SBBD Estendido 2022).*

Dang, T. T., Ngan, H. Y., and Liu, W. (2015). Distance-based k-nearest neighbors outlier detection method in large-scale traffic data. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 507–510. DOI: 10.1109/ICDSP.2015.7251924.

Esen, E. and Celik Kecili, M. (2022). Economic growth and health expenditure analysis for turkey: evidence from time series. *Journal of the knowledge economy*, 13(3):1786–1800.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD96, page 226–231. AAAI Press. DOI: 10.5555/3001460.3001507.

Han, J., Pei, J., and Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

Heiler, J. G., Viana, J. P. S. L., and Santos, R. D. d. (2016). O custo da política subnacional: a forma como o dinheiro é gasto importa? relação entre receita, despesas e sucesso eleitoral. *Opinião Pública*, 22:56–92.

Khol, B. (1992). Brazil: Summary of inflation indicators.

Kriegel, H.-P., Schubert, M., and Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 444–452. DOI: 10.1145/1401890.1401946.

Krishna, K. and Murty, M. N. (1999). Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3):433–439.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE.

Mueen, A. and Keogh, E. (2016). Extracting optimal performance from dynamic time warping. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2129–2130.

Papadimitriou, S., Kitagawa, H., Gibbons, P., and Faloutsos, C. (2003). Loci: fast outlier detection using the local correlation integral. In *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405)*, pages 315–326. DOI: 10.1109/ICDE.2003.1260802.

Paranhos, R., Júnior, J. A. d. S., Filho, D. B. F., and Rocha, E. C. d. (2015). 513 deputados e um segredo: Cota para exercício de atividade parlamentar - ceará e paraíba. *Revista Cadernos de Estudos Sociais e Políticos*, 4(8).

Ravisankar, P., Ravi, V., Raghava Rao, G., and Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50(2):491–500. DOI: https://doi.org/10.1016/j.dss.2010.11.006.

Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171. DOI: 10.1109/ICDCSW.2011.20.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., and Woods, E. (2020). Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6.

Thiago Alencar Gomes, Rommel N. Carvalho, R. S. C. (2017). Identifying anomalies in parliamentary expenditures of brazilian chamber of deputies with deep autoencoders. *16th IEEE International Conference on Machine Learning and Applications*.

Zhao, Y., Nasrullah, Z., and Li, Z. (2019). Pyod: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research*, 20(96):1–7.