Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.7072

Research Paper

Detecting and Analyzing Anomalies in City Expenditure Time Series

Marco Túlio Dutra [Federal University of Ouro Preto | *marco.dutra@aluno.ufop.edu.br*]
Lucas G. L. Costa [Federal University of Minas Gerais | *lucas-lage@ufmg.br*]
Gabriel P. Oliveira [Federal University of Minas Gerais | *gabrielpoliveira@dcc.ufmg.br*]
Mariana O. Silva [Federal University of Minas Gerais | *mariana.santos@dcc.ufmg.br*]
Gisele L. Pappa [Federal University of Minas Gerais | *glpappa@dcc.ufmg.br*]

Department of Computer Science, Federal University of Minas Gerais, Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG, 31270010, Brasil.

Abstract. This paper presents an enhanced approach for detecting and analyzing anomalies in public expenditure time series. The method combines statistical techniques and machine learning models to identify unusual spending behaviors and to rank expenditure items according to both the frequency of anomalies and the monetary impact involved. The statistically generated anomalies produce suspicion alerts of potential fraud, with the objective of prioritizing cases and directing human audit efforts more effectively. In this extended version, we introduce a finer-grained analysis that examines spending patterns at a more detailed level within the governmental budget structure, allowing the detection of irregularities that may not be visible under more aggregated views. The approach is validated on a real-world dataset comprising more than one million city expenditure records from the state of Minas Gerais, and the results demonstrate its ability to reveal irregularities that may remain hidden under higher levels of aggregation.

Keywords: Outlier detection, Time series, Ranking Approach, E-government

Edited by: Edna Dias Canedo | **Received:** 01 December 2025 • **Accepted:** 30 April 2026 • **Published:** 21 May 2026

1 Introduction

Transparent management of public resources is essential to ensure the efficiency and integrity of government administration. In Brazil, Federal Law No. 12.527/11,[1] known as the "Access to Information Law", establishes that data on public expenditures must be accessible to citizens, enabling the monitoring of resource allocation. However, the large volume and complexity of these records pose significant challenges for identifying irregularities, such as unjustified expenditures and overpriced contracts [Silva *et al.*, 2023].

Analyzing the temporal evolution of public expenditures is fundamental for detecting deviations that may indicate irregularities, with time series analysis playing a central role in this process. In this context, anomaly detection techniques have emerged as a promising approach for identifying unusual patterns [Darban *et al.*, 2024]. In the domain of public expenditures, such anomalies may correspond to administrative inconsistencies or potential fraud. Nevertheless, this analysis remains challenging due to the high data volume and the heterogeneity of spending patterns across cities and budget functions. Traditional audit methods, which rely primarily on manual inspection, are insufficient to handle this scale, motivating the adoption of automated approaches for detecting potential anomalies [Gomide *et al.*, 2023].

In this context, we propose a unified approach for detecting and analyzing anomalies in municipal public expenditure time series. The proposed framework integrates statistical, machine learning, and probabilistic techniques to identify anomalous patterns and rank expenditures based on both their frequency of occurrence and monetary impact. The detected anomalies are interpreted as statistical alerts of potential irregularities, supporting the prioritization of cases and guiding human audit efforts more effectively.

This work extends our previous work published at the 13th Latin American Symposium on Digital Government [Dutra *et al.*, 2025]. The original approach focused on anomaly detection at the expenditure function level of the Brazilian budget classification system. In this extended version, we advance the analysis to a finer level of granularity by applying the same methodology to expenditure subfunctions.

In the Brazilian public budget framework, an expenditure function represents a broad policy domain to which public resources are allocated, while a subfunction provides a more detailed specification within that domain [Brazil, 2025]. These classifications act as hierarchical aggregators of expenditure types, enabling both comparative and longitudinal analyses of municipal spending. Table 1 presents examples of expenditure functions and their corresponding subfunctions.

This finer-grained perspective enables a more precise analysis of spending dynamics and increases the ability to identify irregularities that may be obscured at higher levels of aggregation. While function-level analysis may dilute anomalies across broad categories, subfunction-level analysis allows the identification of localized irregularities within specific administrative processes. In light of these aspects, our main contributions are:

- We formally define a data-driven methodology for detecting statistical anomalies in city expenditure time series structured according to the Brazilian public budget classification system.
- We extend prior work by introducing a finer-grained

[1] Law No. 12,527/11: `https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm`. Accessed on 18 May 2026.

analysis at the subfunction level, enabling the identification of irregularities that may remain hidden at higher aggregation levels.
- We propose a ranking mechanism designed to support audit prioritization by combining anomaly frequency and financial impact.

The remainder of this work is structured as follows. Section 2 reviews related work on anomaly detection and public expenditure auditing. Section 3 describes the proposed methodology and the anomaly detection and ranking processes. Section 4 presents the experimental results using expenditure data from cities in the State of Minas Gerais. Section 5 introduces the extended analysis at the subfunction level. Section 6 presents a real-world case study illustrating the behavior of the proposed framework. Finally, Section 7 discusses the conclusions, limitations, and directions for future work.

2 Related Work

Combating fraud in the public sector is a global priority, as it is essential for ensuring the integrity and accountability of government institutions [Silva *et al.*, 2020]. A common challenge in this domain is the need to process large volumes of heterogeneous data, which has motivated the adoption of artificial intelligence and machine learning techniques for fraud detection and audit support [Mongwe *et al.*, 2020; Handoko *et al.*, 2022]. In the Brazilian context, recent studies have investigated fraud detection procedures and the impact of technological solutions on anti-fraud awareness in the public sector [Silva *et al.*, 2020].

In the global context, recent studies have explored anomaly detection in audit analytics using large-scale public expenditure data, demonstrating the effectiveness of combining multiple unsupervised methods in governmental scenarios [Li *et al.*, 2025]. Systematic reviews indicate that the adoption of fraud analytics varies significantly across countries, highlighting the need for approaches that generalize beyond specific institutional settings [Alfian *et al.*, 2023]. A central challenge in this domain is the scarcity of labeled data, which limits the applicability of supervised approaches [Schneider dos Santos *et al.*, 2025]. In addition, anomaly detection systems often generate a large number of alerts, requiring prioritization mechanisms to support practical auditing workflows and mitigate alert fatigue [No *et al.*, 2019]. More recently, advances in explainable artificial intelligence have emphasized the importance of interpretability in fraud detection systems, although challenges remain in evaluating the usefulness of explanations for decision-making [Zafar and Wu, 2026].

In the context of monitoring municipal expenditures, the literature presents several works focused on identifying irregularities in public procurement and government spending. For instance, Braz *et al.* [2023, 2024] proposed methodologies to analyze irregularities in public tenders, focusing on small businesses, including checks on revenue limits and detection of links between legal entities. Similarly, Oliveira *et al.* [2023b] introduced a set of 19 audit trails to identify different types of fraud and irregularities in procurement processes. Furthermore, Silva *et al.* [2023, 2024] proposed a framework for disambiguating bid items and statistical analyses to detect overpricing. Along the same line, Costa *et al.* [2024] developed a query system for public procurement items with a focus on overpricing detection.

Specifically in detecting anomalies in city expenditures, Gomide *et al.* [2023] proposed a heuristic based on data mining to identify cities whose expenditures are significantly above the average of cities with similar characteristics. This approach groups cities based on population ranges from the 2010 Census[2] and their geographic location, allowing the detection of suspicious public spending patterns. Additionally, Oliveira *et al.* [2023a] proposed methods to assess the quality of municipal public revenue and procurement data.

Regarding time series analysis, different approaches have been explored for anomaly detection. Mendes *et al.* [2023] applied the Soft-DTW k-means algorithm to cluster time series of revenue from companies whose partners made donations to political campaigns, identifying potential favoritism in tender processes. Blázquez-García *et al.* [2021] conducted a comprehensive review of anomaly detection in time series, focusing on unsupervised methods. In the context of time series forecasting, Oliveira *et al.* [2024] investigated the use of machine learning to predict hotel room prices, analyzing the efficiency of different methods.

More recently, Darban *et al.* [2024] presented a systematic review of advanced deep learning techniques applied to anomaly detection in time series. The study presents a taxonomy based on the detection strategies used, as well as the deep learning models employed. This review highlights the importance of combining different techniques to improve the identification of atypical patterns in complex time series.

Based on these studies, this work proposes a hybrid approach for detecting anomalies in time series of city expenditures, combining statistical methods, machine learning, and probabilistic techniques. The proposed methodology not only identifies anomalies but also prioritizes them using a ranking, allowing for more targeted auditing.

3 Methodology

In this section, we present the methodology adopted to evaluate the performance of different anomaly detection methods in time series. The objective is to identify the most anomalous cases in the temporal data by applying different anomaly detection techniques. Finally, the results from each method are aggregated and ranked to identify the most critical cases. The approach follows three main stages illustrated in Figure 1: data modeling (Section 3.1), anomaly detection in time series (Section 3.2), ranking of results (Section 3.3) and suitability analysis of the anomaly detection methods (Section 3.4).

3.1 Data Modeling

In the Brazilian public budget framework, a budget allocation refers to the amount authorized for specific public expenditures, whose existence is mandatory for any payment.[3] Budget allocations are structured according to four classification dimensions: institutional, functional, by nature of the expendi-

[2] 2010 Census - IBGE: `https://censo2010.ibge.gov.br/`. Accessed on 18 May 2026.

[3] `https://www12.senado.leg.br/manualdecomunicacao/guia-de-economia/dotacao-orcamentaria`. Accessed on 18 May 2026.

Table 1. Examples of expenditure functions and their three most common subfunctions.

Government Function	Most Common Subactions	Government Function	Most Common Subactions
Administration	Maintenance of Public Assets and Infrastructure Procedural Management of Legal Affairs Strategic Direction of Municipal Policy	Agriculture	Maintenance of the Agriculture and Livestock Coordination Office Maintenance of the Municipal Department of Agriculture and Livestock Maintenance of Agricultural Mechanization Activities
Culture	Administrative and Infrastructure Management of the Department Maintenance of Cultural Promotion Activities Infrastructure Management	Education	Administration of Elementary Education Personnel Management Maintenance of Basic Education Activities
Health	Human Resources Management Technical and Hospital Management Primary Health Care Services	Housing	Technical and Support Services Participatory Budget Housing Projects Other Interventions in Informal Settlements
Judiciary	Administrative Management Payroll Management Court Costs and Judicial Sentences	Urban Development	Strategic Urban Policy Management Execution of Urban Planning and Policy Management Administrative Management and Maintenance of the Department of Public Works and Urban Development

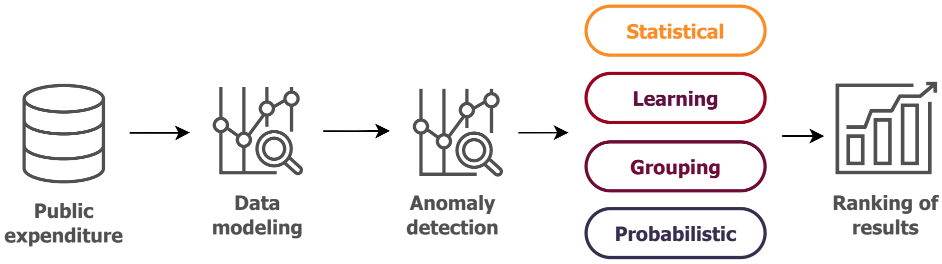

Figure 1. Overview of anomaly detection in expenditure time series.

ture, and by source of resources. In this work, we focus on the functional classification, in which expenditures are grouped according to their associated policy domain. Each function represents the institutional mission of a government body, such as education, health, culture, transport, and citizens' rights.

Thus, the proposed modeling is grounded in set theory and relies on four fundamental sets representing the key entities of interest: Public Expenditure (D), Month (M), City (C), and Function (F).

- $D = \{d \mid d \in \textit{Public Expenditure}\}$: represents all public expenditures;
- $M = \{m \mid m \in Months\}$: set of all months;
- $C = \{c \mid c \in \textit{City}\}$: set of all analyzed cities;
- $F = \{f \mid f \in \textit{Function}\}$: set of all functions of the budget allocation.

Furthermore, each public expenditure d has the following attributes:

- $d[m] \in M$: Month of the public expenditure;
- $d[c] \in C$: City of the public expenditure;
- $d[f] \in F$: Function of the public expenditure;
- $d[v] \in \mathbb{R}^+$: Value of the public expenditure.

In this work, a time series of public expenditures corresponds to a sequence, ordered over time, of the values of these expenditures. Since each expenditure is associated with a specific city and function, there is one time series for each city-function pair, defined as $S_{c,f} = ord_m\left(\{d \in D \mid d[c] = c \wedge d[f] = f\}\right)$, where the function ord_m returns the sequence of expenditures ordered increasingly by the month attribute. Thus, the complete set of time series can be formally defined as $S = \{S_{c,f} \mid c \in C,\ f \in F\}$.

3.2 Anomaly Detection

Time series anomaly detection methods aim to identify patterns or atypical values that deviate from the expected behavior [Darban *et al.*, 2024]. Each anomaly detection method (md_x) is associated with a value function ($fv_x(d)$) and a threshold function ($fl_x(D, \dots)$). The value function calculates a specific metric for each item d in the time series, and the threshold function defines a maximum or minimum limit to determine if an item is anomalous. The value function is always applied individually to each item d, while the threshold function can take into account a broader context, such as the entire series or multiple time series.

An expenditure d is considered an anomaly for a method md_x if the result of its value function exceeds the result of the threshold, i.e., $d \in A_x \iff fv_x(d) > fl_x(D, \dots)$, where A_x is the set of anomalies identified by md_x.

For the proposed approach, we use nine anomaly detection methods, categorized into four main groups: *Statistical*, *Learning*, *Clustering*, and *Probabilistic*. Table 2 presents the nine anomaly detection methods considered in this work, along with the formal definitions of the value function and the threshold function for each one. Then, each method is described in detail, highlighting its main characteristics and advantages, as well as its value and threshold functions.

Z-Score (ZSC). Z-Score is a statistical method that calculates the distance between a value and the mean in terms of standard deviations. This method is very simple, sensitive to outliers, and not effective for non-normal data. In this methodology, values with a Z-Score greater than 3 are considered anomalies. Therefore, the value function ($fv_1(d)$) is the expenditure value itself ($d[v]$). And there is a threshold for each function - month pair (f, m), thus, the threshold function ($fl_1(D, f, m)$) is calculated considering a subset of all expenditures of the same function and month ($D^{f,m} \subset D$). Therefore, the threshold function is the mean of the values of the elements in the subset ($\mu(D^{f,m}[v])$), plus 3 times the standard deviation of the values of the elements in the subset ($\sigma(D^{f,m}[v])$). In other words, the data is grouped by function and month, and records with a Z-Score greater than 3 within each group are considered anomalies.

Interquartile Range (IQR). The IQR is defined as the difference between the third quartile (Q3) and the first quartile (Q1). This method uses this difference to detect anomalies. It is more robust than Z-Score because it is not affected by outliers. As with the Z-Score method, the value function ($fv_2(d)$) is the expenditure value itself, i.e., the element of the time series ($d[v]$). And there is a threshold for each function - month pair (f, m). Thus, the threshold function ($fl_2(D, f, m)$) is calculated considering a subset of all expenditures of the same function and month ($D^{f,m} \subset D$). Therefore, the threshold function is the third quartile of the values of the elements in the subset ($q3\left(D^{f,m}[v]\right)$), plus 1.5 times the IQR of the values of the elements in the subset ($iqr\left(D^{f,m}[v]\right)$).

Seasonal Decomposition Using Moving Averages (SDMA) [Hyndman and Athanasopoulos, 2021, Chapter 6]. This method uses seasonal decomposition on the data series to decompose the series into trend, seasonality, and residuals, and then applies moving averages to the residuals to detect anomalies. It is efficient for data with strong trends and seasonality. In this methodology, only records classified as residuals are considered in anomaly detection. Therefore, the value function ($fv_3(d)$) is the function $rd(d[v])$ which returns the value of $d[v]$ itself if it is considered a residual, or returns zero otherwise. And there is a threshold per time series, i.e., for each city - function pair (c, f). Thus, the threshold function ($fl_3(D, c, f)$) is calculated considering a subset of all expenditures of the same city and function ($D^{c,f} \subset D$). Therefore, the threshold function is the mean of the residual values of the elements in the subset ($\mu\left(rd\left(D^{c,f}[v]\right)\right)$) plus 3 times the standard deviation of the residual values of the elements in the subset ($\sigma\left(rd\left(D^{c,f}[v]\right)\right)$).

Level Shift Anomaly Detection (LSAD) [Hethu Avinash *et al.*, 2024]. This method detects anomalies using a technique to detect abrupt changes in a time series. These changes occur when there is a sudden and significant change in the average value of the time series. The time series is divided into sliding windows, which are consecutive and overlapping time intervals. Each window (w) has a fixed size, defined by the user. The anomaly calculation is done by aggregating the values within each window. Thus, the value function ($fv_4(d)$) is the expenditure value itself ($d[v]$), and there is a threshold for each window of each time series, i.e., a threshold for each window - city - function trio (w, c, f). Thus, the threshold function ($fl_4(D, w, c, f)$) is calculated considering a subset of all expenditures of the same window, city, and function ($D^{w,c,f} \subset D$). Therefore, the threshold function is the third quartile of the values of the elements in the subset ($q3\left(D^{w,c,f}[v]\right)$) plus 1.5 times the IQR of the values of the elements in this subset ($iqr\left(D^{w,c,f}[v]\right)$).

Isolation Forest (ISF) [Liu *et al.*, 2008]. This method uses isolation trees to detect anomalies. Each tree is trained to isolate a data instance, and instances that are isolated more frequently are considered anomalies. The number of splits required to isolate an anomaly determines the path length (cc) to the leaf node of the instance. The anomaly calculation is done by comparing the expected path length for an instance (cce), with the average path length across all isolation trees ($\mu(cc)$). Thus, the value function ($fv_5(d)$) is the expected path length for the expenditure ($cce(d[v])$). And there is a threshold for each time series, i.e., a threshold for each city - function pair (c, f). Therefore, the threshold function ($fl_5(D, c, f)$) is calculated considering a subset of all expenditures of the same city and function ($D^{c,f} \subset D$). Thus, the threshold function is the average path length across all isolation trees ($\mu\left(cc\left(D^{c,f}[v]\right)\right)$).

Long Short-Term Memory (LSTM) [Malhotra *et al.*, 2015]. LSTM is a type of recurrent neural network capable of learning complex patterns in time series. The LSTM is trained to predict the next value in a time series, and the prediction error (ep) is used in the anomaly detection calculation. The idea is that the model learns the pattern and tries to predict the value following this pattern. If the actual value is too far from the predicted value, it may indicate an anomaly. Therefore, the value function ($fv_6(d)$) is the prediction error of the expenditure value ($ep(d[v])$). And there is a threshold for each time series, i.e., a threshold for each city - function pair (c, f). Thus, the threshold function ($fl_6(D, c, f)$) is calculated considering a subset of all expenditures of the same city and function ($D^{c,f} \subset D$). Therefore, the threshold function is the mean of the prediction errors of the values of the elements in the subset ($\mu\left(ep\left(D^{c,f}[v]\right)\right)$) plus 3 times the standard deviation of the prediction errors of the values of the elements in the subset ($\sigma\left(ep\left(D^{c,f}[v]\right)\right)$).

Local Outlier Factor (LOF) [Breunig *et al.*, 2000]. LOF is a method that calculates the local density of each data point and compares it with the local density of its neighbors. The anomaly calculation is done by calculating the local density (lrd) of each data point, and those with a lrd lower than their neighbors are considered anomalies. Therefore, the value function ($fv_7(d)$) is the opposite of the local density of the expenditure value ($-lrd(d[v])$). And there is a threshold for each expenditure of each time series, i.e., a threshold for each expenditure - city - function trio (d, c, f). Therefore, the threshold function ($fl_7(D, d, c, f)$) is calculated considering a subset of all expenditures that are neighbors of another expenditure and are from the same city and function ($D^{d,c,f} \subset D$). Thus, the threshold function is the opposite of the average local density of all expenditures neighboring d in the same time series ($-\mu\left(lrd\left(D^{d,c,f}[v]\right)\right)$). For both the value function and the threshold function, the opposite of the

Table 2. Anomaly detection methods, along with the formal definitions of their value and threshold functions.

x	Method md_x	Value Function $fv_x(d)$	Threshold Function $fl_x(D, \dots)$
		Statistical	
1	Z-Score (ZSC)	$fv_1(d) = d[v]$	$fl_1(D, f, m) = \mu\left(D^{f,m}[v]\right) + 3\,\sigma\left(D^{f,m}[v]\right)$
2	Interquartile Range (IQR)	$fv_2(d) = d[v]$	$fl_2(D, f, m) = q3\left(D^{f,m}[v]\right) + 1.5\,iqr\left(D^{f,m}[v]\right)$
3	Seasonal Decomposition Using Moving Averages (SDMA)	$fv_3(d) = rd\,(d[v])$	$fl_3(D, c, f) = \mu\left(rd\left(D^{c,f}[v]\right)\right) + 3\,\sigma\left(rd\left(D^{c,f}[v]\right)\right)$
4	Level Shift Anomaly Detection (LSAD)	$fv_4(d) = d[v]$	$fl_4(D, w, c, f) = q3\left(D^{w,c,f}[v]\right) + 1.5\,iqr\left(D^{w,c,f}[v]\right)$
		Learning	
5	Isolation Forests (ISF)	$fv_5(d) = cce\,(d[v])$	$fl_5(D, c, f) = \mu\left(cc\left(D^{c,f}[v]\right)\right)$
6	Long Short-Term Memory (LSTM)	$fv_6(d) = ep\,(d[v])$	$fl_6(D, c, f) = \mu\left(ep\left(D^{c,f}[v]\right)\right) + 3\,\sigma\left(ep\left(D^{c,f}[v]\right)\right)$
7	Local Outlier Factor (LOF)	$fv_7(d) = -lrd\,(d[v])$	$fl_7(D, d, c, f) = -\mu\left(lrd\left(D^{d,c,f}[v]\right)\right)$
		Clustering	
8	Soft-DTW k-Means (SDKM)	$fv_8(d) = d[v]$	$fl_8(D, k, f, m) = \mu\left(D^{k,f,m}[v]\right) + 3\,\sigma\left(D^{k,f,m}[v]\right)$
		Probabilistic	
9	Autoregressive Integrated Moving Average (ARIMA)	$fv_9(d) = ep\,(d[v])$	$fl_9(D, c, f) = \mu\left(ep\left(D^{c,f}[v]\right)\right) + 3\,\sigma\left(ep\left(D^{c,f}[v]\right)\right)$

values is used because, in our approach, for an expenditure to be considered an anomaly, its value function must be greater than the threshold function. In LOF, for the expenditure to be considered an anomaly, its density must be lower than that of its neighbors. Thus, to invert the inequality sign, it is necessary to use the opposite values.

Soft-DTW k-Means (SDKM) [Cuturi *et al.*, 2017]. Soft-DTW k-Means is a clustering algorithm that combines the Dynamic Time Warping (DTW) technique with the k-Means algorithm. DTW is a technique used to compare time series sequences. k-Means is a clustering algorithm that divides a dataset into k groups based on the similarity between the data. In the context of this methodology, a clustering round is performed for each function. Thus, each series will be grouped into a group (k). The anomaly calculation is done by aggregating the values by group (k). So, the value function ($fv_8(d)$) is the expenditure value itself ($d[v]$), and there is a threshold for each group of each function and each month, i.e., a threshold for each group - function - month trio (k, f, m). Thus, the threshold function ($fl_8(D, k, f, m)$) is calculated considering a subset of all expenditures of the same group, function, and month ($D^{k,f,m} \subset D$). Therefore, the threshold function is the mean of the values of the elements in the subset ($\mu\left(D^{k,f,m}[v]\right)$) plus 3 times the standard deviation of the values of the elements in the subset ($\sigma\left(D^{k,f,m}[v]\right)$).

Autoregressive Integrated Moving Average (ARIMA) [Hyndman and Athanasopoulos, 2021, Chapter 8]. ARIMA is a statistical model that combines three components: autoregressive, integrated, and moving average. ARIMA is used to model the time series and predict future values, and from there, detect anomalies. To detect anomalies, we need to compare the actual value of the time series with the value predicted by ARIMA. If the prediction error (ep) between the two values is greater than a certain limit, we can consider the observed value as an anomaly. Therefore, the value function ($fv_9(d)$) is the prediction error of the expenditure value ($ep(d[v])$). And there is a threshold for each time series, i.e., a threshold for each city - function pair (c, f). Thus, the threshold function ($fl_9(D, c, f)$) is calculated considering a subset of all expenditures of the same city and function ($D^{c,f} \subset D$). Therefore, the threshold function is the mean of the prediction errors of the values of the elements in the subset ($\mu\left(ep\left(D^{c,f}[v]\right)\right)$) plus 3 times the standard deviation of the prediction errors of the values of the elements in the subset ($\sigma\left(ep\left(D^{c,f}[v]\right)\right)$).

3.3 Ranking of Results

This section describes the ranking strategy adopted to aggregate the results of each method, originally proposed by Oliveira *et al.* [2023b] and adapted to the context of municipal expenditures. For each expenditure, an alert indicator is generated combining the individual results of each method. This indicator, called I_d, is calculated for the expenditure d as the sum of the anomalies detected by each method x. Here, X represents the total number of anomaly detection methods employed. In the methodology proposed in Section 3.2, we use nine methods ($X = 9$). If new detection methods are incorporated into the methodology, the value of X will be increased accordingly, and this increment will be automatically reflected in the computation of the indicator.

This summation is then multiplied by the base-10 logarithm of the expenditure value. This value is included because auditing higher-value expenditures is considered more of a priority. Formally, this indicator is calculated by Equation 1.

$$I_d = log_{10}\left(d[v]\right) \cdot \sum_{x=1}^{X} \begin{cases} 1, & if\ d \in A_x; \\ 0, & otherwise \end{cases} \quad (1)$$

Where $d[v]$ represents the value of the expenditure and A_x is the set of expenditures identified as anomalous by method x. After calculating the alert indicator I_d for each expenditure d, the expenditures are ranked in descending order according to the value of I_d. The resulting set, $\hat{D}$, contains the expenditures ranked based on the alert indicator. Thus, we have that $\hat{D} \subseteq D$ and $\hat{D} = \{d \mid \forall d \implies I_d > I_{(d+1)}\}$.

3.4 Suitability Analysis of the Anomaly Detection Methods

The suitability of anomaly detection techniques must be evaluated in light of the specific characteristics of municipal public expenditure time series. These series typically exhibit non-negativity, pronounced seasonality driven by fiscal cycles, structural breaks resulting from policy changes, heterogeneous variance across cities, and occasional abrupt spikes associated with large public contracts. Given this structural complexity, no single method is sufficient to capture all relevant anomaly patterns. Statistical methods are effective at identifying extreme deviations, probabilistic models capture temporal dependencies, clustering approaches detect structural and shape-based discrepancies, and learning-based techniques model nonlinear dynamics.

Some of the methods presented in Table 2 rely on fixed thresholds (e.g., $1.5 \times IQR$ and 3σ). Such thresholds follow well-established practices in anomaly detection and are supported by prior work in the context of public expenditure analysis. In particular, the $1.5 \times IQR$ criterion has been used to detect potential overpricing in public procurement data [Silva *et al.*, 2024], while the 3σ rule has been applied to identify anomalous spending patterns [Gomide *et al.*, 2023]. These thresholds provide conservative criteria for identifying extreme deviations from typical behavior while preserving interpretability.

For these reasons, the proposed framework adopts a multi-method strategy that integrates the outputs of heterogeneous techniques into a unified alert indicator. This design improves robustness by reducing dependence on the assumptions of any single method and enables anomaly prioritization based on the consensus across different detection principles.

4 Experimental Analysis

To evaluate the proposed approach, an experiment was conducted using a dataset of city expenditures from the State of Minas Gerais, as detailed in Section 4.1. Next, in Section 4.3, the results of the proposed methods are characterized and analyzed. In Section 4.4, the experimental analysis is complemented by a qualitative examination of the characteristics of the expenditure ranking process based on the alert indicators. Finally, in Section 4.5, we present the weak-label validation using external audit evidence.

The study Gomide *et al.* [2023] conducted an anomaly detection experiment on city expenditures using the same database, therefore it was used as a *baseline* to evaluate the proposed approach. That study proposes two detection methods, which work by grouping cities through heuristics. Once grouped, a statistical calculation is performed to identify cities with expenditures considerably higher than the average for their respective population group. The first method (BL_{fx}) uses a heuristic that groups cities based on the population range from the 2010 Census. However, for this experiment, the ranges were updated according to the 2022 Census. The second method (BL_{geo}) uses a heuristic that groups based on the geographic location of the City.

4.1 Dataset

The data used in this article were extracted from the Computerized System of Municipal Accounts (translation of Sistema Informatizado de Contas dos Municípios - SICOM),[4] which provides an open database maintained by the Court of Accounts of the State of Minas Gerais (translation of Tribunal de Contas do Estado de Minas Gerais - TCE-MG).[5] [6]

This database contains detailed fiscal transparency records from 853 cities in Minas Gerais, including information on tenders, contracts, invoices, expenditures, and revenues. For this experiment, 1,045,189 records of municipal expenditures from January 2018 to July 2024 were considered. The expenditures were grouped by budget function, as defined by the Brazilian budget model, totaling 29 distinct expenditure functions.

Since there is one time series for each city–expenditure function pair, defined as $S_{c,f}$, the complete set of time series corresponds to the Cartesian product of the sets C (cities) and F (expenditure functions). Therefore, the total number of time series is $|C| \times |F| = 853 \times 29 = 24{,}737$.

Given that the analysis covers 79 months, each time series contains 79 elements, i.e., $|S_{c,f}| = 79$. Consequently, each city has 29 time series, one for each expenditure function, allowing for a detailed temporal analysis of public spending patterns. Table 3 summarizes these statistics according to the formalization presented in Section 3.1.

Table 3. Dataset summary statistics.

Set	Size
Public Expenditure	$\lvert D\rvert = 1{,}045{,}189$
Months	$\lvert M\rvert = 79$ (01-2018 to 07-2024)
City	$\lvert C\rvert = 853$
Function	$\lvert F\rvert = 29$
Time Series	$\lvert S\rvert = \lvert C\rvert \times \lvert F\rvert = 24{,}737$
Time Series Length	$\lvert S_{c,f}\rvert = 79$

4.2 Hyperparameter Configuration

All anomaly detection methods were implemented with fixed hyperparameter configurations to ensure reproducibility and comparability across municipalities and expenditure functions. The selected values reflect a balance between methodological robustness and computational feasibility.

[4]SICOM: `https://portalsicom1.tce.mg.gov.br/`. Accessed on 18 May 2026.

[5]The raw SICOM data is publicly available at `https://dadosabertos.tce.mg.gov.br/`. Accessed on 18 May 2026.

[6]The preprocessed dataset, including all constructed time series, is available at `https://github.com/LucasLage/mg_city_expenditure_time_series`. Accessed on 18 May 2026.

For density-based methods, the Local Outlier Factor (LOF) was configured with 10 neighbors and a contamination rate of 2%, using the default Euclidean distance metric. Similarly, the Isolation Forest model employed 20 trees with a contamination rate of 2%, while all other parameters were kept at their default values.

The Level Shift Anomaly Detection method was configured to detect only positive structural changes, using a window size of 5 observations and a sensitivity parameter $c = 1.5$.

For the ARIMA-based approach, a fixed specification $(p, d, q) = (5, 1, 5)$ was adopted, corresponding to five autoregressive terms, first-order differencing, and five moving-average terms.

The Soft-DTW k-Means clustering method used Soft-DTW as the distance metric, with smoothing parameter $\gamma = 0.01$. The maximum number of clustering iterations and barycenter updates was set to 50. The number of clusters k was automatically selected within the interval $[2, 10]$ by minimizing the marginal decrease in inertia, computed using a standard K-Means model with 10 initializations and a fixed random seed.

The LSTM autoencoder architecture consisted of two encoding and two decoding layers, with 64 external LSTM units and 32 internal bottleneck units. ReLU activation was used in all recurrent layers. The model was trained using the Adam optimizer and mean squared error (MSE) as the loss function.

Finally, the Seasonal Decomposition method adopted an additive model with a seasonal period of 12 observations. A minimum of 24 observations (two seasonal cycles) was required, and trend extrapolation was performed based on the series frequency.

4.3 Quantitative Analysis

In this section, we present the quantitative results of the experiment, highlighting the frequency and distribution of anomalies identified in the expenditure time series. Figures illustrate histograms representing the number of anomalies per time series, per city and per function.

In the analysis per time series (Figure 2), the average number of anomalies is 3.51. Given that each series contains 79 months of data, an average of 3.51 months with anomalies corresponds to 4.44% of the series. Some series showed significantly higher rates, with three cases where over 40% of the values were identified as anomalous. These cases may indicate data quality issues in the database, corroborating results from previous studies that analyzed the same dataset [Oliveira *et al.*, 2023a].

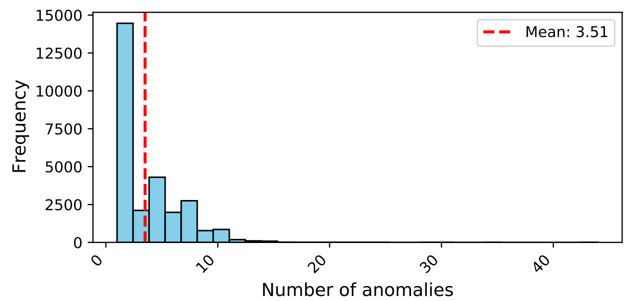

Figure 2. Number of anomalies per time series.

In the analysis per city (Figure 3), each city has 29 time series of 79 records each, totaling 2,291 expenditures analyzed per city. The average number of anomalies detected per city was 113.65, representing 4.96% of the analyzed expenditures. The city with the highest number of anomalies had 174 records classified as anomalous, which equates to 7.60% of its expenditures.

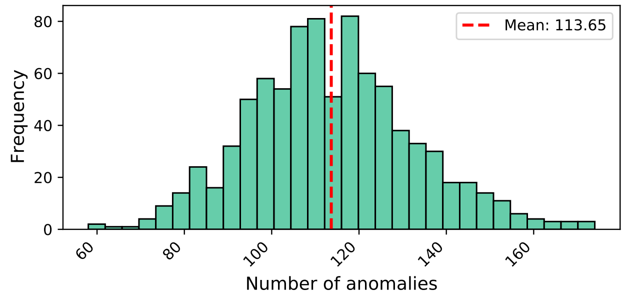

Figure 3. Number of anomalies per city.

Finally, the analysis by function (Figure 4) reveals how anomalies are distributed across different categories of public expenditure. It shows a strong concentration of functions with few anomalous records: 589 of them exhibit up to 185 occurrences, indicating predominantly regular behavior. As the intervals increase, the frequency drops sharply, with only a residual number of functions surpassing one thousand anomalies and very few reaching levels above 3,500 cases. This pattern highlights a highly asymmetric distribution, in which a small number of functions concentrate a disproportionate volume of alerts. As observed in the other levels of aggregation, these extreme cases may reflect not only greater intrinsic variability in the series but also potential inconsistencies specific to certain expenditure categories, suggesting the need for additional domain-oriented investigation.

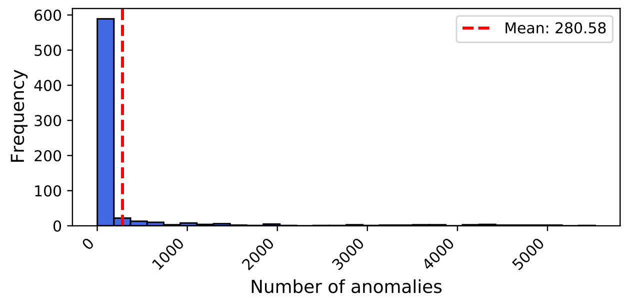

Figure 4. Number of anomalies per function.

In Table 4, the cells on the sides represent the total number of anomalies detected by each method, both from the proposed approach (md_x) and from the *baseline* (BL_{fx}, BL_{geo}). The values in parentheses indicate the percentage of anomalies relative to the total of 1,045,189 expenditures analyzed. The other cells represent the intersection of the number of anomalies detected between the methods of the proposed approach and those of the *baseline*.

The Z-score and IQR methods identified a much lower number of anomalies compared to the other algorithms, which can be attributed to their simplicity. In contrast, the other methods presented quantities of anomalies in the same order of magnitude, suggesting greater robustness in detecting

Table 4. Number of anomalies detected per method, including baselines.

Method	ZSC	IQR	SDMA	LSDA	ISF	LSTM	LOF	SDKM	ARIMA	Total
BL_{fx}	60	232	3,250	1,726	3,872	3,819	2,895	4,675	3,633	19,258 (1.842%)
BL_{geo}	74	298	1,246	1,315	1,764	1,690	1,412	2,012	1,572	28,143 (2.692%)
Total	174 (0.017%)	1,858 (0.178%)	16,113 (1.542%)	35,064 (3.355%)	26,374 (2.523%)	24,583 (2.352%)	24,839 (2.377%)	20,595 (1.970%)	23,353 (2.234%)	

atypical patterns.

In total, 87,779 expenditures were classified as anomalous, i.e., 8.40% of the dataset. Of this total, 44,844 expenditures (4.29%) were considered anomalous by at least one of the *baseline* methods. Furthermore, 9,425 expenditures were detected simultaneously by the proposed approach and the *baseline*, indicating that 21.01% of the anomalies identified by the *baseline* were also captured by the proposed approach.

4.4 Qualitative Analysis

The qualitative results of this experiment focus on prioritizing the most critical anomalous expenditures, using the alert indicator (I_d) defined in Equation 1. Table 5 presents, for illustrative purposes, the most critical anomalous expenditures in a ranking of size 10, according to the indicator I_d.

As mentioned in Section 4.3, the intersection between the anomalies identified by the proposed approach and those detected by the *baseline* methods is only 21.01%. However, this overlap increases significantly when analyzing the most critical anomalies at the top of the ranking.

Figure 5 illustrates the intersection rate between the methods as the ranking size increases. For a ranking size up to 5, the intersection reaches 100%, meaning all ranked Expenditures were also detected by the *baseline*. As the ranking size increases, this intersection gradually decreases, reaching 73% at a ranking size of 100 and 70.6% at 150, indicating that the values converge toward approximately 70%.

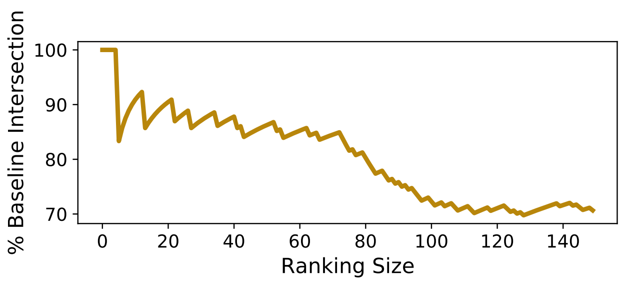

Figure 5. Percentage of intersection between the *baseline* and the proposed approach for different ranking sizes.

Overall, these results indicate that the proposed approach, coupled with the ranking mechanism, is effective in prioritizing the most relevant cases. This enables a more efficient targeting of audit efforts, allowing resources to be allocated to the expenditures with the highest potential for irregularities, in line with previous studies on anomaly detection in public spending.

4.5 Weak-Label Validation using External Audit Evidence

To partially address the lack of ground-truth labels in public expenditure anomaly detection, we conducted an additional validation step based on external evidence from audit-related investigations.

We collected news reports published on the official portal of the Minas Gerais Public Prosecutor's Office (MPMG) that describe investigations involving irregularities in municipal procurement processes. From these reports, we manually identified cases in which the news explicitly referenced the procurement process. This information allowed us to link the investigated procurement processes to the corresponding expenditure records in our dataset. Table 6 presents the eight procurement processes with reported irregularities, including the corresponding city, government function, subfunction, and a link to the news report.

After identifying these procurement processes, we retrieved the related payments and located them within the corresponding municipal expenditure time series. A single procurement process may involve multiple government sectors; therefore, it can generate expenditures associated with different functions. In addition, procurement processes typically involve multiple payments over time, meaning that their total value may appear in expenditures recorded across several months.

In total, the eight investigated procurement processes correspond to 81 payments. Since each expenditure record in our dataset aggregates multiple payments and other expenses, each payment is included within a single expenditure entry. Consequently, 81 expenditure records are associated with procurement processes that were previously investigated for irregularities.

For example, in Table 6, the procurement process #6 generated expenditures associated with three different government functions and payments distributed across eight months (from May to December 2023). Consequently, this single procurement process is associated with $3 \times 8 = 24$ expenditure records in the dataset.

Among the eight procurement processes, five (62.5%) had at least one related expenditure flagged as an anomaly by the proposed framework. Considering the individual payments, 11 out of the 81 expenditures (13.5%) were identified as anomalies. It is important to note that the fraudulent value associated with a procurement process may be diluted across multiple functions and months, since procurement contracts are often executed through several payments over time. Furthermore, the analyzed time series represent aggregated monthly expenditures, which makes the detection of such

Table 5. Top 10 anomalous expenditures according to the alert indicator (I_d).

#	City	Function	Date	Expenditure Value	I_d
1	Belo Horizonte	Transport	07/2023	248,194,900	58.76
2	Montes Claros	Administration	08/2018	20,938,310	51.24
3	Extrema	Administration	03/2022	16,196,690	50.46
4	Itabirito	Sports and Leisure	08/2023	15,266,440	50.28
5	Congonhas	Housing	03/2023	11,324,470	49.38
6	Poços de Caldas	Special Charges	12/2018	8,273,214	48.42
7	Itabirito	Commerce and Services	02/2022	7,127,600	47.97
8	Cambuí	Administration	03/2023	5,880,813	47.38
9	Extrema	Judiciary	06/2019	5,665,788	47.27
10	Araporã	Administration	09/2020	5,392,563	47.12

Table 6. Procurement processes with investigated irregularities, including city, function, subfunction, and reference.

Procurement Process	City	Function	Subfunction	Payments	Reference
#1	Delta	Administration	General Administration	01, 07, 12/2023	MPMG [2024f]
#2	Ijaci	Judiciary	Judicial Action	01–12/2018	MPMG [2024c]
#3	Ipatinga	Administration	General Administration	06–12/2022	MPMG [2024d]
		Education	Primary Education	06–12/2022	
#4	Padre Paraíso	Education	Primary Education	04–12/2022	MPMG [2024b]
#5	Poté	Transport	Road Transportation	05–12/2018	MPMG [2024a]
#6	Pouso Alegre	Education	General Administration	05–12/2023	MPMG [2024e]
		Administration	Urban Infrastructure	05–12/2023	MPMG [2024e]
		Health	Hospital and Outpatient Care	05–12/2023	MPMG [2024e]
#7	Serranos	Health	Prophylactic and Therapeutic Support	02/2021	MPMG [2024g]
#8	Serranos	Health	Prophylactic and Therapeutic Support	03–12/2021	MPMG [2024g]

cases more challenging. Despite these limitations, these investigated procurement processes provide a weak signal that can be used for partial validation of the anomaly detection framework.

Regarding the ranking results, the highest-ranked expenditure associated with a fraudulent procurement process appears at position 540 in the global ranking, corresponding to a position above the 98th percentile. In contrast, the lowest-ranked such expenditure appears at position 67,659, which still lies above the 86th percentile.

However, given the limited number of labeled procurement processes and the inherently weak nature of this supervision signal, we also performed an evaluation using a restricted ranking. In this setting, only expenditures belonging to the same time series as the investigated procurement process are considered, i.e., those corresponding to the same city and government function. As a result, each ranking contains 79 expenditures, corresponding to the length of each time series, as reported in Table 3.

Table 7 reports the position of the highest-ranked expenditure associated with each fraudulent procurement process. These results are substantially higher than those obtained in the global ranking, with some expenditures appearing among the top positions. Based on the position of the highest-ranked relevant expenditure for each procurement process, it is possible to compute the Mean Reciprocal Rank (MRR), defined as the average of the inverse position of the first relevant item. In this restricted evaluation, the MRR reaches 0.497.

Although this procedure does not yield a fully labeled dataset, it provides a form of weak supervision that enables partial grounding of the detected alerts. Specifically, it allows evaluating whether expenditures linked to procurement processes subsequently investigated by authorities are ranked among the most prominent anomalies identified by the proposed method.

This external validation step provides additional evidence about the practical relevance of the detected anomalies and partially addresses the evaluation challenges commonly found in anomaly detection scenarios where confirmed fraud labels are scarce.

5 Analysis at the Subfunction Level

The initial experiment demonstrated the efficacy of our approach for anomaly detection at the function level, a broad categorization of public expenditure. To investigate whether a more granular analysis could yield deeper or more precise insights, we conducted an extended experiment by applying the same methodology at the subfunction level. In the Brazilian budget model, each expenditure function is subdivided into more specific subfunctions, providing a finer-grained view of governmental activities. Table 1 contains examples of eight expenditure functions and their three most frequent subfunctions. This hierarchical structure allows for a more detailed examination of spending patterns, potentially uncovering irregularities that might be averaged out or hidden within the broader function-level aggregates. This section details this extended analysis, exploring the implications of increased data granularity on the anomaly detection process and its results.

5.1 Methodology Extension

In the additional experiment conducted for this extended version of the paper, the methodology was expanded to analyze public expenditures not only at the function level but also at the finer-grained subfunction level.

To integrate this new dimension into our formalization, while preserving the structure and operations de-

Table 7. Position and anomaly count of the highest-ranked expenditure associated with each fraudulent procurement process in the restricted ranking, along with the MRR in this restricted evaluation.

Procurement Process	City	Function	Anomalies	Position
#1	Delta	Administration	1	4
#2	Ijaci	Judiciary	1	4
#3	Ipatinga	Administration	5	1
		Education	3	1
#4	Padre Paraíso	Education	1	3
#5	Poté	Transport	0	46
#6	Pouso Alegre	Education	1	2
		Administration	4	1
		Health	5	1
#7	Serranos	Health	0	17
#8	Serranos	Health	0	17
			Mean Reciprocal Rank (MRR)	**0.497**

fined previously, we adopted a representation strategy in which each subfunction is bound to its corresponding function. Formally, for each function $F = \{f \mid f \in \mathit{Function}\}$, we define a corresponding set of subfunctions : $F' = \{f' \mid f' = \mathit{Function\_Subfunction}\}$

Rather than introducing an additional base set and modifying all operators, we construct a composite functional classification, denoted function-subfunction, which uniquely identifies each (function, subfunction) pair. Formally, we define a new attribute: $d[f'] = f\_ \mid\mid \_f'$, where the symbol " $||$"denotes string concatenation. This composite label replaces the former function attribute in the modeling pipeline.

For example, considering the expenditure functions and their three most common subfunctions presented in Table 1, if an expenditure record d has $d[f] =$ *"Education"*, then $d[f']$ may take values such as *"Education_||_Administration of Elementary Education"*. Or, for $d[f] =$ *"Health"*, $d[f'] =$ *"Health_||_Primary Health Care Services"*.

With this transformation, the entire methodological framework (sets, mappings, time-series construction, and anomaly detection procedures) remains unchanged. Each public expenditure is now associated with a combined function–subfunction classification, and the time series previously defined as $S_{c,f}$ is naturally extended to $S_{c,f'} = ord_m(\{d \in D \mid d[c] = c \wedge d[f'] = f'\})$, following the same ordering by month.

This design choice ensures methodological consistency across both experiments. It also enables deeper analysis of spending behavior while avoiding additional complexity in the formalism. As a result, the proposed framework seamlessly supports analyses at different levels of granularity within the functional classification of public expenditures.

5.2 Extended Experiment

The dataset used in this extended analysis is the same as described in Section 4.1. A key characteristic of this dataset is that it natively contains the subfunction classification for each expenditure record, alongside the broader function classification. This existing data structure allowed for a direct application of our methodology at this more granular level without requiring additional data processing or enrichment.

To ensure a consistent and fair comparison, the two baseline methods (BL_{fx} and BL_{geo}) described in Section 4 were also re-executed using the subfunction-level data. This means that the grouping heuristics for both baselines were applied to the set of expenditures disaggregated by subfunction. This allows for a direct evaluation of how the proposed anomaly detection methods perform against established baselines when operating at a finer level of budgetary detail.

5.3 Extended Quantitative Analysis

The distribution of anomalies per time series (Figure 6) shows a strong concentration in the initial intervals, with most series presenting up to four anomalous records. The frequency decreases steadily across subsequent bins and becomes minimal beyond approximately 13 occurrences. With a mean of 4.87 anomalies per series, the overall behavior remains relatively stable, while a small number of higher-count cases may reflect specific irregular temporal patterns that deserve additional investigation.

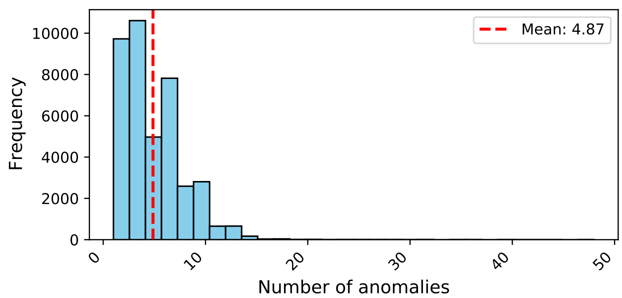

Figure 6. Number of anomalies per time series.

The analysis of anomalies per city (Figure 7) indicates that most cities fall within intermediate ranges, particularly between roughly 180 and 245 anomalous records, where the highest frequencies are observed. As the number of anomalies increases, the frequencies decline, and only a few cities exceed 400 occurrences. The mean of 228.94 anomalies per city suggests a moderately dispersed distribution, with some outliers potentially associated with particular reporting practices or localized data-quality issues.

The distribution of anomalies per function (Figure 8) is relatively sparse, with only a few functions concentrated in the lower and mid-range intervals. Most bins contain few or no occurrences, indicating that only a limited set of func-

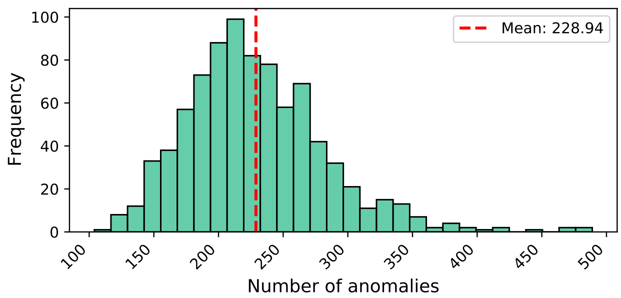

Figure 7. Number of anomalies per city.

tions accumulates larger numbers of anomalies. Although the mean reaches 6,974.5 anomalies, this value is influenced by a small group of functions with substantially higher counts. This asymmetry suggests that deviations are not evenly distributed across expenditure categories and may be linked to structural characteristics of specific functions or to isolated data inconsistencies.

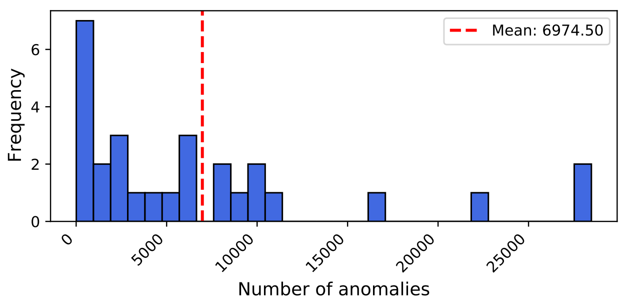

Figure 8. Number of anomalies per function.

The distribution of anomalies per subfunction (Figure 9) is similarly concentrated at the lower end, with 58 subfunctions presenting fewer than 930 anomalous records. Frequencies decline across the remaining bins, and most upper intervals contain no occurrences. Only a few subfunctions appear beyond 8,000 anomalies, reinforcing the asymmetric pattern observed at the function level. With a mean of 1,808.20 anomalies, mostly driven by these higher-count cases, the results indicate that anomalous behavior is concentrated in a subset of subfunctions, whereas the majority maintain comparatively regular dynamics.

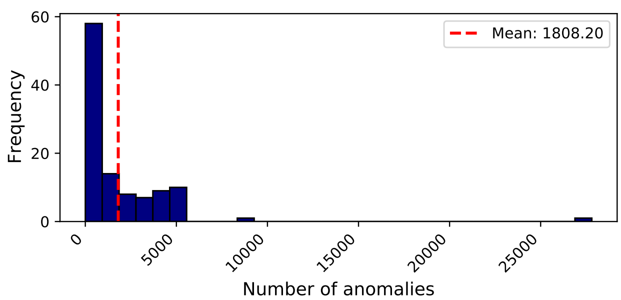

Figure 9. Number of anomalies per subfunction.

The extension of the methodology to the subfunction level allowed for a more granular analysis of public expenditures. This finer granularity, however, resulted in a larger number of time series with fewer data points each, which influenced the anomaly detection performance of several methods. In total, 2,439,369 records were analyzed at the subfunction level, representing a substantial increase compared to the 1,045,189 records at the function level. This broader scope naturally led to a higher absolute number of detected anomalies across most methods.

Comparing the results between the function level (Table 4) and the subfunction level (Table 8) reveals notable shifts in anomaly detection behavior. Simple statistical methods (ZSC, IQR) continued to identify the fewest anomalies, both in absolute and percentage terms, consistent with their limited sensitivity to non-normal distributions and smaller groupings. In contrast, time series–based methods such as SDMA, LSDA, LSTM, and ARIMA maintained strong performance, sustaining high anomaly counts even under increased data fragmentation. Methods like ISF and LOF also remained stable. A marked reduction was observed for SDKM, which was particularly affected by the smaller sample sizes per subfunction, limiting the clustering model's ability to capture consistent patterns.

Despite the overall increase in the detected anomalies, driven primarily by the larger volume of data at the subfunction level, the intersection with the baselines decreased. This reduction suggests that the additional alerts generated at this finer granularity tend to diverge more from the baseline expectations, reflecting either greater heterogeneity across subfunctions or heightened sensitivity of the methods to localized variations that the baselines do not capture.

5.4 Extended Qualitative Analysis

Table 9 presents the top anomalous expenditures according to the alert I_d. The ranked cases reflect a combination of high expenditure values, temporal deviations, and consistent identification across multiple detection methods. As observed in the function-level experiment, the top positions contain records of substantial monetary value, such as the Transport expenditure of Belo Horizonte (07/2023) and the Administration expenditure of the same city (09/2018). These cases are typically characterized by simultaneously high values and strong deviations relative to their respective subfunctions, which explains their prominence in the ranking.

A notable aspect of the extended experiment is the intersection between the Top 10 rankings of the function-level (Table 5) and subfunction-level analyses (Table 9). Three records appear in both tables: (i) *Belo Horizonte – Transport – 07/2023*, (ii) *Itabirito – Commerce and Services – 02/2022*, and (iii) *Araporã – Administration – 09/2020.*

The recurrence of these expenditures across both analyses suggests that their anomalous nature is not dependent on the level of aggregation. In other words, these records stand out both in the broader function context and within their respective subfunctions, reinforcing their relevance as potential priority cases for auditing. Their persistence across different granularities also indicates that the deviations cannot be attributed merely to imbalanced or overly aggregated categories; instead, they likely reflect substantive irregularities or atypical behavior in the underlying expenditure process.

The behavior observed in Figure 10 contrasts with the results obtained at the function level in Figure 5. Whereas the previous experiment showed a declining intersection as the ranking size increased, indicating strong agreement with the baseline only among the highest-ranked anomalies, the

Table 8. Number of anomalies detected per method, including *baselines*.

Method	ZSC	IQR	SDMA	LSDA	ISF	LSTM	LOF	SDKM	ARIMA	Total
BL_{fx}	24	85	616	247	846	762	468	362	834	44,025 (1.805%)
BL_{geo}	13	52	131	77	148	147	86	40	150	28,944 (1.187%)
Total	519 (0.021%)	4,867 (0.200%)	40,939 (1.678%)	79,355 (3.253%)	73,187 (3.000%)	58,302 (2.390%)	64,033 (2.625%)	15,235 (0.625%)	60,221 (2.469%)	

Table 9. Top anomalous expenditures according to the alert indicator (I_d).

#	City	Function	Subfunction	Date	Expenditure Value	I_d
1	Belo Horizonte	Transport	Urban Services	07/2023	248,190,200	50.37
2	Itabirito	Commerce and Services	Trade Promotion	02/2022	7,127,600	47.97
3	Nova Lima	Labor	Worker Protection and Benefits	03/2020	5,152,971	46.98
4	São Gonçalo do Rio Abaixo	Essential to Justice	Legal Representation	12/2020	4,691,108	46.70
5	Araporã	Administration	Financial Administration	09/2020	4,462,142	46.55
6	Manhuaçu	Administration	Financial Administration	01/2018	601,845	46.24
7	São João do Paraíso	Administration	Financial Administration	01/2018	599,842	46.22
8	Belo Horizonte	Administration	General Administration	09/2018	50,303,220	46.21
9	Mariana	Social Security	General Administration	03/2019	3,816,365	46.07
10	Contagem	Essential to Justice	Legal Representation	10/2022	27,420,690	44.63

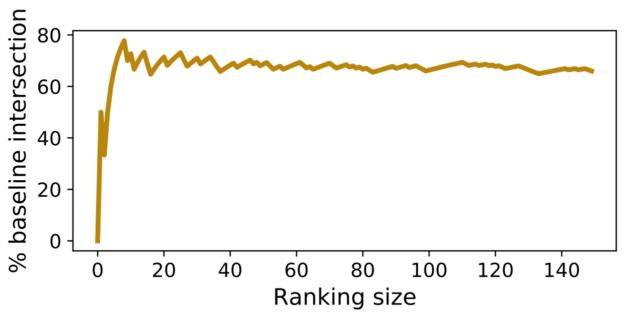

Figure 10. Percentage of intersection between the *baseline* and the proposed approach for different ranking sizes.

subfunction-level curve exhibits the opposite pattern. At smaller ranking sizes, the overlap with the baseline is comparatively low, reflecting greater divergence in the top anomalies detected by the proposed method. However, as the ranking expands, the intersection percentage rises steadily and eventually stabilizes around 68%–69%. This increase suggests that highly ranked anomalies at the subfunction level tend to be more method-specific, while lower-ranked alerts exhibit broader consensus with the baseline.

This trend can be attributed to the finer granularity of the subfunction analysis, which generates a substantially larger number of shorter and more heterogeneous time series. Such fragmentation heightens the sensitivity of the detection methods, resulting in top-ranked anomalies that capture localized or context-specific deviations not always shared with the baseline. As the ranking grows, however, the inclusion of less extreme anomalies incorporates patterns that are more stable and aligned with general expenditure behavior, thereby increasing the overlap. The resulting convergence highlights that, despite methodological differences, both approaches retain a meaningful degree of agreement when a broader portion of the anomaly spectrum is considered.

5.5 Extended Weak-Label Validation using External Audit Evidence

To further assess the effectiveness of the proposed framework, we conduct an additional weak-label validation using the same external audit evidence described in Subsection 4.5. This extended analysis follows the same validation protocol applied to the original ranking, but considers an alternative configuration that incorporates the function–subfunction representation. The evaluation is based on the same eight procurement processes and their corresponding 81 payments presented in Table 6. As previously discussed, these payments map to 81 expenditure records in our dataset, since each payment is associated with a single aggregated expenditure entry.

Among the eight investigated procurement processes, five (62.5%) have at least one related expenditure flagged as anomalous by the proposed framework. At the individual level, 10 out of the 81 corresponding expenditure records (12.3%) are detected as anomalies.

Regarding the ranking results, the highest-ranked expenditure associated with a fraudulent procurement process and flagged as an outlier appears at position 67,659 in the global ranking, corresponding to the top 2%. Conversely, the lowest-ranked such expenditure appears at position 128,827, which still lies within the top 14% of the ranking.

A more fine-grained analysis is obtained by restricting the ranking to expenditures sharing the same city, government function, and subfunction as the investigated procurement process. Under this setting, the positions of relevant expenditures improve substantially. Table 10 summarizes the rank of the top occurrence associated with each fraudulent procurement process, with several cases among the highest-ranked entries. This localized evaluation is particularly appropriate given the scarcity of labeled procurement processes and the inherently weak supervision signal. Based on these positions, we compute the Mean Reciprocal Rank (MRR), defined as the average inverse rank of the first relevant item, which is 0.546.

Overall, this extended validation using the function-subfunction ranking provides additional empirical support for the practical relevance of the detected anomalies. As in the previous evaluation, leveraging external audit evidence helps mitigate the challenge of limited ground-truth labels in fraud detection and public expenditure anomaly analysis.

6 Case Study: City of Araporã – Administration Function

To illustrate the behavior of the proposed anomaly detection framework in a real-world setting, we analyze the expenditure time series of the city of Araporã for the Administration government function and its corresponding subfunctions. These series were selected because they appear among the top 10 anomalous expenditures in both experimental settings, as reported in Tables 5 and 9.

Figure 11 shows the monthly expenditures along with the number of anomaly detection methods that flagged each observation as anomalous. The blue line represents the time series for the Administration function, while the remaining lines correspond to its subfunctions. Table 11 details the specific methods that identified anomalies in the months with the highest number of detected outliers.

6.1 Function-level Analysis

Overall, the main Administration time series exhibits relatively stable expenditure levels with moderate fluctuations over time. However, a prominent spike is observed around the middle of the series, where total expenditure in the Administration function increases sharply compared to the surrounding months. This peak is simultaneously identified by multiple anomaly detection methods. In particular, the highest point in the series is flagged by up to six detectors, indicating strong consensus among the detection techniques that this observation represents an atypical spending pattern.

Beyond detecting anomalous observations, it is essential to understand their underlying characteristics and the reasons why different methods identify them as anomalous. This interpretation is particularly important for auditors, as distinct anomaly patterns may reflect different operational or administrative circumstances.

The most prominent anomalies in the Araporã Administration series occur in September and October 2020. Both months correspond to sharp spikes in expenditures relative to the historical pattern. In September 2020, seven methods identified the observation as anomalous, including forecasting-based models such as ARIMA and LSTM, density-based methods such as Isolation Forest and Local Outlier Factor, clustering-based detection using Soft-DTW k-means, and decomposition-based detection using Seasonal Decomposition with Moving Averages. In addition, the Level Shift detector also flagged this observation, indicating that the magnitude of the increase is large enough to potentially alter the baseline level of the series. This consensus across heterogeneous methods strongly suggests that the spike represents an abrupt and atypical expenditure event.

In October 2020, a second anomaly is detected by six methods, although the Level Shift method no longer signals a structural change. This suggests that the anomaly corresponds to a continuation of unusually high spending rather than a new structural shift in the time series. Forecasting-based models again present large prediction errors, while density-based methods identify the observation as distant from the historical distribution of expenditures. In practical terms, this pattern may indicate that the spending surge observed in the previous month extended over multiple payment installments, a situation that auditors may verify by examining procurement contracts or payment schedules associated with the Administration function.

6.2 Subfunction-level Analysis

The subfunction-level analysis helps explain the origin of these anomalies. Both spikes are primarily driven by the Financial Administration subfunction, which presents anomalies in the same months and is detected by nearly all methods. This indicates that the unusual behavior observed in the aggregated Administration function is largely attributable to expenditures associated with financial management activities. For auditors, this could suggest that relevant evidence may be found in accounting, treasury management, or financial administration contracts executed during this period.

Additional anomalies appear when analyzing individual subfunctions. For example, the General Administration subfunction presents an anomaly in December 2022, detected by forecasting and density-based methods. In this case, the anomaly corresponds to a moderate spike that deviates from the expected seasonal pattern but does not affect the overall Administration function strongly enough to generate an alert at the aggregated level.

Finally, an anomaly is detected in the Information Technology subfunction in May 2019. The methods that identify this observation include forecasting models (ARIMA and LSTM), density-based methods (Isolation Forest and Local Outlier Factor), and seasonal decomposition techniques. This combination of methods suggests a short-term deviation from the expected seasonal behavior of IT-related expenditures.

6.3 Alert Fatigue

The case study also illustrates an important practical aspect of the proposed extension to the subfunction level. While subfunction-based analysis increases the granularity of the monitoring process, it also leads to a larger number of anomaly signals. In Figure 11, several alerts appear at the subfunction level even when the aggregated Administration function does not exhibit a strong anomaly. This situation may potentially lead to alert fatigue if auditors are required to inspect all flagged observations individually.

To mitigate this issue, the ranking mechanism supports prioritization strategies based on consensus among methods and the hierarchical relationship between functions and subfunctions. In practice, auditors can first focus on months in which the aggregated function itself is flagged as anomalous or in which multiple subfunctions simultaneously present anomalies. Such situations indicate coordinated deviations across different expenditure categories and therefore represent stronger signals of unusual behavior. Conversely, isolated anomalies restricted to a single subfunction and detected by only one method can be assigned lower priority and investigated only when additional contextual information suggests potential irregularities.

Table 10. Position and anomaly count of the highest-ranked expenditure associated with each fraudulent procurement process in the restricted ranking incorporating the function–subfunction representation, along with the MRR in this restricted evaluation.

Procurement Process	City	Function	Subfunction	Anomalies	Position
#1	Delta	Administration	General Administration	1	5
#2	Ijaci	Judiciary	Judicial Action	1	4
#3	Ipatinga	Administration	General Administration	4	1
		Education	Primary Education	4	1
#4	Padre Paraíso	Education	Primary Education	1	3
#5	Poté	Transport	Road Transportation	0	15
#6	Pouso Alegre	Education	General Administration	2	1
		Administration	Urban Infrastructure	3	1
		Health	Hospital and Outpatient Care	5	1
#7	Serranos	Health	Prophylactic and Therapeutic Support	0	13
#8	Serranos	Health	Prophylactic and Therapeutic Support	0	13
				Mean Reciprocal Rank (MRR)	**0.546**

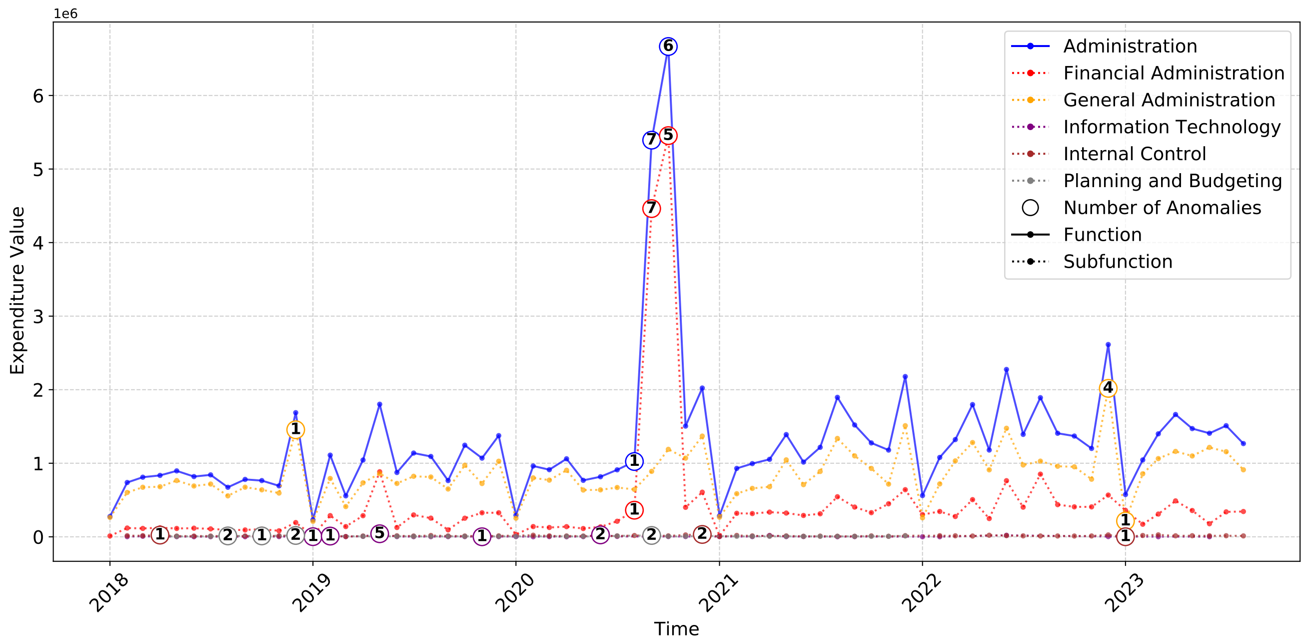

Figure 11. Expenditure time series of the city of Araporã for the Administration function.

In the Araporã case, the most prominent anomaly corresponds to a substantial expenditure spike, in which both the Administration function and the Financial Administration subfunction are flagged by multiple detection methods. Events of this nature naturally rank among the highest-priority cases, whereas smaller deviations restricted to individual subfunctions can be treated as secondary alerts. By leveraging both hierarchical aggregation and consensus across methods, the framework mitigates alert overload and directs audit efforts toward the most relevant expenditure patterns.

7 Conclusion

This work proposes an approach for detecting anomalies in cities' expenditure time series and a method for ranking the most critical cases. The proposed framework aims to identify and prioritize unusual spending patterns, which should be interpreted as potential alerts requiring further investigation by auditors or domain experts. To do so, the approach combines several time series anomaly detection techniques, including statistical, machine learning, and probabilistic methods.

To validate the effectiveness of the approach, we conducted an experiment with an extensive dataset of municipal public expenditures from Minas Gerais. The results demonstrate that the approach is capable of identifying cases with high potential for irregularities, especially when combined with the ranking method. These results suggest that the approach can be a valuable tool for auditors, providing an initial screening analysis in the face of the large volume of available data. Furthermore, the proposed approach highlights the importance of an integrated and multifaceted approach to data analysis in auditing, enabling an automated survey of potential irregularities in city expenditures.

The extended experimental analysis conducted at the subfunction level further validated the robustness and scalability of our methodology. While the finer granularity introduced greater data fragmentation and altered the detection dynamics of some methods, the core approach remained effective. The persistence of key anomalies across both function and subfunction levels, such as the high-value expenditures in Belo Horizonte and Itabirito, underscores that the most criti-

Table 11. Anomaly detection methods responsible for the months with the highest number of flagged outliers in the expenditure time series of Araporã (Administration function and subfunctions).

Subfunction	Date	Anomalies	Methods
–	09/2020	7	ARIMA, ISF, LOF, LSAD, LSTM, SDMA, SDKM
Financial Administration	09/2020	7	ARIMA, ISF, LOF, LSAD, LSTM, SDMA, SDKM
–	10/2020	6	ARIMA, ISF, LOF, LSTM, SDMA, SDKM
Financial Administration	10/2020	5	ISF, LOF, LSTM, SDMA, SDKM
General Administration	12/2022	4	ARIMA, ISF, LOF, LSTM
Information Technology	05/2019	5	ARIMA, ISF, LOF, LSTM, SDMA

cal irregularities are identified consistently regardless of the analysis granularity. This reinforces the reliability of the proposed ranking mechanism for prioritizing audit targets. The differing intersection patterns with the baseline methods between the two experiments highlight the nuanced behavior of anomaly detection at varying levels of detail, yet the overall high agreement for larger ranking sizes confirms the method's stability.

7.1 Limitations

The main limitation of this work is the absence of labeled data indicating confirmed irregularities or auditing outcomes. As a result, the evaluation of the anomaly detection methods relies primarily on statistical and methodological criteria, rather than on supervised validation against known ground-truth cases. While the proposed framework is effective in identifying unusual spending patterns, the detected anomalies should be interpreted as potential alerts requiring further investigation by auditors or domain experts.

Moreover, the applicability of the methodology depends on the availability and comparability of public expenditure data. In particular, the approach requires time series of city (or subnational) expenditures that can be consistently aggregated into comparable categories across jurisdictions and over time. This structure enables meaningful comparisons of spending patterns and the identification of anomalous behavior. However, in some countries, the adoption of such methods may be restricted by weak transparency policies, limited access to fiscal data, or inconsistencies in budget systems across cities.

7.2 Future Work

As future work, we plan to conduct evaluations with expert auditors to assess the relevance of the ranked alerts produced by the proposed framework. In this step, domain specialists will analyze selected alerts and provide feedback on their practical significance in real auditing scenarios. This feedback can be incorporated to refine the methodology, for instance by adjusting detection thresholds, improving the ranking strategy, and prioritizing anomalies that are more likely to correspond to meaningful fiscal irregularities.

Another important direction is the exploration of anomaly detection techniques tailored to multidimensional time series. In addition, we intend to perform a more comprehensive sensitivity and robustness analysis of the proposed framework, examining how the results vary under different detection thresholds and alternative ranking formulations, including variations in the weighting scheme based on the logarithm of expenditure values.

Declarations

Funding

This work was funded by the Prosecution Service of the State of Minas Gerais (in Portuguese, *Ministério Público do Estado de Minas Gerais*, or simply MPMG) through the Analytical Capabilities Program (in Portuguese, *Programa de Capacidades Analíticas*). To CNPq and Amazon Web Services (AWS) for the funding received through the project of the partnership between them. Finally, to CAPES and FAPEMIG.

Authors' Contributions

MTD: Data curation, Software, Visualization, Writing – original draft; **LGLC:** Conceptualization, Formal analysis, Methodology, Writing – original draft; **GPO:** Methodology, Validation, Writing – review & editing; **MOS:** Methodology, Validation, Writing – review & editing; **GLP:** Funding acquisition, Resources, Supervision.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The raw data used in this study is publicly available at `https://dadosabertos.tce.mg.gov.br/`. The preprocessed dataset, including all constructed time series, is available at `https://github.com/LucasLage/mg_city_expenditure_time_series`. Accessed on 18 May 2026.

References

Alfian, A., Ritchi, H., and Adrianto, Z. (2023). Fraud analytics practices in public-sector transactions: a systematic review. *Journal of Public Budgeting, Accounting & Financial Management*, 35(5):685–710. DOI: https://doi.org/10.1108/JPBAFM-11-2022-0175.

Blázquez-García, A. *et al.* (2021). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys*, 54(3):1–33. DOI: https://doi.org/10.1145/3444690.

Braz, C. S., Dutra, M. T., Oliveira, G. P., Costa, L. L., Silva, M. O., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2024). Exploring irregularities in brazilian public bids: An in-depth analysis on small companies. *J. Interact. Syst.*, 15(1):349–361. DOI: https://doi.org/10.5753/JIS.2024.3836.

Braz, C. S. *et al.* (2023). Análise de irregularidades em licitações públicas com foco em empresas de pequeno porte. In *WCGE*, pages 94–105. SBC. Title in English: "Analysis of Irregularities in Public Procurement with Focus on Small Companies". [in Portuguese]. DOI: https://doi.org/10.5753/wcge.2023.230254.

Brazil (2025). Brazilian technical budget manual 2026. Ministry of Planning and Budget, Federal Budget Secretariat https://www1.siop.planejamento.gov.br/mto/lib/exe/fetch.php/mto2026:mto2026.pdf. Accessed on 21 May 2026.

Breunig, M. M. *et al.* (2000). Lof: identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104. DOI: https://doi.org/10.1145/335191.335388.

Costa, L. G. L. *et al.* (2024). Quanto Custa: Banco de Preços de Compras Públicas do Estado de Minas Gerais. In *SBBD DS-CoPS*. SBC. Title in English: "Quanto Custa: Price Database for Public Procurement of the State of Minas Gerais". [in Portuguese]. DOI: https://doi.org/10.5753/sbbd_estendido.2024.243804.

Cuturi, M. *et al.* (2017). Soft-dtw: a differentiable loss function for time-series. In *ICML*, pages 894–903. PMLR. https://dl.acm.org/doi/10.5555/3305381.3305474. Accessed on 18 May 2026.

Darban, Z. Z. *et al.* (2024). Deep learning for time series anomaly detection: A survey. *ACM Comput. Surv.*, 57(1). DOI: https://doi.org/10.1145/3691338.

Dutra, M. T., Costa, L. G., Oliveira, G. P., Silva, M. O., and Pappa, G. L. (2025). Detecçao e análise de anomalias em séries temporais de despesas municipais. In *Latin American Symposium on Digital Government (LASDiGov)*, pages 121–132. SBC. Title in English: "Detection and Analysis of Anomalies in City Expenditure Time Series". [in Portuguese]. DOI: https://doi.org/10.5753/lasdigov.2025.8935.

Gomide, L. D. *et al.* (2023). Mineração de dados sobre despesas públicas de municípios mineiros para gerar alertas de fraudes. In *SBBD*, pages 378–383. SBC. Title in English: "Data Mining on Public Expenditures of Municipalities in the State of Minas Gerais to Generate Fraud Alerts". [in Portuguese]. DOI: https://doi.org/10.5753/sbbd.2023.232788.

Handoko, B. L. *et al.* (2022). The effect of skepticism, big data analytics to financial fraud detection moderated by forensic accounting. In *ICEEG*, page 123–130. DOI: https://doi.org/10.1145/3537693.3537703.

Hethu Avinash, D. *et al.* (2024). Integrating level shift anomaly detection for fault diagnosis of battery management system for lithium-ion batteries. *IEEE Access*, 12:116071–116084. DOI: https://doi.org/10.1109/ACCESS.2024.3445955.

Hyndman, R. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts, 3nd edition. https://otexts.com/fpp3/. Accessed on 21 May 2026.

Li, B., Kaplan, B., Lazirko, M., and Kogan, A. (2025). Unsupervised outlier detection in audit analytics: A case study using usa spending data. DOI: https://doi.org/10.48550/arXiv.2509.19366.

Liu, F. T., Ting, K. M., and Zhou, Z. (2008). Isolation forest. In *ICDM*, pages 413–422. IEEE Comp. Society. DOI: https://doi.org/10.1109/ICDM.2008.17.

Malhotra, P. *et al.* (2015). Long short term memory networks for anomaly detection in time series. In *ESANN*, volume 89, page 94. https://www.esann.org/sites/default/files/proceedings/legacy/es2015-56.pdf. Accessed on 18 May 2026.

Mendes, B. M. A. *et al.* (2023). Impacto de Doações Eleitorais no Faturamento de Empresas: Um Estudo nas Eleições Municipais em Minas Gerais. In *SBBD*, pages 420–425. SBC. Title in English: "Impact of Electoral Donations on Company Revenues: A Study of Municipal Elections in Minas Gerais". [in Portuguese]. DOI: https://doi.org/10.5753/sbbd.2023.233278.

Mongwe, W. T. *et al.* (2020). The efficacy of financial ratios for fraud detection using self organising maps. In *SSCI*, pages 1100–1106. IEEE. DOI: https://doi.org/10.1109/SSCI47803.2020.9308602.

MPMG (2024a). Court accepts complaint filed by mpmg against former mayor of poté for fraud in public procurement. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/justica-recebe-denuncia-do-mpmg-contra-ex-prefeito-de-potepor-fraude-em-licitacoes-2C9F8AAB9881AD90019899EDBEE04504-00.shtml. News report. Accessed on 21 May 2026.

MPMG (2024b). Mpmg charges mayor and education secretary for involvement in procurement fraud in padre paraíso. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/mpmg-denuncia-prefeito-secretaria-de-educacao-e-outras-tres-pessoas-por-envolvimento-em-fraude-a-licitacao-em-padre-paraiso.shtml. News report. Accessed on 21 May 2026.

MPMG (2024c). Mpmg charges mayor and municipal attorney for contract without public procurement in ijaci. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/mpmg-denuncia-prefeito-e-procurador-geral-do-municipio-por-contrato-sem-licitacao-em-ijaci.shtml. News report. Accessed on 21 May 2026.

MPMG (2024d). Mpmg files lawsuit against mayor and others for fraud in public procurement in ipatinga. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/mpmg-propoe-acao-contra-prefeito-de-ipatinga-secretaria-e-ex-secretarios-municipais-empresarios-e-empresas-por-fraude-em-licitacao.shtml. News report. Accessed on 21 May 2026.

MPMG (2024e). Mpmg investigates procurement waiver and fraud in public contract execution in pouso alegre. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/mpmg-investiga-casos-de-dispensa-de-licitacao-e-fraude-na-execucao-de-em-contrato-publico-na-prefeitura-de-pouso-alegre.shtml. News report. Accessed on 21 May 2026.

MPMG (2024f). Operation investigates scheme of fraud in public procurement and public examination in delta, minas gerais. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/operacao-investiga-esquema-de-fraude-a-processo-licitatorio-e-concurso-publico-realizados-em-delta-no-triangulo-mineiro.shtml. News report. Accessed on 21 May 2026.

MPMG (2024g). Public agents and businessman from serranos charged with fraud in procurement of medicines. https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/agentes-publicos-e-empresario-d e-serranos-no-sul-de-mg-sao-denunciados-por-fraude-em-licitacao-de-medicamentos.shtml. News report. Accessed on 21 May 2026.

No, W. G., Lee, K., Huang, F., and Li, Q. (2019). Multidimensional audit data selection (mads): A framework for using data analytics in the audit data selection process. *Accounting Horizons*, 33(3):127–140. DOI: https://doi.org/10.2308/acch-52453.

Oliveira, F. B. *et al*. (2024). Machine learning and time series analysis to forecast hotel room prices. In *BRACIS*, pages 358–371. Springer. DOI: https://doi.org/10.1007/978-3-031-79035-5_25.

Oliveira, G. P. *et al*. (2023a). Assessing data quality inconsistencies in brazilian governmental data. *Journal of Information and Data Management*, 14(1). DOI: https://doi.org/10.5753/jidm.2023.3220.

Oliveira, G. P. *et al*. (2023b). Ranqueamento de licitações públicas a partir de alertas de fraude. In *BraSNAM*, pages 1–12. SBC. Title in English: "Ranking Public Procurement Processes Based on Fraud Alerts". [in Portuguese]. DOI: https://doi.org/10.5753/brasnam.2023.232105.

Schneider dos Santos, E., Machado dos Santos, M., Castro, M., and Tyska Carvalho, J. (2025). Detection of fraud in public procurement using data-driven methods: a systematic mapping study. *EPJ Data Science*, 14(1):52. DOI: https://doi.org/10.1140/epjds/s13688-025-00569-3.

Silva, C. A. *et al*. (2020). Consciência, Prevenção e Detecção contra Fraude: Estudo com Auditores Internos e Contadores do Setor Público Brasileiro. In *USP International Conference on Accounting*. Title in English: "Fraud Awareness, Prevention, and Detection: A Study with Internal Auditors and Accountants in the Brazilian Public Sector". [in Portuguese]. https://congressousp.fipecafi.org/anais/20UspInternational/ArtigosDownload/2630.pdf. Accessed on 18 May 2026.

Silva, M. O., Costa, L. L., Gomide, L. D., dos Santos, G. B., Oliveira, G. P., Brandão, M. A., Lacerda, A., and Pappa, G. L. (2024). Overpricing analysis in brazilian public bidding items. *J. Interact. Syst.*, 15(1):130–142. DOI: https://doi.org/10.5753/JIS.2024.3831.

Silva, M. O. *et al*. (2023). Análise de sobrepreço em itens de licitações públicas. In *WCGE*, pages 118–129. SBC. Title in English: "Analysis of Overpricing in Items of Public Procurement Bids". [in Portuguese]. DOI: https://doi.org/10.5753/wcge.2023.230608.

Zafar, U. and Wu, F. (2026). Methodological challenges in explainable ai for fraud detection: a systematic literature review. *Artificial Intelligence Review*, 59(4):115. DOI: https://doi.org/10.1007/s10462-026-11516-7.