Journal of Information and Data Management, 2024, 15:1, doi: 10.5753/jidm.2024.3571

Indicators and Municipal Data: A Database for Evaluating the Efficiency of Public Expenditures

Paula Guelman Davis [**Governo do Estado de Minas Gerais - Secretaria de Estado de Justiça e Segurança Pública** | ***paula.davis@seguranca.mg.gov.br***]

Secretaria de Estado de Justiça e Segurança Pública - Avenida Papa João Paulo II, n. 4143 - Edifício Minas, 5o andar, Bairro Serra Verde, Belo Horizonte - MG, 31630-900, Brazil

Received: 23 July 2023 • **Published:** 6 April 2024

Abstract This article describes the construction of a database with financial and operational data from Brazilian municipalities. Public data were collected regarding expenses by function (education, health, public security, among others), indicators and other data that reflect the municipal situation in the areas of education, health, public security, development, sanitation and finance. Data from various sources were integrated and transformed to allow studies on the correlation between performance indicators of the effectiveness of public governance, and the corresponding spenditures, to follow up and assess the effects of public policies.

Keywords: Public finances, government planning.

1 Introduction

The Constitution of the Brazilian Republic of 1988, along with laws recognizing transparency and information as citizens' rights and the State's duty, such as the Fiscal Responsibility Law (Complementary Law n. 101/2000), the Transparency Law (Complementary Law n. 131/2009) and the Access to Information Law (Law n. 12.527/2011) has facilitated the opening access of data related to public administration for citizens. These legal frameworks, as well as policies that require transparency from public institutions, enable access to a wide variety and quantity of data, made available through various means and formats. However, often, the data are available by area. Despite the amount of data accessible to the public, it remains a challenge to understand the relationship between different sectors. For instance, health and education data may be presented separately, limiting a comprehensive understanding of their interconnections.

According to Silva (2005), the execution of state functions can take place through three of political organization: (a) total centralization - the central government assumes all functions, which means that all aspects of governance are concentrated in the hands of the central authority; (b) absolute decentralization - local governments enjoy complete autonomy and operates independently, without significant oversight or control; and (c) federative form - combines elements of both centralization and decentralization, in which responsibilities are divided among various governmental entities. This is the case with the Brazilian state, where a federative structure allocates responsibilities between different levels of government.

In 1988, the consolidation of fiscal federalism, as highlighted by Bovo (2001), empowered subnational entities through modifications in the distribution of public revenues. The distribution of tax resources and the autonomy granted to federative entities were fundamental for financing local and regional development policies. However, even with an increase in their own revenue-generating capacity, decentralization worsened the financial situation of states and municipalities, compromising the funding of social policies and investments (Matias-Pereira, 2017).

Therefore, through the collection of data from multiple sources, a database was constructed with the aim of aggregating financial and operational data to enable the visualization and analysis of resource allocation and municipal outcomes. These outcomes are measured through indicators across key areas of State activity. The data collection was conducted from government public sources.

The database, subject of this work, was used in a previous study to analyze the efficiency of municipalities' allocative expenses. Despite some limitations, such as the quality of the available data, relevant results were observed for the areas of education, health, development and fiscal situation (Davis and Souza, 2021). In a more in-depth analysis, it was possible to identify, through data mining algorithms, groups of efficient municipalities - those with lower expenses and better outcomes in certain areas compared to the rest of Brazil. This result also allowed for a geographical visualization of municipalities to identify common characteristics for each cluster. This last aspect was detailed in the present article.

The database can be accessed at the URL `https://zenodo.org/record/6828889#.Yvgb6ezMJTY` in .CSV (comma-separated values) format.

This paper is organized as follows: Section 2 discusses related works that used similar datasets related to public expenditures and socio-economic indicators at the municipal level; Section 3 outlines the adopted methodology for data collection, integration, cleaning and transformation for the construction of the database; Section 4 presents the applications of the database; Section 5 outlines the challenges and limitations of the dataset. Finally, Section 6 presents the conclusions of this work.

2 Related Studies

The collection and utilization of data related to public expenditures and socio-economic indicators at the municipal level have been the focus of various international and national studies. Neduziak and Correia (2017) discuss the consequences of public expenditure allocation in relation to the Gross Domestic Product - GDP (Produto Interno Bruto, PIB) of states. Data on budget execution (expenses by function), PIB per capita and rates of population growth and unemployment were collected.

Machado, Irffi and Benegas (2011) used indicators related to sanitation (urban coverage rate of piped water supply and urban coverage rate of sewage), education (number of early childhood education establishments, early childhood education literacy rate, schooling rate), health (infant and child mortality rate), public safety (homicide rate, bodily injury rate, robbery rate and theft rate) and per capita spending on education, culture, health, sanitation and social assistance. The study focused on municipalities in Ceará, aiming to evaluate public expenditures in education, health and social assistance.

In the field of education, Lourenço, Angotti, Nascimento and Sauerbronn (2017) employed data on liquidated expenses in education, average spending per student, Human Development Index (Índice de Desenvolvimento Humano Municipal, IDHm) and Basic Education Performance Index (Índice de Desenvolvimento da Educação Básica, IDEB) for the initial years (1st to 5th grade) and final years (6th to 9th grade). The objective was to assess the efficiency of financial expenditures in the education sector for the 250 largest Brazilian municipalities in terms of enrollment numbers in elementary education.

In the health sector, studies have focused on data provided through DATASUS, which is the Brazilian Public Unified Health System (SUS) department responsible for providing health data, as seen in the work by Politelo, Rigo and Hein (2014). Or studies have used the Unified Health System Performance Index (Índice de Desempenho do Sistema Único de Saúde, IDSUS) such as Almeida, Vasconcelos, Miranda and Feitosa (2017). According to the Ministry of Brazilian Health, IDSUS is a set of indicators designed to assess the performance of the Unified Health System (SUS) based on results related to access and effectiveness of care in primary and specialized care within municipalities across various areas.

In the field of public security, Resende and Andrade (2011) conducted an exploratory analysis of a dataset organized by the National Secretariat of Public Security (SENASP) of the Ministry of Justice in 2004. This dataset was created with the aim of monitoring, in a more efficient way, the incidence of crimes in municipalities with more than 100,000 inhabitants. The authors sought evidence supporting the economic assumption that income inequality contributes to an increase in criminality.

Regarding the socio-economic development area, it was possible to observe studies that, in addition to obtaining data on expenses in various areas, collected development data through the FIRJAN Municipal Development Index (Índice Firjan de Desenvolvimento Municipal, IFDM), as seen in the work of Silva, Costa, Castro and Silva (2016).

Classification	Source	Collected Data	Year
General Data	Brazilian Institute of Geography and Statistics (IBGE)	Federal Unit	2017
		Municipality	2017
		Population Estimate	2017
		Area	2017
		Mayor's party	2013/2017
		Existence of municipality's development plan	2015
Public Expenditure	National Treasury Secretariat - Municipal Finances (FINBRA)	Expenditures paid by function: Public Security, Social Assistance, Health, Education (elementary education) and Sanitation	2017
Fiscal Indicators	National Treasury Secretariat - Municipal Finances (FINBRA)	Net Consolidated Debt (DCL)	2017
		Total Personnel Expenditure (DTP)	
		Net Current Revenue (RCL)	
Education	National Institute for Educational Studies and Research Anísio Teixeira (INEP)	Basic Education Development Index (IDEB)	2017
		Number of enrollments	
		Number of teachers	
		Number of establishments	
		Number of classes	
		Average students per class	
		Daily average hours of class	
		Age-grade distortion rate	
		Percentage of teachers with higher education	
		Approval, repetition and dropout rates	
Health	Ministry of Health	Unified Health System Performance Index (IDSUS)	2011
	DATASUS	Number of hospitals admissions through the Brazilian Unified Health System (SUS)	2017
		Average length of hospital stay	
		Number of basic care outpatient clinics	
		Number of deaths by occurrence	
		Deaths from avoidable causes (0 to 4 years and 5 to 74 years)	
		Vaccination doses administered	
Public Safety	Violence Atlas	Quantity of homicides	2017
Development	Human Development Atlas	Municipal Human Development Index (IDHm)	2010
	Federation of Industries of Rio de Janeiro (FIRJAN)	FIRJAN Municipal Development Index (IFDM)	2016
Sanitation	National Sanitation Information System (SNIS)	Total population served with water supply	2017
		Total population served with sanitation	

Source: Davis and Souza (2021)

Figure 1. Collected Data and Sources, 2021

Finally, it is worth highlighting the study by Cruz and Afonso (2018, which investigates the relationship between the achievement of fiscal goals and limits of the Fiscal Responsibility Law and the characteristics of fiscal management in terms of planning, transparency and control. The analysis considered a sample of 282 large municipalities.

From the analysis of related studies, it is noticeable that, in most cases, there is a geographical limitation, where only municipalities from a specific region or state are considered, or a restriction to a specific area of analysis, such as health. This study demonstrates the construction of a dataset that includes all municipalities in Brazil through the collection of financial and operational data from various areas of state activity. This enables the use of the database in an even more comprehensive and multidisciplinary manner.

3 Methodology

The construction of the database involved data collection from government public information sources. Figure 1 presents the collected data and their respective sources.

The selection of sources was based on the availability of municipal data in the most recent years for each source. As the objective was to capture, through a single dataset, the reality of municipalities in the main areas of state activity, it was not necessary for all data to be collected for the same year. Additionally, each source has different publication frequen-

cies, making it challenging to collect a substantial volume of data for the same year.

The general data were collected from IBGE. The population estimate and the area of municipalities were obtained from specific annual surveys conducted by them. Information about the winning mayor's party in 2013 and 2017 elections and the existence of the municipality's development plan were extracted from the Municipal Basic Information Survey (MUNIC). MUNIC conducts an annual survey on the functioning, dynamics and structure of public institutions in municipalities.

The municipal public Expenditures classified by function (as established in Law n. 4,320 of March 17, 1964) were obtained through the Annual Account Declarations (DCS) submitted by federative entities to the National Treasury Secretariat (STN), as stipulated in Article 51 of the Complementary Lay n. 101 of May 4th, 2000 (Fiscal Responsibility Law). The STN makes this data available through Brazil's Public Finances System (FINBRA). For the purpose of the presented dataset, the areas related to public expenditures were chosen considering the availability of financial and operational data.

Regarding the education area, considering the competencies of municipal entities, only the Elementary Education subfunction was considered. The other areas were considered by function, which includes a variety of subfunctions.

For operation data in education, the data was extracted from the National Institute for Educational Studies and Research Anísio Teixeira (INEP) for the initial and final years of Elementary Education, which are published annually, with the exception of the Basic Education Development Development Index (IDEB), which is released every two years.

Data related to enrollment numbers, teachers, establishment and classes were extracted from the Statistical Synopses of Basic Education. The indicators, on the other hand, were collected from the Report on Educational Indicators. The average was calculated across the years to obtain information for elementary education as a whole. The Basic Education Development Index (IDEB), in turn, was extracted from a specific report. It is worth noting that IDEB measures the performance of the Brazilian education system for elementary and secondary education. Given that elementary education is a municipal competence and responsibility, only expenses paid in this subfunction were considered for this area.

In the case of Health, the data of expenses was related to basic care, hospital and outpatient assistance, prophylactic and therapeutic support, sanitary and epidemiological surveillance, among others.

For the operational data, the Single Health System Performance Index (Índice de Desempenho do Sistema Único de Saúde, IDSUS) were collected, provided by the Ministry of Health, whose objective is to evaluate the performance, public access to the Unified Health System (Sistema Único de Saúde, SUS), as well as the effectiveness of care related to primary care, outpatient/hospital care, urgencies and emergencies in each municipality, state, region and nation sphere (Ministry of Health, 2014). The IDSUS divides municipalities into Homogeneous Groups (HG), based on: (1) the Socioeconomic Development Index (Índice de Desenvolvimento Socioeconômico, IDSE) - which associates per capita GDP/PIB with the percentage of families in the Bolsa Família program; (2) the Health Conditions Index (Índice de Condições de Saúde, ICS) that considers indicators such as the infant mortality rate; and (3) the Municipal Health System Structure Index (Índice de Estrutura do Sistema de Saúde do Município, IESSM), which focuses on the specialized outpatient and hospital infrastructure of the municipality (Portulhak, Raffaeli and Scarpin, 2018). Currently, there are 6 HGs, as shown in Figure 2.

HG	IDSE	ICS	IESSM	Number of Municipalities
6	Low	Low	No MHC Structure(*)	2,183
5	Medium	Medium	No MHC Structure	2,038
4	Low	Low	Little MHC Structure	587
3	Medium	Medium	Little MHC Structure	632
2	High	Medium	Average MHC Structure	94
1	High	Medium	Average MHC Structure	29

Source: CGMA/Demas/SE/MS
(*) MHC Structure (Medium and High Complexity) refers to the structure of high and medium complexities or specialized outpatient and hospital care structure (emergency).

Figure 2. IDSUS Homogeneous Groups, 2019

Other data was collected from various systems that make up DATASUS, such as the Hospital Information System (Sistema de Informações Hospitalares, SIH), the Mortality Information System (Sistema de Informações sobre Mortalidade, SIM), the National Immunization Program (Programa Nacional de Imunizações, PNI), and the National Register of Health Establishments in Brazil (Cadastro Nacional de Estabelecimentos de Saúde, CNES).

The security function included expenses related to policing, civil defense, information and intelligence. It was possible to collect data on the number of homicides in each municipality. This information was obtained through the Violence Atlas, which is carried out by the Institute of Economic Research (Instituto de Pesquisas Econômicas Aplicadas, IPEA) in collaboration with the Brazilian Public Security Forum.

For sanitation, expenditures cover subfunctions related to urban and rural services. The data on sanitation were collected from the National Information System of Sanitation (Sistema Nacional de Informações sobre Saneamento, SNIS). The total population served with water supply and sewage corresponds to the number of people effectively served with services in urban and non-urban areas (SNIS, 2019).

Finally, social assistance includes expenses related to assistance for de elderly, people with disabilities, children and adolescents, among others.

For fiscal indicators, data related to Net Consolidated Debt (Dívida Consolidada Líquida, DCL), Total Personnel Expenditure (Despesa Total com Pessoal, DTP) and Net Current Revenue (Receita Corrente Líquida, RCL) of municipalities were collected for the year of 2017, in the 3rd quadrimester, which reflects the entire financial year's situation. This collection was carried out based on Fiscal Management Reports (Relatórios de Gestão Fiscal, RGF) made available by the FINBRA system of the National Treasury Secretariat (Secretaria de Tesouro Nacional, STN).

In addition to these areas, data on development were collected. Regarding the measures of municipal development, data were collected from two different sources: (1) IDHm, made available by the Human Development Atlas, carried out by the United Nations Development Programme (UNDP) in partnership with the João Pinheiro Foundation (FJP) and the Institute of Applied Economic Research (IPEA), composed of three indicators: longevity, education and income - which measure citizens' opportunities to lead a long and healthy life, have access to knowledge, and conditions to enjoy a dignified standard of living (UNDP, 2013); and (2) IFDM, published by the Federal of Industries of Rio de Janeiro (FIRJAN), which includes three areas of human development: employment and income, education and health.

The integration of data was done using the municipal codes provided by the IBGE, present in all the databases that were collected. However, it should be noted that, even with this standardization, some sources alter these codes, which ended up complicating the integration, such as DATASUS and SNIS, whose codes excuse some digits considered by IBGE. Additionally, it is worth mentioning that some differences in the spellings of municipality names also made integration challenging. These conflicts were resolved through the analysis of samples from various datasets, where it was possible to perceive the same cutting patterns of the IBGE digits (example: cutting of the first two digits that are equivalent to the state where the municipality is located) as well as the same spelling errors, which were corrected through research in the city halls of each municipality.

3.1 Data Transformation: Formulation of Indicators

Considering that the data was collected from multiple sources, it was possible to observe that some data were already presented in the form of indicators, with a predefined calculation methodology and precalculated values. An example is the IDEB, whose scores range from 0 to 10, with 6 considered the desired average for the federal government.

For data that were not collected in the form of indicators, a transformation process was necessary to eliminate the population size effect of municipalities and normalize their values to assume the format of indicators. This process was applied, for instance, to data related to expenses by function. To turn them into indicators, per capita values were calculated, meaning the investment value in the area was divided by the number of citizens in the municipality. For the education sector, the number of students in elementary school was considered.

For IBGE data, it was not necessary to formulate indicators. Firstly, for the mayor's party chance indicator, data related to the party by which the mayor was elected in the years 2017 and 2017 (when municipal elections took place) were collected. Then, a classification work was done regarding the party chance in the considered years - YES/NO. Regarding the data on the existence of the municipality's development plan, the data format was already binary, meaning YES/NO.

For expenses by function, per capita indicators were calculated, i.e., the investment value in the area per citizen.

Figure 3 shows the existing and created indicators.

Indicator	Reference
Expense per student in elementary education *(DespEdu)*	Savian and Bezerra (2013)
	Machado, Irffi and Benegas (2011)
Basic Education Development Development Index (*IDEB*)	INEP/IDEB
Students per class *(ATEdu)*	INEP/Dictionary of Educational Indicators
Daily class hours *(HAEdu)*	
Age-grade distortion *(DISEdu)*	
Teachers with higher education degree *(DCSEdu)*	
Approval rate *(TApEdu)*	
Reproval rate *(TREdu)*	
Dropout rate *(TAbEdu)*	
Per capita health expenditure *(DespSaude)*	Machado, Irffi and Benegas (2011)
Single Health System Performance Index (*IDSUS*)	Ministry of Health (2014)
	Almeida, Vasconcelos, Miranda and Feitosa (2017)
	Portulhak, Raffaeli and Scarpin (2018)
Hospitalizations per 100,000 inhabitants *(ItSaude)*	Adapted from Politelo, Rigo and Hein (2014)
Average hospital stay *(MPHSaude)*	Suggested by the authors
Outpatient clinics per 100,000 inhabitants *(AbSaude)*	
Deaths from avoidable causes *(OCESaude)*	
Vaccination doses applied *(DAVSaude)*	
Per capita expenditure on Security *(DespSeg)*	
Homicides per 100,000 inhabitants *(HomSeg)*	Adapted from the Violence Atlas
Total Personnel Expenditure *(DTP)*	Fiscal Responsibility Law (Complementary Law n. 101/2000), article 2
Net Consolidated Debt *(DCL)*	Senate Federal Resolution n. 40/2001
Human Development Atlas for municipalities (*IDHm*)	Human Development Atlas
FIRJAN Municipal Development Index (*IFDM*)	Federation of Industries of Rio de Janeiro (FIRJAN)
Per capita expenditure on sanitation*(DespSaneam)*	Machado, Irffi and Benegas (2011)
Population served with water supply *(PAASaneam)*	Suggested by the authors
Population served with sewage *(PESSaneam)*	
Per capita expenditure on social assistance *(DespASoc)*	Machado, Irffi and Benegas (2011)

Source: Davis and Souza (2021)

Figure 3. Indicators considered in the database, 2021

3.2 Data Cleaning and Missing Values

After data collection and transformation, it was necessary to perform data cleaning, which included checking for and detecting inconsistencies. Such issues were expected considering the multiple sources that make up the dataset.

According to Han and Kamber (2006), noisy data can be handled in various ways, such as interpolation, clustering, regression, among others. For the purposes of the database presented in this work, the decision was made to exclude inconsistent data since, given the heterogeneity of municipalities in Brazil, the use of inference or regression methods would not be appropriate.

Inconsistent data were identified through dispensation analyses in the R statistical software. Values that were significantly different from the others were analyzed and excluded, such as per capita expenses with unreal values, for example less than R$ 1.00 per student in elementary school.

In this context, it is worth mentioning the inconsistencies found in public expenses by function, total personnel expenditure (DTP) and net consolidated debt (DCL). Further investigation revealed that such data are declared by the municipalities themselves to the National Treasury Secretariat (STN). It was not possible to verify the existence of checks at the time of constructing the dataset.

Additionally, it is noted that the attributes of TPE and NCD have a significant amount of missing data. This situation can be explained by Article 63 of the Fiscal Responsibility Law (Complementary Law n. 101/2000), which allows municipalities with a population of fewer than 50,000 inhabitants to submit fiscal reports to the STN semi-annually, covering 4,905 municipalities. Of this group, 2,780 have no data on DTP and 3,793 have no data on DCL.

When considering all municipalities, 2,828 have no DTP data and 4,071 have no DCL data. Therefore, municipalities with fewer than 50,000 inhabitants represent 98.30% of missing DTP data and 93.17% of missing DCL data.

Regarding other municipalities with missing data and the identified inconsistencies, it is important to note that on September 20, 2019, the STN published an orientation (Portaria n. 642) which establishes rules for receiving and making financial and accounting data available in the information system used. Among the determinations, Chapter III includes stages related to data validation and verification. Therefore, in future updates to the database, an increase in the availability and quality of fiscal data from municipalities is expected.

4 Applications

The database treated in this work was analyzed in the works of Davis (2019) and Davis and Souza (2021) in order to verify patterns and association rules among municipalities that showed good results in education, health, human development and fiscal status indicators.

Davis (2019) and Davis and Souza (2021), using dataset and employing decision trees and association rules, demonstrated the importance of investing in teacher's higher education degrees and training. They also highlighted a strong relationship between education and health, meaning municipalities with good educational results tend to have better health outcomes.

For this purpose, they performed a reduction in the available indicators in the database. This reduction was carried out through correlation analyses, quantified using the Pearson correlation coefficient, to investigate the relationship between indicators and to avoid redundancies that could harm the analysis. Thus, indicators such as approval rate, failure rate and dropout rate were removed (IDEB was retained due to its more widespread use). Additionally, IFDM was excluded in favor of keeping IDHm, considering the greater availability of data.

Moreover, the results obtained indicated that expenditures exceeding the legal limits can negatively impact the performance of municipalities in education and health. Regarding the database, it is essential to note that the indicators underwent a verification process to remove redundant information that could compromise the analysis.

After removing some indicators, the remaining ones underwent discretization (binning) with the aim of converting numerical attributes into categorical ones (Zaki and Meira Júnior, 2014). As the database contains some noise in the data, this process helped reduce the number of distinct values for the same attribute, allowing for the disregard of small and irrelevant differences in values. It was also possible to categorize the data in most areas included in the database. Each indicator was individually analyzed based on its descriptive characteristics.

After this process of discretization of attributes in the database, three data mining algorithms were used: (1) InfoGainAttributeEval, which allowed identifying attributes with a greater capacity to explain variations in classes and classifying them in ascending order through the ranker analysis method based on a reference indicator; (2) J48, which was used to generate decision trees by examining the attributes that best distinguish the values of the reference indicator; and (3) Apriori for association rule detection. The authors obtained relevant results for the areas of education, health, development and fiscal indicators.

In the field of education, in which IDEB was chosen as the reference variable, it was observed, through the application of the algorithms, that the attribute that best explains variations in IDEB results was IDSUS, followed by the region of the municipalities and IDHm. This result demonstrates the relationship between education and health. Furthermore, the education variables most related to IDEB were the number of students per class and the percentage of teachers with a college degree. This result was corroborated by association rules that showed municipalities with good IDEB results and high percentage of teachers with a college degree tend to have a higher IDHm.

In the health sector, IDSUS was chosen as the reference variable. The attributes that best explain variations in the results of this indicator were the region of the municipality, per capita health expenditure and IDEB. In Brazil, capitals tend to have reference centers for more complex care, which often results in higher per capita spending. It was found that there is a relationship between health outcomes and the volume of resources applied.

For the development area, IDHm was used as the reference variable. The indicators that best explain variations in IDHm results are IDSUS, the region of the municipalities and the percentage of teachers with a college degree. The composition of the IDHm indicator itself corroborates this result. Association rules showed that the results can be predicted based on the results of IDEB and IDSUS.

Regarding fiscal indicators, only personnel expenses presented significant results. The algorithms showed that groups of municipalities with better performance in this indicator had the best results in both IDEB and IDSUS. This shows that excessive spending on personnel can harm the performance of municipalities in the priority areas of government action. As for the net consolidated debt, it was not possible to obtain significant results, and it was observed that its composition does not explicitly include any of the areas analyzed.

Regarding the areas of security and sanitation, it was not possible to obtain relevant results. In the case of public security, a lack of cohesion among indicators was observed, explained by the role of municipalities in public security. Among the institutions that implement public security policies, three are organized and maintained by the federal government (federal police, federal highway police and federal railway police) and two by the states and federal district governments (civil police, military police and fire brigade), as established in article 144 of the Constitution of the Republic of 1988. Separately, article 8 deals with municipalities and establishes that they may have their own municipal guards. According to Ricardo and Caruzzo (2007), municipal guards are located in municipalities with populations between 100,000 and 500,000 inhabitants.

Regarding sanitation, it was found that municipalities have a poorly defined role in the system. According to the National Water Agency (Agência Nacional de Águas, ANA, 2017), the municipality can organize sanitation services indirectly (through delegation of provision to municipal autarchy, state company or private concessionaire) or directly (without an institutionalized service provider). In addition, the Constitution of the Republic of 1988 establishes, through article 21, the competence of the federal government to establish guidelines for basic sanitation.

In summary, the organization of the database and the results obtained through data mining algorithms allowed the identification of groups of municipalities with the best performance in socio-economic indicators and their common characteristics. The formation of groups of municipalities enabled a comparative geographical analysis of municipalities considered efficient and inefficient.

Efficient municipalities, as understood by Davis (2019) and Davis and Souza (2021), are those that showed the best relationship between the volume of expenses and results in socioeconomic indicators, meaning lower expenses and better results compared to other municipalities. Inefficient municipalities, on the other hand, are those that exhibited the highest expenditures and the worst performances in socioeconomic indicators. Figures 4, 5 and 6 show the municipalities that were part of the main groups of efficient and inefficient municipalities in the areas of education, health and development. Other municipalities did not have notable results for any of the groups.

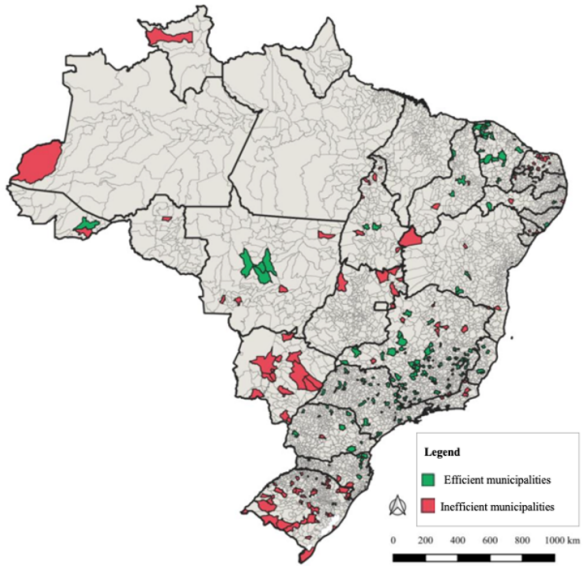

Figure 4. Efficient and inefficient municipalities in resource allocation for education

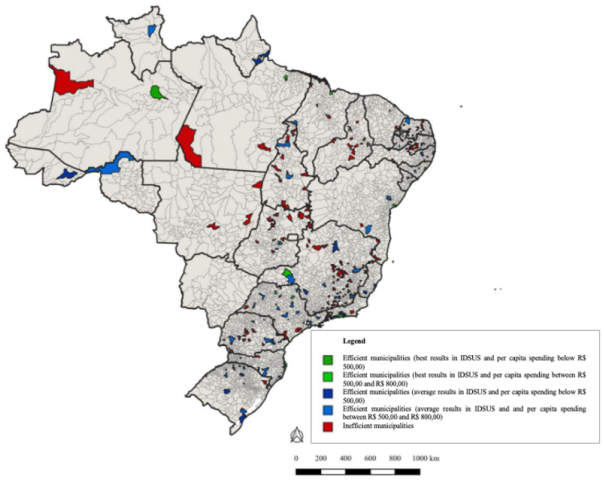

Figure 5. Efficient and inefficient municipalities in resource allocation for health

The visualization of the results from the data mining techniques applied to the database in this study was done comparatively to identify any region or state that stands out positively or negatively. In the example of Figures 4, 5 and 6 it is possible to notice a concentration of cities considered efficient in the states of Minas Gerais, São Paulo and Ceará.

The mentioned studies used data mining techniques for knowledge extraction in the dataset. However, it is possible to employ various other techniques that may yield different relevant results, considering the multidisciplinary nature of the constructed database. It is also worth noting that the future update of indicators, as they are released, will allow obtaining more important results. For example, the expected release of the 2022 census will enable the use of more current data for the IDHm. The last publication of this indicator was in 2010.

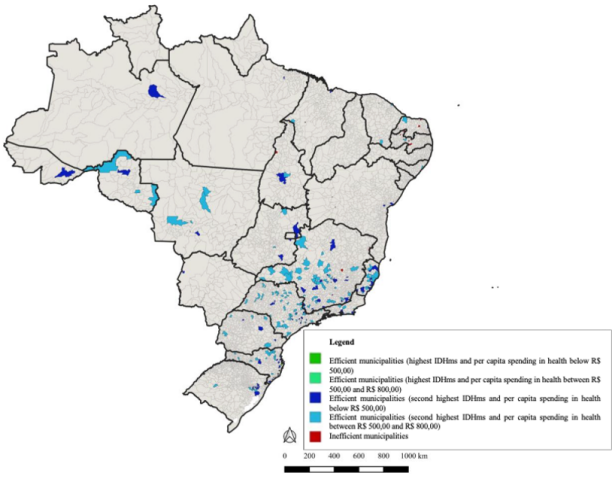

Figure 6. Efficient and inefficient municipalities in resource allocation for development

5 Challenges and Limitations

The main limitation of the presented database is the quality of public data, as each piece of data is presented differently in government sources. Issues related to fixed publication periodicities should also be considered as a constraint.

Furthermore, some inconsistencies were identified in data related to public expenses by function, which are declared by municipalities to the National Treasury Secretariat (STN). It was not possible to identify any type of verification of accuracy.

Some datasets, such as those related to sanitation, could provide better details about the data since broad access may raise questions or even lead to the misuse of data due to lack of understanding.

The main problem with the data sources was the absence of some attributes for many municipalities, some of which are important and large. Porto Alegre (RS) is an example: the 2017 IDEB is not available for the twelfth-largest Brazilian city.

Finally, it is worth mentioning the currentness of the database. As mentioned earlier, the data sources have varied publication practices and intervals, which makes it difficult to collect a significant volume of data for the same year. Some sources, such as IBGE and INEP, update their data annually or biannually. However, for some indicators, such as IDSUS and IDHm, updates have not been issued since 2011 and 2010, respectively. For future work, updates to each part of the database should take place as new data are made available, and provisions should be made in analyses to consider the time difference between sources.

6 Conclusions

The database constructed and analyzed in this study had the goal of combining financial data (public expenses by function) with operational data from municipalities. Data available from public information sources were collected and then underwent processes of integration and transformation. In this process, some weaknesses were observed, as discussed in Section 5.

The database in this study included general indicators - collected from surveys provided by IBGE - and those related to education, health, public security, development, sanitation and fiscal situation. For all these areas, except for development, per capita indicators were formulated. Or, in the case of education, per student indicators were formulated. For the social assistance area, only the expense indicator was considered, as it was not possible to identify a specific function indicator.

In public administration, the government's areas of activity should not be analyzed in isolation. To formulate efficient public policies, decision-makers need knowledge about various areas, as well as the relationship of one area to another, in order to generate the best results for the population.

It is worth noting that it is important to increase the frequency of updating certain databases, such as the one related to IDSUS. Other databases, like SNIS, could provide better details about the meaning of the data. There is also a need for the establishment of historical series over time that may allow, in the future, the conduct of predictive analyses (Davis and Souza, 2021). Despite the limitations, the database presented in this study proves to be relevant since it allows for various other types of analyses and comparative studies between municipalities, states and regions.

7 References

Almeida, J.E.G., Vasconcelos, M.T.C., Miranda, L.C., Feitosa, M.O. (2017) "Gasto público e sua influência no Índice de Desempenho do Sistema Único de Saúde dos municípios pernambucanos" Anais do XXIV Congresso Brasileiro De Custos.

Bovo, J. M. (2001) "Gastos Sociais dos Municípios e Desequilíbrio Financeiro", Revista de Administração Pública. Rio de Janeiro, 35(1), p. 93-117.

Cruz, C. F. and Afonso, L.E. (2018) "Gestão fiscal e pilares da Lei de Responsabilidade Fiscal: evidências em grandes municípios", Revista de Administração Pública, 52(1), 126-148.

Davis, P.G. (2019), "Análise comparativa da eficiência alocativa das despesas públicas dos municípios brasileiros utilizando técnicas de mineração de dados", Dissertação de Mestrado em Administração, UFMG.

Davis, P.G. and Souza, A.A. (2021), "Análise comparativa da eficiência alocativa das despesas públicas municipais utilizando mineração de dados", Administração Pública E Gestão Social, 13(4). Han, J. and Kamber, M. (2006), Data Mining: Concepts and Techniques, 2nd.

Lourenço, R. L., Angotti, M., Nascimento, J.C.H.B., Sauerbronn, F. F. (2017) "Eficiência do gasto público com ensino fundamental: uma análise dos 250 maiores municípios brasileiros", Revista Contabilidade Vista & Revista, 28(1), 89-116.

Machado, S.P., Irffi, G. and Benegas, M. (2011) "Análise da eficiência técnica dos gastos com educação, saúde e assistência social dos municípios cearenses", Planejamento e Políticas Públicas, 1(36), p. 87-113.

Matias-Pereira, J. (2017), Finanças Públicas, Atlas, 7 ed.

Ministry of Health (2014), "Índice de Desempenho do Sistema Único de Saúde", http://idsus.saude.gov.br/documentos.html, Setembro 2019.

Neduziak, L.C.R. and Correia, F.M. (2017) "Alocação dos gastos públicos e crescimento econômico: um estudo em painel para os estados brasileiros", Revista de Administração Pública, 4(51), p. 616-632.

Politelo, L., Rigo, V.P. and Hein, N. (2014) "Eficiência da aplicação de recursos no atendimento do Sistema Único de Saúde (SUS) nas cidades de Santa Catarina" Revista de Gestão em Sistemas de Saúde, 3(2), 45-60.

Portulhak, H., Raffaelli, S.C.D., Scarpin, J.E. (2018) "A Eficiência da Aplicação de Recursos Voltada à Saúde Pública nos Municípios Brasileiros", Revista Contabilidade, Gestão e Governança, 21(1), 21-39.

PNUD - United Nations Development Programme (2013), O Índice de Desenvolvimento Humano Municipal Brasileiro: Atlas do Desenvolvimento Humano do Brasil.

Resende, J. P. and Andrade, M. V. (2011) "Crime Social, Castigo Social: Desigualdade de Renda e Taxas de Criminalidade nos Grandes Municípios Brasileiros", Estudos Econômicos, 41(1), p. 173-195.

Savian, M.P.G. and Bezerra, F.M. (2013) "Análise de eficiência dos gastos públicos com educação no ensino fundamental no estado do Paraná", Economia & Região, 1(1), 26-47.

Silva, M. S. (2005) "Teoria do federalismo fiscal: notas sobre as contribuições de Oates, Musgrave, Shah e Ter-Minassian", Nova Economia, 15(1), p. 117-137.

Silva, M.R.C.S., Costa, E.P.A., Castro, M.A.R.C. and Silva, L.P. (2016) "O Efeito dos Gastos Públicos no Nível de Desenvolvimento de Municípios no Brasil: os Gastos Municipais Importam?", Anais VII Encontro de Administração Pública e Governança.

SNIS - Sistema Nacional de Informações Sobre Saneamento (2019). Glossário de Informações - Água e Esgotos, http://snis.gov.br/glossarios, Setembro de 2019.