

Mineração de Dados sobre Despesas Públicas de Municípios Mineiros para Gerar Alertas de Fraudes

Larissa D. Gomide¹, Guilherme Bezerra dos Santos¹, Lucas L. Costa¹,
Michele A. Brandão^{1,2}, Anísio Lacerda¹, Gisele L. Pappa¹

¹Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG, Brasil

²Instituto Federal de Minas Gerais (IFMG) – Ribeirão das Neves, MG, Brasil

{larissa.gomide, guilhermebezerra}@dcc.ufmg.br, lucas-lage@ufmg.br
michele.brandao@ifmg.edu.br, {anísio, glpappa}@dcc.ufmg.br

Abstract. *This paper proposes a methodology to compare expenses between municipalities and identify possible frauds in the cities' budgets in Minas Gerais. Using a public and government database, the methodology uses data mining techniques to group municipalities and rank expenditures based on population and micro-region. The results reveal anomalies in expenses, indicating possible fraud.*

Resumo. *Este artigo propõe uma metodologia para comparar despesas entre municípios e identificar possíveis fraudes no orçamento de cidades mineiras. Utilizando uma base de dados pública e governamental, a metodologia utiliza técnicas de mineração de dados para agrupar municípios e então ranquear as despesas com base na faixa populacional e microrregião. Os resultados revelam anomalias nas despesas, indicando possíveis fraudes.*

1. Introdução

A Lei de Acesso à Informação (Lei nº 12.527/11)¹ estabeleceu a obrigatoriedade dos poderes governamentais em publicar conjuntos de dados de interesse público para aumentar a transparência e participação civil na política. Esses dados são provenientes de diversas fontes e apresentam diferentes tipos e naturezas, o que dificulta a validação, consistência e interoperabilidade entre sistemas. Além disso, esses dados são volumosos e impactam na velocidade de processamento e extração de informações relevantes.

Nesse contexto, a principal contribuição deste trabalho é uma metodologia, na qual um município é escolhido para ser utilizado como referência e, a partir dessa escolha, são ranqueadas despesas de municípios com dados censitários similares (faixa populacional ou microrregião) e com valores muito discrepantes nas despesas. Vale destacar que os municípios são agrupados por esses dois aspectos para que sejam feitas análises considerando municípios de tamanho similares e municípios geograficamente próximos e, também, por esses aspectos serem tipos de agrupamentos comumente feitos pelo IBGE e essas informações estarem presentes nos dados considerados. Finalmente, é possível gerar alertas sobre a ocorrência de alguma fraude nas despesas de um município.

Trabalhos Relacionados. Existem poucos estudos sobre a análise de despesas públicas como indicativo de fraude. Alguns exemplos são os artigos de [dos Santos et al. 2014] e

¹Lei nº 12.527/11: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

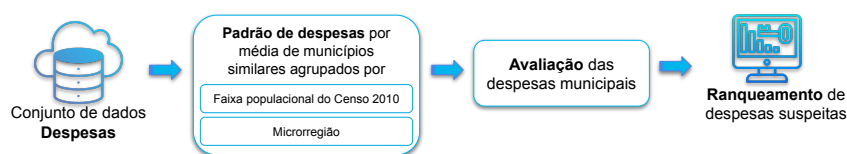


Figura 1. Principais etapas para realização deste trabalho.

[Galante et al. 2015], cujo objetivo é apresentar como a Lei de Responsabilidade Fiscal (Lei nº 101/00) impacta a análise e a transparência das contas públicas. Por outro lado, [Mangravite et al. 2022], [Oliveira et al. 2022] e [Oliveira and Silveira 2018] destacam a importância de se realizar um pré-processamento dos dados selecionados. Levando, portanto, em consideração problemas de inconsistência, falta de padronização e as características de um conjunto de dados volumoso. Além disso, o artigo de [Costa et al. 2022] aborda uma estratégia para identificar indícios de fraudes em licitações, apesar de não explorar como a análise de despesas pode ser utilizada como indicativo de irregularidades. Este é o ponto que este artigo aborda, combinando ideias como o agrupamento de cidades com base em dados censitários [dos Santos et al. 2014] e uma metodologia estruturada com etapas como filtragem dos dados, pré-processamento e cálculos estatísticos para identificar indicadores de fraudes, inspirados no trabalho de [Costa et al. 2022].

2. Metodologia

Esta pesquisa tem caráter exploratório e, dado um conjunto de dados como entrada, foram realizadas três etapas principais para o levantamento de alertas de fraude sobre despesas de municípios mineiros. A Figura 1 sumariza essa metodologia e cada etapa está descrita conforme segue.

Descrição do conjunto de dados. Os dados selecionados neste artigo vêm do Sistema Informatizado de Contas dos Municípios (SICOM),². Nele, estão compiladas informações referentes aos 853 municípios do estado de Minas Gerais, as quais foram organizadas pelo Tribunal de Contas do Estado de Minas Gerais (TCE-MG). Dentre as informações contidas na base de dados, há 40.466.692 despesas, com descrição de 29 funções da dotação orçamentária. É importante destacar que a função é o maior nível de agregação das diversas áreas de atuação do setor público. A função se relaciona com a missão institucional do órgão como, por exemplo, educação, saúde, cultura, transporte, direitos da cidadania, etc. Já a dotação orçamentária é a união de quatro classificações que organizam as despesas públicas, são elas: institucional, funcional, de natureza da despesa e por fonte de recursos. Neste trabalho, a classificação por função da dotação orçamentária é considerada nas análises de despesas dos municípios. Essa função é também chamada de tipo de despesa.

Identificação do padrão de despesa. Nesta etapa, foram utilizados dois aspectos para identificar um padrão de despesa dos municípios, são eles: (i) faixa populacional do Censo 2010³; e (ii) microrregião (conjunto de municípios contíguos com similaridades econômicas e sociais⁴). Ao considerar a faixa populacional censitária, os municípios são

²SICOM: <https://portalsicom1.tce.mg.gov.br/>

³Censo 2010 - IBGE: <https://censo2010.ibge.gov.br/>

⁴Definição de microrregiões: <https://metadados.snirh.gov.br/geonetwork/srv/api/records/e6dd026c-afa7-4a7c-8904-abb86662da5>

Algoritmo 1 Avaliação comparativa de despesas.

Entrada: dados do SICOM, $result \leftarrow []$;
Saída: $result$ com despesas ordenada por diff;
início

```

para cada  $m$  in  $todos\_municipios$  faça
   $t \leftarrow$  todos os municípios similares a  $m$ :
  para cada  $f$  in  $todas\_funções$  faça
    para cada  $ano$  in  $todos\_anos$  faça
      para cada  $mes$  in  $todos\_meses$  faça
         $media \leftarrow$  média de despesas para função  $f$  em  $ano/mes$  de todos os  $t$ ;
         $desvio \leftarrow$  desvio de despesas para função  $f$  em  $ano/mes$  de todos os  $t$ ;
         $soma \leftarrow$  soma de despesas para função  $f$  em  $ano/mes$  do município  $m$ ;
         $diff \leftarrow$  soma - (média + 2 * desvio);
         $result += [m, f, ano, mes, soma, (media + 2 * desvio), diff]$ ;
      fim
    fim
  fim
fim

```

agrupados pelo seguinte critério: até 2.000 pessoas; de 2.001 a 5.000 pessoas; de 5.001 a 10.000 pessoas; de 10.001 a 20.000 pessoas; de 20.001 a 50.000 pessoas; de 50.001 a 100.000 pessoas; de 100.001 a 500.000 pessoas; e mais de 500.000 pessoas. Já no agrupamento por microrregião, são considerados similares os municípios que pertencem a uma mesma microrregião. Assim, é calculada a média de despesa por grupo de município.

Avaliação comparativa das despesas municipais. Nesta etapa, cada município da base de dados tem as despesas comparadas com a média das despesas dos dois grupos ao qual faz parte, um por faixa populacional e outro por microrregião. Essa comparação também considera o mês e o ano de exercício e a função da dotação orçamentária, conforme mostra o Algoritmo 1. Note que, os dados do SICOM são utilizados como entrada e, a partir de instâncias dessa entrada, realiza-se as operações: para cada município presente na lista, realiza-se uma comparação com os demais procurando àqueles que apresentam uma similaridade (segundo critérios censitários e de microrregião) com o município selecionado. Considerando os municípios similares, é feita uma avaliação comparativa das despesas que levam como base informações temporais e o cálculo da média mais duas vezes o desvio padrão, ou seja, o percentil 95 das médias. As despesas com menor diferença (representada por *diff* no algoritmo) entre o valor total dela e o percentil 95 são consideradas menos suspeitas de fraude do que as despesas com maior valor para essa diferença.

Ranqueamento das despesas suspeitas. O critério utilizado para definir se a despesa de um município está fora do padrão, ou seja, se é uma anomalia, é se a despesa deste município for maior que a média mais desvio padrão das despesas dos municípios similares.

3. Resultado

Esta seção apresenta uma caracterização e ranqueamento das despesas municipais.

3.1. Caracterização das Despesas Municipais

As Figuras 2 (a) e (b) mostram características das maiores despesas municipais. Em particular, a Figura 2(a) apresenta os Top 10 municípios mineiros com maiores despesas no período de 2014 até 2022. Por se tratar de dados volumosos, as despesas foram ordenadas pelo valor total delas e, então, foram selecionadas as maiores até completar 10 municípios.

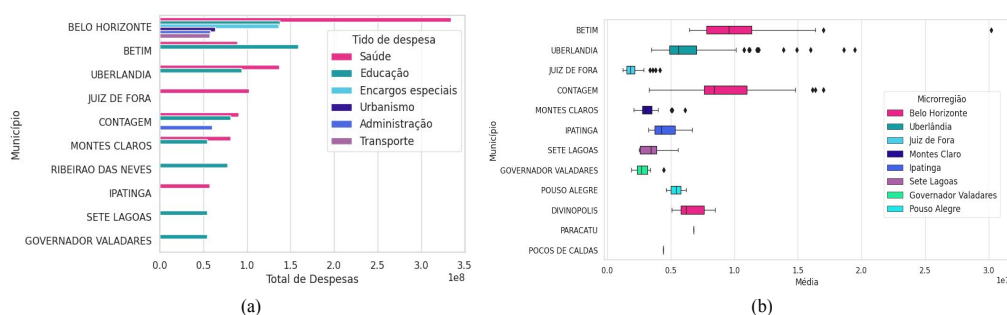


Figura 2. Despesas municipais no período de 2014 até 2022: (a) Despesa máxima dos Top 10 municípios por função (tipo de despesa); e (b) Boxplot da média dos agrupamentos por microrregião das Top 500 despesas.

O gráfico apresenta apenas a despesa máxima do município por tipo de despesa (função da dotação orçamentária) no período considerado. Isso resultou em 20 despesas.

É possível observar que o total de despesa está em notação científica e os valores são da ordem de 100 milhões. Por ser a capital mineira, Belo Horizonte se destaca em quantidade de despesas nas seis funções (saúde, educação, encargos especiais, urbanismo, administração e transporte). Já Betim, Uberlândia e Montes Claros possuem mais gastos com saúde e educação. De forma complementar, Juiz de Fora, Ribeirão das Neves, Ipatinga, Sete Lagoas e Governador Valadares tiveram gastos significativos com saúde ou educação. Finalmente, Contagem além de ter despesas altas para esses dois tipos de despesa, também teve destaque para gastos com a administração. É importante destacar que a ausência de despesas retratadas no gráfico para uma função não significa que tal município não teve esse tipo de despesa. Apenas indica que essa despesa não foi tão alta quanto as apresentadas no gráfico.

Além disso, a Figura 2(b) mostra boxplots das médias das despesas em cada mês e ano de exercício do grupo de municípios definido pela microrregião ao qual faz parte. Para elaborar esta visualização calculou-se a diferença entre o número total e o percentil 95 das despesas, definindo, assim, este cálculo como o critério utilizado para ranquear as despesas, e são filtradas as top 500 despesas com maiores diferenças. Entretanto, Belo Horizonte foi desconsiderada nesse gráfico já que é a capital e possui maiores valores de despesas. Os valores estão em notação científica e na ordem de 10 milhões.

Dentre as análises possíveis na Figura 2(b), nota-se que as cidades com maiores valores de média das despesas (Betim, Contagem e Divinópolis) fazem parte da microrregião de Belo Horizonte. Uma hipótese para isso é Belo Horizonte estar aumentando as médias desse grupo. Outra cidade que chama a atenção é Uberlândia, é a cidade referência de uma microrregião possui valores médios de despesa próximos à capital mineira.

Por fim, ao comparar os resultados das Figuras 2 (a) e (b), observa-se que Ribeirão das Neves aparece no Top 10 municípios com maior total de despesa, mas não aparece no top de municípios com maiores diferença em relação ao grupo de microrregião ao qual faz parte. Esse município faz parte da microrregião de Belo Horizonte, então, em análises futuras, é também importante analisar mais detalhadamente o total de despesa que destoa em cada mês e ano de exercício.

Tabela 1. Top 5 ranqueamento das despesas agrupadas por município, ano, mês e função da dotação orçamentária.

Faixa Populacional	Mês/Ano Exercício	Tipo de Despesa	Município	Total de Despesas	Diferença
Qtd população <= 2.000	07/2020	Saneamento	Pedro Teixeira	931.296,54	339.937,15
	12/2020	Urbanismo	Tapirai	1.203.664,21	330.345,88
Qtd municípios = 19	03/2022	Transporte	Antônio Prado de Minas	708.011,18	226.286,96
	10/2020	Urbanismo	Grupiara	773.423,30	226.010,32
2.001 <=	07/2021	Transporte	Consolação	663.387,28	221.544,11
	06/2020	Assistência Social	Crucilândia	2.525.614,58	2.088.538,29
Qtd população <= 5.000	12/2016	Transporte	Tapira	2.481.117,61	2.023.321,18
	07/2021	Assistência Social	Crucilândia	2.409.492,23	1.977.928,59
Qtd municípios = 221	11/2017	Educação	Catas Altas	2.537.148,66	1.851.595,31
	12/2021	Encargos Especiais	Jaguarapu	2.223.989,19	1.798.249,27
5.001 <=	12/2020	Urbanismo	São Gonçalo do Rio Abaixo	13.288.105,84	10.910.143,51
	09/2014	Urbanismo	São Gonçalo do Rio Abaixo	6.623.690,69	5.592.401,08
Qtd população <= 10.000	10/2014	Urbanismo	São Gonçalo do Rio Abaixo	6.475.999,55	5.464.800,19
	07/2014	Urbanismo	São Gonçalo do Rio Abaixo	6.335.727,34	5.348.259,18
Qtd municípios = 251	10/2020	Administração	Arapora	6.665.708,36	5.275.249,97
10.001 <=	04/2022	Urbanismo	Conceição do Mato Dentro	16.237.252,43	13.124.670,92
	12/2021	Urbanismo	Conceição do Mato Dentro	14.247.290,47	11.236.866,49
Qtd população <= 20.000	09/2021	Urbanismo	Conceição do Mato Dentro	11.767.178,43	9.328.280,02
	12/2020	Urbanismo	Conceição do Mato Dentro	11.552.119,81	8.501.593,32
Qtd municípios = 184	10/2021	Urbanismo	Conceição do Mato Dentro	10.436.350,64	8.244.566,10
20.001 <=	12/2020	Saúde	Brumadinho	22.493.748,99	12.857.609,68
	03/2022	Administração	Extrema	16.196.690,39	11.089.994,16
Qtd população <= 50.000	12/2021	Educação	Brumadinho	22.801.064,42	10.807.666,39
	12/2020	Educação	Brumadinho	19.062.965,03	10.778.377,23
Qtd municípios = 112	10/2015	Administração	Itapecerica	13.040.873,88	9.764.408,18
50.001 <=	02/2022	Administração	Mariana	22.944.404,24	11.921.942,11
	12/2021	Educação	Paracatu	35.234.472,62	11.589.238,67
Qtd população <= 100.000	04/2014	Educação	Nova Lima	19.218.791,45	9.884.554,17
	12/2018	Educação	Nova Lima	22.294.715,13	9.598.657,84
Qtd municípios = 37	03/2022	Educação	Araxá	22.598.654,63	8.706.786,25
100.001 <=	12/2021	Educação	Betim	159.123.026,44	55.169.082,25
	10/2020	Saúde	Betim	88.549.308,79	30.591.824,98
Qtd população <= 500.000	10/2019	Educação	Betim	60.099.003,51	24.148.239,81
	10/2020	Educação	Betim	59.370.783,71	23.372.414,22
Qtd municípios = 25	10/2019	Saúde	Betim	68.961.358,52	21.615.729,73

Tabela 2. Top 5 despesas agrupadas por microrregião.

Mês/Ano Exercício	Tipo de Despesa	Município	Total de Despesas	Média	# Despesas	Diferença	# Municípios
12/2020	Saúde	Juiz de Fora	101.940.327,84	4.150.549,23	31	61.477.550,89	33
10/2021	Saúde	Juiz de Fora	92.680.344,94	3.585.534,33	31	56.011.299,66	33
12/2021	Saúde	Juiz de Fora	90.823.777,88	3.819.927,69	31	54.691.369,90	33
07/2021	Saúde	Juiz de Fora	86.268.572,09	3.351.914,75	31	52.127.584,08	33
12/2021	Educação	Betim	159.123.026,44	30.195.098,72	17	46.979.474,46	24

3.2. Ranqueamento das Despesas Municipais

Esta seção apresenta os resultados da metodologia proposta neste trabalho para ranquear despesas suspeitas de fraude. A Tabela 1 apresenta as Top 5 despesas segundo o critério de ranqueamento e considera o critério censitário e o número de municípios que pertencem a cada faixa populacional. Além disso, também são detalhados o mês e o ano de exercício da despesa, a função da dotação orçamentária, o município ao qual se referem os dados. Nota-se que cidades até 2.000 pessoas e com mais de 500.000 pessoas possuem menor quantidade de municípios. Apesar disso, na faixa de 2.000 a 5.000 habitantes há mais cidades distintas suspeitas de fraude. Já nas faixas populacionais com mais habitantes, há uma recorrência das cidades, como por exemplo, Conceição do Mato, que está no top 5 e as despesas no topo do ranqueamento são em Urbanismo do ano de 2020 até 2022, reforçando o indício de fraude.

De forma complementar, a Tabela 2 apresenta as top 5 despesas ordenadas pelo

campo diferença e a diferença é calculado pela média das despesas do grupo de municípios que pertencem a mesma microrregião. Observa-se que Juiz de Fora aparece no topo para o tipo de despesa Saúde nos anos de 2020 e 2021. Já Betim aparece para o tipo de despesa Educação para o ano de 2021 também. Portanto, vale a pena análises mais detalhadas dessas despesas para verificar se houve alguma fraude.

4. Conclusão

Este trabalho apresentou uma metodologia, que utilizou técnicas de mineração de dados para ranquear despesas públicas de municípios mineiros de forma a auxiliar na detecção de fraudes. Vale destacar que essa metodologia utiliza duas abordagens de ranqueamento, uma agrupando os municípios por faixa populacional e outra por microrregião. Os resultados mostraram que a metodologia proposta pode auxiliar especialistas a filtrar despesas que devem realmente ser analisadas por ter valores suspeitos. Como trabalhos futuros, planeja-se melhor avaliar as duas abordagens de ranqueamento propostas para verificar a diferença entre elas e como podem ser utilizadas de forma agregada para melhorar os resultados. Assim, a metodologia proposta será melhor validada.

Agradecimentos. Ao Ministério Público do Estado de Minas Gerais pelo apoio através do Projeto Capacidades Analíticas. Ao CNPq, CAPES e FAPEMIG pelo apoio, e também pelo financiamento recebido pelo projeto da parceria entre AWS e CNPq.

Referências

- Costa, L., Reis, A., Bacha, C., Oliveira, G., Silva, M., Teixeira, M., Brandão, M., Lacerda, A., and Pappa, G. (2022). Alertas de fraude em licitações: Uma abordagem baseada em redes sociais. In *Anais do XI BraSNAM*, pages 37–48, Porto Alegre, Brasil. SBC.
- dos Santos, D. F., de Paula Bueno, F., Correa, M. D., Antonovz, T., and da Silva, W. V. (2014). Lei de responsabilidade fiscal: um estudo comparativo das despesas com pessoal em municípios paranaenses. *Revista da FAE*, 17(2):6–19.
- Galante, C., Debastiani, L. C., and Oenning, V. (2015). Evolução dos gastos públicos: análise dos municípios da microrregião de chapecó sc. In *Anais do Congresso Brasileiro de Custos-ABC*.
- Mangaravite, V., Carvalho, M., Cantelli, L., Ponce, L., Campoi, B., Nunes, G., Laender, A., and Gonçalves, M. (2022). Dedupegov: Uma plataforma para integração de grandes volumes de dados de pessoas físicas e jurídicas em Âmbito governamental. In *Anais do XXXVII SBBB*, pages 90–102, Porto Alegre, Brasil. SBC.
- Oliveira, E. F. and Silveira, M. S. (2018). Open government data in brazil a systematic review of its uses and issues. In *Procs. of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pages 1–9.
- Oliveira, G., Reis, A., Mendes, B., Bacha, C., Costa, L., Canguçu, G., Silva, M., Caetano, V., Brandão, M., Lacerda, A., and Pappa, G. (2022). Ferramentas open-source de qualidade de dados para licitações públicas: Uma análise comparativa. In *Anais do XXXVII SBBB*, pages 116–127, Porto Alegre, Brasil. SBC.