

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Carolina Moura de Bittencourt Fialho

Análise dos Gastos do Governo do Distrito Federal

**Curitiba
2020**

Carolina Moura de Bittencourt Fialho

Análise dos Gastos do Governo do Distrito Federal

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Eduardo Cunha de Almeida

Curitiba
2020

Análise dos Gastos do Governo do Distrito Federal

Carolina Moura de Bittencourt Fialho¹
Prof. Eduardo Cunha de Almeida²

Resumo

Os dispêndios públicos estão ganhando cada vez mais atenção e, por conseguinte, são questionados pelos contribuintes. Dessa forma, fez-se necessário cumprir o Princípio da Transparência previsto na Constituição de 1988, com o propósito de apresentar onde o dinheiro público está sendo investido. Diante da relevância do tema, o objetivo geral da pesquisa é compreender os gastos públicos do governo do Distrito Federal, através de comparação das receitas com as despesas nas áreas essenciais, bem como gastos com verbas indenizatórias dos servidores. Os resultados sugerem que as despesas com educação e saúde acompanharam a curva de receitas, o que já não ocorreu com segurança pública e administração, que se mantiveram praticamente estáveis ao longo do período. E que as despesas com as cotas diminuíram ao longo dos anos, exceto em 2016; gastos com diárias, passagens e publicidade também não acompanharam as curvas de receitas.

Palavras-chave: gastos públicos, receitas, despesas, transparência.

Abstract

Public spending is gaining even more attention and, therefore, is questioned by taxpayers. Thus, it became necessary to comply the Principle of Transparency provided in the 1988 Constitution with the purpose of presenting where public money is being invested. About the relevance of the theme, the general objective of the research is to understand the public expenditures of the Federal District government, by comparing the revenues with the expenditures in the essential areas, as well as expenses with indemnities from the civil servants. The results suggest that expenditure on education and health followed the revenue curve, which did not happen with public security and administration, which remained practically stable throughout the period. And quota expenses have decreased over the years, except in 2016; expenses with daily rates, tickets and advertising also did not follow the revenue curves.

Keywords: Public spending, revenues, expenditure, Transparency.

¹Aluno do programa de Especialização em Data Science & Big Data, carolfialho@yahoo.com.br.

²Professor do Departamento de Estatística - DEST/UFPR.

1 Introdução

Os gastos públicos são cada vez mais questionados pelos contribuintes. Grandes escândalos de corrupção no Brasil fizeram com que se cumprisse o Princípio da Transparência previsto na Constituição de 1988 com o objetivo de mostrar a toda a população exatamente onde o dinheiro público está sendo investido. Dessa forma, nos dias de hoje é obrigatório que qualquer cidadão brasileiro tenha acesso aos gastos realizados pelos gestores públicos.

1.1 Objetivo

O objetivo geral da pesquisa é compreender os gastos públicos do governo do Distrito Federal, através de comparação das receitas com as despesas nas áreas essenciais, como saúde, educação, segurança pública e administrativas, bem como gastos com verbas indenizatórias, a exemplo de diárias e hospedagens dos servidores. O desafio deste trabalho é verificar se o Governo do Distrito Federal está utilizando o dinheiro público com probidade e eficiência nas áreas essenciais. Extrairemos e trataremos os dados disponíveis em sites do GDF usando a linguagem Python, armazenaremos esses dados no elasticsearch e faremos todas as análises necessárias para a avaliação dos gastos públicos no Kibana.

2 Materiais e Métodos

O presente capítulo é dedicado à apresentação do tipo de estudo, as ferramentas adotadas, a base e tratamento de dados utilizados para identificar e extrair as informações sobre os gastos públicos realizados no governo do Distrito Federal.

2.1 Tipos de Estudo

Nesta pesquisa, optou-se pela coleta de dados secundários, ou seja, em base de dados existentes. Esse método é mais utilizado quando o objetivo é conhecer melhor o problema pesquisado, como uma forma de pesquisa exploratória inicial [6]. Dessa forma, tais dados foram obtidos por meio de documentações, website do governo e registros em arquivos disponibilizados publicamente. Para atingir os objetivos do estudo foi realizada uma

pesquisa descritiva que, de forma cuidadosa e precisa, apresenta as características de determinado fenômeno ou população [1, 10] e, embora sirva de base para explicar os fenômenos que descreve, não tem compromisso de explicá-los [11].

2.2 Ferramentas

Três principais ferramentas foram utilizadas para importar, armazenar e tratar os dados: Python, Elasticsearch e Kibana. Python é uma linguagem de programação que foi lançada em 1991 por Guido van Rossum e, de acordo com o website Python [9], é utilizada em muitos domínios de aplicação, a exemplo do desenvolvimento de web e internet, acesso a base de dados, desenvolvimento de software e jogos, entre outros. O Python é desenvolvido sob uma licença de código aberto aprovada pela Open Source Initiative (OSI), tornando-o livremente utilizável e distribuível, mesmo para uso comercial. Segundo o portal colaborativo da comunidade PyScience-Brasil [8], a linguagem Python suporta múltiplos paradigmas de programação e foi criada para produzir código bom e fácil de manter e dentre suas características, destaca-se o baixo uso de caracteres especiais, o que torna a linguagem parecida com pseudo-código executável; o uso de indentação para marcar blocos; pouco uso de palavras-chave voltadas para a compilação; coletor de lixo para gerenciar automaticamente o uso da memória; entre outros. Neste trabalho, o Python foi essencial para importar dados, tratá-los e armazenar no Elasticsearch. A segunda ferramenta utilizada foi o Elasticsearch. De acordo com o website Elastic [4], o Elasticsearch foi lançado em 2010 e é um mecanismo que permite busca e análise de dados distribuídos e open source para todos os tipos de dados, incluindo textuais, numéricos, geoespaciais, estruturados e não estruturados. O Elasticsearch vai além de recuperar dados e agregar informações para descobrir padrões e tendências em seus dados, a ferramenta pode armazenar e indexar dados e textos de uma forma que suporte pesquisas rápidas. Além disso, à medida que o volume de dados e consulta crescem, a natureza distribuída do Elasticsearch permite que sua implantação cresça perfeitamente junto com ela. Já o Kibana foi desenvolvido em 2013 e é uma aplicação que trabalha com o Elastic Stack, fornecendo recursos de busca e visualização de dados indexados no Elasticsearch [5]. Conhecido como a ferramenta de gráficos, o Kibana permite que o indivíduo explore, visualize e compartilhe percepções sobre seus dados de forma interativa e gerencie e monitore a estrutura/pilha. O Kibana oferece suporte às seguintes funcionalidades:

- Busca e visualização de dados indexados no Elasticsearch e fornecimento em tempo real de visualizações analíticas de grandes volumes de dados por meio da criação de gráficos;
- Monitorar, gerenciar e proteger uma instância do Elastic Stack via interface Web;
- Centralização do acesso para soluções integradas desenvolvidas no Elastic Stack.

2.3 Workflow

A figura 1 ilustra o workflow de todo o projeto facilitando seu entendimento onde cada fase é explicada no item indicado.

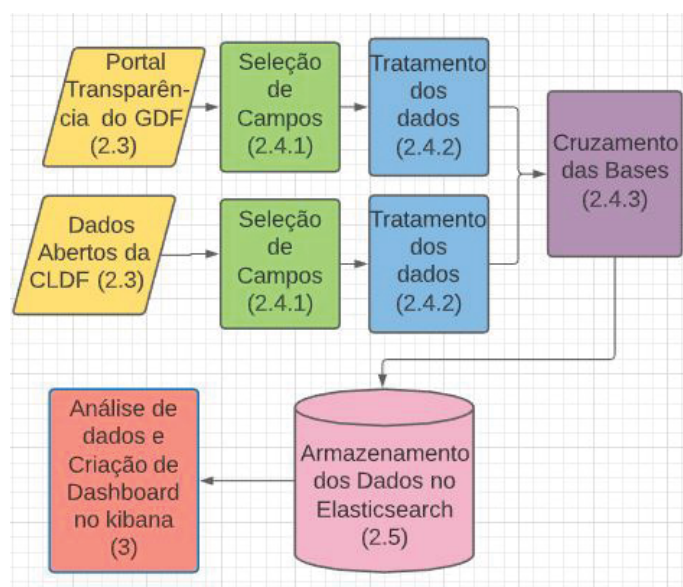


Figura 1: Fluxograma do projeto

2.4 Base de Dados

Os dados utilizados neste projeto estão disponíveis em plataformas de dados abertos. A coleta foi extraída de duas bases de dados, de forma manual porque nenhuma das plataformas disponibiliza uma Interface de Programação de Aplicações (API) para download automatizado, então, os dados foram baixados no formato Comma-separated values (CSV) e Microsoft Excel Open XML Spreadsheet File (XLSX). Uma das bases de dados está disponível no Portal da Transparência do Distrito Federal (DF), [7] que fornece dados abertos do governo do DF. Foram extraídos dados sobre gastos realizados na área da educação, saúde, administração e segurança entre janeiro de 2009 a dezembro de 2019, e as receitas arrecadas no mesmo período. Também foram utilizados dados abertos da Câmara Legislativa do Distrito Federal (CLDF), [2] informações relacionadas às verbas indenizatórias (cotas) dos deputados entre janeiro de 2013 a dezembro de 2019. Segundo o próprio website, as verbas indenizatórias são os recursos que o Poder Legislativo repassa para custear os trabalhos dos gabinetes parlamentares. A verba indenizatória é usada para ressarcir despesas com locação de imóveis e de veículos, material de expediente, combustível e contratação de consultoria, entre outros.

2.5 Tratamento e Análise de Dados

► 2.4.1 Seleção de Campos

Após uma análise exploratória das bases de dados, alguns campos de cada base foram selecionados para compor o conjunto de dados final. Assim sendo, na base de despesas do GDF foram utilizados os campos data, função, subfunção, elemento de despesa, ação e liquidado. Na base receitas do GDF os campos foram: data, categoria econômica, origem e receita realizada. Já na base de dados Cotas Deputados foram selecionados as categorias data e valor.

► 2.4.2 Tratamento de Dados

Também foi necessário realizar alguns procedimentos de tratamento dos dados para cada base de dados, tais como informados a seguir:

- Base do GDF:
 - Campos categóricos foram convertidos para minúsculo;
 - Normalização (padronização) dos campos 'Função', 'Subfunção';
 - Correção dos atributos numéricos que usavam vírgula como separador decimal;
 - Correção de valores NULL (0 para campos numéricos e N/I para campos categóricos);
 - Limpeza de espaços em excesso.
- Base CLDF:
 - Campos categóricos foram convertidos para minúsculo;
 - Normalização (padronização) dos campos 'Função', 'Subfunção';
 - Correção dos atributos numéricos que usavam vírgula como separador decimal;
 - Correção de valores NULL (0 para campos numéricos e N/I para campos categóricos);
 - Correção do campo data, alguns valores inválidos, alguns foram corrigidos outros foram substituídos por N/I;
 - Limpeza de espaços em brancos em excesso. Para todas as bases novos campos foram criados, ano e mês, com base no campo data.

Todo este procedimento foi feito utilizando python (jupyter notebook).

► 2.4.3 Cruzamento das Bases de Dados

Como as bases de dados são heterogêneas e a ferramenta utilizada para o dashboard (kibana) não permite elaborar gráficos com bases diferentes foi feita uma análise e exploração de cada campo em cada uma dessas repositórios para verificar o que poderia ter a mesma informação, mas com nomes diferentes. Então, foram criados novos campos contendo cada informação igual, abaixo será demonstrado o nome do novo campo, o que ele armazena e alguns exemplos de tipo de valores vindo de cada base (despesas, receitas e cotas):

- Tipo: irá armazenar o tipo de informação, podem ser: Despesas, Receitas ou Cotas Deputados;
- Setor: irá armazenar o setor que está envolvido no registro;
 - Despesas: administração, educação, saúde, segurança pública;

- Receita: Receita GDF;
- Cotas Deputados: CLDF;
 - Categoria: irá armazenar qual categoria ("subsetor") do setor;
 - Despesas: ensino fundamental, ensino médio, comunicação social, tecnologia da informação dentre outros;
 - Receitas: receitas correntes, receitas de capital, deduções da receita dentre outros;
 - Cotas Deputados: essa base não possui nenhum tipo de informação dessa natureza, então foi inserido valores N/I;
 - Valor: armazena valores das despesas, receitas ou cotas parlamentares;
 - Todo este procedimento foi feito utilizando python (jupyter notebook).

2.6 Data Lake

O Data Lake pode ser definido como "armazenamento centralizado, consolidado e persistente de dados brutos, não modelados e não transformados de múltiplas fontes, sem um esquema pré-definido explícito e sem metadados definidos externamente." [3] A etapa seguinte ao tratamento dos dados foi a transferência deles para o elasticsearch, que possibilita armazenar grandes volumes de dados tanto dados estruturados como não estruturados. Neste presente trabalho, a ferramenta elasticsearch funcionou como um data lake, pois os dados ficaram armazenados e não receberam nenhum tratamento. Todas as métricas e consultas foram feitas no Kibana.

3 Resultados

Este capítulo fornece uma apresentação dos resultados da investigação que resultaram da análise de dados do Governo do Distrito Federal.

3.1 Apresentação e Análise

Os dados foram analisados na ferramenta Kibana na qual foram gerados gráficos com o objetivo de confrontar as receitas públicas do Distrito Federal com seus respectivos gastos e verificar como o dinheiro é investido nas principais áreas como saúde, educação, segurança pública e administrativas. A primeira análise foi comparar as despesas e receitas realizadas pelo governo. Foi considerado o período de uma década, ou seja, 2009 até 2019, e não foram considerados os dados de 2020, pois as informações disponíveis eram apenas do primeiro trimestre do ano. Gráfico da Figura 2.

De acordo com o gráfico acima, verifica-se que as despesas com educação e saúde acompanharam a curva de receitas, o que já não ocorreu com segurança pública e administração, que se mantiveram praticamente estáveis ao longo do período. Das quatro áreas investigadas, as menores despesas foram realizadas com segurança pública e as maiores com a educação.

Também foi realizada uma comparação entre as receitas e os gastos com educação, dando enfoque aos gastos

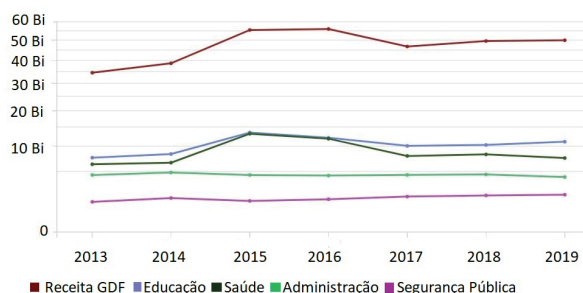


Figura 2: Comparativo entre as despesas e receitas do governo do DF. Fonte: elaborado pela autora (2020) a partir do Portal da Transparência do DF

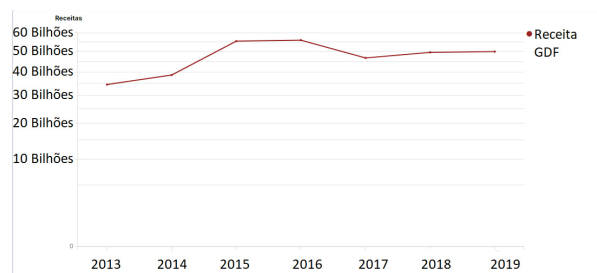


Figura 4: Receitas do Governo do Distrito Federal. Fonte: elaborado pela autora (2020) a partir do Portal da Transparência do DF

com ensino fundamental, ensino médio e ensino profissional. Gráfico da Figura 3.

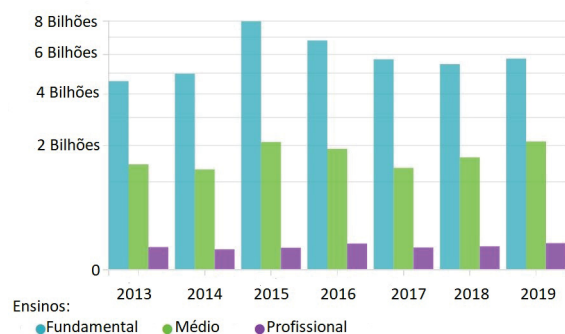


Figura 3: Enfoque aos gastos com educação. Fonte: elaborado pela autora (2020) a partir do Portal da Transparência do DF

De acordo com os dados apresentados, foi observado que os gastos com educação no ensino fundamental e com ensino médio acompanham a curva de receita. Já a educação com ensino profissional se manteve estável ao longo do período. Os maiores investimentos na educação pública do DF foram realizados no ensino fundamental, enquanto o ensino profissional recebeu os menores. Outro importante estudo realizado foi a comparação entre as receitas e as cotas da Câmara Legislativa do Distrito Federal. Foram analisados os gastos totais com cotas, gastos com diárias e passagens e gastos feitos com publicidade, conforme apresentados nos Gráficos das Figuras 4 a 7. Como os dados disponibilizados desta fonte são somente a partir de 2013, os dados apresentados tanto das receitas como das despesas foram a partir de 2013.

De acordo com as figuras de 4 a 7, foi observado que as cotas da CLDF não acompanharam a curva de receita: as despesas com as cotas vêm diminuindo ao longo dos anos, exceto em 2016, que houve um grande aumento nos gastos comparado com o ano anterior.

As despesas com diárias e passagens não acompanharam as curvas de receitas e nem das cotas; verifica-se um grande aumento em 2014, ano de eleição. Em relação aos gastos com publicidade, verifica-se que eles também não acompanharam as curvas de receitas, nem das cotas.

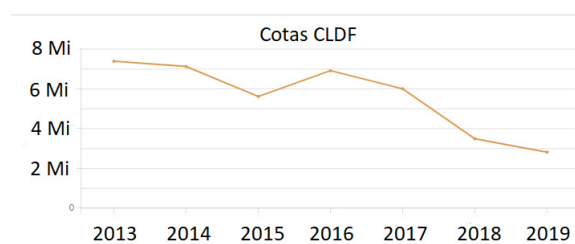


Figura 5: Despesas com as cotas da Câmara Legislativa do Distrito Federal. Fonte: elaborado pela autora (2020) a partir do Portal da CLDF

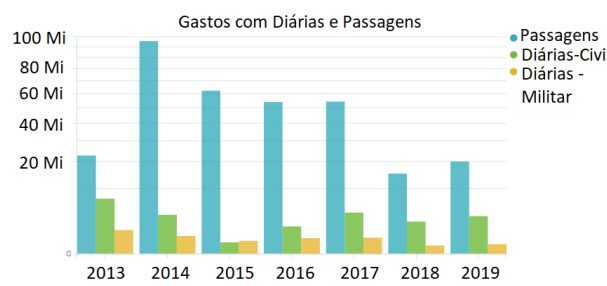


Figura 6: Despesas com diárias e passagens realizadas pela Câmara Legislativa do DF. Fonte: elaborado pela autora (2020) a partir do Portal da CLDF

E ainda apresentaram um aumento nos anos de 2013 e 2017, anos que antecedem as eleições, tendo em vista que em ano eleitoral gastos publicitários são proibidos por lei.

4 Discussão

4.1 Limitação da Pesquisa

Foram encontrados algumas limitações e dificuldades para realizar a pesquisa, principalmente relacionadas a base de dados, tais como:

- Falta de uma API para baixar os dados, foi necessário baixar todos os dados em CSV ou XLSX manualmente;
- Falta de documentação das bases de dados;
- Na base de dados da CLDF (cotas) havia muitos valores errados relacionados ao nome da empresa / for-



Figura 7: Despesas com publicidade realizadas pela Câmara Legislativa do DF. Fonte: elaborado pela autora (2020) a partir do Portal da CLDF

necedor;

- Na base de dados da CLDF (cotas) há muita pouca informação, por exemplo, não há informação com o que o deputado gastou (alimentação, combustível etc). As informações estão bem genéricas;

- O software Kibana não permite fazer gráficos com bases diferentes (cruzamento de bases durante a execução), sendo necessário fazer um trabalho de transformação bem detalhada para deixar as bases iguais antes de armazená-las no elasticsearch.

5 Conclusões

De acordo com os dados relacionados aos gastos públicos do governo do Distrito Federal, dos quatro principais setores pesquisados, verifica-se que as despesas com educação e saúde acompanharam a curva de receitas, o que já não ocorreu com segurança pública e administração, que se mantiveram praticamente estáveis ao longo do período. Realizando a comparação entre as receitas e as cotas da Câmara Legislativa do DF, nota-se que as despesas com as cotas vêm diminuindo ao longo dos anos, exceto em 2016. As despesas com diárias, passagens e publicidade também não acompanharam as curvas de receitas: as maiores despesas aconteceram nos anos de 2014 e 2017 (diárias e despesas) e 2013 e 2017 (publicidade) sendo que aos anos de maiores arrecadações foram 2015 e 2016. De acordo com os dados relacionados aos gastos públicos do governo do Distrito Federal não se percebe nenhum indício de que as despesas públicas foram investidas de forma a corroborar com algum indício de falta de legitimidade. O que pode e deve ser questionado é se os dados estão inseridos nos respectivos portais em sua totalidade e de forma realmente transparente. Será que em um portal onde o governo realiza sua prestação de contas perante toda a população seria adequado colocar notas genéricas, algumas sem a

identificação correta do fornecedor e sem a sua devida destinação do gasto realizado?

6 Trabalhos Futuros

Como sugestão para trabalhos futuros, sugere-se que este presente trabalho seja estendido a outros estados, prefeituras e até mesmo ao Governo Federal realizando os devidos ajustes na Base de Dados. Torna-se necessário conhecer a base de dados de cada ente da Federação, realizar um pré-processamento manual dos dados identificando as colunas utilizadas no dashboard para o posterior armazenamento no elasticsearch.

Agradecimentos

Agradeço aos Professores da UFPR, aos colegas e amigos Arcson Assunção, Antônio C. Júnior, Gustavo Boiko e Paulo Chixaro pelo apoio e suporte durante o curso.

Referências

- [1] BABBIE, Earl R, *The practice of social research.*, (Belmont, Thomson Higher Education. 2007.), 11a ed.
- [2] CÂMARA LEGISLATIVA DO DISTRITO FEDERAL, Disponível em <<http://dadosabertos.cl.df.gov.br/View/conjuntos.html>>, (Acesso em: 01 mar. 2020.)
- [3] Ciência e Dados, Disponível em <<http://www.cienciaedados.com/data-lake-a-evolucao-do-armazenamento-e-processamento-de-dados/>>. , (Acesso em: 28 jul. 2020)
- [4] ELASTIC, Disponível em <<https://www.elastic.co/pt/what-is/elasticsearch>>, (Acesso em: 28 jul. 2020).
- [5] KIBANA, Disponível em <<https://www.elastic.co/pt/what-is/kibana>>, (Acesso em: 28 jul. 2020).
- [6] KUMAR, Ranjit, *Research methodology: a step-by-step guide for beginners*, (London. SAGE. 2005).
- [7] PORTAL DA TRANSPARÊNCIA DO DISTRITO FEDERAL, Disponível em <<http://www.transparencia.df.gov.br/>>., (Acesso em: 01 mar. 2020.).
- [8] PYSCIENCE-BRASIL, Disponível em <<http://pyscience-brasil.wikidot.com/>>. , (Acesso em: 25 jul. 2020.).
- [9] PYTHON, Disponível em <<https://www.python.org/>>. , (Acesso em: 01 ago. 2020.).
- [10] RICHARDSON, Roberto J. , *Pesquisa social: métodos e técnicas*, (São Paulo: Atlas. 1999.) 3.ed.

- [11] VERGARA, Sylva, *Projetos e relatórios de pesquisa em administração.*, (São Paulo: Atlas, 2013), 14.ed.