

CIÊNCIA DE DADOS APLICADA À EXTRAÇÃO E ANÁLISE DO PROGRAMA NOTA FISCAL GOIANA

Jaqueline Souza Medeiros
Graduação em Estatística, UFG
medeirosestatistica@gmail.com

Vitória Nery Lopes Manheze
Graduação em Engenharia de Produção, PUCGO
vitoria.nery9@gmail.com

Douglas Farias Cordeiro
Professor Adjunto da Universidade Federal de Goiás
cordeiro@ufg.br

Resumo

Este artigo tem como objetivo apresentar uma aplicação ETL (*Extract, Transform, and Load*) voltada à automação da coleta de dados da Nota Fiscal Goiana por meio do uso de *web scraping* e consolidação dos registros através da utilização de técnicas de visualização de dados. A aplicação inclui um processo de tratamento e encriptação dos dados, os quais são persistidos em um banco de dados. A visualização é realizada em um relatório online gerado através do software *Power BI*. A solução proposta tem como objetivo tornar a coleta e análise de dados mais eficaz para os usuários interessados na Nota Fiscal Goiana. Os resultados obtidos mostraram que a aplicação desenvolvida é capaz de coletar e processar dados de forma eficaz, permitindo que os usuários visualizem informações importantes em um relatório online de maior usabilidade. O processo de desenvolvimento da aplicação incluiu a seleção de tecnologias adequadas e a realização de testes para verificar a eficiência do programa fiscal.

Palavras-chave: Nota Fiscal Goiana, Web scraping, ETL, Visualização de dados, Banco de dados.

Abstract

This article presents an ETL (Extract, Transform, and Load) application aimed at automating data collection from Invoice Goiana through the use of web scraping and consolidation of records through the use of data visualization techniques. The application includes a data processing and encryption process, which is persisted in a database. The visualization is carried out in an online report generated using Power BI software. The proposed solution aims to make data collection and analysis more effective for users interested in the Goiana Invoice. The results obtained showed that the developed application is capable of collecting and processing data effectively, allowing users to view important information in an online report with greater usability. The application development process included the selection of appropriate technologies and carrying out tests to verify the efficiency of the tax program.

Keywords: Nota Fiscal Goiana, Web scraping, ETL, Data visualization, Database.

1 Introdução

Segundo informações disponibilizadas pelo Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE, 2023), a nota fiscal foi desenvolvida na década de 1970 no Brasil, com o objetivo de diminuir a sonegação fiscal. A emissão de nota fiscal tornou-se obrigatória para todas as empresas envolvidas em atividades comerciais ou registradas no regime Simples Nacional do Brasil a partir de 1º de julho de 2018. Essa data marca a implementação do Sistema de Escrituração Digital das Obrigações Fiscais, Previdenciárias e Trabalhistas (eSocial), o qual se trata de um sistema unificado de envio de informações fiscais, previdenciárias e trabalhistas para o governo brasileiro. O eSocial abrange diversos aspectos das obrigações fiscais, incluindo a emissão de notas fiscais eletrônicas. Neste sentido, a partir de julho de 2018, todas as empresas envolvidas em atividades comerciais, independentemente do regime tributário, passaram a ter a obrigação legal de emissão de notas fiscais eletrônicas. É importante pontuar que a nota fiscal assegura a conformidade legal das operações comerciais e o devido pagamento dos tributos pela empresa. Além disso, ela proporciona maior confiança aos clientes, uma vez que esse documento garante direitos, como o de realizar trocas de produtos.

No contexto tributário, ao qual a nota fiscal se consolidou como principal mecanismo de monitoramento e fiscalização, o ICMS (Imposto sobre Circulação de Mercadorias e Serviços) se consolida como um dos principais impostos. O ICMS é de competência dos estados e do Distrito Federal, sendo que o valor da sua alíquota varia de estado para estado. No estado de Goiás, o ICMS é considerado como a principal fonte de arrecadação do Governo de Goiás (SECRETARIA DE ESTADO DA ECONOMIA - GOVERNO DO ESTADO DE GOIÁS, s.d.). A sonegação do ICMS se reverbera na subtração de recursos que seriam direcionados ao governo para custear serviços públicos. Desta forma, ao demandar a emissão da nota fiscal, o consumidor está auxiliando o governo a garantir que mais empresas estejam cumprindo suas obrigações tributárias (SEE-GO, s.d.)

A Nota Fiscal Goiana (NFG) é um programa de cidadania fiscal promovido pelo governo do estado de Goiás. O programa surgiu através da Lei Nº 18.679 de 3 de dezembro de 2014, com o objetivo de incentivar a população a participar de ações que auxiliam no incremento da arrecadação tributária, verificar a aplicação de recursos públicos, entre outras (GOIÁS, 2014). No início de 2023 para uma pessoa participar do programa basta se cadastrar no portal e ao fazer compras no comércio do estado de Goiás solicitar seu CPF na nota. As notas são enviadas automaticamente para o sistema da NFG e a cada R\$1,00 gasto equivale a um ponto e a cada 100 pontos o cidadão ganha um bilhete para participar do sorteio (GOVERNO DE GOIÁS, 2021). Além dos sorteios, os participantes também adquirem descontos no IPVA com base na quantidade de bilhetes gerados.

No dia 2 de maio de 2019, por meio do Decreto nº 9.440, se determinou a suspensão dos sorteios da NFG, considerando data retroativa a partir do dia primeiro de janeiro de 2019 por tempo indeterminado. Durante esse período, o programa manteve o critério de pontuação com vistas à aplicação de descontos no IPVA. Os sorteios foram retomados em maio de 2021 através do Decreto nº 9.867 (GOIÁS, 2019; GOIÁS, 2021).

Embora a NFG possua considerável importância no apoio à responsabilidade fiscal do estado, a disponibilidade de informações ao cidadão, principalmente em termos de temporalidade e sumarização, é algo pouco explorado, sendo o site essencialmente focado em informações pessoais (privativas) ou informações pontuais de forma isolada. Entende-se que a disponibilidade de informações mais generalizadas pode corroborar para uma melhor compreensão da proposta do programa, auxiliando ainda em questões de transparência. Neste sentido, o presente estudo se dedica a desenvolver uma solução, baseada em ciência de dados, para monitoramento e visualização dos dados da NFG.

Estruturalmente, o estudo se desenvolve por meio de um percurso que perpassa pela explicação das metodologias e ferramentas exploradas até a apresentação dos resultados alcançados. De forma específica, a seção 2 apresenta as metodologias que possibilitam a construção da aplicação

de ETL (*Extract, Transform, Load*) e do relatório em Power BI. Na seção 3 é discutida a aplicação dessas metodologias, quais métodos e passos foram utilizados e a utilização das ferramentas apresentadas na seção 2. Na seção 4 é discutida a construção do pipeline de extração, transformação e carregamento dos dados e a interpretação da visualização do relatório em Power BI. Por fim, a seção 5 apresenta as conclusões acerca da efetividade do programa.

2 Embasamento teórico

Nesta seção apresenta-se uma revisão da literatura com teorias e conceitos importantes que compõem o alicerce do presente trabalho. Ao expor essas bases teóricas espera-se possibilitar um melhor entendimento e interpretação do percurso metodológico adotado, assim como dos resultados obtidos. Neste sentido, se incluem definições que englobam a ciência de dados de uma forma geral, o uso de web scraping para obtenção de dados, ferramentas de *self-service BI*, análise exploratória e teste de hipótese.

2.1 Ciência de dados

A ciência de dados pode ser descrita como uma área interdisciplinar dinâmica, que é composta por conceitos provenientes de programação, estatística, matemática e visualização de dados, além de conhecimentos relacionados a princípios de negócios e sobre o domínio do problema a ser explorado (AMARAL, 2016). Dessa forma, um cientista de dados é conhecido por sua capacidade de extrair conhecimentos e *insights* significativos de uma ampla gama de dados, independentemente de sua estrutura, se alicerçando no emprego de técnicas e abordagens que visam aprimorar a eficiência e a otimização de processos e atividades organizacionais (GRUS, 2019).

Ao estabelecer uma interseção entre várias disciplinas, a ciência de dados não apenas integra conhecimentos técnicos avançados, como também envolve a catalisação de uma diversidade de oportunidades investigativas e inovadoras. Diante disso, a crescente complexidade associada à manipulação e interpretação de dados tem impulsionado a demanda por soluções capazes de amparar o gerenciamento de bancos de dados complexos, assim como o desenvolvimento de algoritmos sofisticados e ferramentas eficazes para a análise de comportamentos e padrões de dados (MARCHIONINI, 2016). De maneira paralela, soluções especializadas em dados desempenham um papel vital ao garantir interações eficazes entre aplicações e para com usuários, fornecendo mecanismos para atender necessidades em constante evolução de informação na sociedade contemporânea.

Para além disso, a rápida evolução tecnológica, característica do fenômeno Big Data, é acompanhada de desafios significativos. A crescente dependência de conjuntos de dados complexos e diversificados se encontra envolta a preocupações sobre questões éticas e de privacidade, além da necessidade premente de aplicação de abordagens em segurança da informação voltadas à salvaguarda da confidencialidade de dados sensíveis (PAULA; CORDEIRO, 2015; MAGRANI, 2019). Ademais, a interpretação e comunicação eficazes dos *insights* obtidos a partir da análise de dados representam em si um desafio adicional, exigindo habilidades que demandam estratégias comunicacionais eficientes para a descoberta de conhecimento (CAO, 2021).

2.2 Web scraping

No contexto da ciência de dados, a aquisição de dados se destaca como uma das etapas fundamentais na construção de soluções robustas, uma vez que é responsável por fornecer os elementos essenciais para os processos de geração de informações e descoberta de conhecimento (AMARAL, 2016). No entanto, frequentemente, os dados de interesse não estão prontamente disponíveis em um formato acessível e integrado, o que requer o desenvolvimento de estratégias específicas para a obtenção desses dados. Neste sentido, é comum o desenvolvimento de abordagens baseadas no uso de APIs (*Application Program Interface*) ou por meio do uso de técnicas de raspagem de dados, comumente conhecidas pelo termo em inglês *web scraping*. Soluções de *web scraping* geralmente são aplicadas para a automatização da extração de conteúdos disponibilizados em páginas da web por meio da identificação de marcadores, conhecidos como *tags* (MITCHELL, 2019).

A aplicação de técnicas de *web scraping* se consolida como uma abordagem versátil e eficaz para coletar dados em larga escala a partir de fontes variadas disponibilizadas na web, sendo aplicada, por exemplo, em investigações no contexto de redes sociais (VIEIRA; CORDEIRO, 2023), portais de notícias (CORDEIRO; MELO, VIEIRA, 2023), estudos financeiros (KROTOV; TENNYSON, 2018). Ao automatizar o processo de coleta de dados, as soluções baseadas em *web scraping* permitem a rápida recuperação de grandes volumes de informações de maneira estruturada ou semi-estruturada, possibilitando a consolidação dos dados em formatos como csv, json ou em bancos de dados relacionais, o que é fundamental para alimentar modelos de ciência de dados.

É fundamental ressaltar que o uso de técnicas de *web scraping*, para além das diversas vantagens elencadas, está inerentemente ligado a questões relacionadas à ética e à legalidade, uma vez que a coleta de dados de fontes na web pode estar sujeita a restrições legais e políticas de privacidade. Diante disso, ao empregar estratégias de *web scraping*, é necessário se investigar as diretrizes éticas e a conformidade regulatória, de maneira a garantir a integridade e a legitimidade da obtenção de dados (KROTOV; SILVA, 2018).

2.2 Ferramentas de self-service bi

Ferramentas de *Self-Service BI* surgiram para permitir usuários de *Business Intelligence* (BI) se tornarem autônomos e menos dependentes de soluções especificamente realizadas por especialistas de tecnologia da informação. Seus quatro objetivos principais são: acesso fácil aos dados de origem para geração de relatórios e análises; ferramentas de BI fáceis de usar e suporte aprimorado para a análise de dados; opções de armazenamento de dados fáceis de implantar e gerenciar; e interfaces de usuário mais simples e personalizáveis (IMHOFF; WHITE, 2011).

Um exemplo de ferramenta de *Self-Service BI* é o Power BI. No presente estudo, tal ferramenta foi escolhida por seus diferenciais no quesito de compartilhamento de informações, onde o relatório desenvolvido pode ser publicado via *web* e visualizado por quais públicos definidos. Tais informações podem ser atualizadas de acordo com a necessidade do desenvolvedor ou da regra de negócio considerada (FRAGA, 2019). Além disso, é uma ferramenta que auxilia na medição de desempenho de um processo ou de um projeto auxiliando em rápidas tomadas de decisão.

2.3 Análise exploratória

A análise exploratória de dados desempenha um papel crucial na descoberta e compreensão de estruturas subjacentes em conjuntos de dados. Sua natureza investigativa é voltada para a identificação de padrões latentes e comportamentos intrínsecos que podem ser imperceptíveis por meio de uma observação superficial (FAVERO, 2017). Essa abordagem analítica oferece uma visão

aprofundada e abrangente dos dados, contribuindo para a extração de informações valiosas que podem alimentar decisões e estratégias informadas (SHARDA; DELEN; TURBAN, 2019).

Nesse contexto, a análise exploratória se destaca como um procedimento multifacetado, que engloba a exploração de diversas dimensões dos dados. Tais aspectos envolvem a investigação da distribuição dos valores, a identificação de tendências de variação e a detecção de possíveis anomalias. A partir disso, a visualização desempenha um papel fundamental nesse processo, representando uma ponte tangível entre os números e a compreensão humana. Gráficos e visualizações elucidativas podem revelar padrões de dispersão, associações temporais e até mesmo comportamentos cíclicos, permitindo um entendimento mais aprofundado e rápido dos dados (KNAFLIC, 2019). Tais estratégias podem ser desenvolvidas por meio de soluções baseadas no uso de ferramentas de *Self-Service BI*.

2.4 Teste de hipótese

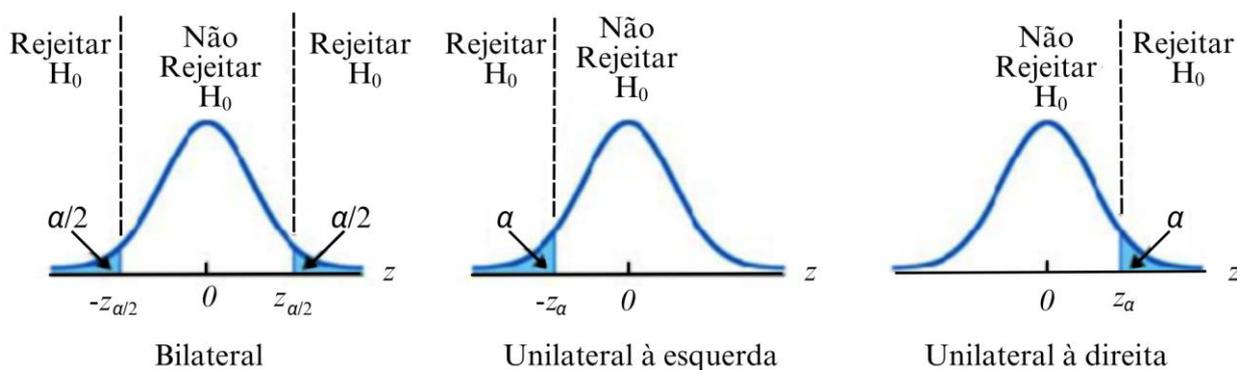
Testes de hipóteses são ferramentas estatísticas que auxiliam na tomada de decisão de rejeitar ou não determinada afirmação usando um conjunto de dados como base. Essas afirmações são chamadas de hipóteses, tem-se a hipótese nula (H_0), também chamada de hipótese de interesse, e a hipótese alternativa (H_1 ou H_a). Mesmo após rejeitar ou não a hipótese nula, a possibilidade de se ter cometido um erro não é descartada. É definido um nível de confiança do teste (α) que é a probabilidade de rejeitar a hipótese nula sendo ela verdadeira (BOLFARINE; SANDOVAL, 2021).

A região crítica, também chamada de região de rejeição, é complementar a região de aceitação, ou não rejeição. Essa região pode ser bilateral ou unilateral, ou seja, se a hipótese nula for definida como $\theta = \theta_0$, então seria bilateral, pois se o valor que se está testando for muito maior ou muito menor ele seria rejeitado. Já no caso unilateral é quando, ao invés de testar se os valores são iguais, testa-se se eles são maiores que ou menores que o outro, formando assim apenas uma região de rejeição. Esse comportamento pode ser observado através da Figura 1 (MAGALHÃES; DE LIMA, 2002).

De acordo com Magalhães e de Lima (2002), as etapas para a realização de um teste de hipótese são:

1. Determinar a hipótese nula e alternativa;
2. Definir o formato da região crítica baseando-se na hipótese alternativa;
3. Identificar a distribuição do estimador e calcular sua estimativa;
4. Fixar o nível de significância α e determinar a região crítica;
5. Realizar o teste com base na estimativa e na região crítica.

Figura 1 - Caldas de Testes de Hipóteses



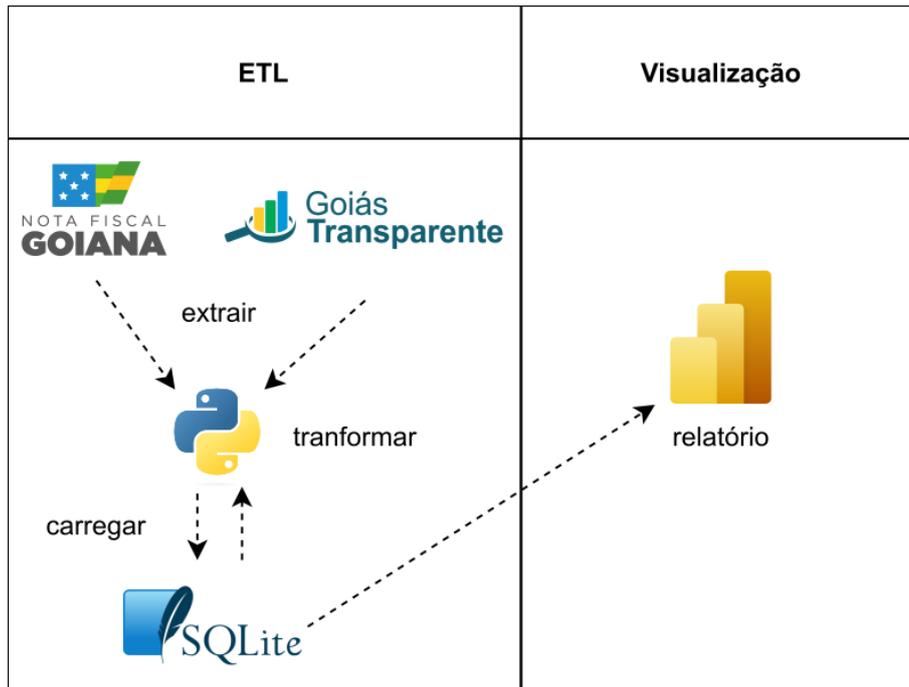
Fonte: adaptado de Fumachi (2022).

De maneira geral, determinar a hipótese nula é definir o que se deseja testar e a hipótese alternativa sendo o seu complemento. Definir o formato da região crítica, nada mais é que definir se seu teste será bilateral ou unilateral, feito isso é preciso identificar a distribuição do estimador e calcular sua estimativa. Ao definir o nível de significância, conseqüentemente está definindo a região crítica, com isso basta comparar se sua estimativa está dentro da região crítica ou não.

3 Metodologia

O projeto é composto por três grandes fases: "Extração, Transformação e Carregamento (ETL)", "Visualização" e "Teste de Hipótese". A seção 3.1 detalha o processo de ETL do projeto, enquanto a seção 3.2 aborda a visualização da informação coletada. Além disso, a seção 3.3 é dedicada à realização do teste de hipótese. Um diagrama ilustrando a arquitetura do projeto, incluindo as etapas de ETL e Visualização, pode ser observado na Figura 2.

Figura 2 - Arquitetura do Projeto



Fonte: Elaborado pelos autores (2023).

3.1 Extração, transformação e carregamento

A etapa de Extração é executada através do uso de técnicas de *web scraping* em Python para obter dados do portal "Nota Fiscal Goiana" e do portal "Goiás Transparente". Após a extração, os dados coletados são submetidos a uma verificação de duplicidade, onde somente aqueles considerados novos são processados e preparados para serem armazenados em um banco de dados. A etapa de Transformação envolve a conversão dos dados extraídos em um formato adequado para análise posterior, o que inclui a limpeza, a padronização e a conversão dos tipos de dados. Finalmente, na etapa de Carregamento, os dados são salvos no banco de dados.

3.2 Visualização

Já a fase de Visualização é implementada por meio de um relatório gerado no Power BI, que se conecta ao banco de dados para apresentar os dados de forma resumida e clara. Foi levantado as informações mais importantes para as análises que estavam sendo feitas. O relatório oferece informações relevantes sobre os dados armazenados, permitindo uma fácil visualização de tendências, padrões e outras análises relevantes sobre o programa da nota premiada.

3.3 Teste de hipótese

Para avaliar a efetividade do programa da Nota Fiscal Goiana optou-se por usar as arrecadações provenientes do ICMS e fazer a comparação antes da implementação do programa e

após sua implementação. Uma forma de fazer essa comparação é através de testes de hipótese, especificamente teste t-pareado.

Esse teste é usado quando deseja-se comparar duas médias de amostras tomadas antes e após uma determinada intervenção. No caso da NFG será comparado a média da variação da arrecadação mensal do ICMS no mesmo período de tempo antes e depois do programa ser implementado (MAGALHÃES; DE LIMA, 2002).

Decidiu-se por usar a variação mensal para diminuir os efeitos da inflação ou possível deflação que possa ter ocorrido durante o período analisado. O cálculo da região crítica e os estimadores são feitos no *software* Python através da biblioteca SciPy (VIRTANEN *et al.*, 2020)

4 Resultados e discussões

O Power BI foi o sistema escolhido para apresentar as análises e os indicadores do trabalho, ele consegue trazer as informações de forma clara e bem entendidos por qualquer público que esteja buscando informações a respeito da Nota fiscal Goiana, após as definições de quais indicadores seriam essenciais para a apresentar o desempenho da Nota Fiscal Goiana foi desenvolvido medidas utilizando linguagem DAX.

A fácil interação de conexões de bancos de dados e a dinâmica de programação de coleta de dados com o seu *gateway* facilitou a escolha do software, e após isso foram incluídos filtros para auxiliar o leitor a fazer sua análises, com filtro de unidades federativas e municípios que os ganhadores residem, e filtros dinâmicos deslizantes com a data de realização e o valor total dos prêmios. A Figura 3 apresenta uma visão macro do relatório, o mesmo também pode ser analisado em tempo real através da *url* disponível¹.

Na Figura 4, se tem uma visão focada do relatório, na qual pode-se identificar os indicadores que foram levantados a partir da extração dos dados do portal da Nota Fiscal Goiana. É possível identificar a quantidade de sorteios que ocorreu no período em que o portal disponibilizou os dados, quando ocorreu o último sorteio realizado e o total sorteado (em reais). Na tabela é possível identificar a quantidade de prêmios que contemplou os ganhadores nos municípios do Brasil e o valor que foi disponibilizado como prêmio.

A Figura 5, foca na parte inferior do relatório, onde é exposto informações das arrecadações totais através do ICMS do estado de Goiás em um gráfico de área com 4 linhas verticais demarcando 4 períodos: Início do período de comparação (01/01/2011), Início do Programa (01/01/2015), Pausa do Programa (01/01/2021), e Retomada do Programa (01/06/2021). Abaixo do gráfico encontram-se as medidas calculadas de acordo com os períodos destacados no gráfico.

Figura 3 - Visualização das Informações do Programa Nota Fiscal Goiana



Fonte: Elaborado pelos autores (2023).

Figura 4 - Indicadores e Tabelas



Fonte: Elaborado pelos autores (2023)

Figura 5 - Gráfico de área e Indicadores



Fonte: Elaborado pelos autores (2023)

5 Conclusão

Além de automatizar todo o processo de coleta dos dados da Nota Fiscal Goiana e os montantes arrecadados pelo Estados de Goiás através do ICMS e apresentar estas informações de uma forma simples e compreensiva, esse trabalho também aplicou um teste de hipótese para testar se o programa cumpriu um dos seus objetivos, que é aumentar a arrecadação do estado.

Considerando a seguinte hipótese nula:

H_0 : A variação média da arrecadação do ICMS antes da aplicação do programa é maior do que a variação média da arrecadação do ICMS após a aplicação do programa.

Avaliando o período anterior ao programa como 01/01/2011 à 01/01/2015 e após o programa como 01/01/2015 à 01/01/2019, com 95% de confiança rejeita-se a hipótese nula (p -valor 0,66). Ou seja, a variação média de arrecadação após a implementação do programa é maior do que a variação média de arrecadação antes da implementação.

Apesar de ainda não ser possível testar o intervalo completo do retorno do programa com a sua pausa, os dados iniciais da variação média da arrecadação do ICMS indicam que a sua efetividade diminuiu. Isso pode ser conduzido por diversas coisas, como a pandemia de COVID-19 durante o período de pausa do programa pode ter causado aumentos do consumo no estado de Goiás e após essa estabilização pós-COVID da economia ter desacelerado as arrecadações, apesar da volta do programa. Ou até mesmo por uma falta de investimento em marketing do programa em meios de comunicação populares.

Trabalhos futuros focados na NFG podem aproveitar dos mecanismos aqui apresentados de extração de informações dos resultados do programa para fazer análises mais aprofundadas da sua eficiência e disseminação. Por fim, destaca-se que todos os *scripts* e dados necessários para reproduzir esses resultados estão disponíveis no repositório do projeto.

Referências

AMARAL, F. **Introdução à Ciência de Dados**: mineração de dados e big data. Rio de Janeiro: Alta Books, 2016.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. SBM, 2001.

CAO, L. Data Science: Profession and Education. *IEEE Intelligent Systems*, v. 34, n. 5, p. 35-44, 2019. doi: <https://doi.org/10.1109/MIS.2019.2936705>.

CORDEIRO, D. F.; MELO, J. P. L.; VIEIRA, L. M. Violência contra a mulher no estado de Goiás com base em análises de mídias digitais: estudo de caso no portal de notícias G1. **Fronteiras – Estudos Midiáticos**, v. 25, n. 1, p. 28-43, 2023. doi: <https://doi.org/10.4013/fem.2023.251.03>.

FAVERO, L. P. **Manual de Análise de Dados: Estatística e Modelagem Multivariada com Excel, SPSS e Stata**. Rio de Janeiro: GEN LTC, 2017.

FRAGA, A. **Microsoft Power BI: Gráficos, Banco de Dados e Configuração de Relatórios**. Rio de Janeiro: Alta Books, 2019.

FUMACHI, Edson Fernando. Teste de hipóteses. *In: Teste de hipóteses*. [S. l.], 9 nov. 2022. Disponível em: <https://fumachi.mat.br/2022/11/09/teste-de-hipoteses/>. Acesso em: 26 jul. 2023.

GOIÁS. **Lei nº 18.679, de 26 de novembro de 2014**. Institui o Sistema Estadual de Cidadania Fiscal e o Programa de Cidadania Fiscal - Nota Fiscal Goiana, e dá outras providências. Goiânia: Palácio do Governo do Estado de Goiás, 3 dez. 2014. Disponível em: https://legisla.casacivil.go.gov.br/pesquisa_legislacao/90940/lei-18679. Acesso em: 28 abr. 2023.

GOIÁS. **Decreto nº 9440, de 2 de maio de 2019**. Dispõe sobre o Programa de Cidadania Fiscal - Nota Fiscal Goiana, instituído pela Lei nº 18.679, de 26 de novembro de 2014. Palácio do Governo do Estado de Goiás, 3 mai. 2019. Disponível em: <https://diariooficial.abc.go.gov.br/portal/visualizacoes/html/3791/#/p:3/e:3791>. Acesso em: 12 jul. 2023.

GOIÁS. **Decreto nº 9867, de 14 de maio de 2021**. Altera o Decreto nº 8.310, de 27 de janeiro de 2015, que regulamenta o Programa de Cidadania Fiscal - Nota Fiscal Goiana, instituído pela Lei nº 18.679, de 26 de novembro de 2014. Palácio do Governo do Estado de Goiás, 3 mai. 2019. Disponível em: <https://diariooficial.abc.go.gov.br/portal/visualizacoes/html/3791/#/p:3/e:3791>. Acesso em: 12 jul. 2023.

GOVERNO DE GOIÁS (Goiás). Secretaria da Economia. **Perguntas e Respostas**. Goiás, 11 jan. 2021. Disponível em: <https://www.economia.go.gov.br/nfgoiana/ajuda-nf.html>. Acesso em: 28 abr. 2023.

GOVERNO DE GOIÁS (Goiás). Secretaria da Economia. **Regulamento**. Goiás, 11 jan. 2021. Disponível em: <https://www.economia.go.gov.br/nfgoiana/regulamento.html>. Acesso em: 27 mai. 2023.

GRUS, Joel. **Data science from scratch: first principles with python**. O'Reilly Media, 2019.

IMHOFF, C.; WHITE, C. **Self-service business intelligence**. Empowering Users to Generate Insights, TDWI Best practices report, TWDI, Renton, WA, 2011.

KNAFLIC, C. N. **Storytelling com dados: Um guia sobre visualização de dados para profissionais de negócios**. Rio de Janeiro: Alta Books, 2019.

KROTOV, V.; SILVA, L. Legality and ethics of web scraping. *In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS*, 24., 2018, New Orleans. **Proceedings** [...]. New Orleans: ERF, 2018.

KROTOV, V.; TENNYSON, M.. Research note: scraping financial data from the web using the R language. **Journal of Emerging Technologies in Accounting**, v. 15, n. 1, p. 169-181, 2018. doi: <https://doi.org/10.2308/jeta-52063>.

LU, Mengwei. **Discovering Microsoft Self-service BI solution** : Power BI. 2014. Haaga-Helia - University of Applied Sciences.

MAGALHÃES, Marcos Nascimento; DE LIMA, Antônio Carlos Pedroso. **Noções de probabilidade e estatística**. Editora da Universidade de São Paulo, 2002.

MAGRANI, E. **Entre dados e robôs: ética e privacidade na era da hiperconectividade**. Porto Alegre: Arquipélago Editorial, 2019.

MARCHIONINI, G. **Information science roles in the emerging field of data science**. Journal of Data and Information Science, Beijing, v. 1, n. 2, p. 1-6, 2016. Disponível em: http://manu47.magtech.com.cn/Jwk3_jdis/Y2016/V1/I2/1.

MITCHELL, Ryan. **Web scraping with Python: Collecting more data from the modern web**. " O'Reilly Media, Inc.", 2018.

PAULA, L. P.; CORDEIRO, D. F. Políticas de segurança da informação em instituições públicas. **RESIGET - Revista Eletrônica de Sistemas de Informação e Gestão Tecnológica**, v. 6, n. 2, p. 58-69, 2015. Disponível em: <http://periodicos.unifacef.com.br/resiget/article/view/1046>. Acesso em: 31 de out. de 2023.

SEBRAE. **A importância da emissão da nota fiscal**. In: SEBRAE. **SEBRAE**. [S.l.]. 16 jan. 2023. Disponível em: <https://sebrae.com.br/sites/PortalSebrae/artigos/a-importancia-da-emissao-da-nota-fiscal,857ed6387eab5810VgnVCM1000001b00320aRCRD>. Acesso em: 2 jun. 2023.

SECRETARIA DE ESTADO DA ECONOMIA - GOVERNO DO ESTADO DE GOIÁS [SEE-GO]. **ICMS - Imposto sobre Circulação de Mercadorias e Serviços**. s. d. Disponível em: <https://www.economia.go.gov.br/receita-estadual/icms.html>. Acesso em: 12 de jul. de 2023.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4. Ed. Porto Alegre: Bookman, 2019.

VIEIRA, L. M.; CORDEIRO, D. F. The dark side of anti-vaccination: analysis of a brazilian anti-vaccine Facebook group. **Revista Famecos**, v. 30, n. 1, e43710, 2023. doi: <https://doi.org/10.15448/1980-3729.2023.1.43710>.

VIRTANEN, Pauli et al. **SciPy 1.0: fundamental algorithms for scientific computing in Python**. Nature methods, v. 17, n. 3, p. 261-272, 2020.

Notas de fim

¹ <https://devmedeiros.com/nota-fiscal-goiana>.