



Instituto Federal da Bahia
Departamento de Computação

Programa de Pós-Graduação em Engenharia de Sistemas de Produtos

**VISUALDATA-NFE: UMA SOLUÇÃO BIG
DATA DE NOTAS FISCAIS ELETRÔNICAS
PARA ANÁLISE VISUAL DE PERFIL
ECONÔMICO DE MUNICÍPIOS**

Evilasio de Sousa Junior

DISSERTAÇÃO DE MESTRADO

Salvador
29 de fevereiro de 2024

EVILASIO DE SOUSA JUNIOR

**VISUALDATA-NFE: UMA SOLUÇÃO BIG DATA DE NOTAS
FISCAIS ELETRÔNICAS PARA ANÁLISE VISUAL DE PERFIL
ECONÔMICO DE MUNICÍPIOS**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Engenharia de Sistemas de Produtos da Instituto Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Engenharia de Sistemas e Produtos.

Orientador: Prof. Dr. Renato Lima Novais

Salvador
29 de fevereiro de 2024

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS DO IFBA, COM OS
DADOS FORNECIDOS PELO(A) AUTOR(A)

S729v Sousa Junior, Evilasio de

Visualdata-NFE: uma solução big data de notas
fiscais eletrônicas para análise visual de perfil
econômico de municípios / Evilasio de Sousa Junior;
orientador Renato Lima Novais -- Salvador, 2024.

95 p.

Dissertação (Programa de Pós-Graduação em
Engenharia de Sistemas de Produtos) -- Instituto
Federal da Bahia, 2024.

1. Big data. 2. Notas fiscais eletrônicas. 3.
Análise visual. 4. Perfil econômico. I. Novais,
Renato Lima, orient. II. TÍTULO.

CDU 657.243.1

INSTITUTO FEDERAL DA BAHIA
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE SISTEMAS E
PRODUTOS – PPGESP

VISUALDATA-NFE: UMA SOLUÇÃO BIG DATA DE NOTAS FISCAIS ELETRÔNICAS PARA ANÁLISE
VISUAL DE PERFIL ECONÔMICO DE MUNICÍPIOS

EVILÁSIO DE SOUSA JÚNIOR

Produto(s) Gerado(s): Dissertação e Manual de Operação Técnica.

Orientador: Prof. Dr. Renato Lima Novais

Banca examinadora:

Prof. Dr. Renato Lima Novais
Orientador – Instituto Federal da Bahia (IFBA)

Prof. Dr. Fabrício Pitombo Leite
Membro Externo – Universidade Federal da Bahia (UFBA)

Prof. Dr. Cleber Jorge Lira de Santana
Membro Interno – Instituto Federal da Bahia (IFBA)

Trabalho de Conclusão de Curso aprovado pela banca examinadora em 29/02//2024



Documento assinado eletronicamente por **RENATO LIMA NOVAIS, Docente Permanente**, em 01/03/2024, às 15:39, conforme decreto nº 8.539/2015.



Documento assinado eletronicamente por **CLEBER JORGE LIRA DE SANTANA, Docente Permanente**, em 01/03/2024, às 15:44, conforme decreto nº 8.539/2015.



Documento assinado eletronicamente por **Fabrício Pitombo Leite, Usuário Externo**, em 04/03/2024, às 11:00, conforme decreto nº 8.539/2015.



A autenticidade do documento pode ser conferida no site http://sei.ifba.edu.br/sei/controlador_externo.php?acao=documento_conferir&acao_origem=documento_conferir&id_orgao_acesso_externo=0 informando o código verificador **3409252** e o código CRC **1561DB8D**.

À Lukinhas Ribeiro (in memoriam), amigo por quase 20 anos, e que pelo destino ou pelos eventos do acaso, teve a existência abreviada... Sempre presente com alegria!

AGRADECIMENTOS

Orgulho de ser o que somos, de falar sobre o que falamos, de ser raiz! É a raiz que possibilita o caule, os galhos e as folhas. Inspirado por Mateus Aleluia, agradeço a Milena Vergne, companheira diária! Desde as batalhas para existirmos e resistirmos, pelo resgate e construção de identidades, até às discussões filosóficas e aos paradoxos existenciais. Por tudo!

A esperança da humanidade está nessas crianças e nas infâncias saudáveis e felizes. Inspirado pelo Prof. José Jerônimo de Moraes, agradeço a Leonardo Vergne, pequeno parceiro de tudo, filho que reconfigurou a minha vida e atualizou as definições de amor na minha existência. É como ter o próprio coração batendo fora do corpo.

É impossível medir a contribuição de cada pessoa no processo histórico da humanidade e a propriedade intelectual pertence à humanidade. Inspirado por Piotr Kropotkin, agradeço ao professor Renato Novais, por compartilhar tanto de si, pelas trocas, pelas descobertas e por ter sido o principal guia nessa jornada que revelou tantos caminhos e tantos aprendizados.

Às minhas mães, de sangue e de coração que a vida, tão gentilmente, me trouxe. Elas iluminaram e ainda iluminam os pensamentos e caminhos desta jornada. Dalva, Maria da Conceição, Lene e Luisa (Del - in memoriam) que comumente me chamava de filho ingrato!

Aos compadres, comadres, irmãs, irmãos, amigas e amigos que também fizeram parte deste percurso, seja com palavras, com gestos, com pensamentos positivos e também com muito amor.

Viver sem compreender a natureza é como adquirir um bilhete falso para um trem com destino a lugar nenhum. Inspirado por Ailton Krenak, agradeço a Maurício Simões, irmão que a vida revelou e que provoca profundas reflexões e tão transformadoras em mim.

Aos professores Ricardo Vieira, Manoel Mendonça Neto, Cleber Santana e Fabrício Pitombo pelas importantes contribuições ao longo desta pesquisa.

AGRADECIMENTOS

A Jadson Bitencourt e equipe da SEFAZ-BA, bem como a Júlia Trindade e a equipe da SDE-BA pelas importantes contribuições durante o processo de construção do produto computacional, parte desta pesquisa.

Ao IFBA - Instituto Federal da Bahia, pela oportunidade de ter ingressado no PPGESP, programa de mestrado de um Instituto Federal de Educação, Ciência e Tecnologia, podendo afirmar a qual raça eu pertença, por fortalecer a minha identidade e por possibilitar ocupar esses espaços que por séculos foram sistematicamente negados ao meu povo, o povo negro! Ainda assim, há muito a se fazer!

'Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda'

Paulo Freire

"O instante do despertar é o instante mais perigoso do dia. . ."

Franz Kafka

—'

RESUMO

Gestores públicos e agentes do setor privado buscam, dentro de um sistema capitalista de produção, o acúmulo de capital e a compreensão dos aspectos socioeconômicos da região onde atuam, para tomarem decisões e encontrar caminhos para o desenvolvimento econômico. Dada a importância do tema, há iniciativas de análise do perfil socioeconômico e também produtivo de municípios e regiões. Essa compreensão visa identificar, dentre outras, de onde o município adquire mercadorias, para quais municípios vende, qual o motivo pelo qual se compra ou vende determinada mercadoria, qual o volume e qual a sazonalidade da aquisição ou venda. A tomada de decisão baseada em dados, geralmente censitários, quase sempre desatualizados e que não refletem a dinâmica econômica nem o perfil atual dos municípios, torna-se imprecisa. Neste trabalho, é apresentado o **VisualData-NFe**: solução de *Big Data* para análise e exploração visual do perfil produtivo e da dinâmica econômica de municípios a partir de uma massa de dados extensa, rica e atual. A solução desenvolvida busca apoiar a tomada de decisão em relação a investimentos, utilizando-se dados de notas fiscais eletrônicas de compra e venda de mercadorias.

Foram realizados estudos experimentais para validação da solução. Neste contexto, com a base que foi disponibilizada e com o ambiente que foi construído, notou-se que as percepções sobre os perfis produtivos foram obtidas e puderam ser executadas por pessoas não especialistas.

Palavras-chave: Big Data. Notas Fiscais Eletrônicas. Análise Visual. Perfil Econômico.

ABSTRACT

Public managers and private sector agents seek, within a capitalist system of production, the accumulation of capital and the understanding of the socioeconomic aspects of the region where they operate, to make decisions and find ways for economic development.

Given the importance of the theme, there are initiatives to analyze the socioeconomic and also productive profile of municipalities and regions. This understanding aims to identify, among others, where the municipality acquires goods, to which municipalities it sells, what the reason for the purchase or sale of a particular commodity, what the volume and what the seasonality of the acquisition or sale.

Decision-making based on data, usually census data, almost always outdated and that do not reflect the economic dynamics or the current profile of municipalities, becomes inaccurate.

In this work, **VisualData-NFe** is presented: Big Data solution for analysis and visual exploration of the productive profile and economic dynamics of municipalities from an extensive, rich and current mass of data. The proposed solution seeks to support decision making in relation to investments, using data from electronic invoices for the purchase and sale of goods.

Experimental studies were carried out to validate the solution. In this context, with the basis that was made available and with the environment that was built, the perceptions about the productive profiles were obtained and could be performed by non-specialists.

Keywords: Big Data. Electronic Invoices. Visual analysis. Economic profile.

RÉSUMÉ

Les gestionnaires publics et les agents du secteur privé recherchent, au sein d'un système capitaliste de production, l'accumulation de capital et la compréhension des aspects socio-économiques de la région où ils opèrent, pour prendre des décisions et trouver des voies de développement économique.

Compte tenu de l'importance du thème, il existe des initiatives d'analyse du profil socio-économique et aussi productif des municipalités et des régions. Cette compréhension vise à identifier, entre autres, d'où la municipalité achète des marchandises, pour quelles municipalités elle vend, pourquoi elle achète ou vend une marchandise, quel est le volume et quelle est la saisonnalité de l'achat ou de la vente.

La prise de décision basée sur des données, généralement des données de recensement, souvent obsolètes et ne reflétant pas la dynamique économique ou le profil actuel des municipalités, devient imprécise.

Dans ce travail, **VisualData-NFe** est présenté: une solution Big Data pour l'analyse et l'exploration visuelle du profil productif et de la dynamique économique des municipalités à partir d'une masse de données vaste, riche et actuelle. La solution proposée vise à soutenir la prise de décision concernant les investissements en utilisant les données des factures électroniques d'achat et de vente de marchandises.

Des études expérimentales ont été réalisées pour valider la solution et, dans ce contexte, sur la base de ce qui a été fourni et avec l'environnement qui a été construit, les perceptions sur les profils de production ont été obtenues et ont pu être réalisées par des non-spécialistes.

Mots-clés: Big Data. Factures électroniques. Analyse visuelle. Profil économique.

LISTA DE FIGURAS

2.1	Sistemática do NCM	32
2.2	Cadastro Nacional de Pessoa Jurídica	33
2.3	CFOP	34
3.1	Visão Geral	50
3.2	Desenho da Solução	51
3.3	NCM API-REST x SISCOMEX	53
3.4	SEFAZ-BA CFOP em PDF	54
3.5	PIB per capita - Bahia	55
3.6	Área e População - Bahia	56
3.7	Dados de Localização - Salvador - BA	57
3.8	Dashboard - VisualData-NFe	59
3.9	Formas de Filtragem	60
3.10	5 UFs que mais venderam para a Bahia - Total Bruto (R\$) Visualização e Detalhamento	61
3.11	10 Municípios que mais venderam para a Bahia - Total Bruto (R\$) Visualização e Detalhamento	61
3.12	10 Municípios que mais compraram na Bahia - Total Bruto (R\$) Visualização e Detalhamento	62
3.13	Mercadorias vendidas por Ilhéus (BA) - Total Bruto	62
3.14	Atividade Econômicas - Vendas por Ilhéus (BA) - Total ICMS	63
4.1	Total Compra - Série Histórica	67
4.2	5 maiores UF Emitentes - Total Bruto e ICMS	68
4.3	Total Compra das UFs PE, MG, RJ e PR	69
4.4	Maiores Municípios Emitentes - SP para BA	70
4.5	Municípios Destinatários - BA	71
4.6	Detalhe - Municípios Destinatários - BA	72
4.7	Municípios que mais venderam a Camaçari-BA	73
4.8	Mercadorias de Taubaté-SP para Camaçari-BA - Total Bruto (%)	73
4.9	Atividades Econômicas - Mercadorias de Camaçari-BA oriundas de Taubaté-SP	75
4.10	Total Bruto e Total de ICMS	76
4.11	Operações Fiscais	77
4.12	UFs Emitentes - Fios de ferro galvanizados	78
4.13	Detalhe - UFs Emitentes - Fios de ferro galvanizados	79

LISTA DE FIGURAS

5.1	Perfis	86
5.2	Resultados - Escala Likert	88

LISTA DE TABELAS

2.1	Dados do Município de Abaíra - BA	35
3.1	Histórias de Usuário	49
4.1	Detalhe - Mercadorias de Taubaté-SP para Camaçari-BA	74
4.2	Métricas para Q1	79
4.3	Métricas para Q2	80
4.4	Métricas para Q3	80
4.5	Métricas para Q4	80
4.6	Métricas para Q5	81
4.7	Métricas para Q6	81
5.1	Resultados	87

LISTA DE SIGLAS

API-REST	<i>Application Programming Interface - Representational State Transfer</i>	52
CFOP	Código Fiscal de Operações e Prestações	54
CNAE	Classificação Nacional de Atividades Econômicas	32
CONFAZ	Conselho Nacional de Política Fazendária	32
CSV	<i>Comma-separated Values</i>	
ETL	Extração-Transformação-Carga	52
GQM	<i>Goal-Question-Metric</i>	39
IBGE	Instituto Brasileiro de Geografia e Estatística	34
ICM	Imposto sobre Circulação de Mercadorias	34
ICMS	Imposto sobre operações relativas à Circulação de Mercadorias e sobre prestações de Serviços de transporte interestadual e intermunicipal e de comunicação	33
IEC	<i>International Electrotechnical Commission</i>	66
IHC	Interação Humano-Computador	38
IPEA	Instituto de Pesquisa Econômica Aplicada	27
IPI	Imposto sobre Produtos Industrializados	33
ISO	<i>International Organization for Standardization</i>	66
JSON	<i>JavaScript Object Notation</i>	66
KDT	<i>Knowledge Discovery from Text</i>	90
KQL	<i>Kibana Query Language</i>	60
NCM	Nomenclatura Comum do Mercosul	52
NF-e	Nota Fiscal Eletrônica	31
PDF	<i>Portable Document Format</i>	54
PIB	Produto Interno Bruto	35
Query-DSL	<i>Query Domain Specific Language</i>	60
RFB	Receita Federal do Brasil	52
SEFAZ-BA	Secretaria da Fazenda do Estado da Bahia	54
SGBD	Sistema de Gerenciamento de Banco de Dados	44
SISCOMEX	Sistema Integrado de Comércio Exterior	52

LISTA DE SIGLAS

SSPL	<i>Server Side Public License</i>	41
URL	<i>Uniform Resource Locator</i>	52
VPS	<i>Virtual Private Server</i>	66

SUMÁRIO

Capítulo 1—Introdução	27
1.1 Contexto e Motivação	27
1.2 Problema e Justificativa	28
1.3 Objetivos	28
1.3.1 Objetivo Geral	28
1.3.2 Objetivos Específicos	28
1.4 Procedimentos Metodológicos	29
1.4.1 Caracterização e Natureza da Pesquisa	29
1.4.2 Cenário de Pesquisa	29
1.4.3 Método	29
1.4.4 Validação	30
1.5 Organização do Texto	30
Capítulo 2—Revisão Bibliográfica	31
2.1 Fundamentação Teórica	31
2.1.1 Enquadramento Conceitual	31
2.1.1.1 Notas Fiscais Eletrônicas e seus dados	31
2.1.1.2 Dados Geográficos e Estatísticos	35
2.1.1.3 Análise de Insumo-Produto	35
2.1.1.4 <i>Big Data</i>	36
2.1.1.5 <i>Visual Analytics</i>	37
2.1.1.6 Painéis de Visualização de Informação - <i>Dashboards</i>	38
2.1.2 Enquadramento Tecnológico	40
2.1.2.1 Mecanismos de Pesquisa: Elasticsearch e Apache Lucene	40
2.1.2.2 Análise Visual Exploratória no Kibana	42
2.1.2.3 Bancos de Dados Não-Relacionais e o MongoDB	43
2.2 Trabalhos Relacionados	44
Capítulo 3—Produto Computacional - VisualData-NFe	47
3.1 Método	47
3.2 Mapeamento das necessidades através de histórias de usuários	48
3.3 Definição da arquitetura da solução e da visão geral da proposta	50
3.4 Desenvolvimento do VisualData-NFe	52
3.4.1 Extração, Transformação, Carga e Indexação dos Dados	52
3.4.1.1 Nomenclatura Comum do Mercosul (NCM)	52

3.4.1.2	Código Fiscal de Operações e Prestações (CFOP)	54
3.4.1.3	Dados econômicos, geográficos e estatísticos	54
3.4.1.4	Carga e Indexação dos Dados de Notas Fiscais Eletrônicas	57
3.4.2	<i>Visualizações</i>	58
3.4.2.1	Pesquisa e Filtragem	60
3.4.2.2	Área de Notas Fiscais	60
3.4.2.3	Área de Mercadorias	62
3.4.2.4	Área de Atividades Econômicas e Operações Fiscais . . .	63
3.5	Desafios Encontrados	63
Capítulo 4—Estudo Exploratório I		65
4.1	Objetivo	65
4.2	Método	65
4.2.1	Definição do Ambiente Experimental	66
4.2.2	Escolha das Métricas	66
4.2.3	Elaboração das Questões e Medições	67
4.2.4	Coleta e Análise de Dados	67
4.3	Resultados	79
4.4	Conclusão	81
Capítulo 5—Estudo Exploratório II		83
5.1	Objetivo	83
5.2	Método	83
5.2.1	Definição do Ambiente Experimental	84
5.2.2	Elaboração da Avaliação da Usabilidade do VisualData-NFe <i>Dashboard</i>	84
5.2.3	Coleta e Análise dos Dados	86
5.3	Resultados	87
5.4	Conclusão	88
Capítulo 6—Considerações Finais		89
6.1	Considerações	89
6.2	Limitações e Trabalhos Futuros	89
Apêndice A—Manual VisualData-NFe		95

Neste capítulo são apresentados o contexto e a motivação para a realização deste trabalho, bem como o problema e a justificativa do estudo. São apresentados também os objetivos, os resultados esperados, os procedimentos metodológicos e as limitações da pesquisa. Por fim, uma apresentação da organização do texto.

INTRODUÇÃO

1.1 CONTEXTO E MOTIVAÇÃO

A dinâmica econômica em uma sociedade capitalista pode ser vista como um processo histórico em constante evolução, onde forças produtivas como tecnologia, conhecimento e equipamentos, podem transformar as relações sociais.

Gestores públicos e agentes do setor privado buscam a compreensão dos aspectos socioeconômicos da região onde atuam, para tomarem decisões e encontrarem caminhos para o desenvolvimento econômico social e dos seus próprios negócios.

De acordo com o estudo de Sperotto et al. (2017), publicado pelo Instituto de Pesquisa Econômica Aplicada (IPEA), em termos macroeconômicos, a variável investimento assume o papel de estímulo ao crescimento econômico. A ampliação da capacidade produtiva exige contratação de mão de obra, aquisição de máquinas, ferramentas e insumos, recolhimento de impostos e possibilita a ampliação do acesso a bens e serviços. Consequentemente, provoca a melhoria da qualidade de vida, e o estudo citado anteriormente associa também a variável investimento ao **dinamismo econômico** e apresenta como “determinante” que as decisões sobre investimentos considerem a oscilação e os cenários conjunturais atuais.

Segundo Gereffi e Fernandez-Stark (2016), o **perfil econômico** de um município ou região pode ser analisado através do fluxo de mercadorias. A análise do fluxo visa compreender de onde o município adquire mercadorias, para quais municípios vende, qual o motivo pelo qual se compra ou vende determinada mercadoria, qual o volume e qual a sazonalidade da aquisição ou venda.

Dada a importância do tema, há iniciativas de análise do perfil socioeconômico e também produtivo de municípios e regiões. A exemplo do Instituto Brasileiro de Geografia e Estatística - IBGE (2017), que utilizou dados do censo agropecuário para investigar informações sobre os estabelecimentos e as atividades agropecuárias neles desenvolvidas.

Essas informações abrangeram características do produtor e do estabelecimento, economia e emprego no meio rural, pecuária, lavoura e agroindústria. Este estudo refere-se ao ano-safra 2017, que estendeu-se entre outubro de 2016 e setembro de 2017.

1.2 PROBLEMA E JUSTIFICATIVA

Ter acesso a análises de perfil econômico pode apoiar na tomada de decisão em relação a investimentos, entretanto é preciso destacar que essas análises em momentos específicos do tempo podem não retratar a situação atual de um local. Portanto, a tomada de decisão baseada em dados desatualizados reduz a precisão e pode não refletir a conjuntura recente da dinâmica econômica, tampouco o perfil produtivo de municípios e regiões.

O que se vende e compra em um município é registrado, em boa parte, pelas notas fiscais. De acordo com a Receita Federal do Brasil (RFB, 2023), até 03 de fevereiro de 2024, tinham sido emitidas 40,87 bilhões de notas fiscais eletrônicas. Através das informações contidas nessa enorme massa de dados, é possível identificar quais mercadorias têm sido adquiridas e comercializadas pelos municípios. Sob essa perspectiva, formulou-se a hipótese que ao analisar os dados de notas fiscais eletrônicas em painéis de exploração visual pode-se proporcionar compreensão sobre a dinâmica econômica e sobre o perfil produtivo de municípios e responder a pesquisas em grandes volumes de dados através de mecanismos de busca textual.

Foi possível identificar uma lacuna existente bem como perceber desafios para tratar grande volume de dados de notas fiscais eletrônicas, relacioná-lo com dados de perfil econômico e disponibilizá-lo em uma ferramenta interativa de análise visual exploratória que possibilite a descoberta de conhecimento.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Desenvolver solução computacional capaz de oferecer mecanismos de análise e exploração visual, que permita a compreensão sobre a dinâmica econômica e sobre o perfil produtivo de municípios, tomando como base pesquisas textuais em grandes volumes de dados de notas fiscais eletrônicas.

1.3.2 Objetivos Específicos

- Coletar as necessidades levantadas pelos usuários reais;
- Apresentar protótipo funcional da solução computacional desenvolvida para usuários reais;
- Validar a eficácia da solução desenvolvida através de estudo exploratório;
- Validar a usabilidade da solução desenvolvida através de estudo exploratório.

1.4 PROCEDIMENTOS METODOLÓGICOS

1.4.1 Caracterização e Natureza da Pesquisa

Quanto à natureza, essa pesquisa caracteriza-se como bibliográfica com revisão da literatura. Com o intuito de compreender o contexto da situação estudada, a abordagem dada à pesquisa caracteriza-se como quantitativa e qualitativa. Já quanto aos objetivos, trata-se de uma pesquisa descritiva, por procurar descrever as características de determinada população, fenômeno e/ou estabelecimento de relações entre variáveis, cabendo ao pesquisador o estudo e a interpretação acurada dos fatos de maneira imparcial (GIL, 2008).

Quanto aos procedimentos técnicos, essa pesquisa caracteriza-se como bibliográfica e documental, pois implica não somente no estudo de artigos, teses, livros e outras publicações como revistas, anais e relatórios, mas também na análise de dados que ainda não foram sistematizados, como arquivos obtidos em órgãos públicos e bancos de dados (WAZLAWICK, 2009).

Destaca-se que esta é uma pesquisa de natureza aplicada que, segundo Gil (2008), apresenta muitos pontos semelhantes à pesquisa pura, pois depende de suas descobertas e se enriquece com o seu desenvolvimento. No entanto, tem-se como principal característica o interesse na aplicação, utilização e consequências práticas dos conhecimentos.

O alcance do objetivo geral desta pesquisa foi concretizado através da implementação de uma solução computacional com desenvolvimento de software. Para tal, utilizou-se a aplicação de boas práticas de métodos ágeis (MARTIN; NEWKIRK; KOSS, 2003), que se desdobra em etapas desenvolvidas de forma incremental e iterativa.

1.4.2 Cenário de Pesquisa

De acordo com Gaba (2004), a simulação é uma técnica para substituir ou amplificar experiências reais com experiências guiadas a fim de evocar ou replicar aspectos substanciais do mundo real de uma forma controlada e interativa.

Para o presente trabalho, utilizou-se de dados sintéticos de notas fiscais, **fictícios, realísticos e anonimizados** de compra e venda de mercadorias do Estado da Bahia. Ao total, foram geradas cerca de trinta e quatro milhões de dados de notas fiscais eletrônicas de pessoas jurídicas, de operações de compra e venda com municípios do estado da Bahia.

1.4.3 Método

Na busca por respostas ao problema de pesquisa, este trabalho foi conduzido por um percurso baseado em métodos ágeis e as etapas foram realizadas de forma incremental e iterativa, descritas a seguir:

- Revisão bibliográfica e identificação de uma lacuna e um problema de pesquisa;
- Mapeamento das necessidades, planejamento e desenvolvimento de solução computacional para o problema identificado;

- Realização de um estudo experimental para avaliar a eficácia do produto computacional;
- Realização de um estudo experimental para avaliar a usabilidade da interface do produto computacional;
- Apresentação dos resultados e considerações.

1.4.4 Validação

A verificação do alcance dos objetivos poderá ser realizada através da análise dos resultados dos estudos experimentais apresentados nos Capítulos 4 e 5.

1.5 ORGANIZAÇÃO DO TEXTO

Além desta introdução, este trabalho está organizado da seguinte forma:

- O Capítulo 2 apresenta o referencial teórico e a pesquisa sobre trabalhos relacionados;
- O Capítulo 3 descreve o planejamento do desenvolvimento, o mapeamento das necessidades e as etapas do processo de construção do produto computacional VisualData-NFe;
- Os Capítulos 4 e 5 apresentam os estudos experimentais;
- O Capítulo 6 reúne as considerações finais, traz as contribuições para a área de pesquisa, cita as limitações e sugere trabalhos futuros.
- O Apêndice A traz o manual do usuário do VisualData-NFe.

Este capítulo traz a fundamentação teórica relacionada aos assuntos abordados ao longo da pesquisa e foi dividida em enquadramento conceitual e enquadramento tecnológico. Ao final, são relacionados os trabalhos correlatos ao problema desta pesquisa.

REVISÃO BIBLIOGRÁFICA

2.1 FUNDAMENTAÇÃO TEÓRICA

2.1.1 Enquadramento Conceitual

Esta seção traz conceitos relacionados às notas fiscais eletrônicas, dados geográficos e estatísticos, grandes volumes de dados e análise visual exploratória.

2.1.1.1 Notas Fiscais Eletrônicas e seus dados

Esta subseção trata das notas fiscais eletrônicas, documento fiscal que registra a compra e venda de bens ou serviços. Traz alguns conceitos sobre os campos de uma nota fiscal eletrônica que são relevantes para essa pesquisa.

O projeto que estabeleceu a Nota Fiscal Eletrônica (NF-e), segundo a RFB (2022c), integrou as Secretarias de Fazenda dos Estados e a Receita Federal do Brasil e buscava possibilitar:

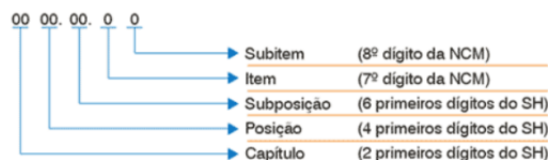
- Melhor intercâmbio e compartilhamento de informações entre os fiscos;
- Redução de custos e entraves burocráticos, facilitando o cumprimento das obrigações tributárias e o pagamento de impostos e contribuições;
- Fortalecimento do controle e da fiscalização.

Dentre os benefícios, a NF-e oportunizou a rapidez no acesso às informações e o cruzamento eletrônico desses dados. A exemplo das mercadorias, que são representadas por um código associado à **Nomenclatura Comum do Mercosul (NCM)**. Essa nomenclatura é utilizada pelos países-membros do Mercosul e o código é composto por oito dígitos. Um exemplo é o NCM 0406.20.00 que representa “queijos ralados ou em pó, de qualquer tipo” (SISCOMEX, 2022).

Cada mercadoria tem a sua classificação estruturada conforme Figura 2.1 a seguir:

Figura 2.1 Sistemática do NCM

A sistemática de classificação dos códigos na Nomenclatura Comum do Mercosul (NCM) obedece à seguinte estrutura:



Exemplo:

Código Produtos/Serviços: 0104.10.11

Animais reprodutores de raça pura, da espécie ovina, prenhe ou com cria ao pé

Esse código é resultado dos seguintes desdobramentos:

Seção	I	Animais vivos e produtos do reino animal
Capítulo	01	Animais vivos
Posição	0104	Animais vivos das espécies ovina e caprina
Subposição	0104.10	Ovinos
Item	0104.10.1	Reprodutores de raça pura
Subitem	0104.10.11	Prenhe ou com cria ao pé

Fonte: RFB, 2022.

Entre os dias 14 e 15 de dezembro de 1970, o Conselho Nacional de Política Fazendária (CONFAZ), formado pelo ministro da fazenda e os secretários de fazenda e finanças dos Estados e do Distrito Federal, definiu um acordo para criar o Sistema Nacional Integrado de Informações Econômico-Fiscais. Foi estabelecido então pelo CONFAZ (1970) que as unidades federadas adotariam códigos que representassem a **Classificação Nacional de Atividades Econômicas (CNAE)**. O código CNAE consta na NF-e, e de acordo com o IBGE (2022), é a classificação oficialmente adotada na produção de estatísticas por tipo de atividade econômica e na identificação da atividade econômica em cadastros e registros de pessoas jurídicas. Como exemplo de atividade econômica, temos a fabricação de produtos do refino de petróleo, que em uma NF-e teria o código CNAE 1921-7/00.

A Figura 2.2 a seguir mostra o cadastro de uma pessoa jurídica, e a escolha dentre as atividades econômicas disponíveis.

Figura 2.2 Cadastro Nacional de Pessoa Jurídica

Fazenda
Ministério da Fazenda

Comprovante de Inscrição e de Situação Cadastral

Contribuinte,
Confira os dados de Identificação da Pessoa Jurídica e, se houver qualquer divergência, providencie junto à RFB a sua atualização cadastral.

REPÚBLICA FEDERATIVA DO BRASIL
CADASTRO NACIONAL DA PESSOA JURÍDICA

NÚMERO DE INSCRIÇÃO: [REDACTED] DATA DE ABERTURA: 17/08/2000
MATRIZ: [REDACTED] COMPROVANTE DE INSCRIÇÃO E DE SITUAÇÃO CADASTRAL
NOME EMPRESARIAL: [REDACTED] **CNAE**

TÍTULO DO ESTABELECIMENTO (NOME DE FANTASIA): [REDACTED]

CÓDIGO E DESCRIÇÃO DA ATIVIDADE ECONÔMICA PRINCIPAL
49.30-2-02 - Transporte rodoviário de carga, exceto produtos perigosos e mudanças, intermunicipal, interestadual e internacional

CÓDIGO E DESCRIÇÃO DAS ATIVIDADES ECONÔMICAS SECUNDÁRIAS
49.30-2-01 - Transporte rodoviário de carga, exceto produtos perigosos e mudanças, municipal
52.50-8-03 - Agenciamento de cargas, exceto para o transporte marítimo
52.50-8-05 - Operador de transporte multimodal - OTM
52.11-7-99 - Depósitos de mercadorias para terceiros, exceto armazéns gerais e guarda-móveis
49.30-2-03 - Transporte rodoviário de produtos perigosos

CÓDIGO E DESCRIÇÃO DA NATUREZA JURÍDICA
206-2 - SOCIEDADE EMPRESARIAL LIMITADA

LOGRADOURO: R CM 14 NÚMERO: 201 COMPLEMENTO: QD 3-A LT 06/11 UF: GO
CEP: 74.463-280 BAIRRO/DISTRITO: CANDIDA DE MORAES MUNICÍPIO: GOIÂNIA

SITUAÇÃO CADASTRAL: ATIVA DATA DA SITUAÇÃO CADASTRAL: 03/11/2005
MOTIVO DE SITUAÇÃO CADASTRAL: [REDACTED]

SITUAÇÃO ESPECIAL: [REDACTED] DATA DA SITUAÇÃO ESPECIAL: [REDACTED]

Fonte: Ministério da Fazenda, 2022.

O CONFAZ (1970), a fim de identificar a natureza da circulação de uma mercadoria ou prestação de serviço de transportes, estabeleceu o **Código Fiscal de Operações e Prestações (CFOP)**. Desta forma, através de códigos de quatro dígitos, é fixada a tributação, e de que forma ela irá acontecer. A circulação de mercadorias ou prestação de serviço de transportes podem ter operações de entrada e saída dentro do próprio Estado, de outros estados ou do exterior. O CFOP 1.102, por exemplo, identifica uma compra de produtos para comercialização, com origem e destino no mesmo estado, e sem substituição tributária. Os CFOPs constam na NF-e e visam aglutinar em grupos homogêneos as operações e prestações realizadas pelos contribuintes do Imposto sobre Produtos Industrializados (IPI) e do Imposto sobre operações relativas à Circulação de Mercadorias e sobre prestações de Serviços de transporte interestadual e intermunicipal e de comunicação (ICMS).

O CONFAZ (1970) definiu a estrutura do CFOP da seguinte forma: o primeiro algarismo refere-se ao trajeto que será percorrido pela mercadoria. Se for uma entrada, seguirá o seguinte padrão:

- 1.000 – ENTRADA E/OU AQUISIÇÕES DE SERVIÇOS DO ESTADO

- 2.000 – ENTRADA E/OU AQUISIÇÕES DE SERVIÇOS DE OUTROS ESTADOS
- 3.000 – ENTRADA E/OU AQUISIÇÕES DE SERVIÇOS DO EXTERIOR

E há o seguinte padrão para saídas:

- 5.000 – SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA O ESTADO
- 6.000 – SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA OUTROS ESTADOS
- 7.000 – SAÍDAS OU PRESTAÇÕES DE SERVIÇOS PARA O EXTERIOR

O segundo algarismo é estabelecido de acordo com a Figura 2.3 a seguir

Figura 2.3 CFOP

2º Algarismo do CFOP - Operação de Entrada

CFOP = x102 - x202 - x302 - x403 - x556 - x653 - x916
1- Compra / Industrialização / Transferências / Serviços
2- Devolução / Energia Elétrica
3- Serviços de comunicação / Serviços de Transporte
4- Mercadoria com substituição tributária / Integração
5- Exportação / Ativo Imobilizado / Material de Uso
6- Crédito / Ressarcimento do ICMS / Combustíveis
9- Outras entradas / Amostras Grátis / Bonificação Concerto / Exposição

Fonte: CEFIS, 2022.

Já os dois algarismos finais definem a situação tributária de cada produto referente ao ICMS cobrado pelos Estados.

Para Marques e Uchôa (2006), o Imposto sobre Circulação de Mercadorias (ICM) foi ratificado pela Constituição de 1988 como ICMS e esse imposto é, para os estados, a mais importante fonte de receita e objeto de preocupação dos seus governos. Para os municípios, as transferências do ICMS se constituem também como parte substancial de suas receitas. Os valores referentes ao ICMS também constam na NF-e, complementando informações sobre a circulação de mercadorias.

Essa circulação registrada na NF-e envolve **municípios** de origem e destino e o Instituto Brasileiro de Geografia e Estatística (IBGE) disponibiliza importantes dados acerca dos municípios.

2.1.1.2 Dados Geográficos e Estatísticos

As informações sobre os municípios brasileiros disponibilizadas pelo IBGE tomam sempre como referência um código composto de sete dígitos, sendo os dois primeiros referentes ao código da Unidade da Federação, a exemplo do município de Salvador, que tem código 2927408, onde 29 é código do estado da Bahia.

O IBGE disponibiliza também as informações sobre a unidade federativa e as regiões em que cada município está inserido. Dispõe também de informações como o centro geográfico (centróide), dimensão em quilômetros quadrados. É possível obter também do IBGE dados de um panorama geral, como por exemplo a população e o Produto Interno Bruto (PIB) *per Capita*, além de diversos outros índices e resultados de pesquisas. A seguir, a Tabela 2.1, com as informações sobre o município de Abaíra:

Tabela 2.1 Dados do Município de Abaíra - BA

Atributo	Informação
Código do Município	2900108
Nome	Abaíra
Região	Nordeste
UF	BA
Mesorregião	Centro Sul Baiano
Região Imediata	Brumado
Região Intermediária	Vitória da Conquista
Microrregião	Seabra
Longitude (Centróide)	-41.7429
Latitude (Centróide)	-13.2966
Área	538.677 Km ²
População (2010)	8.316
PIB Per Capita	7.734,24 R\$

Fonte: IBGE, 2022

2.1.1.3 Análise de Insumo-Produto

Gestores buscam apoio nas tomadas de decisão em nível regional, e dados regionalizados proporcionam essas percepções, seja uma regionalização por municípios, mesorregiões, microrregiões, ou ainda por divisões mais abrangentes, como os territórios de identidade.

Para que seja possível a realização de análises sobre a relação insumos e produtos, são produzidas e disponibilizadas as tabelas de recursos e usos que dão origem às matrizes Insumo-Produto.

Existem pesquisadores e gestores que trabalham com a análise da estrutura produtiva considerando dados que não têm atualizações frequentes, uma vez que consideram que a estrutura produtiva muda lentamente. Por outro lado, é importante considerar uma contextualização mais recente de uma região para se definir investimentos que têm grande impactos na sociedade. Este estudo se propõe a relacionar os conjuntos de dados permitindo a análise de cadeias produtivas através de dados atualizados.

2.1.1.4 Big Data

Em 2001, o termo *Big Data* é dado por Laney et al. (2001), e características essenciais sobre os dados foram propostas, denominadas três “Vs”: volume (que consiste em enormes quantidades de dados); velocidade (que versa sobre a criação de dados em tempo real), e variedade (classificando os dados em estruturados, semiestruturados e não estruturados).

Atualmente dois outros “Vs”: veracidade e valor são considerados e foram incluídos nos aspectos essenciais. Kitchin e McArdle (2015, 2016) reúnem e apresentam características complementares à qualificação de *Big Data*, tais como:

- Exaustividade: que trata da captura de sistemas inteiros em vez da captura de amostras;
- Veracidade: os dados podem ser desarrumados, com ruídos e podem conter incertezas e erros;
- Valor: muitos *insights* podem ser extraídos;
- Variabilidade: dados cujo significado pode ser constantemente alterado em relação ao contexto em que é gerado;
- Relacionalidade: que trata sobre campos comuns que permitem a junção de diferentes conjuntos de dados;
- Extensionalidade: poder adicionar ou alterar novos campos facilmente;
- Escalabilidade: poder expandir rapidamente em tamanho.

Chen e Liu (2014) referem-se a *Big Data* como conjuntos de dados grandes, diversos, complexos, longitudinais e/ou distribuídos e gerados a partir de instrumentos, sensores, transações na Internet, *e-mail*, vídeo, fluxos de cliques e tantas as outras fontes digitais disponíveis. Considera-se que esses conjuntos de dados não podem ser adquiridos, gerenciados e processados por tecnologias da informação e ferramentas de software/hardware tradicionais, dentro de um tempo tolerável.

Big Data é descrito por Cooper e Mell (2012), como a capacidade de extrair informações e *insights* onde anteriormente era economicamente ou tecnicamente impossível fazê-lo. O volume de dados, velocidade de aquisição ou representação de dados (variedade), limita a capacidade de realizar análises eficazes usando abordagens relacionais tradicionais ou requer o uso de dimensionamento horizontal significativo para processamento eficiente.

De acordo com o Statista (2021), a quantidade total de dados criados, capturados, copiados e consumidos globalmente aumentou rapidamente, atingindo 64,2 zettabytes em 2020. Estima-se que até 2025, a criação de dados globais deverá crescer para mais de 180 zettabytes. O crescimento foi maior do que o esperado anteriormente, causado pelo aumento da demanda devido à pandemia do COVID-19, pois mais pessoas trabalharam e aprenderam em casa e usaram as opções de entretenimento doméstico com mais frequência.

De acordo com Keim et al. (2008), o principal objetivo da área de *Visual Analytics* é a análise de grandes quantidades de dados para identificar e destilar visualmente o conteúdo de informações mais valioso e relevante. A representação visual deve revelar padrões estruturais e propriedades de dados relevantes para fácil percepção pelo analista. Estes conceitos serão abordados na subseção a seguir.

2.1.1.5 Visual Analytics

Keim et al. (2008) delineiam que existem importantes desafios científicos para a representação visual das informações e nível de detalhe para a análise visual. É preciso acomodar grandes fluxos de dados, e as soluções apropriadas precisam combinar, de forma inteligente, visualizações de detalhes das análises selecionadas.

Keim et al. (2008) definem análise visual como uma combinação de técnicas de análise automatizada com visualizações interativas que produzam uma compreensão eficaz, raciocínio e tomada de decisão com base em dados complexos e em grande volume. Afirmam ainda que o objetivo da análise visual é criar ferramentas e técnicas para permitir, entre outras, que as pessoas possam sintetizar informações, detectar o esperado, descobrir o inesperado e fornecer avaliações oportunas, defensáveis e compreensíveis.

Keim et al. (2008) descrevem o processo de Análise Visual, definindo que a escolha de um *design* deve, ciclicamente, permitir a análise, mostrar o importante, ampliar, filtrar e permitir analisar mais detalhes sob demanda. Para Keim e Ward (2002), a exploração visual de dados pode ser vista como um processo de geração de hipóteses e as visualizações dos dados permitem que o usuário obtenha informações sobre os dados e crie novas hipóteses. A exploração visual de dados no protótipo construído segue o processo descrito por Keim e Ward (2002) composto de três etapas: disponibilizar primeiramente uma visão geral, seguida de controles de *zoom* e filtro e, posteriormente, detalhamento sob demanda de tal forma que ofereça percepções de padrões ou grupos interessantes nos dados.

É estabelecida também, por Keim et al. (2008), a diferença entre análise visual e visualização de informações, descrevendo que a análise visual é mais do que apenas visualização, podendo ser vista como uma abordagem integral para a tomada de decisão, combinando visualização, fatores humanos e análise de dados. Aponta ainda que o desafio é identificar o melhor algoritmo automatizado para a tarefa de análise, identificar limites que não podem ser automatizados, e desenvolver uma solução que integre técnicas adequadas de visualização e interação.

2.1.1.6 Painéis de Visualização de Informação - Dashboards

O desenvolvimento de *dashboards* e a visualização de dados são tratados por Few (2006) na obra “Criação de *dashboards* informativos: A comunicação visual eficaz dos dados”. Nela foram definidos aspectos que devem ser observados para a construção de painéis de visualização e estes foram descritos a seguir:

- Informações organizadas para apoiar seu significado e uso;
- Consistência mantida para uma interpretação rápida e precisa;
- Visualização esteticamente agradável;
- *Design* para uso como plataforma de lançamento;
- Avaliação da Usabilidade nos painéis criados.

Para Few (2006), painéis de visualização devem ser projetados de acordo com aspectos importantes do design visual e de usabilidade:

Alguns aspectos importantes do *design* visual do painel ainda precisam ser considerados. Um dos mais desafiadores é a necessidade de organizar muitos itens de informação, muitas vezes relacionados apenas pela necessidade do espectador de monitorá-los todos de uma maneira que não resulte em uma bagunça desordenada. Este arranjo deve apoiar as relações intrínsecas entre os vários itens e a maneira pela qual eles devem ser navegados e utilizados para apoiar a tarefa em questão. O design de um painel deve suportar seu uso de forma otimizada e transparente. O todo também deve ser agradável de se ver, ou será ignorado (Few (2006), p. 138).

Few (2006) afirma ainda que a usabilidade é um aspecto a ser observado, e avaliações de usabilidade na área de conhecimento da Interação Humano-Computador (IHC), tem em suas origens as propostas e diretrizes de Nielsen e Molich (1990).

A publicação de Nielsen (1994) diz que heurísticas são princípios gerais para o desenvolvimento de interfaces de usuários, sendo regras gerais e não diretrizes específicas de usabilidade. Foram apresentadas dez heurísticas de usabilidade para desenvolvimento de interfaces de usuário e essas heurísticas foram atualizadas e disponibilizadas no site da *Nielsen Norman Group*¹. A seguir uma breve descrição dessas heurísticas:

1. **Visibilidade do Estado do Sistema:** O sistema deve informar aos usuários sobre o que está acontecendo, através de *feedbacks* apropriados e dentro de um tempo razoável;

¹<https://www.nngroup.com/articles/ten-usability-heuristics/>

2. **Correspondência entre o Sistema e o Mundo Real:** A linguagem utilizada no sistema deve ser familiar aos usuários. Deve refletir o mundo real em vez de termos técnicos;
3. **Controle e Liberdade do Usuário:** O sistema deve oferecer saídas de emergência aos usuários, permitindo que eles desfaçam ações ou saiam de situações indesejadas;
4. **Consistência e Padrões:** Os usuários não devem ter que adivinhar se palavras, situações ou ações diferentes significam a mesma coisa;
5. **Prevenção de Erros:** Criar um *design* cuidadoso a fim de evitar que um problema ocorra. Se um erro ocorrer, mensagens claras e construtivas devem ser fornecidas para ajudar os usuários na compreensão e auxiliar na correção do problema;
6. **Reconhecimento em Vez de Lembrança:** Objetos, ações e opções devem estar visíveis aos usuários ou facilmente acessíveis, eliminando a necessidade de lembrança;
7. **Flexibilidade e Eficiência de Uso:** Deve permitir que usuários experientes usem atalhos. O sistema deve ser eficiente para novos usuários e também para usuários experientes;
8. **Design Estético e Minimalista:** Diálogos não devem conter informações irrelevantes ou raramente necessárias;
9. **Ajuda aos Usuários a reconhecerem, diagnosticarem e recuperarem-se de Erros:** Mensagens de erro devem ser expressas em linguagem clara. Devem indicar precisamente o problema, e sugerir uma solução construtiva.
10. **Ajuda e Documentação:** O conteúdo da ajuda e da documentação deve ser fácil de pesquisar e focado na tarefa do usuário.

As heurísticas de Nielsen são amplamente consideradas como diretrizes valiosas para avaliação de usabilidade, não apenas para interfaces gerais, mas também para *dashboards*. No entanto, Few (2006) desenvolveu um conjunto de heurísticas específicas para a avaliação de painéis de visualização de informações, chamadas de “Heurísticas para *Design de Dashboards*”. Estas heurísticas são orientadas para a apresentação de informações em painéis de visualização de forma clara, eficiente e compreensível. Essas heurísticas são:

1. **Foco:** O painel deve concentrar as informações mais importantes e relevantes para os usuários;
2. **Visibilidade:** As informações devem estar visíveis de maneira clara e sem ambiguidades;
3. **Compreensibilidade:** Os usuários devem compreender as informações apresentadas no *dashboard* sem esforço;

4. **Consistência:** O *dashboard* deve manter uma consistência visual e semântica em sua totalidade;
5. **Eficiência de Uso do Espaço:** O espaço disponível deve ser usado de maneira eficiente, evitando excesso de informações desnecessárias;
6. **Avaliação Comparativa:** A comparação entre diferentes conjuntos de dados ou elementos no *dashboard* deve ser facilitada;
7. **Integração de Elementos Gráficos e Tabelas:** Gráficos e tabelas devem ser integrados de maneira eficaz para uma compreensão aprimorada;
8. **Concisão:** As informações devem ser apresentadas de forma concisa, evitando excesso de detalhes ou elementos;
9. **Contexto Adequado:** As informações devem ser apresentadas no contexto apropriado, a fim de facilitar a interpretação;
10. **Design Estético:** O *dashboard* deve ser focado na clareza e na facilidade de leitura.

Para o desenvolvimento do produto computacional proposto por essa pesquisa, buscou-se adequar estes aspectos de usabilidade definidos por Few (2006).

2.1.2 Enquadramento Tecnológico

Nesta seção foram trazidos os conceitos de enquadramento tecnológico desta pesquisa.

2.1.2.1 Mecanismos de Pesquisa: Elasticsearch e Apache Lucene

Gormley e Tong (2015) definem o Elasticsearch como um mecanismo de pesquisa de código aberto construído sobre Apache Lucene. Este último, por sua vez, é uma biblioteca de mecanismo de pesquisa avançada de texto e de alto desempenho.

Sendo o Lucene apenas uma biblioteca, é preciso integrar sua aplicação diretamente a esta biblioteca. Um fator a ser considerado é que Lucene é muito complexo. Já o Elasticsearch usa o Lucene internamente para indexação e pesquisa, ocultando as complexidades do uso da biblioteca Lucene.

Gormley e Tong (2015) caracterizam ainda o Elasticsearch como:

- Um armazenamento distribuído de documentos em tempo real onde cada campo é indexado e pesquisado;
- Um mecanismo de pesquisa distribuído com análise em tempo real;
- Capaz de escalar para centenas de servidores e petabytes de dados estruturados e não estruturados.

O Elasticsearch pode ser usado e modificado gratuitamente, pois é disponibilizado sob a licença Apache 2 e passou a ser distribuído sob a *Elastic License* em sua segunda versão², uma das licenças de código aberto disponíveis mais flexíveis, pois permite o acesso gratuito e aberto para usar, modificar, redistribuir e colaborar no código-fonte do Elasticsearch e do Kibana. A Elastic License é uma *Server Side Public License* (SSPL) que foi introduzida pela MongoDB Inc no ano de 2018,

O Elasticsearch utiliza *JavaScript Object Notation* (JSON) como formato de serialização para documentos e Gormley e Tong (2015) apresentam características e definem os seguintes termos:

- **Nó** é uma instância em execução do Elasticsearch. Enquanto um **cluster** consiste em um ou mais nós com o mesmo nome, trabalhando em conjunto e compartilhando dados e balanceando carga de trabalho;
- Um nó em um *cluster* é eleito para ser o **nó mestre**, e este passa a ser responsável por gerenciar mudanças em todo o *cluster*, como criar ou excluir um índice ou adicionar ou remover um nó do *cluster*;
- Para adicionar dados ao Elasticsearch, faz-se necessário um **índice**, que é um espaço lógico apontando para um ou mais fragmentos físicos. Um **fragmento** é uma unidade de trabalho de baixo nível que contém apenas uma fatia de todos os dados no índice;
- Um **shard** é uma única instância Lucene e é um mecanismo de pesquisa completo por si só. Os documentos do Elasticsearch são armazenados e indexados em fragmentos, mas as aplicações não se comunicam diretamente com eles. Ao invés disso, elas comunicam-se com um índice;
- Um fragmento pode ser do tipo primário ou uma réplica. Um **fragmento de réplica** é apenas uma cópia de um **fragmento primário**. As réplicas são usadas para fornecer cópias redundantes de seus dados para proteção contra falhas de *hardware* e para atender às pesquisas e recuperação de documentos;
- Um **documento** é um objeto JSON serializado com chaves e valores. Uma chave é o nome de um atributo, e um valor pode ser de tipos variados, como uma *string*, um número, um booleano, outro objeto, uma matriz de valores ou algum outro tipo especializado, como uma geolocalização;
- No Elasticsearch, todos os valores em todos os atributos são indexados por padrão. Ou seja, cada atributo possui um índice invertido dedicado para recuperação rápida de documentos correspondentes a uma pesquisa. Um **índice invertido** - (*Lucene Inverted Index*) consiste em uma lista de todas as palavras únicas que aparecem em qualquer documento e, para cada palavra, uma lista dos documentos nos quais ela aparece;

²<https://www.elastic.co/pt/pricing/faq/licensing>

Gormley e Tong (2015) afirmam que desnormalizar os dados antes de indexá-los, é a maneira de obter o melhor desempenho de pesquisa do Elasticsearch. Ter dados redundantes em cada documento elimina a necessidade de junções.

De acordo com Kimball e Ross (2011), um esquema de numerosas tabelas e junções, geralmente, apresenta desempenho de consulta mais lento em relação a uma abordagem de dados desnormalizados (*denormalized data*).

De acordo com Srivastava (2019), é possível, através do Kibana, aplicar aprendizado de máquina em seus dados para obter anomalias de dados ou tendências futuras analisando o conjunto de dados atual.

2.1.2.2 Análise Visual Exploratória no Kibana

De acordo com a Elastic (2022), a *Elastic Stack* é oriunda da *ELK Stack*, e “ELK” é o acrônimo para três projetos *open source*: *Elasticsearch*, *Logstash* e *Kibana*. O *Elasticsearch* é um mecanismo de busca e análise. O *Logstash* é um pipeline de processamento de dados do lado do servidor que faz a ingestão de dados a partir de inúmeras fontes simultaneamente, transforma-os e envia-os para o *Elasticsearch*.

Por sua vez, o *Kibana* permite que os usuários visualizem dados com diagramas e gráficos sobre a estrutura do *Elasticsearch*. Com a evolução da pilha ELK, novos componentes foram criados, como o *Beats*, agentes de coleta de dados, e o acrônimo já não acompanhava a escalabilidade da pilha, resultando então na Pilha Elastic - *Elastic Stack*, que contém os mesmos produtos *open source* com maior integração e possibilidades.

Elastic (2023) descreve que o Kibana é o portal para a Pilha *Elastic*, permitindo que seja dado forma aos seus dados e navegue no *Elastic Stack*, onde é possível pesquisar, observar e proteger os dados. No Kibana é possível realizar a Análise de Dados, e a busca por *insights* em gráficos, medidores, mapas e gráficos combinando painéis em *dashboards*.

Elastic (2023) informa que os seguintes papéis podem se integrar com o Kibana:

- Administrador, cuja função é gerenciar a Pilha do Elastic, desde a criação de sua implantação até a obtenção de dados do Elasticsearch no Kibana e, em seguida, gerenciar os dados;
- Analista de Dados, que busca visualizar os dados em painéis, encontrar *insights* e compartilhar suas descobertas;
- Usuário de Negócio (*Product Owner*), que pretende visualizar os painéis existentes, se aprofundar nos detalhes dos dados e tomar decisões apoiadas pelo conhecimento oferecido pela análise do dados.

Processos de ETL com um certo grau de complexidade precisam ser realizados em mais de uma fase e, por vezes, de uma forma não linear. Para tanto, cargas e transformações intermediárias podem utilizar mecanismos de armazenamento auxiliares, como bancos de dados. Esta pesquisa trabalha com dados semi-estruturados e um banco de dados não relacional apresenta características mais adequadas para trata-las e armazená-las. A subseção seguinte traz alguns conceitos sobre o tema.

2.1.2.3 Bancos de Dados Não-Relacionais e o MongoDB

De acordo com Sadalage e Fowler (2013), os bancos de dados relacionais oferecem muitas vantagens, mas apresentam o problema de incompatibilidade de impedância, que é a diferença entre o modelo relacional e as estruturas de dados na memória. O modelo de dados relacional não pode conter nenhuma estrutura como um registro aninhado ou uma lista. Ele organiza os dados em uma estrutura de tabelas e linhas, ou mais propriamente, relações e tuplas.

Para Chodorow (2013), o MongoDB é um banco de dados orientado a documentos, flexível e escalável. Combina a capacidade de expansão com recursos como índices secundários, consultas de intervalo, classificação, agregações e índices geoespaciais.

Alguns dos conceitos básicos do MongoDB são definidos por Chodorow (2013):

- Um **documento** é a unidade básica de dados do MongoDB e equivale a uma linha em um banco de dados relacional;
- Uma **coleção**, por sua vez pode ser considerada como uma tabela com um esquema dinâmico;
- Uma **instância** do MongoDB pode hospedar vários bancos de dados independentes, cada um com suas próprias coleções;
- Uma chave especial, o **_id**, é única dentro de uma coleção e identifica o documento.

Esquemas dinâmicos permitem que os documentos dentro de uma única coleção possam ter “formas” diferentes, e a necessidade dessa flexibilidade é algo que permeia os dados que foram tratados. Após conclusão desta fundamentação teórica conceitual e tecnológica, são apresentados na próxima seção os trabalhos relacionados a esta pesquisa.

2.2 TRABALHOS RELACIONADOS

Foram reunidas algumas pesquisas correlatas ao problema de pesquisa que se pretende investigar, particularmente estudos para análises visuais de perfil econômico e pesquisas sobre grandes volumes de dados de notas fiscais eletrônicas.

Santos (2018) propõe uma plataforma distribuída de mineração de dados para *Big Data* e realiza um estudo de caso com notas fiscais de consumidores, disponibilizados pela Secretaria de Tributação do Rio Grande do Norte. Os trabalhos se aproximam em alguns aspectos como a manipulação, processamento e extração de conhecimento a partir de grande volume de dados de notas fiscais. No entanto, diferem em outros, como o tipo de nota fiscal eletrônica usada como base, pois em vez de nota de consumidor, esta pesquisa utilizou notas fiscais eletrônicas de mercadorias entre pessoas jurídicas. Diferem também quanto ao objetivo, pois enquanto o trabalho relacionado busca desenvolver uma plataforma de extração de conhecimento através da classificação de dados por algoritmos de aprendizado de máquina, esta pesquisa tem como objetivo desenvolver uma solução de exploração visual interativa para análise de perfil econômico.

Nocko et al. (2017), tendo como base os registros das notas fiscais eletrônicas, mapeiam a interdependência econômica do Distrito Federal (DF) e apresentam uma análise descritiva do perfil comercial, observando quais são os principais parceiros do Distrito Federal, a origem das suas compras, o destino das suas vendas e os principais produtos transacionados. Este trabalho relacionado é importante para esta pesquisa, pois toma como base o mesmo tipo de nota fiscal eletrônica e tem como objetivo apresentar resultados de análises visuais sobre o mapeamento do perfil econômico do DF. Tais visualizações balizarão a criação de painéis na solução proposta por esse trabalho. A pesquisa relacionada não se preocupa em apresentar as técnicas e tecnologias utilizadas para a construção, pois foca no estudo aplicado e na discussão sobre as explorações realizadas, enquanto esta produção preocupava-se em construir uma solução interativa para a exploração visual, apresentando o produto computacional detalhadamente.

Neto (2021), em sua dissertação, desenvolve um sistema para auxiliar os auditores fiscais tributários durante o processo de análise e identificação de possíveis fraudes nas notas fiscais de serviços. A solução apresentada consiste em extrair informações das notas fiscais e, aplicando algoritmos de classificação, identificar possíveis fraudes. Apesar do trabalho correlato diferir no objetivo, que foca na identificação de fraudes, e no tipo de nota fiscal eletrônica, diverso ao que essa pesquisa se propõe, eles se assemelham no tocante à construção de solução para análise e exploração visual e na busca de conhecimento a partir de dados de notas fiscais eletrônicas.

Mendonça (2022), em sua pesquisa, investiga o desempenho do processamento de notas fiscais eletrônicas em Sistema de Gerenciamento de Banco de Dados (SGBD) Não-Relacionais (NoSQL) para controle fiscal de compras governamentais. O estudo relacionado propõe uma arquitetura baseada em *Application Programming Interface - Representational State Transfer* (API-REST) e assemelha-se a esta pesquisa quando considera a variabilidade e o volume dos dados de notas fiscais como fator relevante para optar pelo uso de bancos de dados não relacionais. Diferem, no entanto, em relação ao objetivo, pois foca na verificação do desempenho das buscas, enquanto este trabalho

propõe viabilizar o acesso ao conhecimento em uma plataforma de pesquisa textual sobre grande volume de dados.

Após revisão bibliográfica dos trabalhos relacionados, não foram encontrados estudos que apresentassem soluções de análise visual exploratória que realizassem busca textual sobre grandes volumes de dados, e que, de forma ampla, genérica e flexível, disponibilizassem compreensão do perfil econômico de municípios sobre dados de notas fiscais eletrônicas.

Neste capítulo é apresentado o produto computacional, o objetivo, as funcionalidades, as etapas de construção e as dificuldades encontradas durante o processo de desenvolvimento da solução. .

PRODUTO COMPUTACIONAL - VISUALDATA-NFE

O VisualData-NFe é uma solução de *Visual Analytics* e *Big Data* com busca textual sobre dados de notas fiscais eletrônicas de compra e venda de mercadorias entre pessoas jurídicas. O objetivo principal da solução é oferecer mecanismos de análise e exploração visual, para compreensão do perfil produtivo e da dinâmica econômica de municípios.

Esta solução é composta por um núcleo de extração, transformação, carga e indexação de dados fiscais, geográficos, estatísticos e econômicos em uma plataforma de *Big Data*. Ela é composta também por um conjunto de visualizações interativas que permitem a exploração.

3.1 MÉTODO

O processo de desenvolvimento foi conduzido por um percurso baseado em métodos ágeis e as etapas foram realizadas de forma incremental e iterativa, descritas nas seções a seguir:

1. Mapeamento das necessidades através de histórias de usuários;
2. Definição da arquitetura da solução e da visão geral da proposta;
3. Desenvolvimento do VisualData-NFe.

No final deste capítulo são apresentados os desafios encontrados durante a construção do produto computacional.

3.2 MAPEAMENTO DAS NECESSIDADES ATRAVÉS DE HISTÓRIAS DE USUÁRIOS

De acordo com Cohn (2004), uma *story user* descreve funcionalidades que serão valiosas para os usuários e são compostas por três aspectos:

1. Uma descrição escrita da história usada para planejamento;
2. Conversas sobre a história que servem para detalhá-la;
3. Definição de critérios de aceitação que podem ser usados para determinar se a implementação de uma história foi completa e testá-la.

Cohn (2004) também descreve as seguintes técnicas para criação de um conjunto de histórias:

- Entrevistas com usuários;
- Questionários;
- Observação;
- Oficinas de redação de histórias.

Este trabalho utilizou-se de entrevistas com usuários reais para a construção do protótipo funcional. Utilizou-se também da observação da interação dos usuários com o software seguida da coleta de *feedbacks*.

A seguir, a Tabela 3.1 traz as necessidades de usuários que foram mapeadas através de *story users*.

Tabela 3.1 Histórias de Usuário

História:	1 - Mercadorias identificadas por NCM
Descrição:	Como usuário gostaria que as mercadorias identificadas pelo NCM tenham a descrição completa de acordo com o Sistema de Comércio Exterior - SISCOMEX
Feedback:	Notas Fiscais com NCMs não informados ou não identificados devem ter a descrição “00000000- NCM NÃO IDENTIFICADO”
História:	2 - Estados e Municípios de origem e destino das mercadorias
Descrição:	Como usuário desejo saber quais os estados e municípios de origem das mercadorias, quais os destinos, quais são as mercadorias e qual o volume. O objetivo é realizar buscas até no nível dos municípios.
Feedback:	Municípios com identificador 9999999 referem-se às exportações, e devem constar com a descrição “EXTERIOR”. Essa identificação é importante para sabermos o quanto cada município está contribuindo nas exportações.
História:	3 - Operações Fiscais
Descrição:	Toda Nota Fiscal Eletrônica (NF-e) traz o código da operação fiscal (CFOP). Como usuário quero saber quais as operações fiscais envolvidas nas entradas e saídas de mercadorias.
Feedback:	Gostaríamos que os CFOPs estivessem categorizados, pois assim facilitaria pesquisas por categorias pré-definidas.
História:	4 - Atividades Econômicas
Descrição:	Como usuário desejo saber quais as atividades econômicas (CNAEs) envolvidas.
Feedback:	De acordo com o esperado.
História:	5 - Filtros de Pesquisa
Descrição:	Como usuário gostaria que fossem disponibilizados filtros sobre as informações fornecidas nas notas fiscais.
Feedback:	É importante ter uma forma de desprezar valores muito pequenos e valores muito grandes.

Fonte: Autor, 2023.

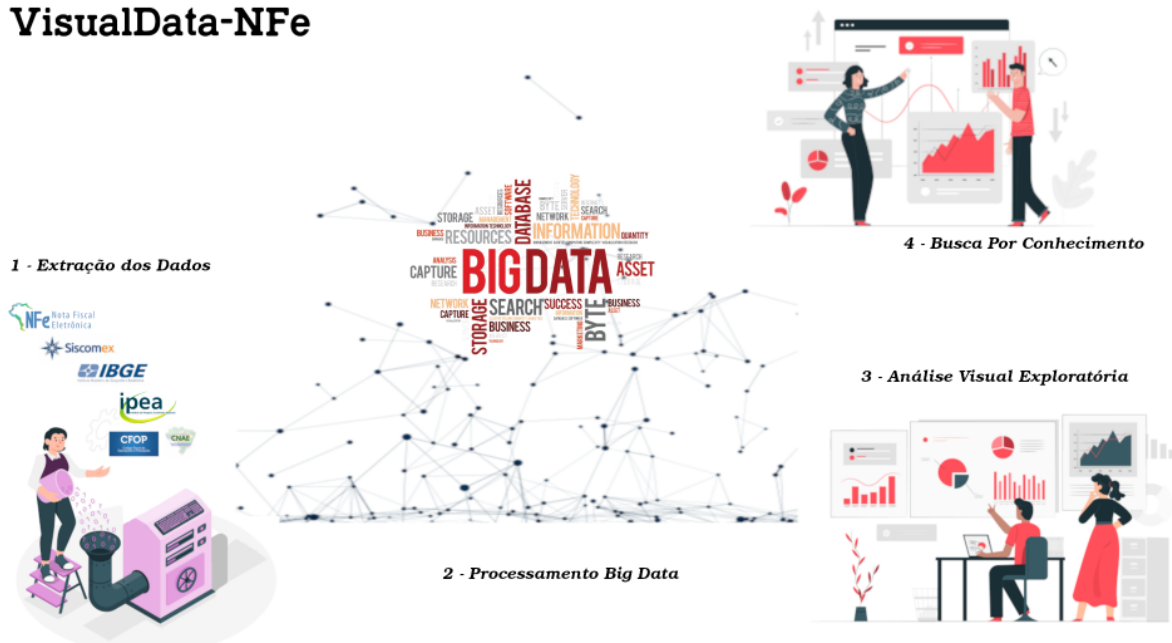
3.3 DEFINIÇÃO DA ARQUITETURA DA SOLUÇÃO E DA VISÃO GERAL DA PROPOSTA

A Figura 3.1 apresenta as etapas do processo de busca por conhecimento através da análise visual exploratória e a visão geral da solução. Este processo é compreendido pelas seguintes etapas:

1. Extração dos Dados: Nesta fase, as variáveis que se pretende analisar são extraídas de diversas fontes. Elas devem estar relacionadas com os dados constantes nas notas fiscais eletrônicas, e devem ser disponibilizadas para a aplicação;
2. Processamento *Big Data*: Nesta etapa ocorre a transformação dos dados heterogêneos e a indexação na plataforma de *Big Data* visando a pesquisa em alto desempenho;
3. Análise Visual Exploratória: Nesta etapa os analistas, pesquisadores e especialistas em geral têm acesso ao *dashboard* da aplicação: um painel contendo diversas visualizações. Estas visualizações são adaptáveis, flexíveis e consistentes com a pesquisa e filtragem dos dados que se pretende analisar;
4. Busca por Conhecimento: Nesta fase concretiza-se a principal função da solução que é tornar-se instrumento de apoio na compreensão do perfil e da dinâmica econômica de municípios.

Figura 3.1 Visão Geral

VisualData-NFe



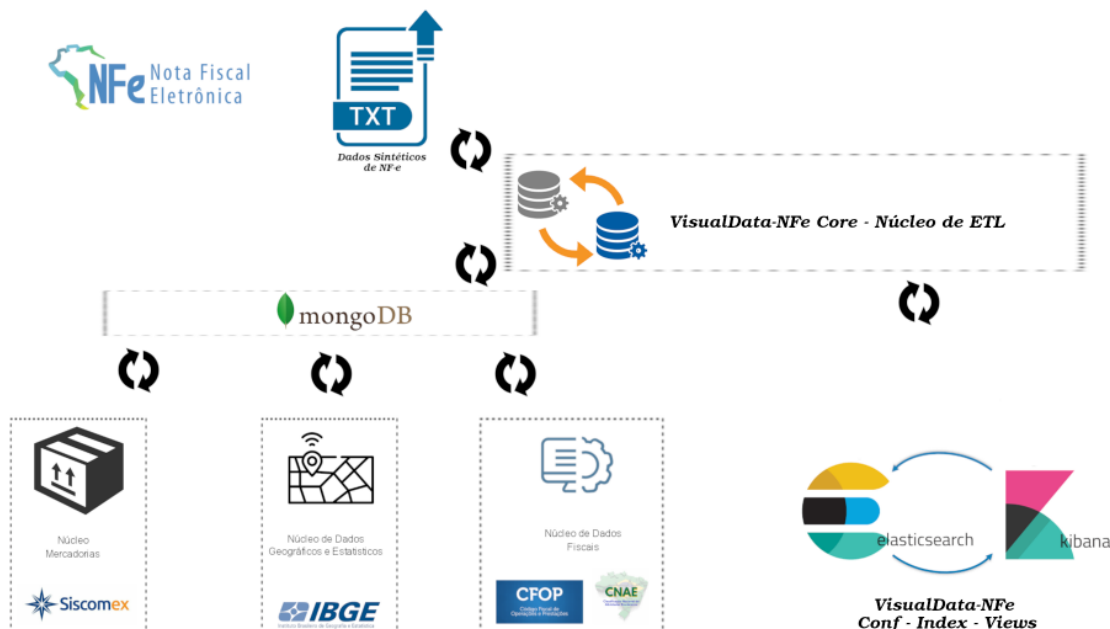
Fonte: Autor, 2023.

Uma solução que envolve *Big Data*, *Visual Analytics* e um *dashboard*, pode ser enquadrada em uma arquitetura cliente-servidor, com elementos de arquitetura orientada a serviços. Os seguintes elementos foram envolvidos nesta proposta:

- Processamento e armazenamento *Big Data* no Elasticsearch;
- Processamento de *Visual Analytics*, visualizações e *dashboards* no Kibana - VisualData-NFe Dashboard;
- API REST utilizada para ingestão, processamento e enriquecimento de dados - VisualData-NFe Core;
- Banco de Dados de apoio (NoSQL) no MongoDB.

A Figura 3.3 traz o desenho da solução e os componentes envolvidos no produto computacional. A solução é composta pela obtenção dos dados sobre regiões e municípios (Ex.: IDHM, PIB, Área, Geolocalização), informações sobre atividades econômicas (CNAE), operações fiscais (CFOP) e mercadorias (NCM). Esses dados são transformados e carregados no MongoDB através da API do VisualData-NFe - Core. A partir do momento que as coleções estão prontas no MongoDB, os dados de notas fiscais podem ser extraídos, tratados e relacionados para posterior carregamento na plataforma de *big data* (*Elasticsearch*). Após os dados de notas fiscais estarem indexados e disponíveis na plataforma, diversas visualizações interativas foram organizadas em um *dashboard*.

Figura 3.2 Desenho da Solução



Fonte: Autor, 2023.

3.4 DESENVOLVIMENTO DO VISUALDATA-NFE

O processo de desenvolvimento do VisualData-NFe compreende duas grandes etapas: a primeira está relacionada aos processos de Extração-Transformação-Carga (ETL) e a segunda diz respeito as visualizações construídas. As subseções a seguir descrevem essas etapas.

3.4.1 Extração, Transformação, Carga e Indexação dos Dados

O processo de ETL envolveu a extração, transformação e carga de dados de notas fiscais, NCM, CFOP, Dados econômicos, geográficos e estatísticos. Este processo é detalhado a seguir.

3.4.1.1 Nomenclatura Comum do Mercosul (NCM) A Receita Federal do Brasil (RFB) disponibiliza os dados de NCM, separados ano a ano, através de uma *Application Programming Interface - Representational State Transfer* (API-REST). Para acessar essas informações, fez-se necessário a obtenção de um *Token*¹ para autorização de acesso. Após obtido o *token*, este foi enviado no cabeçalho da requisição através do atributo *X-CSRF-Token*, para o *end-point*² disponibilizado para consulta. Os anos foram informados como parâmetro na URL através do atributo de mesmo nome, como por exemplo: “ano=2021”.

Feito isso, obteve-se a informação pesquisada no formato de uma lista JSON com atributos como o código, o nome por extenso, o início e o fim da vigência de cada código pertencente àquele ano. Para o presente estudo, foram obtidos os dados do período compreendido entre os anos de 1997 (ano de implantação) e 2023 (ano de realização deste estudo).

Para que se conseguisse um detalhamento da mercadoria foi necessário desenvolver uma solução que fizesse a busca das descrições de forma recursiva em toda a cadeia de classificação a que a mercadoria pertence. Por exemplo, ao buscar a descrição do NCM 0101.29.00, obteve-se como resposta: “– Outros”, que por si só é insuficiente para descrever a mercadoria com precisão.

A Figura 3.3 traz a forma como a API disponibiliza a descrição do NCM e como é a informação é disponibilizada de forma completa no Sistema Integrado de Comércio Exterior (SISCOMEX).

Para contornar tal situação, fez-se necessário tratar os dados de forma recursiva e hierárquica, tornando possível compreender que o código 0101.29.00 refere-se à descrição completa: “01 - Animais vivos; 0101 - Cavalos, asininos e muares, vivos; 0101.2 - Cavalos; 0101.29.00 - Outros”.

¹Token obtido em RFB (2022a)

²Dados extraídos de RFB (2022b)

Figura 3.3 NCM API-REST x SISCOMEX

The image displays a REST client interface for an API call to the Siscomex Nomenclatura API. The URL is `https://portalunico.siscomex.gov.br/classif/api/nomenclatura/consulta?codigo=01012900&nomeExtenso=`. The Headers tab is selected, showing 25 headers. The Body tab shows a JSON response:

```

1 [
2   {
3     "id": 32,
4     "codigo": "01012900",
5     "nomeExtenso": "-- Outros",
6     "notas": []
  }
]

```

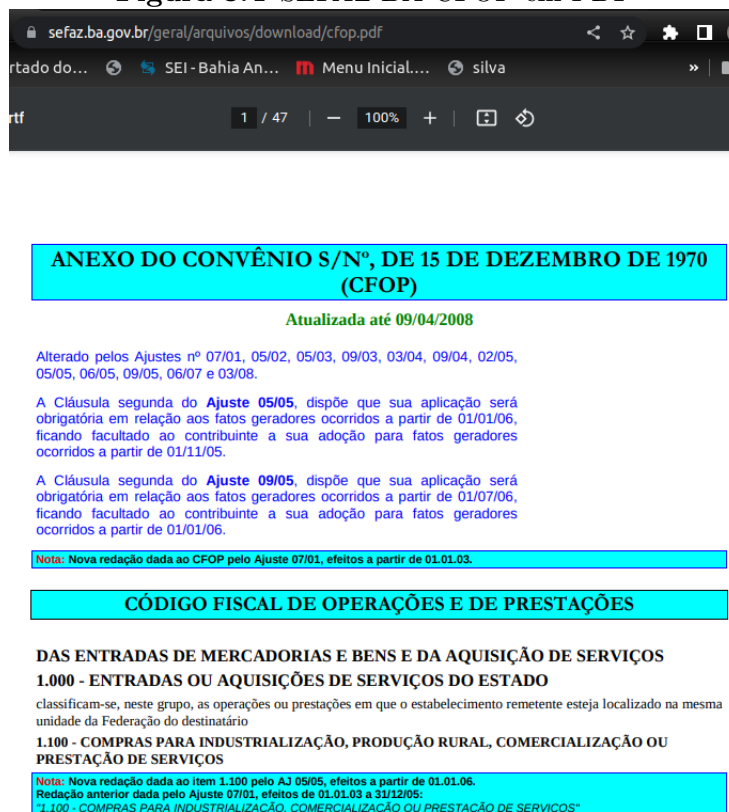
Below the API view, the Siscomex website interface is visible. It includes a warning: "ATENÇÃO! SUA CONSULTA ESTÁ COM DATA RETROATIVA." and a search result for "0101.29.00 - Outros" under the section "SEÇÃO I - ANIMAIS VIVOS E PRODUTOS DO REINO ANIMAL".

Fonte: SISCOMEX, 2023.

3.4.1.2 Código Fiscal de Operações e Prestações (CFOP) Os órgãos e secretarias fazendárias estaduais disponibilizam as informações sobre os CFOPs em diversos formatos. Para este trabalho, a tabela³ utilizada foi a disponibilizada pela SEFAZ-BA. Esses dados foram disponibilizados no formato PDF e seu conteúdo foi extraído e transformado em um arquivo CSV separado por “;”.

A seguir, a Figura 3.4 apresenta o documento que contém os CFOPs disponibilizados pela SEFAZ-BA.

Figura 3.4 SEFAZ-BA CFOP em PDF



Fonte: SEFAZ, 2008.

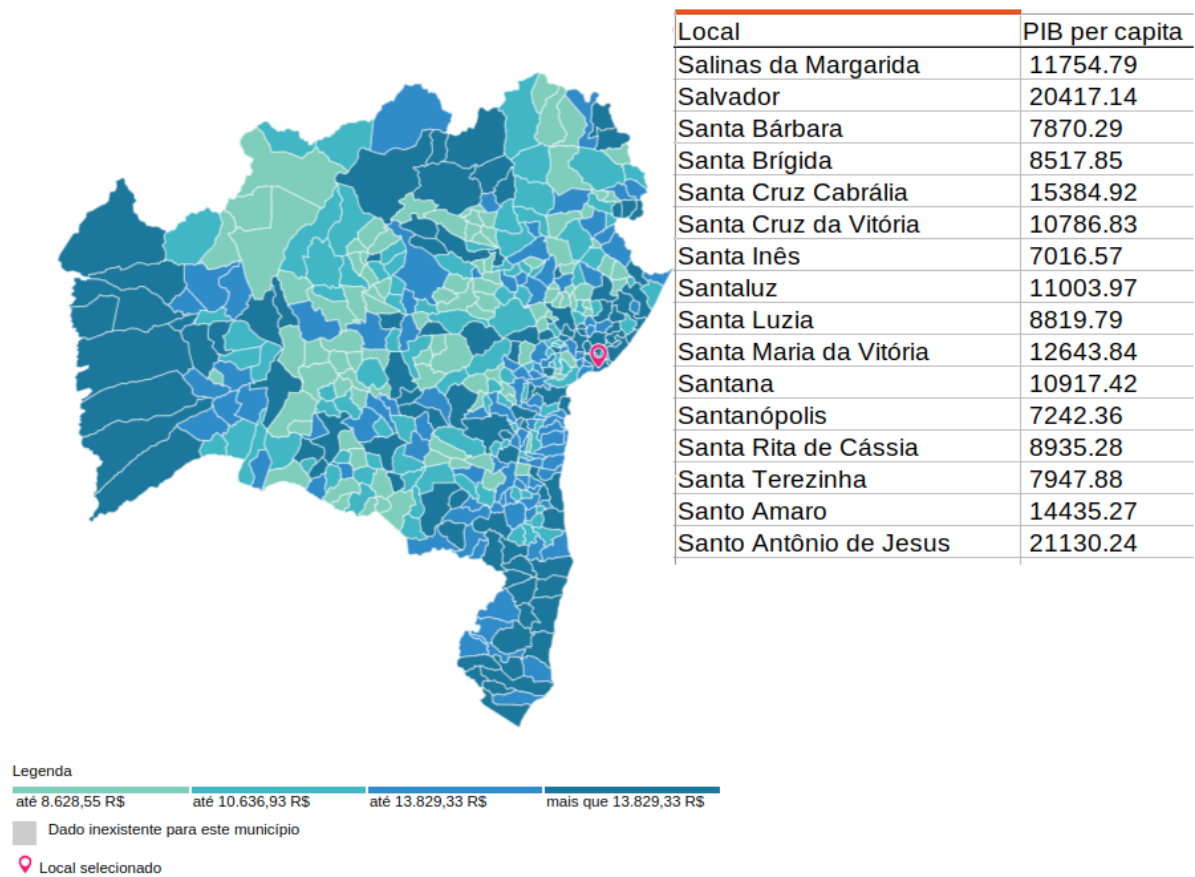
3.4.1.3 Dados econômicos, geográficos e estatísticos O PIB per Capita é a razão entre o PIB e a população. Os dados foram obtidos do IBGE e são estruturados conforme Figura 3.5 a seguir.

De maneira similar, foram obtidas do IBGE as informações territoriais de área dos municípios em km² e da população no último censo divulgado até essa pesquisa (2010). A Figura 3.6 mostra a estrutura dos dados obtidos.

³Dados extraídos de SEFAZ-BA (2008)

Figura 3.5 PIB per capita - Bahia

PIB per capita



Fonte: IBGE, 2020.

Figura 3.6 Área e População - Bahia

1	Local	Área da unidade territorial	1	Local	População no último censo
337	Salvador	693442	337	Salvador	2675656
338	Santa Bárbara	347021	338	Santa Bárbara	19064
339	Santa Brígida	934461	339	Santa Brígida	15060
340	Santa Cruz Cabralia	1462942	340	Santa Cruz Cabralia	26264
341	Santa Cruz da Vitória	284083	341	Santa Cruz da Vitória	6673
342	Santa Inês	379270	342	Santa Inês	10363
343	Santaluz	1623445	343	Santaluz	33838
344	Santa Luzia	824473	344	Santa Luzia	13344
345	Santa Maria da Vitória	1984910	345	Santa Maria da Vitória	40309
346	Santana	1909353	346	Santana	24750
347	Santanópolis	222686	347	Santanópolis	8776
348	Santa Rita de Cássia	6030491	348	Santa Rita de Cássia	26250
349	Santa Terezinha	719257	349	Santa Terezinha	9648
350	Santo Amaro	494502	350	Santo Amaro	57800
351	Santo Antônio de Jesus	261740	351	Santo Antônio de Jesus	90985
352	Santo Estêvão	360334	352	Santo Estêvão	47880
353	São Desidério	15156712	353	São Desidério	27659
354	São Domingos	289963	354	São Domingos	9226
355	São Félix	103226	355	São Félix	14098
356	São Félix do Coribe	1751671	356	São Félix do Coribe	13048
357	São Felipe	222407	357	São Felipe	20305
358	São Francisco do Conde	269715	358	São Francisco do Conde	33183
359	São Gabriel	1146054	359	São Gabriel	18427
360	São Gonçalo dos Campos	294768	360	São Gonçalo dos Campos	33283

Fonte: IBGE, 2010.

Os dados do IBGE sobre localização de municípios trazem informações sobre as regiões em que os municípios estão inseridos (microrregião, mesorregião, região, região-imediata e região-intermediária). Trazem também informações sobre a latitude e a longitude do ponto central do município (centroide), além de informações sobre área em quilômetros quadrados. A Figura 3.7 apresenta os dados semi-estruturados em JSON.

Figura 3.7 Dados de Localização - Salvador - BA

```

{
  "id":2927488,
  "nome":"Salvador",
  "microrregiao":{
    "id":29821,
    "nome":"Salvador",
    "mesorregiao":{
      "id":2985,
      "nome":"Metropolitana de Salvador",
      "UF":{
        "id":29,
        "sigla":"BA",
        "nome":"Bahia",
        "regiao":{
          "id":2,
          "sigla":"NE",
          "nome":"Nordeste"
        }
      }
    }
  },
  "regiao-imediata":{
    "id":298881,
    "nome":"Salvador",
    "regiao-intermediaria":{
      "id":2981,
      "nome":"Salvador",
      "UF":{
        "id":29,
        "sigla":"BA",
        "nome":"Bahia",
        "regiao":{
          "id":2,
          "sigla":"NE",
          "nome":"Nordeste"
        }
      }
    }
  }
},
{
  "id":"2927488",
  "nivel-geografico":"munic\u00EDpio",
  "centroide":{
    "longitude":-38.5147,
    "latitude":-12.8735
  },
  "regiao-limitrofe":[
    {
      "longitude":-38.6993,
      "latitude":-12.7336
    },
    {
      "longitude":-38.3843,
      "latitude":-13.8173
    }
  ],
  "area":{
    "dimensao":"693.442",
    "unidade":{
      "id":"km2",
      "nome":"Quil\u00F4metros quadrados"
    }
  }
}
]

```

Fonte: IBGE, 2022.

3.4.1.4 Carga e Indexação dos Dados de Notas Fiscais Eletrônicas O núcleo de extração, transformação e carga do VisualData-NFe é responsável por agregar os dados de notas fiscais eletrônicas com as variáveis e índices para análise exploratória. Uma vez agrupados, esses dados são carregados na plataforma de *Big Data* de forma a responderem às pesquisas, com reflexo em todos os painéis de visualizações do *dashboard*.

Por se tratar de grande volume de dados, a forma como os estes são armazenados e recuperados formam uma tarefa de alto grau de complexidade computacional e as estratégias de indexação viabilizam respostas rápidas aos filtros, que ao serem aplicados, refletem e atualizam todas as visualizações do *dashboard* em poucos segundos.

3.4.2 Visualizações

Ao acessar a aplicação, é exibido um conjunto de visualizações que compõem o *Dashboard* do VisualData-NFe. Ele fornece *insights* sobre os dados de várias perspectivas e permite que os usuários se aprofundem nos detalhes. Esse *dashboard* é apresentado na Figura 3.8 e é dividido em quatro áreas, a saber:

1. Área de Pesquisa e Filtragem;
2. Área de Notas Fiscais;
3. Área de Mercadorias;
4. Área de Atividades Econômicas e Operações Fiscais.

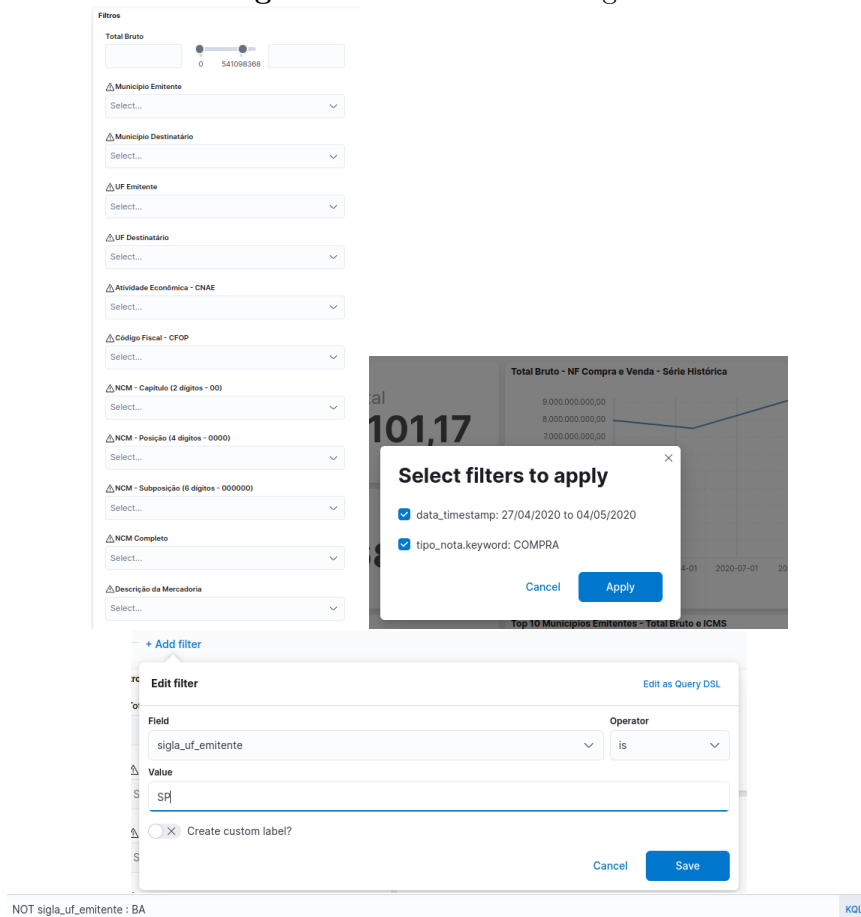
Figura 3.8 Dashboard - VisualData-NFe



Fonte: Autor, 2023.

3.4.2.1 Pesquisa e Filtragem Foram disponibilizados controles de filtragem como listas de opções, calendário e *range slider*. A seguir a Figura 3.9, variadas formas de filtragem dos dados, através do painel de filtros, diretamente pela visualização, através de *Query Domain Specific Language* (Query-DSL) e de *Kibana Query Language* (KQL).

Figura 3.9 Formas de Filtragem

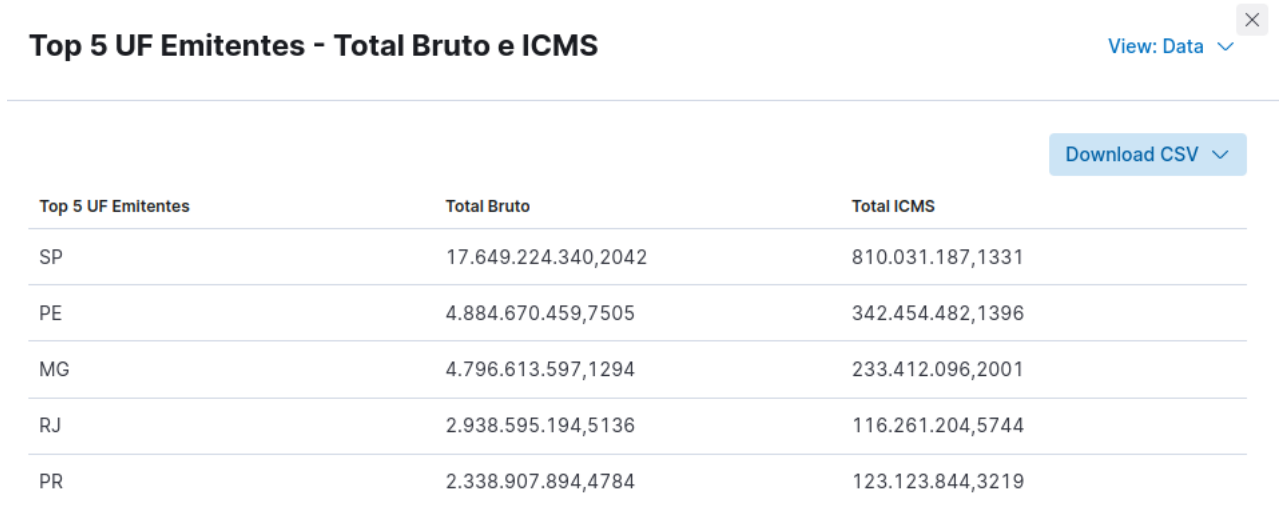


Fonte: Autor, 2023.

Para este trabalho foram adicionadas ao *dashboard* diversas visualizações de dados graficamente representados em colunas, linhas, agregações, *treemaps* e em *pizzas*. Essas visualizações são apresentadas na subseção a seguir.

3.4.2.2 Área de Notas Fiscais Foram construídas visualizações exploratórias que permitem ao usuário buscar compreensões sobre como é a relação de compra e venda de mercadorias, relacionadas ao estado da Bahia, volume de venda e aquisição, bem como as localidades envolvidas. A Figura 3.10 mostra a relação com os cinco estados que mais vendem para a Bahia.

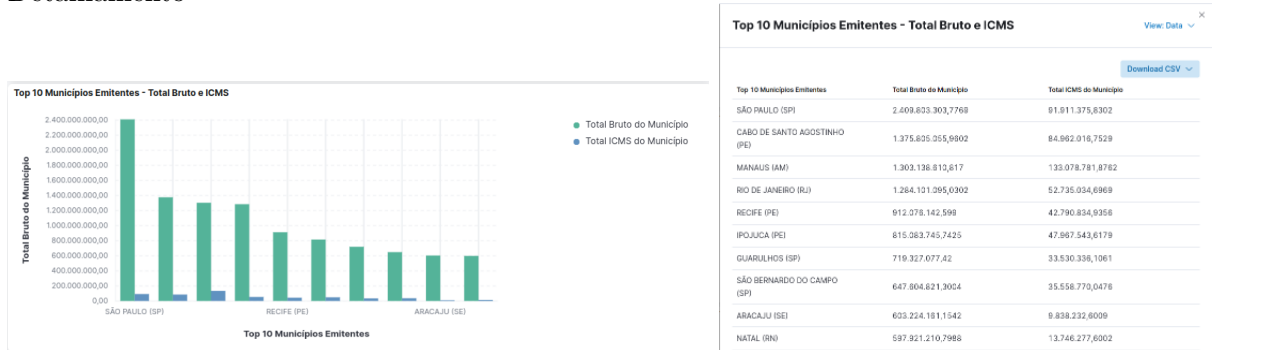
Figura 3.10 5 UFs que mais venderam para a Bahia - Total Bruto (R\$) Visualização e Detalhamento



Fonte: Autor, 2023.

A Figura 3.11 traz uma visão sobre os municípios que mais venderam para a Bahia (por total bruto).

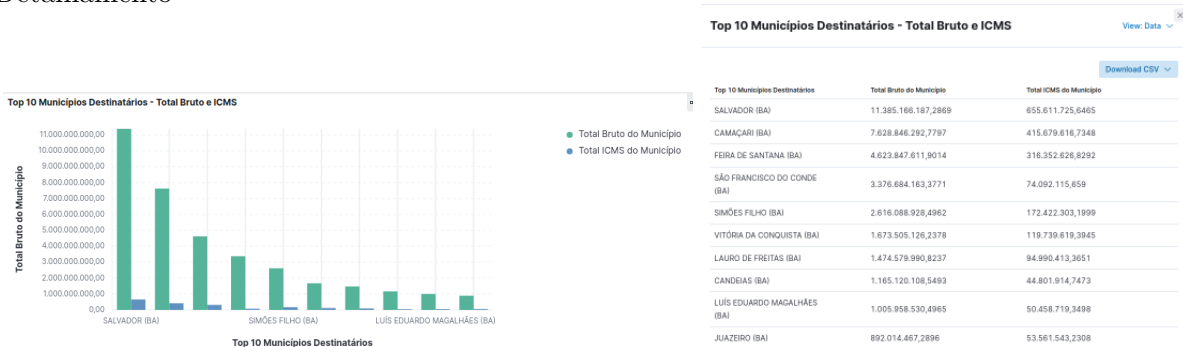
Figura 3.11 10 Municípios que mais venderam para a Bahia - Total Bruto (R\$) Visualização e Detalhamento



Fonte: Autor, 2023.

É possível buscar compreensão sobre as relações de compra sob a ótica dos municípios compradores. A Figura 3.12 traz os municípios baianos que mais compraram de outros estados (por total bruto).

Figura 3.12 10 Municípios que mais compraram na Bahia - Total Bruto (R\$) Visualização e Detalhamento



Fonte: Autor, 2023.

3.4.2.3 Área de Mercadorias As visualizações disponibilizadas nesta área possibilitam ao usuário buscar compreensões sobre as mercadorias negociadas. É possível verificar o percentual do total bruto e total de ICMS comercializado. A Figura 3.13 apresenta as mercadorias e o volume total bruto vendido por Ilhéus (BA) para outros municípios do estado da Bahia.

Figura 3.13 Mercadorias vendidas por Ilhéus (BA) - Total Bruto

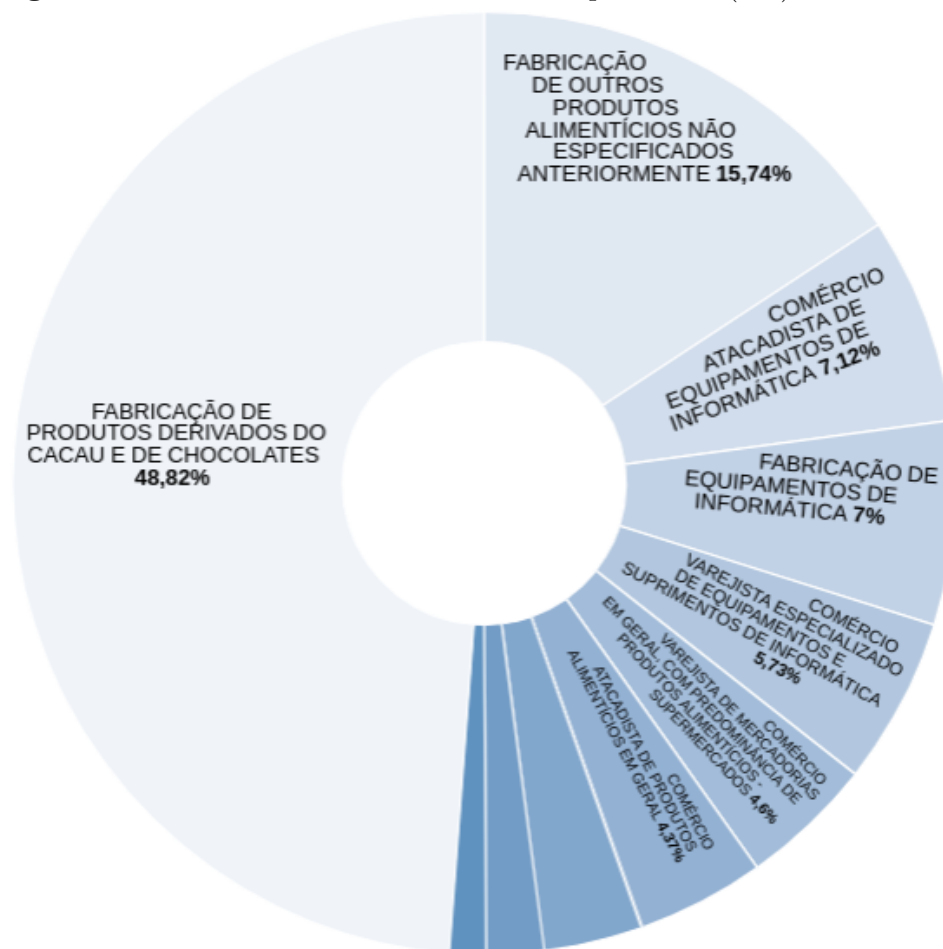


Fonte: Autor, 2023.

3.4.2.4 Área de Atividades Econômicas e Operações Fiscais Esta área traz visualizações que permitem acesso ao panorama geral de associação das operações de compra e venda com a atividade comercial desempenhada e a operação fiscal.

A Figura 3.14 traz o volume total do ICMS sob a ótica das atividades econômicas das operações de venda por Ilhéus (BA) para outros municípios do estado da Bahia.

Figura 3.14 Atividade Econômicas - Vendas por Ilhéus (BA) - Total ICMS



Fonte: Autor, 2023.

3.5 DESAFIOS ENCONTRADOS

No decorrer da pesquisa, diversos desafios foram encontrados e precisaram ser superados. Dentre eles, destaca-se a extração de dados a partir da publicação em formatos que dificultaram o processamento, como por exemplo, arquivos em .PDF como o de CFOP. Mesmo ao encontrar informações publicadas em APIs, a documentação para utilização era escassa ou inexistente. Isso aplica-se também à superficialidade ou falta de metadados que expliquem as estruturas dos dados disponibilizados.

Ao realizar pesquisas e experimentos em *Big Data*, compreende-se que alguns requisitos básicos de infraestrutura são exigidos, e este trabalho dispôs de recursos limitados de

capacidade de processamento e armazenamento. Para superar os desafios impostos pelas limitações dos recursos, precisou-se suprimir, por exemplo, o campo aglutinador do *Elasticsearch: message*, que reduz o tempo de resposta durante as pesquisas, mas que praticamente dobra o tamanho do espaço consumido em disco. Para contornar a questão de desempenho, o esquema de dados precisou ser projetado para indexar dados desnormalizados. Esta medida objetiva a compreensão de perfis produtivos de municípios tomando como base milhões de registros agrupados de notas fiscais.

Este capítulo apresenta o primeiro estudo exploratório realizado a fim de avaliar se VisualData-NFe é eficaz em cumprir seu objetivo.

ESTUDO EXPLORATÓRIO I

4.1 OBJETIVO

Em maio de 2023, seguindo o processo *Goal-Question-Metric* (GQM) de Basili e Rombach (1994), foi realizado o primeiro estudo para avaliação do VisualData-NFe. O **objetivo** foi analisar a ferramenta com a **finalidade** de verificar se a solução é **eficaz** em oferecer compreensão do perfil produtivo e da dinâmica econômica de municípios **sob a ótica** dos pesquisadores da solução, no **contexto** da análise visual exploratória de notas fiscais eletrônicas sintéticas de compra e venda do estado da Bahia.

4.2 MÉTODO

De acordo com Basili e Rombach (1994), a abordagem GQM é um mecanismo para definir e interpretar de forma mensurável, um software operacional. Consiste na especificação de um sistema de medição visando o levantamento de um conjunto particular de questões em relação a um objetivo e na identificação de métricas que proporcionem respostas às questões levantadas.

O desenvolvimento deste estudo experimental ocorreu a partir das seguintes etapas:

1. Definição do Ambiente Experimental;
2. Escolha das Métricas;
3. Elaboração das Questões e Medições.

O detalhamento das etapas está organizado e distribuído nas próximas subseções.

4.2.1 Definição do Ambiente Experimental

Como tecnologia de *Big Data* para armazenamento, indexação e pesquisa foi adotado o *Elasticsearch*. A plataforma pode ser usada e modificada gratuitamente, pois é disponibilizada sob a licença Apache 2, uma das licenças de código aberto mais flexíveis, e utiliza *JavaScript Object Notation* (JSON) como formato de serialização para os documentos indexados.

O VisualData-NFe disponibiliza recursos para exploração visual interativa dos dados acessíveis no *Elasticsearch*. Encontram-se disponíveis os painéis integrados de visualização. Isso significa que, ao adicionar ou retirar um filtro de pesquisa, todas as visualizações são afetadas simultaneamente e de forma coerente.

Essas ferramentas e tecnologias foram implantadas e configuradas em um sistema operacional Ubuntu (Linux/Debian) em um *Virtual Private Server* (VPS).

Este estudo utilizou dados sintéticos fictícios de notas fiscais eletrônicas. Foram gerados e carregados na plataforma cerca de 34 milhões de dados sintéticos de notas, distribuídos ao longo de seis meses, ocupando um espaço de armazenamento de 31,8 GB. Os dados totalizam R\$ 290,8 bi em valores brutos e R\$ 12,2 bi em valores de ICMS.

Os dados gerados foram baseados no perfil produtivo de estados e municípios obtido dos dados publicados pelo Observatório da Agropecuária Brasileira (INMET, 2022) e pela Confederação Nacional da Indústria (CNI, 2022).

4.2.2 Escolha das Métricas

A *International Organization for Standardization* (ISO) (Organização Internacional para Padronização) reúne autoridades para discutir meios de facilitar internacionalmente a coordenação e unificação de padrões industriais. A *International Electrotechnical Commission* (IEC) (Comissão Eletrotécnica Internacional) é uma organização internacional de padronização de tecnologias elétricas, eletrônicas e relacionadas. De acordo com Bevan et al. (2016), a qualidade em uso de software é definida na ISO/IEC 25010 - Sistemas e Modelos de Qualidade de Software. Esta norma classifica as métricas nos componentes eficácia, eficiência e satisfação. Para este estudo experimental, foram utilizadas as seguintes métricas:

- M1 - Objetivos alcançados - Um resultado bem-sucedido da interação com um sistema (eficácia). Obtém-se o valor SIM caso o objetivo da questão seja plenamente atingido, e NÃO em caso contrário;
- M2 - Tempo de tarefa - Tempo para conclusão completa e precisa de uma tarefa. Obtém-se o valor médio em segundos, para conclusão da tarefa com êxito;
- M3 - Independe de proficiência - Avalia se pode ser usado por pessoas que não possuem conhecimentos, habilidades ou experiências específicas, e nestes casos, obtém-se o valor SIM. Ou se deve ser potencialmente utilizável por pessoas com habilidades específicas, e nestes casos, obtém-se valor NÃO.

4.2.3 Elaboração das Questões e Medições

Para Basili e Rombach (1994), um conjunto de questões é utilizado para caracterizar a forma como a avaliação/atingimento de um objetivo específico vai ser realizada com base em algum modelo de caracterização. As perguntas tentam definir o objeto de medição (produto, processo ou recurso) em relação a um problema de qualidade selecionado e determinar sua qualidade do ponto de vista selecionado.

Para este estudo, as questões formuladas são apresentadas a seguir.

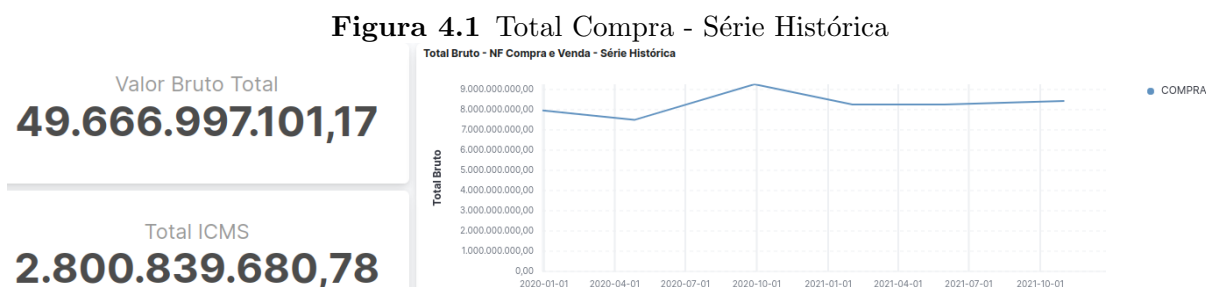
- Q1 - É possível identificar o volume de mercadorias adquiridas de outros estados?
- Q2 - É possível identificar de quais regiões as mercadorias foram adquiridas?
- Q3 - É possível identificar quais os maiores municípios fornecedores de um determinado estado?
- Q4 - É possível identificar as mercadorias negociadas entre municípios específicos?
- Q5 - É possível identificar quanto foi o total de ICMS em um determinado período sobre a comercialização de produtos específicos?
- Q6 - O sistema é capaz de auxiliar na busca de alternativas de substituição de região fornecedora de determinada mercadoria?

4.2.4 Coleta e Análise de Dados

Nesta seção, são apresentadas as questões de pesquisa, seguidas da visualização da ferramenta utilizada para atingir o objetivo da questão, além de uma análise das métricas propostas.

Q1 - É possível identificar o volume de mercadorias adquiridas de outros estados?

Nesta simulação objetiva-se identificar o valor total bruto e o valor total de ICMS em relação às mercadorias que o estado da Bahia comprou de outros estados. Busca-se os valores acumulados e também distribuídos ao longo dos meses, conforme visualizado na Figura 4.1.



Fonte: Autor, 2023.

A Tabela 4.2 traz as métricas apresentadas anteriormente, as unidades de medida e as medições obtidas na simulação.

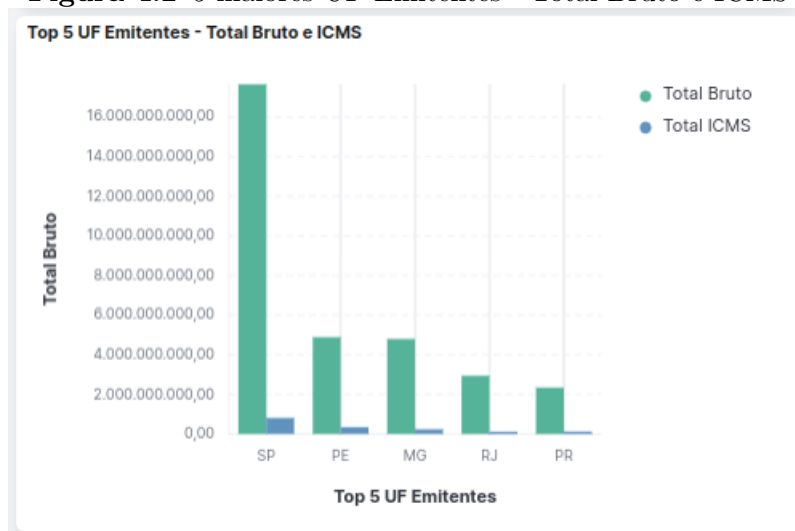
É possível verificar que:

- A Bahia adquiriu um total bruto de R\$ 49,67 bi em mercadorias vendidas por outros estados;
- O total de ICMS foi de R\$ 2,8 bi;
- A média do valor total de aquisição de mercadorias é R\$ 8,28 bi por mês.

Q2 - É possível identificar de quais regiões as mercadorias foram adquiridas?

Nesta simulação busca-se visualizar de quais regiões o estado da Bahia adquire mercadorias. A Figura 4.2, revela as cinco maiores regiões que vendem mercadorias à Bahia.

Figura 4.2 5 maiores UF Emitentes - Total Bruto e ICMS

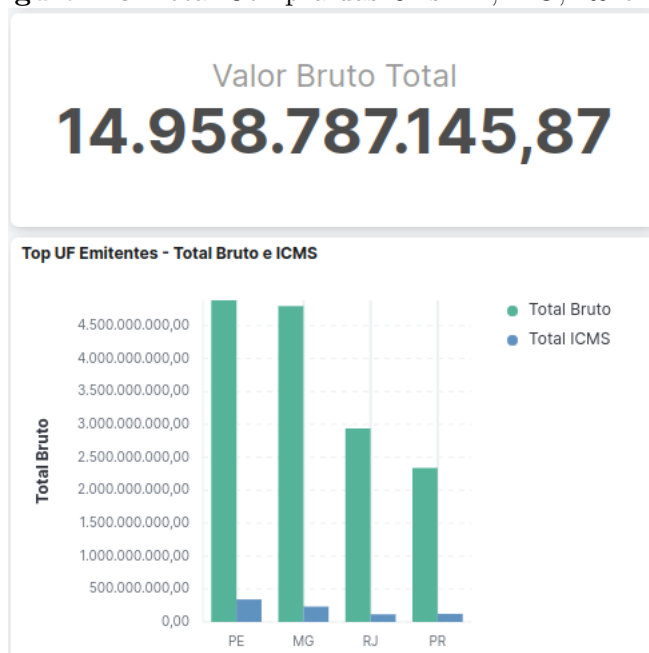


Fonte: Autor, 2023.

Aprofundando as análises da simulação, é possível obter algumas percepções, tais como:

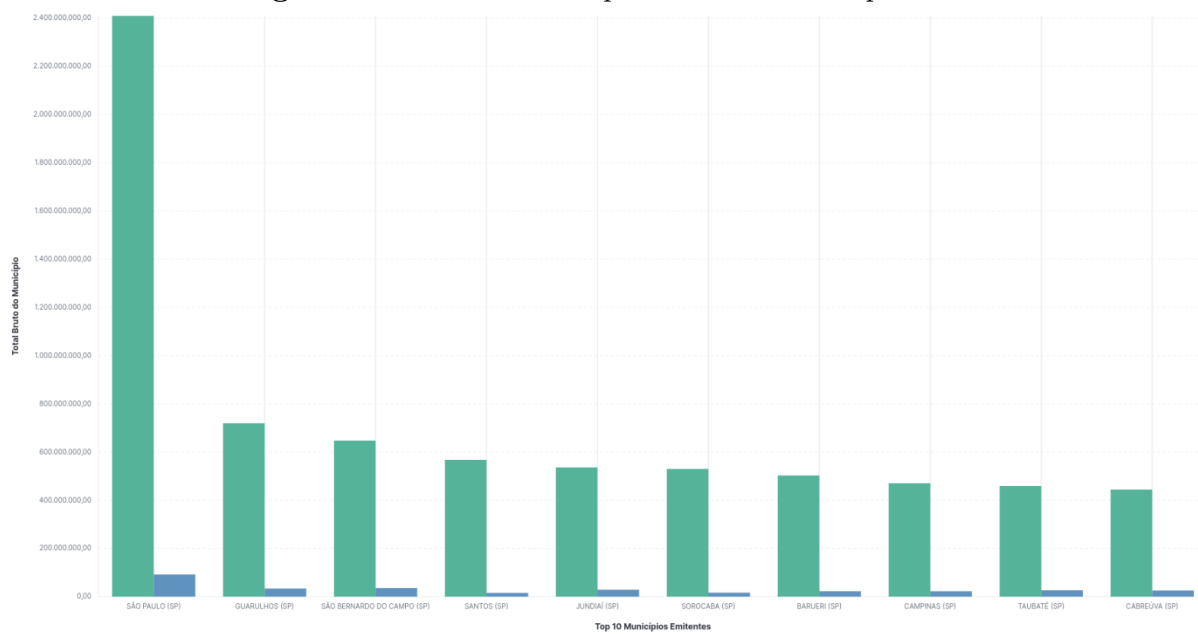
- A UF que mais vende mercadorias para a Bahia é São Paulo (R\$17,65 bi);
- As UFs PE (R\$ 4,88 bi), MG (R\$ 4,79 bi), RJ (R\$ 2,93 bi) e PR (R\$ 2,93 bi), juntas, totalizaram R\$ 14,95 bi conforme Figura 4.3. Esta soma é inferior ao total adquirido de São Paulo.

Q3 - É possível identificar quais os maiores municípios fornecedores de um determinado estado?

Figura 4.3 Total Compra das UFs PE, MG, RJ e PR

Fonte: Autor, 2023.

Nesta simulação busca-se identificar quais os municípios do estado de São Paulo que mais forneceram mercadorias para a Bahia. A Figura 4.4, revela os dez maiores municípios fornecedores.

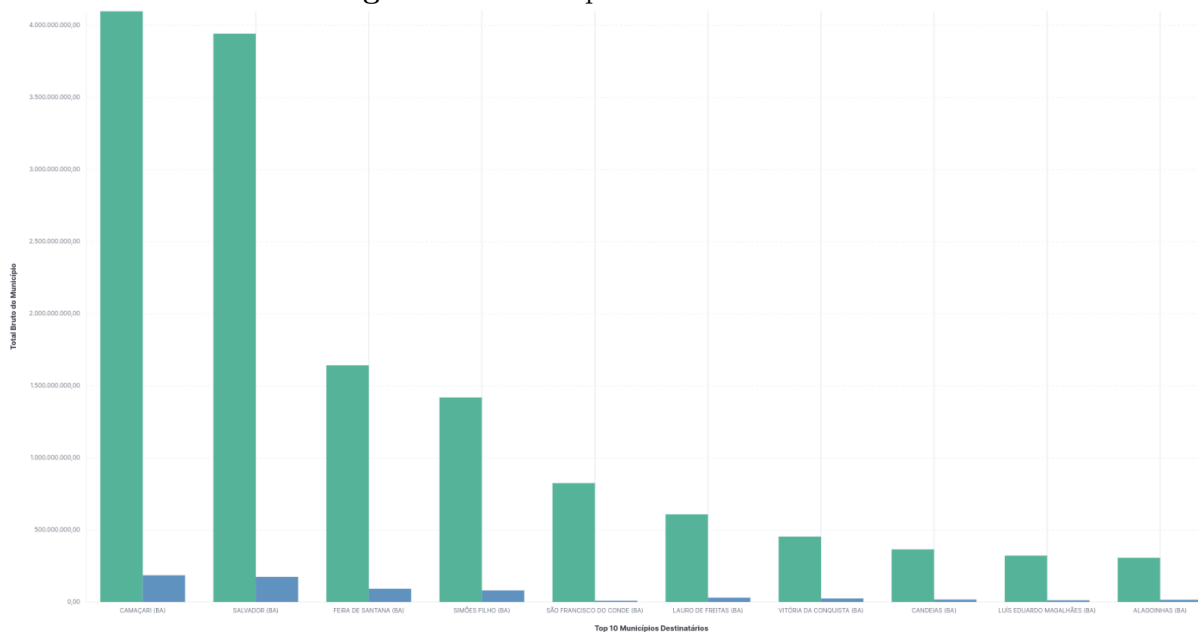
Figura 4.4 Maiores Municípios Emitentes - SP para BA

Fonte: Autor, 2023.

Identifica-se que:

- O município de São Paulo é responsável pela venda de mercadorias que totalizam R\$2,41 bi);
- A Figura 4.5 traz os municípios da Bahia que mais adquirem mercadorias do estado de São Paulo.

Figura 4.5 Municípios Destinatários - BA



Fonte: Autor, 2023.

- A Figura 4.6 traz o detalhamento do valor total bruto e o total de ICMS dos municípios da Bahia que mais adquirem mercadorias do estado de São Paulo.

Figura 4.6 Detalhe - Municípios Destinatários - BA**Top 10 Municípios Destinatários - Total Bruto e ICMS**

Top 10 Municípios Destinatários	Total Bruto do Município	Total ICMS do Município
CAMAÇARI (BA)	4.097.724.709,0315	185.005.148,6687
SALVADOR (BA)	3.942.804.301,2405	174.049.591,0063
FEIRA DE SANTANA (BA)	1.641.708.515,2322	91.612.357,5888
SIMÕES FILHO (BA)	1.418.855.946,2623	80.492.472,6733
SÃO FRANCISCO DO CONDE (BA)	824.645.199,7255	8.379.748,5069
LAURO DE FREITAS (BA)	607.785.284,3769	30.145.995,4533
VITÓRIA DA CONQUISTA (BA)	453.573.449,5317	24.575.348,0769
CANDEIAS (BA)	365.053.164,5408	16.776.514,3367
LUÍS EDUARDO MAGALHÃES (BA)	321.778.044,3562	11.559.809,8946
ALAGOINHAS (BA)	306.381.430,1762	15.049.429,5935

Fonte: Autor, 2023.

- Identifica-se que Camaçari é o município que mais adquiriu mercadorias do estado de São Paulo. Um volume total bruto de R\$ 4,09 bi em mercadorias.
- Observa-se que São Francisco do Conde adquiriu R\$ 824,6 mi em mercadorias e o valor total de ICMS foi R\$ 8,38 mi, e apesar do município de Lauro de Freitas ter adquirido R\$ 607,8 mi em mercadorias, um valor inferior, o total de ICMS foi superior ao de São Francisco do Conde, com o valor de R\$ 30,14 mi.

Q4 - É possível identificar as mercadorias negociadas entre municípios específicos?

Nesta simulação objetiva-se identificar quais as mercadorias obtidas por Camaçari, o município da Bahia que mais adquiriu mercadorias do estado de São Paulo. Para tanto, observou-se que o maior fornecedor é o município de Taubaté, A Figura 4.7 apresenta os dez maiores municípios fornecedores.

Figura 4.7 Municípios que mais venderam a Camaçari-BA

Top 10 Municípios Emitentes	Total Bruto do Município	Total ICMS do Município
TAUBATÉ (SP)	342.327.405,9655	19.912.580,8188
SÃO PAULO (SP)	305.368.748,1705	12.396.003,4868
SÃO BERNARDO DO CAMPO (SP)	275.984.394,7374	13.934.859,8581
SANTOS (SP)	220.754.606,0875	6.356.271,016
LIMEIRA (SP)	217.962.978,1766	9.148.863,917
CAMPINAS (SP)	192.959.114,5519	11.205.288,3238
SOROCABA (SP)	161.812.148,627	3.481.911,5565
GUARULHOS (SP)	152.594.915,6056	6.836.282,2302
JUNDIAÍ (SP)	151.778.069,8257	8.250.441,4303
SANTO ANDRÉ (SP)	149.243.413,4101	8.814.398,5211

Fonte: Autor, 2023.

O *dashboard* traz visualizações que apresentam as mercadorias comercializadas. A Figura 4.8 traz o *Treemap* com as mercadorias compradas pelo município de Camaçari-BA e que foram vendidas por Taubaté-SP, e a Tabela 4.1 traz o detalhamento do volume por mercadoria adquirida.

Figura 4.8 Mercadorias de Taubaté-SP para Camaçari-BA - Total Bruto (%)

Fonte: Autor, 2023.

Tabela 4.1 Detalhe - Mercadorias de Taubaté-SP para Camaçari-BA

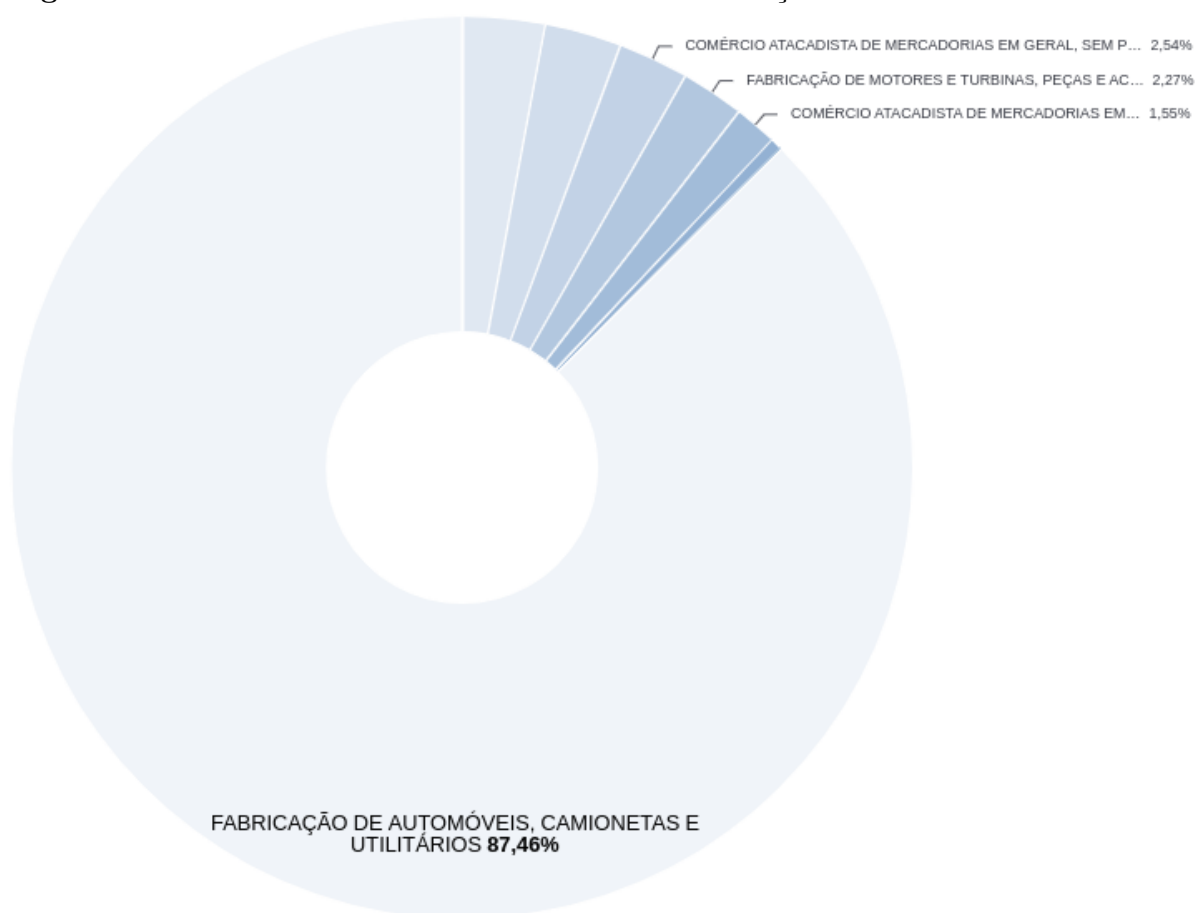
NCM	Mercadoria	Total Bruto (R\$)
87084090	87-VEÍCULOS AUTOMÓVEIS, TRATORES, CICLOS E OUTROS VEÍCULOS TERRESTRES, SUAS PARTES E ACESSÓRIOS. 8708-PARTES E ACESSÓRIOS DOS VEÍCULOS AUTOMÓVEIS DAS POSIÇÕES 87.01 A 87.05. 870840- CAIXAS DE MARCHAS (VELOCIDADES*) E SUAS PARTES 87084090-PARTES	87.490.451,20
73269090	73-OBRAS DE FERRO FUNDIDO, FERRO OU AÇO. 7326-OUTRAS OBRAS DE FERRO OU AÇO. 732690- OUTRAS 73269090-OUTRAS	13.730.661,86
87083090	87-VEÍCULOS AUTOMÓVEIS, TRATORES, CICLOS E OUTROS VEÍCULOS TERRESTRES, SUAS PARTES E ACESSÓRIOS. 8708-PARTES E ACESSÓRIOS DOS VEÍCULOS AUTOMÓVEIS DAS POSIÇÕES 87.01 A 87.05. 870830- FREIOS (TRAVÕES) E SERVOFREIOS SUAS PARTES 87083090-OUTROS	5.622.732,19
39231090	39-PLÁSTICO E SUAS OBRAS. 3923-ARTIGOS DE TRANSPORTE OU DE EMBALAGEM, DE PLÁSTICO ROLHAS, TAMPAS, CÁPSULAS E OUTROS DISPOSITIVOS PARA FECHAR RECIPIENTES, DE PLÁSTICO. 392310- CAIXAS, CAIXOTES, ENGRADADOS E ARTIGOS SEMELHANTES 39231090-OUTROS	5.492.113,48
87087090	87-VEÍCULOS AUTOMÓVEIS, TRATORES, CICLOS E OUTROS VEÍCULOS TERRESTRES, SUAS PARTES E ACESSÓRIOS. 8708-PARTES E ACESSÓRIOS DOS VEÍCULOS AUTOMÓVEIS DAS POSIÇÕES 87.01 A 87.05. 870870- RODAS, SUAS PARTES E ACESSÓRIOS 87087090-OUTROS	4.036.856,39
73209000	73-OBRAS DE FERRO FUNDIDO, FERRO OU AÇO. 7320-MOLAS E FOLHAS DE MOLAS, DE FERRO OU AÇO. 73209000- OUTRAS	3.295.285,18
87089990	87-VEÍCULOS AUTOMÓVEIS, TRATORES, CICLOS E OUTROS VEÍCULOS TERRESTRES, SUAS PARTES E ACESSÓRIOS. 8708-PARTES E ACESSÓRIOS DOS VEÍCULOS AUTOMÓVEIS DAS POSIÇÕES 87.01 A 87.05. 87089- OUTRAS PARTES E ACESSÓRIOS 870899- OUTROS 87089990-OUTROS	950.140,62

Fonte: Autor, 2023.

Observa-se também que:

- Cerca de 33% das mercadorias que Camaçari-BA adquiriu de Taubaté-SP pertencem ao grupo 8708 de Partes/Acessórios de Veículos Automotores, compreendendo um valor total superior a R\$ 112 mi;
- A Figura 4.9 revela que 87% das mercadorias adquiridas foram relacionadas à atividade econômica de fabricação de automóveis, camionetas e utilitários. Essas aquisições totalizaram um valor bruto de R\$ 299,1 mi;

Figura 4.9 Atividades Econômicas - Mercadorias de Camaçari-BA oriundas de Taubaté-SP

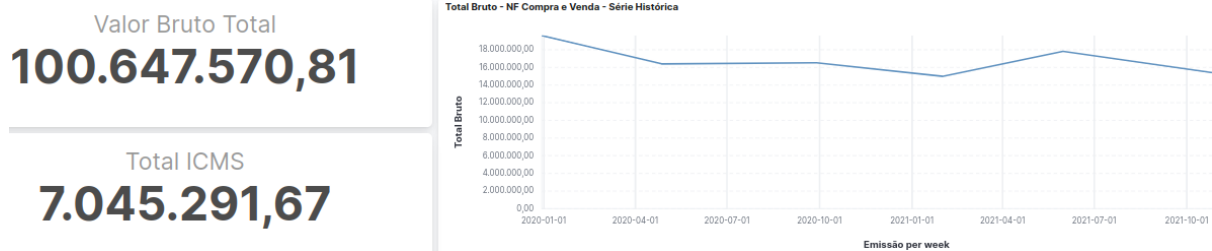


Fonte: Autor, 2023.

Q5 - É possível identificar quanto foi o total de ICMS em um determinado período sobre a comercialização de produtos específicos?

Esta simulação visa apresentar qual o total de ICMS sobre as mercadorias adquiridas ao longo de seis meses. Essas aquisições foram feitas pelos municípios da Bahia e as mercadorias pertenciam ao grupo 721720 - Fios de Aço ou Ferro. A Figura 4.10 traz essa informação e objetiva agregar conhecimento para a projeção de recolhimento de ICMS e avaliação de atração de investimentos.

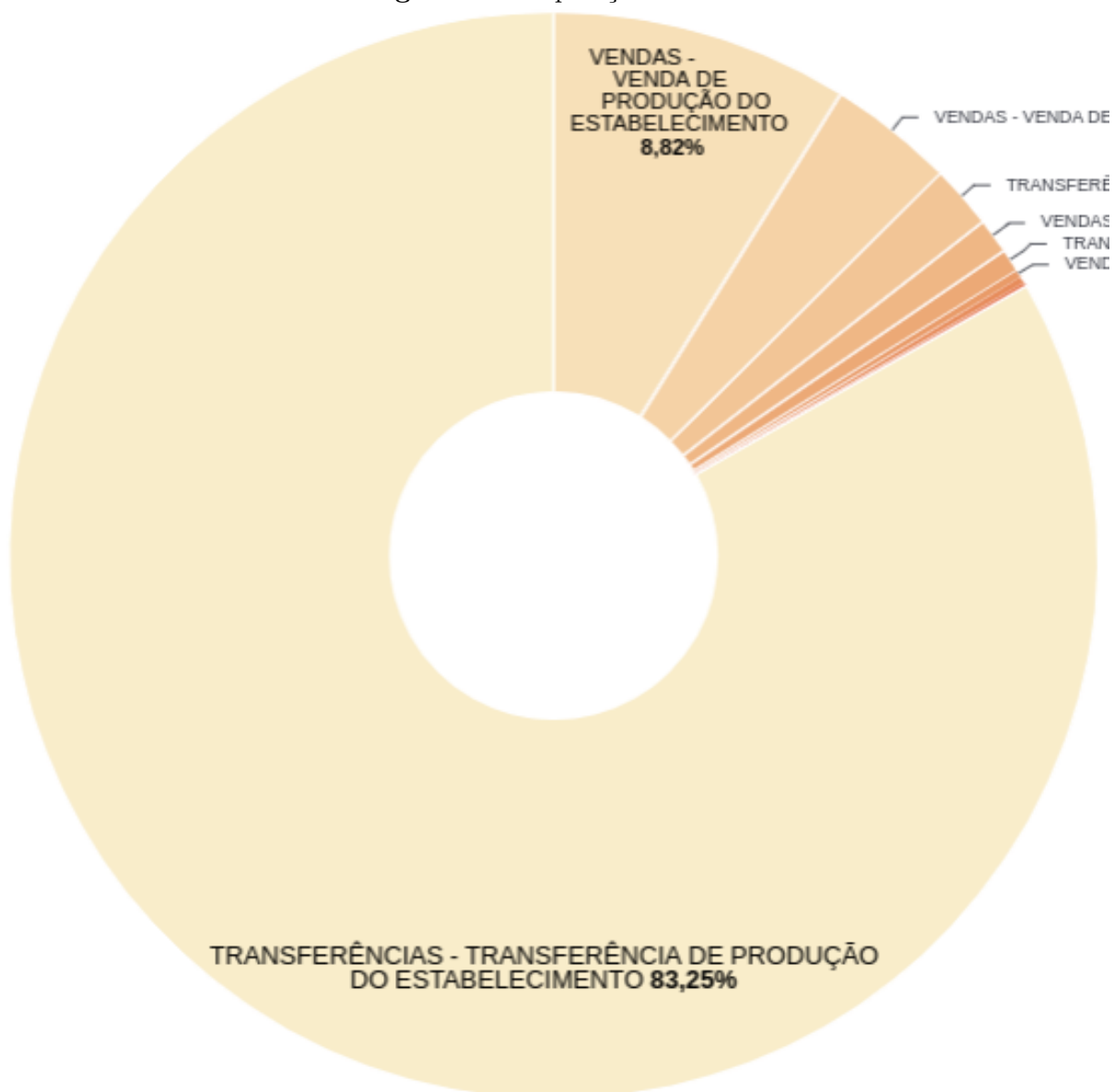
Figura 4.10 Total Bruto e Total de ICMS



Fonte: Autor, 2023.

Observa-se também que 83% do valor total do ICMS das operações fiscais são referentes às transferências de produção do estabelecimento. Essa distribuição pode ser observada na Figura 4.11.

Figura 4.11 Operações Fiscais



Fonte: Autor, 2023.

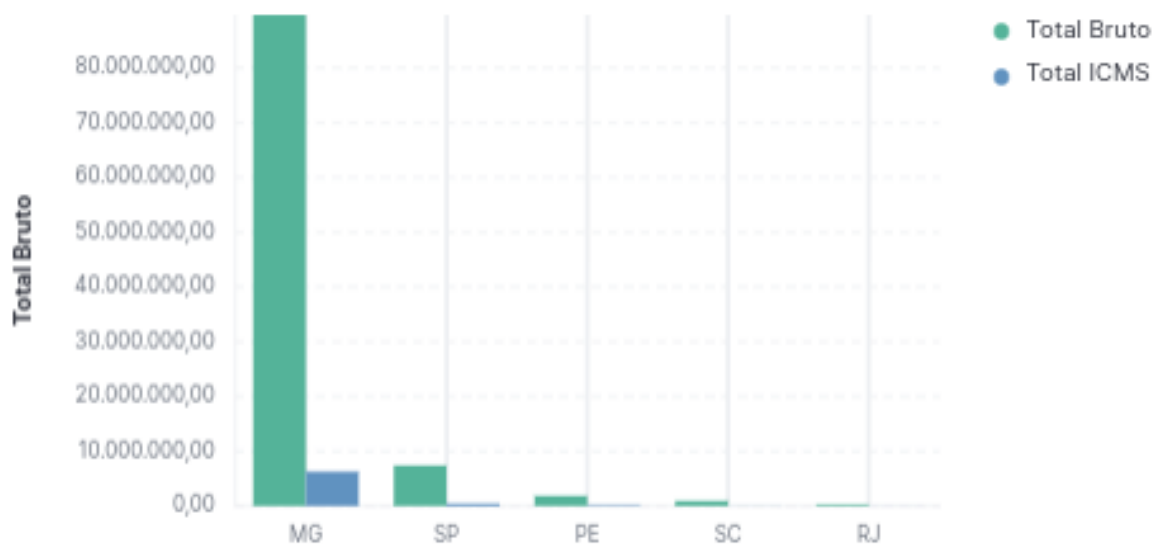
Q6 - O sistema é capaz de auxiliar na busca de alternativas de substituição de região fornecedora de determinada mercadoria?

Esta simulação visa compreender a dependência comercial de uma determinada região e a busca por uma alternativa de regiões fornecedoras de mercadorias específicas.

Na visualização apresentada na Figura 4.12, observa-se que foi adquirido um valor total de R\$ 100 mi em mercadorias do grupo 721720 - Fios de Aço ou Ferro. Observa-se também que o estado de Minas Gerais é o maior fornecedor.

Figura 4.12 UFs Emitentes - Fios de ferro galvanizados

Valor Bruto Total
100.647.570,81

Top UF Emitentes - Total Bruto e ICMS**Top 5 UF Emitentes**

Fonte: Autor, 2023.

Na Figura 4.13, observa-se que os municípios mineiros de Contagem, Sabará e Ponte Nova, somaram um valor total de R\$ 89,57 mi.

Figura 4.13 Detalhe - UFs Emitentes - Fios de ferro galvanizados

Top 10 Municípios Emitentes	Total Bruto do Município	Total ICMS do Município
CONTAGEM (MG)	57.289.131,117	4.004.628,5713
SABARÁ (MG)	31.961.465,9924	2.237.302,7538
ORLÂNDIA (SP)	3.224.973,7892	225.789,7787
HORTOLÂNDIA (SP)	3.032.398,5536	189.884,2892
RECIFE (PE)	1.833.940,5653	220.067,0506
JOINVILLE (SC)	773.414,088	34.624,64
PONTE NOVA (MG)	320.909,3667	22.351,9201
SÃO PAULO (SP)	309.123,3215	2.337,62
RIO DE JANEIRO (RJ)	290.351,1174	19.772,8299
COTIA (SP)	254.702,448	18.564,5599

Fonte: Autor, 2023.

Percebe-se também que é possível realizar a substituição parcial da comercialização de fios com o estado de Minas Gerais, pois outros estados também comercializam essa mercadoria, mas o volume total não é suficiente para uma substituição plena.

4.3 RESULTADOS

Tabela 4.2 Métricas para Q1

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 120s$
M3	Independência de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

A Tabela 4.3 traz as métricas, as unidades de medida e as medições obtidas nesta simulação.

Tabela 4.3 Métricas para Q2

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 120s$
M3	Independe de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

A Tabela 4.4 traz as métricas, as unidades de medida e as medições obtidas nesta simulação.

Tabela 4.4 Métricas para Q3

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 150s$
M3	Independe de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

A Tabela 4.5 traz as métricas, as unidades de medida e as medições obtidas nesta simulação.

Tabela 4.5 Métricas para Q4

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 120s$
M3	Independe de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

A Tabela 4.6 traz as métricas, as unidades de medida e as medições obtidas nesta simulação.

Tabela 4.6 Métricas para Q5

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 180s$
M3	Independência de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

A Tabela 4.7 traz as métricas, as unidades de medida e as medições obtidas nesta simulação.

Tabela 4.7 Métricas para Q6

Métrica	Descrição	Unidade de Medida	Resultado
M1	Objetivos alcançados	Sim/Não	Sim
M2	Tempo de tarefa	Segundos	$\leq 180s$
M3	Independência de proficiência	Sim/Não	Sim

Fonte: Autor, 2023.

4.4 CONCLUSÃO

Após finalizado o primeiro estudo experimental e analisadas as medições obtidas, notou-se que os objetivos de todas as questões foram alcançados e que a busca por respostas às questões levantadas levou, em média, tempo inferior a 180s. Dentro do contexto dessas simulações, com a base que foi disponibilizada.

Considera-se então que esse estudo foi capaz de analisar o VisualData-NFe e verificar que a solução foi eficaz em oferecer compreensão do perfil produtivo e da dinâmica econômica de municípios através da análise visual exploratória.

Entendeu-se, porém, que esta foi uma avaliação preliminar e que havia espaço para evolução na validação da ferramenta, principalmente considerando usuários reais especialistas de domínio.

Este capítulo apresenta o segundo estudo exploratório realizado a fim de avaliar as questões de usabilidade do VisualData-NFe.

ESTUDO EXPLORATÓRIO II

5.1 OBJETIVO

Em dezembro de 2023 foi realizado o segundo estudo para avaliação do VisualData-NFe. Este estudo foi desenvolvido a partir do método proposto por Nielsen (1994) para avaliação da usabilidade de uma interface através de análise heurística, adaptado para uso das heurísticas desenvolvidas por Few (2006), específicas para a avaliação de painéis de visualização de informações.

O objetivo foi analisar a usabilidade do *dashboard* do VisualData-NFe com a finalidade de verificar se o painel de visualização é eficaz em apresentar as informações de forma clara, eficiente e compreensível, no contexto da análise visual exploratória de notas fiscais eletrônicas sintéticas de compra e venda do estado da Bahia.

5.2 MÉTODO

O desenvolvimento deste estudo experimental ocorreu a partir das seguintes etapas:

1. Definição do Ambiente Experimental;
2. Elaboração da Avaliação da Usabilidade do VisualData-NFe *Dashboard*;
3. Coleta e Análise dos Dados.

O detalhamento das etapas está organizado e distribuído nas próximas subseções.

5.2.1 Definição do Ambiente Experimental

Este estudo utilizou dados sintéticos fictícios de notas fiscais eletrônicas. Foram gerados e carregados na plataforma cerca de 36 milhões de dados sintéticos de notas, distribuídos ao longo de seis meses, ocupando um espaço de armazenamento de 33,8 GB. Os dados totalizam R\$ 311,9 bi em valores brutos e R\$ 12,8 bi em valores de ICMS.

Os dados gerados foram baseados no perfil produtivo de estados e municípios obtido dos dados publicados pelo Observatório da Agropecuária Brasileira (INMET, 2022) e pela Confederação Nacional da Indústria (CNI, 2022).

5.2.2 Elaboração da Avaliação da Usabilidade do VisualData-NFe Dashboard

Nesta seção são retomados os aspectos elencados por Few (2006) que nortearam a elaboração do instrumento de coleta de dados, bem como as heurísticas que foram utilizadas para a avaliação da usabilidade do painel de visualização de informações do VisualData-NFe.

- **Informações organizadas para apoiar seu significado e uso:** Grupos de visualizações organizados de acordo com funções de negócios, entidades e uso. Reunir itens que pertencem ao mesmo grupo;
- **Consistência mantida para uma interpretação rápida e precisa:** Tipos de exibição que melhor comunicam os dados;
- **Visualização esteticamente agradável:** Cores escolhidas apropriadamente, imagens gráficas alta resolução e escolha correta do texto;
- **Design para uso como plataforma de lançamento:** Aprofundar os detalhes (*Drill Down*), “fatiar” os dados e permitir o lançamento para o detalhamento a partir de cliques nos próprios dados;
- **Avaliação da Usabilidade nos painéis criados:** Avaliação do bom funcionamento do painel por usuários finais. Painéis preenchidos com dados reais:
 1. **Foco:** O painel deve concentrar as informações mais importantes e relevantes para os usuários;
 2. **Visibilidade:** As informações devem estar visíveis de maneira clara e sem ambiguidades;
 3. **Compreensibilidade:** Os usuários devem compreender as informações apresentadas no *dashboard* sem esforço;
 4. **Consistência:** O *dashboard* deve manter uma consistência visual e semântica em sua totalidade;
 5. **Eficiência de Uso do Espaço:** O espaço disponível deve ser usado de maneira eficiente, evitando excesso de informações desnecessárias;

6. **Avaliação Comparativa:** A comparação entre diferentes conjuntos de dados ou elementos no *dashboard* deve ser facilitada;
7. **Integração de Elementos Gráficos e Tabelas:** Gráficos e tabelas devem ser integrados de maneira eficaz para uma compreensão aprimorada;
8. **Concisão:** As informações devem ser apresentadas de forma concisa, evitando excesso de detalhes ou elementos;
9. **Contexto Adequado:** As informações devem ser apresentadas no contexto apropriado, a fim de facilitar a interpretação;
10. **Design Estético:** O *dashboard* deve ser focado na clareza e na facilidade de leitura.

O instrumento de avaliação foi composto por dez itens de análise objetiva que foram graduados de acordo com a escala proposta por Likert (1932), onde os valores de atribuição variam de 1 a 5. A saber: **1. Discordo totalmente; 2. Discordo parcialmente; 3. Nem concordo, nem discordo (Neutro); 4. Concordo parcialmente e 5. Concordo totalmente.** Os itens de avaliação que compuseram o formulário foram:

1. O painel apresentou informações importantes e relevantes? (Foco)
2. As informações foram apresentadas de forma clara e sem ambiguidades? (Visibilidade)
3. As informações apresentadas foram compreensíveis? (Compreensibilidade)
4. Ao clicar nos elementos do *dashboard*, as tarefas e ações solicitadas foram efetivadas? (Consistência Semântica)
5. Os espaços foram utilizados adequadamente pelas visualizações no painel? (Uso do Espaço)
6. O painel permitiu comparar dados? P.Ex. Comparar os cinco ou dez maiores em um conjunto de dados. (Avaliação Comparativa)
7. Os elementos gráficos no painel se mostraram integrados? (Integração)
8. As informações foram apresentadas de forma concisa? (Concisão)
9. As informações foram apresentadas dentro de um contexto apropriado? (Contexto Adequado)
10. As informações foram exibidas de forma clara e com facilidade para leitura? (*Design Estético*)

Este instrumento foi disponibilizado através de um link encurtado do *Google Forms*¹.

¹<https://acesse.one/vswMw>

5.2.3 Coleta e Análise dos Dados

A coleta dos dados respeitou a resolução CNS n.º 510, de 2016, em seu artigo 2º, XIV, onde participantes não identificados, através de consulta por metodologia específica, foram convidados a avaliar o produto. Este tipo de pesquisa dispensa a apreciação de Comitê de Ética em Pesquisa/Comissão Nacional de Ética em Pesquisa - CEP/CONEP (BRASIL, 2016).

O estudo realizado por Nielsen e Molich (1990) concluiu que os resultados de uma avaliação heurística serão muito melhores se várias pessoas conduzirem a avaliação, e elas provavelmente deverão fazê-lo independentemente umas das outras. O estudo citado recomenda ainda que a avaliação heurística seja feita com **três a cinco avaliadores**.

Esse estudo para avaliação da usabilidade do VisualData-NFe, contou com a participação de cinco usuários com os seguintes perfis:

- 1 realiza atividades e/ou pesquisas em Economia;
- 3 realizam pesquisas e/ou desenvolvimento de painéis de visualização (*Dashboards*);
- 1 realiza pesquisas e/ou desenvolvimento em *Big Data*.

A Figura 5.1 traz o resultado da questão sobre o perfil que melhor correspondia ao da pessoa avaliadora.

Figura 5.1 Perfis

Qual atividade mais se aproxima do seu perfil?

5 respostas



Fonte: Autor, 2023.

5.3 RESULTADOS

A Tabela 5.1 apresenta os resultados sobre os itens de avaliação da usabilidade. Observa-se que não houve respostas “Discordo Parcialmente” nem “Discordo Totalmente” para nenhum dos itens avaliados.

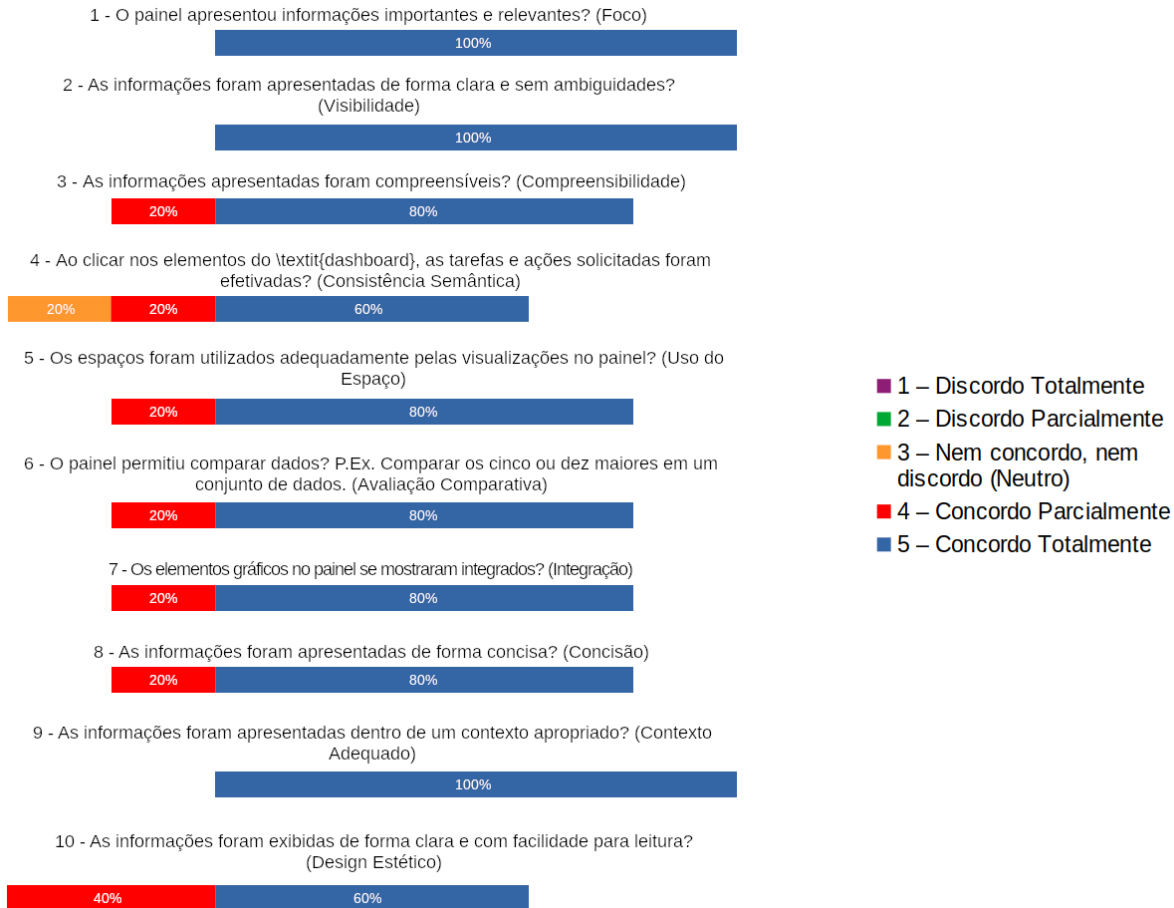
Tabela 5.1 Resultados

Item	Nem concordo, nem discordo (neutro)	Concordo Parcialmente	Concordo Totalmente
1 - O painel apresentou informações importantes e relevantes? (Foco)	0%	0%	100%
2 - As informações foram apresentadas de forma clara e sem ambiguidades? (Visibilidade)	0%	0%	100%
3 - As informações apresentadas foram compreensíveis? (Compreensibilidade)	0%	20%	80%
4 - Ao clicar nos elementos do <i>dashboard</i> , as tarefas e ações solicitadas foram efetivadas? (Consistência Semântica)	20%	20%	60%
5 - Os espaços foram utilizados adequadamente pelas visualizações no painel? (Uso do Espaço)	0%	20%	80%
6 - O painel permitiu comparar dados? P.Ex. Comparar os cinco ou dez maiores em um conjunto de dados. (Avaliação Comparativa)	0%	20%	80%
7 - Os elementos gráficos no painel se mostraram integrados? (Integração)	0%	20%	80%
8 - As informações foram apresentadas de forma concisa? (Concisão)	0%	20%	80%
9 - As informações foram apresentadas dentro de um contexto apropriado? (Contexto Adequado)	0%	0%	100%
10 - As informações foram exibidas de forma clara e com facilidade para leitura? (<i>Design</i> Estético)	0%	40%	60%

Fonte: Autor, 2023.

A Figura 5.2 traz os resultados em escala Likert, através de um gráfico de barras.

Figura 5.2 Resultados - Escala Likert



Fonte: Autor, 2023.

5.4 CONCLUSÃO

Após finalizado o segundo estudo experimental e analisados os dados coletados, entendeu-se que os aspectos de usabilidade, sob a perspectiva das heurísticas e métodos apresentados anteriormente, foram respeitados e atendidos no VisualData-NFe.

Entende-se então, que a partir das avaliações dos usuários de diferentes perfis, o produto computacional proposto cumpre seu objetivo. No entanto, é reconhecido que há espaço para evoluções e melhorias.

Este capítulo traz as considerações finais, as limitações desta pesquisa, sugestões de trabalhos futuros e um plano global atual do mestrado.

CONSIDERAÇÕES FINAIS

6.1 CONSIDERAÇÕES

O presente trabalho teve como objetivo construir e apresentar o **VisualData-NFe**, solução capaz de oferecer mecanismos de análise e exploração visual, em busca de compreensão do perfil produtivo e da dinâmica econômica de municípios. Através desta solução é possível o apoio na tomada de decisão em relação a investimentos. Buscou-se compreensão do perfil através dos dados de notas fiscais eletrônicas em detrimento de dados censitários, geralmente desatualizados.

Foram realizados processos de extração, transformação e carga de dados sintéticos de notas fiscais eletrônicas em uma solução de *Big Data*. Os resultados evidenciaram que esses processos são capazes de oferecer conhecimento e possibilitam a análise visual exploratória através de painéis interativos.

Através dos mecanismos de exploração visual oferecidos pela solução, foi possível identificar o recolhimento de ICMS em um certo período. De posse desta informação, torna-se viável estimar uma projeção de recolhimento e apoiar a tomada de decisão em relação a investimentos. Empresas podem ser atraídas por incentivos fiscais, e a instalação dessas empresas pode gerar renda e empregos diretos e indiretos.

Para fins de validação, foram conduzidos dois estudos experimentais envolvendo cerca de 34 milhões de dados sintéticos de notas fiscais, distribuídos ao longo de seis meses. Os dados totalizaram R\$ 290,8 bi em valores bruto e R\$ 12,2 bi em valores de ICMS. O VisualData-NFe conseguiu atingir o objetivo principal que era oferecer solução eficaz para compreensão do perfil produtivo e da dinâmica econômica de municípios através da análise visual exploratória. O produto computacional conseguiu atender aos aspectos de usabilidade.

6.2 LIMITAÇÕES E TRABALHOS FUTUROS

Este trabalho limita-se a tratar apenas com a estrutura econômica formal de produção, não sendo possível captar a parte informal da estrutura produtiva, onde não são emitidas

notas fiscais eletrônicas.

O produto computacional desta pesquisa também não é capaz de diferenciar a produção verticalmente integrada em conglomerados de empresas. Como por exemplo, uma empresa que produz partes e/ou subprodutos em comparação a uma outra empresa que não produz as partes, estando elas em um mesmo contexto, sendo apresentados cenários distintos na comparação, mas que possivelmente estariam representando a mesma estrutura produtiva.

Ao realizar pesquisas e experimentos em *Big Data*, compreende-se que alguns requisitos básicos de infraestrutura são exigidos. Este trabalho dispôs de recursos limitados de capacidade de processamento e armazenamento. Para superar os desafios impostos pelas limitações dos recursos, precisou-se suprimir, por exemplo, campos aglutinadores do *Elasticsearch* (message), que reduzem o tempo de resposta durante as pesquisa, mas que praticamente dobram o tamanho do espaço consumido em disco. Para contornar a questão do desempenho, o esquema de dados precisou ser projetado para responder de forma eficiente, ao tempo que se visava a entrega de valor e compreensão do perfil produtivo dos municípios tomando como base milhões de registros agrupados de notas fiscais.

Apesar dos resultados de ambos estudos exploratórios terem sido positivos, faz-se necessário evoluções e melhorias. Sugere-se como trabalho futuro a pesquisa e adoção de técnicas de mineração de dados textuais - *Knowledge Discovery from Text* (KDT), em busca da ampliação de conhecimento através de documentos, PDFs e outras fontes de dados não estruturadas para análise de perfil econômico de municípios.

Sugere-se também como proposta de trabalho futuro a realização de estudos de perfil econômico de municípios utilizando-se dados reais de notas fiscais eletrônicas através da solução apresentada neste trabalho.

Sugere-se ainda realizar pesquisa que analise a aplicação de algoritmos de reconhecimento de padrões e aprendizagem de máquina sobre dados reais de notas fiscais eletrônicas que foram tratados na plataforma desenvolvida.

REFERÊNCIAS BIBLIOGRÁFICAS

- BASILI, V. R.; ROMBACH, H. D. The goal question metric approach. *Encyclopedia of software engineering*, p. 528–532, 1994.
- BEVAN, N. et al. New iso standards for usability, usability reports and usability measures. In: SPRINGER. *Human-Computer Interaction. Theory, Design, Development and Practice: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part I 18*. [S.l.], 2016. p. 268–278.
- BRASIL. *Resolução CNS n.º 510, de 07 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais [provides for applicable standards in the Human and Social Sciences]*. [S.l.]: Diário Oficial da União, 2016.
- CHEN, M.; LIU, Y. Big data: a survey, mobile networks and application. hh, 2014.
- CHODOROW, K. *MongoDB: the definitive guide: powerful and scalable data storage*. [S.l.]: O'Reilly Media, 2013.
- CNI. 2022. Perfil Industrial nos Estados. Disponível em: <https://perfilindaindustria.portaldaindustria.com.br/>. Acesso em: 14 mar. 2022.
- COHN, M. *User stories applied: For agile software development*. [S.l.]: Addison-Wesley Professional, 2004.
- CONFAZ. 1970. Convênio s/n. de 15 de dezembro de 1970. Disponível em: https://www.confaz.fazenda.gov.br/legislacao/ajustes/sinief/cvsn/_70. Acesso em: 24 mai. 2022.
- COOPER, M.; MELL, P. 2012. Tackling big data. Disponível em: https://csrc.nist.gov/csrc/media/projects/forum/documents/2012/fcsm/_june2012/_cooper/_mell.pdf. Acesso em: 29 jan. 2023.
- ELASTIC. 2022. O que é o ELK Stack? Por que se chama Elastic Stack? Disponível em: <https://www.elastic.co/pt/what-is/elk-stack>. Acesso em: 13 jul. 2022.
- ELASTIC. 2023. Kibana Guide. Disponível em: <https://www.elastic.co/guide/en/kibana/8.6/introduction.html>. Acesso em: 13 jul. 2022.
- FEW, S. *Information dashboard design: The effective visual communication of data*. [S.l.]: O'Reilly Media, Inc., 2006.

GABA, D. M. The future vision of simulation in health care. *BMJ Quality & Safety*, BMJ Publishing Group Ltd, v. 13, n. suppl 1, p. i2-i10, 2004.

GEREFFI, G.; FERNANDEZ-STARK, K. Global value chain analysis: a primer. Duke CGGC (Center on Globalization, Governance & Competitiveness), 2016.

GIL, A. C. *Métodos e técnicas de pesquisa social*. [S.l.]: 6. ed. Editora Atlas SA, 2008. 200 p.

GORMLEY, C.; TONG, Z. *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. [S.l.]: "O'Reilly Media, Inc.", 2015.

IBGE. 2017. 1 p. Censo Agropecuário. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/21814-2017-censo-agropecuario.html?=&t=o-que-e>. Acesso em: 27 jul. 2022.

IBGE. 2022. Comissão Nacional de Classificação - CONCLA/IBGE. Disponível em: <https://concla.ibge.gov.br/busca-online-cnae.html?view=estrutura>. Acesso em: 16 jan. 2022.

INMET. 2022. Estatísticas do Observatório da Agropecuária Brasileira. Disponível em: <http://observatorio.agropecuaria.inmet.gov.br/plataformas/estatistica/>. Acesso em: 19 fev. 2022.

KEIM, D. et al. Visual analytics: Definition, process, and challenges. p. 157, 03 2008.

KEIM, D. A.; WARD, M. O. *Visual data mining techniques*. 2002.

KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the complete guide to dimensional modeling*. [S.l.]: John Wiley & Sons, 2011.

KITCHIN, R.; MCARDLE, G. The diverse nature of big data. programmable city working paper 15. Programmable City Working Paper, 2015.

KITCHIN, R.; MCARDLE, G. What makes big data, big data? exploring the ontological characteristics of 26 datasets. *Big Data & Society*, SAGE Publications Sage UK: London, England, v. 3, n. 1, p. 2053951716631130, 2016.

LANEY, D. et al. 3d data management: Controlling data volume, velocity and variety. *META group research note*, Stanford, v. 6, n. 70, p. 1, 2001.

LIKERT, R. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

MARQUES, C. A. G.; UCHÔA, C. F. A. Estimação e previsão do icms na bahia. *Revista Desenharia*, v. 3, n. 5, p. 196, 2006.

MARTIN, R. C.; NEWKIRK, J.; KOSS, R. S. *Agile software development: principles, patterns, and practices*. [S.l.]: Prentice Hall Upper Saddle River, NJ, 2003.

MENDONÇA, A. d. R. B. *Avaliação experimental de uma arquitetura de microsserviços para o gerenciamento de notas fiscais eletrônicas*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2022.

NETO, J. C. L. *Audita-NFSe: sistema auxiliar de auditoria em notas fiscais de serviços eletrônicas*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2021.

NIELSEN, J. *Heuristic evaluation, w: Nielsen J., Mack RL (eds.), usability inspection methods*. [S.l.]: John Wiley & Sons, New York, NY, 1994.

NIELSEN, J.; MOLICH, R. Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. [S.l.: s.n.], 1990. p. 249–256.

NOCKO, L. M. et al. Levantamento do perfil comercial do distrito federal baseado nas notas fiscais eletrônicas. 2017.

RFB. 2022. Nomenclatura Comum do MERCOSUL. Disponível em: <https://portalunico.siscomex.gov.br/classif/proxy/user?checkLogout=true&tabId=&perfil=publico>. Acesso em: 10 jan. 2022.

RFB. 2022. Nomenclatura Comum do MERCOSUL. Disponível em: <https://portalunico.siscomex.gov.br/classif/api/nomenclatura/consulta?ano=2021&codigo=&nomeExtenso=>. Acesso em: 10 jan. 2022.

RFB. *Sped - Sistema Público de Escrituração Digital - NF-e - O que é*. 2022. Disponível em: <http://sped.rfb.gov.br/pagina/show/1328>.

RFB. 2023. Notas Fiscais Eletrônicas - Informações sobre Estatísticas. Disponível em: <https://www.nfe.fazenda.gov.br/portal/infoEstatisticas.aspx>. Acesso em: 16 mar. 2023.

SADALAGE, P. J.; FOWLER, M. *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. [S.l.]: Pearson Education, 2013.

SANTOS, D. S. d. *Uma plataforma distribuída de mineração de dados para big data: um estudo de caso aplicado à Secretaria de Tributação do Rio Grande do Norte*. Dissertação (Mestrado) — Brasil, 2018.

SEFAZ-BA. *Código Fiscal de Operações e Prestações*. 2008.

SISCOMEX. 2022. NCM - Nomenclatura Comum do Mercosul. Disponível em: <https://www.gov.br/siscomex/pt-br/servicos/aprendendo-a-exportarr/planejando-a-exportacao-1/nomenclatura-comum-do-mercosul>. Acesso em: 21 ago. 2022.

SPEROTTO, F. Q. et al. O perfil econômico dos municípios da região de fronteira do mato grosso do sul, mato grosso, paran e rio grande do sul e o investimento produtivo e em infraestrutura na regio de fronteira do rio grande do sul. Instituto de Pesquisa Econmica Aplicada (Ipea), 2017.

SRIVASTAVA, A. *Kibana 7 Quick Start Guide: Visualize Your Elasticsearch Data with Ease*. [S.l.]: Packt Publishing Ltd, 2019.

STATISTA. 2021. Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. Disponível em: <https://www.statista.com/statistics/871513/worldwide-data-created/>. Acesso em: 09 fev. 2023.

WAZLAWICK, R. S. *Metodologia de pesquisa para ciência da computação*. [S.l.]: Elsevier, 2009.

Apêndice

A

MANUAL VISUALDATA-NFE



VisualData-NFe

Manual do Usuário

1 de outubro de 2023

Sumário

1	Apresentação	2
2	Pesquisa e Filtragem	4
2.1	Utilizando o Painel de Filtros	5
2.2	Filtrando através das Visualizações	5
2.3	Aplicando Filtro Manualmente	6
2.4	Filtrando através da KQL	7
3	Detalhamento e Exportação de Dados	8
4	Busca por Conhecimento	9
4.1	Como identificar o volume de mercadorias adquiridas de outros estados?	9
4.2	Como identificar de quais regiões as mercadorias foram adquiridas?	11
4.3	Como identificar quais os maiores municípios fornecedores de um determinado estado?	12
4.4	Como identificar as mercadorias negociadas entre municípios específicos?	13
4.5	Como identificar as atividades econômicas relacionadas as mercadorias que foram adquiridas?	14
4.6	Como identificar as operações fiscais relacionadas as mercadorias que foram adquiridas?	15

1 Apresentação

O VisualData-NFe é uma solução de *Visual Analytics* e *Big Data* com busca textual sobre dados de notas fiscais eletrônicas de compra e venda de mercadorias entre pessoas jurídicas. O objetivo principal da solução é oferecer mecanismos de análise e exploração visual, para compreensão do perfil produtivo e da dinâmica econômica de municípios.

Este manual visa ajudar o usuário a navegar no sistema e extrair o melhor do que a ferramenta de exploração visual tem a oferecer.

Ao acessar a aplicação, é exibido o *Dashboard* do VisualData-NFe. Ele fornece *insights* sobre os dados de várias perspectivas e permitem que os usuários se aprofundem nos detalhes. Este *Dashboard* é apresentado na próxima figura e é dividido em quatro áreas a saber:

1. **Área de Pesquisa e Filtragem:** Foram disponibilizados controles de filtragem como listas de opções, calendário e *range slider*;
2. **Área de Notas Fiscais:** Foram construídas visualizações exploratórias que permitem ao usuário buscar compreensões sobre como é a relação de compra e venda de mercadorias, volume de venda e aquisição, bem como as localidades envolvidas;
3. **Área de Mercadorias:** As visualizações disponibilizadas nesta área possibilitam ao usuário buscar compreensões sobre as mercadorias negociadas. É possível verificar o percentual do total bruto e total de ICMS comercializado;
4. **Área de Atividades Econômicas e Operações Fiscais:** Esta área traz visualizações que permitem acesso ao panorama geral de associação das operações de compra e venda com a atividade comercial desempenhada e a operação fiscal.

Para confecção deste manual, utilizou-se **dados sintéticos fictícios** de notas fiscais eletrônicas. Estes dados foram gerados e carregados na plataforma para servirem como simulação.



2 Pesquisa e Filtragem

A Figura a seguir, apresenta as variadas formas de filtragem dos dados, através do painel de filtros, diretamente pela visualização, e através do campo de consulta, que aceita a *KQL - Kibana Query Language*.

The image displays the Kibana search interface. On the left, a 'Filtros' (Filters) sidebar contains various filter categories, each with a 'Select...' dropdown menu:

- Total Bruto (with a range slider from 0 to 541099368)
- Município Emitente
- Município Destinatário
- UF Emitente
- UF Destinatário
- Atividade Econômica - CNAE
- Código Fiscal - CFOP
- NCM - Capítulo (2 dígitos - 00)
- NCM - Posição (4 dígitos - 0000)
- NCM - Subposição (6 dígitos - 000000)
- NCM Completo
- Descrição da Mercadoria

In the center, a visualization titled 'Total Bruto - NF Compra e Venda - Série Histórica' is shown. A modal dialog box titled 'Select filters to apply' is overlaid on the chart, with the following checked filters:

- data_timestamp: 27/04/2020 to 04/05/2020
- tipo_nota.keyword: COMPRA

Buttons for 'Cancel' and 'Apply' are visible at the bottom of the dialog.

Below the visualization, an 'Edit filter' dialog is open, showing the configuration for a filter:

- Field: sigla_uf_emitente
- Operator: is
- Value: SP
- Option: Create custom label?

Buttons for 'Cancel' and 'Save' are at the bottom of the dialog.

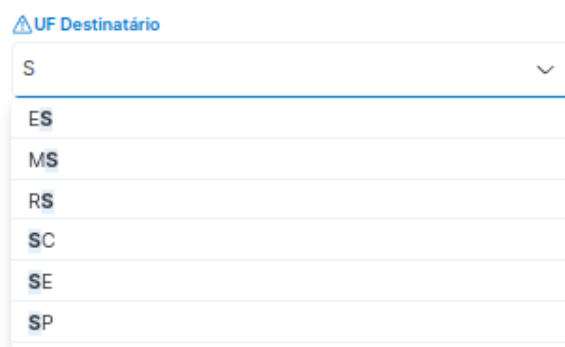
At the bottom of the page, the search query is displayed as 'NOT sigla_uf_emitente : BA' on the left and 'KQL' on the right.

2.1 Utilizando o Painel de Filtros

O painel de filtros está localizado do lado esquerdo, e reúne componentes visuais como listas de opções e um *range slider* para limitar os valores do total bruto.



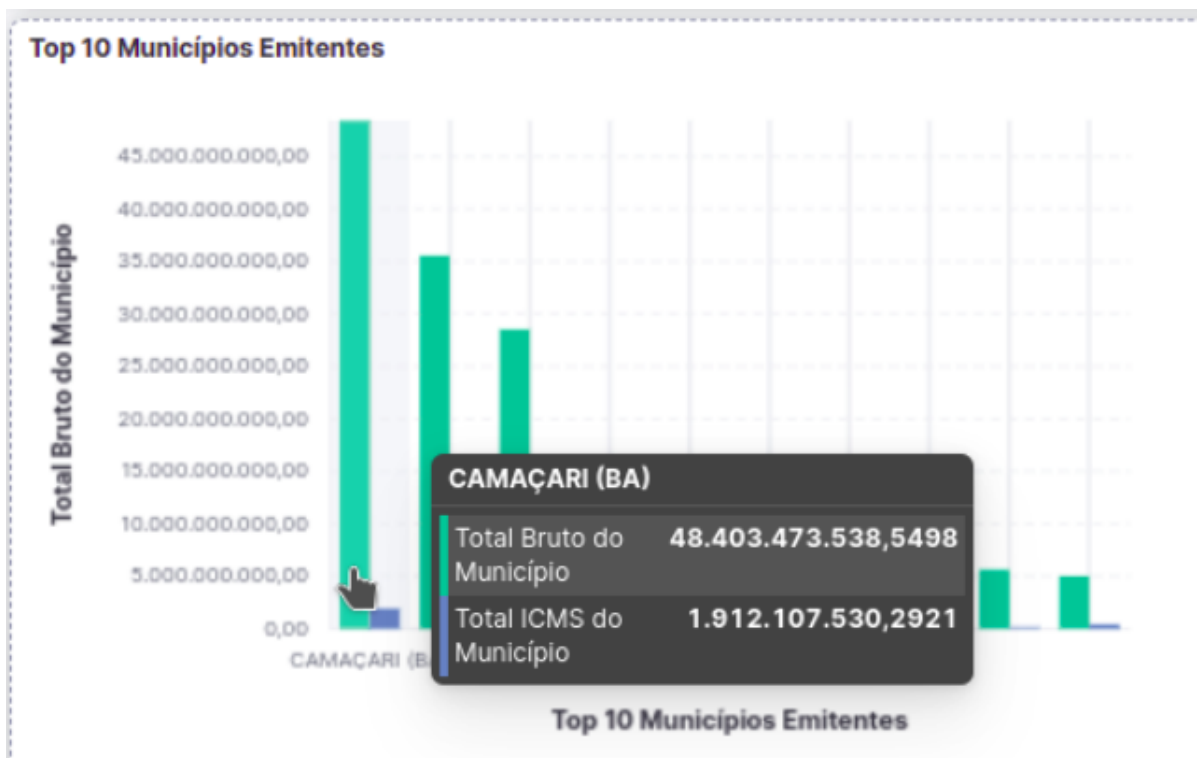
Nos campos como os de UFs e Municípios, o preenchimento deve ser feito com letras **MAIÚSCULAS**.



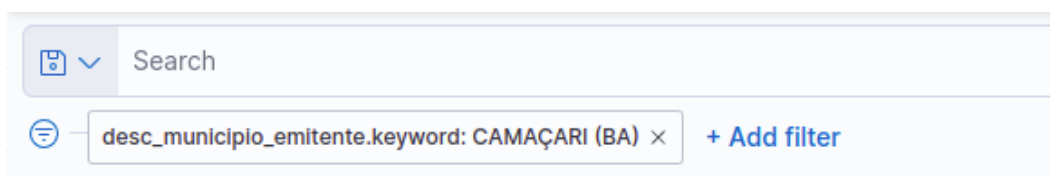
Uma vez aplicado um filtro de pesquisa, todos os painéis de visualização disponíveis no *Dashboard* serão afetados e atualizados de forma consistente.

2.2 Filtrando através das Visualizações

É possível aplicar filtros ao clicar diretamente nas visualizações. Como exemplo, segue uma figura que apresenta os 10 municípios com os maiores volumes de saída de mercadorias.



Ao clicar na barra que representa o volume do município, o filtro será aplicado, e percebe-se que fará parte do conjunto de filtros da barra de pesquisa na área superior do *Dashboard*.



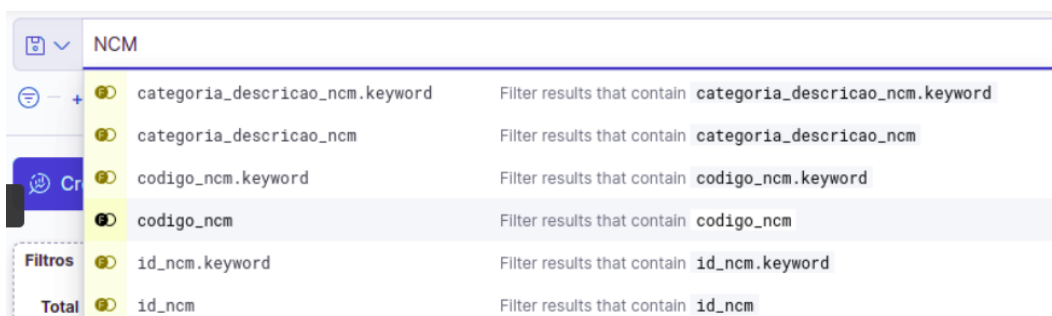
2.3 Aplicando Filtro Manualmente

Na parte superior do *Dashboard* é possível criar um filtro manualmente, através do botão *Add Filter*. Selecionar o atributo, o operador lógico e o valor. Clicar no botão *Save* e atualizar o *Dashboard* através do botão *Refresh*.

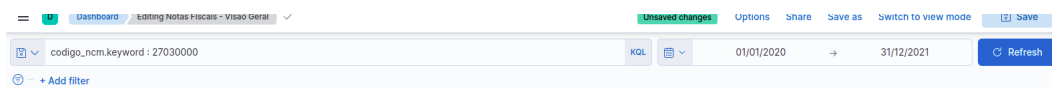


2.4 Filtrando através da KQL

A barra de pesquisa que fica localizada na parte superior, é capaz de realizar filtragem através da KQL, a linguagem de consulta do Kibana.

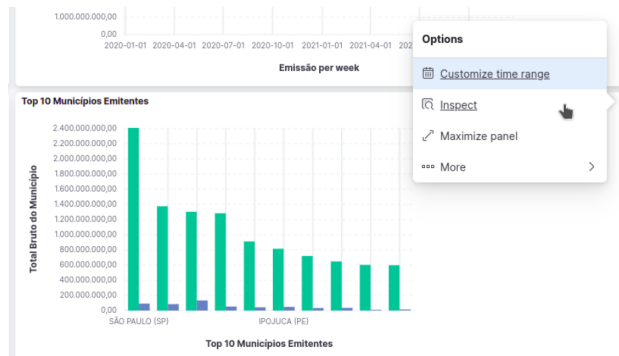


Após escolher o atributo do filtro, pode-se informar o ":" e o valor que deseja ser pesquisado. Em sequência, é preciso clicar em *refresh* para que todas as visualizações sejam sincronizadas com a pesquisa.



3 Detalhamento e Exportação de Dados

É possível obter informações detalhadas sobre os dados representados visualmente nos painéis. Ao passar o ponteiro do mouse sobre o canto superior direito de cada painel de visualização, será apresentado um pequeno botão "...". Ao clicá-lo, será exibido um menu suspenso com a opção *Inspect*.



Cada visualização apresentará os dados de acordo com os dados que são exibidos no painel. Para a inspeção de detalhamento do painel de municípios os dados são exibidos como a figura a seguir:

Top 10 Municípios Emitentes View: Data

[Download CSV](#)

Top 10 Municípios Emitentes	Total Bruto do Município	Total ICMS do Município
SÃO PAULO (SP)	2.409.803.303,7769	91.911.375,8302
CABO DE SANTO AGOSTINHO (PE)	1.375.805.055,9602	84.962.016,7529
MANAUS (AM)	1.303.138.610,817	133.078.781,8762
RIO DE JANEIRO (RJ)	1.284.101.095,0302	52.735.034,6969
RECIFE (PE)	912.076.142,599	42.790.834,9356
IPOJUCA (PE)	815.083.745,7425	47.967.543,6179
GUARULHOS (SP)	719.327.077,42	33.530.336,1061
SÃO BERNARDO DO CAMPO (SP)	647.604.821,3004	35.558.770,0476
ARACAJU (SE)	603.224.161,1542	9.838.232,6009
NATAL (RN)	597.921.210,7988	13.746.277,6002

Data not ready

Para exportar dados para um arquivo .CSV (separado por vírgulas), estando na tela do *inspect*, basta clicar no botão "Download CSV".

4 Busca por Conhecimento

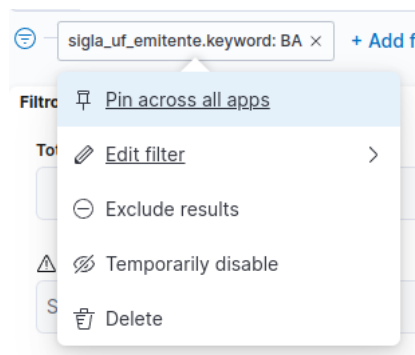
Esta seção visa apresentar ao usuário algumas possibilidades de configuração de filtros na busca por *insights* e conhecimento. Para todas as simulações, tomaremos como base, entradas e saídas de mercadorias no estado da Bahia.

4.1 Como identificar o volume de mercadorias adquiridas de outros estados?

Nesta simulação objetiva-se identificar o valor total bruto e o valor total de ICMS em relação às mercadorias que o estado da Bahia comprou de outros estados. Busca-se os valores acumulados e também distribuídos ao longo dos meses.

Configuração dos Filtros:

1. No Painel de Filtros do lado esquerdo, clique no campo UF Origem.
2. Escolha ou digite: **BA**. O filtro será aplicado e estará na parte superior do *Dashboard*.
3. Clique sobre o filtro e no menu suspenso que será exibido, clique em *Edit Filter*.



4. Modifique a operador **Is** para **Is Not** e clique em salvar.

The screenshot shows a dialog box titled "Edit filter" with a close button (X) and a "+ Add filter" button. The dialog contains the following fields:

- Field:** sigla_uf_emitente.keyword
- Operator:** is not
- Value:** BA
- Create custom label?

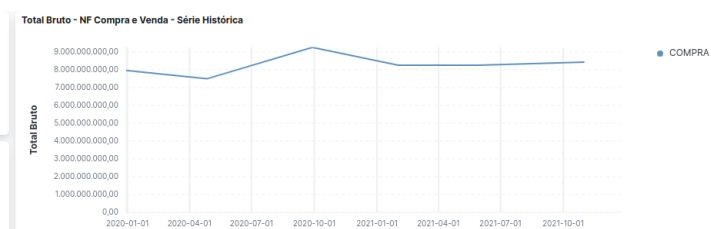
At the bottom right, there are "Cancel" and "Save" buttons.

5. Se necessário, clique em *Refresh* no canto superior direito, e aguarde os painéis serem atualizados.

Para onde olhar?

Valor Bruto Total
49.666.997.101,17

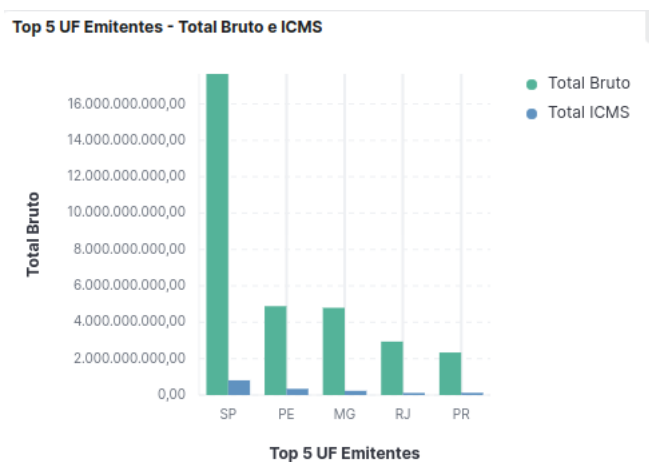
Total ICMS
2.800.839.680,78



4.2 Como identificar de quais regiões as mercadorias foram adquiridas?

Configuração dos Filtros: Repetir a configuração dos filtros da seção 4.1.

Para onde olhar?



Passar o ponteiro do mouse sobre a visualização e clicar no botão “...” que aparecerá no canto superior do painel , e clicar em *Inspect* para visualizar os detalhes de aquisição de mercadorias dos 5 estados que mais venderam à Bahia.

Top 5 UF Emitentes - Total Bruto e ICMS View: Data

[Download CSV](#)

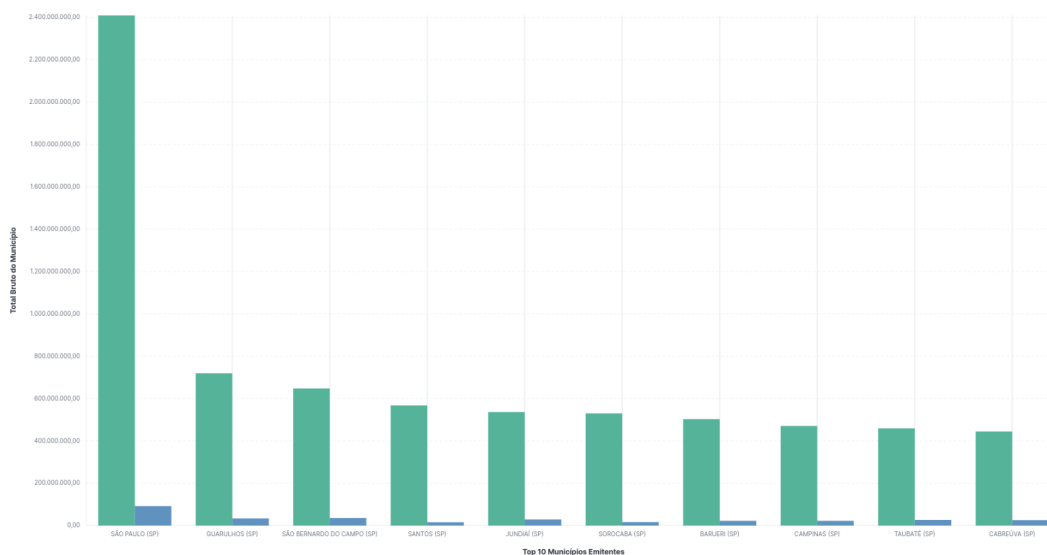
Top 5 UF Emitentes	Total Bruto	Total ICMS
SP	17.649.224.340,2042	810.031.187,1331
PE	4.884.670.459,7505	342.454.482,1396
MG	4.796.613.597,1294	233.412.096,2001
RJ	2.938.595.194,5136	116.261.204,5744
PR	2.338.907.894,4784	123.123.844,3219

4.3 Como identificar quais os maiores municípios fornecedores de um determinado estado?

Configuração dos Filtros: Repetir a configuração dos filtros da seção 4.1 e seguir os seguintes passos:

1. Na visualização de UF Origem, clicar na barra que representa **SP**, na simulação, o estado com o maior volume de vendas para a Bahia.
2. Se necessário, clique em *Refresh* no canto superior direito, e aguarde os painéis serem atualizados.

Para onde olhar?



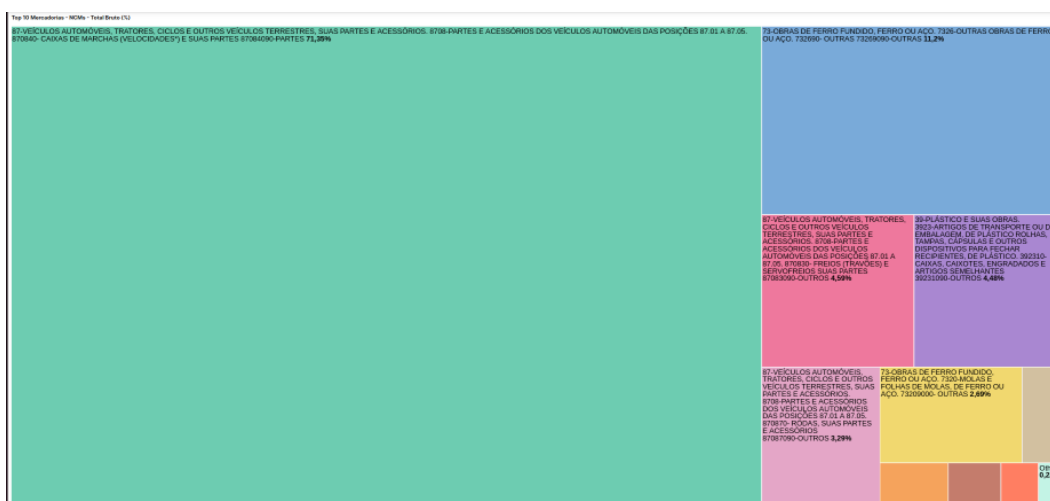
O detalhamento poderá ser acessado através da opção *Inspect* já informada anteriormente.

4.4 Como identificar as mercadorias negociadas entre municípios específicos?

Configuração dos Filtros: Repetir a configuração dos filtros da seção 4.3 e seguir os seguintes passos:

1. Na visualização de Município Origem, clicar na barra que representa **Taubaté (SP)**.
2. Na visualização de Município Destino, clicar na barra que representa **Camaçari (BA)**.
3. Se necessário, clique em *Refresh* no canto superior direito, e aguarde os painéis serem atualizados.

Para onde olhar?

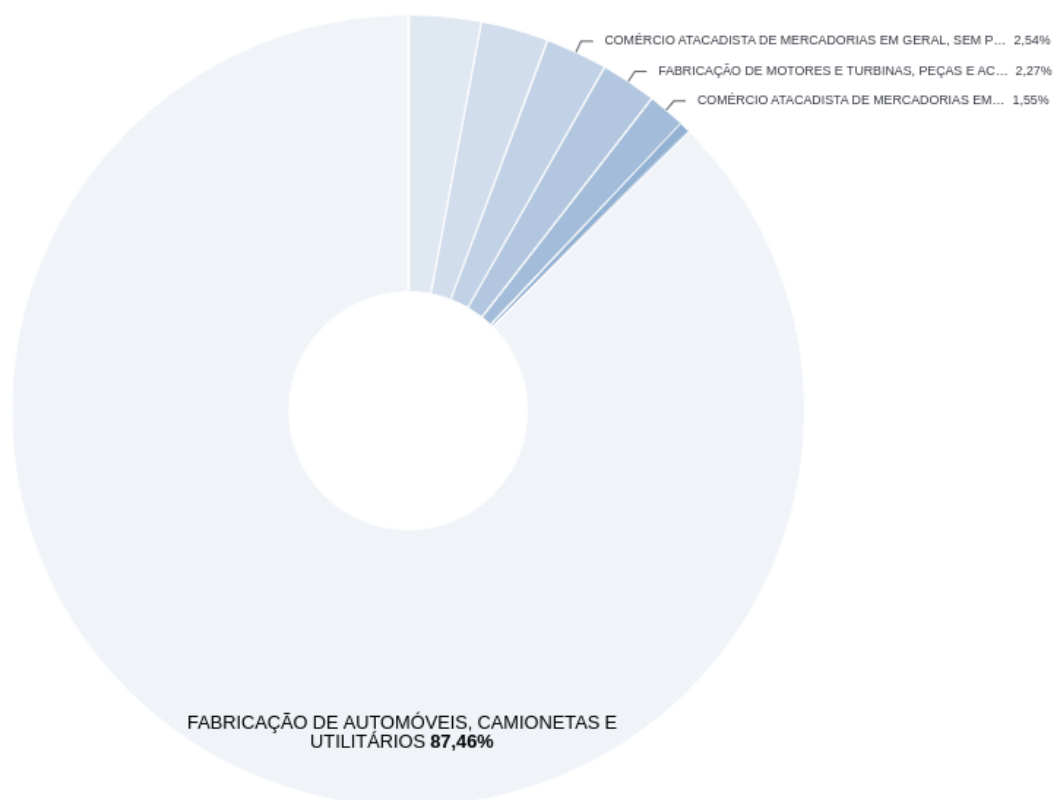


O detalhamento poderá ser acessado através da opção *Inspect* já informada anteriormente.

4.5 Como identificar as atividades econômicas relacionadas as mercadorias que foram adquiridas?

Configuração dos Filtros: Repetir a configuração dos filtros da seção 4.4.

Para onde olhar?



4.6 Como identificar as operações fiscais relacionadas as mercadorias que foram adquiridas?

Configuração dos Filtros: Repetir a configuração dos filtros da seção 4.4.

Para onde olhar?

