

**INSTITUTO  
FEDERAL**  
Pernambuco

INSTITUTO FEDERAL DE PERNAMBUCO  
CAMPUS BELO JARDIM  
BACHARELADO EM ENGENHARIA DE SOFTWARE

GABRIEL VASCONCELOS MARQUES

**DATAHARBOR: CAPACITANDO O ENGAJAMENTO CIDADÃO POR MEIO DA  
VISUALIZAÇÃO DE DADOS ABERTOS NA GOVERNANÇA MUNICIPAL**

Belo Jardim, Pernambuco  
2024

GABRIEL VASCONCELOS MARQUES

**DATAHARBOR: CAPACITANDO O ENGAJAMENTO CIDADÃO POR MEIO DA  
VISUALIZAÇÃO DE DADOS ABERTOS NA GOVERNANÇA MUNICIPAL**

**Trabalho de conclusão de Curso (TCC)** apresentado como requisito parcial para obtenção do grau de Bacharel em Engenharia de Software. TCC aprovado no curso de Engenharia de Software do IFPE - Campus Belo Jardim

**Banca de Qualificação:**

José Fernando da Silva - IFPE - Campus Belo Jardim  
Jobson Tenório do Nascimento - IFPE - Campus Pesqueira  
Wellyson Fernando Nunes Souza - IFPE - Campus Garanhuns

Dados Internacionais de Catalogação - CIP

M357d Marques, Gabriel Vasconcelos  
DataHarbor: capacitando o engajamento cidadão por meio da visualização de dados abertos na governança municipal / Gabriel Vasconcelos Marques. – Belo Jardim-PE, 2024.  
72f.: il.

Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Software) – Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, Campus Belo Jardim- PE, 2024.

Orientador: Prof.º José Fernando da Silva.

Inclui referências.

1. Desenvolvimento de aplicações. 2. Análise de dados. 3. Dados públicos - análise. 4. Banco de dados. I. Título. II. Silva, José Fernando da. III. Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco.

CDD 005

GABRIEL VASCONCELOS MARQUES

**DATAHARBOR: CAPACITANDO O ENGAJAMENTO CIDADÃO POR MEIO  
DA VISUALIZAÇÃO DE DADOS ABERTOS NA GOVERNANÇA  
MUNICIPAL**

Trabalho aprovado. Belo Jardim, 25/04/2024.

José Fernando da Silva

---

Professor Orientador

Jobson Tenório do Nascimento

---

Convidado 1

Wellyson Fernando Nunes Souza

---

Convidado 2

Belo Jardim, Pernambuco  
2024

## **AGRADECIMENTOS**

Gostaria de expressar meus sinceros agradecimentos a Deus, por ser a fonte de toda sabedoria, força e orientação ao longo desta jornada acadêmica.

À minha família, especialmente aos meus tios Cícero e Petrônio, à minha mãe Valéria, e a minha namorada Rayssa que sempre estiveram ao meu lado, oferecendo seu apoio incondicional e amor que tornaram possível a realização deste sonho.

Agradeço também a todos os meus amigos e familiares, por compartilharem comigo cada desafio, alegria e conquista ao longo deste caminho.

A todos vocês, meu profundo obrigado por fazerem parte desta conquista e por tornarem possível a realização deste artigo. Suas contribuições e apoio foram fundamentais para meu crescimento pessoal e acadêmico.

Gostaria de agradecer também à instituição de ensino IFPE Campus Belo Jardim, ao meu orientador Prof. José Fernando da Silva, pela disponibilidade, apoio e orientação ao longo deste trabalho, e ao meu coordenador João Almeida pelo suporte oferecido, pelas oportunidades disponibilizadas e por acreditar nos meus projetos, bem como a todos os profissionais que contribuíram indiretamente para a realização deste trabalho.

Por fim deixo aqui o versículo bíblico "Esta é a vitória que vence o mundo: a nossa fé" (João 5:4), que representa a importância da fé e da perseverança em superar os obstáculos e alcançar os objetivos.

## RESUMO

A crescente disponibilidade de dados públicos proporciona oportunidades importantes para a compreensão e a tomada de decisões informadas; no entanto, a complexidade e a heterogeneidade destes dados podem criar barreiras à sua interpretação e utilização eficaz. A falta de padronização e a dispersão dos dados podem dificultar a compreensão dos padrões e tendências subjacentes, tornando fundamental o desenvolvimento e utilização de plataformas de visualização que simplifiquem a análise e aumentem a transparência. Ao destacar esses desafios e enfatizar a importância da transparência e visualização dos dados públicos, percebemos a necessidade de soluções e ferramentas que facilitem o acesso e a compreensão dessas informações a públicos de diversos graus de instrução, promovendo maior envolvimento dos cidadãos e tomadas de decisão mais responsáveis. Este trabalho visa apresentar uma solução prática para os desafios mencionados, através do desenvolvimento de uma aplicação web especializada em visualização de dados públicos disponibilizados no portal da prefeitura da cidade de Belo Jardim, no agreste pernambucano. A plataforma tem como objetivo fornecer *dashboards* detalhados e interativos que oferecem *insights* valiosos sobre os dados públicos, facilitando o entendimento e a interpretação das informações para a população em geral. Além disso, a solução busca promover a transparência e a prestação de contas, permitindo que os usuários acessem dados relevantes de maneira fácil e intuitiva. Com interfaces amigáveis, a plataforma disponibiliza informações transparentes e úteis para o acompanhamento de políticas públicas, monitoramento de serviços governamentais e participação ativa na tomada de decisões. Espera-se que este trabalho contribua para uma maior democratização do acesso à informação e para o fortalecimento da política de transparência, ao mesmo tempo em que estimule o desenvolvimento de mais soluções aplicadas a diferentes órgãos públicos, resolvendo os desafios inerentes à análise e compreensão de dados públicos abertos.

Palavras-chave: análise de dados, dados públicos governamentais, sistema web

## **ABSTRACT**

The increasing availability of public data provides significant opportunities for understanding and making informed decisions; however, the complexity and heterogeneity of these data can create barriers to their interpretation and effective use. The lack of standardization and data dispersion can hinder the understanding of underlying patterns and trends, making it crucial to develop and utilize visualization platforms that simplify analysis and enhance transparency. By highlighting these challenges and emphasizing the importance of transparency and visualization of public data, we recognize the need for solutions and tools that facilitate access and understanding of this information for audiences with varying levels of education, promoting greater citizen engagement and responsible decision-making. This work aims to present a practical solution to the mentioned challenges through the development of a specialized web application for visualizing public data available on the municipality's portal of the city of Belo Jardim, in the Agreste region of Pernambuco. The platform aims to provide detailed and interactive dashboards that offer valuable insights into public data, facilitating understanding and interpretation of information for the general population. Furthermore, the solution seeks to promote transparency and accountability by allowing users to access relevant data in an easy and intuitive manner. With user-friendly interfaces, the platform provides transparent and useful information for monitoring public policies, government services, and active participation in decision-making processes. It is expected that this work will contribute to greater democratization of access to information and strengthen transparency policies, while also encouraging the development of more solutions tailored to different public agencies, addressing inherent challenges in the analysis and understanding of open public data.

**Keywords:** data analysis, government public data, web system

## LISTA DE FIGURAS

<b>Figura 1: Api's disponibilizadas</b>	<b>30</b>
<b>Figura 2: Fluxograma de interação - Página Inicial</b>	<b>35</b>
<b>Figura 3: Página inicial DataHarbor</b>	<b>36</b>
<b>Figura 4: Fluxograma de interação - Registrar Usuário</b>	<b>37</b>
<b>Figura 5: Tela de registrar usuário</b>	<b>38</b>
<b>Figura 6: Fluxograma de interação - Realizar Login</b>	<b>39</b>
<b>Figura 7: Tela de login no sistema</b>	<b>39</b>
<b>Figura 8: Tela de home</b>	<b>40</b>
<b>Figura 9: Fluxograma de interação - Dashboard Despesas</b>	<b>41</b>
<b>Figura 10: Dashboard de despesas públicas carregando dados</b>	<b>42</b>
<b>Figura 11: Dashboard de despesas públicas carregados (parte 1)</b>	<b>43</b>
<b>Figura 12: Dashboard de despesas públicas carregados (parte 2)</b>	<b>43</b>
<b>Figura 13: Dashboard de despesas públicas carregados (parte 3)</b>	<b>44</b>
<b>Figura 14: Fluxograma de interação - Dashboard Licitação</b>	<b>45</b>
<b>Figura 15: Dashboard de licitação carregando dados</b>	<b>45</b>
<b>Figura 16: Dashboard de licitação carregado (parte 1)</b>	<b>46</b>
<b>Figura 17: Dashboard de licitação carregado (parte 2)</b>	<b>47</b>
<b>Figura 18: Filtro com ano preenchido</b>	<b>48</b>
<b>Figura 19: Filtro com ano e nome da função preenchidos</b>	<b>48</b>
<b>Figura 20: Card Valor Pago Total preenchido</b>	<b>49</b>
<b>Figura 21: Card Valor Liquidado Total preenchido</b>	<b>49</b>
<b>Figura 22: Card Valor Empenhado Total preenchido</b>	<b>50</b>
<b>Figura 23: Card QTD de Despesas Registradas preenchido</b>	<b>50</b>
<b>Figura 24: Gráfico obtido a partir da análise de valores por subfunção</b>	<b>51</b>
<b>Figura 25: Gráfico obtido a partir da análise de valor pago por subfunção (por programa)</b>	<b>52</b>
<b>Figura 26: Gráfico obtido a partir da análise de valor pago por elemento (Top 10)</b>	<b>53</b>
<b>Figura 27: Gráfico obtido a partir da análise de valor pago por categoria</b>	<b>54</b>
<b>Figura 28: Gráfico obtido a partir da análise de valor pago por unidade</b>	<b>55</b>
<b>Figura 29: Gráfico obtido a partir da análise de somas de despesas por data de empenho</b>	<b>56</b>
<b>Figura 30: Gráfico obtido a partir da análise de contagem de despesas registradas por data de empenho</b>	<b>57</b>
<b>Figura 31: Filtro com ano preenchido</b>	<b>58</b>
<b>Figura 32: Card Valor Total Adjudicado preenchido</b>	<b>58</b>
<b>Figura 33: Card QTD Geral de Licitações preenchido</b>	<b>59</b>
<b>Figura 34: Card QTD Geral de Licitações Concluídas preenchido</b>	<b>59</b>
<b>Figura 35: Gráfico obtido a partir da análise de soma do valor total adjudicado por natureza</b>	<b>60</b>

<b>Figura 36: Gráfico obtido a partir da análise de soma do valor adjudicado por modalidade</b>	<b>61</b>
<b>Figura 37: Gráfico obtido a partir da análise de soma do valor total adjudicado por descrição objeto (Top 10)</b>	<b>62</b>
<b>Figura 38: Gráfico obtido a partir da análise de somas do total adjudicado de acordo com o tempo</b>	<b>63</b>
<b>Figura 39: Informação disponibilizada acerca dos campos retornados como resposta da api de despesas públicas</b>	<b>64</b>
<b>Figura 40: Informação retornada como resposta da api de despesas públicas</b>	<b>64</b>

## **LISTA DE QUADROS**

**Quadro 1 – Quadro Metodológico**

**25**

## LISTA DE ABREVIATURAS E SIGLAS

<b>API's</b>	Application Programming Interfaces, interfaces de programação que permitem a comunicação entre diferentes sistemas.
<b>CSS</b>	Cascading Style Sheets, utilizada para estilizar o conteúdo HTML.
<b>ETL</b>	Extração, Transformação e Carga (Extraction, Transformation, Loading).
<b>HTML</b>	Linguagem de marcação utilizada para criar páginas web.
<b>HTTP</b>	Hypertext Transfer Protocol, protocolo de comunicação utilizado na web.
<b>JSON</b>	JavaScript Object Notation, formato de dados utilizado para troca de informações.
<b>LAI</b>	Lei de Acesso à Informação, legislação que regula o acesso a informações públicas no Brasil.
<b>QTD</b>	Quantidade.
<b>UI</b>	User Interface (Interface de Usuário), a parte visual com a qual os usuários interagem em um sistema.

## TERMOS E DEFINIÇÕES

<b>Backend</b>	Parte de um sistema que lida com a lógica e processamento de dados.
<b>Bootstrap</b>	Framework de front-end para desenvolvimento ágil e responsivo de sites e aplicativos web.
<b>Dash</b>	Framework para construção de aplicativos web interativos em Python.
<b>DataFrame</b>	Estrutura de dados tabular do Pandas.
<b>datetime</b>	Tipo de dado em Python para representar datas e horários.
<b>fillna</b>	Método do Pandas para preencher valores ausentes em um DataFrame.
<b>float</b>	Tipo de dado numérico em Python para representar números de ponto flutuante.
<b>Flask</b>	Microframework em Python para desenvolvimento de aplicações web.
<b>Frontend</b>	Parte de um sistema ou aplicação que interage diretamente com o usuário, geralmente referindo-se à interface gráfica.
<b>Insights</b>	Percepções ou descobertas obtidas a partir da análise de dados.
<b>JavaScript</b>	Linguagem de programação utilizada principalmente para desenvolvimento web.
<b>login</b>	Processo de autenticação de usuários em sistemas.
<b>merge</b>	Método do Pandas para combinar DataFrames.
<b>MySQL</b>	Sistema de gerenciamento de banco de dados relacional.
<b>Pandas</b>	Biblioteca de análise de dados em Python.
<b>Plotly.js</b>	Biblioteca JavaScript para criação de visualizações interativas.
<b>Python</b>	Linguagem de programação de alto nível amplamente utilizada em desenvolvimento de software e análise de dados.
<b>React.js</b>	Biblioteca JavaScript de código aberto para criação de interfaces de usuário.
<b>Requests</b>	Biblioteca em Python para fazer requisições HTTP.
<b>Scrum</b>	Framework ágil para gerenciamento de projetos, baseado em ciclos iterativos de trabalho.
<b>Sprint</b>	Período de tempo fixo (geralmente de duas semanas) em que uma equipe desenvolve e entrega um incremento do produto.
<b>strings</b>	Sequências de caracteres em Python.
<b>sum</b>	Método do Pandas para somar valores em um DataFrame.
<b>to_datetime</b>	Função do Pandas para converter dados para o tipo datetime.

**to\_numeric** Função do Pandas para converter dados para o tipo numérico.

**value\_counts** Método do Pandas para contar valores únicos em uma coluna.

**Web** Abreviação para World Wide Web, o sistema de informação interconectado que permite o acesso a recursos na internet.

## SUMÁRIO

<b>1. Introdução</b>	<b>15</b>
<b>1.1 Contextualização</b>	<b>15</b>
<b>1.2 Problemática do trabalho</b>	<b>16</b>
<b>1.3 Objetivos</b>	<b>16</b>
<b>1.3.1 Objetivo geral</b>	<b>16</b>
<b>1.3.2 Objetivos específicos</b>	<b>16</b>
<b>1.4 Estrutura</b>	<b>17</b>
<b>2. Referencial Teórico</b>	<b>19</b>
<b>2.1 Importância da consolidação e integração de dados dispersos para uma visão holística do negócio</b>	<b>19</b>
<b>2.2 Oportunidades e benefícios</b>	<b>21</b>
<b>2.3. Desenvolvimento de Soluções Baseadas em Dados Abertos Governamentais</b>	<b>23</b>
<b>3. Metodologia</b>	<b>25</b>
<b>3.1 Definições da abordagem</b>	<b>25</b>
<b>4. O sistema Dataharbor</b>	<b>27</b>
<b>4.1. Resumo do projeto</b>	<b>27</b>
<b>4.2 Objetivos da ferramenta</b>	<b>27</b>
<b>4.3 Origem dos dados</b>	<b>28</b>
<b>4.4 Processo de ETL</b>	<b>28</b>
<b>4.4.1 Processo de coleta dos dados</b>	<b>28</b>
<b>4.4.2 Limpeza e preparação de dados</b>	<b>31</b>
<b>4.5 Desenvolvimento da aplicação</b>	<b>32</b>
<b>4.5.1 Tecnologias Utilizadas</b>	<b>32</b>
<b>4.5.1.1 Frontend</b>	<b>32</b>
<b>4.5.1.2 Backend</b>	<b>32</b>
<b>4.5.1.3 Gerenciamento de banco de dados</b>	<b>33</b>
<b>4.5.2 Metodologia de desenvolvimento</b>	<b>33</b>
<b>4.6 Casos de uso</b>	<b>34</b>
<b>4.6.1 Página inicial</b>	<b>34</b>
<b>4.6.2 Registrar usuário</b>	<b>36</b>
<b>4.6.3 Login no sistema</b>	<b>38</b>
<b>4.6.4 Home</b>	<b>40</b>
<b>4.6.5 Dashboard de despesas públicas</b>	<b>41</b>
<b>4.6.6 Dashboard de licitação</b>	<b>44</b>
<b>4.7 Análise e objetivos dos componentes da dashboard: Detalhamento das visualizações de dados</b>	<b>47</b>
<b>4.7.1 Componentes da dashboard de despesas públicas</b>	<b>47</b>
<b>4.7.1.1 Filtros</b>	<b>47</b>

<b>4.7.1.1.1 Filtro de ano</b>	<b>47</b>
<b>4.7.1.1.2 Filtro de nome de função</b>	<b>48</b>
<b>4.7.1.2 Cards</b>	<b>48</b>
<b>4.7.1.2.1 Card Valor Pago Total</b>	<b>48</b>
<b>4.7.1.2.2 Card Valor Liquidado Total</b>	<b>49</b>
<b>4.7.1.2.3 Card Valor Empenhado Total</b>	<b>49</b>
<b>4.7.1.2.4 Card QTD de Despesas Registradas</b>	<b>50</b>
<b>4.7.1.3 Gráficos</b>	<b>50</b>
<b>4.7.1.3.1 Análise de Valores por Subfunção</b>	<b>50</b>
<b>4.7.1.3.2 Valor Pago por Subfunção (por Programa)</b>	<b>51</b>
<b>4.7.1.3.3 Valor Pago por Elemento (Top 10)</b>	<b>52</b>
<b>4.7.1.3.4 Valor Pago por Categoria</b>	<b>53</b>
<b>4.7.1.3.5 Valor Pago por Unidade</b>	<b>54</b>
<b>4.7.1.3.6 Somas de Despesas Por Data de Empenho</b>	<b>55</b>
<b>4.7.1.3.7 Contagem de Despesas Registradas por Data de Empenho</b>	<b>56</b>
<b>4.7.2 Componentes da dashboard de licitação</b>	<b>57</b>
<b>4.7.2.1 Filtros</b>	<b>57</b>
<b>4.7.2.1.1 Filtro de ano</b>	<b>57</b>
<b>4.7.2.2 Cards</b>	<b>58</b>
<b>4.7.2.2.1 Card Valor Total Adjudicado</b>	<b>58</b>
<b>4.7.2.2.2 Card QTD Geral de Licitações</b>	<b>58</b>
<b>4.7.2.2.3 Card QTD Geral de Licitações Concluídas</b>	<b>59</b>
<b>4.7.2.3 Gráficos</b>	<b>59</b>
<b>4.7.2.3.1 Soma do Valor Total Adjudicado por Natureza</b>	<b>59</b>
<b>4.7.2.3.2 Soma do Valor Adjudicado por Modalidade</b>	<b>60</b>
<b>4.7.2.3.3 Soma do Valor Total Adjudicado por Descrição Objeto (Top 10)</b>	<b>61</b>
<b>4.7.2.3.4 Somas do Total Adjudicado de acordo com o tempo</b>	<b>62</b>
<b>5. Conclusões e Trabalhos Futuros</b>	<b>64</b>
<b>5.1 Desafios Enfrentados e Restrições da Aplicação</b>	<b>64</b>
<b>5.2 Conclusões</b>	<b>66</b>
<b>5.3 Trabalhos Futuros</b>	<b>66</b>
<b>6. Referências</b>	<b>68</b>

## 1. Introdução

Neste capítulo serão abordados a contextualização do trabalho, a problemática e sua estrutura, expondo as motivações para o estudo e definindo os objetivos a serem alcançados.

### 1.1 Contextualização

As grandes quantidades de dados gerados em vários campos da sociedade moderna desempenham um papel fundamental na compreensão e no avanço do conhecimento humano, revolucionando o mundo. Deste modo a análise de dados tornou-se uma ferramenta indispensável para extrair *insights* significativos de grandes quantidades de informação, permitindo uma compreensão mais profunda dos aspectos sociais, econômicos, ambientais e culturais que moldam a nossa sociedade (CERVANTES, 2016).

Um aspecto importante da análise de dados é a sua importância na compreensão da população. Vivemos numa era cada vez mais digital, onde quase todas as interações e transações deixam uma pegada digital (GRUS, 2016), tornando os dados um recurso valioso para compreender os comportamentos, preferências e necessidades das pessoas. Ao analisar cuidadosamente esses dados, você pode descobrir padrões ocultos, identificar tendências emergentes e prever o que o futuro reserva.

A disponibilidade de dados abertos por órgãos governamentais têm desempenhado um papel fundamental na transformação do cenário da informação, abrindo novas oportunidades e possibilitando uma visão mais abrangente e transparente das atividades do governo (OECD, 2013).

Dados governamentais abertos são entendidos como conjuntos de dados fornecidos por órgãos públicos de forma gratuita e acessível, representando uma valiosa fonte de informação que pode ser utilizada para diversos fins, fornecendo vários benefícios (DINIZ, 2010). Para os cidadãos, estes dados proporcionam oportunidades para o exercício da cidadania ativa, permitindo o acompanhamento das ações governamentais, a participação em processos democráticos e a participação em iniciativas de transparência e responsabilização (Thorsby et al., 2017). Para as empresas, os dados governamentais abertos podem fornecer informações valiosas sobre os mercados, as tendências econômicas, as oportunidades de negócio e as necessidades dos consumidores, promovendo assim o desenvolvimento de produtos e serviços inovadores e melhorando a competitividade do mercado. Para as organizações da sociedade civil, estes dados são ferramentas poderosas para a defesa de

direitos, o desenvolvimento de políticas públicas e a monitorização de questões sociais, ambientais e económicas.

## **1.2 Problemática do trabalho**

Apesar do potencial dos dados governamentais abertos, a sua utilização eficaz enfrenta frequentemente desafios significativos, tais como dispersão e fragmentação da informação, falta de normalização e interoperabilidade, e dificuldades na análise e interpretação dos dados, onde segundo Bico et al. (2012), é necessário se possuir prévio conhecimento sobre formatos ou tecnologias para se extrair informações desses dados.

Este trabalho visa abordar a integração de dados como um elemento-chave para uma compreensão abrangente das atividades governamentais e do impacto das políticas públicas. Ao unificar fontes de dados díspares e criar interfaces intuitivas e interativas, pode ser criada uma visão unificada e abrangente das operações governamentais, facilitando a identificação de padrões, tendências e correlações relevantes.

Neste contexto, iremos propor uma solução inovadora sob a forma de uma aplicação *web* desenhada com o propósito de resolver os desafios acima mencionados e explorar todo o potencial desta informação. A ferramenta foi projetada para fornecer uma interface intuitiva e acessível para explorar, visualizar e analisar dados governamentais de maneira eficaz e eficiente, ajudando assim a aumentar a transparência, o envolvimento e o envolvimento dos cidadãos. Neste trabalho, detalhamos o desenvolvimento e operação da solução, destacando o seu potencial para promover uma governança mais transparente, responsável e baseada em dados.

É importante ressaltar que, ao longo deste trabalho, utilizaremos o nome fictício DataHarbor para ilustrar o conceito de uma aplicação de visualização de dados. Este nome é apenas um exemplo hipotético e não está associado a nenhuma aplicação real existente.

## **1.3 Objetivos**

### **1.3.1 Objetivo geral**

O objetivo principal deste trabalho é desenvolver uma solução *web* para visualização de dados públicos abertos, visando contribuir para a transparência, democratização da informação e tomada de decisão dos cidadãos e demais interessados.

### **1.3.2 Objetivos específicos**

1. Avaliar a disponibilidade, atualização e qualidade dos dados públicos relevantes para o estudo de caso, identificando possíveis desafios e limitações.
2. Utilizar técnicas de análise de dados, tais como visualização dos dados, análise estatística e outras metodologias relevantes para explorar e extrair *insights* dos dados públicos abertos.
3. Desenvolver uma solução *web* para visualização dos dados:  
Desenvolver uma aplicação *web* com *dashboards* que permitam a visualização intuitiva e interativa dos dados públicos abertos, utilizando ferramentas e tecnologias adequadas para a análise e apresentação dos dados.
4. Demonstrar como a disponibilização e visualização dos dados públicos abertos através da solução *web* contribuem para a transparência governamental e o acesso democrático à informação.

#### **1.4 Estrutura**

Este trabalho está dividido em 5 capítulos, abordando os conteúdos a seguir:

- Capítulo 1: Introdução: Abordagem inicial do trabalho, problemática, abordagem dos objetivos gerais e específicos e a estruturação.
- Capítulo 2: Referencial teórico: Iniciando com a abordagem da importância da consolidação e integração de dados dispersos para uma visão holística do negócio, logo após contextualizando a cerca das oportunidades e benefícios oferecidos pelos dados abertos governamentais para cidadãos, empresas e organizações da sociedade civil e por fim a importância da consolidação e integração de dados dispersos para uma visão holística do negócio.
- Capítulo 3: Metodologia: Serão abordados as definições da abordagem de estudo, o processo de origem, coleta, limpeza e preparação dos dados e tópicos referentes ao desenvolvimento da solução, como tecnologias e metodologias utilizadas
- Capítulo 4: O sistema Dataharbor: Será abordado um breve resumo sobre a aplicação desenvolvida, destacando seus objetivos, apresentação dos casos de uso da ferramenta

e uma análise completa dos componentes de cada *dashboard*.

- Capítulo 5: Conclusões, desafios enfrentados e trabalhos futuros: Apresenta as conclusões do trabalho, os desafios enfrentados juntamente com as restrições da aplicação e o que se espera que sejam realizados nas etapas futuras.

## **2. Referencial Teórico**

### **2.1 Importância da consolidação e integração de dados dispersos para uma visão holística do negócio**

A análise de dados tornou-se uma ferramenta essencial não só para as organizações públicas e privadas modernas, como também demais entidades, permitindo-lhes extrair *insights* valiosos e tomar decisões informadas, baseada na exposição de resultados (Ignácio, 2012). No entanto, muitas dessas organizações enfrentam o desafio de lidar com dados dispersos e fragmentados, armazenados em diferentes sistemas e formatos. A consolidação e integração desses dados dispersos são fundamentais para uma visão holística do negócio, pois permitem que as mesmas obtenham uma compreensão abrangente de suas operações, clientes e mercado. Através da análise integrada de dados, podem se identificar padrões, tendências, oportunidades ou melhorias que de outra forma poderiam passar despercebidas, onde de acordo com (Costa & Coelho, 2006), a partir da análise de dados levantados no nível estratégico, é montada a especificação de um novo processo ou da melhoria/ inovação de um processo existente.

A integração de dados envolve agregar informações de múltiplas fontes em um repositório centralizado. Isso pode incluir dados de transações, dados de clientes, dados de vendas, dados de marketing, etc. A integração de dados de fontes e formatos díspares apresenta dificuldades (BATISTA, 2020 pg 170) e, com a proliferação de dados em sistemas e plataformas díspares, muitas organizações enfrentam o desafio de unificar e integrar estes dados de forma coesa. Isto pode ser particularmente problemático quando os dados estão fragmentados em sistemas legados ou armazenados em formatos não suportados, dificultando às organizações a obtenção de uma visão abrangente e integrada das suas operações.

A falta de padronização e consistência nos dados pode levar a problemas de qualidade dos dados (BATISTA, 2020 pg 173), como duplicações, erros e omissões, comprometendo a integridade e a confiabilidade das informações. Além disso, a falta de uma estrutura clara dificulta a identificação de relações e padrões nos dados, tornando difícil para as organizações obter uma visão abrangente e integrada de suas operações.

Quando uma base de dados não possui estrutura suficiente e não há uma forma clara de visualização, isso pode gerar uma série de problemas e desafios significativos para a organização. Um dos principais problemas é a dificuldade em extrair informações úteis e relevantes dos dados, pois podem ser confusos, incompletos ou inconsistentes. Isto pode

dificultar a análise e a tomada de decisões baseadas em dados, impactando diretamente o desempenho e a competitividade de uma organização.

Além de facilitar a análise desses dados, a consolidação e integração de dados díspares oferece uma série de benefícios às organizações. Primeiro, ao criar um repositório de dados centralizado, as empresas podem reduzir a redundância e inconsistências de dados e garantir a precisão e integridade das informações (Rozza, 2023). Isto é especialmente importante para organizações onde vários departamentos ou unidades de negócios coletam e mantêm dados separadamente. Além disso, a consolidação de dados pode melhorar a eficiência operacional, reduzindo o tempo e os recursos necessários para acessar e analisar informações.

Outro benefício significativo da integração de dados é a capacidade de desenvolver uma compreensão mais completa e precisa dos clientes e mercados. Ao integrar dados de diferentes fontes, através de um método de clusterização, as empresas podem criar perfis de clientes mais detalhados e compreender melhor os seus comportamentos, preferências e necessidades (C. GUILHERME, 2023). Isso permite uma segmentação de mercado mais precisa e a personalização de produtos e serviços para atender às necessidades específicas dos clientes. A falta de uma abordagem clara para a visualização de dados pode dificultar a comunicação e a colaboração entre os membros da equipe e as partes interessadas. Segundo Nascimento & Ferreira (2011) “a meta da visualização é ajudar também na descoberta de novas informações”, onde sem uma representação visual dos dados adequada, pode ser difícil para os utilizadores compreendê-las e interpretá-las, o que limita a capacidade de uma organização partilhar conhecimento e tomar decisões informadas.

Além dos benefícios operacionais e estratégicos, a consolidação e integração de dados díspares têm implicações mais amplas para a capacidade de uma organização se adaptar e inovar num ambiente de negócios em constante mudança. Ao obter uma visão abrangente do negócio, as empresas podem identificar melhor as oportunidades de mercado, antecipar tendências emergentes e responder rapidamente às mudanças nas condições do mercado. Isto pode proporcionar uma vantagem competitiva significativa, permitindo que as organizações diferenciam-se num mercado saturado e crescente. No puro conceito defendido do empreendedorismo, conforme Bernardi (2012, p.67):

A ideia de um empreendimento surge da observação, da percepção e análise de atividades, tendências e desenvolvimentos, na cultura, na sociedade, nos hábitos sociais e de consumo. As oportunidades detectadas ou visualizadas, racional ou intuitivamente, das necessidades e das demandas prováveis, atuais e futuras, e necessidades não atendidas definem a ideia do empreendimento.

À medida que as organizações lidam com volumes cada vez maiores de dados, a padronização melhora a qualidade e a precisão das visualizações, tornando-as mais intuitivas e acessíveis para o público, gerando maior facilidade na aplicação de técnicas de visualização através de *dashboard*, permitindo uma rápida identificação de tendências, relações ou comparações entre as informações, trazendo clareza acerca das informações.

## **2.2 Oportunidades e benefícios**

Os dados governamentais abertos representam uma fonte rica e vasta de informações que pode ser acessada e utilizada por cidadãos, empresas e organizações da sociedade civil para diversos fins. Os dados abrangem uma vasta gama de informações, desde estatísticas demográficas e económicas até informações sobre saúde, transportes e ambiente. Segundo a (OECD, 2013) uma das principais oportunidades oferecidas pelos dados governamentais abertos é a capacidade de promover a transparência e a responsabilização governamental, dando aos cidadãos acesso a informações sobre as atividades e decisões governamentais. Por exemplo, portais de dados abertos como dados.gov.br no Brasil fornecem uma riqueza de dados relacionados com políticas públicas, gastos governamentais e outros aspectos da administração pública, permitindo que os cidadãos acompanhem e fiscalizem as ações do governo.

A disponibilidade de dados governamentais abertos é muitas vezes regida por leis e políticas específicas concebidas para melhorar a transparência e o acesso à informação. Em muitos países, as leis de acesso à informação garantem o direito dos cidadãos de solicitar e receber informações na posse do governo, o que muitas vezes se estende aos dados governamentais. Além disso, algumas jurisdições têm leis específicas que exigem que os governos divulguem e acessem publicamente certos tipos de dados.

Um exemplo bem conhecido é a Lei de Acesso à Informação (LAI) do Brasil, que entrou em vigor em 2012 e oferece aos cidadãos o direito de acessar informações em poder do governo. A LAI também impõe às autoridades públicas a obrigação de fornecer informações de interesse coletivo ou geral por meio de portais de dados abertos (Lei de Acesso à Informação - Lei nº 12.527 Art. 8º). Esta legislação é um marco importante na promoção da transparência e responsabilização no país, incentivando os governos a disponibilizar uma vasta gama de dados de forma aberta e acessível, onde segundo Andrade e Raupp (2017), a transparência por parte do governo é um fator chave na questão democrática,

pois a construção de um governo envolve a sociedade num papel fiscalizador das ações dos seus líderes.

Como praticamente todo dado governamental é público, é fundamental que os governos implementem políticas para disponibilizá-los” (ENAP, 2017), ao defender a ideia de “Todo dado público tem vocação para ser dado aberto”, quando se trata de dados municipais, as prefeituras desempenham um papel vital no fornecimento de dados abertos relevantes aos cidadãos e às empresas locais. Muitos governos municipais adotaram políticas de dados abertos e lançaram portais de dados abertos para fornecer informações sobre uma variedade de tópicos, incluindo transporte, educação, saúde, segurança pública e muito mais.

Por exemplo, cidades como São Paulo, Rio De Janeiro possuem portais robustos de dados abertos que fornecem acesso a uma vasta gama de conjuntos de dados, desde informações sobre transportes públicos e infraestruturas urbanas até dados sobre serviços de saúde e educação. Os criadores de aplicações, investigadores, jornalistas e cidadãos interessados podem utilizar estes dados para diversos fins, desde o desenvolvimento de aplicações móveis úteis até à realização de análises e pesquisas sobre questões urbanas.

No entanto, embora as prefeituras tenham feito progressos na disponibilização de dados abertos, ainda existem desafios que precisam de ser superados para garantir que os dados estejam disponíveis a todos os membros da comunidade de uma forma completa, atualizada e acessível (ENAP, 2017, pg. 29). É importante que as mesmas trabalhem em estreita colaboração com os cidadãos e as partes interessadas locais para identificar as necessidades de informação da comunidade e garantir que os dados são fornecidos de uma forma que satisfaça eficazmente essas necessidades. Além disso, deve ser assegurado que os dados são fornecidos de uma forma que proteja a privacidade e a segurança das informações pessoais dos cidadãos, onde de acordo com Vaz (2007), segurança e privacidade são valores que devem estar associados à utilização de qualquer sistema de informação, aderindo às melhores práticas de anonimato e proteção de dados. Ao fazê-lo, os municípios podem aproveitar todo o potencial dos dados abertos para promover a transparência, a inovação e o envolvimento dos cidadãos a nível local.

Além da transparência e da responsabilização, os dados governamentais abertos oferecem oportunidades importantes para promover a inovação e o empreendedorismo (MANUAL, 2011). Ao fornecer dados sobre todos os aspectos da sociedade, os governos criam um ambiente propício para que empresas e organizações da sociedade civil desenvolvam soluções e serviços inovadores.

Esses dados têm o potencial de promover o envolvimento e a participação dos

cidadãos na tomada de decisões públicas (OECD, 2013) e também o desenvolvimento económico e social de um país, estimulando a inovação, o empreendedorismo e a criação de emprego. Ao fornecer informações e recursos valiosos, os dados abertos criam um ambiente favorável ao crescimento económico e ao desenvolvimento de novas indústrias e setores, onde podem ser utilizados para identificar oportunidades de negócios, desenvolver novos produtos e serviços e otimizar operações (Magalhaes et al., 2013).

Ao fornecer informações sobre políticas, programas e serviços governamentais, os dados abertos podem permitir que os cidadãos participem mais ativamente no processo democrático, proporcionando-lhes os recursos de que necessitam para compreender e participar no debate público (OECD, 2013). Com a utilização de dados abertos, podem ser realizadas consultas públicas sobre questões de interesse comunitário, permitindo que os cidadãos expressem as suas opiniões e contribuam para a formulação de políticas públicas.

É importante reconhecer que o sucesso de uma iniciativa de dados abertos depende não apenas da disponibilidade dos dados, mas também do acesso equitativo aos dados, da capacidade de utilizar os dados de forma eficaz e da garantia de que a privacidade e a segurança da informação sejam respeitadas, mantidas e protegidas, para que as empresas e as organizações da sociedade civil consigam trabalhar em conjunto para aproveitar todo o potencial dos dados governamentais abertos e garantir que estes benefícios sejam partilhados de forma ampla e equitativa

### **2.3. Desenvolvimento de Soluções Baseadas em Dados Abertos Governamentais**

O desenvolvimento de soluções baseadas em dados governamentais abertos representa uma área promissora para a inovação cidadã e o avanço tecnológico. A utilização destes dados em aplicações *web* oferece oportunidades únicas para aumentar a transparência governamental, promover o envolvimento dos cidadãos e criar serviços públicos mais eficientes. Como cita a teoria de Isotani (2015, p.42):

Dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de citar a fonte original e compartilhar com as mesmas licenças em que as informações foram inicialmente apresentadas.

Tendo em vista tal abrangência, observa-se uma sólida abertura para o desenvolvimento de soluções inovadoras que atendam essa necessidade social.

Existe como principal benefício do desenvolvimento de aplicações *web* baseadas em

dados abertos governamentais a capacidade de aumentar a transparência e a responsabilização dos órgãos públicos, permitindo um maior envolvimento do cidadão em análises e avaliações (Gurstein, 2010). Ao adotar a política de dados abertos e disponibilizar essas informações, sejam elas de qualquer natureza, tais como finanças públicas, despesas governamentais, licitações e muito mais, o governo permite tanto que cidadãos, organizações e empresas possam analisar e questionar suas ações e decisões. O portal da transparência fornece ao público um acesso direto a informações referentes a despesas, receitas, licitações, contratos, serviços, contribuindo para um controle social mais eficaz (Portal da Transparência, 2024). Estas soluções não só promovem um governo mais aberto e responsável, mas também incentivam uma maior participação dos cidadãos na tomada de decisões políticas, criando um ambiente mais participativo, fornecendo oportunidades para inovação e desenvolvimento econômico, através da criação de novos produtos, serviços ou novas estratégias de marketing (Gurstein, 2010).

Através dessas oportunidades, podem ser criadas aplicações específicas que atendam diversos setores da sociedade, em tempo real, melhorando a qualidade de vida dos cidadãos e impulsionando o crescimento de novas empresas e empregos, contribuindo para a economia digital e a promoção da cidadania e serviços públicos (O'REILLY, 2011). A longo prazo, estas iniciativas podem levar a mudanças profundas na relação entre o governo e os cidadãos, aproximando mais ambos, estreitando as relações e tornando o cidadão parte das decisões e ações governamentais. Num mundo cada vez mais conectado e digital, a capacidade de acessar, analisar e utilizar dados abertos torna-se uma ferramenta poderosa para resolver problemas sociais complexos, desde a gestão de recursos naturais até respostas a crises de saúde pública.

Ao defender a ideia de acesso à informação “A finalidade principal da transparência só pode ser alcançada mediante a conversão da informação em instrumento de ação do cidadão, o que nem sempre é algo fácil de alcançar”, Gruman, M. (2012, p.104) nos apresenta que o governo e a sociedade, ao promoverem uma cultura de dados abertos, podem trabalhar juntos de forma mais eficaz no desenvolvimento de soluções que satisfaçam as necessidades atuais e futuras, proporcionando oportunidades sem precedentes para melhorar a governação, estimular a economia e promover o bem-estar social. Ao aproveitar o potencial destes dados, podemos não só resolver desafios imediatos, mas também preparar o caminho para uma sociedade mais informada, empenhada e resiliente.

### 3. Metodologia

#### 3.1 Definições da abordagem

Este trabalho adotará uma abordagem exploratória e descritiva para compreender o impacto da utilização de dados abertos da prefeitura de Belo Jardim na construção de *dashboards* interativos e na tomada de decisões dos usuários da aplicação *web*. A abordagem exploratória baseia-se numa investigação preliminar e extensa do tema, visando compreender as nuances e complexidades dos dados disponíveis e a sua relevância para o utilizador final.

Durante a fase de exploração, foi realizada uma análise preliminar do conjunto de dados fornecidos pela API da prefeitura de Belo Jardim para identificar padrões, tendências e possíveis insights que possam orientar a criação do *dashboard*. Isto incluiu a exploração de diferentes tipos de dados, tais como dados numéricos, temporais e textuais, para cobrir diferentes áreas de interesse e necessidades do utilizador.

Essa abordagem possibilitará a identificação de oportunidades de melhoria na apresentação dos dados e na funcionalidade da ferramenta, contribuindo para uma melhor utilização dos recursos públicos disponíveis e para uma maior transparência das informações.

Quadro 1 – Quadro Metodológico

<i>Quanto à Natureza</i>	<i>Aplicada</i>
<i>Objetivos</i>	<i>Exploratório</i>
<i>Procedimentos</i>	<i>Estudo de Caso</i>
<i>Forma de abordagem</i>	<i>Quantitativa</i>

Fonte: Próprio Autor (2024)

#### 3.2 Classificação Metodológica

Neste estudo, foi adotada uma abordagem predominantemente quantitativa. A metodologia visou analisar os dados disponíveis na API da prefeitura de Belo Jardim de forma objetiva e quantificável. Isto envolve a recolha e análise de dados digitais, tais como estatísticas, indicadores socioeconómicos e informações de serviço público.

Os métodos quantitativos buscam identificar padrões, tendências e *insights* a partir de conjuntos de dados coletados. Não foram coletadas opiniões individuais de usuários ou

aspectos subjetivos sobre a usabilidade da aplicação *web*, o foco principal está na análise de dados brutos para gerar *insights* e informações úteis para construção de *dashboards* e tomada de decisões.

### **3.3 Quanto à Natureza**

A natureza deste estudo foi predominantemente aplicada, tendo como foco principal a criação de uma solução prática (aplicação *web*) para usabilidade e visualização de dados públicos. O objetivo é fornecer aos moradores de Belo Jardim uma ferramenta útil e fácil de usar que aumente a transparência e promova a tomada de decisões informadas pelos cidadãos.

### **3.4 Quanto ao Objetivo**

A metodologia adotada foi exploratória, buscando compreender em profundidade como a utilização de dados abertos da prefeitura de Belo Jardim pode impactar a construção de *dashboards* interativos e influenciar na tomada de decisões dos usuários da aplicação *web*. O objetivo principal foi investigar as potenciais contribuições desses dados para a melhoria da gestão pública e o engajamento da comunidade local.

### **3.5 Quanto aos Procedimentos**

Este estudo seguiu um delineamento de estudo de caso, onde foram coletadas informações diretamente da API disponibilizada pela prefeitura de Belo Jardim. Essas informações incluíram dados relacionados a serviços públicos, indicadores socioeconômicos, eventos e outras informações relevantes para a comunidade. A coleta desses dados permitiu uma análise detalhada da sua estrutura, qualidade e potencial de uso na criação dos *dashboards* interativos.

### **3.6 Quanto à Abordagem**

Quanto à abordagem, a metodologia é classificada como quantitativa, onde foi buscado compreender de forma abrangente como a utilização dos dados abertos da prefeitura de Belo Jardim pode ser aproveitada na construção de *dashboards* interativos, a fim de identificar oportunidades de melhorias e potenciais benefícios para os moradores.

## 4. O sistema Dataharbor

### 4.1. Resumo do projeto

O projeto DataHarbor é um sistema *web*, projetado para transformar a maneira como cidadãos, entidades e analistas acessam e interpretam dados de licitações e despesas da prefeitura da cidade de Belo Jardim, no agreste pernambucano. Através da utilização de *dashboards* dinâmicos e atualizados em tempo real, esta plataforma busca democratizar o acesso a dados públicos governamentais disponibilizados de forma gratuita, promovendo uma maior transparência e eficiência na gestão dos seus recursos. A estratégia central do projeto envolve a integração direta com a base de dados fornecida pelo site da prefeitura, garantindo que as visualizações sejam não apenas interativas, mas também reflitam as informações mais atuais disponíveis. Os usuários poderão filtrar, explorar e analisar os dados de licitações e despesas através de uma interface intuitiva, compostas por *dashboards* personalizados que oferecem uma variedade de visualizações de dados, permitindo análises com facilidade de interpretação. A implementação do DataHarbor não só alinha a cidade de Belo Jardim com as melhores práticas em termos de dados abertos e governança digital, mas também incentiva que outras cidades venham a ter uma abordagem mais proativa no gerenciamento de dados públicos, fortalecendo a confiança nas instituições governamentais, melhorando a qualidade dos serviços, aumentando a transparência e aproximando o cidadão do governo.

### 4.2 Objetivos da ferramenta

1. Promoção da transparência governamental: Facilitar o acesso aos dados públicos da prefeitura da cidade de Belo Jardim, de forma estruturada, coesa e centralizada, permitindo que a sociedade, entidades e organizações da sociedade civil tenham uma visão clara das operações de licitações e despesas governamentais através de uma visualização intuitiva dessas informações.
2. Fomento à participação cidadã: Ao tornar os dados públicos da prefeitura da cidade de Belo Jardim acessíveis e compreensíveis, o sistema DataHarbor visa estimular uma maior participação dos cidadãos nos processos decisórios e de fiscalização, promovendo uma cultura de engajamento cívico.
3. Incentivo à inovação e desenvolvimento econômico: Através da visualização centralizada e estruturada dos dados públicos, serão geradas novas percepções, permitindo que qualquer interessado, como pesquisadores, empreendedores ou

empresas desenvolvam novos produtos, serviços ou soluções que atendam às necessidades específicas levantadas.

4. Fortalecimento da confiança pública: Ao demonstrar compromisso com a transparência e a responsabilidade, o sistema DataHarbor pode ajudar a construir e restaurar a confiança entre os cidadãos e o governo, essencial para uma sociedade democrática saudável.
5. Suporte à tomada de decisão baseada em dados: Poderão ser fundamentadas decisões estratégicas e operacionais através do acesso a dados confiáveis e atualizados, oferecendo oportunidades para otimizar resultados e mitigar riscos.

### **4.3 Origem dos dados**

Os dados utilizados para coleta, limpeza, tratamento, preparação e construção das *dashboards* foram extraídos do site da prefeitura da cidade de Belo Jardim, através do link <https://belojardim.pe.gov.br/>

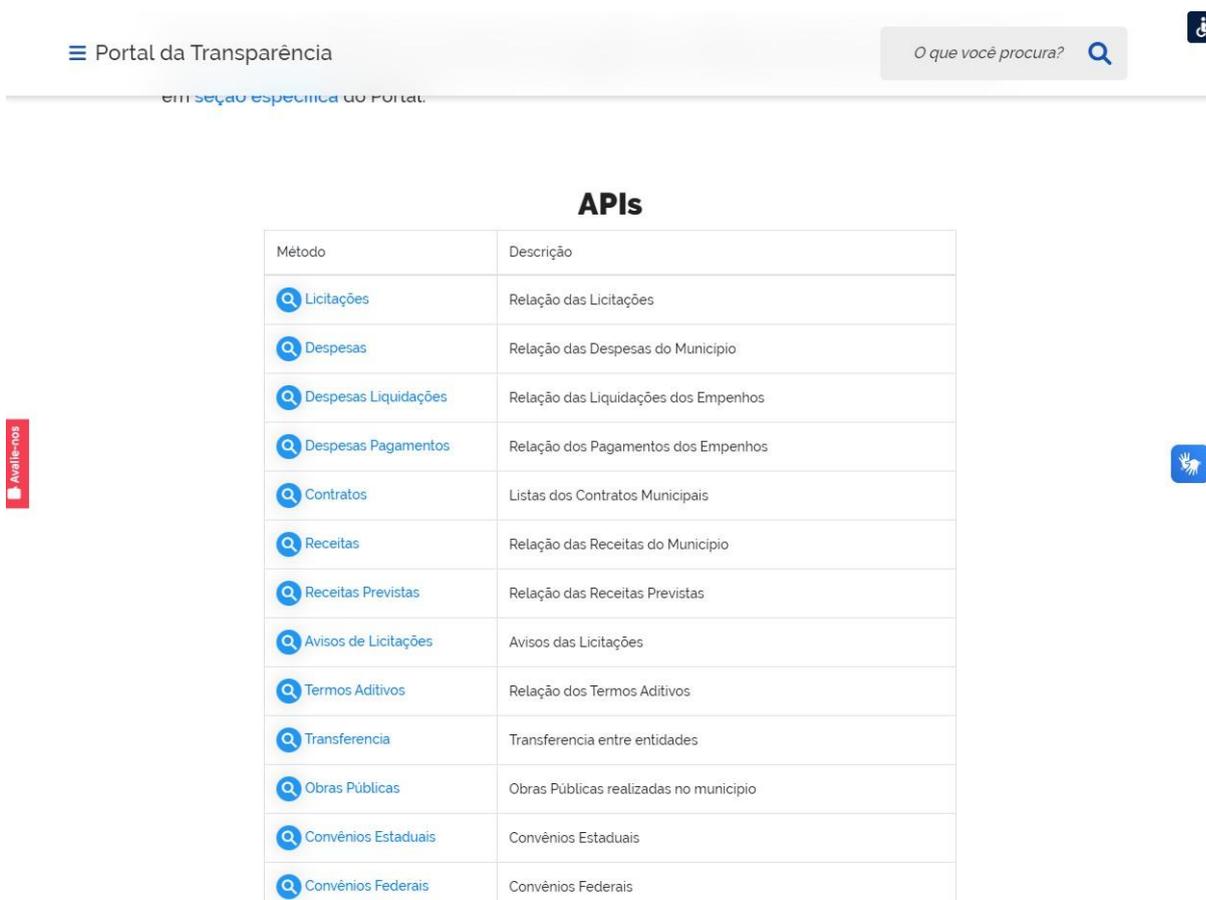
### **4.4 Processo de ETL**

Conforme descrito pelo Google Cloud, ETL é a sigla para o processo de extrair, transformar e carregar dados, sendo uma peça fundamental no processo de manipulação e análise de dados no campo da que envolve a ciência e análise de dados. A necessidade de integrar dados de diversas fontes, transformá-los em formatos adequados e carregá-los em um local centralizado é uma realidade enfrentada na era em que a informação é abundante e essencial para a tomada de decisões fundamentadas. Como parte do desenvolvimento da aplicação Dataharbor, este processo foi separado e descrito nas etapas a seguir, de Processo de coleta de dados e Limpeza e preparação de dados.

#### **4.4.1 Processo de coleta dos dados**

As api's disponibilizadas se apresentam através do link: <https://belojardim.pe.gov.br/transparencia/dados-abertos/lista-api?ug-codigo=&ano=> , onde é possível obtermos acesso a diversas bases, referentes a diversos segmentos do governo:

Figura 1: Api's disponibilizadas



Método	Descrição
Licitações	Relação das Licitações
Despesas	Relação das Despesas do Município
Despesas Liquidações	Relação das Liquidações dos Empenhos
Despesas Pagamentos	Relação dos Pagamentos dos Empenhos
Contratos	Listas dos Contratos Municipais
Receitas	Relação das Receitas do Município
Receitas Previstas	Relação das Receitas Previstas
Avisos de Licitações	Avisos das Licitações
Termos Aditivos	Relação dos Termos Aditivos
Transferencia	Transferencia entre entidades
Obras Públicas	Obras Públicas realizadas no município
Convênios Estaduais	Convênios Estaduais
Convênios Federais	Convênios Federais

Fonte: [https://belojardim.pe.gov.br/transparencia/dados-abertos/lista-api?ug-codigo=&ano=\(2024\)](https://belojardim.pe.gov.br/transparencia/dados-abertos/lista-api?ug-codigo=&ano=(2024))

Para coleta de dados referentes às licitações, foi utilizado o endereço <https://belojardim.pe.gov.br/transparencia/api/licitacoes> e para coleta de dados referentes às despesas, foi utilizado o endereço <https://belojardim.pe.gov.br/transparencia/api/despesas>. Para tais coletas, foi utilizado a linguagem *Python*, juntamente com a biblioteca *Requests*, como uma solução eficaz para automatizar a coleta de dados de licitações e despesas a partir das fontes especificadas. Este processo envolveu a criação de uma função em *Python* que é responsável por acessar o endereço fornecido, onde os dados estão disponíveis para realizar a coleta de informações pertinentes.

A função desenvolvida para coleta utiliza a biblioteca *Requests* para enviar solicitações HTTP ao endereço em questão, capturando os dados retornados, que são estruturados em um formato JSON, permitindo a extração eficiente dos dados de licitação e despesas para posteriormente, serem aplicadas técnicas de análise, processamento e visualização de dados, onde segundo MCKINNEY (2018, p. 20) “Em conjunto com a

robustez de *Python* para uma engenharia de software de propósito geral, é uma excelente opção como uma linguagem principal para a construção de aplicações de dados”.

#### 4.4.2 Limpeza e preparação de dados

Quando se lida com um grande volume de dados, temos como principais etapas a limpeza e preparação dos dados, pois as informações obtidas podem apresentar inconsistências, valores ausentes, nulos e dados em formatos variados que não condizem com o que a informação de fato deveria trazer (MCKINNEY, 2018). A utilização da linguagem *Python*, em conjunto com a biblioteca *Pandas*, fornece um conjunto poderoso de ferramentas para realizar essas tarefas de maneira eficiente e eficaz.

Na elaboração do projeto DataHarbor, foi realizado o processo de limpeza e preparação dos dados referente às bases de dados de Despesas Públicas e Licitações. O processo se iniciou com a utilização do método *'fillna'*, onde os dados faltantes das colunas *'VAPAG'*, *'VALOR\_GLOBAL'* e *'LIQUIDADO'*, na base de dados de Despesas Públicas, que representam dados de valor na moeda real, foram substituídos pelo valor 0; o mesmo tratamento foi aplicado na coluna *'VALOR\_ESTIMADO'* na base de dados de Licitações, pois o tratamento de valores ausentes, especialmente em campos financeiros onde a precisão é essencial, onde valores faltantes podem distorcer análises e levar a conclusões errôneas.

O próximo passo para limpeza dos dados foi a conversão de dados de texto em números, pois muitas vezes, informações numéricas são importadas como *strings* (texto), o que impede a realização de cálculos e análises estatísticas, pois esses dados acabam vindo num formato incorreto para tais análises (MCKINNEY, 2018 pag. 248). Com a utilização de métodos do *Pandas*, esses dados são facilmente convertidos para tipos numéricos, como *float* ou *integer*, permitindo assim sua manipulação matemática e inclusão em análises quantitativa, para essa manipulação, foi utilizado o método *'to\_numeric'* nas mesmas colunas citadas acima, convertendo seu tipo de *string* para *integer*.

Outra transformação realizada refere-se à manipulação de datas, foi aplicada a conversão em que dados em texto que representam datas são convertidos para o formato *datetime* do *Python*, realizando assim uma representação especializada que facilita a realização de operações como comparações, extração de componentes específicos da data (dia, mês, ano) e cálculo de intervalos de tempo (Awari, 2023). Esta conversão é crucial para análises temporais e séries temporais, onde a dimensão tempo desempenha um papel central. Para este tratamento foi utilizado o método do *Pandas* *'to\_datetime'*, sendo aplicado na coluna

‘DATA\_EMPENHO’, pertencente à base de dados de Despesas Públicas e nas colunas de ‘DATA\_PUBLICACAO\_HOMOLOGACAO’ e ‘DATA\_PUBLICACAO\_HABILITACAO’ nas colunas pertencentes à base de dados de Licitações.

Ao final desta etapa, teremos os dados tratados e prontos para utilização na aplicação, podendo fornecer uma melhor análise descritiva com maior precisão e confiabilidade.

## 4.5 Desenvolvimento da aplicação

### 4.5.1 Tecnologias Utilizadas

#### 4.5.1.1 Frontend

*Frontend:* Visualização de Dados com Dash, HTML, CSS e Bootstrap

No *frontend*, a utilização da biblioteca *Dash*, construída sobre *Plotly.js*, *Flask* e *React.js*, permite o desenvolvimento de aplicações *web* de análise de dados altamente interativas, sem a necessidade de escrever extenso código *JavaScript*, tornando as aplicações *react* naturalmente performáticas, devido à sua simplicidade, rapidez e flexibilidade (Prata, 2023). *Dash* facilita a criação de *dashboards* de dados com componentes ricos em funcionalidades, que podem ser personalizados e estilizados usando *HTML* e *CSS*. A integração com *Bootstrap* potencializa a responsividade e estética da aplicação, fornecendo uma vasta gama de templates e componentes para uma interface de usuário (UI) elegante e funcional. Essa abordagem não apenas melhora a acessibilidade e a usabilidade da aplicação, mas também permite que cientistas de dados concentrem-se na lógica e análise, sem comprometer a qualidade da apresentação dos dados (Abreu, 2023).

#### 4.5.1.2 Backend

*Backend:* Processamento e Análise de Dados com *Python*

No *backend*, a aplicação utiliza amplamente do ecossistema *Python*, especialmente das bibliotecas *Requests*, *Flask*, *Pandas* e *Plotly*. *Flask* serve como uma *microframework* para aplicações *web*, oferecendo a flexibilidade necessária para suportar as requisições do lado do cliente, gerenciar sessões de usuário e integrar-se com sistemas de banco de dados (Gonzalo, 2023) como o *MySQL*. Ainda segundo McKinney, *Pandas* é utilizado para manipulação e

análise de dados, fornecendo estruturas de dados rápidas, flexíveis e expressivas, ideais para trabalhar com dados tabulares ou heterogêneos (McKinney, 2011). *Plotly*, por sua vez, é empregado para gerar gráficos interativos que podem ser incorporados diretamente nas *dashboards Dash*, facilitando a visualização de dados complexos de forma intuitiva.

#### 4.5.1.3 Gerenciamento de banco de dados

Integração com *MySQL*: Gerenciamento de Dados

A escolha do *MySQL* como sistema de gerenciamento de banco de dados reflete a necessidade de uma plataforma confiável e eficiente para armazenamento, recuperação e manipulação de grandes volumes de dados, apresentando também vantagens como custo gratuito, excelentes recursos de autenticação e segurança e um grande suporte da comunidade (Oracle, s.d.) A integração do *MySQL* com *Python*, através de bibliotecas como *MySQL Connector/Python*, permite que a aplicação execute operações de banco de dados complexas dentro do contexto de análise de dados, garantindo que os dados sejam facilmente acessíveis e gerenciáveis.

#### 4.5.2 Metodologia de desenvolvimento

No contexto do desenvolvimento da aplicação, a metodologia ágil *Scrum* foi adotada para garantir uma abordagem ágil, iterativa e eficiente (Dessoldi, 2019). Foram definidas *sprints* específicas para cada fase do processo, começando pela coleta de dados através das APIs disponibilizadas pela prefeitura. Durante essas *sprints*, foram realizadas análises preliminares dos dados coletados para identificar possíveis inconsistências e garantir a qualidade dos dados a serem utilizados. Em seguida, foi dedicada uma *sprint* para a limpeza e preparação dos dados, incluindo a remoção de duplicatas, tratamento de valores ausentes e padronização dos formatos.

Posteriormente, foram criadas *sprints* individuais para a elaboração de gráficos e visualizações separadas, cada uma focada em um aspecto específico dos dados, como gráficos de barras e de pizzas para dados categóricos e gráficos de linha para tendências temporais, como citam as documentações da IBM (IBM, 2023). Essa abordagem permitiu uma atenção detalhada a cada tipo de visualização, garantindo a clareza e a relevância das informações apresentadas nos dashboards interativos.

Houve também a criação de uma *sprint* dedicada ao desenvolvimento da aplicação

*web* utilizando o framework *Flask*, que oferece uma estrutura leve e flexível para construção de aplicações *web* em *Python* (Gonzalo, 2023). Durante essa fase, foram implementadas as funcionalidades de interatividade nos *dashboards*, permitindo aos usuários explorar os dados de forma dinâmica e intuitiva.

Por fim, as *sprints* anteriores foram integradas em uma sprint de finalização, na qual todos os componentes da aplicação, incluindo a coleta e preparação de dados, as visualizações criadas e a aplicação *Flask*, foram unidos e testados em conjunto para garantir o funcionamento integrado e a qualidade da aplicação final.

## 4.6 Casos de uso

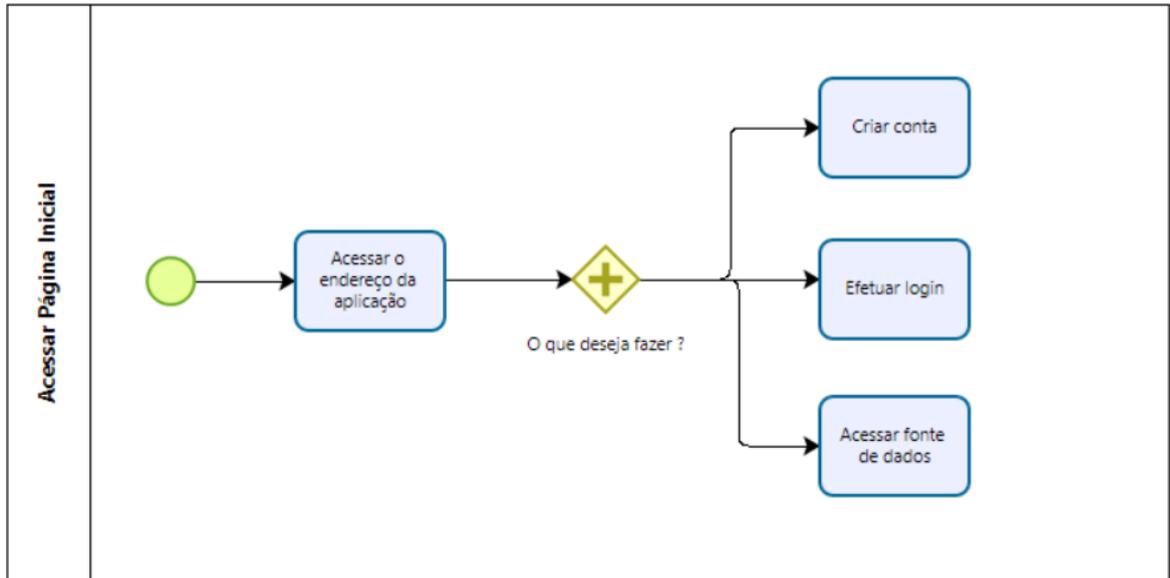
### 4.6.1 Página inicial

Funcionalidade: A tela da página inicial é o ponto de entrada para os usuários da aplicação, que contém elementos projetados para orientar o usuário, tanto na navegação, quanto fornecendo uma introdução clara e direta do objetivo da plataforma.

Fluxo de interações:

1. O usuário acessa a aplicação *web* e é direcionado para tela da página inicial
2. O usuário visualiza o conteúdo principal da página, contendo informações claras do objetivo da aplicação
3. O usuário tem a opção de criar uma conta para acessar os dados, caso ainda não tenha se registrado. Esses dados serão provenientes das telas seguintes e requerem um *login* válido. Alternativamente, o usuário pode fazer o *login* se já tiver uma conta registrada. Ambas as opções são facilmente acessíveis no cabeçalho localizado no topo da aplicação.
4. O usuário também pode ser encaminhado para a fonte dos dados utilizados na aplicação ao clicar no link destacado em vermelho, identificado como "aqui" na página.

Figura 2: Fluxograma de interação - Página Inicial



Fonte: Próprio Autor (2024)

Figura 3: Página inicial DataHarbor



Fonte: Próprio Autor (2024)

#### 4.6.2 Registrar usuário

**Funcionalidade:** Permite que novos usuários que desejem acessar a plataforma façam registro de uma conta, preenchendo e enviando o formulário, seus dados serão processados e armazenados de forma segura no banco de dados da aplicação.

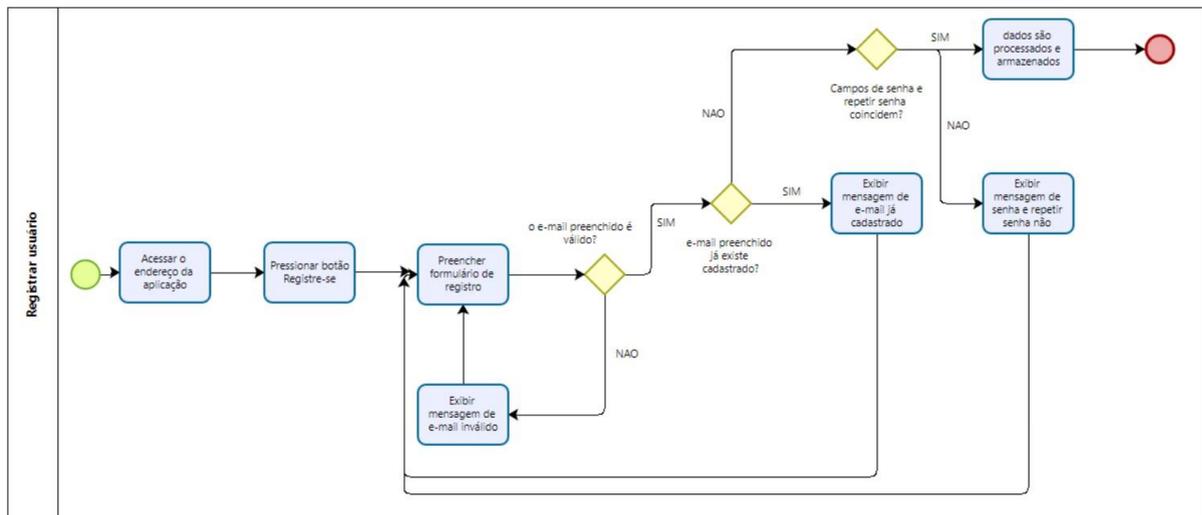
**Fluxo de interações:**

1. O usuário acessa aplicação e clica no botão “Registre-se” na página inicial
2. O usuário preenche o formulário de registro, com um e-mail válido, sua senha, e deverá repetir a senha e clicar em “Cadastrar”.
3. O sistema processa e armazena seus dados de forma segura no banco de dados

**Validações:**

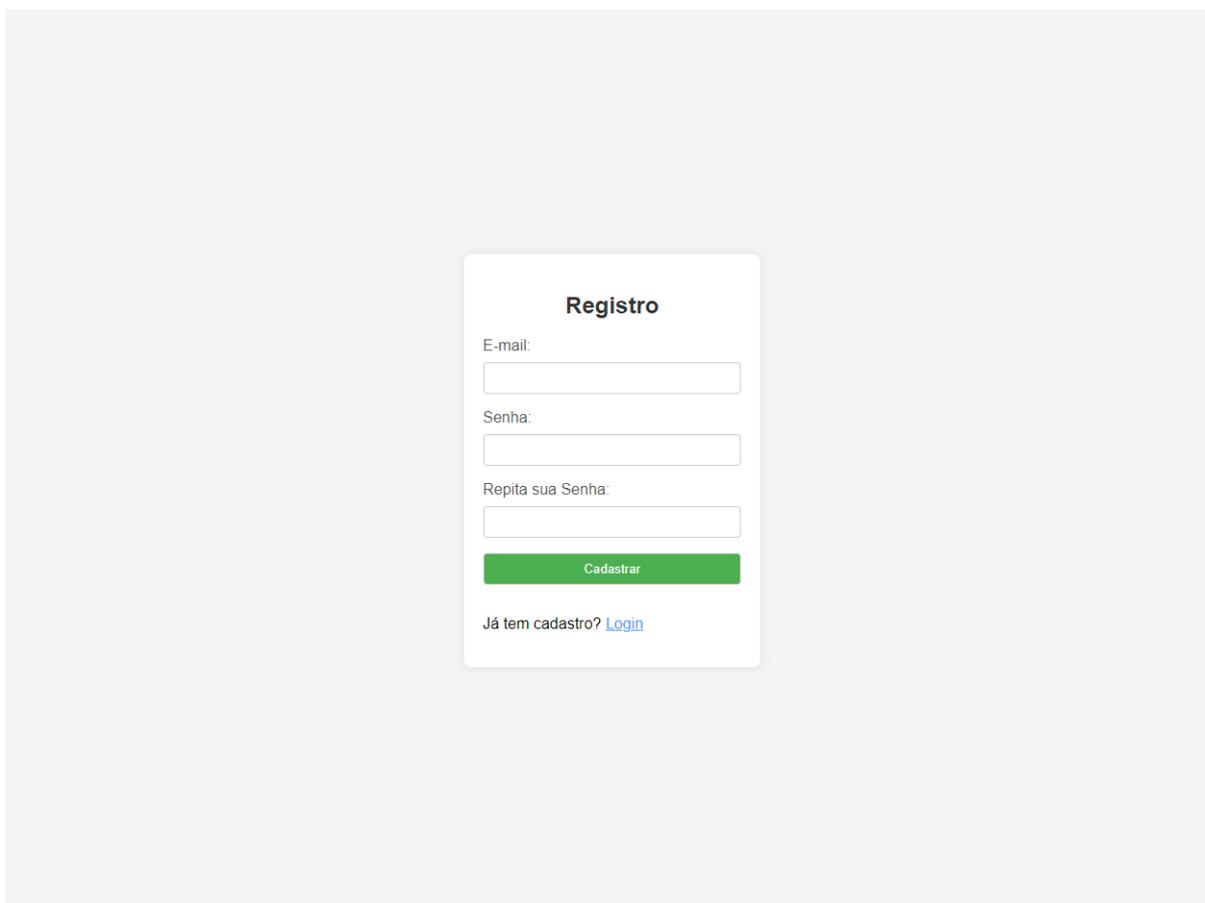
1. Caso o usuário insira um email inválido no campo de email, uma mensagem será exibida solicitando que um email válido seja fornecido, e o cadastro não será concluído.
2. Se o usuário fornecer um e-mail que já está cadastrado no sistema, uma mensagem informando que esse e-mail já está em uso será exibida, e o cadastro não será permitido.
3. Se as senhas nos campos de senha e de repetição de senha não coincidirem, uma mensagem de erro será exibida, impedindo que o cadastro seja realizado.

Figura 4: Fluxograma de interação - Registrar Usuário



Fonte: Próprio Autor (2024)

Figura 5: Tela de registrar usuário



A imagem mostra uma tela de registro de usuário. No topo, o título "Registro" está centralizado. Abaixo dele, há três campos de entrada de texto: "E-mail:", "Senha:" e "Repita sua Senha:". Cada campo é precedido por seu respectivo rótulo. Abaixo dos campos, há um botão verde com o texto "Cadastrar". Na base da tela, há o texto "Já tem cadastro? [Login](#)".

Fonte: Próprio autor (2024)

### 4.6.3 Login no sistema

Funcionalidade: Permite que usuários cadastrados se autentiquem no sistema para acessar as áreas protegidas da aplicação.

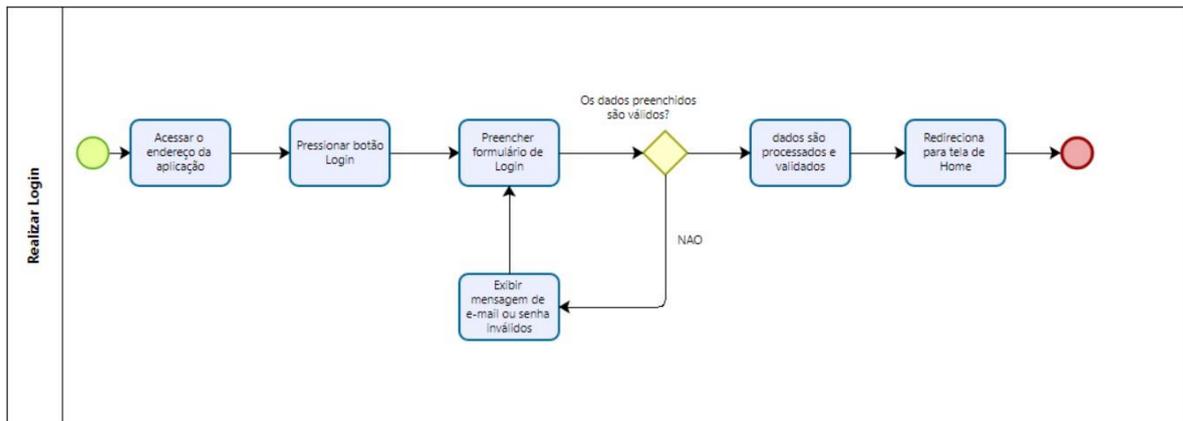
Fluxo de interações:

1. O usuário acessa aplicação e clica no botão “*Login*” na página inicial
2. O usuário preenche o formulário de login, com um e-mail válido, cadastrado posteriormente, insere sua senha e clica em “*Login*”.
3. O sistema processa e valida seus dados, realizando o redirecionamento para tela de *Home*.

Validações:

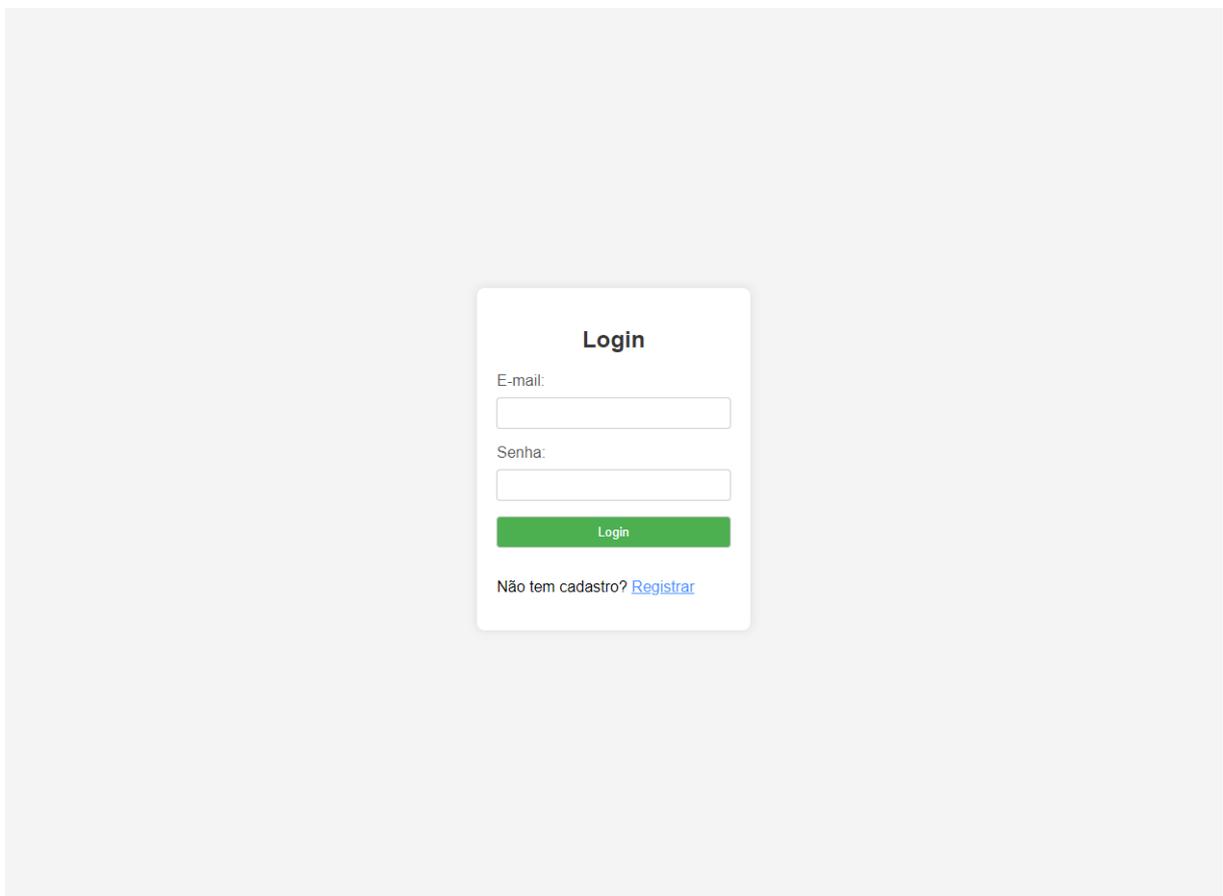
1. Caso o usuário insira um e-mail ou senha incorretos, uma mensagem será exibida indicando que o login ou senha estão incorretos, e o processo de *login* não será concluído.

Figura 6: Fluxograma de interação - Realizar *Login*



Fonte: Próprio autor (2024)

Figura 7: Tela de *login* no sistema



Fonte: Próprio autor (2024)

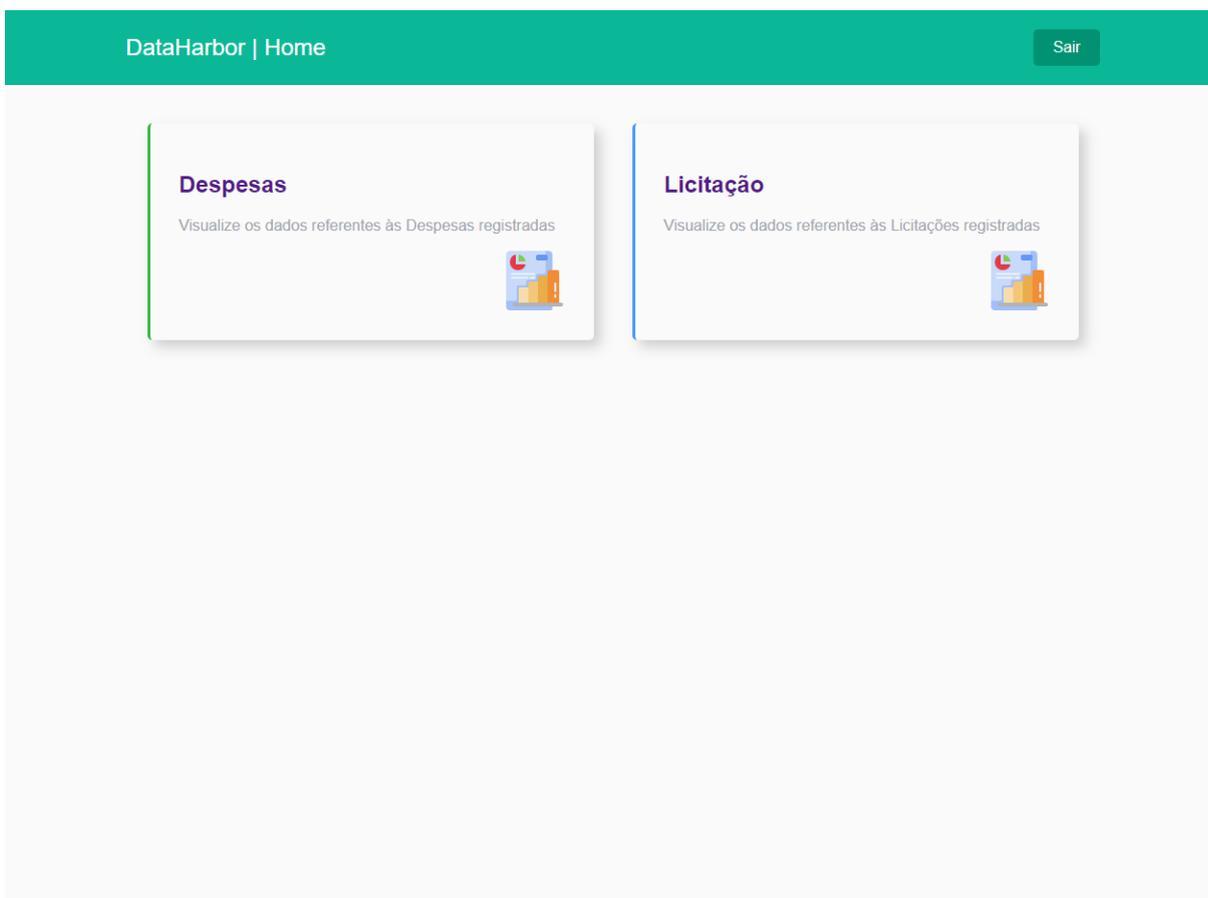
#### 4.6.4 Home

Funcionalidade: A tela de *home* serve como ponto de partida para que os usuários após o *login*, possam ter uma visão geral das *dashboards* disponíveis para acesso das informações.

Fluxo de interações:

1. O usuário acessa aplicação e clica no botão “*Login*” na página inicial
2. O usuário deverá ter feito seu *login* com sucesso
3. O usuário poderá clicar no card de “Despesas” para acessar a *dashboard* de Despesas Públicas ou no *card* “Licitação” para acessar a *dashboard* de Licitações.

Figura 8: Tela de *home*



Fonte: Próprio autor (2024)

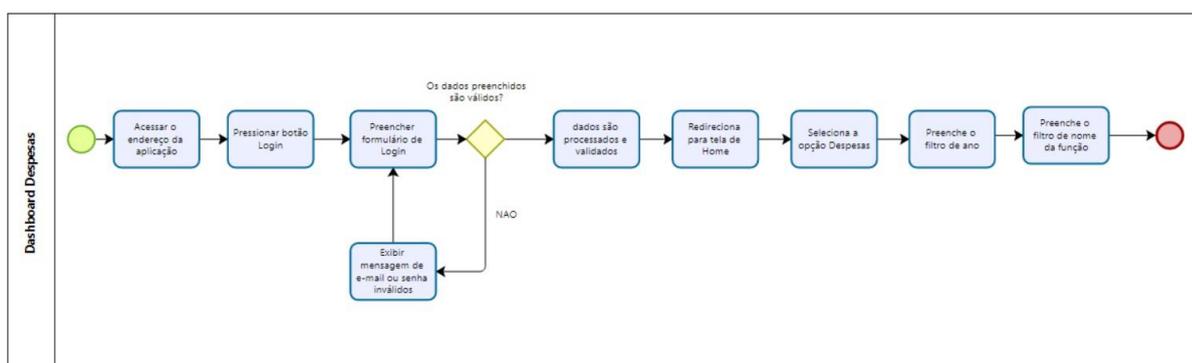
#### 4.6.5 Dashboard de despesas públicas

Funcionalidade: Visa fornecer uma visualização clara da distribuição dos dados, baseada nas filtragens e interações do usuários feitas de acordo com o ano e nome da função, permitindo comparações diretas entre os gráficos em diferentes períodos de tempo, sendo uma ferramenta essencial para fornecer aos gestores e a população informações importantes, relevantes e de fácil entendimento. As *dashboards* são projetadas para serem altamente interativas, permitindo uma fácil exploração através de seus filtros citados, para uma melhor análise dos dados de forma visual e rápida.

Fluxo de interações:

1. O usuário acessa aplicação e clica no botão “Login” na página inicial
2. O usuário deverá ter feito seu *login* com sucesso
3. Na tela de *Home* o usuário deverá selecionar “Despesas”
4. Após o preenchimento de filtro de Ano, o usuário deverá aguardar o sistema carregar os dados
5. Deverá selecionar as opções disponíveis dentro do filtro de Nome da Função

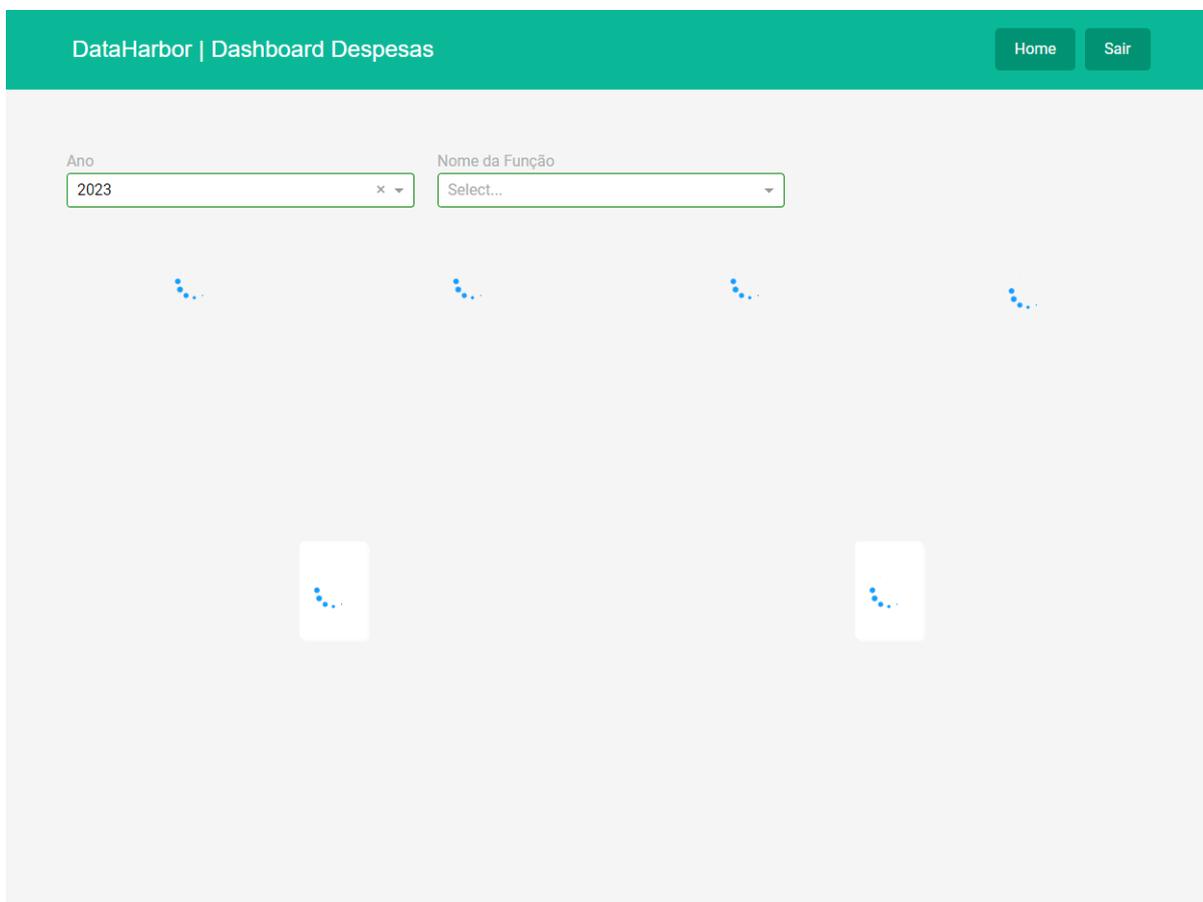
Figura 9: Fluxograma de interação - *Dashboard* Despesas



Fonte: Próprio autor (2024)

Imagem com os dados sendo carregados:

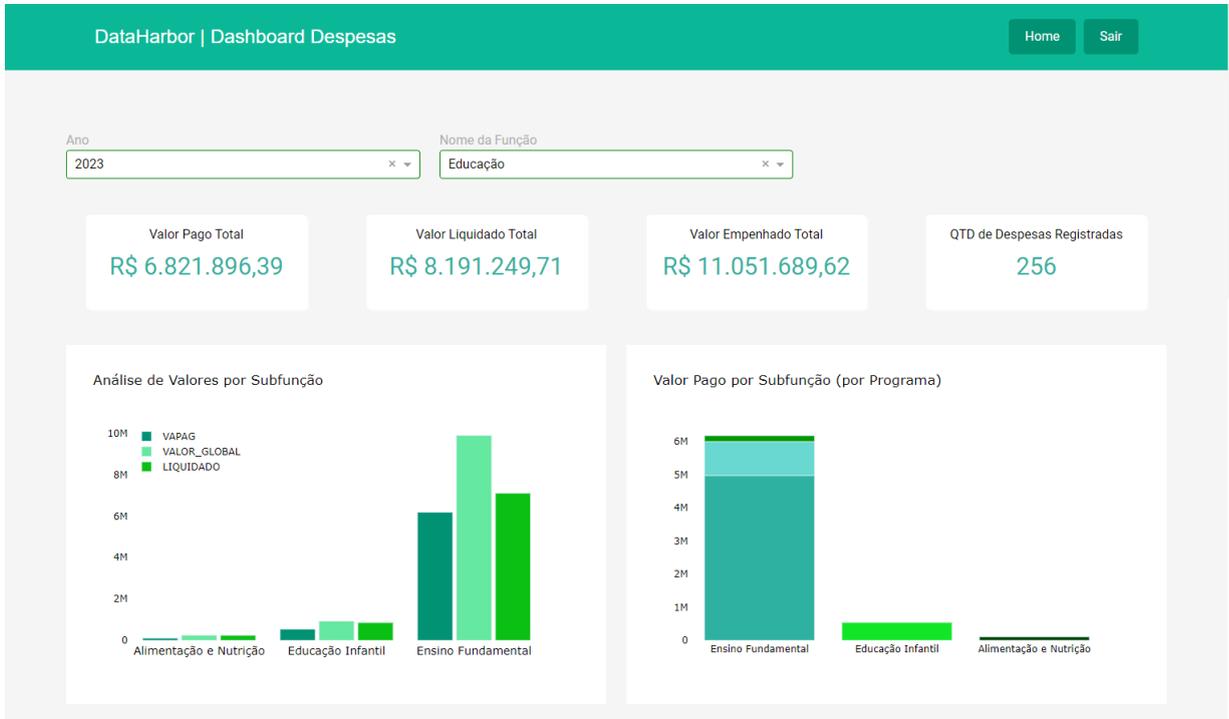
Figura 10: *Dashboard* de despesas públicas carregando dados



Fonte: Próprio autor (2024)

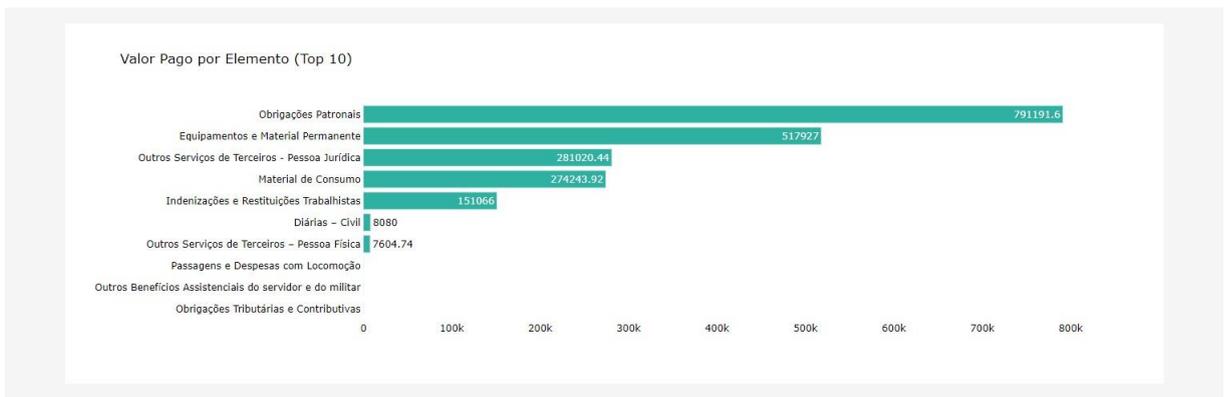
Dashboard após carregamento:

Figura 11: Dashboard de despesas públicas carregados (parte 1)



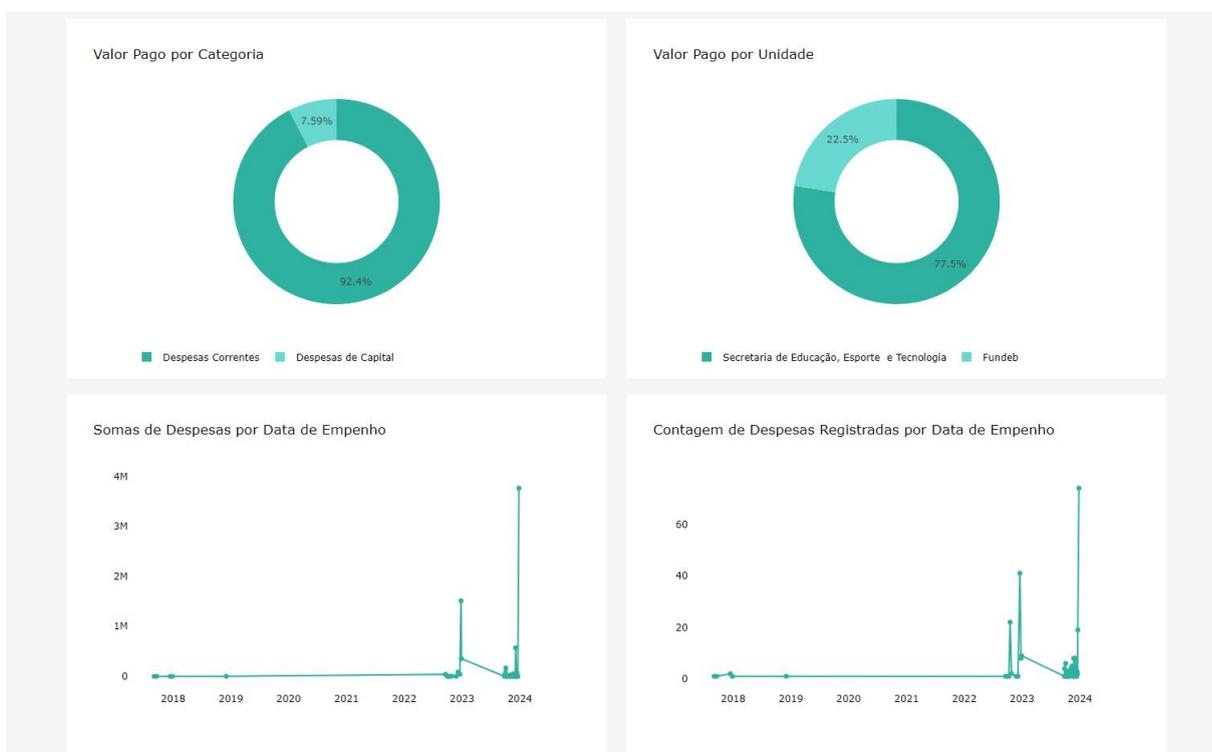
Fonte: Próprio autor (2024)

Figura 12: Dashboard de despesas públicas carregados (parte 2)



Fonte: Próprio autor (2024)

Figura 13: *Dashboard* de despesas públicas carregados (parte 3)



Fonte: Próprio autor (2024)

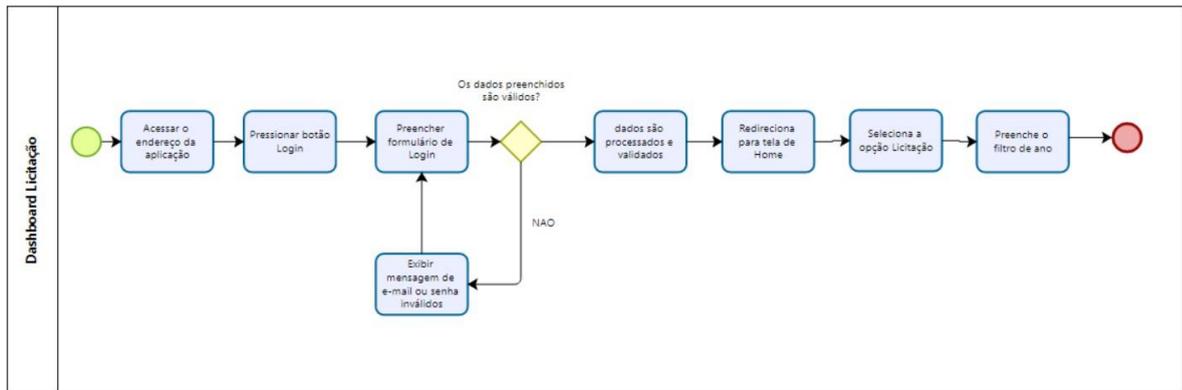
#### 4.6.6 Dashboard de licitação

Funcionalidade: Visa fornecer uma visualização clara da distribuição dos dados, baseada nas filtragens e interações dos usuários feitas de acordo com o ano, permitindo também comparações diretas entre os gráficos em diferentes períodos de tempo, trazendo os mesmos benefícios da *dashboard* de despesas.

Fluxo de interações:

1. O usuário acessa aplicação e clica no botão “*Login*” na página inicial
2. O usuário deverá ter feito seu *login* com sucesso
3. Na tela de *Home* o usuário deverá selecionar “Licitação”
4. Após o preenchimento de filtro de Ano, o usuário deverá aguardar o sistema carregar os dados

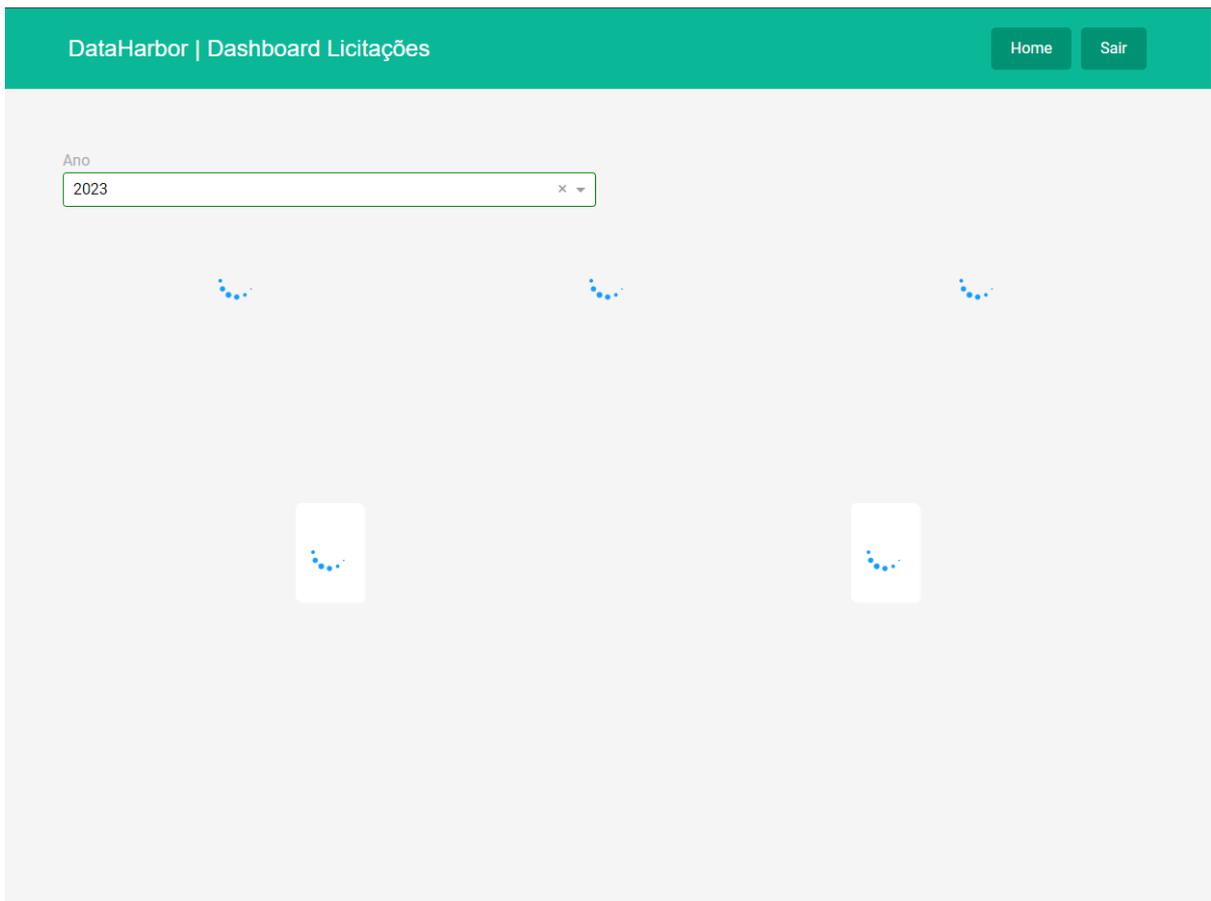
Figura 14: Fluxograma de interação - *Dashboard Licitação*



Fonte: Próprio autor (2024)

Imagem com os dados sendo carregados:

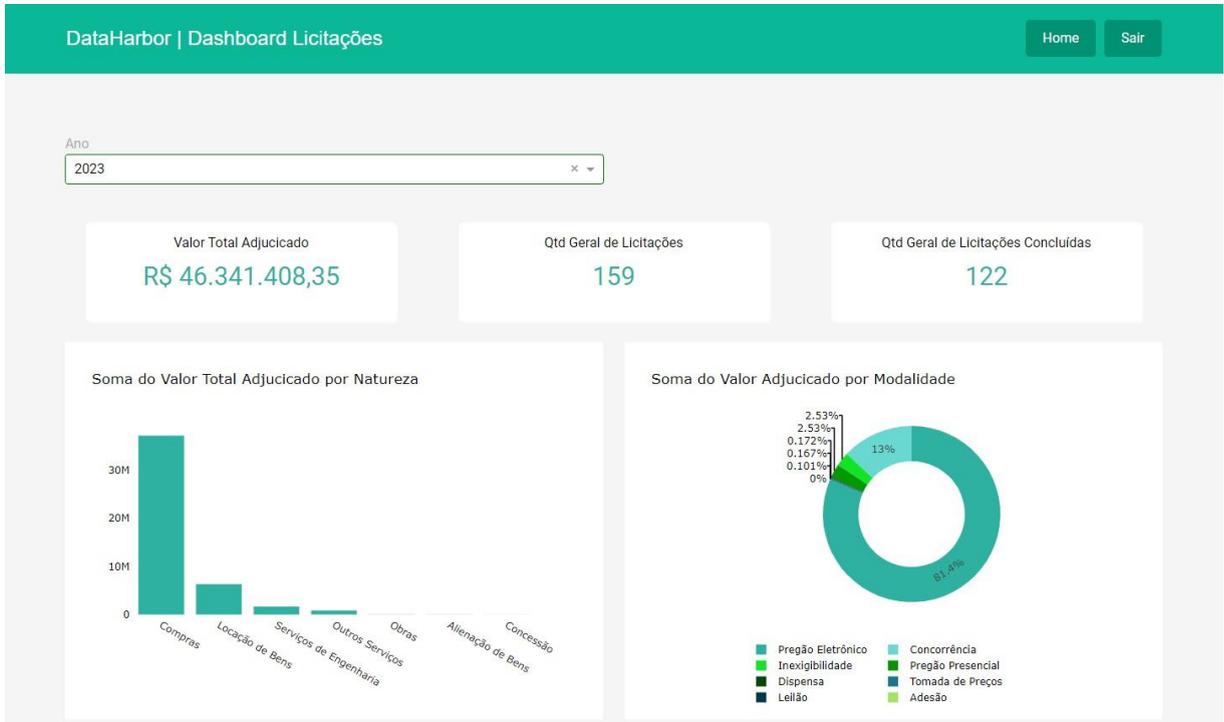
Figura 15: *Dashboard* de licitação carregando dados



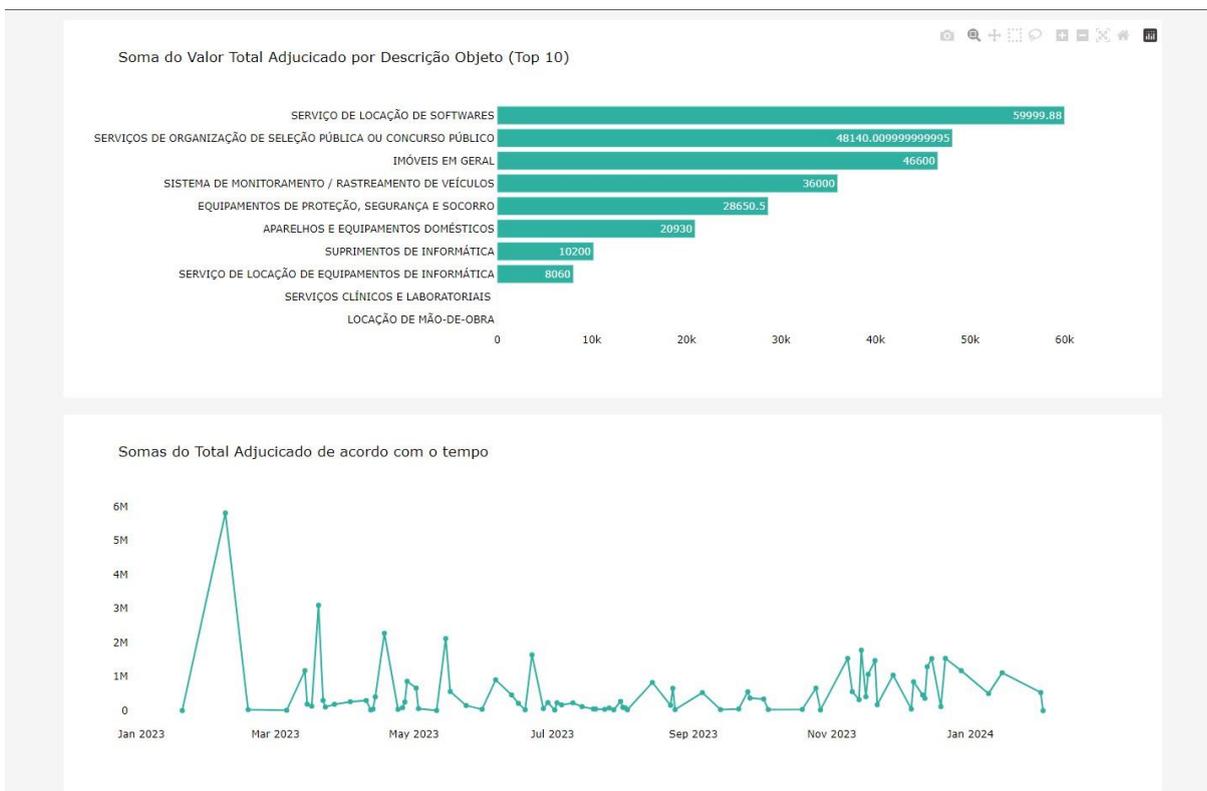
Fonte: Próprio autor (2024)

Dashboard após carregamento:

Figura 16: Dashboard de licitação carregado (parte 1)



Fonte: Próprio autor (2024)

Figura 17: *Dashboard* de licitação carregado (parte 2)

Fonte: Próprio autor (2024)

## 4.7 Análise e objetivos dos componentes da dashboard: Detalhamento das visualizações de dados

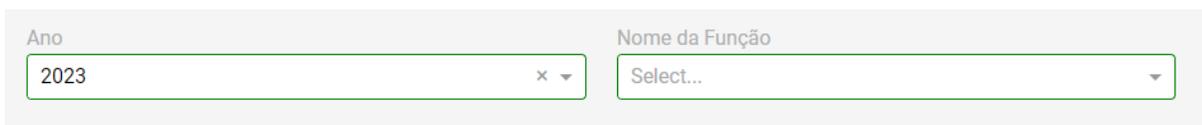
### 4.7.1 Componentes da dashboard de despesas públicas

#### 4.7.1.1 Filtros

##### 4.7.1.1.1 Filtro de ano

Este filtro permite que os usuários selecionem um ano específico para análise dos dados, servindo como ponto de partida para a exploração dos dados. A seleção de um ano ativa automaticamente o filtro subsequente, garantindo que a análise seja focada e relevante para o período escolhido.

Figura 18: Filtro com ano preenchido



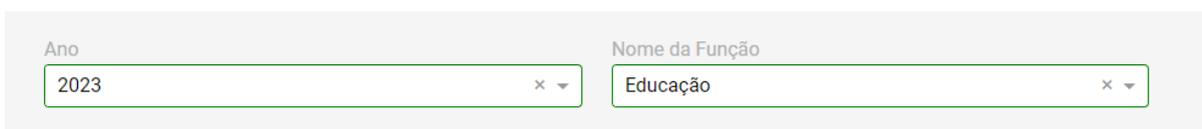
The image shows a filter interface with two dropdown menus. The first dropdown, labeled 'Ano', has '2023' selected. The second dropdown, labeled 'Nome da Função', has 'Select...' selected. Both dropdowns have a small 'x' icon and a downward arrow.

Fonte: Próprio autor (2024)

#### 4.7.1.1.2 Filtro de nome de função

Após a seleção de um ano específico, este filtro apresenta todas as funções cadastradas que tiveram despesas registradas no ano em questão. Esse nível de detalhamento é crucial para os usuários que desejam analisar como os recursos foram alocados para diferentes funções.

Figura 19: Filtro com ano e nome da função preenchidos



The image shows a filter interface with two dropdown menus. The first dropdown, labeled 'Ano', has '2023' selected. The second dropdown, labeled 'Nome da Função', has 'Educação' selected. Both dropdowns have a small 'x' icon and a downward arrow.

Fonte: Próprio autor (2024)

Para demonstrações das imagens abaixo, tratadas como exemplos, foram utilizados valores da primeira página de consulta à api de despesas públicas, utilizando como parâmetro o ano de 2023 e a função de nome Educação, como mostra imagens dos filtros aplicados acima, no dia 06 de março de 2024.

#### 4.7.1.2 Cards

##### 4.7.1.2.1 Card Valor Pago Total

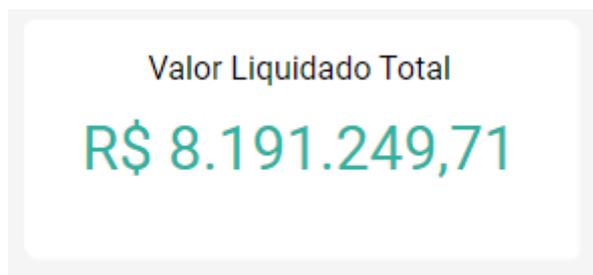
A informação apresentada no *card* "Valor Total Pago" origina-se da coluna específica denominada 'VAPAG' dentro do conjunto de dados utilizado na *Dashboard* de Despesas Públicas. De acordo com o dicionário de dados, a coluna 'VAPAG' representa os valores individuais pagos em cada transação ou registro de despesa. Para calcular o valor total pago, que é a soma de todos os valores individuais registrados na coluna 'VAPAG', foi aplicado um método *'sum'* nos valores dessa coluna, para calcular a soma total desses valores, resultando no valor total pago.

Figura 20: *Card* Valor Pago Total preenchido

Fonte: Próprio autor (2024)

#### 4.7.1.2.2 Card Valor Liquidado Total

Em relação à informação apresentada no *card* "Valor Liquidado Total", a mesma origina-se da coluna denominada 'LIQUIDADO' dentro do mesmo conjunto de dados. A coluna 'LIQUIDADO' representa valores individuais liquidados em cada transação ou registro de despesa. O mesmo método *'sum'* foi aplicado nos valores dessa coluna, resultando na soma total dos valores liquidados.

Figura 21: *Card* Valor Liquidado Total preenchido

Fonte: Próprio autor (2024)

#### 4.7.1.2.3 Card Valor Empenhado Total

A informação apresentada no *card* "Valor Empenhado Total" origina-se da coluna denominada 'VALOR\_GLOBAL', também pertencente ao mesmo conjunto de dados. De acordo com o dicionário de dados, onde anteriormente à atualização, seu valor vinha através da coluna 'VALOR\_EMPENHADO', representa valores individuais empenhados em cada

transação ou registro de despesa. Seguindo o mesmo padrão dos *cards* anteriores, foi aplicado o método `‘.sum’` nos valores dessa coluna, resultando na soma total dos valores empenhados.

Figura 22: *Card* Valor Empenhado Total preenchido



Fonte: Próprio autor (2024)

#### 4.7.1.2.4 Card QTD de Despesas Registradas

A informação apresentada no *card* “QTD de Despesas Registradas” origina-se do cálculo do número de linhas no *DataFrame* filtrado, onde cada linha representa uma despesa registrada que satisfaz os critérios de filtragem. Portanto, o valor apresentado reflete o número total de despesas para a função e ano selecionados, oferecendo uma métrica quantitativa direta das despesas ocorridas sob essas condições específicas.

Figura 23: *Card* QTD de Despesas Registradas preenchido



Fonte: Próprio autor (2024)

#### 4.7.1.3 Gráficos

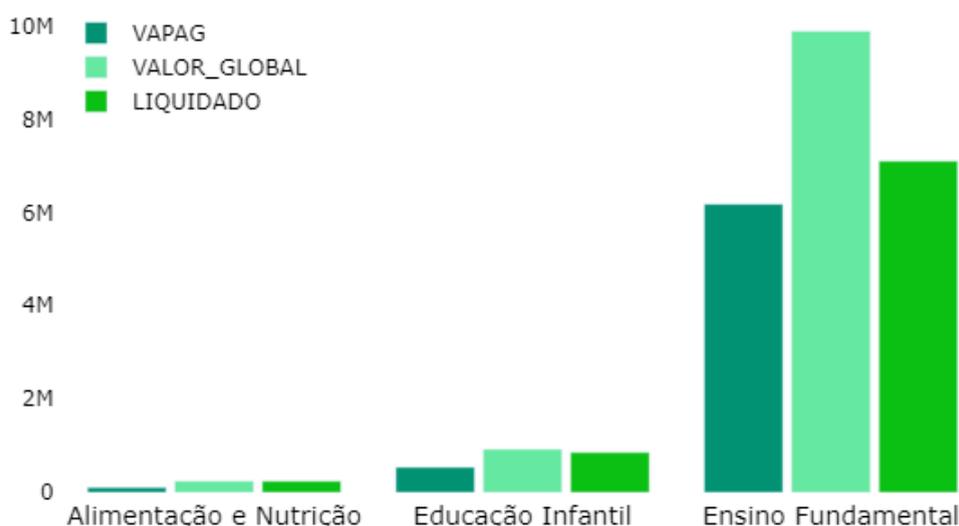
##### 4.7.1.3.1 Análise de Valores por Subfunção

Nesta visualização, cada entidade apresentada no eixo x refere-se ao tipo de subfunção, sendo gerada de acordo com os filtros aplicados. Para cada variável do eixo x, temos três representações em coluna, que se referem a soma de valor pago, valor liquidado e valor

empenhado de cada, permitindo uma comparação entre os três tipos de valores aplicados a cada subfunção, possibilitando uma compreensão mais abrangente e detalhada dos padrões e tendências financeiras. Essa visualização integrada facilita a identificação de padrões de gastos em diferentes áreas da organização. Para visualização deste gráfico, foi aplicado o método do *Pandas* chamado de *'groupby'*, criando uma nova base de dados levando em consideração a coluna de *'NOME\_SUBFUNCAO'*, agregando a ela os valores de soma de *'VAPAG'*, *'VALOR\_GLOBAL'* e *'LIQUIDADO'*.

Figura 24: Gráfico obtido a partir da análise de valores por subfunção

#### Análise de Valores por Subfunção



Fonte: Próprio autor (2024)

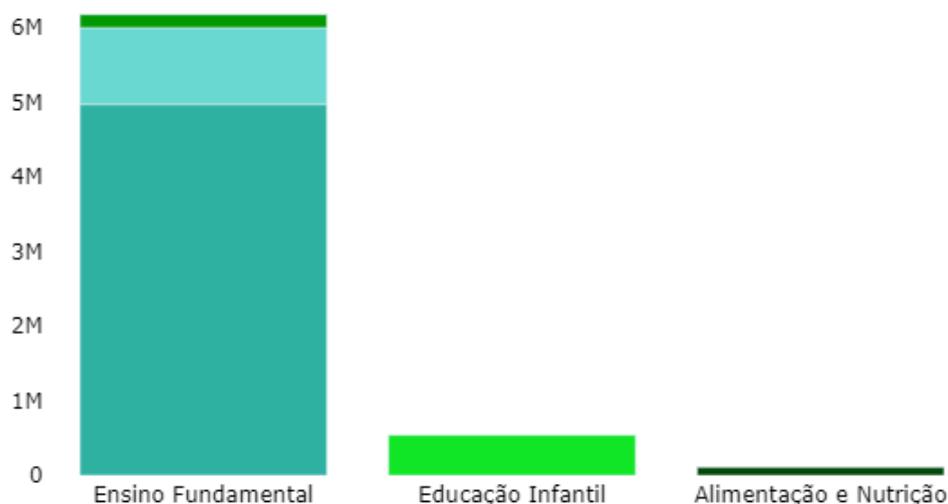
#### 4.7.1.3.2 Valor Pago por Subfunção (por Programa)

Este gráfico de colunas oferece uma visualização complementar ao gráfico anterior, fornecendo uma análise mais detalhada das subfunções. Aqui, as colunas representam a soma dos valores pagos por subfunção, com uma distinção de cores dentro de cada coluna, associada aos respectivos programas aos quais pertencem. Para produzir este gráfico, foi gerado um novo conjunto de dados utilizando o método *'groupby'*, que agregou as somas dos

valores pagos, formando grupos com base nos nomes das subfunções e dos programas correspondentes.

Figura 25: Gráfico obtido a partir da análise de valor pago por subfunção (por programa)

Valor Pago por Subfunção (por Programa)

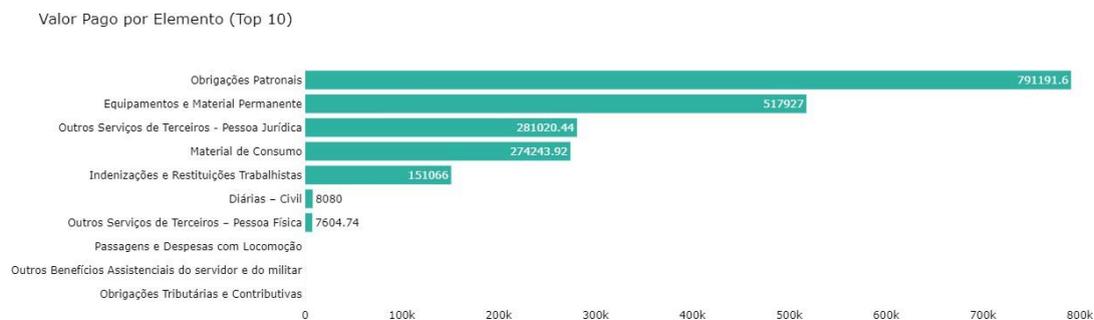


Fonte: Próprio autor (2024)

#### 4.7.1.3.3 Valor Pago por Elemento (Top 10)

Este gráfico de colunas, apresentado em formato horizontal, exibe informações sobre a soma dos valores pagos pelos dez elementos principais, juntamente com o número de ocorrências registradas para cada elemento, destacado ao lado de cada coluna. Para criar essa visualização, foram combinadas duas novas bases de dados filtradas. A primeira foi gerada usando o método `'value_counts'` do *Pandas* para contar o número de registros por elemento na coluna `'NOME_ELEMENTO'`. Já a segunda base de dados foi criada com o método `'groupby'`, que agregou a soma dos valores pagos de acordo com o grupo da coluna `'NOME_ELEMENTO'`. Em seguida, as duas bases de dados foram mescladas utilizando o método `'merge'`, unindo-as com base na coluna `'NOME_ELEMENTO'`.

Figura 26: Gráfico obtido a partir da análise de valor pago por elemento (Top 10)



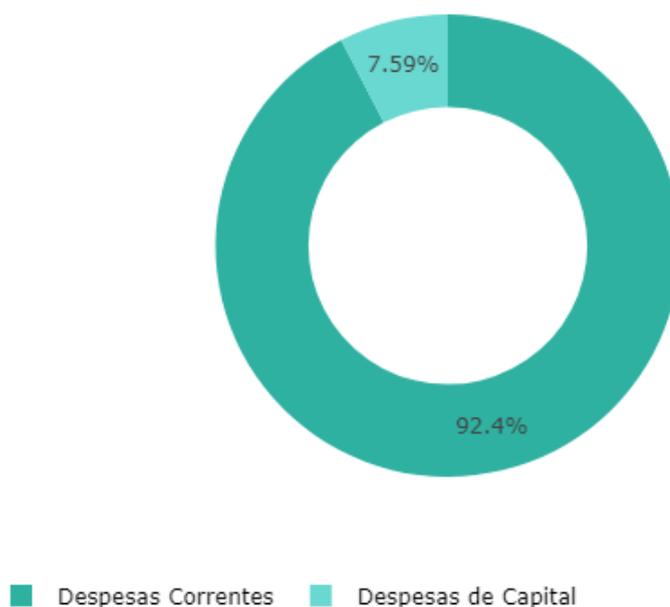
Fonte: Próprio autor (2024)

#### 4.7.1.3.4 Valor Pago por Categoria

Este gráfico de pizza foi gerado utilizando o método '*groupby*', que agregou a soma dos valores pagos com base nos grupos da coluna 'NOME\_CATEGORIA'. Na legenda do gráfico, são listadas as entidades registradas em cada categoria, enquanto as fatias da pizza representam a porcentagem correspondente de cada uma delas.

Figura 27: Gráfico obtido a partir da análise de valor pago por categoria

## Valor Pago por Categoria



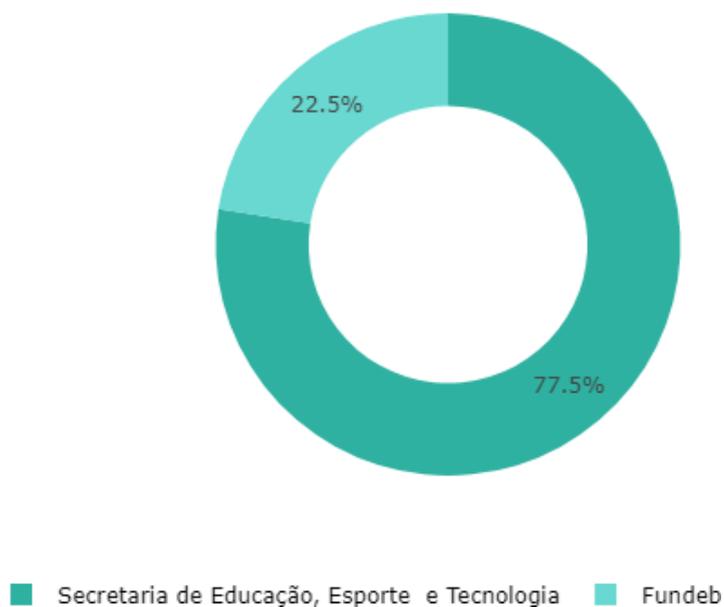
Fonte: Próprio autor (2024)

**4.7.1.3.5 Valor Pago por Unidade**

Gráfico de pizza gerado semelhante ao gráfico de Valor Pago por Categoria, utilizando o método '*groupby*', que agregou a soma dos valores pagos com base nos grupos da coluna 'NOME\_UNIDADE'. Na legenda do gráfico, são listadas as entidades registradas em cada unidade, enquanto as fatias da pizza representam a porcentagem correspondente de cada uma delas.

Figura 28: Gráfico obtido a partir da análise de valor pago por unidade

## Valor Pago por Unidade



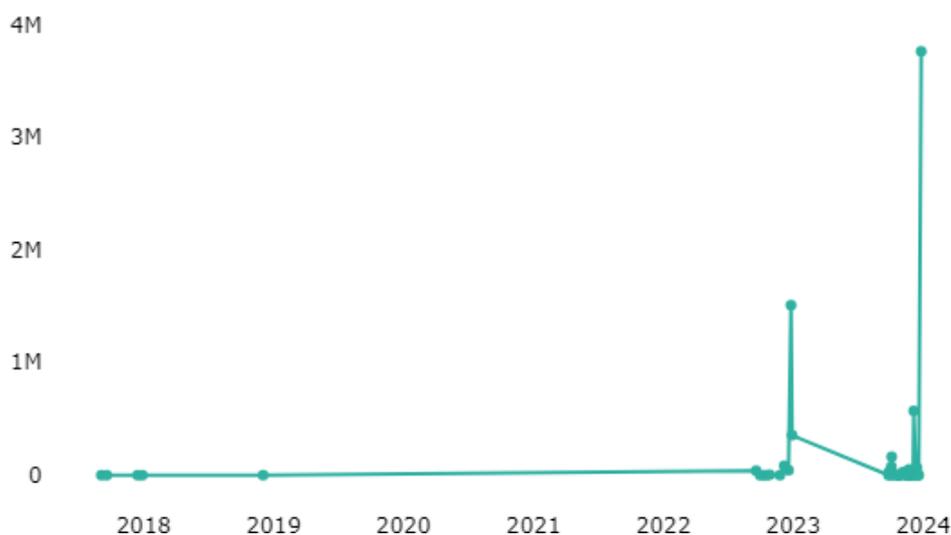
Fonte: Próprio autor (2024)

#### 4.7.1.3.6 Somas de Despesas Por Data de Empenho

O gráfico de linhas apresenta a soma dos valores pagos ao longo do tempo, com base nas datas de empenho fornecidas. Ele destaca as despesas associadas ao ano passado por parâmetro no filtro, mas que podem incluir datas de empenho de anos anteriores ou do ano atual, proporcionando uma visualização da distribuição dos gastos ao longo do tempo. Para criar esta representação, um novo conjunto de dados foi criado através de uma operação *'groupby'* na coluna *'DATA\_EMPENHO'*, agregando a soma dos valores pagos pela coluna *'VAPAG'* utilizando o método *'sum'*. No eixo x, estão os períodos de datas retirados da coluna *'DATA\_EMPENHO'*, enquanto no eixo y, estão as somas dos valores pagos, obtidas através do método *'sum'* aplicado à coluna *'VAPAG'*.

Figura 29: Gráfico obtido a partir da análise de somas de despesas por data de empenho

## Somadas de Despesas por Data de Empenho



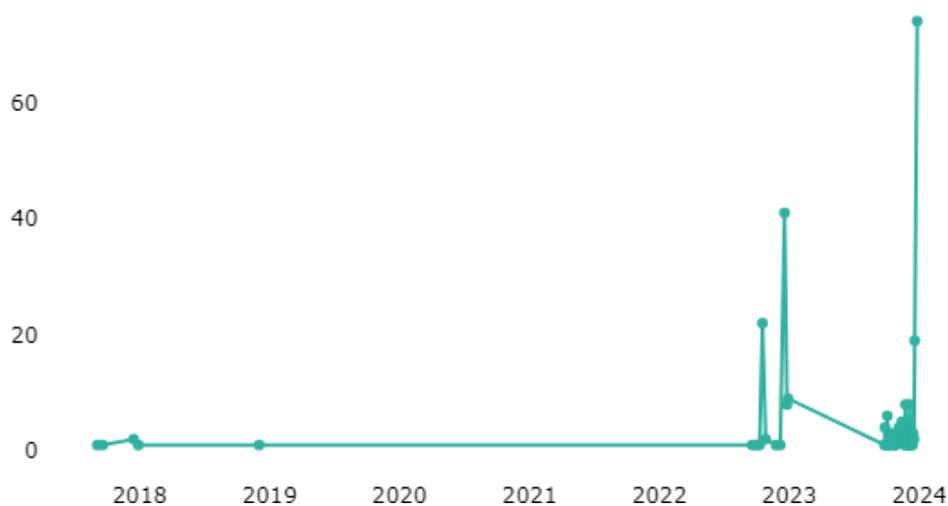
Fonte: Próprio autor (2024)

#### 4.7.1.3.7 Contagem de Despesas Registradas por Data de Empenho

O gráfico de linhas apresenta a contagem da frequência dos valores pagos ao longo do tempo, com base nas datas de empenho fornecidas. Ele destaca as despesas associadas ao ano passado por parâmetro no filtro, mas que podem incluir datas de empenho de anos anteriores ou do ano atual, proporcionando uma visualização da contagem dos gastos ao longo do tempo. Para criar esta representação, um novo conjunto de dados foi criado através de uma operação 'groupby' na coluna 'DATA\_EMPENHO', agregando a contagem dos valores pagos utilizando o método 'size' do Pandas. No eixo x, estão os períodos de datas retirados da coluna 'DATA\_EMPENHO', enquanto no eixo y, estão as quantidades dos valores registrados, obtidas através do método 'size'.

Figura 30: Gráfico obtido a partir da análise de contagem de despesas registradas por data de empenho

Contagem de Despesas Registradas por Data de Empenho



Fonte: Próprio autor (2024)

## 4.7.2 Componentes da dashboard de licitação

### 4.7.2.1 Filtros

#### 4.7.2.1.1 Filtro de ano

Este filtro permite que os usuários selecionem um ano específico para análise dos dados, servindo como ponto de início para a exploração dos dados.

Figura 31: Filtro com ano preenchido

A imagem mostra um campo de filtro com o rótulo "Ano" no canto superior esquerdo. O campo de entrada contém o texto "2023" e possui um ícone de "x" e uma seta para baixo no canto superior direito, indicando que é um campo de seleção ou de busca com opção de limpeza.

Fonte: Próprio autor (2024)

#### 4.7.2.2 Cards

##### 4.7.2.2.1 Card Valor Total Adjudicado

A informação apresentada no *card* "Valor Total Adjudicado" origina-se da coluna específica denominada 'TOTAL\_ADJUDICADO\_LICITACAO' dentro do conjunto de dados utilizado na *Dashboard* de Licitação. De acordo com o dicionário de dados, a coluna 'TOTAL\_ADJUDICADO\_LICITACAO' representa os valores individuais pagos em cada registro de licitação. Para calcular o valor total pago, que é a soma de todos os valores individuais registrados na coluna 'TOTAL\_ADJUDICADO\_LICITACAO', foi aplicado um método 'sum' nos valores dessa coluna, para calcular a soma total desses valores, resultando no valor total pago.

Figura 32: Card Valor Total Adjudicado preenchido



Fonte: Próprio autor (2024)

##### 4.7.2.2.2 Card QTD Geral de Licitações

A informação apresentada no *card* "QTD Geral de Licitações" origina-se do cálculo do número de linhas no DataFrame filtrado, onde cada linha representa uma licitação registrada que satisfaz os critérios de filtragem baseado no ano. Portanto, o valor apresentado reflete o

número total de licitações para o ano selecionado, oferecendo uma métrica quantitativa direta das licitações ocorridas.

Figura 33: *Card* QTD Geral de Licitações preenchido



Fonte: Próprio autor (2024)

#### 4.7.2.2.3 Card QTD Geral de Licitações Concluídas

A informação apresentada no *card* “QTD Geral de Licitações Concluídas” origina-se do cálculo do número de linhas no *DataFrame* filtrado, onde cada linha representa uma licitação registrada que satisfaz os critérios de filtragem baseado no ano e no filtro aplicado na base de dados para obtenção de todas as licitações onde a coluna ‘NOME\_SITUACAO’ seja igual a ‘Concluído’. Portanto, o valor apresentado reflete o número total de licitações concluídas para o ano selecionado, oferecendo uma métrica quantitativa direta das licitações ocorridas de acordo com os filtros informados.

Figura 34: *Card* QTD Geral de Licitações Concluídas preenchido



Fonte: Próprio autor (2024)

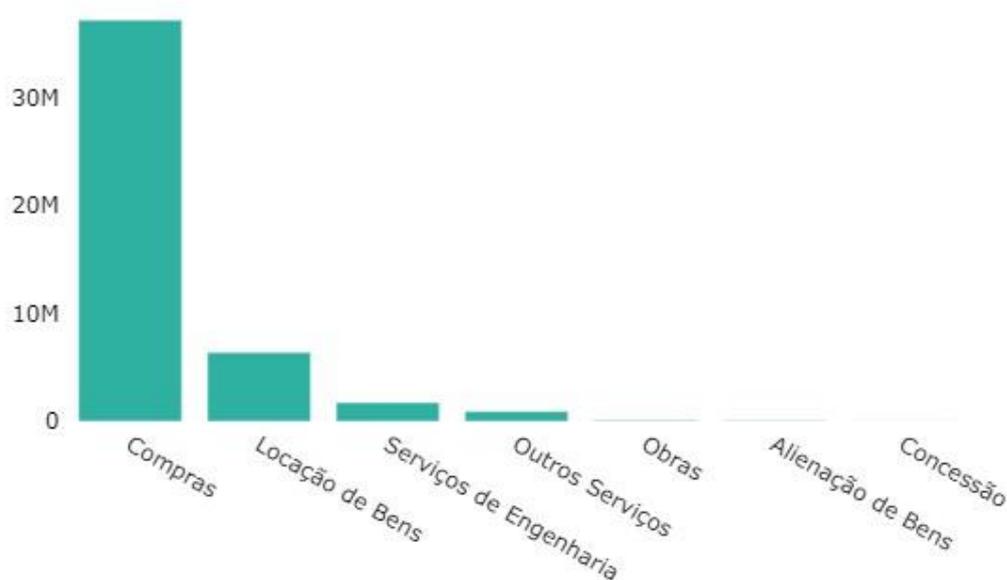
#### 4.7.2.3 Gráficos

##### 4.7.2.3.1 Soma do Valor Total Adjudicado por Natureza

Este gráfico de barras proporciona uma visualização em que as entidades da coluna 'NOME\_NATUREZA' estão dispostas no eixo x, enquanto no eixo y são apresentadas as somas correspondentes, exibidas em ordem decrescente. Cada coluna representa a soma dos valores pagos por natureza. Para criar esta representação, um novo conjunto de dados foi gerado usando o método '*groupby*', que agregou as somas dos valores pagos, formando grupos com base no campo da natureza.

Figura 35: Gráfico obtido a partir da análise de soma do valor total adjudicado por natureza

#### Soma do Valor Total Adjudicado por Natureza



Fonte: Próprio autor (2024)

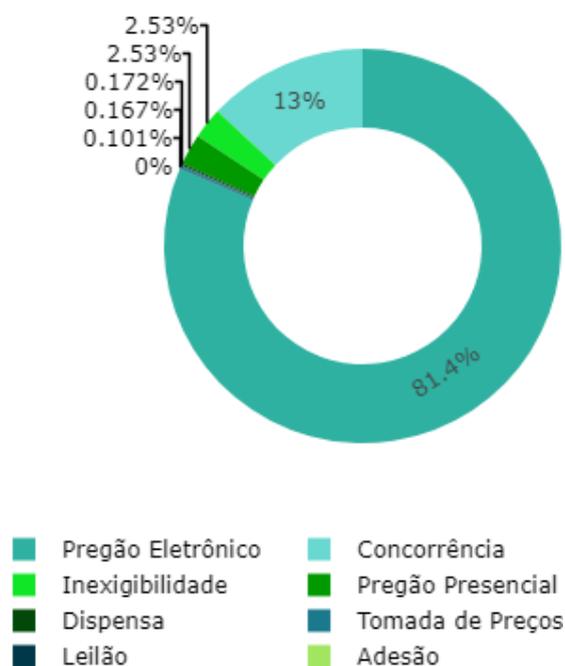
#### 4.7.2.3.2 Soma do Valor Adjudicado por Modalidade

Este gráfico de pizza foi gerado utilizando o método '*groupby*', que agregou a soma dos valores adjudicados com base nos grupos da coluna 'NOME\_MODALIDADE'. Na legenda do

gráfico, são listadas as entidades registradas em cada modalidade, enquanto as fatias da pizza representam a porcentagem correspondente de cada uma delas.

Figura 36: Gráfico obtido a partir da análise de soma do valor adjudicado por modalidade

#### Soma do Valor Adjudicado por Modalidade



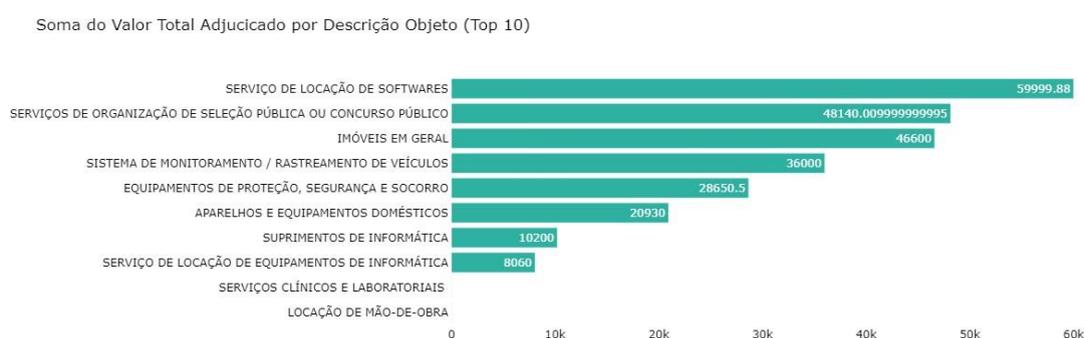
Fonte: Próprio autor (2024)

#### 4.7.2.3.3 Soma do Valor Total Adjudicado por Descrição Objeto (Top 10)

Gráfico de colunas, apresentado em formato horizontal, exibe informações sobre a soma dos valores adjudicados pelas dez descrições de objetos principais, juntamente com o número de ocorrências registradas para cada objeto, destacado ao lado de cada coluna. Para criar essa visualização, foram combinadas duas novas bases de dados filtradas. A primeira foi gerada usando o método '*value\_counts*' do *Pandas* para contar o número de registros por nome de objeto na coluna '*NOME\_DESCRICAOOBJETO*'. Já a segunda base de dados foi criada com o método '*groupby*', que agregou a soma dos valores pagos de acordo com o grupo da coluna

'NOME\_DESCRICAOOBJETO'. Em seguida, as duas bases de dados foram mescladas utilizando o método *'merge'*, unindo-as com base na coluna 'NOME\_DESCRICAOOBJETO'.

Figura 37: Gráfico obtido a partir da análise de soma do valor total adjudicado por descrição objeto (Top 10)

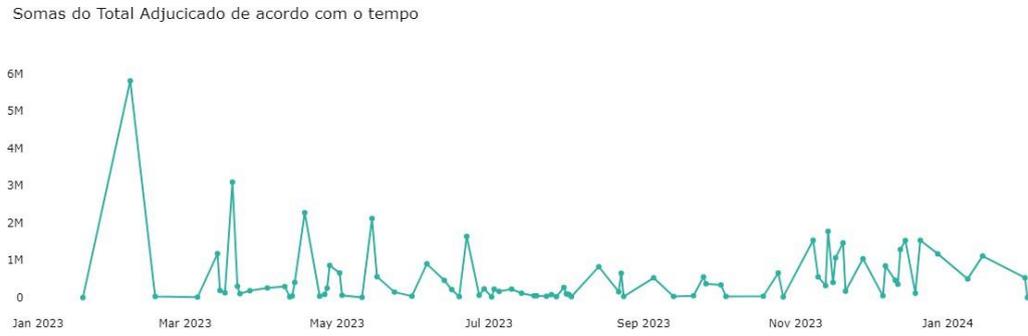


Fonte: Próprio autor (2024)

#### 4.7.2.3.4 Somas do Total Adjudicado de acordo com o tempo

Este gráfico de linhas apresenta a soma dos valores adjudicados ao longo das datas de homologação, conforme especificado nos parâmetros de filtragem por ano. No eixo x, estão dispostos os meses do ano em questão, enquanto no eixo y, são exibidos os totais das somas dos valores adjudicados. Essa representação proporciona uma visualização clara e informativa do padrão temporal dos valores adjudicados ao longo do ano filtrado.

Figura 38: Gráfico obtido a partir da análise de somas do total adjudicado de acordo com o tempo



Fonte: Próprio autor (2024)

## 5. Conclusões e Trabalhos Futuros

### 5.1 Desafios Enfrentados e Restrições da Aplicação

Em relação aos desafios enfrentados, o projeto foi criado acessando as apis fornecidas pela prefeitura municipal da cidade de Belo Jardim, onde seguindo os dicionários dos dados de Despesas Públicas, fornecida através do link <https://belojardim.pe.gov.br/transparencia/dados-abertos/api/despesas>, temos a seguinte informação disponibilizada:

Figura 39: Informação disponibilizada acerca dos campos retornados como resposta da api de despesas públicas

VALOR_EMPENHADO	Decimal Decimal (20,2)	Valor Empenhado
VALOR_PAGO	Decimal Decimal (20,2)	Valor Pago
VALOR_LIQUIDADO	Decimal Decimal (20,2)	Valor Liquidado

Fonte: <https://belojardim.pe.gov.br/transparencia/dados-abertos/api/despesas> (2024)

Porém, os dados disponibilizados via API sofreram uma alteração, onde tais campos passaram a vir como segue imagem abaixo, na consulta de exemplo: <https://belojardim.pe.gov.br/transparencia/api/despesas?limit=1000&ano=2023&page=1> acessado dia 05/03/2024

Figura 40: Informação retornada como resposta da api de despesas públicas

```
"VALOR_GLOBAL": "2150.00",
"VAPAG": "2150.00",
"LIQUIDADO": "2150.00",
```

Fonte:

<https://belojardim.pe.gov.br/transparencia/api/despesas?limit=1000&ano=2023&page=1>  
(2024)

Dessa forma, ocorreu a modificação dos campos "VALOR\_EMPENHADO" para "VALOR\_GLOBAL", "VALOR\_PAGO" para "VAPAG" e "VALOR\_LIQUIDADO" para "LIQUIDADO". Até a presente data da redação deste artigo, essa discrepância persiste no

dicionário de dados, exigindo a refatoração do projeto, originalmente desenvolvido com base nos campos "VALOR\_EMPENHADO", "VALOR\_PAGO" e "VALOR\_LIQUIDADO".

Um outro desafio enfrentado foi o aumento no tempo de resposta das requisições feitas à API de despesas públicas. Inicialmente, estávamos recebendo respostas em apenas 2 segundos, porém, posteriormente, esse tempo aumentou significativamente, chegando a demorar vários minutos. Devido à limitação de retorno da API, o sistema DataHarbor realiza solicitações de dados em lotes de 1000 registros por vez, para utilizar esses dados na construção dos gráficos das *dashboards*. Contudo, devido ao aumento no tempo de resposta, as imagens apresentadas neste projeto, referentes à *dashboard* de despesas públicas, foram baseadas nos primeiros 1000 registros solicitados no ano de 2023, conforme a requisição a seguir.

<https://belojardim.pe.gov.br/transparencia/api/despesas?limit=1000&ano=2023&page=1> .

Em relação às restrições do projeto, podemos citar as seguintes:

1. Acesso limitado: Um projeto *web* sendo executado localmente só pode ser acessado em dispositivos conectados à mesma rede local. Isso significa que ele não está disponível para acesso público na internet.
2. Disponibilidade Limitada: Projetos *web* locais podem não estar disponíveis 24/7, pois estão sujeitos ao tempo de atividade da máquina local. Se a máquina for desligada ou reiniciada, o projeto *web* também ficará inacessível temporariamente.
3. Limites de Armazenamento: Por utilizar um gerenciador de banco de dados gratuito, geralmente existem limites de armazenamento em termos de quantidade de dados que podem ser armazenados, implicando numa quantidade restrita de usuários cadastrados.
4. Quebra de Compatibilidade: Alterações na API podem resultar na quebra de compatibilidade com versões anteriores. Isso significa que como o sistema depende da API disponibilizada pelo governo, pode deixar de funcionar corretamente após a atualização, exigindo modificações para se adaptar às novas alterações.
5. Interrupções no Serviço: Durante o processo de atualização ou implantação de alterações na API, pode ocorrer interrupções no serviço, resultando em tempo de inatividade para os consumidores da API. Isso pode afetar a disponibilidade e o desempenho da aplicação por conta dessa dependência.

## 5.2 Conclusões

No decorrer deste trabalho, abordamos as dificuldades enfrentadas ao analisar dados públicos abertos e quais ganhos podem ser alcançados, provenientes das oportunidades e benefícios com a análise dessas informações. Um ponto central discutido e abordado no artigo foi a importância da disponibilidade de uma plataforma online dedicada à visualização e análise desse conjunto de informações, pois uma plataforma desse tipo oferece recursos e ferramentas visuais que permitem ao usuário uma experiência mais acessível, intuitiva e eficiente e uma maior participação sobre o que acontece no ambiente ao seu redor, de acordo com informações reais disponibilizadas.

Apresentamos a solução do sistema DataHarbor, uma solução *web* com uma interface amigável e intuitiva, que fornecem funcionalidades de visualizações de dados, permitindo que pesquisadores, analistas, e qualquer interessado possa explorar e interpretar as informações disponibilizadas da forma mais eficaz, facilitando identificação de padrões e tendências, fomentando o pensamento crítico e analítico da população em geral, contribuindo assim para o avanço e promoção da transparência pública e participação cívica.

## 5.3 Trabalhos Futuros

Com a conclusão deste trabalho e com a finalização da aplicação DataHarbor, foram identificadas oportunidades futuras, que são:

1. Desenvolvimento de ferramentas analíticas avançadas: Trazer ao sistema mais funcionalidades que possam fornecer análises mais avançadas e sofisticadas, lidando com inteligência artificial e aprendizado de máquina para identificação de padrões ou tendências ocultas nas bases de dados.
2. Melhoria de usabilidade da plataforma: Através da realização de estudos referentes a usabilidade da plataforma, com a coleta de *feedbacks* dos usuários, realizar um aprimoramento da experiência de navegação e usabilidade, implementando *interfaces* que sejam mais intuitivas e personalizáveis.
3. Integração de novas fontes de dados: Expandir a aplicação para que passe a coletar informações de novas fontes de dados públicos, abrangendo diferentes áreas de interesse, como dados referentes à saúde, dados ambientais, de segurança e demais informações do interesse público.

4. Aumento da exploração de interatividade nas visualizações: Permitir uma exploração mais profunda e envolvendo acerca dos dados, trazendo painéis de controle personalizáveis, permitindo mais interatividade nas explorações dos dados.
5. Expansão da plataforma para diferentes contextos municipais: Adaptação da aplicação para abranger informações de outras cidades, permitindo assim uma análise mais ampla, aumentando a utilidade e relevância da aplicação.

Diante das possibilidades e recursos apresentados neste trabalho é possível acreditar que a solução DataHarbor tenha o potencial de se tornar escalável, onde a expansão da plataforma possa ampliar sua utilidade, permitindo que governos e cidadãos locais tenham acesso a ferramentas e recursos para explorar e compreender os dados públicos de suas respectivas cidades e regiões, contribuindo para uma sociedade mais empenhada e informada.

## 6. Referências

ABREU, Janderson. **Dash com Plotly: A Arte da Personalização no Mundo dos Painéis Interativos.** Digital Innovation One, Set. 2023. Disponível em: <https://www.dio.me/articles/dash-com-plotly-a-arte-da-personalizacao-no-mundo-dos-paineis-interativos>. Acesso em: 02/04/2024.

ANDRADE, R. G. de; RAUPP, F. M. Transparência do legislativo local à luz da Lei de Acesso à Informação: evidências empíricas a partir dos maiores municípios brasileiros. **Desenvolvimento em Questão**, [S. l.], v. 15, n. 41, p. 85–130, 2017. DOI: 10.21527/2237-6453.2017.41.85-130. Disponível em: <https://www.revistas.unijui.edu.br/index.php/desenvolvimentoemquestao/article/view/4213>. Acesso em: 02/04/2024.

AWARI. Python: Convertendo uma string de data em datetime. Como converter uma string de data em datetime em Python? A conversão de uma string de data em um objeto datetime em Python é essencial para realizar operações avançadas com datas e horários. Disponível em: <https://awari.com.br/python-convertendo-uma-string-de-data-em-datetime/>. **Ciência de Dados**, Nov. 2023. Acesso em: 02/04/2024.

BATISTA, Natércia A. et al. **Dados de Múltiplas Fontes da Web: coleta, integração e pré-processamento.** Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2020.

BERNARDI. Luiz Antônio. **Manual de empreendedorismo e Gestão: Fundamentos, Estratégias e Dinâmicas.** 2. Ed. São Paulo. Editora Atlas S.A, 2012.

BICO, F. C.; TRINDADE, L. N.; CARACCILO, R. A. R.; PAIVA JR, R. J. S.; PERES, S. M. Legibilidade em Dados Abertos: uma Experiência com os Dados da Câmara Municipal de São Paulo. *In: VIII Simpósio Brasileiro de Sistemas de Informação, 2012, São Paulo. Anais do VIII Simpósio Brasileiro de Sistemas de Informação: Trilhas Técnicas*, 2012. v. 1. p. 678-689.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011.** Diário Oficial da União, Brasília, DF, 18 nov. 2011. Disponível em:

[https://www.transparenciacultura.sp.gov.br/wp-content/uploads/2018/12/Lei\\_de\\_acesso\\_a\\_informacao\\_n\\_12.527\\_2011.pdf](https://www.transparenciacultura.sp.gov.br/wp-content/uploads/2018/12/Lei_de_acesso_a_informacao_n_12.527_2011.pdf). Acesso em: 28/02/2024.

C., Guilherme. **Business Intelligence na Análise de Clientes e Clusterização**: Explorando o Potencial dos Dados. 10 fev. 2023. Disponível em: <https://www.linkedin.com/pulse/business-intelligence-na-an%C3%A1lise-de-clientes-e-o-dos-dados-campanez/>. Acesso em: 04/03/2024.

COSTA, L., PILATTI, L. A., & COELHO JR, T. **Gerenciamento de Processos de Negócios**: Uma visão holística. Simpósio de Engenharia de Produção, XIII. Disponível em: [https://www.simpep.feb.unesp.br/anais/anais\\_13/artigos/1073.pdf](https://www.simpep.feb.unesp.br/anais/anais_13/artigos/1073.pdf). Acesso em: 06/03/2024.

CURTY, Renata Gonçalves; CERVANTES, Brígida Maria Nogueira. Data science: ciência orientada a dados. **Informação & Informação**, v. 21, n. 2, p. 1-4, 2016.

DESSOLDI, Flávia. Método SCRUM: Um resumo de tudo o que você precisa saber. **Medium**, Jun. 2019. Disponível em: <https://medium.com/reprogramabr/scrum-um-breve-resumo-f051e1bc06d9>. Acesso em: 02/04/2024.

DINIZ, Vagner. **Como conseguir dados governamentais abertos**. In: CONGRESSO CONSAD DE GESTÃO PÚBLICA, III, Brasília, 2010.

ENAP. **Elaboração de Plano de Dados Abertos**. Disponível em: <http://repositorio.enap.gov.br/handle/1/3152>. Acesso em: 02/03/2024.

GONZALO, Javier Galante. **Flask vs Django**. Revelo Community, Mar. 2023. Disponível em: <https://community.revelo.com.br/flask-vs-django/#:~:text=Vantagens%20do%20Flask&text=%C3%89%20muito%20%C3%A1cil%20de%20usar,os%20quais%20pode%20ser%20utilizado>. Acesso em: 02/04/2024.

Google Cloud. **O que é ETL?**. Disponível em: <https://cloud.google.com/learn/what-is-etl?hl=pt-br>. Acesso em: 10/05/2024.

GRUMAN, M. Lei de acesso à informação: Notas e um breve exemplo. **Revista Debates**, 6(3), 97. Disponível em: <https://doi.org/10.22456/1982-5269.34229>. Acesso em: 10/03/2024.

GRUS, Joel. **Data Science do Zero**. Alta Books, 2016.

GURSTEIN, Michael. Open data: Empowering the empowered or effective data use for everyone? **First Monday**, Volume 16, Number 2, 7 February 2011. Disponível em: <https://firstmonday.org/ojs/index.php/fm/article/download/3316/2764>. Acesso em: 10/03/2024.

IBM. **Gráficos de barras**. Disponível em: <https://www.ibm.com/docs/pt-br/spss-modeler/18.4.0?topic=types-bar-charts>. Acesso em: 02/04/2024.

IBM. **Gráficos de linhas**. Disponível em: <https://www.ibm.com/docs/pt-br/spss-modeler/18.4.0?topic=types-line-charts>. Acesso em: 02/04/2024.

IGNÁCIO, S. A. Importância da Estatística para o Processo de Conhecimento e Tomada de Decisão. **Revista Paranaense De Desenvolvimento - RPD**, (118), 175–192. Disponível em: <https://ipardes.emnuvens.com.br/revistaparanaense/article/view/89>. Acesso em: 05/04/2024.

INPI - Instituto Nacional da Propriedade Industrial. **Manual dos Dados Abertos - Governo Federal**. Disponível em: <https://www.gov.br/inpi/pt-br/acesso-a-informacao/dados-abertos/arquivos/documentos/diversos/ManualdosDadosAbertosGoverno.pdf/view>. Acesso em: 02/03/2024.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados abertos conectados**: em busca da web do conhecimento. Novatec Editora, 2015.

MAGALHAES, G.; ROSEIRA, C.; STROVER, S. **Open government data intermediaries**: a terminology framework. *In*: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance. ACM, 2013. p. 330-333.

MCKINNEY, Wes et al. pandas: a foundational Python library for data analysis and statistics. **Python for high performance and scientific computing**, v. 14, n. 9, p. 1-9, 2011.

MCKINNEY, Wes. **Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython**. Novatec Editora, 2018.

NASCIMENTO, H.; FERREIRA, C. Uma introdução à visualização de informações. **VISUALIDADES**, Goiânia v.9 n.2 p. 13-43, jul/dez.2011.

OECD (2013), "**Open government data**", in Government at a Glance 2013, OECD Publishing, Paris, 2013. Disponível em: [https://doi.org/10.1787/gov\\_glance-2013-48-en](https://doi.org/10.1787/gov_glance-2013-48-en). Acesso em: 16/03/2024.

O'REILLY, T. Government as a Platform. **Innovations: Technology, Governance, Globalization**, v. 6, n. 1, p. 13–40. 2011.

Oracle. **O que é o MySQL?**. Disponível em: <https://www.oracle.com/br/mysql/what-is-mysql/>. Acesso em: 02/04/2024.

**Portal da Transparência**. Disponível em: <https://portaldatransparencia.gov.br/>. Acesso em: 10/03/2024.

PRATA, Ramon. **React: por que considerá-lo no seu projeto?**. DTI Digital, Jun. 2019. Disponível em: <https://www.dtidigital.com.br/blog/considere-react-no-seu-projeto>. Acesso em: 02/04/2024.

ROZZA, Giovanni. **Organizando um banco de dados usando as formas normais**. 04 jul. 2023, 20:26. Disponível em: <https://www.dio.me/articles/organizando-um-banco-de-dados-usando-as-formas-normais>. Acesso em: 04/03/2024.

THORSBY, J., STOWERS, G. N. L., WOLSLEGEL, K., & TUMBUAN, E. (2017). Understanding the content and features of open data portals in American cities. Jan, 2017. **Government Information Quarterly**, 34(1), 53–61.

VAZ, Ana. **Segurança da Informação, Proteção da Privacidade e dos Dados Pessoais**. Instituto da Defesa Nacional, 2007. Disponível em: <http://hdl.handle.net/10400.26/1218>. Acesso em: 08/03/2024.