



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**VICTOR ARRUDA CÂMARA VIRGOLINO**

**UrnaLogs:  
Facilitando a Análise de Logs de Urnas Eletrônicas para  
Promover Transparência no Processo Eleitoral**

**CAMPINA GRANDE - PB**

**2023**

**VICTOR ARRUDA CÂMARA VIRGOLINO**

**UrnaLogs:**

**Facilitando a Análise de Logs de Urnas Eletrônicas  
para Promover Transparência no Processo Eleitoral**

**Trabalho de Conclusão Curso apresentado  
ao Curso Bacharelado em Ciência da  
Computação do Centro de Engenharia  
Elétrica e Informática da Universidade  
Federal de Campina Grande, como requisito  
parcial para obtenção do título de Bacharel  
em Ciência da Computação.**

**Orientador : MAXWELL GUIMARÃES DE OLIVEIRA**

**CAMPINA GRANDE - PB**

**2023**

**VICTOR ARRUDA CÂMARA VIRGOLINO**

**UrnaLogs:**

**Facilitando a Análise de Logs de Urnas Eletrônicas  
para Promover Transparência no Processo Eleitoral**

Trabalho de Conclusão Curso apresentado ao Curso Bacharelado em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

**BANCA EXAMINADORA:**

**MAXWELL GUIMARÃES DE OLIVEIRA**  
Orientador – UASC/CEEI/UFCG

**ROBERTO MEDEIROS DE FARIA**  
Examinador – UASC/CEEI/UFCG

**Francisco Vilar Brasileiro**  
Professor da Disciplina TCC – UASC/CEEI/UFCG

**Trabalho aprovado em: 29 de JUNHO de 2023.**

**CAMPINA GRANDE - PB**

## RESUMO

O processo de votação eletrônica no Brasil, utilizando urnas eletrônicas e reconhecimento biométrico, trouxe facilidades e agilidade para as eleições. No entanto, problemas técnicos e a falta de transparência têm gerado desconfiança e questionamentos sobre a contabilização correta dos votos. Nesse contexto, os logs das urnas eletrônicas se tornam uma fonte valiosa de dados, registrando todas as operações realizadas pelos eleitores e mesários. No entanto, a grande parcela da população desconhece a existência desses logs e como acessá-los. Este trabalho visa tornar mais acessível a obtenção e análise dos dados armazenados nos logs das urnas das eleições, disponibilizados no site do Tribunal Superior Eleitoral (TSE). A ferramenta possui 4 módulos: o primeiro é responsável por realizar a coleta de todos os logs das urnas da eleição através do uso de Web Scraping; o segundo efetua o processo de Extract, Transform, Load (ETL) para tornar os dados estruturados; o terceiro é responsável por armazenar os dados; e, por fim, o quarto disponibiliza um dashboard no qual é possível visualizar e filtrar os dados de cada seção com o suporte de ferramentas visuais que facilitem a interpretação.

# **UrnaLogs: Facilitating the Analysis of Electronic Ballot Box Logs to Promote Transparency in the Electoral System**

## **ABSTRACT**

The electronic voting process in Brazil, using electronic ballot boxes and biometric recognition, brought facilities and agility to the elections. However, technical problems and lack of transparency have generated distrust and questions about the correct counting of votes. In this context, electronic voting machine logs become a valuable source of data, recording all operations carried out by voters and polling station officials. However, a large portion of the population is unaware of the existence of these logs and how to access them. This work aims to make it more accessible to obtain and analyze the data stored in the ballot box logs, available on the Superior Electoral Court (TSE) website. The tool has 4 modules: the first is responsible for collecting all logs from the election ballot boxes through the use of Web Scraping; the second performs the Extract, Transform, Load (ETL) process to make the data structured; the third is responsible for storing the data; and, finally, the fourth provides a dashboard in which it is possible to visualize and filter the data of each section with the support of visual tools that facilitate the interpretation.

# UrnaLogs:

## Facilitando a Análise de Logs de Urnas Eletrônicas para Promover Transparência no Processo Eleitoral

Victor Arruda Câmara Virgolino

victor.virgolino@ccc.ufcg.edu.br

Universidade Federal de  
Campina Grande

Maxwell Guimarães de Oliveira

maxwell@computacao.ufcg.edu.br

Universidade Federal de  
Campina Grande

### RESUMO

O processo de votação eletrônica no Brasil, utilizando urnas eletrônicas e reconhecimento biométrico, trouxe facilidades e agilidade para as eleições. No entanto, problemas técnicos e a falta de transparência têm gerado desconfiança e questionamentos sobre a contabilização correta dos votos. Nesse contexto, os logs das urnas eletrônicas se tornam uma fonte valiosa de dados, registrando todas as operações realizadas pelos eleitores e mesários. No entanto, a grande parcela da população desconhece a existência desses logs e como acessá-los. Este trabalho visa tornar mais acessível a obtenção e análise dos dados armazenados nos logs das urnas das eleições, disponibilizados no site do Tribunal Superior Eleitoral (TSE). A ferramenta possui 4 módulos: o primeiro é responsável por realizar a coleta de todos os logs das urnas da eleição através do uso de Web Scraping; o segundo efetua o processo de Extract, Transform, Load (ETL) para tornar os dados estruturados; o terceiro é responsável por armazenar os dados; e, por fim, o quarto disponibiliza um dashboard no qual é possível visualizar e filtrar os dados de cada seção com o suporte de ferramentas visuais que facilitem a interpretação.

### PALAVRAS-CHAVE

UrnaLogs, urnas eletrônicas, logs, transparência, processo eleitoral

### 1. INTRODUÇÃO

A participação ativa do cidadão nas eleições é um dos pilares fundamentais da democracia, permitindo a escolha de representantes políticos que defenderão seus interesses nos âmbitos municipal, estadual e nacional. No Brasil, apesar dos avanços tecnológicos aplicados ao processo de votação, como o uso de urnas eletrônicas e reconhecimento biométrico, ainda existem desafios que impactam a eficiência e a confiabilidade do sistema eleitoral. Um problema recorrente, especialmente durante o primeiro turno, é a ocorrência de atrasos e problemas técnicos que resultam em longas filas e extensos períodos de espera, comprometendo a experiência do eleitor [1].

No entanto, as críticas e ataques frequentes direcionados à tecnologia das urnas eletrônicas, como mencionado por Ferrão et al. [2], têm contribuído para a desconfiança da população em relação ao processo eletrônico de votação. Essa desconfiança é

alimentada pela falta de transparência e pela ausência de mecanismos de auditoria efetivos. Muitas pessoas sugerem a evolução do sistema eletrônico ou até mesmo o retorno ao modelo tradicional, baseado exclusivamente em cédulas de papel, como forma de solucionar esses problemas.

Assim sendo, os logs das urnas eletrônicas emergem como uma fonte valiosa de dados que, embora públicos, são amplamente desconhecidos pela maioria da população. Esses logs registram todas as operações realizadas pelos eleitores e mesários, incluindo data e hora, desde os testes realizados dias antes das eleições para verificação e a adequação das urnas até o encerramento da votação e a impressão dos Boletins de Urna (BU). No presente trabalho, concentramos nosso estudo nos logs das urnas utilizadas no primeiro e segundo turno das eleições de 2022 no estado da Paraíba, totalizando 223 cidades, 68 zonas eleitorais, 9.603 seções e mais de 145 milhões de registros.

Com o intuito de solucionar os desafios mencionados acima, foi desenvolvida a ferramenta UrnaLogs, que visa facilitar a obtenção e o estudo dos dados contidos nos logs das urnas eletrônicas. A UrnaLogs propõe uma abordagem simples e automatizada para extrair os dados diretamente do site do TSE<sup>1</sup>, apresentando-os de forma estruturada em tabelas e fornecendo uma interface de visualização e análise através de um dashboard interativo.

A ferramenta UrnaLogs está disponível de forma gratuita no repositório GitHub<sup>2</sup>, onde é possível encontrar o passo a passo para utilização, bem como demonstrações das funcionalidades. Além disso, uma playlist de vídeos exemplificando o uso da ferramenta também está disponível no YouTube<sup>3</sup>.

O restante deste trabalho está estruturado da seguinte maneira: na Seção 2, realizamos a fundamentação teórica de conceitos importantes. Na seção 3 apresentamos a ferramenta em linhas gerais. Na Seção 4, apresentamos a arquitetura da ferramenta. Na Seção 5, demonstramos casos de uso da ferramenta. Na Seção 6 discutimos sobre a experiência de desenvolvimento e desafios

<sup>1</sup> Site do Tribunal Superior Eleitoral: <https://resultados.tse.jus.br>

<sup>2</sup> GitHub da Ferramenta Desenvolvida: <https://github.com/VictorVirgolino/UrnaLogs>

<sup>3</sup> Vídeos demonstrativos da ferramenta: <https://www.youtube.com/playlist?list=PLsIRppJ2pVH7lqM-Zc15VffsEIX9TrJm>

encontrados na construção da ferramenta. E, por fim, concluímos com uma discussão sobre as limitações da ferramenta e as possíveis direções para trabalhos futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, estão apresentados os conceitos de Web Scraping, ETL, SGBD, Dados Relacionais, ORM e Dashboard, que são essenciais para uma compreensão adequada deste trabalho.

### 2.1 Web Scraping

Segundo Mitchell [3], Web scraping é a prática de coletar dados da internet de forma automatizada, utilizando web crawlers ou bots para consultar servidores web, solicitar dados em formato HyperText Markup Language (HTML) e outros arquivos que compõem páginas web, e extrair as informações necessárias a partir desses dados. Essa técnica é útil quando não há uma Application Programming Interface (API) disponível para os dados desejados ou quando se deseja coletar informações de várias fontes diferentes ao mesmo tempo.

### 2.2 ETL

O ETL [9] é um processo utilizado na área de gerenciamento de dados para extrair dados de diversas fontes, transformá-los em um formato adequado e carregá-los em um destino específico, como um banco de dados ou um data warehouse.

O processo de ETL abrange três etapas principais: extração, transformação e carga. Na etapa de extração, os dados são coletados e extraídos de várias fontes, que podem incluir bancos de dados, sistemas legados, arquivos de texto, planilhas, páginas web, entre outros. A extração pode envolver a leitura direta dos dados de suas fontes originais ou a obtenção dos dados por meio de interfaces, APIs, serviços da web ou Web Scraping.

Após a extração dos dados, eles passam por um processo de transformação, onde são limpos, filtrados, padronizados, enriquecidos e integrados. Essas etapas visam garantir a consistência e a qualidade dos dados, bem como adequá-los ao formato desejado. Por fim, na etapa de carga, os dados transformados são carregados no destino final, que pode ser um banco de dados relacional, um data warehouse ou qualquer outro sistema de armazenamento apropriado. A carga geralmente envolve a definição de esquemas, tabelas e estruturas de dados no destino, além da execução de processos para inserir, atualizar ou deletar os dados existentes.

### 2.3 SGBD e Banco de Dados Relacional

O Sistema de Gerenciamento de Banco de Dados (SGBD) é um sistema de software responsável por gerenciar um ou mais banco de dados. Ele atua como uma interface que permite aos usuários realizar operações como inserção, manipulação e consulta de dados previamente armazenados.

O Banco de Dados Relacional [10] é um tipo de banco de dados que armazena e fornece acesso de dados relacionados entre si. Eles são baseados no modelo relacional, uma maneira direta e intuitiva de representar os dados na forma de tabelas e que utiliza Structured Query Language (SQL) para interagir com os dados. Uma característica essencial desse tipo de banco de dados é sua conformidade com as propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade) em suas transações,

garantindo processamento e armazenamento seguros e eficientes dos dados.

### 2.4 ORM

ORM (Object Relational Mapping) [7] ou Mapeamento Objeto Relacional é uma técnica de programação que conecta os objetos de uma aplicação com as tabelas de um banco de dados relacional. Essa conexão é estabelecida por meio de um mapeador objeto-relacional, que normalmente é uma biblioteca ou framework especializado.

Ao utilizar um ORM, a manipulação e consulta dos dados armazenados no banco de dados são simplificadas. O ORM abstrai a necessidade de escrever consultas SQL manualmente, o que resulta em maior segurança, redução do tempo de desenvolvimento e menos linhas de código necessárias em comparação com o uso direto de SQL.

### 2.5 Dashboard

Dashboard [11] é uma interface visual que reúne e apresenta informações relevantes de maneira clara e concisa. Ele é projetado para fornecer uma visão panorâmica de dados, métricas e indicadores essenciais para realização de análises e para auxiliar na tomada de decisões. O Dashboard pode combinar gráficos, tabelas, indicadores-chave de desempenho (KPIs) e outros elementos visuais para facilitar a compreensão e interpretação dos dados.

Além disso, dashboards geralmente são interativos, permitindo que os usuários explorem os dados de diferentes maneiras, apliquem filtros, realizem análises mais detalhadas e obtenham insights adicionais. Isso facilita a exploração dos dados e o aprofundamento nas informações relevantes.

## 3. URNALOGS

O UrnaLogs é uma ferramenta desenvolvida em Python que oferece uma solução completa para extrair, estruturar e visualizar os dados contidos nos logs das urnas eletrônicas disponibilizados pelo TSE.

Com o UrnaLogs, é possível identificar padrões, tendências e eventuais anomalias nos registros, facilitando a detecção de problemas ou falhas durante o processo de votação. Ademais, a ferramenta oferece recursos de filtragem, busca e visualização dos logs de forma clara e intuitiva, permitindo que estudiosos, cidadãos e demais interessados possam analisar uma variedade de informações relacionadas às seções eleitorais de suas cidades.

O UrnaLogs é uma solução poderosa para garantir a transparência, confiabilidade e integridade do sistema de votação eletrônica, contribuindo para a segurança e eficiência dos processos democráticos.

## 4. ARQUITETURA DA FERRAMENTA

A arquitetura da ferramenta é dividida em 4 módulos conforme ilustrado no fluxograma da Figura 1.



Figura 1 : Arquitetura do sistema

#### 4.1 Módulo de Web Scraping

O módulo de Web Scraping é responsável pela extração dos dados das eleições disponibilizados no site do TSE. Para isso, ele se utiliza da ferramenta Selenium WebDriver [4] para automatizar o controle de um navegador Chrome.

Após ser fornecido o site inicial da eleição escolhida, o módulo realiza automaticamente uma varredura de todas as seções, realizando o download dos arquivos contendo os dados de cada seção. Esses arquivos são organizados em ordem alfabética e seguindo uma estrutura hierárquica, composta por Estado, Município, Zona e Seção, conforme exemplificado na Figura 2.

```

Paraíba - PB
├── Aguiar
│   └── Zona 0066
│       └── Seção 0001
│           ├── DadosDeUrna_o00406-1903800660001_1666216668525.zip
│           ├── logd.dat
│           ├── o00406-1903800660001.bu
│           ├── o00406-1903800660001.imgbu
│           ├── o00406-1903800660001.logjez
│           ├── o00406-1903800660001.rdv
│           └── o00406-1903800660001.vscmr
│       └── Seção 0002
│           ├── DadosDeUrna_o00406-1903800660002_1666216673957.zip
│           ├── logd.dat
│           ├── o00406-1903800660002.bu
│           ├── o00406-1903800660002.imgbu
│           ├── o00406-1903800660002.logjez
│           ├── o00406-1903800660002.rdv
│           └── o00406-1903800660002.vscmr
  
```

Figura 2: Estrutura de sistema de arquivos

Cada seção possui um arquivo compactado (.zip) que contém os seguintes arquivos:

- O arquivo \*.bu que contém um arquivo binário referente aos dados do BU
- O arquivo \*.imgbu é um print da BU impressa na seção
- O arquivo \*.logjez é o arquivo criptografado que contém o arquivo com os logs das urna
- O arquivo \*.rdv corresponde ao registro digital do voto, onde está registrado de modo anônimo todos os votos registrados na urna
- O arquivo \*.vscmr contém informações criptografadas acerca do hardware urna.
- O arquivo logd.dat é o obtido da decriptografia do arquivo \*.logjez, é nele que está contido os logs da urna

Uma vez estruturados dessa forma, o módulo de Web Scraping finaliza e passa a execução para o módulo de ETL.

#### 4.2 Módulo de ETL

Após a conclusão da coleta de dados pelo módulo de Web Scraping, é acionado o módulo de ETL. Esse módulo percorre o sistema de arquivos descompactando e decriptografando os arquivos .logjez com o auxílio da biblioteca py7zr [5] para ter, desse modo, acesso ao arquivo logd.dat, que é um arquivo de texto o qual contém os logs da urna utilizada naquela seção. A Figura 3 apresenta um exemplo de trecho dos logs armazenados na urna durante o primeiro turno das eleições de 2022.



02/10/2022 09:36:27	INFO	2032321	VOTA	Aguardando digitação do título	544B12386D19D91E
02/10/2022 09:36:58	INFO	2032321	VOTA	Título digitado pelo mesário	9E5461CB316648D7
02/10/2022 09:37:01	INFO	2032321	VOTA	Solicita digital. Tentativa [1] de [4]	345F5FA9914F9974
02/10/2022 09:37:03	INFO	2032321	VOTA	Capturada a digital. Tentativa [1] de [4]	640BB17FE8700842
02/10/2022 09:37:05	INFO	2032321	VOTA	Dedo reconhecido e o score para habilitá-lo. Polegar direito - Score [128]	F940DCFD5F1362B7
02/10/2022 09:37:08	INFO	2032321	VOTA	Tipo de habilitação do eleitor [biométrica]	0AE404B990E40972
02/10/2022 09:37:08	INFO	2032321	VOTA	Eleitor foi habilitado	A6A8F6F94FC04910
02/10/2022 09:37:32	INFO	2032321	VOTA	Voto confirmado para [Deputado Federal]	2998A0DF1965CE6F
02/10/2022 09:37:42	INFO	2032321	VOTA	Voto confirmado para [Deputado Estadual]	EFF4DB61B01CFD4E
02/10/2022 09:37:51	INFO	2032321	VOTA	Voto confirmado para [Senador]	2E383737CD5DF48D
02/10/2022 09:38:00	INFO	2032321	VOTA	Voto confirmado para [Governador]	12D17740611E14D0
02/10/2022 09:38:04	INFO	2032321	VOTA	Voto confirmado para [Presidente]	EA2320052BF32F31
02/10/2022 09:38:05	INFO	2032321	VOTA	O voto do eleitor foi computado	51A0095120A99ECC
02/10/2022 09:38:06	INFO	2032321	VOTA	Aguardando digitação do título	7A3E8E113042450A
02/10/2022 09:38:50	INFO	2032321	VOTA	Título digitado pelo mesário	6F6B0C197252B1B6
02/10/2022 09:38:53	INFO	2032321	VOTA	Solicita digital. Tentativa [1] de [4]	69E01A41172AFAA2
02/10/2022 09:38:55	INFO	2032321	VOTA	Capturada a digital. Tentativa [1] de [4]	B0CF7EF0CCE526F5
02/10/2022 09:38:57	INFO	2032321	VOTA	Dedo reconhecido e o score para habilitá-lo. Polegar direito - Score [36]	C1036A43BA5B1E20
02/10/2022 09:39:00	INFO	2032321	VOTA	Tipo de habilitação do eleitor [biométrica]	38B64AAC704B1814
02/10/2022 09:39:00	INFO	2032321	VOTA	Eleitor foi habilitado	785F01C2AD9B69E5
02/10/2022 09:39:09	INFO	2032321	VOTA	Voto confirmado para [Deputado Federal]	58D7651CA3F140CA
02/10/2022 09:39:14	INFO	2032321	VOTA	Tecla indevida pressionada	31D6174A6F9BF4AA
02/10/2022 09:39:17	INFO	2032321	VOTA	Voto confirmado para [Deputado Estadual]	23DE184410EBE0B2
02/10/2022 09:39:23	INFO	2032321	VOTA	Voto confirmado para [Senador]	751188699C68A18E
02/10/2022 09:39:27	INFO	2032321	VOTA	Voto confirmado para [Governador]	2B283F33B72E930A
02/10/2022 09:39:30	INFO	2032321	VOTA	Voto confirmado para [Presidente]	3EE2F99E4CF9A86D
02/10/2022 09:39:31	INFO	2032321	VOTA	O voto do eleitor foi computado	68B458ABC6E5B158

Figura 3: Trecho dos Logs de uma Urna Eletrônica (versão 1)

É importante destacar que o log da urna é complexo e de difícil compreensão para os humanos, requerendo um tratamento especial para que os dados possam ser corretamente analisados. Nesse sentido, o módulo de ETL do UrnaLogs lê minuciosamente linha por linha do arquivo de log, extraindo os dados relevantes. Após o ajuste necessário, esses dados são inseridos no módulo de persistência utilizando a ferramenta ORM Peewee. Esse processo de extração e persistência dos dados permite que as informações sejam armazenadas de forma estruturada e segura, sendo acessíveis para consultas e análises posteriores.

### 4.3 Módulo de Persistência

O módulo de persistência do UrnaLogs é responsável por armazenar os dados processados após a conclusão do módulo de ETL. Utilizando a ORM Peewee, esse módulo oferece suporte a diversos bancos de dados relacionais, como Postgresql, MySQL e SQLite, entre outros. As instruções para configurar essa integração podem ser encontradas na documentação da ORM Peewee [6].

Ao estruturar o banco de dados, foi dado um foco especial aos relacionamentos entre as seções, zonas, municípios e estados, a fim de proporcionar uma análise abrangente dos dados dos logs. Essa estrutura permite explorar as informações a partir de diferentes perspectivas e facilita a compreensão das relações existentes entre os elementos.

O esquema do banco de dados segue o diagrama ER exibido na Figura 4. Neste diagrama, as tabelas estão organizadas da seguinte forma:

Na tabela “Eleição” são armazenados o ano e o turno das eleições cadastradas. Essa tabela está relacionada com a tabela “Estado”, que registra os estados envolvidos nas eleições. Cada Estado possui suas respectivas zonas e municípios. É importante ressaltar que uma zona pode abranger mais de um município, e um município pode conter mais de uma zona. Esse relacionamento é representado pela tabela “MunicípioZona”.

Na tabela “Seção”, são armazenados o nome, o ID da zona e o ID do município relacionados a cada seção específica. Essa estrutura permite rastrear e associar corretamente os dados de cada seção às respectivas zonas e municípios.

Por fim, a tabela “Logs” contém informações detalhadas sobre os logs das urnas eletrônicas. Ela inclui a data e hora da ocorrência do log, o tipo de ação registrada, o ID da urna, o modo de operação da urna, a descrição do log e uma hash. Cada registro de log está relacionado a uma única seção, possibilitando a análise precisa e individualizada dos eventos ocorridos em cada urna durante o processo eleitoral.

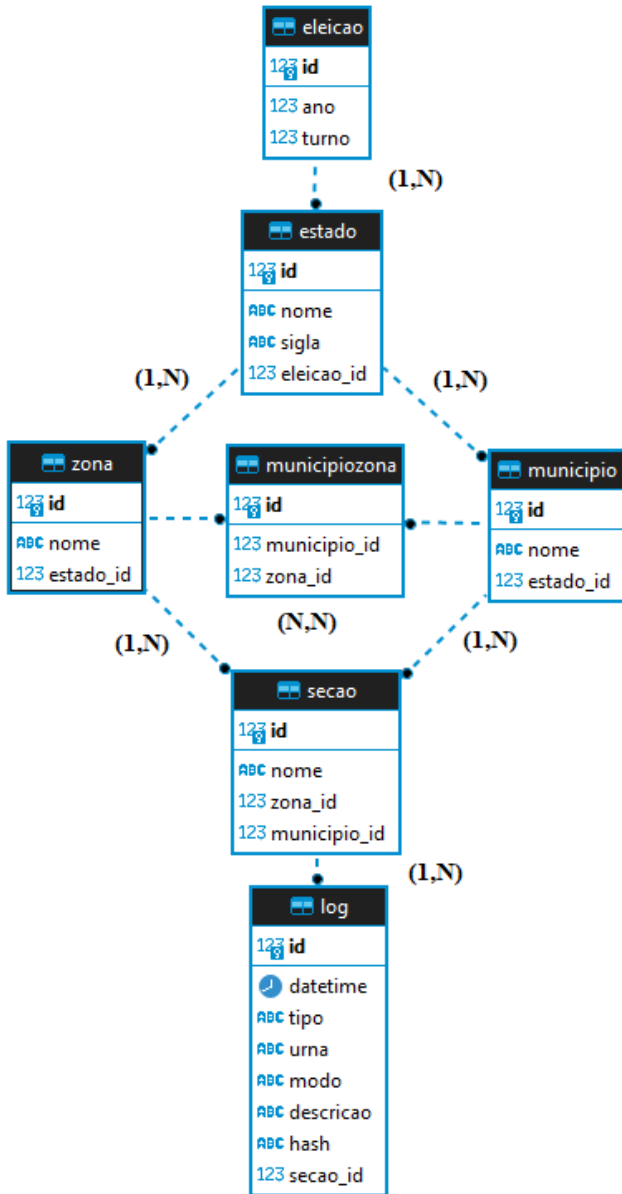


Figura 4: Modelo de dados.

#### 4.4 Dashboard

Após o término do processo de ETL e a inserção dos dados na camada de persistência, o módulo de dashboard disponibiliza uma forma fácil e interativa de explorar os dados. Tal módulo foi implementado utilizando um Python Notebook, responsável por realizar consultas no banco de dados e plotar gráficos dinâmicos com base nos filtros de entrada selecionados pelo usuário. As opções dos filtros de entrada estão representadas na Figura 5.

Eleição	Selecione a Eleição...	▼
Estado	Selecione o Estado...	▼
Município	Selecione o Município...	▼
Zona	Selecione a Zona...	▼
Seção	Selecione a Seção...	▼
Tipo	Histograma	▼
X	Tempo em Votação	▼
Y	Tempo na Urna	▼
Cor	#7eb0d5	<input type="color" value="#7eb0d5"/>
Paleta	Accent	▼

Figura 5: Entradas do dashboard

No painel de entradas, representado na Figura 5, é possível filtrar as informações da seção utilizando os campos de eleição, estado, município, zona e seção. Esses filtros permitem selecionar os dados específicos que serão visualizados no dashboard.

Os campos restantes são utilizados para configurar os gráficos. No campo "Tipo", é possível escolher entre três opções de gráfico: Histograma, Dispersão ou Calor. Essa escolha determina a forma como os dados serão representados visualmente.

Nos campos "X" e "Y", é possível selecionar os eixos dos gráficos. Dependendo da opção escolhida no campo "Tipo", esses eixos podem representar diferentes valores. Por exemplo, para os gráficos de Histograma, é possível escolher entre os valores de Votos por Hora, Porcentagem por Hora, Tempo em Votação, Tempo na Urna, Erros na Digital ou Erros na Votação.

Por fim, nos campos "Cor" e "Paleta", é possível personalizar a aparência dos gráficos. No campo "Cor", é possível selecionar a cor dos gráficos de Histograma e Dispersão, permitindo uma personalização visual mais detalhada. Já no campo "Paleta", é possível escolher a paleta de cores dos gráficos de Calor, adicionando mais variedade e distinção aos dados apresentados.

Essas configurações das entradas permitem ao usuário explorar e visualizar os dados de forma personalizada e adaptada às suas necessidades, auxiliando na análise e compreensão das informações disponíveis no dashboard.

## 5. CASOS DE USO

Com o intuito de avaliar o uso da ferramenta, foram selecionados os dados correspondentes ao primeiro e segundo turnos das eleições de 2022 no Estado da Paraíba. Essa seleção resultou em um cenário abrangente, composto por 1 estado, 68 zonas eleitorais, 223 municípios, 9.603 seções eleitorais e aproximadamente 145 milhões de registros de logs.

Para analisar os dados, foram desenvolvidos dois casos de uso distintos. O primeiro caso de uso concentra-se no primeiro turno da seção 0026, zona 0016 no município de Campina Grande

durante as eleições de 2022 no estado da Paraíba. Nesse caso, a votação ocorreu em 2 de outubro de 2022, das 8h às 17h, e os eleitores puderam votar para os cargos de presidente, governador, senador, deputado estadual e deputado federal.

O segundo caso de uso está relacionado ao segundo turno da seção 0026, zona 0016 no município de Campina Grande durante as eleições de 2022 no estado da Paraíba. Essa etapa ocorreu em 30 de outubro de 2022, das 8h às 17h, e envolveu a votação para os cargos de presidente e governador.

Cada caso de uso será detalhado a seguir, abordando as análises específicas realizadas em cada turno das eleições. Ao final, será apresentada uma conclusão que comparará os resultados e insights obtidos nos dois casos de uso.

### 5.1 Caso 1 - Análise dos dados da seção 0026 / zona 0016 (1º turno - 2022)

Os dados a seguir são referentes ao primeiro turno das eleições de 2022, na seção 0026, zona 0016, localizada no município de Campina Grande. Com base nos registros de logs da urna, foram analisadas algumas métricas que podem ser visualizadas na Figura 6. São elas: Quantidade de Votos Computados; Horário de Início das Atividades na seção; Horário de Término das Atividades na seção; Horário em que o Primeiro Voto foi Computado na Seção; Horário em que o Último Voto foi Computado na Seção; Média de Votos por Hora na Seção; Tempo Médio de Votação do Eleitor; e Tempo Médio na Urna do Eleitor.

É importante esclarecer que o Tempo em Votação foi contabilizado a partir do momento em que o título do eleitor foi digitado pelo mesário até o voto ser computado. Já o Tempo na Urna foi cronometrado desde o momento em que o eleitor foi habilitado para votação até o instante em que o voto foi computado.

Conforme podemos verificar na Figura 6, nesta seção foram computados 294 votos. O período de utilização da urna foi de aproximadamente 10 horas, iniciando às 7:13 e encerrando às 17:09. O primeiro voto da seção foi registrado às 8:03 e o último voto às 16:53. Além disso, essa seção apresentou uma média de 32 votos por hora. O tempo médio gasto por cada eleitor para votar foi de 1 minuto e 15 segundos, sendo que desses, 54 segundos foi o tempo gasto na urna durante o processo de votação.

Ao analisarmos o tempo médio gasto por eleitor e a quantidade de votos computados na seção, chegamos à conclusão de que, durante o período de votação das 8 às 17 horas, totalizando 9 horas, os eleitores utilizaram a urna por aproximadamente 4 horas e meia, o que corresponde à metade do tempo disponível. Além disso, a seção esteve em atividade por cerca de 6 horas, o que implica que há aproximadamente 3 horas restantes correspondentes a períodos sem atendimento ou intervalos entre os atendimentos aos eleitores.

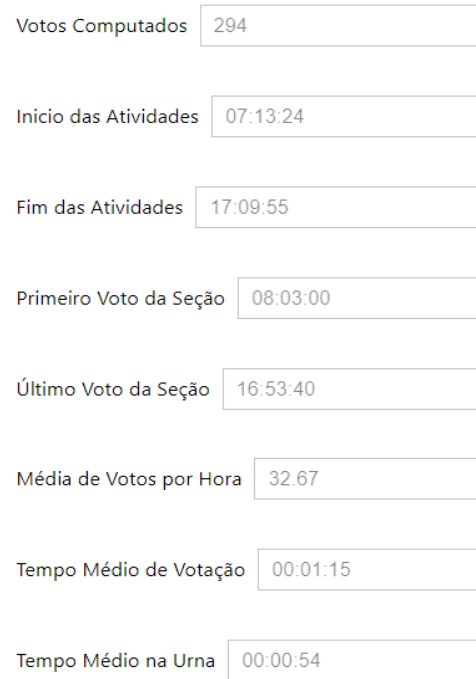


Figura 6: Dados da seção

Além das métricas gerais da seção (Figura 6), o dashboard também oferece a possibilidade de gerar gráficos para facilitar a análise dos dados.

Na Figura 7, podemos observar o gráfico que representa a porcentagem de votantes por hora da seção, disponibilizado no Dashboard da ferramenta. A análise desse gráfico nos permite chegar à conclusão de que os horários de pico da seção ocorreram entre 11 e 15 horas, concentrando 63,61% dos votantes dessa seção. Além disso, podemos verificar que o período das 8 horas foi o momento com menor concentração de votantes, representando apenas 6,46% do total de votantes da seção.

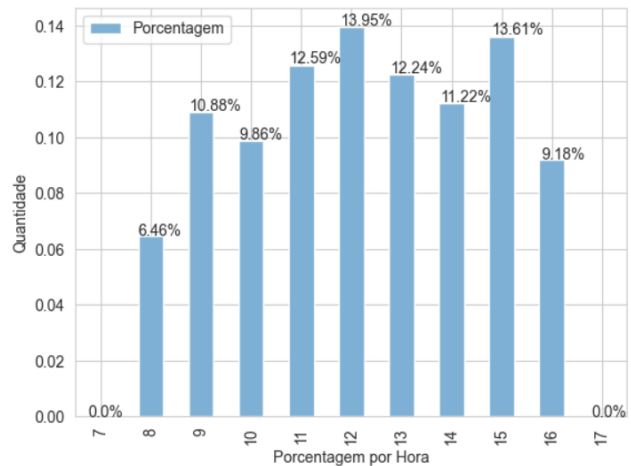
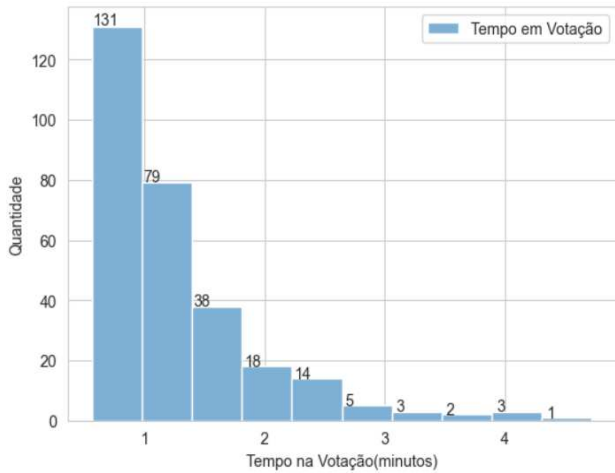


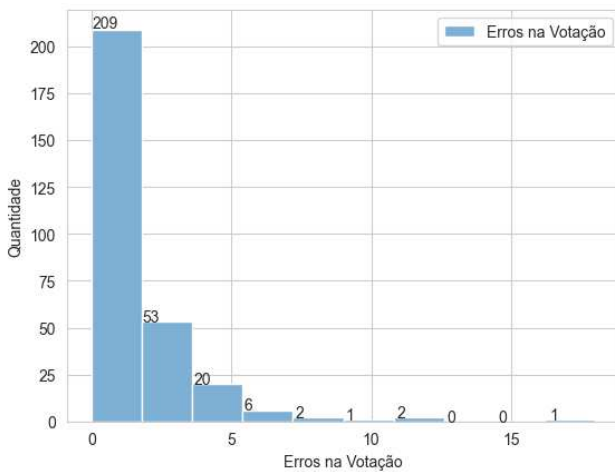
Figura 7: Porcentagem por hora

A Figura 8 apresenta a quantidade de votantes agrupados de acordo com o tempo necessário para concluir o processo de votação. Ao analisar o gráfico, podemos observar que a maioria esmagadora dos eleitores levou cerca de 1 a 2 minutos para concluir sua votação. No entanto, existem alguns casos excepcionais (outliers) em que alguns eleitores levaram mais de 3 a 4 minutos para finalizar o processo de votação.



**Figura 8: Tempo na votação em minutos**

Na Figura 9, o gráfico representa a quantidade de erros cometidos pelos eleitores durante a utilização da urna eletrônica. Considera-se um erro quando uma tecla é pressionada incorretamente. Observa-se que a maioria dos eleitores nesta seção tem média de 1.31 erros. No entanto, existem alguns casos atípicos em que alguns eleitores cometeram mais de 10 erros, sendo o máximo registrado de 18 erros durante o processo de votação.

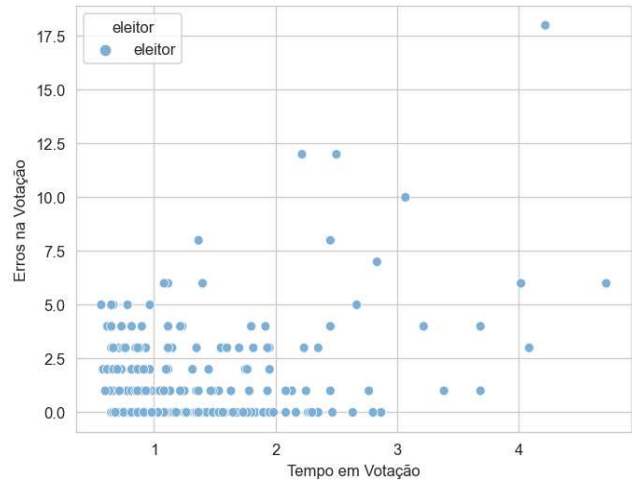


**Figura 9: Erros na votação**

O dashboard também oferece a funcionalidade de gerar gráficos de dispersão, permitindo correlacionar duas análises da seção eleitoral.

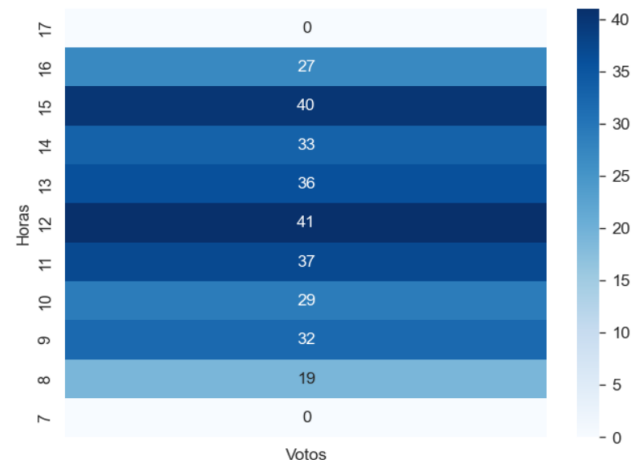
Na Figura 10, o gráfico ilustra a relação entre o tempo de votação dos eleitores e a quantidade de erros cometidos durante o processo. É possível observar que a maioria dos eleitores teve

uma média de 1,31 erros e levou aproximadamente 1 minuto e 15 segundos para concluir a votação. Além disso, não há uma correlação óbvia entre esses dois fatores, pois nem sempre os eleitores que cometeram mais erros foram aqueles que gastaram mais tempo no processo de votação. Essa falta de correlação pode ser atribuída a diversos fatores, como a familiaridade dos eleitores com o sistema de votação, a clareza das instruções fornecidas e a atenção dos eleitores durante o processo.



**Figura 10: Tempo na votação X Erros na votação**

Por fim, o dashboard também oferece a opção de gerar gráficos de calor relacionados às opções de votos por hora. Na Figura 11, o gráfico de calor fornece uma representação visual clara dos períodos de maior movimentação da seção. Foi observado que o período da tarde apresentou uma maior atividade em comparação ao período da manhã, com destaque para os horários das 12 e 15 horas, que registraram 41 e 40 votos, respectivamente. Por outro lado, o horário das 8 horas apresentou apenas 19 votos, indicando uma menor movimentação nesse período.



**Figura 11: Votos por hora**

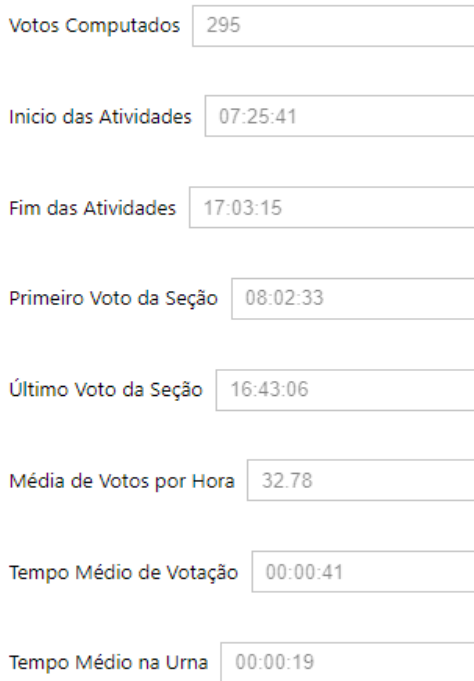
## 5.2 Caso 2 - Análise dos dados da seção 0026 / zona 0016 (2º turno - 2022)

Com o intuito de comparar os dados previamente apresentados, selecionamos, neste exemplo, os dados da mesma seção, Seção

0026 da Zona 0016 do município de Campina Grande, porém referentes ao segundo turno das eleições de 2022.

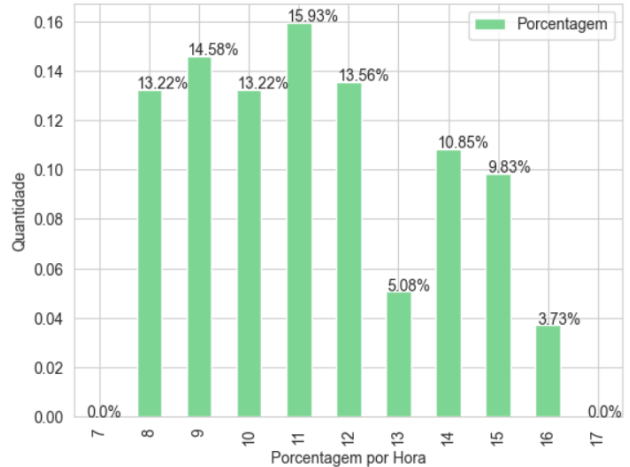
Na Figura 12, temos as mesmas métricas apresentadas no primeiro turno (Figura 6) e podemos observar que o comparecimento foi equivalente ao primeiro turno, com 295 votos computados. Além disso, o período de utilização da urna diminuiu um pouco, iniciando às 07:25 e encerrando às 17:03. O primeiro voto da seção foi registrado às 8:02 e o último às 16:43. A média de votos por hora permaneceu a mesma, com 32 votos. No entanto, o tempo médio de votação e o tempo médio na urna diminuíram para 41 segundos e 19 segundos, respectivamente.

Considerando a quantidade de votos computados, o tempo médio gasto por eleitor e o período de votação das 8h às 17h, totalizando novamente 9 horas, houve uma redução significativa no tempo de utilização da urna pelos eleitores, aproximadamente 1 hora e meia. Além disso, a seção esteve em atividade por cerca de 3 horas e 20 minutos, quase a metade em comparação ao primeiro turno. As 6 horas restantes correspondem a períodos sem atendimento ou intervalos entre os atendimentos aos eleitores.



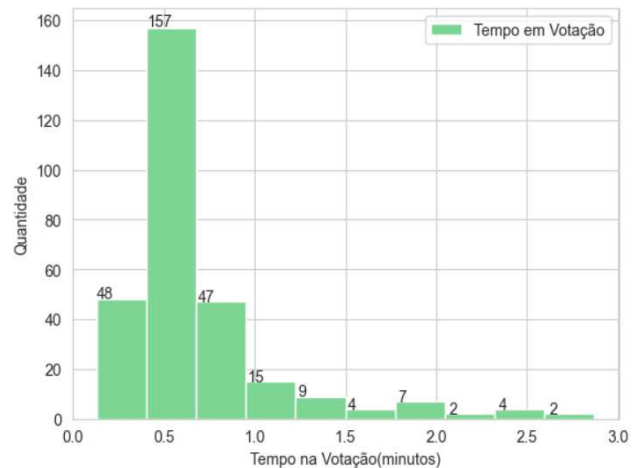
**Figura 12: Dados da seção**

Na Figura 13, o gráfico de porcentagem de votantes por hora revela uma inversão em relação ao primeiro turno. Diferentemente do que ocorreu anteriormente, o turno da manhã, das 8 às 12 horas, foi o mais movimentado, representando mais de 70% do total de eleitores dessa seção. Um destaque especial é o intervalo das 11 horas, que concentrou 15,83% dos votos nesse período. Por outro lado, o período da tarde foi consideravelmente mais tranquilo, com o horário das 16 horas equivalendo a apenas 3,73% do total de votos.



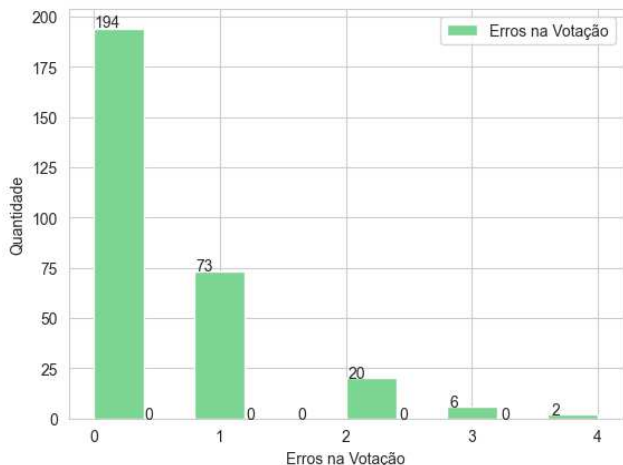
**Figura 13: Porcentagem por Hora**

Na Figura 14, os dados apresentados revelam uma redução significativa no tempo gasto em votação em comparação ao primeiro turno, provavelmente devido à diminuição no número de cargos e candidatos por cargo. A maioria dos eleitores levou menos de um minuto para concluir sua votação, evidenciando um processo mais ágil. O tempo máximo observado também diminuiu, ficando um pouco abaixo de três minutos. Essa redução no tempo de votação pode ser atribuída à simplificação do processo eleitoral no segundo turno.



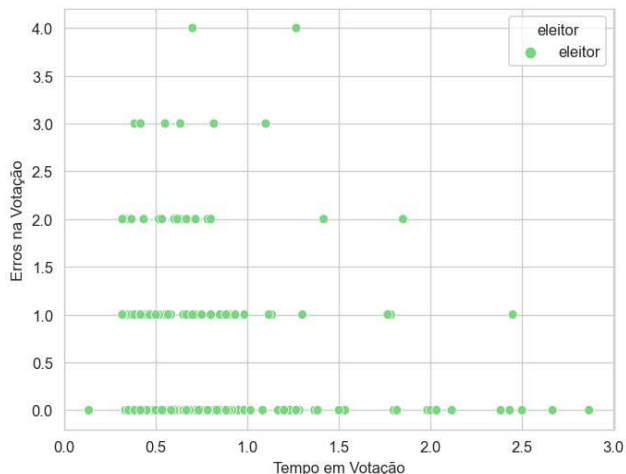
**Figura 14: Tempo na votação**

Na Figura 14, ao analisar os dados apresentados no gráfico, fica evidente uma redução na quantidade de erros durante a votação em comparação ao primeiro turno. A média de erros por eleitor foi de 0,47, sendo que o máximo registrado foi de 4 erros. Essa redução representa uma diminuição significativa em relação à média de 1,31 erros e ao máximo de 18 erros observados no primeiro turno.



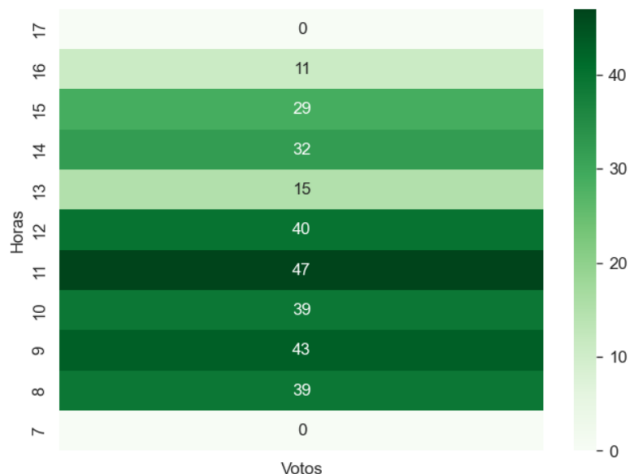
**Figura 15: Erros na Votação**

Na Figura 15, é possível observar uma diminuição na quantidade média de erros e no tempo médio de votação, que foram de 0,47 erros e 41 segundos, respectivamente. Novamente, constata-se a falta de uma correlação forte entre o número de erros e o tempo de votação, uma vez que os eleitores que levaram mais tempo para votar estão entre aqueles que não cometeram erros ou os que cometeram apenas um erro. Isso indica que o tempo gasto pelos eleitores na votação não está diretamente relacionado à ocorrência de erros.



**Figura 16: Tempo na votação X Erros na Votação**

Na Figura 16, podemos observar uma inversão em relação ao primeiro turno nas preferências dos votantes em relação ao período de votação. Diferentemente do primeiro turno, a maioria dos votantes optou por votar durante o período da manhã. Por outro lado, o período da tarde se mostrou mais tranquilo, com menos votos por hora. O horário de pico de votação ocorreu às 11 horas, com 47 votos registrados, enquanto o período de menor movimento foi às 16 horas, com apenas 11 votos.



**Figura 17: Votos por hora**

### 5.3 Discussão dos Casos

Comparando os dois casos de uso apresentados, podemos observar algumas diferenças significativas. No primeiro turno, notamos que a maioria dos eleitores cometeu poucos erros durante a votação, em média 1,31 erros, e levou cerca de 1 minuto e 15 segundos para concluir o processo de votação, conforme representado pela Figura 10. Não houve uma correlação clara entre o tempo de votação e a quantidade de erros, pois nem sempre os eleitores que cometeram mais erros foram aqueles que gastaram mais tempo no processo de votação.

Já no segundo turno, conforme demonstrado na Figura 14, ocorreu uma redução significativa tanto na quantidade média de erros quanto no tempo médio de votação. A média de erros por eleitor diminuiu para 0,47, sendo que o máximo registrado foi de 4 erros. Além disso, o tempo médio de votação diminuiu para 41 segundos. Mais uma vez, não foi observada uma correlação forte entre o número de erros e o tempo de votação. Os eleitores que demoraram mais para votar estavam entre aqueles que não cometeram erros ou que cometeram apenas um.

Essas diferenças entre os dois casos de uso sugerem possíveis melhorias no processo de votação entre o primeiro e o segundo turno. A redução tanto na quantidade de erros quanto no tempo de votação indica que os eleitores estavam mais familiarizados e eficientes no uso da urna no segundo turno. Essa melhoria também pode ser atribuída a diversos fatores, como a experiência adquirida durante o primeiro turno, a maior familiaridade com o processo de votação e a redução do número de cargos a serem votados.

Esses resultados destacam a importância de fornecer um sistema de votação rápido e intuitivo, visando reduzir a ocorrência de erros por parte dos eleitores e diminuir o tempo de espera. Além disso, a análise desses dados pode auxiliar na identificação de possíveis áreas de melhoria no sistema de votação, visando aprimorar a eficiência e a experiência dos eleitores durante o processo eleitoral.

## 6. EXPERIÊNCIA DE DESENVOLVIMENTO

Inicialmente, a estrutura do site do TSE foi analisada para determinar a viabilidade da automação da extração de dados e se os dados estavam em um formato legível. Em seguida, foram selecionadas as tecnologias a serem utilizadas na ferramenta, levando em consideração a necessidade de lidar com grandes volumes de dados e a experiência prévia adquirida em trabalhos anteriores.

O desenvolvimento do UrnaLogs foi baseado no método ágil SCRUM [8], embora não tenha sido aplicado integralmente. Essa metodologia foi escolhida devido à sua abordagem de entregas recorrentes em Sprints, permitindo o aprimoramento contínuo da ferramenta. As Sprints ocorriam em ciclos de 1 ou 2 semanas cada, e as reuniões de acompanhamento com o orientador desempenhavam um papel fundamental. Essas reuniões ajudavam a definir os objetivos de cada ciclo, identificar eventuais erros e avaliar se a implementação estava alinhada com a proposta da ferramenta.

Durante o desenvolvimento da ferramenta, nos deparamos com um desafio significativo relacionado a um vazamento de memória no navegador Chrome durante a execução do módulo de web scraping. Para resolver esse problema, implementamos um mecanismo que reinicia o navegador caso ocorra algum erro ou se atinja o limite de downloads, permitindo assim que a extração de dados seja retomada a partir do ponto em que parou. Outro desafio que enfrentamos foi o desenvolvimento do Dashboard interativo no notebook Python. Esse desafio foi superado com sucesso graças ao uso da biblioteca ipywidgets, que tornou a dashboard mais intuitiva e fácil de ser utilizada.

## 7. CONCLUSÃO

O UrnaLogs foi desenvolvido com o objetivo de extrair e analisar dados eleitorais do site do TSE de forma automatizada. Através desse processo, foi possível obter insights valiosos sobre o tempo de votação, quantidade de erros e padrões de comportamento dos eleitores.

A escolha das tecnologias utilizadas, considerando a necessidade de lidar com grandes volumes de dados, mostrou-se acertada, garantindo a eficiência e versatilidade da ferramenta. A aplicação do método ágil SCRUM, mesmo que adaptado, permitiu um desenvolvimento iterativo e contínuo, resultando em entregas recorrentes e melhorias progressivas.

Ao longo do projeto, a comunicação constante com o orientador foi fundamental para alinhar os objetivos, identificar eventuais erros e garantir que a implementação estivesse alinhada com as expectativas e necessidades da ferramenta. Essa interação também possibilitou uma maior compreensão dos desafios e oportunidades relacionados à análise dos dados eleitorais.

O UrnaLogs oferece uma solução valiosa para a análise de dados eleitorais, permitindo uma compreensão mais profunda e precisa dos resultados das eleições. Sua utilização pode contribuir para a

identificação de padrões, tendências e possíveis melhorias no processo eleitoral, auxiliando na tomada de decisões informadas e no aprimoramento contínuo do sistema democrático. Em suma, o UrnaLogs representa um avanço significativo na área de análise de dados eleitorais, proporcionando uma visão mais clara e detalhada do processo eleitoral. Seu potencial impacto na transparência, eficiência e confiabilidade das eleições é promissor, abrindo caminho para futuras pesquisas e desenvolvimentos nesse campo.

Durante o desenvolvimento da ferramenta, algumas limitações foram identificadas e não tiveram como serem abordadas devido às restrições de tempo e escopo. A ferramenta de web scraping apresenta uma limitação inerente relacionada à possibilidade de mudanças na estrutura do site do TSE. Caso ocorram alterações, será necessário refatorar o código para se adequar à nova estrutura. Ademais, outra limitação da ferramenta está relacionada à disponibilidade e estrutura dos arquivos que contém os logs das urnas. A estrutura desses arquivos pode ser modificada em eleições futuras, o que pode afetar a funcionalidade da ferramenta.

O trabalho realizado teve como foco o estudo dos logs das urnas eleitorais. No entanto, existem ainda alguns pontos que não foram aprofundados e que podem ser explorados em trabalhos futuros. Um desses pontos é a análise dos logs referentes aos testes realizados dias antes das eleições, nos quais seria possível identificar se a urna estava em conformidade com os pré-requisitos necessários para ser utilizada. Além dos logs das urnas, o arquivo compactado contém outros arquivos que poderiam ser estudados, como o boletim de urna. Esse boletim contém informações como o total de votos por partido, total de votos por candidato, total de votos em branco, total de comparecimento, total de votos nulos, entre outros.

Outro aspecto interessante a ser explorado é o Registro Digital do Voto (RDV), que permite validar os votos armazenados na urna. Esses votos são ordenados de forma aleatória para manter a anonimidade dos eleitores. Além disso, seria possível, também, desenvolver outras visualizações dos dados, permitindo comparar várias seções eleitorais ou explorar outros componentes, como estado, município e zona. Essas análises ampliariam ainda mais o entendimento e a utilização dos dados obtidos. Esses são alguns dos possíveis trabalhos futuros que podem contribuir para a evolução do projeto UrnaLogs.

## 8. AGRADECIMENTOS

Desejo expressar minha profunda gratidão ao meu orientador, Maxwell Guimarães de Oliveira, por sua paciência e dedicação em me guiar durante este trabalho. Seu conhecimento e orientação foram de valor inestimável para o amadurecimento e progresso deste projeto.

Também gostaria de agradecer à minha família, que esteve ao meu lado incondicionalmente, fornecendo apoio emocional e encorajamento durante todo o percurso. Sua presença e amor foram fundamentais para o meu sucesso.

Por fim, quero estender meus agradecimentos a todos os professores do curso, cujas instruções e conhecimentos

compartilhados ao longo do curso foram essenciais para a minha formação acadêmica. Agradeço pela sua dedicação em transmitir seu conhecimento e pela inspiração que proporcionaram em minha jornada educacional.

## 9. REFERÊNCIAS

- [1] ELEITORES da região enfrentam longas filas com espera de até 4 horas para votar. G1 GLOBO. São José dos Campos, 02 de outubro de 2022.  
<https://g1.globo.com/sp/vale-do-paraiba-regiao/eleicoes/2022/noticia/2022/10/02/eleitores-da-regiao-enfrentam-longas-filas-com-espera-de-ate-4h-para-votar.ghtml>  
Último acesso em: 03 de junho 2023
  - [2] FERRÃO et al. Urnas Eletrônicas no Brasil: linha do tempo, evolução e falhas e desafios de segurança. 2019. Artigo original. Publicado na Revista Brasileira de Computação Aplicada, julho de 2019, p.2. Disponível em:  
<http://seer.upf.br/index.php/rbca/article/view/9056>  
Último acesso em: 03 de junho 2023
  - [3] MITCHELL, Ryan. Web Scraping com Python: Coletando mais dados da web moderna. 2019, Novatec Editora, p. 9-12. Disponível em:  
[https://www.google.com.br/books/edition/Web\\_Scraping\\_com\\_Python/95qODwAAQBAJ?hl=pt-BR&gbpv=0](https://www.google.com.br/books/edition/Web_Scraping_com_Python/95qODwAAQBAJ?hl=pt-BR&gbpv=0)  
Último acesso em: 03 de junho 2023
  - [4] SELENIUM Webdriver Disponível em:  
<https://www.selenium.dev/documentation/webdriver/>  
Último acesso em: 03 de junho 2023
  - [5] PY7ZR Disponível em: <https://pypi.org/project/py7zr/>  
Último acesso: 03 de junho de 2023
  - [6] ORM Peewee Disponível em:  
<https://docs.peewee-orm.com/en/latest/peewee/database.html#initializing-a-database>  
Último acesso em: 03 de junho 2023
  - [7] O que é ORM? Disponível em:  
<https://renan04-marques.medium.com/oque-%C3%A9-orm-3432889661f6>  
Último acesso em: 03 de junho 2023
  - [8] WHAT is Scrum? Disponível em:  
Disponível em:  
[https://www.scrum.org/resources/what-scrum-module?gclid=Cj0KCQjw7PCjBhDwARIsANo7CgknuczdkzXizzJ-N6MF#W2AuUwbPNSxwi\\_psPIFriH\\_aIFo9fpmijygaAmnhEALw\\_wcB](https://www.scrum.org/resources/what-scrum-module?gclid=Cj0KCQjw7PCjBhDwARIsANo7CgknuczdkzXizzJ-N6MF#W2AuUwbPNSxwi_psPIFriH_aIFo9fpmijygaAmnhEALw_wcB)  
Último acesso em: 03 de junho 2023
  - [9] O que é ETL? Disponível em:  
<https://www.oracle.com/br/integration/what-is-etl/>  
Último acesso em: 07 de junho 2023
  - [10] O que é um banco de dados relacional(RDBMS)? Disponível em:  
<https://www.oracle.com/br/database/what-is-a-relational-database/>  
Último acesso em: 03 de junho 2023
  - [11] O que é Dashboard? Disponível em:  
<https://neilpatel.com/br/blog/dashboard-o-que-e/>  
Último acesso em: 07 de junho 2023
-