



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

**MARIA CLARA MORAES MONTEIRO DANTAS**

**MONITORA SANTA CRUZ:  
UMA FERRAMENTA WEB PARA AUXILIAR O CIDADÃO NO  
MONITORAMENTO DOS GASTOS DA PREFEITURA DE SANTA  
CRUZ DO CAPIBARIBE**

**CAMPINA GRANDE - PB**

**2021**

**MARIA CLARA MORAES MONTEIRO DANTAS**

**MONITORA SANTA CRUZ:  
UMA FERRAMENTA WEB PARA AUXILIAR O CIDADÃO NO  
MONITORAMENTO DOS GASTOS DA PREFEITURA DE SANTA  
CRUZ DO CAPIBARIBE**

**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 Bacharela em  
Ciência da Computação.**

**Orientador: Professor Dr. Herman Martins Gomes.**

**CAMPINA GRANDE - PB**

**2021**



D192m Dantas, Maria Clara Moraes.

Monitora Santa Cruz: uma ferramenta web para auxiliar o cidadão no monitoramento dos gastos da Prefeitura de Santa Cruz do Capibaribe. / Maria Clara Moraes Dantas. - 2021.

12f.

Orientador: Prof. Dr. Herman Martins Gomes.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Aplicação web. 2. Dados financeiros públicos. 3. Santa Cruz do Capibaribe - PE - gastos. 4. Prefeitura Municipal de Santa Cruz do Capibaribe - PE. 5 Portal da transparência. 6. Transparência de gastos públicos. 7. Aplicação web. 8. Acesso à informação. 9. Visualização de dados. 10. Linguagem de programação R. I. Gomes, Herman Martins. II. Título.

CDU:004.6(045)

**Elaboração da Ficha Catalográfica:**

Johnny Rodrigues Barbosa  
Bibliotecário-Documentalista  
CRB-15/626

**MARIA CLARA MORAES MONTEIRO DANTAS**

**MONITORA SANTA CRUZ:  
UMA FERRAMENTA WEB PARA AUXILIAR O CIDADÃO NO  
MONITORAMENTO DOS GASTOS DA PREFEITURA DE SANTA  
CRUZ DO CAPIBARIBE**

**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 Bacharela em  
Ciência da Computação.**

**BANCA EXAMINADORA:**

**Professor Dr. Herman Martins Gomes  
Orientador – UASC/CEEI/UFCG**

**Professor Dr. Cláudio E. C. Campelo  
Examinador – UASC/CEEI/UFCG**

**Professor Tiago Lima Massoni  
Professor da Disciplina TCC – UASC/CEEI/UFCG**

**Trabalho aprovado em: 20 de Outubro de 2021.**

**CAMPINA GRANDE - PB**

## **ABSTRACT**

According to the Transparency Law, information about public income and expenses should be available in real-time, with a maximum deadline of 24 hours, in a website commonly denominated Transparency Portal. Although data availability represents a step forward in a more transparent and auditable government by its citizens, data monitoring is still a complex task. The Transparency Portals are not obligated to make available any visualization that helps the users to inspect and understand the data, and for the specific case of the Santa Cruz do Capibaribe city, in the state of Pernambuco, the data is only available in the form of tables. Thus, for the majority of the people, it is not possible to extract any value or relevant information about the data. In this context, the present graduation conclusion work aims to empower the citizen providing a web application that makes the automatic acquisition of the expenses data available in the city's Transparency Portal, consolidates the important information of that data, and provides multiple interactive visualizations that assist the task of understanding the expenses made by any citizens from all backgrounds.

# Monitora Santa Cruz: Uma ferramenta web para auxiliar o cidadão no monitoramento dos gastos da prefeitura de Santa Cruz do Capibaribe

Maria Clara Moraes Monteiro  
Dantas  
Universidade Federal de Campina Grande  
Campina Grande, Paraíba, Brasil  
maria.dantas@ccc.ufcg.edu.br

Orientador: Herman Martin  
Gomes  
Universidade Federal de Campina Grande  
Campina Grande, Paraíba, Brasil  
hmg@computacao.ufcg.edu.br

## Resumo

De acordo com a lei da transparência, informações sobre receitas e gastos públicos devem ser divulgadas em tempo real, com prazo máximo de 24 horas, em um sítio da Internet, comumente denominado portal da transparência. Apesar da disponibilidade dos dados ser um passo na direção de um governo mais transparente e auditável pela população, a monitoração desses dados ainda é uma tarefa complexa. Os portais não são obrigados a fornecer quaisquer meios de visualização que facilitem a inspeção e a compreensão dos dados e, no caso específico do portal da transparência do município de Santa Cruz do Capibaribe, em Pernambuco, os dados estão disponíveis apenas em tabelas. Desse modo, para a maior parte da população, não é possível extrair valor ou informações relevantes dos dados. Neste contexto, o presente trabalho de conclusão de curso tem como objetivo empoderar o cidadão, provendo uma aplicação web que faz a aquisição automática dos dados de despesas, consolida informações relevantes destes dados e fornece diferentes visualizações interativas que auxiliam a compreensão dos gastos por cidadãos comuns.

## Palavras-Chave

visualização de dados, aplicação web, portal da transparência, acesso à informação.

## Repositórios

[Crawler para capturar os dados do portal.](#)

[Aplicação web para visualização dos dados.](#)

## Link para Acesso à Ferramenta Desenvolvida

<https://clara-moraes-dantas.shinyapps.io/monitora-scc/>

## 1. INTRODUÇÃO

O portal da transparência do município de Santa Cruz do Capibaribe [1], PE, disponibiliza seus dados de despesas diárias apenas por meio tabelas, e conta, até o dia 28 de setembro de 2021, com 61.835 linhas que datam desde 03/01/2013. Apesar de existirem muitos dados como, por exemplo, valor do gasto, nome, cpf ou cnpj do favorecido, categoria etc., não é possível extrair muita informação sem ferramentas que permitam visualizar e

sumarizar esses dados. Este trabalho se propõe a resolver este problema construindo uma aplicação web que fornece visualizações interativas dessas informações, onde o usuário pode aplicar filtros para explorar o comportamento dos gastos em determinado período de tempo, comparar os gastos por categoria (educação, cultura, urbanismo etc.), explorar a distribuição dos gastos, dentre outras funcionalidades.

A execução desse trabalho foi dividida em duas grandes etapas, a coleta dos dados e o desenvolvimento de uma aplicação web contendo as visualizações. Para a realização do primeiro passo, foi necessária a construção de um *web crawler*<sup>2</sup> pois, apesar de o portal da transparência dever, obrigatoriamente, fornecer um meio de exportar os dados em formato de arquivo csv (tipo de arquivo em que os valores são separados por vírgula), foram encontrados alguns problemas na exportação fornecida. O arquivo gerado pela exportação oferecida pelo portal da transparência de Santa Cruz do Capibaribe é mal formatado, possui informações faltantes, e não contém cabeçalho [Figura 1.1 e 1.2]. Além disso, não é possível exportar a seção que fornece mais detalhes sobre o pagamento como informações sobre a categoria, histórico do empenho, cpf ou CNPJ do favorecido etc.

**FIGURA 1.1.** Exemplo do arquivo gerado pela exportação das despesas diárias detalhadas. Colunas de A à C.

	A	B	C
1	24/09/2021	<span style="display:flex;flex-direction:column;"> <i class="fas fa-external-link-alt"></i>2021NE0000652</span>	18852
2	23/09/2021	<span style="display:flex;flex-direction:column;"> <i class="fas fa-external-link-alt"></i>2021NE0000153</span>	374
3	22/09/2021	<span style="display:flex;flex-direction:column;"> <i class="fas fa-external-link-alt"></i>2021NE0000734</span>	18781

Após o *crawler* coletar todos os dados de interesse, esses dados são estruturados em um arquivo em formato csv, que em seguida é consumido pela aplicação web.

A solução final desenvolvida por este trabalho tem como resultado uma ferramenta que empodera o cidadão santa-cruzense gerando visualizações que o ajudam a entender como a prefeitura utiliza seus recursos financeiros, e fornece meios de fiscalização pública.

<sup>2</sup> Um *web crawler* é um *software* que pesquisa informações de maneira sistemática na Internet.

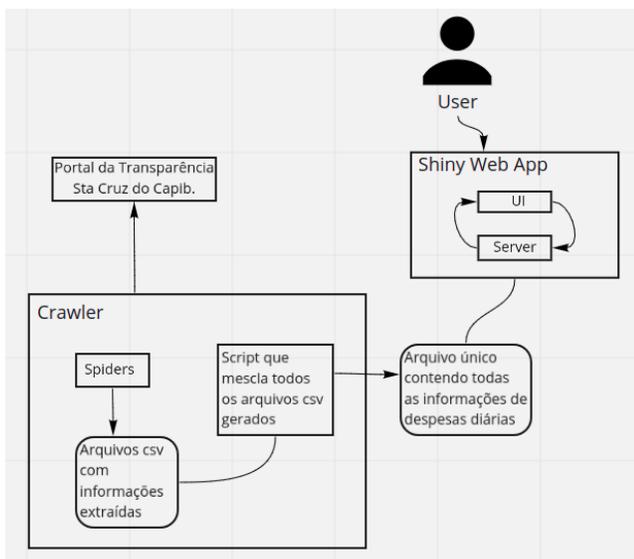
**FIGURA 1.2.** Exemplo do arquivo gerado pela exportação das despesas diárias detalhadas. Colunas D à F.

D	E	F
Global	FERNANDO COSTA DO NASCIMENTO	44,7
Estimativo	FMSSCC PROG COMBATE COVID 19 - EFETIVOS	453,6
Global	BARBOSA & COUTO ADVOGADOS ASSOCIADOS	180

## 2. Arquitetura e Projeto da Solução

Conforme mencionado anteriormente, a solução foi dividida em dois grandes passos, o primeiro é a coleta dos dados que foi realizada com a implementação de um *crawler* que processa e extrai as informações necessárias provenientes do portal da transparência de Santa Cruz; o segundo é a aplicação web que implementa as visualizações para os dados coletados. A aplicação web por sua vez é dividida entre *back-end* e *front-end*.

**FIGURA 2.** Arquitetura da solução.



### 2.1 Crawler

O crawler foi desenvolvido na linguagem de programação Python, com os frameworks Scrapy<sup>3</sup> e Selenium<sup>4</sup>. Foi necessário usar Selenium em conjunto ao Scrapy, pois o portal da transparência em questão possui elementos que carregam de modo dinâmico e, apesar de existir uma alternativa para que o Scrapy lide com esse tipo de conteúdo, o Selenium foi empregado para resolver essa questão de forma mais simples, além de fornecer um *webdriver*<sup>5</sup> que facilita o acompanhamento do trabalho que o *crawler* está realizando, e o processo de depuração ao longo do desenvolvimento.

<sup>3</sup> Página oficial: <https://scrapy.org/>

<sup>4</sup> Página oficial: <https://www.selenium.dev/>

<sup>5</sup> Um *webdriver* é navegador nativo usado pelo Selenium que age como um usuário humano navegando na *web*.

Como exemplificado [Figura 2], o *crawler* é responsável por navegar no portal da transparência, coletar os dados necessários, e gerar arquivos no formato csv. Esse passo-a-passo poderia ser realizado de forma direta, porém existe uma limitação de tempo de permanência na página de pagamentos do portal, e esse tempo é insuficiente para que o *crawler* obtenha as informações de pagamento [Figura 3] e detalhamento do pagamento [Figura 4].

**FIGURA 3.** Listagem de pagamentos do portal da transparência de Santa Cruz do Capibaribe.

Código de verificação: 056CWA540681 | Última Atualização: 24/09/2021 09:52:53 | Emitido em, 29 Setembro de 2021

Data	Empenho	Parcela	Tipo	Favorecido(a)	Valor
24/09/2021	2021NE0000652	0018852	Global	FERNANDO COSTA DO NASCIMENTO	44,70
23/09/2021	2021NE0000153	0000374	Estimativo	FMSSCC PROG COMBATE COVID 19 - EFETIVOS	453,60
22/09/2021	2021NE0000143	0000007	Estimativo	CELPE - COMPANHIA ENERGETICA DE PERNAMBUCO	533,72

**FIGURA 4.** Detalhamento do pagamento de empenho 2021NE0000652.

Empenho Nº: 000000652

DADOS GERAIS	CLASSIFICAÇÃO
Empenho: 000000652	Função: Administração
Tipo do Empenho: 3 - Global	Subfunção: Administração Geral
Espécie do Empenho: Original	Programa: ADMINISTRAÇÃO GERAL
Unidade Jurisdiccionada: Prefeitura Municipal de Santa Cruz do Capibaribe	Ação: GESTÃO TÉCNICA E ADMINISTRATIVA DA SECRETARIA
Unidade Orcamentária: SECRETARIA DE PLANEJAMENTO E GESTÃO DE PESSOAS	Categoria Econômica: Despesas Correntes
Histórico Empenho: VALOR QUE SE EMPENHA REFERENTE, A LOCAÇÃO DE IMÓVEL SITUADO DA RUA GUILHERME JOSÉ DA COSTA, Nº 48 - BAIRRO MALAQUIAS CARDOSO, SANTA CRUZ DO CAPIBARIBE, COM O CONTRATO Nº 011/2021, PROCESSO DE DISPENSA DE LICITAÇÃO Nº 013/2021, DISPENSA DE LICITAÇÃO Nº 004/2021. ( SUBSTITUIÇÃO DA NE 390).	Grupo de Despesa: Outras Despesas Correntes
Data Empenho: 16/06/2021	Modalidade de Aplicação: APLICAÇÕES DIRETAS
CPF/CNPJ do Credor: 00046194274468	Elemento de Despesa: Outros Serviços de Terceiros - Pessoa Física
Nome/Razão Social: FERNANDO COSTA DO NASCIMENTO	Subelemento de Despesa: LOCAÇÃO DE IMÓVEIS
Tipo: Pessoa Física	LICITAÇÃO
Fonte de Recurso: 01	Modalidade: Dispensa por outros motivos
	Procedimento / Sequencial da Modalidade Nº: 0004/2021

Sendo assim, para conseguir coletar todos os 61.835 registros de pagamentos, foram implementados dois *Spiders*<sup>6</sup>, um que é responsável por coletar os dados da tabela que lista os pagamentos [Figura 3] e outro *Spider* responsável pela coleta dos detalhamentos de cada pagamento [Figura 4]. Além disso, para não esgotar o tempo da sessão, o *Spider* de pagamentos só é capaz de coletar 10.000 elementos por vez, e o *Spider* de detalhamentos apenas 410. O número do último *Spider* é menor pois, enquanto o *spider* de pagamentos necessita apenas iterar nas linhas da tabela, o *spider* de detalhamento precisa, para cada elemento na tabela, clicar no link do empenho e esperar uma nova janela, com os dados do detalhamento, carregar. Ambos os números de cada conjunto de coleta foram determinados empiricamente.

Após os crawlers terminarem de percorrer todos os elementos, eles geram, para cada iteração, um arquivo no formato csv com os

<sup>6</sup> *Spiders* são classes do *framework* Scrapy que definem como um sítio ou conjunto de sítios da Internet serão processados pelo *crawler*.

dados capturados de forma estruturada. Ao todo, foram gerados 7 arquivos para os pagamentos, e 151 arquivos para os detalhamentos de pagamentos. Para ser mais simples processar esses dados e gerar as visualizações e sumários da aplicação *web*, todos os arquivos gerados foram mesclados em um único arquivo de formato csv. Com o objetivo de realizar o mapeamento entre pagamento e detalhamento, foi usado o campo do número do empenho e nome do favorecido. Na tabela abaixo (Tabela 1), estão descritas todas as colunas do conjunto de dados.

**TABELA 1.** Descrição das colunas do conjunto de dados gerado pelo crawler.

Coluna	Descrição
data	Data do pagamento
empenho	Identificador do empenho formado por <ano_empenho>NE<numero_empenho>
favorecido	Nome do favorecido
parcela	Identifica a parcela do pagamento do empenho
tipo	Tipo do empenho (Global, Ordinário, Estimativo, por exemplo)
valor	Valor do pagamento
acao	Ação promovida pelo pagamento (ex. "Manutenção das ações de comunicação institucional e cerimonial")
categoria_economica	Categoria a qual o gasto pertence (ex. Despesas Correntes, Despesas de Capital)
cpf_cnpj	CPF ou CNPJ do favorecido
data_empenho	Data do empenho relacionado ao pagamento

elemento_despesa	Demais elementos da despesa (ex. Outros Serviços de Terceiros – Pessoa Jurídica, Obras e Instalações)
funcao	Setor beneficiado pelo pagamento (ex. Comércio e Serviços, Educação, Urbanismo)
grupo_despesa	Grupo ao qual a despesa pertence (ex. Pessoal e Encargos Sociais, Investimento, Amortização, Refinanciamento da Dívida)
historico_empenho	Descrição detalhada sobre o pagamento, qual a motivo do gasto.
licitacao_modalidade	Se houve licitação, especifica qual a modalidade da mesma.
licitacao_sequencial_modalidade	String no formato <numero_procedimento>/<numero_sequencial_da_modalidade> que identifica o procedimento e modalidade
modalidade_aplicacao	Modalidade da aplicação (ex. Transferências a Consórcios Públicos)
nome_razao_social	Nome ou razão social do favorecido (proveniente da janela de detalhamento do pagamento)
num_empenho	Número identificador do empenho (proveniente da janela de detalhamento do pagamento)
programa	Programa ao qual o empenho pertence (ex. Administração geral, Iluminação pública)
subelemento_despesa	Subelementos referentes à despesa (ex. Uniformes, Serviços Técnicos Profissionais)
subfuncao	Subsetor beneficiado pelo pagamento (ex. Assistência Comunitária, Desporto Comunitário)
tipo_1	Indica se o favorecido é uma pessoa física ou jurídica
unidade_jurisdiccional	Jurisdição (ex. Prefeitura Municipal de Santa Cruz do Capibaribe)

unidade_orcamentaria	Unidade orçamentária (ex. Gabinete do Prefeito, Secretaria de Desenvolvimento Urbano)
----------------------	---

## 2.2 Aplicação Web

A ferramenta para a visualização dos dados foi escrita na linguagem de programação R, uma vez que é uma linguagem voltada para o uso em estatística e comumente utilizada para análise de dados. O pacote empregado para gerar os gráficos foi o Plotly<sup>7</sup>, e ele foi escolhido em detrimento ao ggplot (um dos principais pacotes R para criação de gráficos), pois a geração de gráficos interativos é abstraída pelo Plotly, de modo que todo gráfico produzido pelo pacote já é interativo e possui funcionalidades como *zoom* (você pode aumentar determinada área da plotagem), fazer *download* do gráfico em formato de arquivo pdf, movimentar os eixos do gráfico, remover valores de determinada categoria da plotagem a partir da seleção de valores na legenda, dentre outras utilidades.

Como o intuito do trabalho não é apenas gerar visualizações para os dados dos pagamentos realizados pela prefeitura, como também fazer com que as visualizações e análises sejam acessíveis à população, se faz necessário a construção de uma aplicação *web* que interaja com o código R que manipula os dados para gerar os gráficos. Por essa razão, o pacote Shiny<sup>8</sup> foi escolhido. Esse pacote provê um *framework* para a construção de aplicações *web* usando R, tornando as análises em páginas *web* interativas. Para utilizar o framework devem ser implementados os componentes de *UI* e servidor [Figura 2]. O primeiro componente de interface com o usuário contém o código HTML e referências para renderização dos gráficos, além de componentes para entradas de usuários como o seletor de datas e campos de *checkbox* etc. Já o servidor é responsável por realizar as operações necessárias no conjunto de dados, como filtros, mutações, e ordenações, e renderizar os gráficos gerados pelo Plotly. Por fim, outro fator que teve impacto na decisão do uso do Shiny foi o fato de a comunidade disponibilizar gratuitamente o armazenamento da aplicação em servidores, de modo que é possível realizar o *deploy* do projeto sem custos adicionais.

### 2.2.1 Funcionalidades e Apresentação da Aplicação

Visando conferir maior inclusão, a aplicação *web* utiliza-se o máximo possível de linguagem simples, e existe um esforço em tentar explicar como ler e interpretar cada tipo de gráfico para que as informações contidas na página sejam compreendidas por todas as pessoas, independente do seu grau de escolaridade.

Na tela inicial [Figura 5] está descrito o propósito da página, bem como demais informações sobre como os dados foram coletados e um link para o repositório do Github<sup>9</sup> contendo o código responsável pela coleta dos dados para maiores inspeções. Outra informação relevante nessa tela inicial é a instrução de como

interagir com os gráficos, aplicar *zoom* e fazer *download* do gráfico, por exemplo.

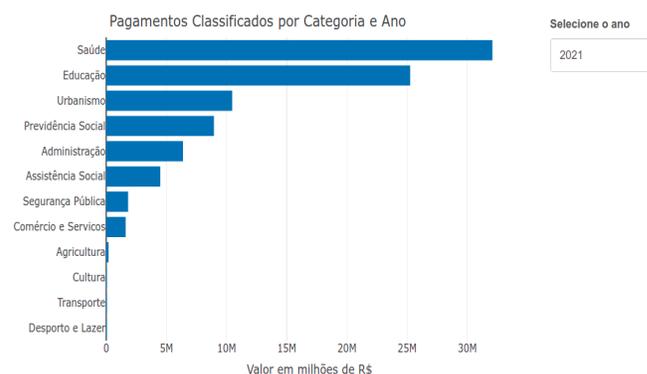
As seções seguintes da aplicação foram estruturadas de modo a responder perguntas específicas, como por exemplo “qual o investimento cada área recebeu por ano?”, “quais CNPJs mais receberam pagamentos da prefeitura?”.

FIGURA 5. Tela inicial do Monitora SCC.



A segunda seção da aplicação [Figura 6] encarrega-se de responder a pergunta “Quanto foi gasto por categoria e ano?”. Para responder a essa pergunta, os pagamentos foram agrupados por ano e categoria, e tiveram seus valores somados. O gráfico escolhido para essa visualização foi um gráfico de barras ordenadas pelo valor da soma dos pagamentos para cada categoria e ano, de modo a facilitar a comparação e o ranqueamento de cada categoria (saúde, educação, administração) em determinado ano.

FIGURA 6. Tela que responde a pergunta “qual o valor total gasto por categoria e ano?”



Ainda sobre a segunda seção da aplicação, um gráfico complementar [Figura 7] foi gerado para que fosse possível realizar a comparação do valor total despendido por cada categoria em diferentes períodos de tempo. Para gerar essa visualização, foram aplicados os mesmos procedimentos anteriores de agrupamento dos dados. O tamanho de cada ponto é proporcional ao total gasto na respectiva área e ano. Com essa

<sup>7</sup> Sítio oficial: <https://plotly.com/r/>

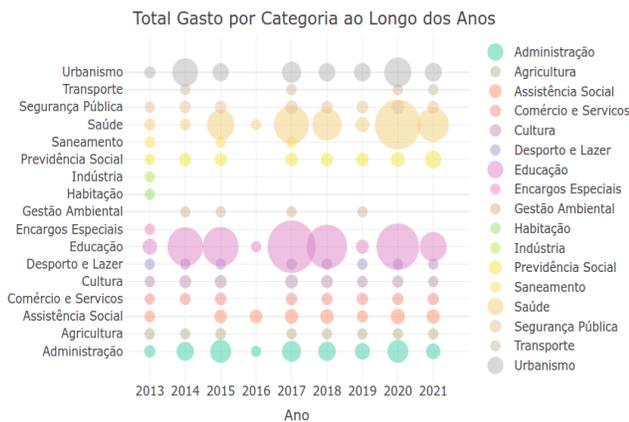
<sup>8</sup> Sítio oficial: <https://shiny.rstudio.com/>

<sup>9</sup> Sítio oficial: <https://github.com/>

visualização, é possível identificar as áreas que comumente recebem mais investimento, e as que menos recebem, como é o caso de Cultura, e Desporto e Lazer. O gráfico gerado permite que o usuário interaja fornecendo a possibilidade de remover a visualização de determinadas linhas na plotagem, devendo para tanto, remover a seleção da categoria desejada na coluna da legenda posicionada ao lado direito do gráfico.

Um fato interessante que é ressaltado com essa visualização é que, apesar da categoria Educação ser uma das categorias em que a prefeitura investe mais dinheiro, não parece haver uma correlação entre esse investimento e o IDEB, Índice de Desenvolvimento de Educação Básica, métrica do Inep<sup>10</sup> que avalia a qualidade da educação básica pública pois, desde 2013 até a data da última avaliação registrada para as escolas públicas municipais, em 2019, a meta do IDEB estabelecida para o município nunca é atingida para as séries de 1º ao 9º ano (não existem dados para o ensino médio).

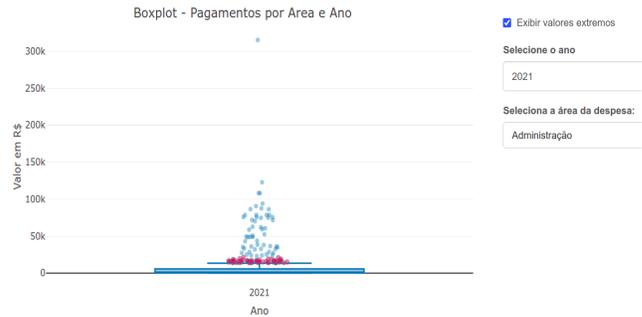
**FIGURA 7.** Valor investido por categoria desde 2013.



A terceira seção ocupa-se em analisar a distribuição dos valores e indicar algumas medidas de sumário. Foram utilizadas três visualizações diferentes, sendo elas o *boxplot*, gráfico de densidade, e gráfico de dispersão, que consideram os pagamentos realizados para determinada categoria e ano. No *boxplot* é exibido uma medida de centro, a mediana, primeiro e terceiro quartis, valores máximos e mínimos, e possíveis *outliers* (valores extremos que não se assemelham à distribuição dos dados em que foram amostrados) destacados em vermelho [Figura 8], e o usuário pode selecionar no *checkbox*, localizado à direita do gráfico, se deseja ou não visualizar os possíveis *outliers*. Para gerar esse *boxplot* os dados foram apenas agrupados por ano e categoria.

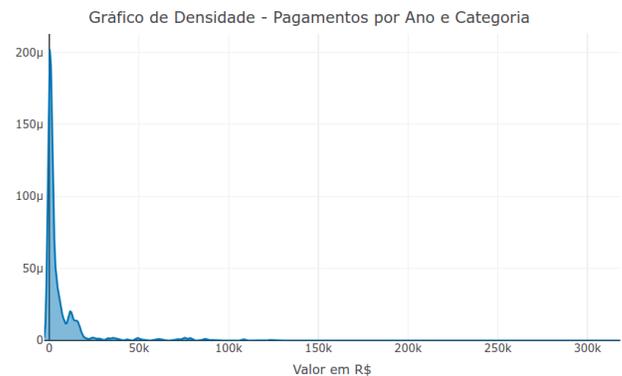
<sup>10</sup> O Inep, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, é uma autarquia federal vinculada ao Ministério da Educação.

**FIGURA 8.** Boxplot dos pagamentos realizados por categoria e ano.



O gráfico de densidade [Figura 9] foi utilizado para exibir a distribuição dos pagamentos realizados pela prefeitura para uma determinada categoria e ano. Como a variável de interesse, valores dos pagamentos, é quantitativa, o gráfico de densidade é mais adequado que um histograma para visualizar a distribuição. Percebe-se que, para a categoria Administração, no ano de 2021, a distribuição dos valores é enviesada à direita, e concentrada em valores menores que R\$21.000,00, possuindo ainda uma cauda longa à esquerda.

**FIGURA 9.** Gráfico de densidade para os valores de pagamentos realizados na categoria Administração, no ano de 2021.

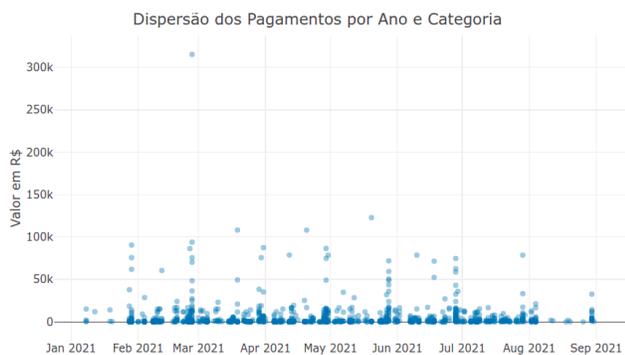


Por fim, é fornecido um gráfico de dispersão [Figura 10] para que seja possível ter uma visão mais granular de como os gastos se comportam ao longo do ano. Os dados foram também agrupados por categoria e ano, sem quaisquer transformações adicionais. É possível inspecionar os pagamentos realizados mês a mês, bem como ver a data dos pagamentos cujos pontos foram sinalizados como possíveis *outliers* no gráfico de *boxplot*.

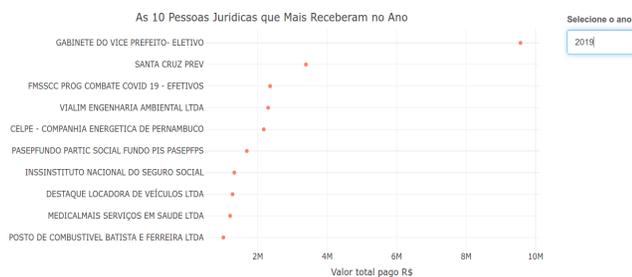
Por fim, a última seção da ferramenta responde a pergunta “Qual CNPJ mais recebeu dinheiro da prefeitura por ano?” [Figura 11]. Para isso, o conjunto de dados foi filtrado de modo que contivesse

apenas os pagamentos realizados à pessoa jurídica, em seguida fez-se o agrupamento desses valores por CNPJ e ano, e uma coluna sintética para esse novo conjunto de dados foi criada, indicando a soma dos valores recebidos por cada CNPJ em cada ano.

**FIGURA 10.** Gráfico de dispersão dos pagamentos realizados na categoria Administração no ano de 2021.



**FIGURA 11.** CNPJs que mais receberam da prefeitura em 2019.



Informações relevantes extraídas dessa visualização foram, por exemplo, que no ano de 2019 o CNPJ que mais recebeu dinheiro da prefeitura foi o correspondente ao favorecido Gabinete do Vice Prefeito - Eletivo, equivalente a um total de R\$9.57 milhões, sendo, para fins de comparação, aproximadamente 2.8 vezes maior que o valor recebido pela construtora Vialim Engenharia Ambiental LTDA. Outro ponto interessante é o valor gasto com gasolina onde, no ano de 2019, um único posto de gasolina, Posto de Combustível Batista e Ferreira LTDA, recebeu um total de R\$1.01 milhões, estando entre os dez CNPJs que mais receberam. O valor pago pela prefeitura a esse mesmo posto até o dia 24/08/2021, foi de R\$1.15 milhões, esse valor é curioso por não ser muito maior que o valor registrado em 2019, chamando atenção porque em 2021 o Brasil ainda passa por uma pandemia e espera-se que carros do SAMU (Serviço de Atendimento Móvel de Urgência) circulem mais, o que aumentaria o gasto com combustível. Além disso, o ano de 2021 apresentou aumentos substanciais no preço dos combustíveis e, segundo o site da Petrobras [11], o valor de revenda do litro da gasolina estava entre R\$5.792 e R\$6.648 para a semana de 26/09/2021 à 02/10/2021 [12].

### 3. Avaliação

Para avaliar a ferramenta produzida por este trabalho, foi desenvolvido um questionário com o objetivo de mensurar o

impacto da aplicação percebido pelos usuários. Esse questionário foi realizado através da plataforma do Google, e enviado por WhatsApp para os usuários juntamente com o link para a ferramenta. Trinta residentes da cidade de Santa Cruz do Capibaribe foram convidados para testar o Monitora SCC e responder a seguinte pergunta contida no questionário:

1. Numa escala de 1 à 5, onde 1 significa discordo totalmente, e 5 concordo totalmente, avalie as seguintes afirmações:
  - a. Consegui compreender completamente as explicações contidas nos textos do Monitora SCC.
  - b. Minha visão sobre a administração da cidade mudou após ter acesso ao Monitora SCC.
  - c. Achei que as visualizações contidas no Monitora SCC foram eficientes para transmitir as informações desejadas.
  - d. Minha curiosidade sobre como a prefeitura da cidade de Santa Cruz do Capibaribe utiliza seus recursos financeiros aumentou após ter acesso ao Monitora SCC.

Os resultados para cada uma das afirmações acima foram os seguintes:

- a. 81.8% das pessoas deram nota 5 (concordam totalmente), e 18.2% das pessoas deram nota 4.
- b. 72.7% das pessoas deram nota 5, 18.2% nota 4, e 9.1% nota 3.
- c. 81.8% das pessoas deram nota 5, e 18.2% nota 4.
- d. 36.4% das pessoas deram nota 5, 36.4% nota 4 e 27.3% nota 3.

Apesar do tamanho da amostra de usuários que participaram da pesquisa ter sido pequeno demais para ter relevância estatística, foi um resultado bastante positivo apontando que, para as pessoas que utilizaram a ferramenta, todas parecem ter extraído algum tipo de valor.

## 4. Experiência

Desenvolver uma aplicação que possui um impacto social é, por si só, uma experiência bastante enriquecedora. Ademais, o contato com as tecnologias utilizadas nesse trabalho definitivamente agregou bastante valor, em especial conhecer e estudar o pacote Shiny, que abre um leque de oportunidades únicas para compartilhar análise de dados e engajar as pessoas que estão lendo, por meio da interatividade promovida.

### 4.1 Processo de Desenvolvimento

O processo de desenvolvimento teve início com a avaliação de quais dados provenientes do portal da transparência seriam utilizados. Em seguida, foi necessário elaborar uma estratégia para coletar os dados de interesse, o que incluiu analisar quais tecnologias seriam mais adequadas para a realização da tarefa. Com os dados coletados, foram formuladas perguntas a serem respondidas através da análise do conjunto de dados, bem como quais visualizações seriam as mais apropriadas para auxiliar nas

respostas. Por último, foram definidas as tecnologias que viriam a ser empregues na implementação da aplicação *web*, que foi então desenvolvida de acordo com as especificações geradas através dos passos anteriores.

## 4.2 Principais Desafios

Definitivamente, os principais desafios se concentraram no processo de coleta de dados. Como comentado em seções passadas, um dos primeiros obstáculos encontrados foi o fato de o portal da transparência possuir conteúdos carregados dinamicamente na página, o que tornou mais complexo o uso exclusivo do Scrapy para navegar no portal, pois o *framework* não é muito bem adaptado para lidar com essas situações. A primeira tentativa de solução foi buscar descobrir a requisição HTTP geradora da página de interesse (página que contém o *iframe*<sup>13</sup> que lista os pagamentos realizados pela prefeitura) uma vez que, possuindo a URL e demais *headers*<sup>14</sup> da requisição, seria possível realizar a mesma usando apenas o Scrapy, não sendo necessário percorrer o portal da transparência desde a página inicial até a de interesse, o que resolveria a questão de o portal usar elementos que carregam de forma dinâmica no DOM<sup>15</sup>, *Document Object Model*. A requisição foi descoberta e foi possível replicá-la com sucesso pelo Postman<sup>16</sup>, porém essa mesma requisição era negada quando realizada pelo Scrapy, resultando em uma resposta HTTP de *status* 500. A segunda tentativa foi então utilizar um segundo *framework*, o Selenium, que além de disponibilizar diversos meios para aguardar que certas condições sejam cumpridas para continuar o processamento do portal (elementos estejam carregados, por exemplo), fornece também um *web driver* que permite acompanhar, de forma mais agradável, o andamento do trabalho realizado pelo *crawler*; o que facilita o processo de depuração.

Um outro desafio foi que existe um tempo máximo de permanência na página “Despesas Detalhadas Diárias”, onde encontra-se a seção de pagamentos, de 15 minutos. Quando essa sessão expira, o usuário é redirecionado para a página inicial do portal da transparência do município. Esse tempo de 15 minutos não é o suficiente para o *crawler* capturar as todas as informações disponíveis na tabela de listagem dos pagamentos, ou as informações adicionais mais detalhadas, que são encontradas numa janela diferente [Figura 4] após o usuário clicar no *link* localizado na coluna de Empenho da tabela de pagamentos [Figura 3]. Como explicado na seção 2.1, para solucionar esse problema, a captura dos dados foi dividida em duas partes, uma que encarrega-se de coletar os dados da tabela, e outra que coleta os dados detalhados dos pagamentos. Para cada parte, o Spider

responsável é executado com o objetivo de coletar um segmento do conjunto de dados, e o processo é repetido até que todos os dados sejam coletados. Para o Spider responsável pela tabela de pagamentos, cada iteração tinha duração em torno de 13 minutos, e era possível capturar 10.000 linhas da tabela por vez. Já para o segundo Spider, que é mais demorado devido ao fato de ter que abrir, para cada linha da tabela, uma janela contendo os detalhes do pagamento, além de ter que esperar essa janela nova carregar, só era possível capturar os dados para 410 linhas da tabela de pagamentos por vez. Essa solução de segmentar a captura dos dados com o objetivo de burlar o tempo máximo de permanência gerou, no entanto, um segundo desafio; se apenas um processo em um computador executa um Spider, para concluir a coleta dos dados da tabela de pagamentos, seriam necessárias 1.75 horas, e para concluir a coleta dos dados detalhados, 37.43 horas. Na tentativa de otimizar esse tempo, a solução foi executar processos em paralelo para os dois Spiders. Cada processo estava responsável por um intervalo de páginas, por exemplo: o processo 1 captura as páginas de 1 à 21, o processo 2 as páginas de 22 à 43, e assim por diante. Ainda existiu um trabalho manual de checar se houveram erros sempre que um processo terminava de executar para determinada página, uma vez que se algo inesperado acontece e para a execução do *crawler*, informações deixam de ser coletadas, e o processo deveria executar novamente para a mesma página. Algo que costumava provocar termos inesperados de execução era quando uma janela demorava demais a carregar, fazendo com que o *crawler* tentasse encontrar elementos que ainda não estavam no DOM, uma vez que a requisição não tinha sido finalizada, gerando um erro.

## 4.3 Trabalhos Futuros

Como trabalhos futuros, o principal seria construir um software observador que monitora atualizações nos dados de pagamentos disponíveis no portal da transparência de Santa Cruz do Capibaribe, e reflete essas atualizações nos dados coletados para este trabalho. O observador deve não apenas capturar as informações novas adicionadas, como também checar se houveram mudanças nos dados já existentes no conjunto de dados extraídos por este trabalho. Indo além, também é possível adicionar várias outras visualizações e análises na aplicação *web* Monitora SCC como, por exemplo, analisar de forma esmiuçada os gastos com educação, observando as subcategorias das despesas ou analisar melhor a evolução da folha de pagamento dos funcionários das escolas da rede pública municipal.

## 5. Referências

- [1] Portal da Transparência de Santa Cruz do Capibaribe, PE. Disponível em: [https://santacruzdocapibaribe.pe.tenosoftsistemas.com.br/portal/v81/p\\_index\\_entidades/?municipio=47&represent=1](https://santacruzdocapibaribe.pe.tenosoftsistemas.com.br/portal/v81/p_index_entidades/?municipio=47&represent=1). Acesso em: 23 de Setembro de 2021.
- [11] Petróleo Brasileiro S.A., Petrobras, é uma empresa de capital aberto cujo acionista majoritário é o Governo Brasileiro.
- [12] Preços de venda de combustíveis. Disponível em: [https://petrobras.com.br/pt/nossas-atividades/precos-de-venda-de-combustiveis/?gclid=CjwKCAjwtfqKBhBoEiwAZuesiHayfla82bp\\_-TY13eufuRiz3F\\_2W75Qsacl2dFoFihS4cHhL7xFgBoCLE8QAvD\\_BwE](https://petrobras.com.br/pt/nossas-atividades/precos-de-venda-de-combustiveis/?gclid=CjwKCAjwtfqKBhBoEiwAZuesiHayfla82bp_-TY13eufuRiz3F_2W75Qsacl2dFoFihS4cHhL7xFgBoCLE8QAvD_BwE). Acesso em: 08 de Outubro de 2021.

<sup>13</sup> Um *iframe* é uma *tag* HTML que embute um outro documento HTML dentro do documento corrente.

<sup>14</sup> *Headers* permitem que o servidor e o cliente passem informações adicionais nas requisições.

<sup>15</sup> Segundo a definição disponível na página [MDN Web Docs](#), DOM é a representação de dados dos objetos que compõem a estrutura e o conteúdo de documentos na web.

<sup>16</sup> Postman é uma plataforma API para construir e usar outras APIs. Sítio oficial: <https://www.postman.com/>

[16] Operação Serenata de Amor. Disponível em: <<https://serenata.ai/>>. Acesso em: 13 de Agosto de 2021.

[17] IDEB - Inep. Disponível em: <<http://ideb.inep.gov.br/resultado/>>. Acesso em: 01 de Outubro de 2021.

[18] From Data to Viz. Disponível em: <<https://www.data-to-viz.com/>>. Acesso em: 25 de Setembro de 2021.