



UNIVERSIDADE ESTADUAL DO PIAUÍ
CAMPUS DRA. JOSEFINA DEMES
CURSO DE GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

WILLIAN KENNEDY DOS SANTOS BRITO

**FLORIANO STATVIEW: UMA PLATAFORMA WEB DE ANÁLISE UTILIZANDO BIG
DATA E INFORMAÇÕES DO IBGE SOBRE A CIDADE DE FLORIANO**

FLORIANO

2025

WILLIAN KENNEDY DOS SANTOS BRITO

FLORIANO STATVIEW: UMA PLATAFORMA WEB DE ANÁLISE UTILIZANDO BIG
DATA E INFORMAÇÕES DO IBGE SOBRE A CIDADE DE FLORIANO

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciências da Computa-
ção da Universidade Estadual do Piauí, como
requisito parcial à obtenção do grau de bacharel
em Ciências da Computação.

Orientadora: Prof. Ma. Edna Yoshiko
Senzako.

FLORIANO

2025

Floriano StatView: Uma Plataforma Web de Análise Utilizando Big Data e Informações do IBGE Sobre a Cidade de Floriano

Willian Kennedy dos S. Brito¹, Edna Yoshiko Senzako¹

¹ Ciência da Computação – Universidade Estadual do Piauí (UESPI)
Floriano – PI – Brasil

`willian.k.dos.s.b@aluno.uespi.com.br`

Abstract. *This work presents the development of the Floriano StatView web platform, aimed at the visualization and analysis of public data from the municipality of Floriano, Piauí. The initiative emerged from the identification of a gap in the availability of tools that facilitate access to and understanding of local socioeconomic information. To this end, a literature review was conducted, followed by the implementation of the platform using Python, Plotly, Dash, and Werkzeug. The platform provides charts and indicators in the areas of economy, demographics, and education, with the objective of supporting public administration and facilitating access to information. Thus, the proposed solution contributes to fostering a data-driven culture in the municipality.*

Resumo. *Este trabalho apresenta o desenvolvimento da plataforma web Floriano StatView, voltada à visualização e análise de dados públicos do município de Floriano, Piauí. A iniciativa surgiu da identificação de uma lacuna quanto à disponibilidade de ferramentas que facilitem o acesso e a compreensão das informações socioeconômicas locais. Para isso, realizou-se uma revisão bibliográfica e em seguida a implementação da plataforma utilizando Python, Plotly, Dash e Werkzeug. A plataforma disponibiliza gráficos e indicadores das áreas de economia, demografia e educação, com o objetivo de apoiar a gestão pública e facilitar o acesso à informação. Dessa forma, a solução proposta contribui fomentando uma cultura orientada a dados no município.*

1. Introdução

Segundo o Mackynsey Global Institute (2011), empresas geram grandes volumes de dados transacionais que incluem informações sobre clientes, fornecedores e operações. Além das empresas, outras fontes contribuem significativamente para a geração de dados em larga escala, como redes sociais, sensores e dispositivos conectados a internet.

Esse volume, a variedade de dados e a velocidade com que eles são gerados, são as principais características de *Big Data* (TAURION, 2013). Ademais dessas 3 características, existem outros aspectos a serem considerados, como a veracidade dos dados e o valor que os dados podem entregar (CARVALHO; MENEZES; BONIDIA, 2024). Juntas, essas características formam os 5 V's do *Big Data*.

Para Taurion (2013), *Big Data* pode ser analisada para produzir informações valiosas a todos os setores da raça humana, já existindo iniciativas nas mais diversas áreas que

utilizam *Big Data* para produzir algo de valor. Ponto ressaltado por Marquesone (2017), pois a autora destaca que o mais importante aspecto de *Big Data* é a capacidade de extrair valor a partir dos dados.

Tendo isto em vista, já se observam aplicações relevantes de *Big Data* em diferentes domínios, como o estudo de Xavier et al. (2024), em que são utilizadas técnicas de *Big Data* para investigar a relação entre indicadores de contexto e indicadores estruturais de cidades do estado de São Paulo com o resultado da campanha contra a dengue.

Enquanto na área da economia, Libório, Ekel e Martins (2022) verificam o impacto que a redução da quantidade de pessoas visitando locais de varejo/entretenimento teve nos casos de COVID-19 e na taxa de desemprego. Os autores também estudam a relação do interesse sobre o futuro, diplomas universitários, pesquisas e publicações científicas com a renda média das famílias brasileiras.

Apesar da ampla utilização de *Big Data* em diversas áreas, observa-se a escassez de iniciativas que forneçam ferramentas de análise voltadas especificamente à cidade de Florianópolis, tendo isso em vista, o objetivo deste artigo é desenvolver uma plataforma web que permita a visualização e análise dinâmica de dados da cidade de Florianópolis, disponibilizando gráficos interativos que facilitem a compreensão e acesso às informações locais.

Este artigo está estruturado da seguinte forma: a Seção 2 apresenta a fundamentação teórica; a Seção 3 descreve os métodos adotados; a Seção 4 apresenta os resultados obtidos; e, por fim, a Seção 5 expõe as conclusões do trabalho.

2. Fundamentação Teórica

Big Data é um campo de estudo multidisciplinar que abrange diversas áreas do conhecimento, como computação, estatística e disciplinas específicas relacionadas a cada tipo de problema. Além disso, diferentes profissões estão envolvidas no trabalho com *Big Data*, sendo a Engenharia de Dados e a Ciência de Dados algumas das mais relevantes. Assim, esta seção tem como objetivo explicar o que é *Big Data*, Engenharia de Dados, Ciência de Dados e suas características, visando esclarecer a relação entre estes três conceitos.

2.1. Engenharia de Dados

Muito se confunde Engenharia de Dados e Ciência de Dados, porém elas são áreas distintas que se complementam. Reis e Housley (2022) definem Engenharia de Dados como o desenvolvimento, implementação e manutenção de sistemas, processos que coletam dados e produzem informações que serão úteis em usos futuros. Já Gavin (2020) afirma que a essência da Engenharia de Dados é a movimentação, manipulação e o gerenciamento de dados.

A Engenharia de Dados, independente da definição, segue processos que envolvem os dados, sendo o principal desses processos o gerenciamento do ciclo de vida dos dados. Este ciclo é composto de três passos chave: *Extract* (Extrair), *Transform* (Transformar) e *Load* (Carregar). Esse ciclo é comumente chamado de processo ETL, e, no contexto de *Big Data*, pode ter sua ordem alterada e ser chamado de processo ELT (GAVIN, 2020).

Os processos da Engenharia de Dados têm como objetivo tornar *datasets* úteis e acessíveis para os consumidores de dados, que podem ser pessoas ou sistemas que ne-

cessitam de acesso aos dados de uma organização (EAGAR, 2023). Esses consumidores podem ser diretores e outros funcionários, analistas e cientistas de dados e até mesmo outros sistemas.

2.2. Ciência de Dados

Ciência de Dados é um termo que possui diversas definições. Escovedo e Koshiyama (2020) definem Ciência de Dados como a coleta de dados de várias fontes para análise, com o objetivo de auxiliar na tomada de decisões utilizando grandes volumes de dados. Para cumprir esse propósito, a Ciência de Dados usa métodos científicos, processos, algoritmos e sistemas para extrair conhecimento e percepções dos dados, tratando-se de uma área multidisciplinar (CUADRADO-GALLEGO; DEMCHENKO, 2020), com os principais conhecimentos necessários sendo das áreas de Machine Learning (ML), Estatística, Programação e Banco de Dados (WAGH; BHENDE; THAKARE, 2021).

O conhecimento extraído é usado para resolver problemas, resultando em aplicações de Ciência de Dados, cada uma dessas aplicações podem ser classificadas por algum critério, os mais comuns sendo: a área de conhecimento do problema abordado, o setor econômico ou a tarefa de modelagem resolvida, que pode ser preditiva, descritiva ou prescritiva (CARVALHO; MENEZES; BONIDIA, 2024).

A resolução desses problemas traz benefícios para o negócio, alguns dos benefícios que podem ser trazidos pela Ciência de Dados são: Percepções sobre o negócio; Produtos inteligentes, como sistemas de recomendação personalizados; Construção de *Frameworks* de Análise; entre outros benefícios (CADY, 2021).

2.3. Big Data

Com a popularização da internet, surgiram diversos serviços online que auxiliam atividades humanas, a exemplo, existem serviços que auxiliam no planejamento de viagens, na realização do trabalho, no lazer, na compra, entre outras atividades. Tantos serviços geram muitos dados, o que levou ao surgimento da chamada “era dos dados” (MARQUESONE, 2017).

Com a chegada dessa era, o volume de dados produzidos aumentou, esse grande volume de dados é uma das características de *Big Data*, mas não é a única. Juntamente ao volume, a variedade, velocidade, veracidade e valor formam o conceito dos 5 V’s e caracterizam o conceito de Big Data (MORAIS et al., 2018).

Essas características fazem com que processar *Big Data* usando sistemas tradicionais de banco de dados seja difícil. Isso ocorre porque esses dados são complexos e os sistemas tradicionais apresentam limitações na quantidade de dados que conseguem processar, além de se tornarem muito caros e lentos à medida que o volume cresce (BALUSAMY et al., 2021).

Por isso, Carvalho, Menezes e Bonidia (2024) entendem que *Big Data* se refere também ao estudo e desenvolvimento de tecnologias para lidar com dados, envolvendo a coleta, armazenamento, processamento e transmissão de dados. Algumas das tecnologias criadas para lidar com *Big Data* são o Hadoop, Spark, MapReduce e Bancos de Dados NoSQL.

2.4. Artigos Relacionados

O Instituto Brasileiro de Geografia e Estatística (IBGE) é o principal provedor de dados e informações do país (IBGE, 2024), sendo amplamente utilizado em estudos acadêmicos e aplicações. Diversos estudos também demonstram como os dados disponibilizados pelo IBGE podem ser utilizados juntamente a abordagens de *Big Data* para entender e resolver os mais diversos problemas.

Nascimento, Antunes e Lima (2022) demonstram em seu estudo como o Sistema IBGE de Recuperação Automática (SIDRA) disponibiliza dados que podem auxiliar no ensino e aprendizagem de Geografia, destacando como este recurso é uma ferramenta formidável de ensino e aprendizagem.

Quanto a análise de infraestrutura urbana, Gusmão (2018) utiliza dados do SIDRA para analisar a infraestrutura do esgotamento sanitário na área urbana de Marituba, cidade do estado do Pará. Com isso foi possível identificar a inadequação do esgotamento sanitário nessa cidade e também que o acesso a rede de esgoto e de fossas sépticas é desigual, com apenas uma pequena parcela da população tendo acesso a esse tipo de recurso. Esse estudo demonstra como os dados do IBGE podem ajudar na identificação de problemas e características de uma localidade.

Quanto ao uso de *Big Data* e Ciência de Dados, Saldanha, Barcellos e Pedroso (2021) afirmam que não é difícil pensar em possibilidades de uso em estudos populacionais e de saúde, alguns desses usos que eles citam são: monitoramento; predição de casos e situações de saúde-doença na população, bem como qual a relação dos casos e situações com suas causas socioambientais e demográficas. Neste estudo eles ainda elencam diversas iniciativas que utilizam *Big Data* e Ciência de Dados para estimar dados e indicadores de saúde-doença por meio de dados públicos.

Xavier et al. (2024) também demonstram o uso de *Big Data* e Ciência de Dados na análise do resultado da campanha contra a Dengue no estado de São Paulo, identificando quais as características em comum das cidades que tiveram o pior resultado. Os autores ainda destacam que a gestão pública poderia levar em conta as informações encontradas para melhor direcionar os investimentos, levando em conta as necessidades de cada localidade.

Esses estudos mostram como os dados do IBGE e o uso *Big Data* podem auxiliar na análise das características de uma região, permitindo análises mais profundas que as técnicas tradicionais.

3. Métodos

O estudo consiste na criação de um protótipo de uma plataforma denominada Floriano StatView, cujo objetivo é fornecer visualizações de dados sobre a cidade de Floriano. Como preparação inicial, foi realizada uma revisão bibliográfica sobre os temas *Big Data*, Ciência de Dados, Engenharia de Dados e também sobre outras plataformas e pesquisas com objetivos semelhantes.

Concluída a revisão, iniciou-se o desenvolvimento da plataforma, dividido em duas fases principais: preparação e implementação. Na fase de preparação, foram realizados testes, demonstrações e estudos sobre a plataforma SIDRA e as tecnologias se-

lecionadas. Nesse estágio, também ocorreu a coleta de dados para testes e a criação de protótipos utilizando diferentes arquiteturas.

Com os aprendizados obtidos, iniciou-se a fase de desenvolvimento, que envolveu a definição dos dados a serem coletados e exibidos, a escolha da arquitetura do código e, finalmente, a implementação da plataforma. Definiu-se que a primeira versão exibiria informações gerais sobre o município.

Destaca-se que, por se tratar de um protótipo, a quantidade de dados processados não é volumosa, não caracterizando um sistema de *Big Data* em termos operacionais. Ainda assim, os conceitos de *Big Data* foram importantes como referencial teórico e auxiliaram na definição de boas práticas para a coleta, organização e visualização das informações.

Os dados selecionados para exibição na plataforma, bem como seus respectivos anos de disponibilidade, estão apresentados na Tabela 1.

Tabela 1. Dados exibidos na plataforma e anos de disponibilidade

Dado exibido	Anos disponíveis
População total de Floriano	2010, 2022; estimativas nos demais anos
PIB total de Floriano	Até 2021
PIB <i>per capita</i>	Até 2021
Distribuição da população (zona urbana e rural)	2022
Distribuição racial da população	2010, 2022
Taxa de alfabetização (município, estado e país)	2022
Cidades mais populosas do Piauí	2010, 2022; estimativas nos demais anos
Distribuição etária da população de Floriano	2010, 2022
Produção agrícola de Floriano (10 maiores)	2010 a 2023

As tecnologias utilizadas para a construção do *dashboard* foram: Python como linguagem principal; Flask¹ como *framework web*, pela simplicidade de suas rotas e integração com Dash; Dash² e Plotly³ para construção de gráficos interativos e responsivos; e o *middleware Werkzeug*⁴ para a integração entre Flask e Dash.

A arquitetura do sistema seguiu o modelo de três camadas. Da camada inferior para a superior: camada de dados, responsável pela coleta e tratamento das informações; camada de gráficos, encarregada de montar e disponibilizar os gráficos e métricas formatadas; e camada de layout, que organiza o *dashboard* a partir dos elementos fornecidos pela camada de gráficos. Cada camada se comunica apenas com a imediatamente inferior, garantindo uma estrutura modular e organizada.

Em relação ao tratamento dos dados coletados, seguiu-se um padrão: inicialmente foram descartadas colunas irrelevantes, como códigos de variáveis e nomes repetitivos, retornados pela API do SIDRA. A primeira tupla dos dados, que apenas descrevia o título das colunas, também foi removida. Em seguida, todos os valores numéricos foram convertidos de *strings* para seus respectivos tipos numéricos.

¹Documentação do Flask: <https://flask.palletsprojects.com/en/stable/>

²Documentação do Dash: <https://dash.plotly.com/>

³Documentação do Plotly: <https://plotly.com/python/>

⁴Documentação do Werkzeug: <https://werkzeug.palletsprojects.com/en/stable/middleware/>

Algumas consultas específicas exigiram tratamentos adicionais. Para o PIB total da cidade, foi necessário multiplicar os valores retornados por mil, devido à unidade utilizada pela API (milhares de reais). Para a identificação das cidades mais populosas do estado, os dados foram ordenados de forma decrescente pelo valor da população, selecionando-se apenas os dez municípios com maior população.

O último tratamento personalizado foi aplicado aos dados referentes à produção agrícola de Floriano. Inicialmente, foram removidos os registros com unidades de medida diferentes de toneladas, devido à sua baixa representatividade e à incompatibilidade na comparação — apenas dois produtos se enquadravam nessa condição: o abacaxi e o coco-da-baía. Em seguida, os dados passaram pelos seguintes tratamentos: descarte de produtos com valores zerados; agrupamento por ano; ordenação pela quantidade produzida; e seleção das dez maiores produções.

Por fim, calculou-se o PIB *per capita* de Floriano. Embora esse dado não esteja disponível diretamente pela API, ele pode ser obtido dividindo-se o PIB total pela população total, conforme apresentado na equação abaixo:

$$\text{PIB}_{\text{pc}} = \frac{\text{PIB}_{\text{total}}}{P} \quad (1)$$

4. Resultados

Seguindo os procedimentos descritos anteriormente, foi desenvolvida uma versão inicial da plataforma, focada em apresentar dados gerais da cidade. Esta versão proporciona uma visão ampla e abrangente, com informações sobre população, etnia, economia e educação. A seguir, descreve-se o *dashboard* (Figura 1).

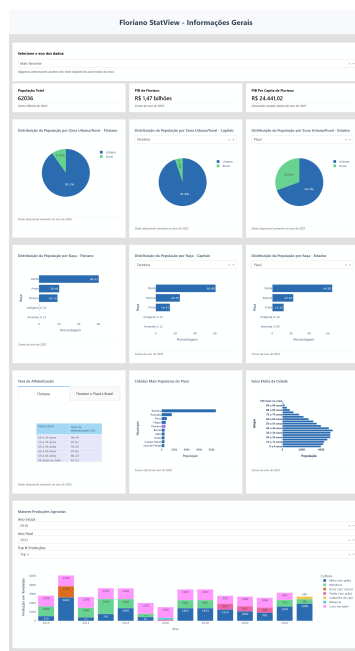


Figura 1. *Dashboad* de informações gerais.

Após o cabeçalho contendo o título do *dashboard*, há um *dropdown* (Figura 2) que permite a seleção do ano para o qual os dados devem ser exibidos. Uma alternativa

considerada seria permitir a escolha do ano diretamente em cada *card* de informação; entretanto, optou-se pela seleção global no início do *dashboard* para proporcionar uma experiência de usuário mais simples e direta, ainda que menos personalizável.

O período disponível para seleção abrange de 2010 até 2025, ou até o dado mais recente disponível. Contudo, nem todas as informações estão disponíveis para todos os anos nesse intervalo. Quando necessário, a plataforma busca exibir dados referentes ao ano mais próximo do selecionado.

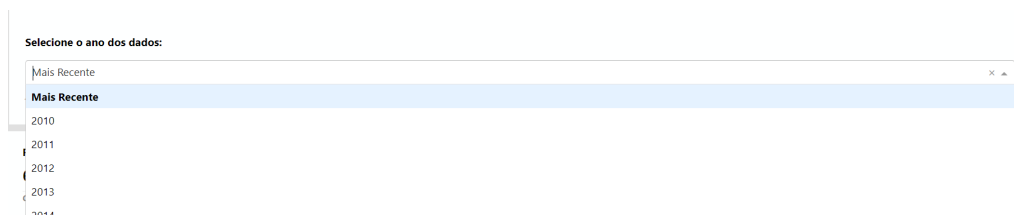


Figura 2. Dropdown de seleção de ano.

Abaixo do *dropdown* de seleção de ano, são exibidos *cards* que apresentam diferentes métricas sobre a cidade (Figura 3). A primeira métrica corresponde à população total, embora essa informação possua uma limitação: apenas os anos de 2010 e 2022 contam com valores oficiais disponibilizados pelo IBGE. Para os demais anos, são utilizados dados estimados disponíveis na plataforma.

A segunda métrica apresentada é o Produto Interno Bruto (PIB) total da cidade, com dados disponíveis até o ano de 2021. Neste *card*, os valores são formatados nas unidades de milhões e bilhões. Uma possível melhoria futura seria permitir que o usuário escolhesse entre visualizar os valores formatados ou os valores completos.

A terceira métrica disponível é o PIB *per capita* da cidade, calculado conforme o procedimento descrito anteriormente.

População Total 62036 <small>Censo Oficial de 2022</small>	PIB de Floriano R\$ 1,47 bilhões <small>Censo do ano de 2021</small>	PIB Per Capita de Floriano R\$ 24.441,02 <small>Calculado usando dados do ano de 2021</small>
--	--	---

Figura 3. Cards de métricas.

A linha seguinte (Figura 4) apresenta três *cards* com a distribuição da população entre as zonas rural e urbana em diferentes contextos geográficos para fins comparativos. O primeiro exibe os dados referentes a Floriano; o segundo, às capitais das unidades federativas brasileiras; e o terceiro, aos estados do país.

Os próximos gráficos exibidos (Figura 5) tratam da distribuição racial da população, seguindo a mesma lógica dos gráficos anteriores. O primeiro exibe os dados de Floriano; o segundo, refere-se às capitais das unidades federativas brasileiras; e o terceiro, aos estados do país.

Na Figura 6, observam-se os *cards* da terceira linha. A parte superior da figura exibe um gráfico de linhas comparativo entre o município, o estado e o país. Na parte inferior, há três *cards*: o primeiro apresenta uma tabela com a taxa de alfabetização da

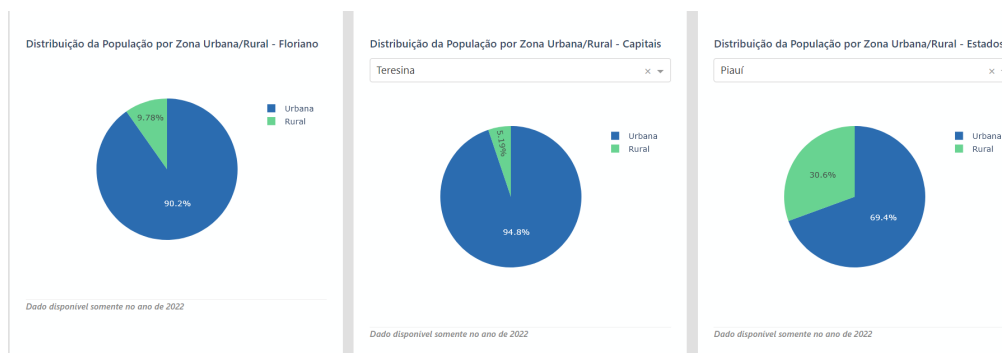


Figura 4. Gráficos de distribuição da população por Zona Rural e Urbana.



Figura 5. Gráficos das distribuições da população por raça.

cidade por faixas etárias; o segundo mostra as dez cidades mais populosas do Piauí, utilizando dados oficiais e estimativas, assim como o *card* de população da cidade de Floriano; e o terceiro exibe a distribuição etária da cidade.

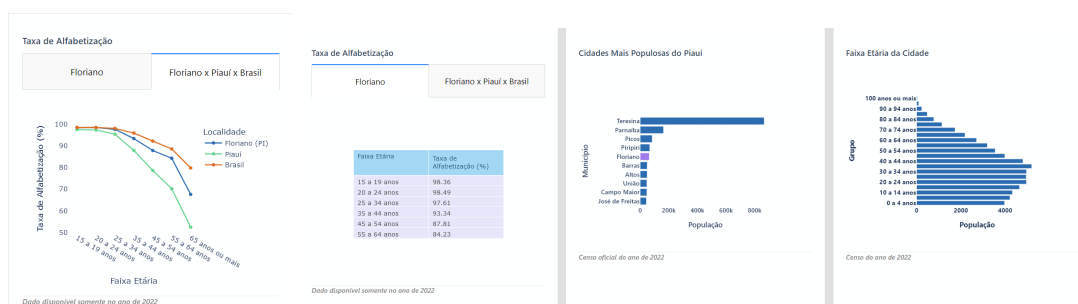


Figura 6. Cards com informações sobre a taxa de alfabetização, cidades mais populosas do Piauí e Faixa Etária da cidade.

Na Figura 7, observa-se o último gráfico do *dashboard*, que apresenta as maiores produções agrícolas de Floriano. Nesse gráfico, o usuário pode selecionar o período dos dados exibidos, com o intervalo máximo compreendendo os anos de 2010 a 2023 (dados mais recentes disponíveis). Além disso, é possível escolher a quantidade de produtos a serem exibidos, desde o de maior produção até as dez maiores produções agrícolas.

Tendo em vista os resultados obtidos e os procedimentos executados, identificam-se duas limitações neste estudo. A primeira refere-se à disponibilidade de dados na plataforma SIDRA: algumas pesquisas não são realizadas anualmente, e algumas não incluem

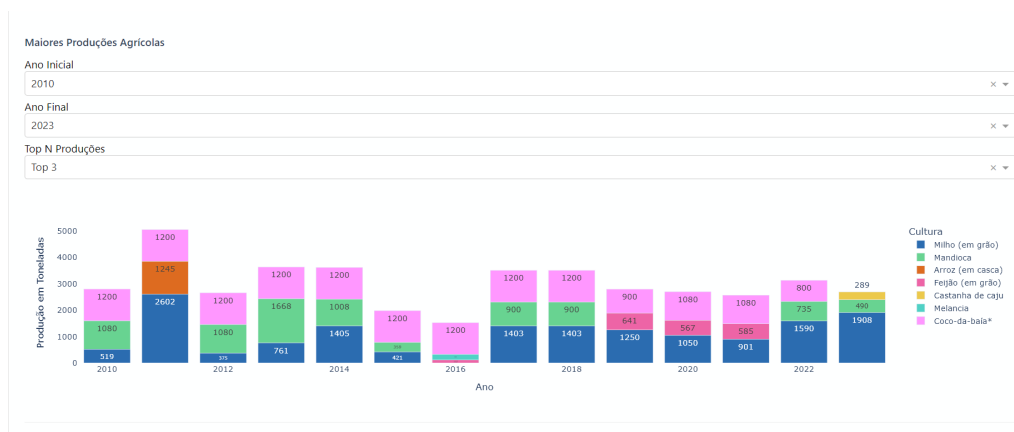


Figura 7. Gráfico de produção agrícola de Floriano

o município de Floriano. A segunda, e mais significativa, diz respeito à ausência de parcerias com instituições públicas locais, o que impossibilitou a elaboração de *dashboards* personalizados que atendessem às necessidades específicas de cada entidade.

5. Conclusão

Este artigo se propôs a enfrentar a escassez de ferramentas de análise de dados voltadas à cidade de Floriano, Piauí. Seu objetivo central foi o desenvolvimento de uma plataforma *web* que permitisse a visualização e análise dinâmica de dados municipais, facilitando a compreensão e o acesso às informações locais.

Com base nesse propósito, foi desenvolvida a plataforma Floriano StatView, cuja funcionalidade principal consiste na disponibilização de dados das áreas de economia, demografia e educação por meio de gráficos, tabelas e *cards* informativos. Considera-se que a plataforma atende ao objetivo proposto, uma vez que facilita a busca e a compreensão de dados específicos da localidade, preenchendo a lacuna identificada no acesso às informações municipais.

A principal contribuição deste estudo está na disponibilização de um recurso tecnológico capaz de colaborar de forma significativa com a tomada de decisão por parte da gestão pública, além de auxiliar em pesquisas acadêmicas e promover o acesso à informação de maneira clara e objetiva para os cidadãos. Dessa forma, busca-se fomentar a transparência administrativa e a consolidação de uma cultura orientada a dados no município. Uma demonstração do recurso pode ser encontrada em vídeo ⁵ e no GitHub com instruções de execução ⁶

No entanto, foram identificadas algumas limitações neste estudo: a indisponibilidade de dados para determinados tópicos e períodos, e a ausência de contato com agentes locais, o que dificultou a elaboração de visualizações personalizadas conforme as necessidades de figuras institucionais, como a administração pública.

Como sugestões para trabalhos futuros, propõe-se a incorporação de novas fontes de dados municipais, a coleta de *feedback* da população e da gestão local — com vistas

⁵Link para o vídeo: <https://youtu.be/MVInL2jA-RE>

⁶Link para o repositório: https://github.com/kennedy-brito/floriano_statview

à personalização e melhoria contínua da plataforma — e a realização de estudos de caso sobre o impacto do seu uso tanto pela população quanto pela administração pública. Além disso, a metodologia empregada pode ser adaptada ao desenvolvimento de plataformas semelhantes em outros municípios.

Em suma, este trabalho não apenas entregou uma solução funcional para uma necessidade identificada em Floriano, como também pavimentou o caminho para futuras pesquisas voltadas à ampliação do acesso e da análise de informações locais, contribuindo para uma gestão pública mais eficiente e para uma sociedade mais bem informada.

Referências

BALUSAMY, B. et al. *Big Data: Concepts, Technology, and Architecture*. 1. ed. [S.l.]: John Wiley & Sons, Inc., 2021.

CADY, F. *Data science : the executive summary : a technical book for non-technical professionals*. [S.l.]: WILEY-BLACKWELL, 2021. ISBN 9781119544081.

CARVALHO, A. C. P. L. F. de; MENEZES, A. G.; BONIDIA, R. P. *Ciência de Dados - Fundamentos e Aplicações*. 1. ed. Grupo GEN, 2024. ISBN 9788521638766. Disponível em: <https://grupogen.vitalsource.com/reader/books/9788521638766>.

CUADRADO-GALLEGO, J. J.; DEMCHENKO, Y. *The data science framework: A view from the EDISON project*. [S.l.]: Springer International Publishing, 2020. 1-194 p. ISBN 9783030510237.

EAGAR, G. *Data Engineering with AWS - Second Edition: Acquire the skills to design and build AWS-based data tr.* [S.l.]: Packt Publishing, 2023. ISBN 9781804614426.

ESCOVEDO, T.; KOSHIYAMA, A. *Introdução a data science : algoritmos de machine learning e métodos de análise*. [S.l.]: Casa do Código, 2020. ISBN 9788572540544.

GAVIN, L. *What Is Data Engineering? A Role for Data Science Enablers*. 2020. Disponível em: <http://oreilly.com>.

GUSMÃO, L. H. A. Avaliação da infraestrutura em esgotamento sanitário na área urbana de marituba (pa) evaluation of infrastructure in sanitary sewage in the urban area of marituba (pa). *GeoAmazonia*, v. 6, p. 161–176, 11 2018. ISSN 2358-1778. Disponível em: <http://www.geoamazonia.net/index.php/revista/indexeISSN:2358-1778>.

IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2024. Disponível em: <https://www.ibge.gov.br/acesso-informacao/institucional/o-ibge.html>.

LIBÓRIO, M. P.; EKEL, P. I.; MARTINS, C. A. P. da S. Economic analysis through alternative data and big data techniques: what do they tell about brazil? *SN Business & Economics*, Springer Science and Business Media LLC, v. 3, 12 2022.

Mackynsey Global Institute. *Big data: The next frontier for innovation, competition, and productivity*. 2011. Disponível em: www.mckinsey.com/mgi.

MARQUESONE, R. *Big Data: Técnicas e tecnologias para extração de valor dos dados*. [S.l.]: Casa do Código, 2017. ISBN 9788555192319.

MORAIS, I. S. de et al. *Introdução a Big Data e Internet das Coisas (IOT)*. [S.l.]: Sagah, 2018. ISBN 9788595027640.

NASCIMENTO, D. T. F.; ANTUNES Ítalo R.; LIMA, G. D. O. Sistema ibge de recuperação automática de dados: orientações para uso e o ensino-aprendizagem em geografia. *Revista de Geografia*, Universidade Federal de Pernambuco, v. 39, p. 172, 4 2022. ISSN 2238-6211.

REIS, J.; HOUSLEY, M. *Fundamentals of Data Engineering: Plan and Build Roubust Data Systems*. [S.l.]: O'Reylli Media, Inc, 2022. ISBN 9781098108304.

SALDANHA, R. de F.; BARCELLOS, C.; PEDROSO, M. de M. Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde? *Cadernos Saúde Coletiva*, FapUNIFESP (SciELO), v. 29, p. 51–58, 2021. ISSN 1414-462X.

TAURION, C. *Big Data*. Brasport Livros e Multimídia Ltda, 2013. ISBN 9788574526089. Disponível em: <www.brasport.com.br>.

WAGH, S. J.; BHENDE, M. S.; THAKARE, A. D. *Fundamentals of Data Science*. 1. ed. [S.l.]: Chapman and Hall/CRC, 2021. ISBN 1138336181.

XAVIER, F. et al. Big data-planetary health approach for evaluating the brazilian dengue control program. *Revista de Saúde Pública*, Universidade de São Paulo. Museu de Zoologia, v. 58, 2024. ISSN 00348910.