

BigDataFortaleza - Análise e Tomada de Decisão para o Desenvolvimento Urbano Sustentável

Rossana M. C. Andrade^{1 2 3 5}, Élcio Batista⁴, Ismayle S. Santos³, Allberson B. O. Dantas^{3 5}, Tales P. Nogueira^{3 5}, Pedro A. M. Oliveira^{3 6}, Pedro F. A. Neto³, Pedro R. L. Fernandes^{3 4}, Victória T. Oliveira^{3 6}

RESUMO

Serviços de cidades inteligentes, como aqueles ligados à saúde, economia, turismo, educação, têm se beneficiado com iniciativas de transformação digital, que promovem o desenvolvimento de soluções para ajudar a gestão pública na tomada de decisões. Cada uma dessas áreas possui suas próprias características e desafios. Entretanto, ao observar o contexto dinâmico da cidade, é necessário aos gestores públicos entender qual o impacto de suas decisões. Diante disso, este projeto visa desenvolver uma plataforma de análise de dados e predição utilizando Big Data e Machine Learning, chamada BigDataFortaleza, com o objetivo de ajudar no planejamento estratégico de soluções para o desenvolvimento urbano sustentável de Fortaleza, no contexto de atuação do Instituto de Planejamento de Fortaleza (Iplanfor). Um dos desafios que surgem nesse cenário é a obtenção e análise dos dados e predições para tomada de decisão com foco na melhoria da colaboração e engajamento dos cidadãos, em consequência, da melhoria da qualidade de vida, especialmente a longo prazo. O desafio de entender melhor a demanda por creches, ligado ao domínio da Primeira Infância, foi escolhido para ser explorado inicialmente, uma vez que a gestão inteligente de dados pode auxiliar na tomada de decisões que podem contribuir com a melhoria do desenvolvimento infantil. Até o momento, diversas análises exploratórias e preditivas foram realizadas para responder questões essenciais desse domínio que foram levantadas em discussões com gestores da Prefeitura Municipal de Fortaleza. Além disso, foram criadas propostas de arquitetura e de interfaces de usuário para a plataforma. Com a finalização da implementação do cenário da Primeira Infância, outros cenários devem ser abordados, por exemplo, desafios ligados à habitação. Como potencial inovador do projeto, destaca-se o desenvolvimento das análises com foco na predição de demandas futuras.

PALAVRAS-CHAVE: Cidades Inteligentes, Big Data, Machine Learning, Análise e Visualização de Dados, Computação em Nuvem.

1. INTRODUÇÃO

Fortaleza é uma das principais capitais do Nordeste, tendo em 2021 um total de 2.703.391 habitantes⁷, sendo a 5ª maior população do Brasil⁸. Ela é formada por uma área de 314.930 km² e está dividida administrativamente em 12 Secretarias Executivas Regionais que abrigam 121 bairros. Na prefeitura de Fortaleza, uma das autarquias municipais é o Instituto de Planejamento de Fortaleza (Iplanfor) que tem o propósito de atuar na geração do conhecimento, monitoramento e avaliação de políticas públicas, articulação do planejamento estratégico e fomento de iniciativas inovadoras. Destaca-se, então, o papel do Iplanfor no planejamento urbano e na modernização da cidade de Fortaleza.

Considerando a cidade Fortaleza e a atuação do Iplanfor, a coleta e análise dos dados é essencial para o diagnóstico e o apoio à tomada de decisões por parte dos gestores municipais. Por meio de uma análise inteligente desses dados, é possível, por exemplo, aprimorar a qualidade dos serviços prestados aos cidadãos ou prever o impacto de mudanças na infraestrutura da cidade. É essencial, portanto, uma infraestrutura robusta capaz de coletar, armazenar e processar dados provenientes de bases de dados de diferentes sistemas da cidade. Essa estrutura permitirá a análise dos dados da cidade de maneira eficiente e inteligente, suportando a tomada de decisões pelos gestores de Fortaleza e trazendo benefícios advindos da implantação de cidades inteligentes, tais como: integração entre os diferentes sistemas da cidade, sustentabilidade, melhoria na prestação dos serviços aos cidadãos, melhor eficiência dos recursos públicos e *insights* para criação de novas soluções (OSMAN et al., 2019, LI et al., 2019).

Rana *et al.* (2019) define cidades inteligentes reforçando o papel das Tecnologias de Informação e Comunicação (TIC) e as descreve como territórios tecnologicamente avançados e modernizados com certa capacidade intelectual para lidar com vários aspectos sociais, técnicos e econômicos baseado em técnicas de computação inteligente para desenvolver componentes inteligentes (e.g. prédios inteligentes, hospitais

¹ Cientista Chefe de Desenvolvimento Urbano Sustentável (Iplanfor)

² Departamento de Computação, Universidade Federal do Ceará (UFC)

³ Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)

⁴ Instituto de Planejamento de Fortaleza - Iplanfor

⁵ Instituto de Engenharias e Desenvolvimento Sustentável, Universidade da Integração Internacional da Lusofonia Afro-Brasileira (Unilab)

⁶ Programa de Mestrado e Doutorado em Ciência da Computação (MDCC) da Universidade Federal do Ceará (UFC)

⁷ <https://www.fortaleza.ce.gov.br/a-cidade>

⁸ https://www.fortaleza.ce.gov.br/images/0001/Apresentacao_Fortaleza_1.pdf

inteligentes) e serviços de qualidade superior. Importante citar também o trabalho das Nações Unidas em trazer as pessoas para o centro das transformações digitais com as cidades inteligentes⁹.

Cidades inteligentes apresentam três características essenciais (OSMAN et al. 2019): (i) uso de Tecnologia de Informação e Comunicação (TIC) para, entre outros pontos, conectar diferentes sistemas da cidade reunindo dados dos mesmos, os quais podem ser analisados para possibilitar a extração de informações úteis e *insights* sobre a cidade usando inteligência artificial; (ii) uma visão global da cidade, onde a inter-relação entre os principais sistemas da cidade deve ser considerada e nenhum destes sistemas funciona isoladamente; e (iii) sustentabilidade, que no contexto de cidades inteligentes está relacionada à habilidade de continuar e crescer sem deterioração significativa dos recursos. Deste modo, cidades inteligentes podem ser a solução para reduzir as contribuições negativas atreladas ao crescimento das cidades (LI et al. 2019) e tem o potencial de oferecer um melhor padrão de vida para futuras gerações (RAZMJOO et al., 2021).

Nesse sentido, o uso extensivo das tecnologias digitais em cidades inteligentes e a difusão dessas tecnologias na vida cotidiana das pessoas impulsionaram as interações humano-humano, humano-máquina, e máquina-máquina que produzem volumes massivos de dados (Big Data) (OSMAN et al. 2019). Para extrair conhecimentos dos conjuntos de dados gerados a partir dos vários domínios da cidade, é necessário que eles sejam integrados e analisados com o uso de técnicas de Inteligência Artificial.

Diante do que foi exposto, este artigo apresenta o projeto intitulado “BigDataFortaleza: Análise e Tomada de Decisão para o Desenvolvimento Urbano Sustentável”, que visa desenvolver uma plataforma de dados integrados, com capacidade de visualização e predição para fins de tomada de decisão e planejamento de estratégias por parte dos gestores da Prefeitura de Fortaleza. Espera-se que os resultados deste projeto contribuam para apoiar as atividades do Iplanfor, especialmente no que diz respeito: (i) à melhoria na tomada de decisões por parte dos gestores de Fortaleza; (ii) ao apoio na integração de dados entre os sistemas de Fortaleza; (iii) à criação de soluções inovadoras para os cidadãos de Fortaleza; (iv) à geração de novos conhecimentos sobre a cidade; e (v) ao fornecimento de insumos para uma Fortaleza Inteligente.

Este projeto iniciou em julho de 2022 e a prefeitura escolheu a CESPI (Coordenadoria Especial da Primeira Infância) como primeiro órgão a ter seus principais desafios explorados no projeto BigDataFortaleza. Após a aplicação de técnicas de levantamento de necessidades, foi identificado o desafio de entender melhor a demanda de vagas em creches, o que envolveu também a Secretaria Municipal de Educação (SME).

As seções que se seguem trazem de forma resumida as principais questões pautadas no projeto, tais como, motivações, metodologia e resultados auferidos até o momento. Na seção 2, o desafio piloto, o da Primeira Infância em Fortaleza, é introduzido. Na seção 3, a metodologia empregada no projeto é apresentada. Já na seção 4, os principais resultados alcançados no projeto até o presente momento são apresentados. A seção 5 traz discussões sobre os resultados, à luz das necessidades estratégicas da PMF. Por fim, a seção 6 traz as conclusões finais e aponta direcionamentos futuros para o projeto BigDataFortaleza.

2. PILOTO BIG DATA FORTALEZA PARA A PRIMEIRA INFÂNCIA

As Nações Unidas, em sua Agenda 2030 para o Desenvolvimento Sustentável (UNGA, 2015), delineou 17 Objetivos de Desenvolvimento Sustentável (ODS) e 169 metas que devem servir para nortear países na definição de políticas para melhoria da vida das pessoas. Alguns dos ODS incluem erradicar a pobreza; alcançar a segurança alimentar e melhoria da nutrição; assegurar uma vida saudável e promover o bem-estar para todos, em todas as idades; assegurar a educação inclusiva e equitativa e de qualidade, e promover oportunidades de aprendizagem ao longo da vida para todos; e reduzir a desigualdade nos países e entre eles.

Ainda na Agenda 2030, a importância de dados é evidenciada quando diz que “serão necessários dados desagregados de qualidade, acessíveis, oportunos e confiáveis para ajudar na medição do progresso e para garantir que ninguém fique para trás. Tais dados são fundamentais para a tomada de decisões. Os dados e informações dos mecanismos de relatórios existentes devem ser utilizados sempre que possível.” A partir desse recorte, pode-se identificar a importância de políticas voltadas para a educação para atingir a concretização de parte importante desses objetivos e que dados são uma fonte primordial para subsidiar governos na tomada de decisão na construção de políticas efetivas para a realização de seus objetivos.

A educação inicial, infantil, pré-escolar, pré-primária ou da primeira infância (alguns dos diversos nomes em países da América Latina para a educação de crianças, do nascimento aos seis anos de idade, aplicáveis a diversas divisões por idade dentro dessa faixa etária) tem percorrido um árduo caminho de reconhecimento e inserção nos sistemas educacionais, o que levou a um aumento lento em sua cobertura, apesar de, nos últimos anos, ter sido apontada como um dos principais focos das políticas educacionais (PERALTA ESPINOSA, 2022).

⁹ <https://unhabitat.org/programme/people-centered-smart-cities>

Sendo assim, o planejamento e a implementação de políticas e serviços voltados à primeira infância são fundamentais para assegurar oportunidades de aprendizado a longo prazo e melhores resultados em educação, emprego e saúde entre indivíduos. Pesquisas em neurociência, psicologia do desenvolvimento e ciência cognitiva revelaram que uma educação de qualidade na primeira infância, comunidades de apoio e um ambiente familiar positivo servem como importantes blocos de construção para promover um desenvolvimento saudável entre bebês e crianças pequenas (UNESCO, 2022).

Estratégias-chave para o pleno aproveitamento de resultados de políticas públicas voltadas para a primeira infância foram sugeridas pela UNESCO (2022) e incluem o impulsionamento do uso de dados, o monitoramento e a avaliação para responsabilização dos entes envolvidos. Idealmente, governos devem desenvolver sistemas de monitoramento que ajudem a acompanhar o progresso e permitir a responsabilização durante o processo que visa atingir as metas dos ODS relacionados à primeira infância.

Nesse sentido, o desenvolvimento de plataformas como o BigDataFortaleza, que permitam o desenvolvimento de análises utilizando dados de diferentes domínios relacionados às políticas voltadas para a primeira infância, mostra-se importante para o monitoramento contínuo e a avaliação de indicadores, de preferência seguindo normas internacionalmente comparáveis, além de prover *insights* mais precisos para tomadas de decisão baseada em dados e evidências. Sendo assim, a prefeitura escolheu como foco inicial do nosso projeto o domínio da primeira infância.

3. METODOLOGIA UTILIZADA

Esta seção traz os materiais e métodos técnico-científicos utilizados no desenvolvimento da plataforma BigDataFortaleza. Através dos métodos Canvas de Dados, Metodologia Fundacional da Ciência de Dados e a Metodologia Scrum, os materiais foram coletados junto aos atores da PMF para o cumprimento dos propósitos do projeto.

3.1. Canvas de Dados

O Canvas de Dados é uma metodologia que orienta gestores na navegação através do pensamento analítico, a partir de objetivos e desafios bem definidos, para pautar projetos e decisões e dar os primeiros passos para criar uma solução de dados (ÍRIS LAB, SOCIAL GOOD BRASIL, 2021). A razão para a proposição da técnica, pelo Social Good Brasil¹⁰, aconteceu pelo fato de muitas vezes resultados efetivos não serem alcançados nos projetos devido ao fato de não se ter um objetivo claro em relação ao projeto, fazendo com que os cientistas de dados envolvidos no projeto não consigam captar e materializar em dados as respostas necessárias aos questionamentos levantados pelos gestores do projeto. De acordo com um estudo do grupo Gartner, 35,85% dos projetos de dados falham, e uma das principais razões para isso é a falta de um objetivo claro¹¹. O que se percebe na prática é que durante a etapa de definição dos objetivos do projeto, os gestores costumam se inundar de dados, fazendo com que os avanços do projeto sejam lentos ou que, ao final do projeto, não se consiga aferir se as decisões tomadas ao longo do projeto foram as melhores.

O Canvas de Dados é tanto uma ferramenta estratégica como prática que pode ser utilizada pelo gestor público. Ela o orienta a navegar pelo pensamento analítico e estratégico. ÍRIS LAB, SOCIAL GOOD BRASIL (2021) apresenta as etapas para construção do Canvas de Dados, segundo a metodologia, bem como a ferramenta Canvas de Dados, original da Social Good Brasil. Como se percebe, parte-se inicialmente do propósito, ou seja, o objetivo a ser alcançado ou desafio que se deseja resolver. Esse propósito será o foco em um universo de dados que ultrapassa a capacidade humana de consumo. Os dados são então identificados na terceira etapa do processo. Assim, só depois da definição de um objetivo claro e das perguntas para tomada de decisão é que são definidos os dados relevantes.

3.2. Metodologia Fundacional da Ciência de Dados

Com relação à análise dos dados está sendo utilizada a **Metodologia Fundacional de Ciência de Dados**, proposta por (IBM, 2015). Ela é importante porque neste projeto há a necessidade de lidar com grandes volumes de dados coletados. Nessa metodologia, 10 estágios determinam um processo iterativo para a descoberta de conhecimento através dos dados. São eles:

¹⁰ <https://socialgoodbrasil.org.br/blog/canvas-da-proposta-de-valor>

¹¹ <https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence>

1. **Entendimento do negócio:** consiste na definição do desafio a ser resolvido, dos objetivos do projeto e nos requisitos da solução. Nesse estágio, a participação de especialistas é essencial.
2. **Abordagem analítica:** após definir o desafio, os pesquisadores podem definir a abordagem analítica para solucioná-lo. Fazer isso envolve expressar o desafio no contexto de técnicas estatísticas e de mineração de dados para que os pesquisadores possam identificar aquelas mais adequadas para alcançar o resultado esperado.
3. **Requisitos dos dados:** a abordagem analítica escolhida determina os requisitos dos dados. Os métodos analíticos requerem certos conteúdos, formatos e representações dos dados tratados.
4. **Coleta de dados:** pesquisadores identificam e coletam dados que sejam relevantes para o domínio do desafio. Caso encontrem dificuldades na coleta de dados, os pesquisadores podem ter que revisar os requisitos de dados e coletar mais informações.
5. **Entendimento dos dados:** estatísticas descritivas e técnicas de visualização podem ajudar no entendimento dos dados, avaliar a qualidade dos mesmos e levar às primeiras descobertas sobre a solução do desafio.
6. **Preparação dos dados:** engloba todas as atividades de construção de conjuntos de dados a serem usados no estágio de modelagem. Essas atividades incluem a limpeza dos dados, a combinação com dados de fontes diversas e a transformação dos dados para serem mais úteis e fáceis de lidar.
7. **Modelagem:** Iniciando com um conjunto de dados preparado nos estágios anteriores, pesquisadores usam um subconjunto de treinamento (dados históricos onde o resultado de interesse é conhecido nos casos de aprendizado supervisionado) para desenvolver modelos preditivos utilizando a abordagem analítica já descrita. O estágio de modelagem é altamente iterativo.
8. **Avaliação:** a qualidade do modelo é avaliada sendo verificado se o mesmo atende ao desafio apropriadamente. Esse estágio envolve a aplicação de métricas de diagnóstico bem como outros recursos como tabelas e gráficos com a utilização do conjunto de dados de teste nos casos de modelos preditivos.
9. **Implantação:** após a aprovação do modelo proposto pelos stakeholders, é realizada a implantação no ambiente de produção ou em um ambiente de testes equivalente. Tal implantação é geralmente limitada inicialmente para permitir uma avaliação do desempenho da solução.
10. **Feedback:** através da coleta de resultados do modelo implantado, a organização captura feedback sobre o desempenho e observa como o ambiente de implantação é afetado. Analisar esse feedback permite que os pesquisadores refinem o modelo, melhorando sua acurácia.

3.3. Scrum

O projeto BigDataFortaleza segue o framework **Scrum** (SCHWABER; SUTHERLAND, 2013), no que tange às atividades de desenvolvimento de software. Scrum é um método de desenvolvimento de software incremental baseado no conceito de *Sprints*, iterações de trabalho de tamanho fixo. Ele define quatro tipos de reuniões: (i) *A Reunião Diária*, que acontece diariamente e cujo objetivo é disseminar conhecimento sobre o que foi feito no dia anterior, identificar impedimentos e descrever o trabalho futuro a curto prazo; (ii) *Reunião de Revisão*, realizada ao final da Sprint para apresentar as funcionalidades implementadas; (iii) *Reunião de Planejamento*, planejamento para definir as demandas que serão realizadas da próxima sprint; e (iv) *Reunião de Retrospectiva*, cujo objetivo é rever os erros e acertos da última Sprint, identificando assim as lições aprendidas. Além disso, o Scrum também define três papéis: (i) Product Owner, representa os interesses de todos os envolvidos (Stakeholders), define as funcionalidades do produto e prioriza os itens de Product Backlog; (ii) Time de desenvolvimento, composto pelos profissionais que fazem o trabalho para entregar as versões incrementais do produto no final de cada Sprint; e (iii) Scrum Master, responsável por garantir que o Scrum é entendido e que o Time de desenvolvimento e o Product Owner sigam as práticas e regras do Scrum.

Dentro das *Sprints* são utilizadas boas práticas de Engenharia de Software, como a aplicação de técnicas de coleta de dados contextuais para a elicitação de requisitos, a utilização de Integração Contínua, a realização de testes de software e inspeções de usabilidade. Em relação à elicitação de requisitos, são utilizadas técnicas como concepção, elicitação, elaboração, especificação e validação, permitindo, assim, coletar informações sobre o produto desejado (DARIN et al., 2020). Sobre testes, MYERS et al. (2011) define como sendo o processo de executar um programa com o objetivo de encontrar defeitos. Os testes serão conduzidos em três níveis (BOURQUE e FAIRLEY, 2014): (i) Testes unitários, nos quais as unidades do software são testadas isoladamente; (ii) Testes de integração, que valida a integração entre as unidades do software; e Testes de Sistema, que é preocupado com o comportamento do sistema como um todo. A integração contínua (MEYER, 2014), por sua vez, é uma prática de desenvolvimento que indica a integração frequente de código, de modo mais eficiente, através de builds e testes automatizados. Quanto à inspeção heurística, trata-se de um método de avaliação com especialistas, no qual um pesquisador ou um grupo de pesquisadores com vasto

conhecimento pode avaliar um produto com base em um conjunto de parâmetros (PREECE et al., 2004), como, por exemplo, as 10 heurísticas de usabilidade de Nielsen (NIELSEN, 2005).

Atualmente, o projeto BigDataFortaleza está na 7.^a sprint (finalizada no dia 18/11/2022), e, ao longo do projeto, 215 issues já foram concluídas. Até o presente momento 10,6GB de dados foram processados, gerando 183 MB de dados pré-processados que foram armazenados em nuvem Amazon AWS ¹².

4. RESULTADOS

Os principais resultados do projeto até o momento são discutidos a seguir e um relatório completo de todas as entregas está disponível na web¹³.

4.1. Workflow da Plataforma

O *workflow* da plataforma BigDataFortaleza - apresentado na Figura 1 - tem como elemento central o *desafio* a ser definido pelos gestores da PMF. Esses desafios podem estar relacionados a um cenário de diminuição das desigualdades sociais como proposto no plano Fortaleza 2040¹⁴.

O fluxo de trabalho inicia a partir da definição do desafio por parte do gestor público e registro desse desafio na plataforma por parte do IPLANFOR. Esse registro auxiliará no acompanhamento das atividades a serem realizadas pelos cientistas de dados, tais como a compreensão do contexto, identificação das perguntas estratégicas, definição dos produtos de dados, aquisição, processamento e análise de dados, além da validação das respostas com Big Data. Após a validação das análises, a plataforma permitirá ao cientistas de dados a inclusão de dois conjuntos de scripts: i) scripts responsáveis pela execução do pipeline de dados (seguindo a arquitetura Delta Lake (ARMBRUST, et al., 2020)) e ii) os scripts responsáveis pela execução dos algoritmos de Machine Learning. O primeiro conjunto é fundamental para a limpeza dos dados brutos, pré-processamento e transformação do dataset, e aplicação da lógica de negócio requerida para construir os analíticos (tabelas, gráficos e indicadores). O segundo conjunto de scripts permitirá ao gestor estimar cenários por meio de previsões e inferências.

Ao final do fluxo, será possível definir os perfis de acesso para cada análise e o gestor poderá analisar os produtos de dados para auxiliar a tomada de decisão em relação ao desafio inicialmente proposto.

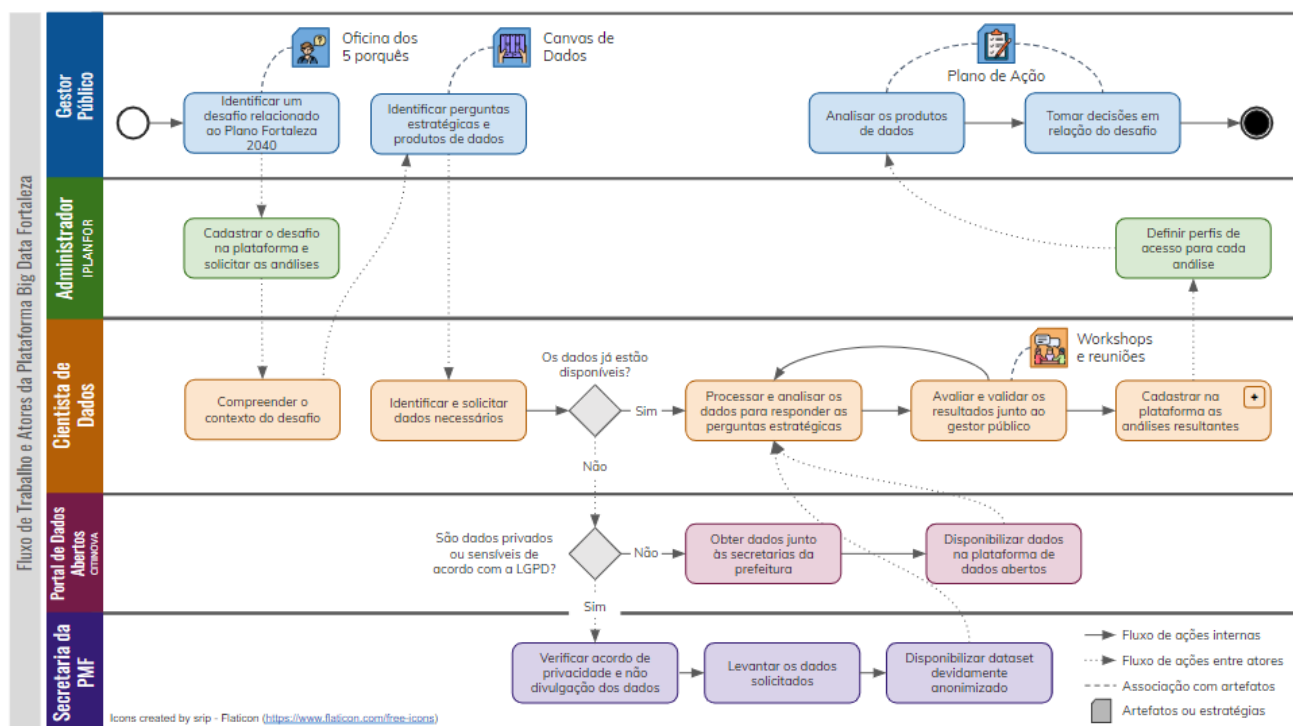


Figura 1: Workflow de trabalho da plataforma BigDataFortaleza. **Fonte:** Os autores.

¹² <https://socialgoodbrasil.org.br/blog/canvas-da-proposta-de-valor>

¹³ shorturl.at/jxzOV

¹⁴ <https://fortaleza2040.fortaleza.ce.gov.br/site/>

4.2. Oficina dos 5 porquês e Canvas de Dados

Seguindo a metodologia Canvas de Dados, foi realizada uma oficina para entendimento das principais questões relacionadas ao domínio escolhido pela PMF, no caso, a Primeira Infância em Fortaleza. A oficina, ilustrada por meio da Figura 2, foi realizada no dia 18 de Julho de 2022 com a participação de mais de 25 *stakeholders* relacionados à primeira infância, tais como, Coordenadoria Especial da Primeira Infância (CESPI), Secretaria Municipal de Educação (SME), Secretaria Municipal de Saúde (SMS), Secretaria do Planejamento, Orçamento e Gestão (SEPOG), Secretaria Municipal da Juventude (SEJUV), Fundo das Nações Unidas para a Infância (UNICEF), dentre outros. Cada grupo elegeu uma “dor” relativa à primeira infância e, em seguida, respondeu a cinco porquês a fim de encontrar a causa raiz para aquele domínio. Ao final, cada equipe apresentou seus resultados e todos realizaram uma votação para priorizar os desafios. No dia 8 de Setembro de 2022, foi realizada a Oficina do Canvas de Dados. A Figura 3 apresenta o resultado do Canvas de Dados da primeira infância de Fortaleza consolidado.



Figura 2: Oficina dos 5 Porquês do desafio Primeira Infância. Fonte: os autores

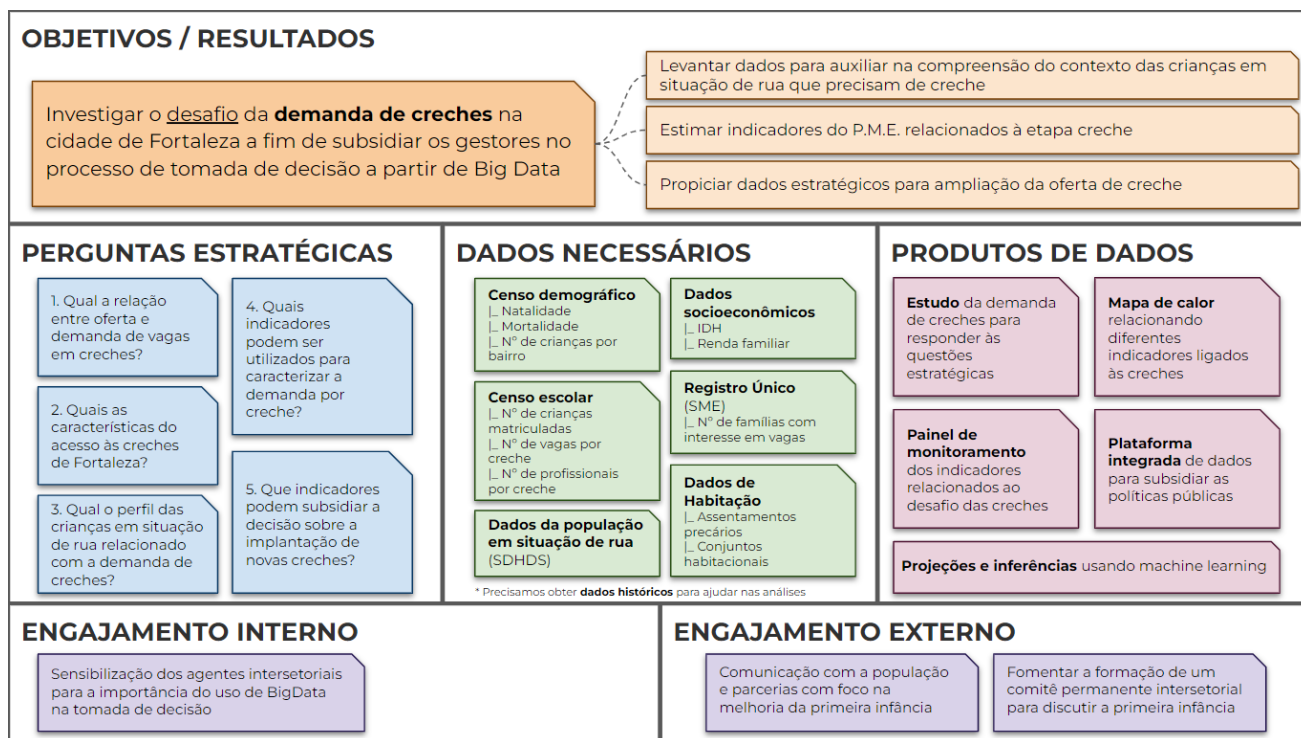


Figura 3: Canvas de Dados do desafio da Primeira Infância em Fortaleza. Fonte: os autores

4.3. Arquitetura

Seguindo a metodologia Fundacional de Ciência de Dados, propôs-se uma arquitetura de software em Big Data adequada aos propósitos do projeto. A arquitetura da plataforma foi inspirada primordialmente na arquitetura de processamento Big Data em nuvem Databricks Delta Lake (ARMBRUST, et al., 2020) e implementada por meio de serviços Amazon AWS, como mostra a Figura 4.

Podem-se citar diversos benefícios na utilização de nuvens computacionais para implementação de serviços de análises de dados, tais como alta disponibilidade, velocidade de processamento, custo reduzido (KIRAN, et al., 2015). Assim, pretende-se também com este projeto, popularizar no âmbito da gestão da PMF a computação em nuvem para sistemas de análises de dados, ao passo que atualmente a maioria desses sistemas é disposta em servidores *on premises*, ou seja, servidores físicos mantidos pela própria PMF.

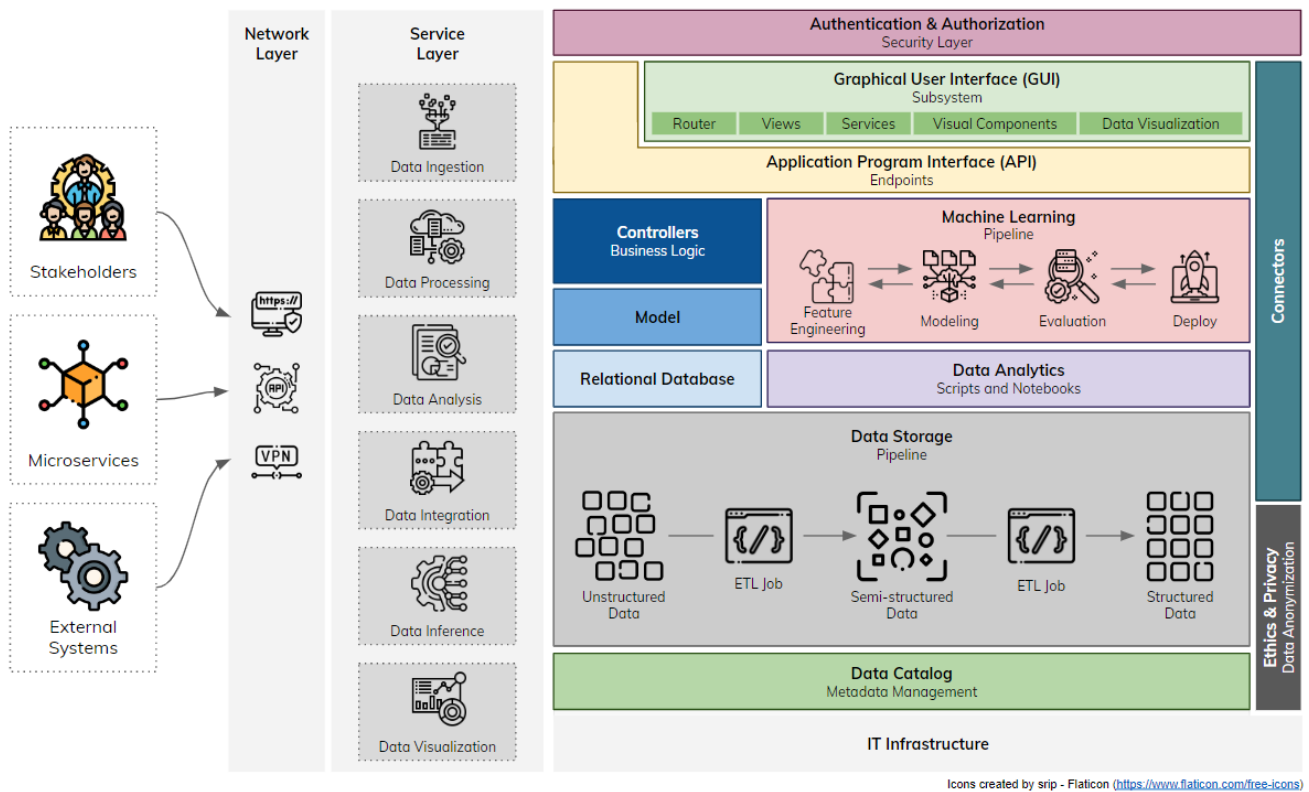


Figura 4: Arquitetura Big Data em nuvem da plataforma BigDataFortaleza. **Fonte:** os autores

4.4. Plataforma

Além da metodologia Scrum, o desenvolvimento da plataforma segue alguns princípios da Engenharia de Software, tais como Engenharia de Requisitos, Integração Contínua, Testes e Validação de Software e Experiência e Interface do Usuário (UI/UX Design). Vale ressaltar que a plataforma tem por finalidade atender tanto a gestores da PMF quanto ao público em geral, por isso, proporcionar um alto grau de usabilidade, bem como fornecer uma experiência rica e intuitiva ao usuário, através de técnicas de UI/UX, são elementos essenciais da plataforma. A título de ilustração, a Figura 5 apresenta a página inicial da plataforma e um exemplo de análise do desafio da primeira infância.

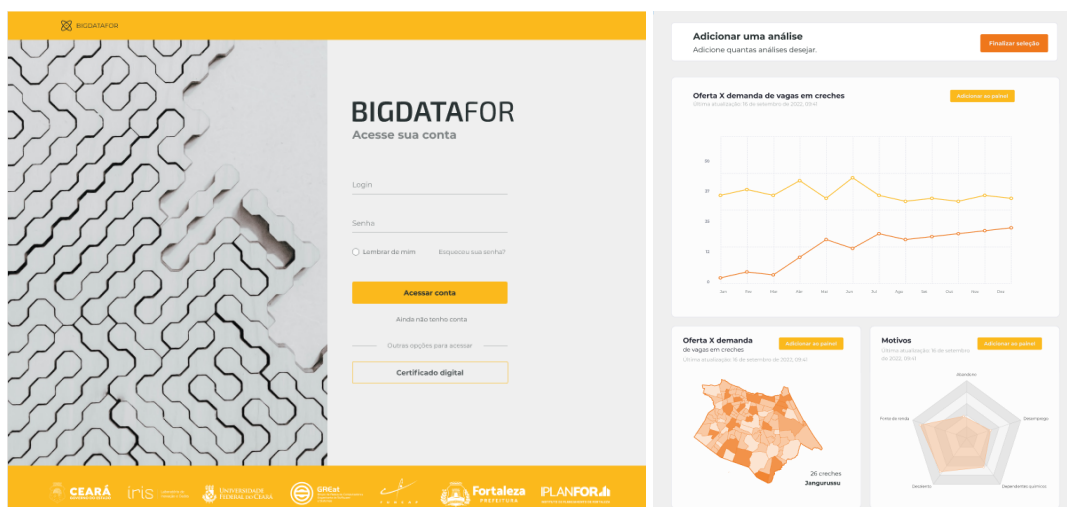


Figura 5: Página inicial da plataforma BigDataFortaleza (esq.) e Página com um exemplo de análise (dir.) acerca do desafio da Primeira Infância. **Fonte:** os autores

5. DISCUSSÃO

O Projeto BigDataFortaleza começou em Julho de 2022 com uma previsão de duração de dois anos e meio. Embora em estágio inicial, a equipe do projeto empreendeu diversas iniciativas para apresentar resultados concretos nos três primeiros meses do projeto, a seguir: a definição da equipe, a proposição da arquitetura de Big Data, a construção do Canvas de Dados sobre a Primeira Infância, e o desenvolvimento das primeiras funcionalidades da plataforma.

No primeiro semestre de 2023, os esforços serão direcionados ao término de outros desafios que envolvem a primeira infância e a saúde, incluindo aí a vacinação e os medicamentos, e à análise de outro domínio: habitação em Fortaleza. No início de 2023, os cientistas de dados do projeto realizarão um novo Canvas de Dados com os gestores da PMF envolvidos nos desafios da habitação, e então proporão um conjunto de perguntas cujas respostas serão dadas através de dados na plataforma. As respostas serão materializadas através de funcionalidades de visualização e sobretudo predição de demandas futuras. Espera-se que este cenário leve em torno de seis meses para conclusão.

Seguindo a métrica dos três primeiros domínios (primeira infância, saúde e habitação) abordados na plataforma, acredita-se que em torno de 5 domínios serão abordados ao longo do projeto, trazendo um claro benefício às políticas de gestão da PMF, ao passo que 5 grandes desafios poderão ser analisados na integralidade, além de remanescer na plataforma uma capacidade preditiva a esses cenários.

Pode-se, por fim, relacionar os seguintes obstáculos para o desenvolvimento do projeto: articulação com as secretarias da PMF, visto que são muitas secretarias e necessitamos de agendas com vários gestores ao mesmo tempo, aquisição de dados (abertos e sensíveis) e dados desatualizados. Pode-se apontar ainda a necessidade de uma cultura de dados planejada em toda a PMF e de uma cultura de integração de dados e sistemas que exige uma articulação com os stakeholders dos cenários de análise da plataforma, integrantes de diversas secretarias. Entretanto, até o momento, tanto os gestores quanto os técnicos da PMF, com a ajuda do Iplanfor, tem se mostrado muito solícitos.

5.1. Perguntas Estratégicas

No domínio da primeira infância, as seguintes perguntas estratégicas foram levantadas, conforme mostra o Canvas de Dados consolidado (Figura 3), as quais devem ser respondidas com a análise dos dados:

1. Qual a relação entre oferta e demanda de vagas em creches?
2. Quais as características do acesso às creches de Fortaleza?
3. Qual o perfil das crianças em situação de rua relacionado com a demanda de creches?
4. Quais indicadores podem ser utilizados para caracterizar a demanda por creche?
5. Que indicadores podem subsidiar a decisão sobre a implantação de novas creches?

5.2. Análises de Dados para responder às Perguntas Estratégicas

De acordo com a metodologia do projeto, os principais resultados se materializam através de análises de dados cadastrados na plataforma pelos cientistas de dados envolvidos no projeto para responder às perguntas estratégicas de cada desafio. Algumas das análises realizadas na plataforma são apresentadas nas subseções que se seguem.

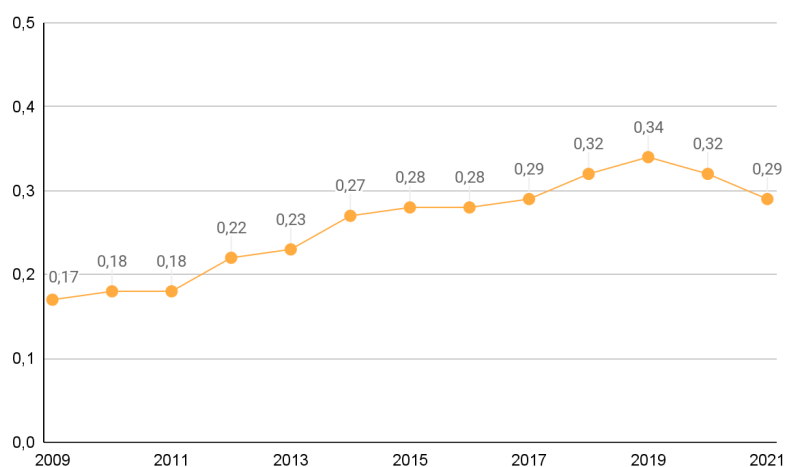


Figura 6: Evolução da Taxa de Escolarização Bruta na Etapa Creche em Fortaleza. **Fonte dos dados:** Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (SMS).

5.2.1. Taxa de Escolarização Bruta

A análise de dados ilustrada na Figura 6 diagnosticou a situação histórica de 2009 a 2021 da Taxa de Escolarização Bruta (TEB) na etapa creche em Fortaleza. Essa taxa representa o percentual de crianças de 0 a 3 anos que estão matriculadas em creches, cuja análise da queda nessa taxa, observada em 2020 e 2021, pode ser explicada pelo fechamento de creches privadas durante a pandemia de COVID-19¹⁵. Isso implica um maior esforço da gestão pública para ampliação de vagas, visando o cumprimento da meta do Plano Municipal de Educação (PME), o qual é de 50% das crianças de 0 a 3 anos matriculadas em creches.

5.2.2. Relação entre Oferta e Demanda

A análise, apresentada na Figura 7, traz a relação entre a oferta e a demanda de vagas por creche. Considerando como referência o indicador Taxa de Escolarização Bruta do PME, apresentado na Figura 6, percebe-se que seriam necessárias mais de 20 mil novas vagas para alcançar a meta do PME em 2021.

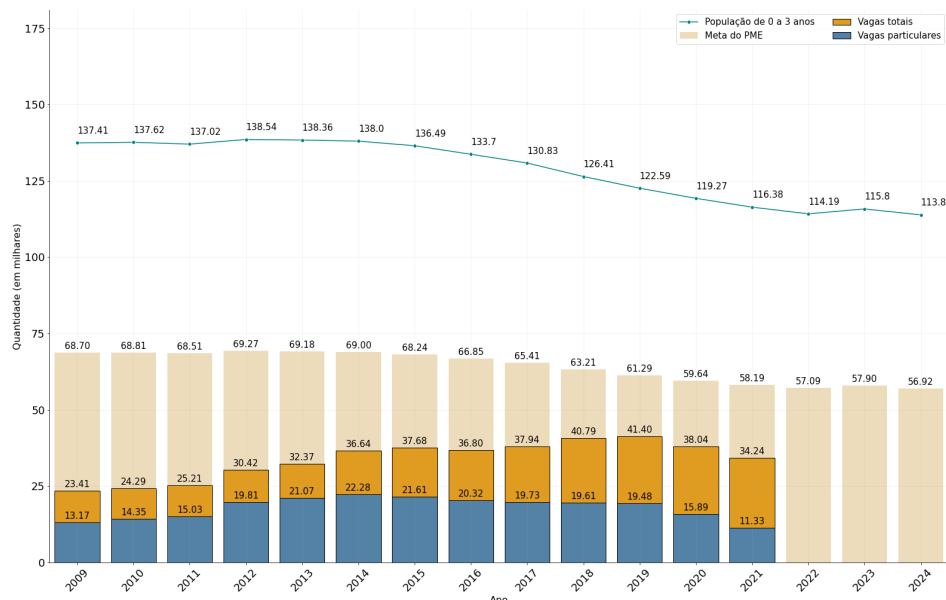


Figura 7: Relação entre oferta e demanda de creches em Fortaleza. **Fonte dos dados:** Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (SMS).

5.2.3. Predição da população de 0 a 3 anos e vagas ofertadas ao longo dos anos

O analítico apresentado na Figura 8 traz i) os dados históricos e a projeção da população de 0 a 3 anos (já removendo o número de óbitos registrados nessa mesma faixa etária) e ii) os dados históricos e a projeção da oferta de vagas incluindo tanto as creches públicas quanto privadas. A decisão de incluir o número de vagas ofertadas nas creches privadas fundamenta-se na forma como o Plano Municipal de Educação calcula seus indicadores e, nesse cenário, a população de 0 a 3 anos representa o público potencial das vagas em creches, sendo utilizada para caracterizar a demanda potencial.

De posse da série histórica com dados registrados entre os anos de 2009 e 2021, modelou-se uma regressão linear para estimar o comportamento da série temporal em relação aos anos futuros. No gráfico, utilizou-se como condição de parada para a estimativa o ano no qual a PMF alcançaria a universalização no número de vagas ofertadas, ou seja, uma vaga em creche para cada criança nascida em Fortaleza. Analisando essa estimativa, observou-se que se a taxa de crescimento das ofertas de vagas em creche for mantida, a meta do PME de 50% das crianças de 0 a 3 anos matriculadas em creche será alcançada em 2030 (seis anos após 2024 que foi o ano previsto pelo PME). No entanto, essa análise precisa ser ponderada considerando alguns aspectos:

- O modelo de regressão linear foi a estratégia inicial utilizada para modelar as séries temporais. No entanto, modelos com maior robustez, como a média móvel integrada autoregressiva (ARIMA), estão sendo avaliados para garantir maior confiabilidade às previsões feitas na plataforma.

¹⁵<https://g1.globo.com/educacao/noticia/2022/02/13/brasil-perdeu-mais-de-25-mil-creches-particulares-durante-a-pandemia-especialistas-veem-risco-de-sobrecarga-na-rede-publica.ghtml>

- Para estimar o ano no qual a PMF atingirá a meta do PME, não foram consideradas as vagas criadas ao longo de 2022 com a construção de novas creches. Esse fator deve reduzir o tempo necessário para alcançar 50% das crianças de 0 a 3 anos matriculadas em creches.

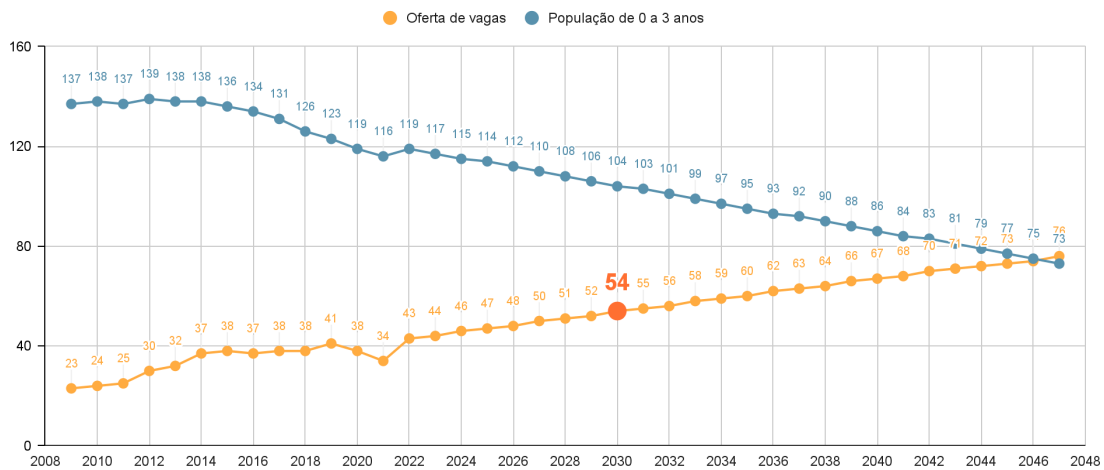


Figura 8: Predição em regressão linear da população de 0 a 3 anos e vagas ofertadas ao longo dos anos. **Fonte dos dados:** Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (SMS).

A predição da população dos bairros foi realizada, permitindo análises com menor granularidade. A Figura 9 traz um mapa coroplético em que cores mais intensas representam a maior população com idade de 0 a 3 anos, projetada para 2024. Essa estimativa usa a série histórica de dados registrados entre os anos de 2009 e 2021. A Figura 10 mostra a evolução da população de 0 a 3 anos nos bairros Bom Jardim e Jangurussu.

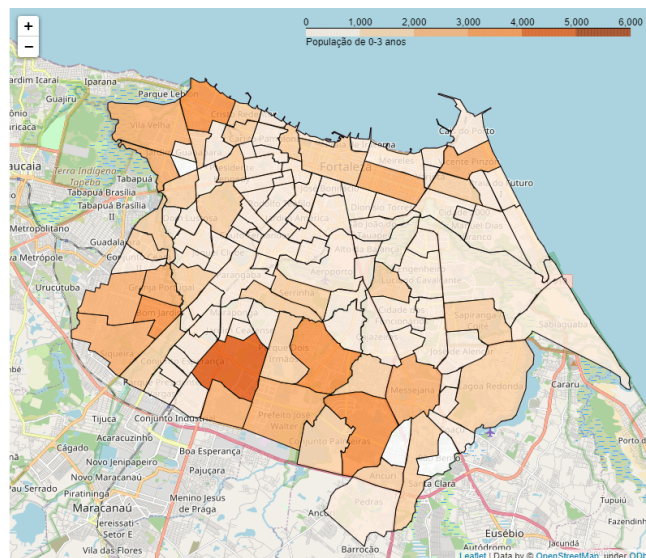


Figura 9: Mapa coroplético de predição em regressão linear da população de 0 a 3 anos nos bairros de Fortaleza. **Fonte dos dados:** SINASC/SIM extraído do TABNET (SMS).

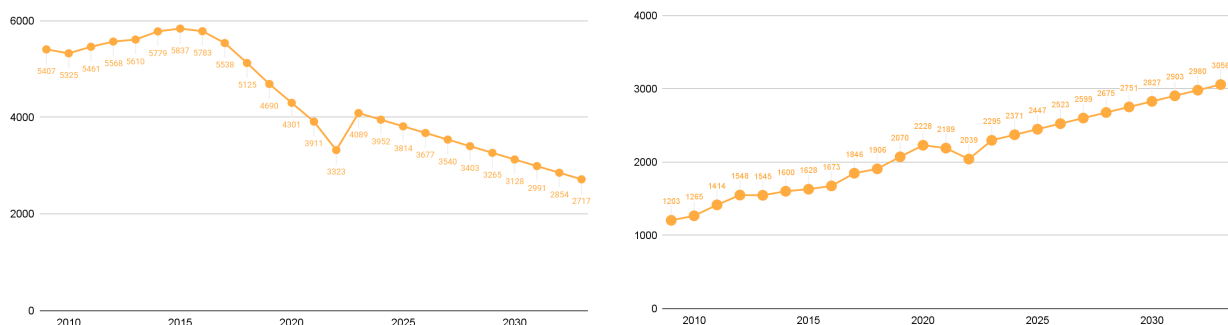


Figura 10: Evolução da população (predição em regressão linear) da população de 0 a 3 anos nos bairros Bom Jardim (esquerda) e Siqueira (direita). **Fonte dos dados:** SINASC/SIM extraído do TABNET (SMS).

5.2.4. Relação entre número de creches e IDH dos bairros

A relação entre o número de creches e o Índice de Desenvolvimento Humano (IDH) dos bairros também foi analisada (Figura 11). Nesse analítico, foi possível observar que bairros com IDH mais baixo (cores mais claras) possuem maior concentração de creches municipais. Por outro lado, bairros com os melhores IDH, tais como Aldeota, possuem poucas creches públicas. Na plataforma desenvolvida, ao clicar em um balão azul, informações sobre a creche, tais como número de matrículas e endereço, são exibidas.

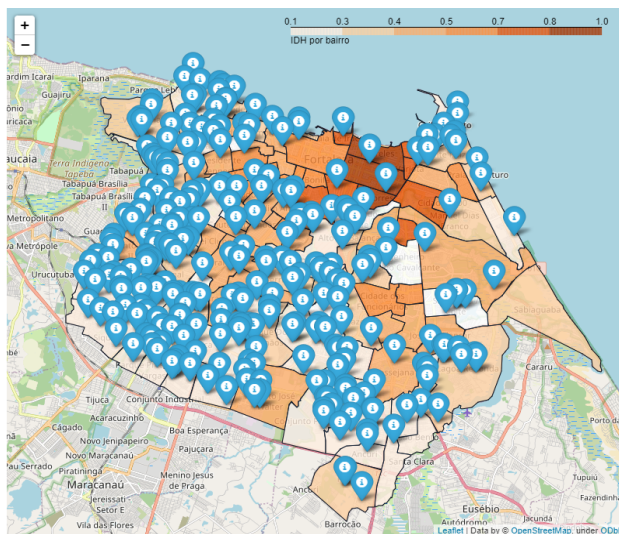


Figura 11: Relação entre número de creches e IDH dos bairros.

Fonte dos dados: Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (SMS).

6. CONCLUSÃO

A plataforma BigDataFortaleza, por definição, é um ambiente de software que visa impactar diariamente o processo de tomada de decisão dos gestores públicos da PMF nos desafios abordados. Ao ir além das plataformas de análise em Big Data tradicionais, focando não somente em visualizações de dados, mas prioritariamente em predições para previsão de demandas futuras, espera-se que, nos cenários impactados, os gestores da PMF possam ter subsídios diferenciados, tanto para analisar fidedignamente os cenários atuais (através da predição de dados faltantes, por exemplo), quanto para a predição de demandas futuras, possibilitando a visão de um panorama mais estruturado a médio e longo prazo de desafios oriundos de cenários de interesse do plano Fortaleza 2040. Acredita-se ainda que a arquitetura de processamento em Big Data proposta para o projeto, bem como o catálogo de dados do projeto, possam ser úteis a outras iniciativas da PMF.

De uma maneira geral, o projeto, ao entregar uma plataforma e rotinas de integração e análise de dados de forma mais eficiente, contribuirá não somente com uma predição atual e futura mas também com o máximo de precisão possível em relação ao planejamento municipal. Por exemplo, no desafio da demanda de creches, o gestor conseguirá prever quantas creches serão necessárias em determinados bairros mesmo antes do nascimento das crianças. Esse tipo de informação é parte integrante do ciclo de políticas públicas implementadas em nível de município e mesmo de estado, que tenha foco em Fortaleza.

Destacam-se como pontos fortes do projeto o enfrentamento a cenários desafiadores às metas do Plano Fortaleza 2040 (e.g., oferta de creches, oferta de moradias populares). A plataforma permite responder perguntas levantadas pelos gestores da PMF com dados, processamento eficiente, integração com outras iniciativas de análise de dados da PMF e capacidade preditiva de cenários futuros através de modelos de Machine Learning. O ponto fraco a mencionar é o fato de que, por limitações orçamentárias do projeto, não é possível implementar todos os cenários desafiadores ao cumprimento das metas do Plano Fortaleza 2040.

Para além dessa contribuição global, o BigDataFortaleza terá como foco temas específicos definidos de maneira cooperativa com os órgãos e entidades da administração municipal. Neste sentido, no primeiro ano o projeto tem como alvo responder perguntas relacionadas à Primeira Infância, que também englobam a Educação e a Saúde. Mais especificamente, o projeto buscará analisar o desafio da demanda de vagas em creches, respondendo a perguntas tais como: quais regiões concentram esse desafio? Qual a magnitude do desafio? Como iniciar sua resolução?

No segundo ano, buscar-se-á, através do uso dos dados, o estabelecimento de análises para apoiar a tomada de decisões nos desafios de déficit habitacional. Conforme dados do Plano Local de Habitação de

Interesse Social ¹⁶, e estimativas recentes da Habitafor, avalia-se que aproximadamente um milhão de pessoas residem em aglomerados subnormais. Lugares, em geral, sem acesso à água potável, sem conexão com a rede de esgotos, e com ruas não urbanizadas. Em geral, os residentes dos aglomerados também não têm documentos de propriedade e são carentes de outros serviços de assistência fornecidos pela municipalidade. Identificar os aglomerados, avaliar suas condições, e estimar a demanda por moradia digna que eles representam são tarefas que o BigDataFortaleza pode endereçar.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ARMBRUST, Michael et al. **Delta lake: high-performance ACID table storage over cloud object stores**. Proceedings of the VLDB Endowment, v. 13, n. 12, p. 3411-3424, 2020.
- BOURQUE, P.; FAIRLEY, R. (Ed.). **Guide to the Software Engineering Body of Knowledge**, Version 3.0. [S.l.]: IEEE Computer Society, 2014
- DARIN, T. G. R. et al. **Contextual Requirements Elicitation through the Combination of Interviews, Scenarios and Visual Artifacts**. In: 19TH Brazilian Symposium on Software Quality. São Luis, Brazil: Association for Computing Machinery, 2020. (SBQS'20).
- HOLLANDS, R. G. (2008). **Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?**. City, 12(3), 303-320.
- ÍRIS LAB, Social Good Brasil, AWS Educate. **A Era dos Dados para o Setor Público. Uma nova Cultura Organizacional Analítica**. 2021. Disponível em <https://irislab.ce.gov.br/wp-content/uploads/2021/09/LIVRO-DIGITAL-A-Era-dos-Dados-para-o-Setor-Publico%81blico.pdf>.
- KIRAN, Mariam et al. **Lambda architecture for cost-effective batch and speed big data processing**. In: 2015 IEEE International Conference on Big Data (Big Data). IEEE, 2015. p. 2785-2792.
- IBM. **Foundational methodology for data science**. 2015.
- LI, Xia, et al. **Towards sustainable smart cities: An empirical comparative assessment and development pattern optimization in China**. Journal of Cleaner Production 215 (2019): 730-743.
- MEYER, Mathias. **Continuous integration and its tools**. IEEE software, v. 31, n. 3, p. 14-16, 2014.
- MYERS, G., SANDLER, C. e BADGETT, T. (2011). **The Art of Software Testing**. 3 Ed. John Wiley & Sons.
- NIELSEN, J. **Ten usability heuristics**. (2005).
- OSMAN, Ahmed M. Shahat. **A novel big data analytics framework for smart cities**. Future Generation Computer Systems 91 (2019): 620-633.
- PERALTA ESPINOSA, V. **Análise da política curricular para a primeira infância na América Latina: estudo comparativo no Chile, Equador, México e Uruguai**. Technical Report. 61 pp. 2022.
- PREECE, J., Rogers, Y., and Sharp, H. (2004). **Interaction design**. Apogeo Editore.
- RANA, N.P., Luthra, S., Mangla, S.K. et al. **Barriers to the Development of Smart Cities in Indian Context**. Inf Syst Front 21, 503–525 (2019).
- RAZMJOO, Armin, et al. **Effective policies to overcome barriers in the development of smart cities**. Energy Research & Social Science 79 (2021): 102175.
- SANTANA, E. F. Z., CHAVES, A. P., GEROSA, M. A., KON, F., & MILOJICIC, D. S. (2016). **Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture**. ACM Computing Surveys (CSUR), 50(6), 78.
- SCHWABER, K.; SUTHERLAND, J., 2013. **The Scrum Guide - The Definitive Guide to Scrum: The Rules of the Game**. Disponível em: <<http://www.Scrumguides.org/docs/Scrumguide/v1/Scrum-guide-us.pdf>>. Acesso em: 16 set. 2022.
- UNESCO. **Global Partnership Strategy for early childhood, 2021-2030**. Technical Report. 46 pp. 2022.
- UNGA General Assembly (UNGA). **Transforming our world: the 2030 Agenda for Sustainable Development**. Resolut. A/RES/70/1, 25 1–35. 2015.

¹⁶<https://acervo.fortaleza.ce.gov.br/download-file/documentById?id=fcd18692-a091-4677-ac71-346c5cff1010>