

Big Data Fortaleza: Plataforma Inteligente para Políticas Públicas Baseadas em Evidências

Ismayle S. Santos^{1,2}, Pedro Almir M. Oliveira¹, Victória T. Oliveira¹,
Tales P. Nogueira^{1,3}, Allberson B. O. Dantas^{1,3}, Larissa M. Menescal⁴,
Élcio Batista⁴, Rossana M. C. Andrade¹

¹ Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)
Universidade Federal do Ceará (UFC)
Fortaleza – CE – Brasil

²Universidade Estadual do Ceará (UECE)
Fortaleza – CE – Brasil

³Universidade da Integração Internacional da Lusofonia Afro-Brasileira (Unilab)
Redenção – CE – Brasil

⁴Gabinete da Vice-Prefeitura de Fortaleza
Instituto de Planejamento de Fortaleza (IPLANFOR)
Fortaleza – CE – Brasil

rossana@ufc.br, elcio.batista@gabvice.fortaleza.ce.gov.br,
larissa.menescal@iplanfor.fortaleza.ce.gov.br, ismayle.santos@uece.br,
{pedromartins, victoria.oliveira}@great.ufc.br,
{allberson, tales}@unilab.edu.br

Abstract. *Public administration must understand the impact of their decisions by observing the dynamic context of the city. In this scenario, this paper presents a data analysis and prediction platform using Big Data and Machine Learning, called Big Data Fortaleza, intending to support the strategic planning (developed by the Fortaleza Planning Institute - IPLANFOR) for the sustainable urban development of Fortaleza. To conduct a pilot study with this platform, the challenge of better understanding the demand for daycare centers linked to the Early Childhood domain was chosen to be explored initially, since intelligent data management can assist in decision-making that can contribute to improving child development. So far, several exploratory and predictive analyses have been performed to answer critical questions of this domain raised in discussions with the public administration of Fortaleza.*

Resumo. *Ao observar o contexto dinâmico da cidade, é necessário aos gestores públicos entender qual o impacto de suas decisões. Diante disso, este artigo apresenta uma plataforma de análise de dados e previsão utilizando Big Data e Machine Learning, chamada Big Data Fortaleza, 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). 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.

1. Introdução

O último relatório da Organização das Nações Unidas (ONU) apontou que, em 100 anos (de 1950 a 2050), o percentual de pessoas vivendo em regiões urbanas passará de 30% para 68%. Por conta disso, o desenvolvimento sustentável das cidades dependerá de um profícuo gerenciamento do processo de urbanização, especialmente nos países em desenvolvimento, nos quais é esperada uma rápida urbanização até 2050¹.

Esse intenso processo de urbanização traz inúmeros desafios para as grandes cidades, tais como uso consciente dos recursos naturais, redução das desigualdades sociais e ampliação do acesso à tecnologia, além do fornecimento de serviços públicos com qualidade e que entendam a particularidade de cada cidadão [Sánchez-Corcuera et al. 2019]. Nesse contexto, a aplicação de tecnologias como *Big Data*, Internet das Coisas (IoT, do inglês *Internet of Things*), Computação em Nuvem e Computação Móvel tem se tornado uma excelente alternativa para lidar com tais desafios [Lai et al. 2020], constituindo a espinha dorsal das Cidades Inteligentes.

Para Osman *et al.* (2019), Cidades Inteligentes apresentam três características essenciais: (i) uso de tecnologia para conectar diferentes sistemas da cidade e coletar *insights* sobre como otimizar tais sistemas; (ii) ter uma visão global da cidade, na qual a inter-relação entre os sistemas é primordial; e (iii) crescer com sustentabilidade, ou seja, com uso consciente dos recursos naturais. Deste modo, tais cidades tem o potencial de oferecer um melhor padrão de vida para futuras gerações [Razmjoo et al. 2021].

Essa integração de sistemas das cidades, proposta por Osman *et al.* (2019), requer a coleta e processamento de volumes massivos de dados. Para isso, tecnologias como *Big Data* são requeridas uma vez que esse cenário deve considerar: i) o (v)olume de dados gerado; ii) a (v)elocidade necessária para armazenar, analisar e processar tais dados; iii) a (v)ariabilidade dos dados, ou seja, seus diferentes tipos; iv) a (v)ariabilidade, a qual refere-se a estrutura e semântica dos dados; e, por fim, v) o (v)alor que cada conjunto de dados possui [Fan and Bifet 2013].

No entanto, ainda há uma série de desafios a serem enfrentados quando o gestor de uma cidade decide investir no uso de *Big Data*. Por exemplo, diversas fontes de dados com diversos formatos; dificuldades para viabilizar a logística de compartilhamento de dados entre diferentes departamentos da cidade; baixa qualidade dos dados (presença de ruídos, dados faltantes, dados preenchidos incorretamente, etc.); segurança e privacidade; e alto custo das soluções já existentes [Osman 2019].

Este artigo apresenta o “Big Data Fortaleza” - uma plataforma inteligente para integração de dados, com a capacidade de fornecer visualizações e predições para fins de tomada de decisão e planejamento de estratégias por parte dos gestores da Prefeitura de Fortaleza. O projeto iniciou em Julho de 2022 tendo como alvo inicial a Coordenadoria Especial da Primeira Infância (CESPI) dada a relevância que a Primeira Infância tem em

¹ Key Facts about WUP: population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf.

diversos serviços (*e.g.*, educação, saúde, assistência social). Além de apresentar a estrutura geral da plataforma e os processos de análise de dados aplicados ao setor público, são discutidas lições aprendidas que podem auxiliar pesquisadores e gestores a mitigar algumas dificuldades na implementação de projetos que usem *Big Data* para políticas públicas baseadas em evidências.

2. Cenário da Cidade de Fortaleza

Fortaleza é uma das principais capitais do Nordeste, tendo, em 2021, um total de 2.703.391 habitantes (a quinta maior população do Brasil). 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.

Recentemente, a Prefeitura de Fortaleza criou um instrumento para o planejamento da cidade com estratégias a serem implementadas no curto, médio e longo prazo (tendo como horizonte o ano 2040). Assim, surgiu o Fortaleza 2040, um plano estratégico participativo, que busca integrar o desenvolvimento físico-territorial ao desenvolvimento social e econômico, ao passo que busca articular a discussão da cidade sob os diversos olhares e setores, territórios e esferas de governo. O Fortaleza 2040 apresenta desafios atuais que precisam ser solucionados, tais como o atraso educacional e baixa qualificação, jovens vulneráveis e pobreza e desigualdade social.

Dentre os cenários de desafios, existe a Primeira Infância, que trata de um domínio de atuação com foco nas crianças de zero a seis anos. Esse domínio é transversal porque requer ações no âmbito da saúde, educação, assistência social, dentre outros. No âmbito da educação, os gestores públicos devem oportunizar o acesso às creches e escolas voltadas para educação infantil. Quanto à saúde, é importante cuidar dessas crianças desde a gestação com o acompanhamento das consultas pré-natal, passando pelos cuidados após o nascimento e aplicação das vacinas. E nos direitos humanos e desenvolvimento social, é importante mapear as famílias com vulnerabilidade socio-econômica e/ou em situação de rua e auxiliá-las com benefícios sociais para reduzir as desigualdades no acesso aos bens e serviços que atendam aos direitos da criança.

Considerando os desafios da 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.

O desenvolvimento de plataformas como o Big Data Fortaleza, permite então análises utilizando dados de diferentes domínios relacionados às políticas voltadas para a Primeira Infância. Isso é 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.

3. Trabalhos Relacionados

Na literatura, diferentes trabalhos abordam soluções relacionadas à *Big Data* para resolver desafios do Governo. A seguir são descritos alguns destes trabalhos.

Furtado *et al.* (2023) apresentam um *framework* teórico para suportar a transformação digital de Cidades Inteligentes. Neste estudo, os autores descrevem como a transformação digital do estado do Ceará vem ocorrendo por meio de sete iniciativas. Dentre elas, há o Big Data Social, um sistema Web que agrega, armazena e processa múltiplas base de dados do governo que têm informações sobre famílias de baixa renda beneficiadas com o programa social “Mais Infância”. Por meio de analíticos, mineração de dados e *dashboards*, o sistema apresenta uma visão dos dados que auxilia os gestores a entender o impacto das políticas sociais. Outro sistema é o *Cerebrum*, que também integra diferentes bases de dados do governo, porém ele tem foco na segurança pública.

Ramos *et al.* (2022) apresentam uma arquitetura de *data lake* lógico aplicada na integração de sistemas e dados governamentais do estado de Alagoas. Os autores descrevem como essa arquitetura foi usada para analisar a distribuição geográfica dos usuários de três sistemas governamentais relacionados à capacitação em TI, assistência social e incentivo à prática de esportes.

Inácio *et al.* (2021) discutem sobre a criação do Sistema de Informações e Monitoramento Inteligente (SIMI), o qual faz parte da Sala de Situação criada durante o auge da pandemia do coronavírus na sede do Governo do Estado de São Paulo. Esse sistema agrega de forma anonimizada, dados epidemiológicos, dados sociais do IBGE e SEADE², taxa de ocupação de leitos nos hospitais e emissão de notas fiscais eletrônicas.

Santos *et al.* (2021) apresentam uma ferramenta de visualização de dados abertos disponibilizados no Portal de Transparência pela Prefeitura Municipal de Florianópolis. Esses dados eram originalmente disponibilizados pela Prefeitura em arquivos individuais em formato PDF. Inicialmente, esses arquivos em PDF foram convertidos no formato *Comma-Separated Values* (CSV) com o apoio da Tabula³ para então serem transferidos para o banco de dados. Com isso, a ferramenta, desenvolvida em PHP, disponibiliza *dashboards* com as informações sobre os gastos dos vereadores da cidade de Florianópolis.

Oliveira *et al.* (2020) apresentam um estudo utilizando *Big Data* para caracterizar os padrões de mobilidade dos usuários do sistema de transporte público por ônibus de Fortaleza. Para tanto, os autores utilizaram dados do Sistema Integrado de Transporte de Fortaleza, bem como os dados de cadastro do bilhete único (utilizado para integração tarifária fora dos terminais físicos de integração dos ônibus).

Desta forma, observa-se na literatura soluções criadas para monitor situações específicas [Inácio et al. 2021], visualização de dados abertos [Santos et al. 2021], integração de dados do governo [Ramos et al. 2022] e também o uso *Big Data* para resolver os problemas da cidade [Furtado et al. 2023, Oliveira et al. 2020]. Neste contexto, o diferencial deste trabalho está na proposição de uma plataforma de *Big Data* com o intuito de agregar dados de diferentes secretarias da Prefeitura de Fortaleza. Além disso, esta plataforma utiliza inteligência artificial para fazer previsões que podem auxiliar na tomada das decisões dos gestores da cidade.

²Sistema Estadual de Análise de Dados: <https://www.seade.gov.br>

³Website Tabula: <https://tabula.technology/>

4. Análise de Dados Aplicada ao Setor Público

A construção de uma solução de *Big Data* com o setor público tem desafios relacionados a heterogeneidade de sistemas, formatos e disponibilidades de dados, bem como os diferentes *stakeholders* envolvidos. Nesse sentido, apesar do Big Data Fortaleza ser extensível para qualquer dado da Prefeitura de Fortaleza, era importante uma temática inicial para que fosse possível definir um escopo para o *Minimum Viable Product* (MVP).

A Figura 1 descreve – na forma de *timeline* – o processo geral de definição do problema a ser abordado até a construção do MVP do Big Data Fortaleza. O primeiro marco do projeto foi a definição da temática alvo da plataforma. Nesse caso, a temática definida pelos gestores de Fortaleza foi “Primeira Infância”.

Em seguida, foi realizada uma *Oficina dos 5 porquês* com gestores de diferentes secretarias da Prefeitura. A oficina teve o intuito de levantar problemas e possíveis causas. Com a lista de problemas definidos, cientistas de dados fizeram uma análise preliminar da disponibilidade dos dados e da viabilidade em responder às questões identificadas. Diante dessa análise, e após uma nova rodada de reuniões com os gestores da cidade, foi escolhido o desafio inicial a ser abordado: “Demanda por Vagas na Etapa Creche”.



Figura 1. *Timeline* da execução das análises sobre a Primeira Infância

Com o desafio definido, no dia 08 de setembro de 2022, os *stakeholders* da Prefeitura relacionados à Primeira Infância se reuniram, bem como toda a equipe do Big Data Fortaleza, totalizando uma quantidade de 31 participantes, para a criação do Canvas de Dados. Os participantes foram divididos em grupos com o intuito de elaborarem perguntas estratégicas relacionadas ao desafio e indicar onde os dados poderiam ser encontrados, bem como os possíveis resultados esperados. Por exemplo, uma das perguntas obtidas foi “Qual a relação entre oferta e demanda de vagas de creches?”, e entre os dados relacionados estavam “número de creches” e “taxa de natalidade”.

Com a concretização do Canvas, realizou-se um conjunto de análises exploratórias de dados da educação básica, tendo como fontes o Censo Escolar do INEP, IPLANFOR e Secretaria Municipal de Educação, dentre outros. A partir da análise desses dados e da definição inicial dos requisitos, surgiu o MVP da plataforma Big Data Fortaleza, o qual é apresentado nas seções seguintes.

5. Plataforma Big Data Fortaleza

O *workflow* da plataforma Big Data Fortaleza (Figura 2) tem como foco os desafios definidos pelos gestores da Prefeitura. A definição do desafio é o ponto de partida do fluxo, sendo este posteriormente cadastrado na plataforma por um administrador. Logo após, os cientistas de dados iniciam o processo de compreensão do domínio associado ao desafio,

identificando perguntas estratégicas junto aos demais atores (e.g. reuniões estruturadas utilizando técnicas como a oficina dos 5 Porquês), definindo os produtos de dados, como os dados serão adquiridos, processados e analisados.

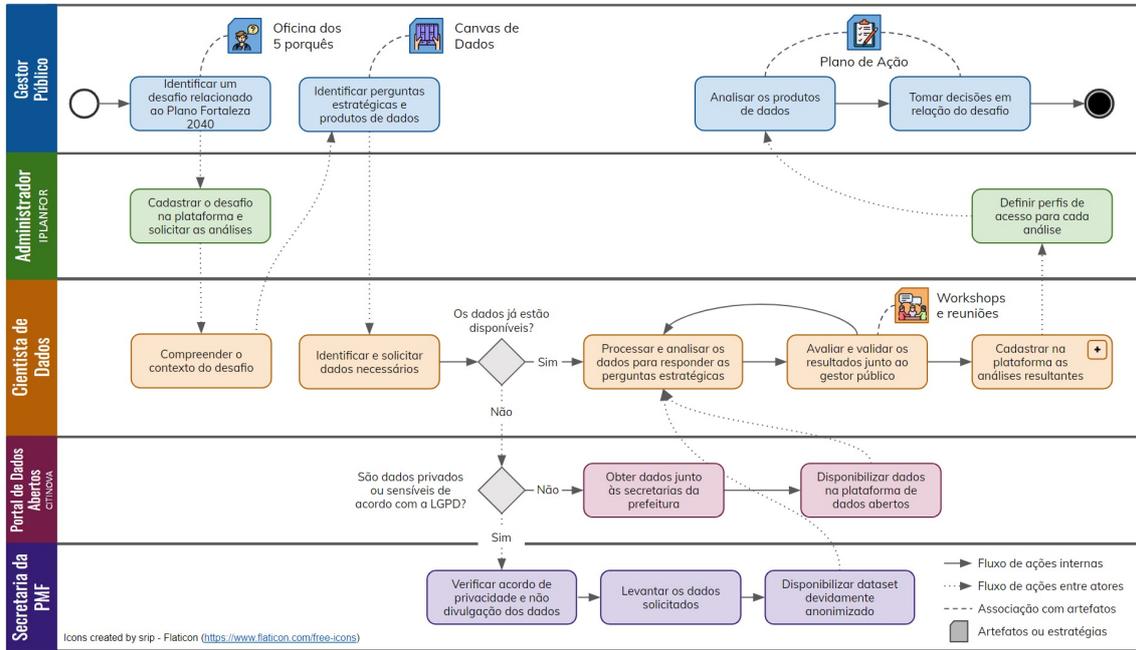


Figura 2. Workflow da plataforma Big Data Fortaleza

Os dados são então obtidos junto às secretarias da Prefeitura ou outras fontes e os analíticos são produzidos pela equipe do projeto, seguindo as normas da Lei Geral de Proteção de Dados Pessoais (LGPD). Após a validação das análises, a plataforma permite que os cientistas de dados incluam dois tipos de *scripts*: i) *scripts* responsáveis pela execução do *pipeline* de dados e ii) os *scripts* responsáveis pela execução dos algoritmos de *Machine Learning*. O primeiro conjunto de *scripts* pode ser utilizado para tarefas de pré-processamento (limpeza e transformação) e para aplicação da lógica de negócio requerida para construir os analíticos (tabelas, gráficos e indicadores). O segundo conjunto de *scripts* permite ao cientista de dados criar cenários preditivos a serem explorados pelos gestores. Ao final, definem-se os perfis de acesso para cada análise e os gestores podem analisar os analíticos a fim de auxiliar a tomada de decisão em relação ao desafio inicialmente proposto.

A arquitetura da plataforma (Figura 3) foi inspirada na arquitetura de processamento Delta Lake [Armbrust et al. 2020] e implementada com serviços Amazon AWS. O fluxo de dados inicia-se no carregamento manual dos dados ou na conexão direta com algum serviço que realiza a transferência de dados para o primeiro *bucket* (chamado de *Bronze*) no serviço S3. Tarefas de ETL (*Extract, Transform, and Load*) são realizadas para limpeza e transformação dos dados brutos – em geral, são aplicados filtros para remoção de linhas e colunas que não dizem respeito ao desafio estudado e é realizada normalização de colunas, como nomes de bairros, etapa fundamental para a produção de mapas coropléticos. Os resultados da primeira etapa são escritos em um segundo *bucket* (*Silver*). Então, uma função especializada nas regras de negócio do desafio que esteja sendo trabalhado realiza tarefas de preparação (e.g. agregação) para que os dados sejam

utilizados pelos analíticos no *front-end* da plataforma. Esses dados são armazenados no terceiro *bucket* (*Gold*).

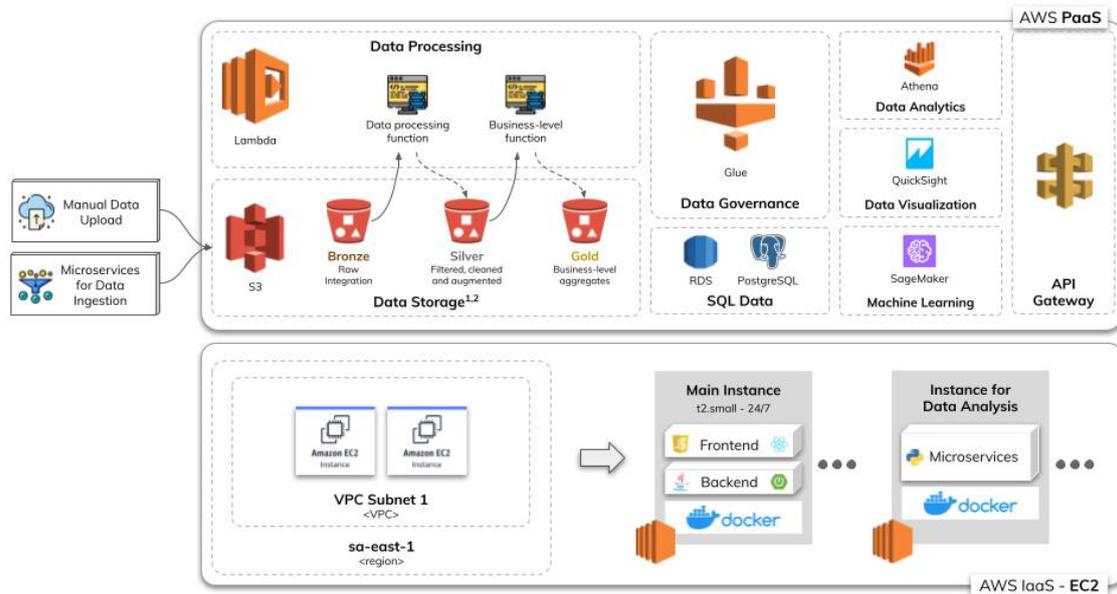


Figura 3. Arquitetura *Big Data* em nuvem da plataforma Big Data Fortaleza

A plataforma utiliza outros serviços gerenciados pela AWS, como o RDS para armazenar dados relativos ao funcionamento da plataforma, como o controle de acesso, perfis e a organização dos analíticos. O AWS Glue tem papel importante na integração de dados pela sua capacidade de criar abstrações de tabelas sobre os diversos dados fornecidos como entrada na Plataforma, que podem estar em diversos formatos. Serviços como SageMaker, QuickSight e Athena são usados em tempo de desenvolvimento, pelos cientistas de dados, para dar maior agilidade na produção dos analíticos e funções de processamento e de regras de negócio.

A plataforma foi desenvolvida utilizando a metodologia ágil Scrum e seguindo boas práticas de Engenharia de *Software* no que concerne Engenharia de Requisitos, Integração Contínua, Testes e Validação de *Software* e Experiência e Interface do Usuário (UI/UX Design). Em termos de tecnologias utilizadas, a plataforma, executada na infraestrutura de nuvem da AWS, utiliza o *framework* Java Spring Boot como principal tecnologia no *back-end* e ReactJS como principal tecnologia no *front-end*.

6. Resultados do Estudo de Caso

O projeto Big Data Fortaleza começou em julho de 2022 e definiu-se como primeiro desafio a ser atacado a “Demanda de Vagas em Creche”. Então, a equipe empreendeu diversas iniciativas para apresentar resultados concretos nos três primeiros meses do projeto para a entrega do MVP. No primeiro trimestre de 2023, os esforços foram direcionados a outros desafios que envolvem a Primeira Infância, educação e a saúde.

A partir do Canvas de Dados, perguntas estratégicas foram elencadas. Conforme a metodologia do projeto, os principais resultados se materializam por meio de análises

de dados cadastrados na plataforma para responder às perguntas estratégicas do desafio. Algumas das análises realizadas na plataforma são apresentadas a seguir.

Relação entre Oferta e Demanda A Figura 4 traz a relação entre oferta e demanda de vagas por creche tendo como referência o indicador Taxa de Escolarização Bruta do Plano Municipal de Educação (PME). Percebe-se que seriam necessárias mais de 20 mil novas vagas para alcançar a meta do PME em 2021.

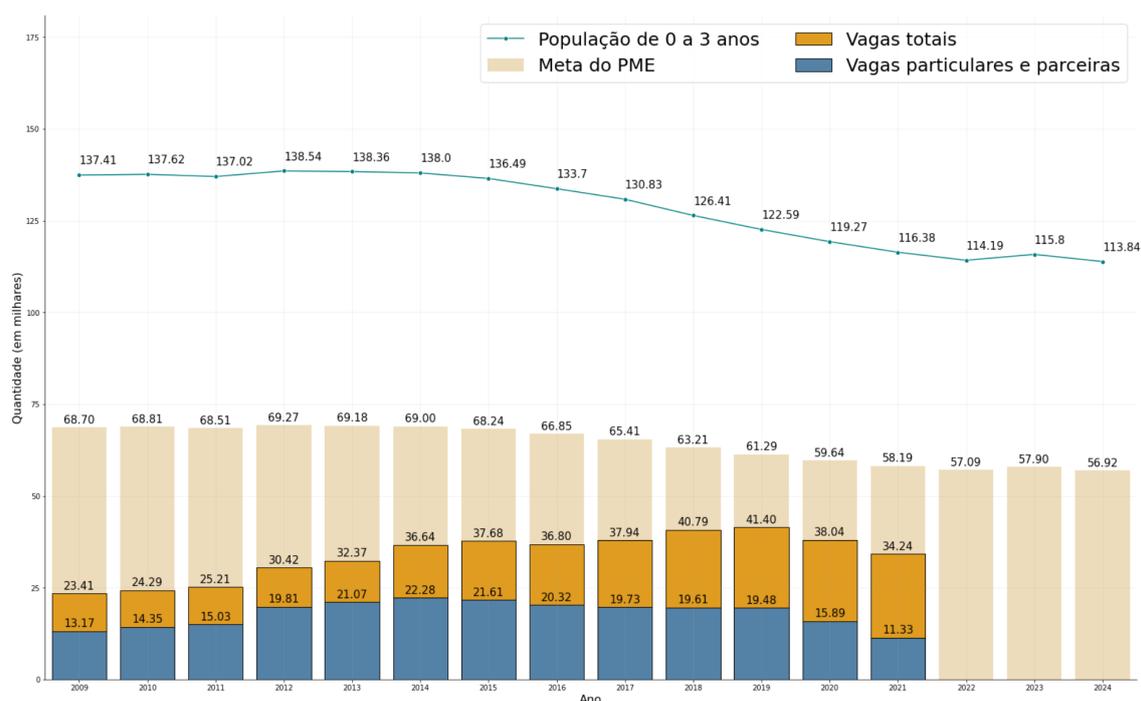


Figura 4. Relação entre oferta e demanda de creches em Fortaleza. Fonte: Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (sistema da Secretária Municipal de Saúde - SMS).

Predição da população de 0 a 3 anos e vagas ofertadas ao longo dos anos O analítico apresentado na Figura 5 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. Assim, a população de 0 a 3 anos representa o público potencial das vagas em creches, sendo utilizada para caracterizar a demanda.

De posse da série histórica de 2009 a 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 será alcançada apenas em 2030 (seis anos após 2024, que foi o ano previsto pelo PME).

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

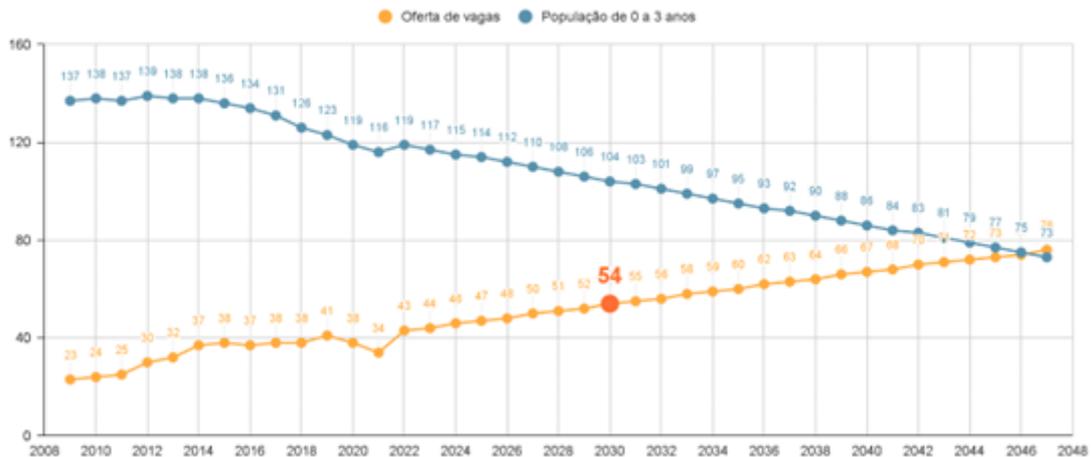


Figura 5. Regressão linear da população de 0 a 3 anos e vagas ofertadas.

integrada autorregressiva (ARIMA), estão sendo avaliados para garantir maior confiabilidade às previsões feitas na plataforma. Para estimar o ano no qual a PMFatingirá 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 a meta do PME.

Predição da população de 0 a 3 anos A predição da população dos bairros foi realizada, permitindo análises com menor granularidade. A Figura 6 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 2009 e 2021. A Figura 7 mostra a evolução da população em bairros distintos da cidade.

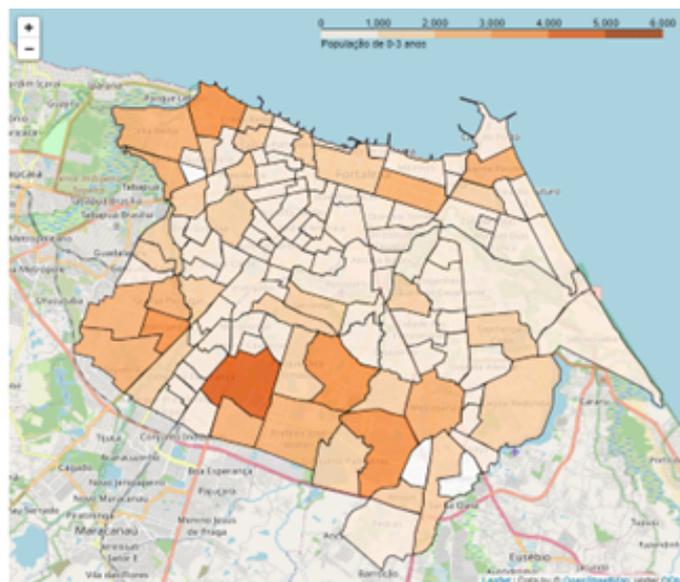


Figura 6. Mapa coroplético com população de 0 a 3 anos nos bairros de Fortaleza.
Fonte: SINASC/SIM extraído do TABNET (SMS).

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 anali-

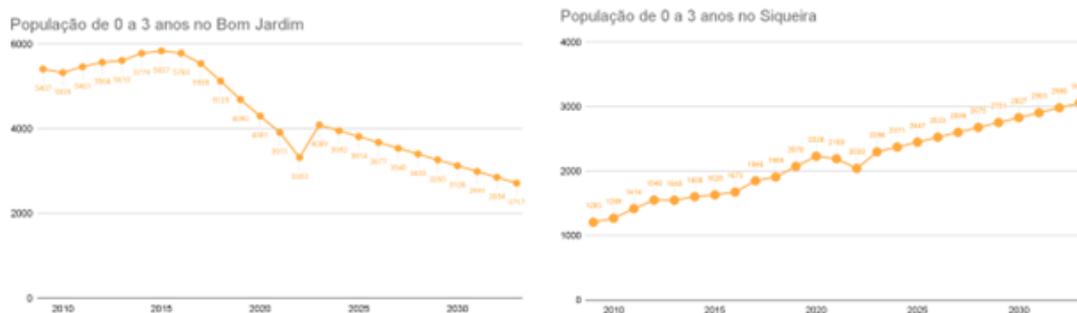


Figura 7. Evolução da população (predição em regressão linear) da população de 0 a 3 anos nos bairros Bom Jardim (esquerda) e Siqueira (direita). Fonte dos dados: SINASC/SIM extraído do TABNET (SMS).

sada (Figura 8). 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 IDH mais alto, possuem poucas creches públicas.

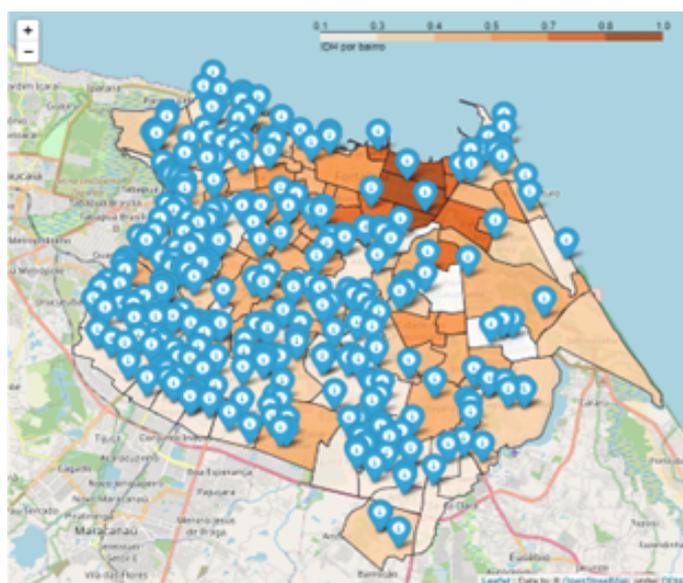


Figura 8. Relação entre número de creches e IDH dos bairros. Fonte dos dados: Censo Escolar (INEP) e SINASC/SIM extraído do TABNET (SMS).

7. Discussão

Durante o desenvolvimento da plataforma, diferentes desafios tiveram que ser mitigados com o propósito de entregar esse sistema de *Big Data* para a cidade de Fortaleza. Neste cenário, são descritos abaixo alguns desses desafios e lições aprendidas.

Ter um foco em um problema específico. O primeiro desafio foi definir a primeira secretaria parceira do projeto, pois, a partir dela, aplicaríamos oficinas para a escolha do projeto-piloto. Por meio das oficinas, percebeu-se que na Prefeitura existem muitos desafios e com o decorrer do projeto os gestores solicitaram demandas que destoavam da temática do projeto-piloto. Desse modo, é importante garantir que

as ações sejam direcionadas com um propósito em particular. No estudo de caso deste artigo, por exemplo, foi essencial o foco no problema de demanda de creches para concentrar os dados de diferentes secretarias em torno desse problema.

Burocracia relacionada ao acesso de dados. No início do projeto, todos os analíticos foram desenvolvidos com dados públicos, ao decorrer dos meses, viu-se a necessidade de utilizar dados sensíveis/privados para a elaboração de analíticos mais robustos, assertivos e personalizados. Então, a equipe do Big Data Fortaleza e do IPLANFOR consultaram uma equipe jurídica e elaboraram termos de confidencialidade (em inglês, *Non Disclosure Agreement* - NDA) que foram assinados entre os membros do projeto e as secretarias envolvidas. Esse termo foi elaborado seguindo as normas da LGPD.

Inconsistência dos dados. Após o recebimento dos dados, a equipe de ciência de dados fez uma Análise Exploratória de Dados (EDA). A EDA é usada por cientistas de dados para analisar conjuntos de dados e resumir suas principais características. A EDA permite determinar a melhor forma de controlar as fontes de dados para obter as respostas que você precisa, tornando mais fácil para os cientistas de dados descobrir padrões, detectar anomalias, testar uma hipótese ou verificar suposições. Após a EDA, constatou-se que existem inúmeros dados duplicados e inconsistentes, por exemplo, idade negativa, data de nascimento futura, etc.

Diferentes stakeholders. O projeto Big Data Fortaleza tem duração de 2 anos e 6 meses. Visto isso, ao longo do projeto serão envolvidos diversos desafios, secretarias e *stakeholders*. Diferentes *stakeholders* geram diferentes demandas, influenciam nas restrições e nos requisitos a serem colocados em prática na gestão de projetos. Com isso, para evitar inconsistências e tratar requisitos conflitantes, é importante aplicar boas práticas de gestão dos requisitos, além de tentar identificar dados em comum que atendem a demandas de diferentes secretárias.

8. Conclusão

Serviços de cidades inteligentes, como aqueles ligados à saúde, economia, turismo, educação, têm se beneficiado com a transformação digital. Nesse contexto, soluções que envolvem *Big Data* são essenciais para apoiar os tomadores de decisão.

A plataforma Big Data Fortaleza, apresentada neste artigo, é uma solução de software que visa impactar diariamente o processo de tomada de decisão dos gestores públicos da Prefeitura Municipal de Fortaleza (PMF). Ao ir além das plataformas de análise tradicionais, focando não somente em visualizações de dados, mas prioritariamente em predições para previsão de demandas futuras, espera-se que os gestores da PMF possam ter subsídios diferenciados, tanto para analisar fidedignamente os cenários atuais, quanto para a predição de demandas futuras.

Isso possibilita 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 proposta, bem como o catálogo de dados do projeto, possam ser úteis a outras iniciativas da PMF. Além disso, as lições aprendidas discutidas no artigo podem auxiliar na implantação deste tipo de solução em outras cidades.

Como trabalhos futuros, espera-se expandir as análises para apoiar a tomada de decisões nos desafios de déficit habitacional e as iniciativas de combate e prevenção contra uso de drogas na cidade de Fortaleza.

Agradecimentos

Ao CNPq pela bolsa de produtividade da Rossana M. C. Andrade 1D (306362/2021-0) e à Funcap/Iplanfor pelo apoio financeiro na execução do projeto “Big Data Fortaleza”.

Referências

- Armbrust, M., Das, T., Sun, L., Yavuz, B., Zhu, S., Murthy, M., Torres, J., van Hovell, H., Ionescu, A., Łuszczak, A., et al. (2020). Delta lake: high-performance acid table storage over cloud object stores. *Proceedings of the VLDB*, 13(12):3411–3424.
- Fan, W. and Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD explorations newsletter*, 14(2):1–5.
- Furtado, L. S., da Silva, T. L. C., Ferreira, M. G. F., de Macedo, J. A. F., and Moreira, J. K. d. M. L. C. (2023). A framework for Digital Transformation towards Smart Governance: using big data tools to target SDGs in Ceará, Brazil. *Journal of Urban Management*, 12(1):74–87. Digital Technologies in Urban Planning and Urban Management.
- Inácio, A., Leal, A., Gava, V., and Santos, A. (2021). Salas de Situação: Implantação de monitoramento inteligente como suporte às Políticas Públicas. In *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, pages 283–290. SBC.
- Lai, C. S., Jia, Y., Dong, Z., Wang, D., Tao, Y., Lai, Q. H., Wong, R. T., Zobia, A. F., Wu, R., and Lai, L. L. (2020). A review of technical standards for smart cities. *Clean Technologies*, 2(3):290–310.
- Oliveira, J. L. A., Parente, R. G. A., Silva, I. M. F., and Oliveira Neto, F. M. d. (2020). Análise da evolução dos padrões de mobilidade no sistema de transporte coletivo por ônibus em Fortaleza utilizando Big Data. In *Anais do Congresso de Pesquisa e Ensino em Transportes*, pages 2030–2041, [s.l.]. Associação Nacional de Pesquisa e Ensino em Transportes.
- Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, 91:620–633.
- Ramos, G., Fernandes, D., Coelho, J., and Aquino, A. (2022). Data lakes lógicos como plataformas para dados governamentais em sociedades e cidades inteligentes. In *Anais Workshop de Computação Aplicada em Governo Eletrônico*, pages 215–226. SBC.
- Razmjoo, A., Østergaard, P. A., Denai, M., Nezhad, M. M., and Mirjalili, S. (2021). Effective policies to overcome barriers in the development of smart cities. *Energy Research & Social Science*, 79:102175.
- Sánchez-Corcuera, R., Nuñez-Marcos, A., Sesma-Solance, J., Bilbao-Jayo, A., Mulero, R., Zulaika, U., Azkune, G., and Almeida, A. (2019). Smart cities survey: Technologies, application domains and challenges for the cities of the future. *International Journal of Distributed Sensor Networks*, 15(6):1550147719853984.
- Santos, M. T., Gasparini, I., Frigo, L., and Mulle, L. D. (2021). Ferramenta de visualização de Dados Abertos do Portal de Transparência da Câmara Municipal da Cidade de Florianópolis. In *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, pages 71–82, Porto Alegre, RS, Brasil. SBC.