



Centro Universitário de Brasília (CEUB)

Faculdade de Tecnologia e Ciências Sociais Aplicadas (FATECS)

Thales Rassi Porto de Matos - 22400186

Gabriel Marques da Rocha - 22451254

Gabrielle Gutierrez - 22350026

Pedro Klein - 22105154

Matheus de Moraes - 22352763

Henrique Lessa - 22402204

Relatório de Análise Exploratória (EDA)

Brasília

2025

Thales Rassi Porto de Matos

Gabriel Marques da Rocha

Gabrielle Gutierrez

Pedro Klein

Matheus de Moraes

Henrique Lessa

Relatório de Análise Exploratória (EDA)

Atividade avaliativa apresentada à Faculdade de Tecnologia e Ciências Sociais Aplicadas (FATECS)

, do Centro Universitário de Brasília (CEUB) como parte integrante do currículo da disciplina Projeto Integrador 1, da graduação em Ciência da computação.

Professora responsável: Kadidja Valeria Reginaldo de Oliveira

Brasília

2025

SUMÁRIO

Passo 1: Sessão Preliminar

1.1. Resumo Executivo

1.2. Introdução

Passo 2: Metodologia e Fonte de Dados

2.1. Descrição da Fonte de Dados

2.2. Preparação e Limpeza de Dados

Passo 3: Estatísticas Descritivas Chave

3.1. Análise Univariada (Variável Única)

3.2. Análise Bivariada (Relação entre Duas Variáveis)

Passo 4: Apresentação e Interpretação dos Gráficos

4.1. Integração com a Dashboard

Passo 5: Conclusão e Próximos Passos

5.1. Conclusão

5.2. Próximos Passos e Limitações

Passo 1: Sessão Preliminar

1.1. Resumo Executivo

Este relatório apresenta os insights estratégicos derivados da Análise Exploratória de Dados (EDA) do projeto "Dashboard Radiologia DF", que mapeou a demanda e acessibilidade de serviços de diagnóstico por imagem no Distrito Federal. A análise revelou três padrões estruturais críticos. Primeiramente, há uma acentuada desigualdade territorial na distribuição de equipamentos, com alta concentração no núcleo central de Brasília (como evidenciado pelo PET/CT e Raio X de baixa potência). Em segundo lugar, existe uma disparidade significativa entre os setores público e privado, onde a iniciativa privada detém uma parcela majoritária de tecnologias de alto custo, comprometendo a equidade assistencial para a população dependente do SUS (exemplo: Gama Câmara e PET/CT). Por fim, foram identificadas fragilidades estruturais na acessibilidade digital do portal DataSUS, com alta incidência de erros de contraste e alertas que prejudicam a navegação e o acesso equitativo à informação pública. As conclusões indicam a necessidade urgente de planejamento estratégico para ampliar a oferta em regiões periféricas e fortalecer a rede pública.

1.2. Introdução

O projeto, que culminou na criação do dashboard interativo "Radiologia DF", teve como objetivo principal realizar uma análise abrangente e integrada para caracterizar o panorama da oferta, demanda e acessibilidade dos serviços de diagnóstico por imagem no Distrito Federal.

O desafio central abordado era identificar e quantificar os padrões estruturais que impactam a eficiência e a equidade no acesso à saúde na região. A pergunta de pesquisa que guiou o projeto pode ser resumida em: Quais são as desigualdades territoriais, operacionais e tecnológicas que caracterizam a infraestrutura de diagnóstico por imagem no Distrito Federal, e como elas comprometem o acesso equitativo à saúde para a população usuária?

Essas questões são relevantes devido à concentração tecnológica e à necessidade de fornecer subsídios que revelem implicações importantes para a gestão pública e para a população.

Passo 2: Metodologia e Fonte de Dados

2.1. Descrição da Fonte de Dados

A Análise de Dados do projeto "Dashboard Radiologia DF" foi realizada a partir da consolidação de múltiplas fontes de dados abertos, fornecidas por órgãos federais. As fontes primárias utilizadas incluem:

CNES (Cadastro Nacional de Estabelecimentos de Saúde), IBGE (Instituto Brasileiro de Geografia e Estatística), PNAD-IPEDF (Instituto de Pesquisa e Estatística do Distrito Federal) e SISCAN (Sistema de Informação do Câncer).

2.2. Preparação e Limpeza de Dados

Embora parte dos dados utilizados já estivesse organizada por se tratar de bases oficiais de acesso público do governo, a coleta em si foi realizada manualmente a partir das plataformas oficiais, o que exigiu atenção e validação contínua dos registros. Além disso, diversas inconsistências e desalinhamentos ainda estavam presentes nos datasets, e para tratá-los foram desenvolvidos scripts de ETL em Python utilizando a biblioteca Pandas, responsáveis por limpar, padronizar, transformar e estruturar as informações. Todo esse material também foi reorganizado de forma adequada para possibilitar a carga no banco de dados com integridade e consistência. No futuro, pretende-se automatizar tanto o processo de coleta quanto o pipeline de tratamento, garantindo atualizações mais rápidas, escaláveis e menos suscetíveis a erros humanos.

Passo 3: Estatísticas Descritivas Chave

3.1. Análise Univariada (Variável Única)

Variável	Métrica Chave	Valor/Resultado	Implicação
Acessibilidade Digital (AIM Score)	Mediana	Próxima de 4	Indica que a maioria das páginas do portal DataSUS atinge um nível insuficiente ou intermediário de qualidade estrutural.
Acessibilidade Digital (Erros WAVE)	Mediana de Erros/ Dispersão (IQR)	Aproximadamente 35 /Muito Alta (caixa extremamente ampla na métrica Errors	Aponta para uma alta heterogeneidade na qualidade estrutural das páginas, com algumas próximas a falhas críticas de construção.
Indisponibilidade (Raio X 100 a 500 mA)	% Em Funcionamento	96.2%	Demonstra alta disponibilidade operacional neste segmento.
Indisponibilidade (Mamógrafo)	% Parado	18.2%	Indica um gargalo potencial, com quase um quinto dos equipamentos fora de operação.

3.2. Análise Bivariada (Relação entre Duas Variáveis)

Esta análise quantifica as relações e as desigualdades mais impactantes, as quais definem o cenário de acesso à saúde no DF.

Disparidade Setorial (Setor Privado x Setor Público)

Equipamento (Alta Complexidade)	Desigualdade (%)	Posição Estratégica
Gama Câmara	3000.00%	A posse da rede privada é 30 vezes maior que a pública.

PET/CT	800.00%	A posse da rede privada é 8 vezes maior que a pública.
Raio X para Densitometria Óssea	662.50%	Disparidade substancial, com 61 equipamentos privados contra 8 públicos.
Raio X de 100 a 500 mA	2.50%	Demonstra a menor disparidade percentual, com 82 privados e 80 públicos.

Eficiência Operacional (Tempo de Espera para Mamografias)

UF	% Exames em 0–10 dias	% Exames em >30 dias	Conclusão
Distrito Federal (DF)	53.1% ¹⁶¹⁶¹⁶¹⁶	22.8% ¹⁷¹⁷¹⁷¹⁷	Maior proporção de exames realizados no curto prazo, indicando eficiência superior a SP e RJ ¹⁸ .
São Paulo (SP)	33.1% ¹⁹	45.3% ²⁰	
Rio de Janeiro (RJ)	42.2% ²¹	36.4% ²²	

O DF, embora tenha uma proporção superior de exames executados em até 10 dias, ainda apresenta um percentual significativo (22.8%) de exames com espera superior a 30 dias.

Passo 4: Apresentação e Interpretação dos Gráficos

Eixo Temático	Gráfico Apresentado	Objetivo Analítico
I. Desigualdade Territorial e Concentração	Mapa de Calor: Concentração de PET/CT na região central de Brasília.	Evidenciar a assimetria espacial da oferta de tecnologias avançadas, reforçando a desigualdade territorial no acesso à saúde.
	Mapa de Calor: Raio X de 100 a 500 mA (Distribuição menos concentrada) ³	Mapear a distribuição de equipamentos de média complexidade para identificar áreas de relativa atenuação das desigualdades.
II. Disparidade Setorial e Eficiência	Tabela/Matriz de Comparação: Disparidade Setor Público vs. Setor Privado (em %).	Quantificar o desequilíbrio na posse de tecnologias de alto custo (ex: Gama Câmara 3000.00%), que compromete a equidade assistencial do SUS.
	Gráfico de Barras Comparativas: Tempo de Espera para Mamografias (DF x SP).	Avaliar a eficiência operacional do DF, comparando a proporção de exames realizados em diferentes intervalos de tempo (ex: 53.1% em 0–10 dias no DF).
III. Acessibilidade Estrutural	Boxplots: Distribuição de Erros, Contrast Errors e Alerts (WAVE).	Diagnosticar o padrão de acessibilidade insuficiente no portal DataSUS (mediana elevada de erros e baixa mediana do AIM Score)
IV. Tendências e Disponibilidade	Gráfico de Linha: Tendência temporal de profissionais de radiologia por categoria.	Analisar o crescimento quantitativo da força de trabalho ao longo do tempo.

	Gráfico de Barras: Status Operacional (Ex: Raio X de 100 a 500 mA).	Visualizar a proporção de equipamentos em funcionamento vs. parados para identificar gargalos de manutenção e disponibilidade.
--	---	--

4.1. Integração com o Dashboard

A integração de todos os gráficos e dados em uma plataforma única (<https://radiologia-df.streamlit.app/>)

Passo 5: Conclusão e Próximos Passos

5.1. Conclusão

A análise integrada das visualizações, métricas e indicadores do projeto revelou que o panorama dos serviços de diagnóstico por imagem no Distrito Federal é caracterizado por uma heterogeneidade estrutural, onde avanços pontuais coexistem com desigualdades sistêmicas. Os principais padrões estruturais identificados demonstram que existe uma acentuada desigualdade territorial, confirmada pela análise cartográfica que mostra a concentração de equipamentos de alto custo (como o PET/CT e Raio X de baixa potência) majoritariamente no núcleo central de Brasília, penalizando regiões periféricas e sobrecarregando as unidades centrais. Além disso, foi confirmada uma disparidade extrema entre o Setor Público e o Privado, com uma discrepância alarmante na posse de tecnologias de alta complexidade, onde a iniciativa privada detém uma parcela exponencialmente maior da oferta, atingindo o pico em equipamentos como a Gama Câmara (3000.00%) e o PET/CT (800.00%), o que compromete diretamente a equidade assistencial para a população dependente do SUS. As investigações também apontaram para fragilidades de acessibilidade digital no portal DataSUS, onde a avaliação das métricas WAVE revelou problemas estruturais e sistêmicos, com a mediana do AIM Score próxima de quatro, o que compromete a transparência e o acesso equitativo à informação pública. Por fim, em termos de eficiência operacional relativa, o DF demonstrou um desempenho favorável na dinâmica de tempo de espera para mamografias, com 53.1% dos exames executados em 0–10 dias; no entanto, a

persistência de gargalos específicos, com 22.8% dos exames excedendo 30 dias, sugere a necessidade de vigilância contínua para aprimoramento. Em suma, os achados fornecem uma base sólida para a formulação de políticas públicas mais direcionadas, com foco na ampliação da oferta em regiões periféricas e no fortalecimento estratégico da rede pública.

5.2. Próximos Passos e Limitações

A continuação do projeto "Dashboard Radiologia DF" está focada em capitalizar os achados da Análise Exploratória de Dados, aprimorando a robustez, escalabilidade e interatividade da plataforma. As próximas etapas buscam transformar o projeto em uma ferramenta de análise de dados de atualização contínua e maior acessibilidade.

As próximas etapas, que constituem as Implementações Futuras, preveem avanços em três grandes frentes. Na área de Automação e Integração de Dados, os esforços serão direcionados para a integração com a API do CNES (DATASUS), desenvolvendo rotinas para consumir o serviço SOA-CNES, o que substituirá a extração manual e garantirá atualizações mais rápidas e seguras diretamente da fonte oficial. Complementarmente, será implementado o Web Scraping para capturar dados essenciais de portais sem API pública, como SISCAN e SISMAMA, aumentando a profundidade e a abrangência dos indicadores. Todo esse processo culminará na Automação do Pipeline de ETL, com a criação de um fluxo automatizado de Extração, Transformação e Carregamento utilizando ferramentas como Airflow ou Prefect, garantindo que os dados estejam sempre atualizados e padronizados.

Por fim, o Aprimoramento da Interatividade e Acessibilidade será o foco da experiência do usuário, com a integração de um Assistente de Inteligência Artificial, que funcionará como um módulo de IA com interação por voz (LLM), permitindo que o usuário converse com o painel para solicitar informações e receber explicações detalhadas sobre gráficos e métricas, reduzindo barreiras de acessibilidade e fornecendo uma experiência conversacional e assistiva.