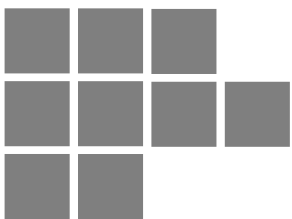




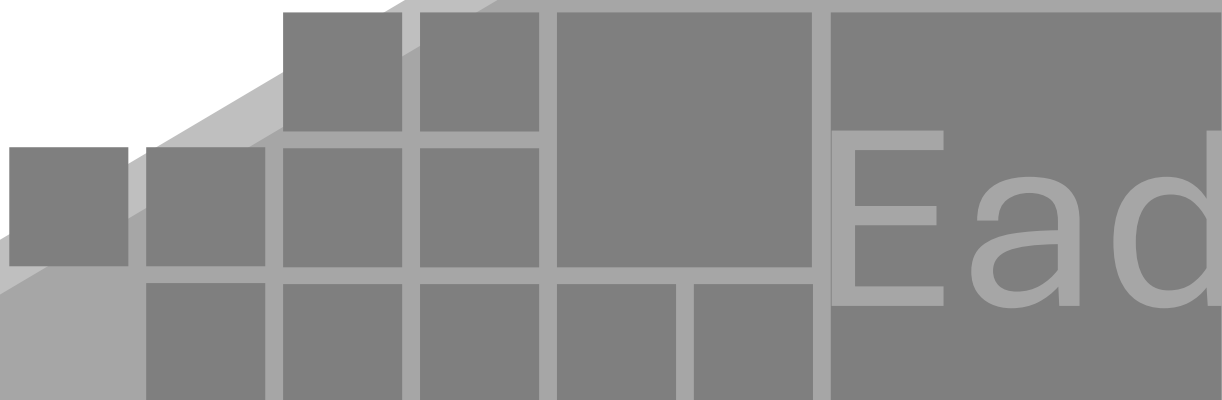
Universidade Presbiteriana

Mackenzie

CIÊNCIA DE DADOS



Projeto **Aplicado** **01**

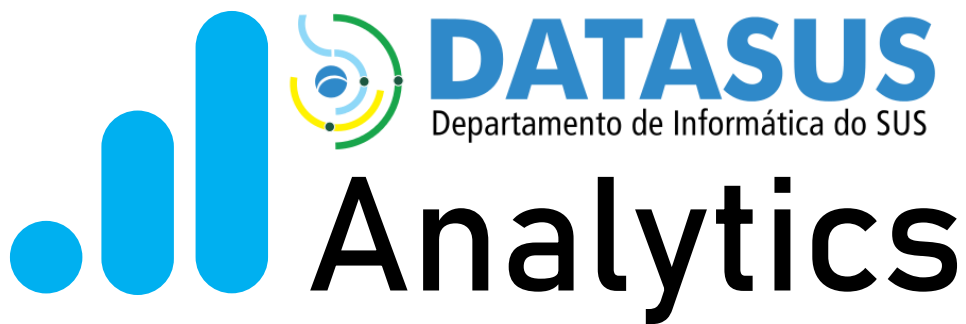
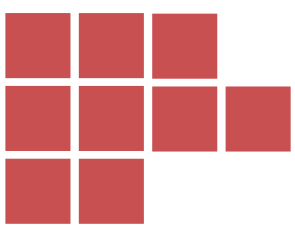


Curso Ciência de Dados

Universidade Presbiteriana Mackenzie

2º semestre – ano letivo: 2025

CIÊNCIA DE DADOS



Este Projeto tem como objetivo principal propor uma análise crítica através de um estudo prático do banco de dados do DataSUS. O estudo é baseado no uso de técnicas e ferramentas desenvolvidas no curso de Ciência de Dados da Universidade Presbiteriana Mackenzie com a orientação do professor Lucas Cerqueira Figueiredo.

Sumário

Apresentação.....	4
Contexto do Estudo	5
Referências de Aquisição do Dataset	6
Origem dos Dados:.....	7
Limitações de Uso:.....	7
Período da Coleta:	7
Descrição da Origem:.....	8
Datasets em vigor:	8
Pipeline	9
Análise Exploratória	11
1. Análise Descritiva: "O que aconteceu?"	12
2. Análise Diagnóstica: "Por que aconteceu?"	15
3. Descrição das Variáveis da Análise:	17
4. Considerações e Próximos Passos	18
Considerações Finais	19
Glossário	20
Link Repositório ONLINE	22
Referências Bibliográficas	23
Figuras	23
Tabelas	24
Links	24



Apresentação

O Sistema Único de Saúde (SUS) configura-se como a maior rede pública de saúde do Brasil, garantindo acesso universal e gratuito a serviços de saúde em todos os níveis de atenção, desde a atenção primária até procedimentos de alta complexidade. A escolha do SUS como objeto de estudo se justifica pela vasta disponibilidade de dados públicos e confiáveis, acessíveis através do DataSUS. Esta plataforma governamental centraliza e disponibiliza informações cruciais sobre a saúde da população brasileira, incluindo dados epidemiológicos, assistenciais, demográficos e socioeconômicos. A riqueza e a granularidade desses dados oferecem um potencial significativo para a realização de análises complexas e aprofundadas, permitindo a identificação de padrões, tendências e relações que podem subsidiar a formulação de políticas públicas mais eficazes, a otimização da gestão dos serviços de saúde e a melhor compreensão dos determinantes sociais da saúde no país. Além disso, a transparência e a acessibilidade dos dados do DataSUS fomentam a pesquisa científica independente e a participação da sociedade civil no acompanhamento e na avaliação do sistema de saúde.

Contexto do Estudo

O pré-natal configura-se como um pilar fundamental na assistência à saúde materno-infantil, desempenhando um papel crucial na detecção precoce e no manejo adequado de potenciais complicações gestacionais, como a prematuridade, que podem impactar significativamente tanto a saúde da mãe quanto a do recém-nascido.

Apesar de sua reconhecida importância, a cobertura e a qualidade do acompanhamento pré-natal no Brasil apresentam heterogeneidades significativas, frequentemente associadas às disparidades socioeconômicas e à infraestrutura de saúde disponível nas diferentes regiões do país. Áreas com recursos limitados e menor acesso a serviços de saúde tendem a apresentar índices inferiores de adesão ao pré-natal, o que, por sua vez, pode contribuir para elevadas taxas de mortalidade materna e fetal, evidenciando a urgência de investigar e compreender os fatores que influenciam o acesso e a efetividade desse cuidado essencial.



Figura 1 - Pré-Natal

Referências de Aquisição do Dataset

Os dados utilizados neste estudo serão provenientes de bases de dados públicas e confiáveis, geridas pelo Ministério da Saúde através do Departamento de Informática do SUS (DataSUS) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE).



Figura 2 - SUS



Figura 3 – IBGE

Origem dos Dados:

Sistema de Informações sobre Nascidos Vivos (SINASC) e Sistema de Informações sobre Mortalidade (SIM): Geridos pelo DataSUS, órgão responsável pela coleta, processamento e disseminação de informações sobre saúde no Brasil. Os dados são originados de Declarações de Nascido Vivo (DNV) e Declarações de Óbito (DO) preenchidas em todo o território nacional por profissionais de saúde em hospitais, maternidades e outros estabelecimentos de saúde, além de registros de óbitos em cartórios.

Infraestrutura Hospitalar: Dados provenientes de levantamentos e pesquisas do IBGE, como o Perfil dos Municípios Brasileiros (Munic) e outras pesquisas relacionadas à saúde e infraestrutura. O IBGE é o principal produtor e disseminador de informações estatísticas e geográficas do Brasil.

Limitações de Uso:

Os dados do DataSUS e do IBGE são públicos e de acesso gratuito para fins de pesquisa e estudo. No entanto, é fundamental observar as notas e a documentação de cada base de dados para compreender as especificidades da coleta, as possíveis limitações de cobertura e qualidade, e as recomendações para a interpretação dos resultados. É importante citar adequadamente a fonte dos dados em qualquer publicação ou divulgação dos resultados.

Período da Coleta:

A coleta dos dados abrangerá um período específico a ser definido, buscando a maior abrangência temporal possível para identificar tendências e variações relevantes. A seleção do período considerará a disponibilidade e a comparabilidade dos dados nas diferentes bases, visando uma análise robusta e significativa.

Descrição da Origem:

DataSUS (SINASC e SIM): O DataSUS é uma organização governamental vinculada ao Ministério da Saúde, responsável por produzir e disseminar informações essenciais para a formulação, acompanhamento e avaliação de políticas de saúde no Brasil. Os dados do SINASC e SIM são gerados no contexto do registro administrativo de eventos vitais (nascimentos e óbitos) em todo o país, sendo cruciais para a vigilância epidemiológica e a gestão da saúde.

IBGE (Infraestrutura Hospitalar): O IBGE é uma fundação pública federal responsável por produzir informações estatísticas, geográficas e ambientais do Brasil. Os dados sobre infraestrutura hospitalar são gerados através de censos, pesquisas amostrais e levantamentos específicos, visando fornecer um panorama detalhado das características e da distribuição dos recursos de saúde no território nacional.

Datasets em vigor:

Sistema de Informações sobre Nascidos Vivos (SINASC): Contém informações detalhadas sobre todos os nascimentos ocorridos no Brasil, incluindo características do recém-nascido (peso ao nascer, idade gestacional), da mãe (idade, escolaridade, local de residência) e do pré-natal (número de consultas realizadas). A proposta deste dataset é fornecer dados para o monitoramento da natalidade, das características dos nascidos vivos e da assistência pré-natal.

Sistema de Informações sobre Mortalidade (SIM): Registra informações sobre todos os óbitos ocorridos no Brasil, incluindo dados sobre o falecido (idade, sexo, causa da morte codificada pela CID-10), o local da ocorrência e informações sobre a mãe em casos de óbito fetal ou materno. A proposta é fornecer dados para a vigilância da mortalidade, a identificação de causas de morte e a avaliação de políticas de saúde.

Infraestrutura Hospitalar (IBGE): Apresenta dados sobre o número de hospitais e maternidades por município, a quantidade de leitos disponíveis, a distribuição de profissionais de saúde e outras características da infraestrutura de saúde. A proposta é fornecer informações sobre a disponibilidade e a distribuição dos recursos de saúde no país.

Esses datasets registram fenômenos relacionados à natalidade, mortalidade e à organização do sistema de saúde, permitindo analisar a relação entre a realização do pré-natal, a incidência de complicações na gestação (inferida a partir dos dados de nascimento e óbito) e a disponibilidade e características da infraestrutura hospitalar.

Pipeline



Etapa 1: Definição e Planejamento (Dia 04 de Março a 03 de Abril):

- Definição do problema e objetivos do projeto
- Identificação da organização e área de atuação
- Montagem do grupo e distribuição de responsabilidades
- Elaboração do cronograma de atividades
- Discussão sobre pensamento computacional no contexto organizacional

Etapa 2: Análise Exploratória e Proposta Analítica (Dia 10 de Março a 3 de Abril):

- Definição da proposta analítica
- Coleta e organização do dataset
- Início das análises exploratórias de dados (EDA)
- Identificação de padrões e tendências iniciais
- Ajustes nos dados para futuras modelagens

Etapa 3: Apresentação e Data Storytelling (Dia 04 a 28 de Abril):

- Desenvolvimento dos resultados analíticos
- Aplicação de técnicas de Data Storytelling
- Construção de gráficos e visualizações impactantes
- Elaboração do relatório com insights principais

Etapa 4: Finalização e Apresentação (Dia 29 de Abril a 26 de Maio):

- Ajustes finais e refinamento da análise
- Gravação do vídeo de apresentação (5 minutos)
- Organização da entrega final
- Apresentação dos resultados

Análise Exploratória

A Análise Exploratória de Dados (AED) é uma abordagem fundamental no estudo, focada em investigar conjuntos de dados para sumarizar suas principais características, frequentemente utilizando métodos visuais e estatísticas sumárias. Seu objetivo principal não é testar hipóteses formais ou alcançar conclusões definitivas, mas sim ganhar familiaridade e compreensão sobre os dados. A AED visa identificar padrões, detectar anomalias (outliers), verificar suposições e descobrir relações potenciais entre variáveis, podendo gerar hipóteses para investigações futuras.

Existem quatro tipos principais de análise de dados: descritiva, diagnóstica, preditiva e prescritiva. Num primeiro momento deste estudo, o foco recairá sobre as duas primeiras.

Para esta Análise Exploratória inicial, foram utilizados dados populacionais e de gestantes no Brasil, predominantemente do ano de 2014, obtidos junto ao DataSUS e ao IBGE. É importante ressaltar que, neste estágio preliminar, houve desafios no acesso a dados mais recentes sobre o tema, sendo os registros de 2014 os mais completos disponíveis na fonte consultada até o momento. Embora essa limitação temporal restrinja conclusões sobre a situação atual, esta análise é crucial para a familiarização com os dados e a identificação de padrões preliminares. Estratégias para incorporar dados mais atuais ou complementares serão exploradas em fases subsequentes (em desenvolvimento).

Conforme mencionado, esta fase da Análise Exploratória abrange duas vertentes principais:

1. Análise Descritiva: "O que aconteceu?"

Este é o tipo mais básico de análise, concentrando-se em resumir e descrever dados históricos para entender o que ocorreu em um determinado período. Utiliza métricas simples, estatísticas básicas (quantidades, porcentagens, médias) e visualizações (gráficos de barras, histogramas etc.) para apresentar os dados de forma compreensível e identificar suas características principais. Serve como base para as demais análises.

Nesta análise descritiva inicial, o objetivo foi caracterizar e sumarizar as informações básicas disponíveis. Registrou-se um total de 25.392.922 mulheres em idade fértil (20-34 anos) em 2015 (dado populacional de referência mais próximo). No ano anterior, 2014, houve 2.974.376 gestantes (rede pública e particular) que realizaram ao menos uma consulta de pré-natal. Apesar da defasagem de um ano entre os dados e a natureza distinta das fontes (população geral vs. usuárias de pré-natal) impedir o cálculo preciso de taxas de gravidez ou cobertura, essa comparação inicial oferece uma estimativa da ordem de grandeza da população gestante que interagiu com os serviços de saúde (pública) no período, fornecendo um panorama quantitativo essencial.

Taxa de Consultas de Pré-Natal Realizadas por Gestantes - 2014				
Ranking De Acuracidade	UF	Baixo (01 - 03 Consultas)%	Médio (04 - 06 Consultas)%	Alto (Acima de 6 Consultas)%
1	PR	50,11	30,94	18,96
2	SP	49,23	33,46	17,31
3	SC	50,07	33,10	16,83
4	MG	55,11	29,82	15,07
5	RS	55,64	29,57	14,79
6	MS	54,69	30,90	14,41
7	TO	50,64	35,20	14,16
8	RO	52,16	34,14	13,70
9	MT	55,50	30,84	13,66
10	CE	53,22	33,64	13,14
11	ES	57,09	30,84	12,06
12	RJ	63,14	26,50	10,36
13	PI	60,95	29,80	9,25
14	PE	65,25	25,58	9,17
15	PB	64,11	27,01	8,87
16	AL	65,67	26,68	7,65
17	BA	67,28	25,69	7,03
18	AC	65,01	28,24	6,75
19	RN	70,92	22,43	6,66
20	PA	72,83	21,52	5,65
21	AM	75,71	19,42	4,87
22	SE	79,28	16,00	4,72
23	GO	76,34	18,99	4,67
24	MA	76,11	19,47	4,42
25	RR	76,74	19,18	4,08
26	AP	97,20	2,59	0,21
27	DF	99,68	0,32	0,00
Total de gestantes:		1.701.920,00	877.804,00	394.652,00
		2.974.376,00		

**Tabela 01 - Taxa de Consultas de Pré-Natal Realizadas por Gestantes - 2014*

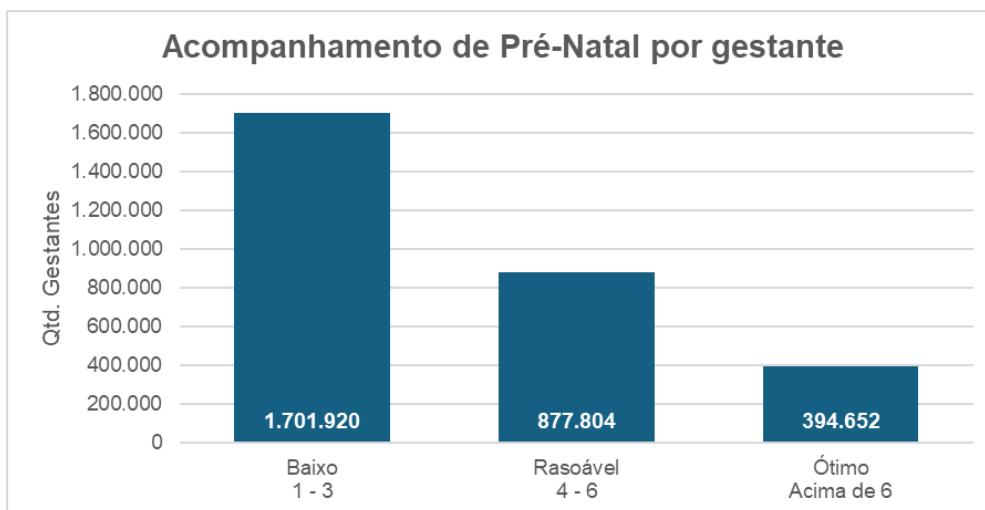


Figura 4 - Gráfico Acompanhamento de Pré Natal por Gestante

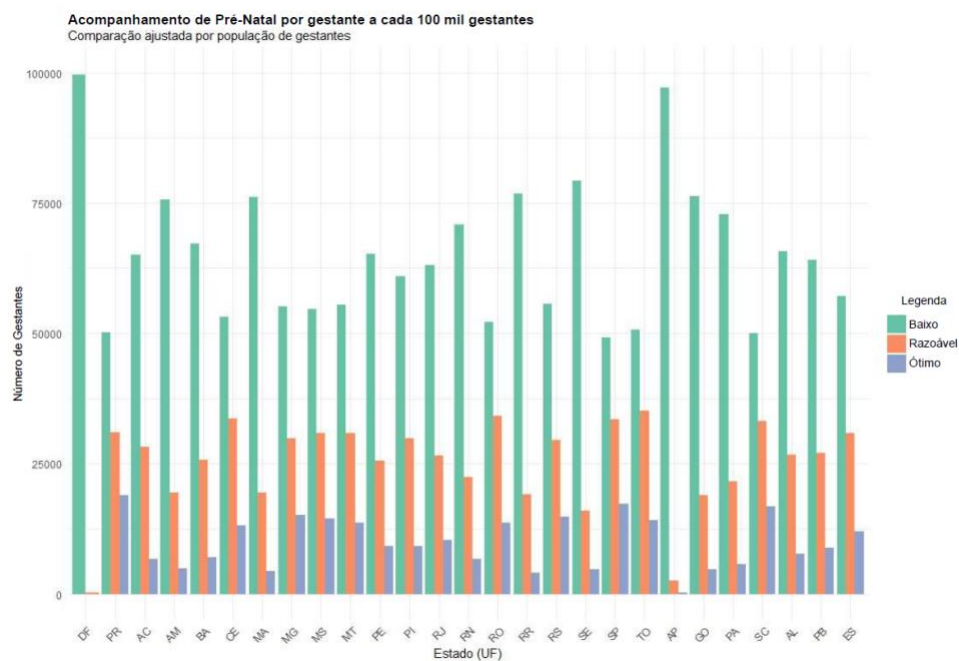


Figura 5-Gráfico Acompanhamento de Pré Natal a cada 100k gestantes

2. Análise Diagnóstica: "Por que aconteceu?"

Indo além da descrição, a análise diagnóstica busca entender as causas raízes dos padrões e resultados observados. Envolve uma investigação mais profunda dos dados, procurando relações, correlações, dependências e identificando fatores que influenciaram um evento ou tendência. Técnicas como análise de correlação e descoberta de padrões são comuns nesta fase.

Média de Consultas de Pré-Natal por Estado (Brasil)

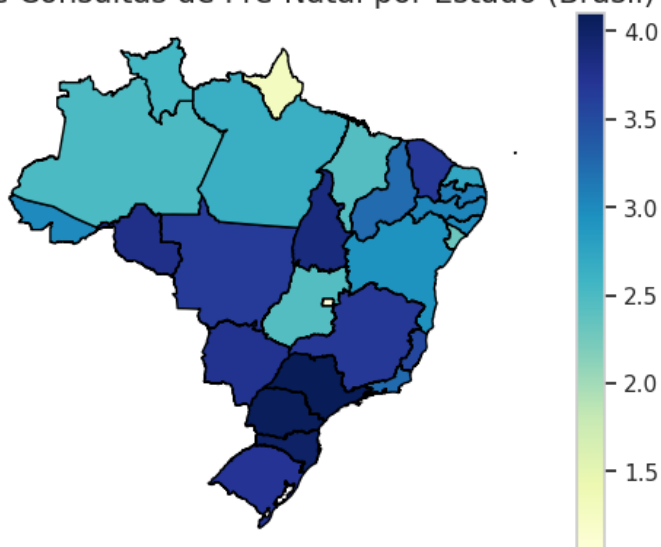


Figura 6 - Média de Consultas de Pré-Natal UF Mapa

Um exemplo prático dessa abordagem é a observação da distribuição das taxas de consultas pré-natais entre as diferentes Unidades Federativas (conforme tabela "Taxa de Consultas de Pré-Natal - 2014"). A identificação de heterogeneidades regionais significativas funciona como um diagnóstico inicial, sinalizando potenciais disparidades no acesso, adesão ou qualidade do acompanhamento pré-natal. Essas variações podem estar correlacionadas a fatores socioeconômicos, diferenças na infraestrutura de saúde local, entre outras variáveis a serem exploradas futuramente.

Previsão de Habitantes por UF para 2015		
UF	DESCRIÇÃO	POPULAÇÃO
SP	São Paulo	21,72%
MG	Minas Gerais	10,21%
RJ	Rio de Janeiro	8,09%
BA	Bahia	7,44%
RS	Rio Grande do Sul	5,50%
PR	Paraná	5,46%
PE	Pernambuco	4,57%
CE	Ceará	4,36%
PA	Pará	4,00%
MA	Maranhão	3,38%
SC	Santa Catarina	3,34%
GO	Goiás	3,23%
PB	Paraíba	1,94%
AM	Amazonas	1,93%
ES	Espírito Santo	1,92%
RN	Rio Grande do Norte	1,68%
AL	Alagoas	1,63%
MT	Mato Grosso	1,60%
PI	Piauí	1,57%
DF	Distrito Federal	1,43%
MS	Mato Grosso do Sul	1,30%
SE	Sergipe	1,10%
RO	Rondônia	0,86%
TO	Tocantins	0,74%
AC	Acre	0,39%
AP	Amapá	0,37%
RR	Roraima	0,25%
Total	204450649	100,00%

**Tabela 02 - Previsão de Habitantes por UF para 2015*

3. Descrição das Variáveis da Análise:

A título de informação adicional, algumas dados foram registrados na execução dos scripts documentados no GitHub:

Tipos de Dados	
CP_UF_IBGE	int64
QT_CONSULT	float64
dtype:	object

*Tabela 03 - Tipos de Dados

Valores ausentes por coluna	
CP_UF_IBGE	0
QT_CONSULT	0
dtype:	int64

*Tabela 04 - Valores ausentes por coluna

Estatísticas descritivas		
	CO_UF_IBGE	QT_CONSULT
count	2.974.376.000.000	2.974.376.000.000
mean	32.433.580	3.568.284
std	9.114.490	2.494.923
min	11.000.000	1.000.000
25%	26.000.000	1.000.000
50%	32.000.000	3.000.000
75%	41.000.000	5.000.000
max	53.000.000	36.000.000

*Tabela 05 - Estatísticas descritivas

4. Considerações e Próximos Passos

Reconhece-se a análise da cobertura e qualidade do pré-natal como um eixo central para a saúde materno-infantil no Brasil. O estudo prático dos dados do DataSUS e do IBGE permite uma análise crítica dessa realidade. A abrangência do DataSUS oferece potencial significativo para análises futuras mais complexas, como a correlação entre o número de consultas pré-natais e com indicadores aos níveis de acesso ao atendimento público sua qualidade, a conscientização do tema e indicadores de mortalidade materna e fetal.

Portanto, esta análise exploratória inicial, mesmo centrada em dados de 2014, cumpre o papel crucial de fornecer uma visão geral quantitativa (descritiva) e apontar possíveis áreas de foco e disparidades (diagnóstica). Ela estabelece a base necessária para investigações futuras mais detalhadas e para o desenvolvimento de estratégias que permitam análises mais atuais. O próximo passo será desenvolver esses dados gerando insights para análises mais eficazes, aprofundando a compreensão dos determinantes sociais da saúde materno-infantil no país.

Considerações Finais

Ao longo deste semestre, o aprofundamento nos conhecimentos do curso de ciência de dados proporcionará as ferramentas necessárias para abordar o tema proposto. Com a justificativa da escolha do DataSUS como fonte de dados robusta e a contextualização da importância do pré-natal, o estudo possui uma base sólida de dados para o seu desenvolvimento.

Todo o processo de coleta, tratamento, análise e visualização dos dados, utilizando as bases e os indicadores públicos, será meticulosamente documentado e disponibilizado em um repositório no GitHub, garantindo a transparência e a replicabilidade da pesquisa. Espera-se que, ao final deste semestre, este trabalho contribua para uma melhor compreensão das dinâmicas envolvidas, possa oferecer insights valiosos para aprimorar as políticas e práticas de saúde materno-infantil no país e contribuir com a comunidade científica, tanto de dados como da área da saúde.

Glossário

Sistema Único de Saúde (SUS): Principal sistema público de saúde do Brasil, que oferece acesso universal e gratuito a serviços de saúde em todos os níveis de complexidade.

DataSUS: Departamento de Informática do SUS, órgão do Ministério da Saúde responsável pela coleta, processamento e disseminação de informações e dados sobre saúde no Brasil.

Dados Epidemiológicos: Informações sobre a ocorrência, distribuição e controle de doenças em populações.

Dados Demográficos: Informações sobre a população, como tamanho, distribuição geográfica, idade, sexo e taxa de natalidade.

Dados Socioeconômicos: Informações sobre as condições sociais e econômicas da população, como renda, escolaridade e acesso a serviços básicos.

Granularidade: Nível de detalhe dos dados. Dados com alta granularidade são mais detalhados e específicos.

Determinantes Sociais da Saúde: Fatores sociais, econômicos, culturais e ambientais que influenciam a saúde das pessoas e das populações.

Pré-natal: Acompanhamento médico e de outros profissionais de saúde durante a gravidez, essencial para a saúde da mãe e do bebê.

Prematuridade: Nascimento do bebê antes de completar 37 semanas de gestação.

Heterogeneidades: Diferenças ou variações significativas.

Mortalidade Materna: Morte de mulheres durante a gravidez, parto ou puerpério (período pós-parto) por causas relacionadas à gestação.

Mortalidade Fetal: Morte do feto antes ou durante o parto.

Dataset: Conjunto de dados estruturado para análise.

Ministério da Saúde: Órgão do governo federal responsável pela formulação e gestão das políticas de saúde no Brasil.

Instituto Brasileiro de Geografia e Estatística (IBGE): Órgão responsável pela produção e disseminação de informações estatísticas, geográficas e ambientais do Brasil.

Sistema de Informações sobre Nascidos Vivos (SINASC): Sistema de informação do DataSUS que coleta dados sobre todos os nascimentos ocorridos no Brasil.

Sistema de Informações sobre Mortalidade (SIM): Sistema de informação do DataSUS que coleta dados sobre todos os óbitos ocorridos no Brasil.

Declarações de Nascido Vivo (DNV): Documento preenchido no momento do nascimento, contendo informações sobre o recém-nascido e a mãe.

Declarações de Óbito (DO): Documento preenchido no momento do óbito, contendo informações sobre o falecido e a causa da morte.

Perfil dos Municípios Brasileiros (Munic): Pesquisa do IBGE que coleta informações sobre as características dos municípios brasileiros, incluindo dados sobre saúde.

Vigilância Epidemiológica: Conjunto de ações que visam o conhecimento e a detecção de mudanças nos fatores determinantes e condicionantes da saúde individual e coletiva, com a finalidade de recomendar e adotar medidas de prevenção e controle de doenças e agravos.

CID-10: Classificação Internacional de Doenças e Problemas Relacionados à Saúde, 10ª revisão, utilizada para codificar as causas de morte.

Software: Software é uma coleção de programas que dizem a um computador como executar tarefas específicas.

Repositório: Um repositório de software é um local de armazenamento de onde pacotes de software podem ser recuperados e instalados em um computador.

GitHub: Plataforma online para armazenar e compartilhar códigos, documentos e outros arquivos de projetos, facilitando a colaboração e a transparência na pesquisa.

Insights: Compreensões ou revelações importantes obtidas a partir da análise dos dados.

Link Repositório ONLINE

<https://github.com/GrupoMackenzie/ProjetoAplicado01-DataScience-Mackenzie-2025.git>

Referências Bibliográficas

Figuras

FIGURA 1 – PRE-NATAL: UNIQUE BABY. [Ilustração sobre a importância da consulta pré-natal]. Unique Baby, s.d. Disponível em: <https://uniquebaby.com.br/blog/gravidez/consulta-pre-natal/>. Acesso em: 20 mar. 2025.

FIGURA 2 – SUS: Wikipedia. [logotipo SUS]. Wikipedia. Disponível em: https://pt.wikipedia.org/wiki/Sistema_%C3%9Anico_de_Sa%C3%BAde. Acesso em: 20 mar. 2025.

FIGURA 3 – IBGE: Instituto Brasileiro de Geografia e Estatística (IBGE). [logotipo SUS]. Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <https://www.infoescola.com/geografia/instituto-brasileiro-de-geografia-e-estatistica-ibge/>. Acesso em: 20 mar. 2025.

Figura 4 - NAGEM, Alberto, 2025. Gráfico Acompanhamento de Pré Natal por Gestante.

Figura 5 - NAGEM, Alberto, 2025. Gráfico Acompanhamento de Pré Natal a cada 100k gestantes.

Figura 6 - NAGEM, Alberto, 2025. Gráfico Acompanhamento de Pré Natal por Gestante UF Mapa.

Tabelas

Tabela 01- MENDES, Gabriel, 2025. Taxa de Consultas de Pré-Natal Realizadas por Gestantes – 2014.

Tabela 02 - MENDES, Gabriel, 2025. Previsão de Habitantes por UF para 2015.

Tabela 03 - NAGEM, Alberto, 2025. Tipos de Dados.

Tabela 04 - NAGEM, Alberto, 2025. Valores ausentes por coluna.

Tabela 05 - NAGEM, Alberto, 2025. Estatísticas descritivas.

Links

DATA SUS. Site da govbr, 2025. Disponível em: <https://datasus.saude.gov.br/>. Acesso em: 10 de março de 2025.

IBGE. Site da govbr, 2025. Disponível em: <https://www.ibge.gov.br/>. Acesso em: 10 de março de 2025.

Curso de Git e GitHub [20 Horas]. Site do CURSOEMVIDEO, 2025. Disponível em: <https://www.cursoemvideo.com/curso/curso-de-git-e-github/>. Acesso em: 10 de março de 2025.

POPULATION PYRAMID, 2025.
<https://www.populationpyramid.net/pt/brasil/2014/>. Acesso em: 03/04/2025