

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL SILVA PEREIRA
IGOR FERREIRA MURRAS

Suporte a construção de indicadores de *near miss* materno no Brasil utilizando dados do
DataSUS

RIO DE JANEIRO
2024

GABRIEL SILVA PEREIRA
IGOR FERREIRA MURRAS

Suporte a construção de indicadores de *near miss* materno no Brasil utilizando dados do
DataSUS

Trabalho de conclusão de curso de graduação
apresentado ao Instituto de Computação da
Universidade Federal do Rio de Janeiro como
parte dos requisitos para obtenção do grau de
Bacharel em Ciência da Computação.

Orientador: Profa. Valéria Menezes Bastos

RIO DE JANEIRO
2024

P436s

Pereira, Gabriel Silva

Suporte a construção de indicadores de near miss materno no Brasil utilizando dados do DataSUS / Gabriel Silva Pereira e Igor Ferreira Murras. – 2024.

77 f.

Orientadora: Valéria Menezes Bastos.

Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)-
Universidade Federal do Rio de Janeiro, Instituto de Computação, Bacharel em
Ciência da Computação, 2024.

1. Saúde pública. 2. DataSUS. 3. Near miss. 4. ETL. 5. ELK. I. Murras,
Igor Ferreira. II. Bastos, Valéria Menezes (Orient.). III. Universidade Federal do
Rio de Janeiro, Instituto de Computação. IV. Título.


GABRIEL SILVA PEREIRA
IGOR FERREIRA MURRAS

Suporte a construção de indicadores de *near miss* materno no Brasil utilizando dados do DataSUS


Trabalho de conclusão de curso de graduação
apresentado ao Instituto de Computação da
Universidade Federal do Rio de Janeiro como
parte dos requisitos para obtenção do grau de
Bacharel em Ciência da Computação.

Aprovado em 17 de Outubro de 2024


BANCA EXAMINADORA:

Documento assinado digitalmente
 VALERIA MENEZES BASTOS
Data: 26/12/2024 12:47:46-0300
Verifique em <https://validar.iti.gov.br>

Valéria Menezes Bastos
D. Sc. (Instituto de Computação / UFRJ)

Documento assinado digitalmente
 ADRIANA SANTAROSA VIVACQUA
Data: 02/01/2025 12:06:24-0300
Verifique em <https://validar.iti.gov.br>

Adriana Santarosa Vivacqua
D. Sc. (Instituto de Computação / UFRJ)

Documento assinado digitalmente
 JULIANA BAPTISTA DOS SANTOS FRANÇA
Data: 21/01/2025 10:27:18-0300
Verifique em <https://validar.iti.gov.br>

Juliana Baptista dos Santos França
D. Sc. (Instituto de Computação / UFRJ)

Dedicamos esse trabalho a todas as mães, profissionais da saúde e pesquisadores.

AGRADECIMENTOS

Gostaríamos de expressar nossa sincera gratidão a todos os envolvidos neste trabalho, cujas contribuições foram fundamentais para a sua realização.

Primeiramente, expressamos nosso reconhecimento aos pesquisadores envolvidos na pesquisa de iniciação científica original que serviu de base para este trabalho. As descobertas e insights provenientes de seu trabalho foram essenciais para nossa investigação, fornecendo um alicerce sólido para nossas próprias análises. Agradecemos especialmente a professora Rosa Domingues, Claudia Medina, Rejane Sobrinho, Valeria Saraceni e Marcos Dias por sua dedicação e expertise no campo.

Além disso, somos imensamente gratos à nossa professora Valéria Bastos, por suas orientações, apoios e conselhos valiosos ao longo deste projeto. Suas contribuições foram essenciais para o desenvolvimento original da pesquisa e também no aprimoramento deste trabalho, e somos profundamente gratos por seu comprometimento conosco. Queremos também agradecer aos nossos colegas de pesquisa, cuja colaboração e troca de ideias foram enriquecedoras e contribuíram significativamente para o desenvolvimento deste trabalho.

Expressamos também nossa profunda gratidão ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e à Fundação Bill e Melinda Gates pelo auxílio financeiro fornecido ao longo do desenvolvimento deste trabalho. O apoio dessas instituições, bem como seu compromisso contínuo com a promoção da pesquisa científica e da inovação tecnológica, foram fundamentais para a realização deste estudo.

Por fim, queremos estender nossos agradecimentos às nossas famílias, cujo apoio e compreensão foram fundamentais durante todo o processo de pesquisa e redação. Agradecemos pelo incentivo, paciência e amor incondicional que nos proporcionaram, tornando possível a realização deste trabalho. Suas contribuições, mesmo que indiretas, são inestimáveis e dignas de reconhecimento.

“A mão que balança o berço é a mão que governa o mundo”

William Ross Wallace

RESUMO

Este trabalho aborda o apoio à construção de indicadores de near missing materno no Brasil, utilizando dados do DataSUS. Por meio de uma exploração detalhada de conceitos, tecnologias, desenvolvimento, infraestrutura, experimentos e resultados, o documento oferece uma visão abrangente sobre o tema. A utilização de tecnologias como *Python*, *Pandas* e *Elasticsearch* aprimora significativamente a análise de dados de saúde materna. Ao integrar fontes de dados-chave e conduzir experimentos, o estudo lança luz sobre a importância da resolução final de indicadores para a vigilância da saúde materna no Brasil, além de detalhar todo o método utilizado durante a pesquisa original.

Palavras-chave: Saúde Pública; DataSUS; *Near Miss*; ETL; ELK.

ABSTRACT

This work addresses support for the construction of maternal near missing indicators in Brazil, using data from DataSUS. Through a detailed exploration of concepts, technologies, development, infrastructure, experiments and results, the document offers a comprehensive overview of the topic. Using technologies such as Python, Pandas and Elasticsearch significantly improves maternal health data analysis. By integrating key data sources and conducting experiments, the study sheds light on the importance of final resolution of indicators for maternal health surveillance in Brazil, in addition to detailing the entire method used during the original research.

Keywords: Public Health; DataSUS; Near Miss;. ETL; ELK.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráficos de Linhas/Barras do Indicador “Consultas pré-natal maior que 7”	40
Figura 2 – <i>Boxplot</i> do Indicador “Consultas pré-natal maior que 7”	40
Figura 3 – Gráfico Linhas/Barras do Indicador “Taxa de mortalidade neonatal precoce”	41
Figura 4 – <i>Boxplot</i> do Indicador “Taxa de mortalidade neonatal precoce”	41
Figura 5 – Gráfico Linhas/Barras do Indicador “Proporção de óbitos maternos investigados”	42
Figura 6 – <i>Boxplot</i> do Indicador “Proporção de óbitos maternos investigados”	42
Figura 7 – Gráfico Linhas/Barras do Indicador “Razão de mortalidade materna”	43
Figura 8 – <i>Boxplot</i> do Indicador “Razão de mortalidade materna”	43
Figura 9 – Gráfico Linhas/Barras do Indicador “Razão de mortalidade materna direta”	44
Figura 10 – <i>Boxplot</i> do Indicador “Razão de mortalidade materna direta”	44
Figura 11 – Gráfico de Pizza do Indicador “Grupo de Robson na taxa global de cesariana”	45
Figura 12 – Gráfico de barras empilhadas do Indicador “Contribuição do Grupo de Robson para a taxa global de cesariana”	46
Figura 13 – Mapa do Indicador “Proporção de nascidos vivos de mulheres no seu primeiro parto”	46
Figura 14 – Gráfico de barras do Indicador “Taxa de fecundidade em menores de 20 anos”	47

LISTA DE CÓDIGOS

Código 1	Consultas em SQL aos dados do SINASC referentes ao Bloco 1 . . .	55
Código 2	Consultas em SQL aos dados do SINASC referentes ao Bloco 2 . . .	57
Código 3	Consultas em SQL aos dados do SINASC referentes ao Bloco 3 . . .	58
Código 4	Consultas em SQL aos dados do SINASC referentes ao Bloco 4 . . .	58
Código 5	Consultas em SQL aos dados do SINASC referentes ao Bloco 5 . . .	61
Código 6	Módulo de Funções Comuns em Python	62
Código 7	Ajustes dos indicadores de “deslocamento ao parto”	66

LISTA DE TABELAS

Tabela 1 – Lista de indicadores preliminares	70
Tabela 2 – Quantidade de municípios por faixas de nascidos vivos	73
Tabela 3 – Lista de indicadores decididos	74

LISTA DE ABREVIATURAS E SIGLAS

CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
GCE	Grand Challenges Explorations
TI	Tecnologia da Informação
DO	Declaração de Óbito
DN	Declaração de Nascido Vivo
SIM	Sistema de Informação sobre Mortalidade
SINASC	Sistema de Informações sobre Nascidos Vivos
CNES	Cadastro Nacional de Estabelecimentos de Saúde
SIH	Sistema de Informações Hospitalares
SIAB	Sistema de Informação da Atenção Básica
SINAN	Sistema de Informação de Agravos de Notificação
ELK	Elasticsearch, Logstash, Kibana
ETL	Extract, Transform, Load
API	Application Programming Interface
PCDaS	Plataforma de Ciência de Dados aplicada à Saúde
SUS	Sistema Único de Saúde
NoSQL	Not only SQL
SQL	Structured Query Language
MS	Ministério da Saúde
JSON	JavaScript Object Notation

SUMÁRIO

1	INTRODUÇÃO	14
1.1	MOTIVAÇÃO	14
1.2	OBJETIVO	16
1.3	TRABALHOS RELACIONADOS	16
1.4	ESTRUTURA DO DOCUMENTO	17
2	ESTADO DA ARTE	18
2.1	CONCEITOS	18
2.1.1	Ciência de Dados	18
2.1.2	Transformação de Dados	18
2.1.3	ETL	18
2.1.4	Visualização de Dados	19
2.1.5	API	19
2.1.6	SQL	20
2.1.7	Banco de Dados Não Relacional (NoSQL)	20
2.1.8	DataSUS	20
2.1.9	PCDaS	21
2.1.10	Morte Materna	22
2.1.11	<i>Near Miss</i> Materno	22
2.2	TECNOLOGIAS	22
2.2.1	<i>Python</i>	22
2.2.2	<i>Pandas</i>	22
2.2.3	<i>Matplotlib</i>	23
2.2.4	<i>Jupyter Notebook</i>	23
2.2.5	<i>JupyterHub</i>	23
2.2.6	<i>Elasticsearch</i>	24
3	DESENVOLVIMENTO E ORGANIZAÇÃO	25
3.1	OBJETIVO GERAL E ESPECÍFICO DA PESQUISA	25
3.2	PROPOSTA INICIAL DE INDICADORES	26
3.3	DISCUSSÕES A RESPEITO DOS INDICADORES	26
3.4	RESOLUÇÃO FINAL DE INDICADORES	29
4	INFRAESTRUTURA E PROCEDIMENTOS	31
4.1	PRINCIPAIS FONTES DE DADOS	31
4.1.1	SIM	31

4.1.2	SINASC	32
4.2	OUTRAS FONTES DE DADOS	33
4.2.1	SIAB	33
4.2.2	Tabnet DataSUS	33
4.2.3	Tabnet ANS	34
4.2.4	MS, IBGE, CNES e SIH	34
4.3	AMBIENTE DE COLABORAÇÃO NO PCDas	35
4.3.1	<i>JupyterHub</i>	35
4.3.2	API do <i>Elasticsearch</i>	35
4.4	CÁLCULO DOS INDICADORES	35
4.4.1	Funções Comuns	36
4.4.2	Bloco 1	37
4.4.3	Bloco 2	37
4.4.4	Bloco 3	37
4.4.5	Bloco 4	37
4.4.6	Bloco 5	37
4.4.7	Ajustes do “Deslocamento ao Parto”	38
5	EXPERIMENTOS E RESULTADOS	39
5.1	EXPERIMENTOS	39
5.1.1	Gráficos iniciais feitos no <i>matplotlib</i>	39
5.1.2	Experimento de visualizações no <i>Tableau</i>	45
5.2	RESULTADOS	47
5.2.1	<i>Design</i> do painel com a <i>Odd Studio</i>	47
6	CONCLUSÃO	49
6.1	TRABALHOS FUTUROS	50
	REFERÊNCIAS	51
	APÊNDICE A – CÓDIGOS EM PYTHON E CONSULTAS SQL	55
	ANEXO A – LISTA DE INDICADORES PRELIMINARES	70
	ANEXO B – QUANTIDADE DE MUNICÍPIOS POR FAIXAS DE NASCIDOS VIVOS	73
	ANEXO C – LISTA DE INDICADORES DECIDIDOS	74

1 INTRODUÇÃO

Com o amadurecimento dos campos de conhecimento relacionados à Computação e às Tecnologias de Informação, a interdisciplinaridade se torna evidente, latente e permeável em estudos e aplicações diversificadas. A Ciência de Dados, em particular, destaca-se como uma área da Computação que exemplifica e aplica essa interdisciplinaridade de maneira profunda e abrangente.

Além disso, a saúde é um dos campos mais importantes para a aplicação e suporte da Ciência de Dados. Nesse contexto, com a digitalização dos prontuários médicos, houve um avanço significativo, pois estes, que originalmente são dados não estruturados e textuais (MURDOCH, 2013), passaram a ser analisados de maneira mais eficaz. A vasta quantidade de informações geradas nesse campo pode caracterizar-se como Big Data, permitindo uma análise aprofundada e a extração de métricas valiosas.

Dado isso, o trabalho aborda os conceitos mencionados, aplicados a um contexto de pesquisa acadêmica. Tal pesquisa foi financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)¹ e a Fundação Bill & Melinda Gates². A pesquisa faz parte de uma iniciativa da fundação: “*Grand Challenges Explorations* (GCE)”³. Essa iniciativa visa promover inovação e propostas práticas para solução de problemas nas áreas da saúde, agricultura e desenvolvimento. Também oportuniza a parceria entre pesquisadores, provendo suporte e integração entre cientistas novos e experientes.

1.1 MOTIVAÇÃO

A saúde materna é um indicador vital do desenvolvimento social de uma nação. No entanto, a mortalidade e morbidade maternas continuam como desafios significativos, cuja melhoria depende crucialmente da capacidade de monitorar, analisar e agir sobre dados de saúde de maneira eficaz e oportuna. Nesse contexto, a consolidação/agregação de dados é indispensável para fundamentar uma vigilância robusta da saúde materna.

O projeto surge da necessidade de pesquisadores da saúde obterem percepções sobre *near miss* materno no Brasil a partir do emprego de ciência de dados nas bases do SUS. A consolidação teve como premissa executar a extração e tratamento de informações, bem como o desenvolvimento de métricas, por meio da integração de tecnologias de ciência de dados e computação. Este trabalho é uma parte integrante de um projeto maior que visa aprimorar a capacidade de vigilância da saúde materna. Nesse contexto, as bases de

¹ CNPq. Disponível em <<https://www.gov.br/cnpq/pt-br>>. Acessado em: 07 de Setembro 2023.

² Fundação Bill & Melinda Gates. Disponível em <<https://www.gatesfoundation.org>>. Acessado em: 05 de Setembro 2023.

³ *Grand Challenges Exploration* (GCE). Disponível em <<https://gcgh.grandchallenges.org/challenge/grand-challenges-explorations-brazil-data-science-approaches-improve-maternal-and-child-0>>. Acessado em: 04 de Setembro 2023.

dados administrativos de sistemas nacionais, como SIM, SINASC, CNES, SIH, SIAB e SINAN, contêm informações valiosas, mas muitas vezes subutilizadas devido aos desafios na extração, integração e análise desses dados. A utilização eficaz dessas bases pode revelar padrões importantes de morbidade e mortalidade materna, permitindo intervenções mais precisas e oportunas.

A pesquisa no qual esse trabalho retrata é: “Uso de dados do mundo real e aprendizado de máquina na vigilância da mortalidade materna e da morbidade materna grave”. Dado esse cenário, é notória a oportunidade de se aplicar conhecimentos computacionais para dar suporte no desenvolvimento desse projeto. Com isso, a pesquisa foi desenvolvida em três etapas, onde uma delas (a terceira) tinha como objetivo desenvolver um painel de vigilância da saúde materna com uso de indicadores provenientes de diversos sistemas de informação disponíveis no Brasil.

Nesse contexto, a estratégia de elaboração desses indicadores envolvia a construção de um modelo teórico dos fatores de risco do óbito materno, por parte dos pesquisadores da área da saúde. Os indicadores seriam definidos a partir da avaliação do modelo em relação aos dados disponíveis nos sistemas de informação recuperados pelos pesquisadores da área da computação.

Além do processo de contraste entre modelo e experimento, que envolve os dados e seus processos de cálculo, agregação e compilação, o desenho do painel foi desenvolvido em colaboração com uma equipe especializada em projetos de visualização. Antes dessa etapa, foram realizados testes de visualização utilizando diferentes ferramentas.

Dados esses detalhes de alguns dos escopos e seções da composição do projeto, é interessante listar os desafios que foram previstos anteriormente, divididos em dois grupos: principais e operacionais.

Desafios principais:

- Bases de dados administrativos de sistemas de informações nacionais (SIM, SINASC, CNES, SIH, SIAB, SINAN) com implementação, tempo, propósito, cobertura e qualidade distintas.
- Documentação limitada e dispersa, tornando difícil o uso desses sistemas.
- Grande volume de dados.

Desafios operacionais:

- Pandemia de COVID-19.
- Rotatividade da equipe.
- Tempo prorrogado, mas sem financiamento complementar.

1.2 OBJETIVO

O objetivo principal do projeto é analisar os dados disponíveis nos sistemas de informação em saúde existentes para a vigilância da saúde materna. Para isso, os objetivos específicos incluem validar o uso das bases de dados administrativos de sistemas de informações nacionais (SIM, SINASC, CNES, SIH, SIAB, SINAN) para análise de morbidade materna, definir indicadores com base nos modelos de dados existentes e executar tratamentos e validações para a criação de uma base de dados consolidada.

Como todo trabalho é fundamentado em projetos anteriores, é mencionável que outros projetos finais – (CARVALHO; ALMEIDA, 2017)(SAMPAIO, 2019)(COSTA; PEREIRA, 2021) – trataram de problemas semelhantes. Destaca-se o projeto (COSTA; PEREIRA, 2021), que aborda a tecnologia usada em uma plataforma que facilitou a pesquisa, demonstrando o potencial das ferramentas empregadas.

Portanto, a finalidade desta dissertação é discutir a pesquisa com um olhar da área de computação, traçando relações com os trabalhos anteriores mencionados. Isso promoverá a comprovação dos conceitos e contribuirá para a atualização das práticas no campo de Ciência de Dados.

1.3 TRABALHOS RELACIONADOS

No trabalho (CARVALHO; ALMEIDA, 2017), foi explorado o carregamento de grandes volumes de dados em bancos de dados relacional e não relacional. Foram utilizadas as bases de dados do SUS: SINASC, SIM e outras auxiliares, com dados referentes ao Estado do Rio de Janeiro no ano de 2013, como base para a investigação. Foi demonstrado o contraste de desempenho entre o *PostgreSQL*⁴ e o *MongoDB*⁵ frente a consultas e geração do ambiente. O trabalho propôs a criação de uma interface onde fosse possível gerar gráficos para visualização desses dados, e as ferramentas do pacote ELK⁶, onde o *Kibana*⁷ cumpre esse propósito de visualização.

No trabalho (SAMPAIO, 2019), o sistema de informação abordado foi o SIH, com dados referentes ao Estado do Rio de Janeiro em um espaço cronológico dos anos de 2010 a 2015, realizando o carregamento apenas no *MongoDB* porém, utilizando métodos diferentes. Também foi mencionada a visualização como uma prospecção interessante, sugerindo ferramentas como *Tableau*⁸. Várias propostas de trabalhos futuros foram discutidas, dentre elas: projetos relacionados com inteligência computacional, mineração de dados, estatística utilizando a base carregada e automatização do ETL para abarcar mais bases, anos e estados brasileiros.

⁴ *PostgreSQL*. Disponível em <<https://www.postgresql.org/>>. Acessado em: 16 de Setembro 2023.

⁵ *MongoDB*. Disponível em <<https://www.mongodb.com/>>. Acessado em: 13 de Setembro 2023.

⁶ ELK. Disponível em <<https://www.elastic.co/elastic-stack/>>. Acessado em: 09 de Setembro 2023.

⁷ *Kibana*. Disponível em <<https://www.elastic.co/kibana>>. Acessado em: 02 de Setembro 2023.

⁸ *Tableau*. Disponível em <<https://www.tableau.com/>>. Acessado em: 25 de Setembro 2023.

No trabalho (COSTA; PEREIRA, 2021) foi aprimorado o conjunto de técnicas relativas a carregamento de grandes volumes de dados, ainda com enfoque em um banco de dados não relacional, somado a instrumentos de visualização, como sugerido nos trabalhos anteriores. Foram produzidas visualizações referentes aos sistemas de informação SIH, SIM e SINASC, e indicado como trabalho futuro a disponibilização desses dados a partir dessa pilha de ferramentas, através de uma API. Assim, é notável que a pesquisa utiliza exatamente uma API de um carregamento de dados semelhante ao realizado pelo PCDaS, que também utilizou a pilha de ferramentas ELK. Este trabalho se posiciona como uma evolução das metodologias apresentadas em (CARVALHO; ALMEIDA, 2017) e (SAMPAIO, 2019), incorporando melhorias técnicas e aplicacionais baseadas nas sugestões desses estudos, e se aproveitando do que foi feito como base para desenvolver os entregáveis.

No trabalho (SILVA, 2022) – um trabalho fora do escopo da computação, mas dentro do contexto da pesquisa, sendo de saúde coletiva – foi avaliada a qualidade de informação do SINASC nos anos de 2014 e 2019 apenas no município do Rio de Janeiro, e foi abordada a classificação de Robson dentro do escopo de variáveis existentes no SINASC. Tal conjunto de variáveis faz parte dos indicadores referentes à pesquisa do *near miss* materno. A qualidade da informação das variáveis é um indicador auxiliar a todos os indicadores compilados, assinalando a importância dessa preocupação com os dados oriundos dos sistemas de informações do DataSUS.

1.4 ESTRUTURA DO DOCUMENTO

No capítulo 1 introduzimos a ideia do projeto *Near Miss* e seu contexto. No capítulo 2 são descritos os conceitos e tecnologias abordados e utilizados. No capítulo 3 são detalhados o andamento e a discussão sobre os indicadores propostos a serem utilizados e os que foram decididos a serem utilizados. No capítulo 4 são explicados os códigos, bibliotecas e ferramentas utilizadas para a construção dos indicadores. No capítulo 5 são exibidos os gráficos e discutidos seus resultados. No capítulo 6 endereçamos uma conclusão e prospecção sobre ferramentas de análise e construção de *big data* a partir de dados do DataSUS.

2 ESTADO DA ARTE

Dado o objetivo deste trabalho, é importante destrinchar os conceitos que baseiam o modo de pensar e operar das tarefas realizadas. Também, é relevante descrever as ferramentas utilizadas no desenvolvimento das atividades. Logo abaixo, em dois agrupamentos – Conceitos e Tecnologias – serão detalhados os elementos envolvidos.

2.1 CONCEITOS

2.1.1 Ciência de Dados

Ciência de dados é um campo interdisciplinar focado em extrair conhecimento de grandes conjuntos de dados, consequentemente aplicando a informação obtida para resolver problemas em uma ampla gama de aplicações. Também integra domínios de conhecimento fundamentais como matemática, estatística e ciência da computação. As principais atividades na área são exploração, preparação, representação, transformação, computação, modelagem, visualização, apresentação dos dados. Colocando de forma sucinta, é a ciência de aprender com dados, no qual estuda os métodos envolvidos na análise e processamento, que propõe tecnologias para aprimorar os métodos em um modo baseado em evidências(DONOHO, 2017).

2.1.2 Transformação de Dados

O principal procedimento ao se trabalhar com dados é a sua transformação, pois é corriqueiro não estarem em um formato adequado para serem processados e examinados. Com os diferentes tipos de dados, há diversas formas de transformá-los. O intuito é reestruturá-los para serem melhor analisados. Um exemplo clássico dessa metodologia é a junção de duas bases de dados contendo uma característica em comum. O processo de transformação pode ser resumido nas seguintes etapas: descoberta, interpretação, mapeamento, codificação, a transformação em si, revisão e envio, não necessariamente nessa ordem e, sendo possível dispensar algumas etapas . Outros métodos para reestruturação dos dados são: integração, filtragem, correção, discretização, agrupamento, reconstrução e normalização (KIMBALL; CASERTA, 2004).

2.1.3 ETL

ETL é a sigla para “*Extract, Transform, and Load*”, ou seja, Extrair, Transformar e Carregar. É caracterizado como uma metodologia adotada no contexto de armazéns de dados. Esse conjunto de procedimentos vigora no cenário onde existem múltiplas fontes

dispersas no qual o objetivo é concentrá-las em um único local para análises futuras. As três etapas do processo se resumem em:

- Extrair: nesse estágio os dados são recuperados das mais diversas fontes distintas sem que os afete na origem, sendo possível automatizar essa recuperação com a finalidade de integração contínua, melhorando a qualidade do projeto ao reduzir erros de procedimento manual.
- Transformar: nessa parte tudo que foi mencionado no item anterior (transformação de dados) se aplica, porém, com o foco de adequá-los – podendo aplicar validações e regras – para serem guardados no sistema centralizado.
- Carregar: nessa etapa os dados são carregados no sistema, podendo ser feito antes ou depois das transformações. A dificuldade neste ponto depende do ambiente do sistema. Assim, as características: volume, estrutura, complexidade, e frequência determinam a natureza dessa dificuldade (KIMBALL; CASERTA, 2004).

2.1.4 Visualização de Dados

O principal objetivo de visualizar dados e estatísticas é obter informações adicionais. No contexto visual é necessário exibir de forma gráfica os dados envolvidos e o melhor gráfico a ser escolhido é aquele que proporciona uma maior compreensão do cenário. Existem vários modelos de gráficos, e é necessário saber avaliar quais gráficos irão fornecer um tipo de ilustração que melhor se adequa ao contexto da investigação. Em análise exploratória dos dados é essencial realizar uma avaliação da qualidade, e para isto, é necessário corrigir erros, encontrar valores discrepantes, identificar tendências, localizar padrões, avaliar modelos e apresentar resultados. Na visualização de dados é essencial conhecer o contexto, a origem dos dados, como foram coletados, e como os especialistas irão interpretá-los. É importante destacar que as visualizações e gráficos em si não são o suficiente, ou seja, são complementadas e complementam o conteúdo da pesquisa (UNWIN, 2020).

2.1.5 API

Uma *Application Programming Interface* (API), ou – em tradução livre – Interface de programação de aplicação, está presente na maioria dos programas modernos. A sua finalidade é fornecer uma funcionalidade como se fosse uma “caixa opaca”, ou seja, escondendo detalhes de implementação. Geralmente, é especificado como deve ser a interação entre as partes do programa para resolver o problema que ela foi designada. É importante destacar que o papel de uma API é ser uma interface bem definida para servir a outras partes de um sistema. Normalmente, é distribuída como uma biblioteca, sendo capaz de ser – por exemplo – um recurso externo na internet, permitindo ser usada em diversas aplicações. Em essência, são partes reutilizáveis que comportam vários componentes

funcionais a serem integrados em um programa que um usuário irá utilizar (REDDY, 2011).

2.1.6 SQL

SQL – *Structured Query Language* (Linguagem de Consulta Estruturada, em tradução livre) – é uma linguagem declarativa e relacional com a proposta de ser usada para manipulação e definição de dados em banco de dados. Em essência, as consultas descrevem a informação a ser definida/encontrada, ao invés de ter uma sequência de passos de como efetuar-la, por isso é uma linguagem declarativa. Possui diversas funcionalidades tais como diferentes tipos de junções, apelidos para variáveis, funções e tipos customizáveis, ações passíveis de serem engatilhadas, dentre outras (CHAMBERLIN, 2012).

2.1.7 Banco de Dados Não Relacional (NoSQL)

Sistemas gerenciadores de bancos de dados são aplicações comuns no contexto de manipulação de dados. Bancos de dados NoSQL dão suporte a gerenciamento de dados em grandes volumes de armazenamento distribuídos por diversos servidores, e em alguns é possível realizar consultas através da linguagem SQL. Os primeiros bancos de dados NoSQL não possuíam tal funcionalidade, dessa forma o “NoSQL” significava diretamente “Não SQL”. Uma redefinição possível é que a sigla seja “*Not Only SQL*” (Não apenas SQL, em tradução livre). As motivações para existências desse tipo de bancos de dados são: implantação rápida, dados em larga escala, escalabilidade, tolerância a falhas e armazenamento intermediário. As possíveis categorias de formato de armazenamento são: chave-valor, orientado a registro, orientado a coluna, orientado a documentos (geralmente guardados como JSON) e orientado a grafos. São características comuns encontradas nesse estilo de banco de dados: fragmentação, elasticidade, réplicas assíncronas, arquiteturas que não compartilham entre si, dentre outras (CHANDRA, 2015).

2.1.8 DataSUS

O DataSUS desempenha um papel central no sistema de informações de saúde pública no Brasil, sendo responsável por consolidar, padronizar e disponibilizar dados fundamentais para o planejamento e gestão de políticas públicas de saúde. Criado em 1991, o departamento surgiu com a missão de fornecer suporte tecnológico e informatizado aos órgãos do Sistema Único de Saúde (SUS), facilitando o acesso a bases de dados essenciais. Suas competências incluem a integração de sistemas, a manutenção de um acervo abrangente de informações e a garantia de acesso a essas ferramentas por gestores e pesquisadores. A seguir, apresenta-se uma citação que detalha a origem e as principais atribuições do DataSUS, destacando sua relevância para a integração e análise de dados no contexto brasileiro. A seguir encontra-se um trecho retirado direto do site oficial do DataSUS.

O Departamento de Informática do Sistema Único de Saúde (DATASUS) surgiu em 1991 com a criação da Fundação Nacional de Saúde (Funasa), pelo Decreto 100 de 16.04.1991, publicado no D.O.U. de 17.04.1991 e retificado conforme publicado no D.O.U. de 19.04.1991. [...] Foi então formalizada a criação e as competências do DATASUS, que tem como responsabilidade prover os órgãos do SUS de sistemas de informação e suporte de informática, necessários ao processo de planejamento, operação e controle. (SUS - Sistema Único de Saúde, 2023)

São estas as competências definidas para o DATASUS pelo Decreto:[...] IV. definir padrões para a captação e transferência de informações em saúde, visando à integração operacional das bases de dados e dos sistemas desenvolvidos e implantados no âmbito do SUS; V. manter o acervo das bases de dados necessárias ao sistema de informações em saúde e aos sistemas internos de gestão institucional; VI. assegurar aos gestores do SUS e órgãos congêneres o acesso aos serviços de informática e bases de dados, mantidos pelo Ministério; (SUS - Sistema Único de Saúde, 2023)

Essas competências garantem que diversos pesquisadores sejam capazes de acessar dados gerados pelo SUS sobre a população brasileira. Dessa forma, é reforçada a necessidade da criação do departamento e sua necessidade em tempos cujo a digitalização é, progressivamente, presente em nosso cotidiano.

2.1.9 PCDaS

A Plataforma de Ciência de Dados Aplicada à Saúde (PCDaS) surge como uma ferramenta inovadora para o fortalecimento da gestão e análise de dados em saúde pública no Brasil. Desenvolvida pela Fundação Oswaldo Cruz (Fiocruz) em parceria com o Laboratório Nacional de Computação Científica (LNCC), a PCDaS combina tecnologia de ponta e infraestrutura acessível para oferecer uma solução integrada para a gestão de grandes volumes de dados. Desde seu lançamento em 2016, a plataforma tem sido fundamental para pesquisadores, gestores e educadores, promovendo o uso de ferramentas open source para o armazenamento, processamento e disseminação de informações relacionadas à saúde e seus determinantes sociais. A seguir, apresenta-se uma citação que detalha os objetivos e a relevância da PCDaS no contexto da ciência de dados em saúde, de acordo com o portal oficial.

A Plataforma de Ciência de Dados aplicada à Saúde (PCDaS) é um projeto de pesquisa e desenvolvimento tecnológico do Laboratório de Informação em Saúde (LIS) do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT) da Fundação Oswaldo Cruz (FIOCRUZ) do Ministério da Saúde (MS), em parceria com o Laboratório Nacional de Computação Científica (LNCC) do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC).

Lançada em 2016, tem como objetivo principal desenvolver e disponibilizar Plataforma de Ciência de Dados aplicada à Saúde (conceito de *PaaS – Platform as a Service*) pública e gratuita com utilização de ferramentas open source para armazenamento, gestão, análise e disseminação de grandes quantidades de dados de saúde e seus determinantes socioambientais para pesquisadores, docentes e discentes de instituições de ensino e pesquisa, bem como gestores governamentais. (BIONETWORKS, 2021)

2.1.10 Morte Materna

Dentro da literatura da área da saúde, a morte materna é definida como o caso em que mulher que estava grávida, parindo ou até 42 dias após o parto veio à óbito. Só é classificado dessa maneira se a circunstância está relacionada com a gravidez, ou seja, motivos acidentais ou externos ao contexto da gravidez não são considerados. (NATIONS, 2017)

2.1.11 *Near Miss* Materno

Near miss pode ser entendido e traduzido como um acidente que quase aconteceu. Já a OMS define o *Near Miss* Materno como o caso de uma mulher que sobreviveu a condições e complicações onde o desfecho poderia ter sido uma morte materna. E esse conceito é definido dado os aspectos em comum entre as que sobreviveram e faleceram nesse contexto. Em outras palavras, o *near miss* materno é o cenário de sobrevivência de uma mulher que passa por condições de ameaça de sua existência e do bebê. Para as que sobreviveram e tiveram agravos na saúde, ou surgiram doenças, são caracterizadas como morbidade materna grave.

2.2 TECNOLOGIAS

2.2.1 *Python*

Python é uma poderosa linguagem de programação com diversas características que sustentam essa qualidade. A programação pode ser feita de forma procedural, orientada a objetos ou funcional. É uma linguagem de uso geral, de alto nível, interativa e interpretada com uma sintaxe simples e uma vasta gama de bibliotecas padronizadas que abarcam diversas áreas, desde processamento de texto, interfaces com o sistema operacional e protocolos de internet, até aplicações científicas (Python Software Foundation, 2023). Nesse sentido, dada a sua maturidade e a quantidade de bibliotecas disponíveis – uma vez que a linguagem foi criada em 1991 – cada vez mais pesquisadores utilizam bibliotecas numéricas e de tratamento de dados como *NumPy*, *SciPy*, *Pandas* e outras (MCKINNEY, 2011).

2.2.2 *Pandas*

Pandas é uma biblioteca com o intuito de ser uma camada fundamental para estatística em *Python*. Desenvolvida desde 2008, também pretende conectar as ferramentas de análises de dados feitas em *Python* a diversas plataformas de estatística e banco de dados. O nome da biblioteca vem de **panel data** (painel de dados, em tradução livre), um termo comum em estatística e econometria. Atualmente, é um bom complemento para ferramentas científicas e contém outros tipos de manipulação de dados encontrados no R,

por exemplo. Possui estruturas de dados bem desenvolvidas, funções claras e adaptadas para manipulação em vários conjuntos de dados comuns em áreas como finanças, ciências sociais, estatística e outras. Algumas características e funcionalidades relevantes são: acesso de dados baseado em etiquetas, manipulação de dados faltantes, integração com Matplotlib, remodelagem avançada, agrupamento e agregação de dados, tabelas com fácil estilo de planilha e desempenho em grandes conjuntos de dados (MCKINNEY, 2011).

2.2.3 *Matplotlib*

Matplotlib é uma biblioteca escrita em *Python* projetada com o pensamento de ser possível criar gráficos – em 2D a partir de listas – com poucos comandos. Criada a partir do contexto do MATLAB, provavelmente a origem da inspiração do nome, foi projetada a partir dos seguintes requisitos: codificação fácil para extensão, incluível em interfaces gráficas de aplicações, qualidade estética dos gráficos e facilidade na criação dos mesmos. Apesar de a biblioteca ser feita só em *Python*, se usa bastante o *NumPy* e outras ferramentas para melhorar o desempenho em grandes conjuntos de dados (Matplotlib development team, 2023).

2.2.4 *Jupyter Notebook*

Cadernos (*notebooks*) computacionais podem ser resumidos como laboratórios para computação científica. É um formato interessante onde é fácil executar, modificar e visualizar o resultado de códigos de forma iterativa. Dessa forma, é possível gerar uma narrativa computacional, onde gráficos, imagens, textos, códigos e seus resultados se complementam nesse tipo de documento. Outra facilidade provida por esse instrumento é o acesso remoto a dados que podem ser difíceis de baixar. O *Jupyter Notebook* é uma ferramenta gratuita, de código aberto e que pode ser utilizada em uma página *web*. Possui dois componentes principais: uma interface acessível por um navegador, onde se pode colocar os códigos e outros elementos, e um núcleo (*kernel*) onde o código é executado. Cadernos computacionais já existiam desde a década de 80, porém o *Jupyter* conseguiu uma popularidade elevada, se tornando um padrão no mercado. O nome da ferramenta é inspirado em nomes de linguagens de programação muito utilizadas no campo da ciência de dados, Julia (Ju), *Python* (Pyt) e R (PERKEL, 2018).

2.2.5 *JupyterHub*

O *JupyterHub* possibilita que diversos usuários usem um *Jupyter Notebook*. Esses usuários podem ser estudantes ou pesquisadores que trabalham em uma mesma equipe e precisam que seus trabalhos sejam compartilhados em um mesmo ambiente (PERKEL, 2018). Neste cenário, onde os recursos são compartilhados e vários grupos de usuários podem acessá-los, a plataforma permite que os acessos sejam administrados pela gestão da

pesquisa. Permite o acesso a um ambiente computacional ao mesmo tempo em que libera o usuário de tarefas de instalação e manutenção. A plataforma pode ser executada na nuvem ou em máquinas próprias da instituição, sendo possível um ambiente já configurado para qualquer usuário com acesso a internet no mundo. Algumas características chaves são: ser customizável, flexível, escalável e portátil (Project Jupyter, 2023).

2.2.6 *Elasticsearch*

Elasticsearch é um motor de busca que possui as funcionalidades de ser distribuído para vários usuários, com capacidade de busca por texto completo. Possui uma interface *web* e documentos JSON sem esquema/modelo. A sua principal potencialidade está na pesquisa e análise de tempo real para vários tipos de dados, pois, na medida que o volume dos dados e das pesquisas crescem, a ferramenta permite que a base de dados e os recursos computacionais possam crescer na mesma medida. Os dados alimentados podem ser estruturados ou não, onde serão armazenados e indexados de forma eficiente, permitindo buscas rápidas para descoberta de tendências e padrões (Elasticsearch team, 2023c).

3 DESENVOLVIMENTO E ORGANIZAÇÃO

A proposta e motivação desse projeto está associada à Fundação Bill & Melinda Gates e seu programa GCE, que é aplicado em diversos países do mundo. Dentro do escopo da chamada do programa “Ciência de Dados Para Melhorar a Saúde Materno Infantil, Saúde da Mulher e Saúde da Criança no Brasil”, foram pleiteados 12 projetos no qual estavam abarcados em 4 grupos: saúde materno-infantil, saúde da mulher, nutrição materno-infantil e saúde da criança. O projeto em questão, que esse trabalho trata – “Uso de dados do mundo real e aprendizado de máquina na vigilância da mortalidade materna e da morbidade materna grave” – se encontrava no grupo saúde materno-infantil. Neste capítulo será discutido o quadro da pesquisa e como se decorreu a evolução dos indicadores propostos, que serão apresentados a seguir.

3.1 OBJETIVO GERAL E ESPECÍFICO DA PESQUISA

A pesquisa foi em torno da exploração de dados do Sistema de Informações do SUS a fim de examinar a morbidade e mortalidade materna, dados os critérios da OMS, além de analisar indicadores provenientes desses sistemas de informação, aliados a outros dados públicos, para ampliação da vigilância dos gestores de saúde para prevenção e mitigação dos óbitos maternos.

Neste sentido, o objetivo geral era analisar os sistemas de informação existentes para a vigilância de saúde materna. Os objetivos específicos se dividiram na validação do uso do Sistema de Informações Hospitalares do SUS (SIH/SUS) para análise da morbidade materna, na definição de indicadores baseados em um modelo teórico de dados existentes em sistemas de informação brasileiros e na criação de um painel de visualização de informações para vigilância de saúde materna.

O projeto original foi dividido em três etapas para dar base aos objetivos. A primeira etapa envolveu o desenvolvimento de um algoritmo preditivo para identificar morbidade materna a partir do SIH usando análise exploratória dos dados e aprendizado de máquina. A segunda etapa se debruçou sobre a validação da identificação de *near miss* materno a partir do SIH, levando em consideração os dados e o estudo “Nascer no Brasil”¹. Por fim, na terceira etapa – no qual esse trabalho se baseia – o objetivo foi a criação de um painel com indicadores demográficos, socioeconômicos, de morbidade, de recursos e de cobertura de serviços de diversos sistemas de informação para vigilância de saúde materna onde seria aplicado o ETL nesses sistemas para a construção desses indicadores.

¹ Nascer do Brasil. Disponível em <<https://nascernobrasil.ensp.fiocruz.br/>>. Acessado em: 26 de Agosto 2023.

A conveniência da criação deste painel se deu pela inovação da maneira com que os dados serão visualizados por gestores de saúde, sendo possível comparar indicadores – de forma rápida em uma localidade – com metas, municípios e UFs diferentes. Os indicadores que indicam a possibilidade de óbito materno poderão identificar regiões vulneráveis, apesar de serem locais com baixa ou nenhuma ocorrência de óbitos. Possíveis utilizadores do painel vão desde pesquisadores até pessoas que acessam o SUS. O objetivo do painel é fornecer suporte à instrução de ações de promoção e atenção à saúde materna, com capacidade de impactar a saúde infantil.

3.2 PROPOSTA INICIAL DE INDICADORES

O estudo do modelo teórico feito pelas pesquisadoras da área da saúde foi iterado em diversas reuniões sobre o assunto. Inicialmente foi proposto um conjunto de indicadores iniciais que seriam avaliados durante a pesquisa. Esses indicadores versam sobre os fatores de risco para o óbito materno. Nessa proposta existiam 5 grupos de indicadores que estavam sendo avaliados:

1. Características demográficas e obstétricas
2. Acesso à contracepção/gestações não intencionais
3. Assistência pré-natal
4. Assistência ao parto
5. Morbidade materna

Além desses grupos existiam outros indicadores em discussão que não tinham um grupo definido. Após diversas análises, foi decidido quais indicadores seriam coletados e calculados para serem averiguados. Nessa tabela constavam o nome do indicador, a forma de cálculo, a fonte dos dados, as variáveis envolvidas na base de dados, a descrição, observações, dúvidas e responsabilidades. Na tabela 1 em anexo 6.1 é apresentado um recorte desse quadro inicial, totalizando 34 indicadores.

3.3 DISCUSSÕES A RESPEITO DOS INDICADORES

A construção e apuração dos indicadores ocorreram em um processo de análise exploratória dos dados. Foi averiguado que a maior parte das fontes dos indicadores poderiam ser obtidas a partir das bases de dados enriquecidas do PCDaS. Outras fontes incluíam TABNET, SIAB, SINAN, ANS, IBGE, CNES, etc.

O indicador “cobertura de equipes de saúde da família (ESF)”, foi avaliado pelas pesquisadoras da área da saúde, onde foi analisada a diferença entre a cobertura segundo a

atenção básica e segundo as ESF, em que estimou-se uma considerável semelhança entre as duas. Essa análise serviu para concluir a incorporação do indicador cobertura de ESF. Foi decidido usar a média anual a partir da análise do coeficiente de variação, desvio padrão dividido pela média.

No indicador “taxa de fecundidade de mulheres de 35 anos ou mais”, onde se utiliza dados do TABNET, foi verificada uma elevação no período analisado, inicialmente indicando problemas nos cálculos. A resolução dos casos de indicadores com problemas será apresentada ao fim desta seção.

No indicador “proporção de óbitos maternos investigados” houve um equívoco ao utilizar uma variável que não correspondia exatamente ao filtro de idade de mulheres de 10 a 49 anos. A variável mais adequada no SIM seria “IDADE”, ou “idade_obito_anos” enriquecida pelo PCDaS. Em reunião com a PCDaS foi informado que a variável “IDADEMAE” só é preenchida nos casos de óbito fetal e neonatal.

No indicador "proporção de mulheres com número adequado de consultas", os dados do ano de 2013 vieram zerados, identificando um problema na base de dados. Também foi percebido esse problema nas variáveis relacionadas aos indicadores de número de consulta pré-natal e grupo de Robson. O PCDaS confirmou ser um problema na base de dados do DataSUS, ao invés de ser um problema no processo de ETL.

Na exploração dos dados de indicadores cuja proporção é crítica pelo número baixo de ocorrências – por exemplo, o indicador de razão de mortalidade materna – foi levantada a questão de municípios onde o número de nascimentos é extremamente baixo ou ausente, levantando a necessidade de ser informado no painel que os dados devem ser analisados com cuidado. Nesse sentido, foi avaliada a distribuição de nascidos vivos por município. Na tabela 2 em anexo 6.1 esses dados são mostrados por faixa de valores, onde 82.4% dos municípios possuíam menos de 500 nascidos vivos.

Foi explorada a questão de como certas variáveis categóricas seriam apresentadas no painel, se apenas as categorias de risco ou todas, no final optou-se por apresentar todas, por ex., variáveis como cor de pele e escolaridade. E nesse contexto, a qualidade da informação também foi discutida, uma vez que variáveis associadas à quantidade de partos era notório ser de baixo preenchimento. Variável em relação à gestação gemelar foi excluída da lista de possíveis indicadores por conta da baixa frequência e estar associada a regiões mais desenvolvidas.

Os indicadores foram analisados em relação ao município do Rio de Janeiro, onde as pesquisadoras já possuíam domínio e entendimento do comportamento desses dados. diversos indicadores foram percebidos problemas tais como: razão de mortalidade materna (RMM) – total e direta –, proporção de óbitos de mulheres em idade fértil, óbitos maternos investigados, taxa de mortalidade (neonatal e fetal) e taxa de fecundidade. Listando a deliberação desses temos:

- O indicador de taxa de mortalidade neonatal precoce foi decidido pela não inclusão

porque o Ministério da Saúde não havia disponibilizado os fatores de correção.

- Para o indicador de taxa de mortalidade fetal seria necessário usar a base de dados do “SIM-DOFET” que não estava disponível no PCDaS.
- Os indicadores de taxa de fecundidade foram averiguados de diversas formas – pois estavam discrepantes em relação a Rede Interagencial de Informações para a Saúde (RIPSA) – até chegarem à conclusão de mantê-los apenas como taxa de fecundidade específica em mulheres menores de 20 anos.
- Para os indicadores “proporção de óbitos de mulheres em idade fértil” e “óbitos maternos investigados” a fonte de informação foi alterada para o TABNET, pois nem todos os locais atualizam essa informação no SIM.
- E o cálculo dos indicadores de RMM foram repassados para a equipe do Observatório Obstétrico², pois tiveram extensas análises por conta de inconsistências calculadas.

Existiram diversas questões no decorrer do desenvolvimento dos indicadores que envolviam:

- A definição do período em anos - 2012 a 2019, 2014 a 2019 ou 2012 a 2020?
- A menor unidade de análise - município ou hospital.
- Tipos de análises que seriam permitidas.
- Dados os tipos de análise como seriam geridos esses dados em termos de consulta e armazenamento.
- Precaução em não copiar análises já existentes.

Em resumo, alguns indicadores foram retirados da lista, outros mudaram a fonte de dados e um grupo de indicadores adicionais já estavam calculados ao serem recuperados, bastando colocá-los no formato que foi formalizado. Na maior parte do tempo deixávamos uma coluna com o indicador já calculado ao invés de apenas as colunas referentes às variáveis utilizadas para o cálculo. Mas uma vez que seriam utilizadas apenas as colunas das variáveis pela equipe do Observatório Obstétrico no ambiente em que desenvolveram o painel, foi decidido não colocar as colunas com os cálculos dos indicadores. Por fim, diversas correções pontuais na formatação da entrega dos indicadores – incluindo a documentação do que foi produzido – foram feitas e, ao final do projeto, foi dado o alinhamento com a equipe do Observatório Obstétrico.

² Observatório Obstétrico Brasileiro. Disponível em <<https://observatorioobstetricobr.org>>. Acessado em: 05 de Setembro 2023.

3.4 RESOLUÇÃO FINAL DE INDICADORES

A construção do modelo teórico dos determinantes da mortalidade materna ficou organizado em 7 campos, onde se avaliam diversas particularidades. Esses campos e suas especificidades são:

- Condições socioeconômicas e de acesso a serviços de saúde: IDH-M, escolaridade materna, cor da pele, cobertura SUS, cobertura ESF.
- Planejamento reprodutivo/prevenção de gestações indesejadas e de alto risco: Taxa de fertilidade em mulheres <20 anos, proporção de grandes múltiparas, taxa de aborto inseguro/1000 mulheres em idade fértil, taxa de aborto inseguro por 100 nascidos vivos.
- Assistência pré-natal: Cobertura de assistência pré-natal, proporção de gestantes com início do PN até 12 semanas de gestação, proporção de gestantes com mais de 7 consultas, taxa de incidência de sífilis congênita.
- Assistência ao parto: Deslocamento para assistência ao parto, proporção de cesarianas, proporção de mulheres segundo grupo de Robson, contribuição proporcional de cada grupo de Robson para a taxa de cesariana.
- Condições de nascimento: Proporção de nascimentos prematuros, promoção de nascimentos termo precoce, proporção de baixo peso ao nascer, incidência de morbidade materna grave.
- Desfecho Mortalidade materna: Razão de Mortalidade Materna, Proporção de óbitos maternos diretos.
- Qualidade dos sistemas de informação: Cobertura SINASC, Cobertura SIM, investigação de óbitos de Mulheres em idade fértil e de óbitos maternos, completude de variáveis do SIM e do SINASC.

No encerramento da pesquisa foram documentados em uma planilha os indicadores separados em seus respectivos blocos. Apenas o último bloco teve a nomenclatura um pouco modificada, “Mortalidade Materna e Morbidade Materna Grave”, considerando também a morbidade além da mortalidade.

Essa planilha contém o responsável pelo indicador, nome do indicador, fonte dos dados, descrição, nome da base de dados, a forma de cálculo a partir das colunas dos dados compilados, cobertura do sistema, incompletude de informação, nome da base de dados da incompletude, descrição do que é incompleto, a forma de cálculo da incompletude. É possível notar a existência de mais colunas do que na tabela anterior, pelo motivo da

existência do grupo de qualidade dos sistemas de informação, que incide sobre cada indicador. Da coluna “cobertura do sistema” até a coluna “forma de cálculo da incompletude” contém informações relacionadas à qualidade.

Na tabela 3 em anexo 6.1 é apresentado um recorte dessa planilha, totalizando 43 indicadores. É importante notar que o ano de 2020 passa a ser incluído, uma vez que no PCDaS, os dados – oriundos do SIM e do SINASC – são disponibilizados no final do projeto para serem consumidos.

Dentro do grupo 4, existem indicadores relacionados ao deslocamento do parto, mas não foram colocados na tabela 3 pois são apresentados de uma maneira diferente no painel. São indicadores cuja fonte de dados são o SINASC e o CNES, e foram compilados por uma pesquisadora especialista do IBGE.

4 INFRAESTRUTURA E PROCEDIMENTOS

A compilação dos dados para a constituição dos indicadores ocorreu utilizando diversas ferramentas de desenvolvimento em ciência de dados. Foi utilizado um ambiente de desenvolvimento disponibilizado pelo PCDaS. Esse ambiente foi através do *JupyterHub* para acessar os dados tratados que vieram do DataSUS e carregados no *Elasticsearch* da plataforma (PCDaS, 2023b). Nesse ambiente foram codificados em cadernos *Jupyter* as consultas ao *Elasticsearch*, a análise exploratória dos dados utilizando o *Pandas* e o *Matplotlib*, através da linguagem de programação *Python*.

Com a parceria com o PCDaS durou alguns meses após o início da pesquisa, permitindo efetuar análises exploratórias e estudos da base de dados, a partir do trabalho apresentado em (COSTA; PEREIRA, 2021). Essas análises envolviam mais estudos da etapa 1, onde parte desse estudo embasou alguns indicadores, como os da morbidade materna, por exemplo.

4.1 PRINCIPAIS FONTES DE DADOS

Como pode ser notado nas tabelas 1 e 3, a maior parte dos indicadores são derivados do SINASC e do SIM, sendo a grande maioria deles provenientes do SINASC. Assim se faz necessário explicar cada um desses sistemas de informação e mencionar as variáveis envolvidas. Essas duas principais fontes de dados foram acessadas a partir da base de dados no Elasticsearch do PCDaS, cujo os dados foram tratados e enriquecidos a partir do dicionário de dados (SAÚDE, 2012a)(SAÚDE, 2012b). Foram criadas novas colunas, assim facilitando a análise dos dados (PCDaS, 2023b).

4.1.1 SIM

O Sistema de Informação sobre Mortalidade (SIM) começou a ser desenvolvido em 1975 e passou a ser informatizado em 1979. Os dados de mortalidade são coletados a partir da declaração de óbito (DO), documento que deve ser emitido por um médico. O Ministério da Saúde incentiva o uso desses dados para criação de indicadores epidemiológicos para traçar estratégias e elaborar planos para reduzir mortes evitáveis (SUS, 2023c).

Essa base de dados possui diversos campos, ou colunas, referentes a dados de mortalidade padronizados pela DO. As variáveis que foram utilizadas apenas na análise exploratória dos dados são:

- Código do município de residência (CODMUNRES).
- Causa básica do óbito, classificados como CID-10 (CAUSABAS).

- Óbito na gravidez (OBITOGRAV).
- Óbito no puerpério (OBITOPUERP).
- Tipo do óbito, classificando entre fetal e não fetal (TIPOBITO).
- Óbito investigado (TPPOS). (SAÚDE, 2012a)

4.1.2 SINASC

O Sistema de Informações sobre Nascidos Vivos (SINASC) foi inaugurado em 1990 com o objetivo de coletar e fornecer dados de nascimentos em todo território brasileiro e em todos os níveis do SUS. Os dados do nascimento são coletados a partir de uma declaração de nascido vivo (DN) e podem ser preenchidos por profissionais de saúde ou parteiras tradicionais. Esses dados podem ser utilizados para estudos epidemiológicos e o Ministério da Saúde também incentiva que os gestores utilizem esses dados para o monitoramento do cenário de saúde daquela população (SUS, 2023d).

Como a grande maioria dos indicadores calculados pela equipe inicial estão associadas a essa base de dados, é importante listar as variáveis utilizadas:

- Código do município de residência (CODMUNRES).
- Idade da mãe (IDADEMAE).
- Tipo de raça/cor da mãe (RACACORMAE).
- Quantidade de partos vaginais (QTDPARTNOR).
- Quantidade de partos cesáreos (QTDPARTCES).
- Escolaridade da mãe em anos de estudo concluídos (ESCMAE).
- Quantidade de consultas pré-natal (CONSPRENAT).
- Mês da gestação que iniciou as consultas de pré-natal (MESPRENAT).
- Tipo de parto (PARTO).
- Código do grupo de Robson (TPROBSON).
- Peso ao nascer em gramas (PESO).
- Duração da gestação em faixas de semanas (GESTACAO).
- Quantidade de semanas de gestação (SEMAGESTAC). (SAÚDE, 2012b)

4.2 OUTRAS FONTES DE DADOS

Para alguns indicadores, os dados procederam de fontes diversas. Nessas fontes, os dados estavam ou organizados em tabelas no formato “.xlsx”, onde bastava passar para o formato convencionado, ou eram tabelas que podiam ser acessadas via *web*, cujo tratamento e organização dos dados foi realizado de forma automatizada através de procedimentos em *Python*.

4.2.1 SIAB

O Sistema de Informação da Atenção Básica (SIAB) tem como objetivo acompanhar os trabalhos e resultados das equipes do programa de saúde da família. É um instrumento gerencial que complementa e incorpora a reorganização do SUS, além de promover avanços significativos em informação em saúde. Nesse sistema é possível obter dados sobre cadastro das famílias, condição de moradia, situação sanitária e constituição das equipes de saúde (DATASUS, 2023b).

O indicador que utilizou essa fonte foi o “cobertura populacional com equipes de saúde da família”. O site “e-Gestor AB”¹ foi o local onde esses dados foram obtidos. As variáveis envolvidas são:

- Quantidade de população atendida por equipes de saúde da família (QT_COBERTURA_SF).
- Quantidade da população total estimada (QT_POPULACAO).

4.2.2 Tabnet DataSUS

O Tabnet DataSUS é um domínio onde o DataSUS disponibiliza informações de diferentes bases de dados de maneira tabular, nele é possível escolher a forma como a informação será disponibilizada, com opção de ser exportada nos formatos csv e Tabwin. Os dados abrangem informações demográficas, epidemiológicas, assistenciais, mórbidas e vitais. (DATASUS, 2023a)

O site do “Tabnet DataSUS”² é a origem dos dados extraídos dos procedimentos em *Python*. A informação buscada era associada à estimativa populacional das faixas de 10 a 19 e de 10 a 49 anos. Os indicadores relacionados a essa fonte de dados são:

- Proporção mulheres 10 a 49 anos usuárias exclusivas do SUS
- Taxa de fertilidade específica em mulheres menores de 20 anos

¹ e-Gestor AB. Disponível em <<https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCobertura>>. Acessado em: 16 de Setembro 2023.

² TabNet Win32 3.2: População Residente - Estudo de Estimativas Populacionais por Município, Idade e Sexo 2000-2021 - Brasil. Disponível em <<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popsvsbr.def>> . Acessado em: 17 de Setembro 2023.

- Taxa de abortos inseguros por 100 nascidos vivos
- Taxa de abortos inseguros por 1000 mulheres em idade fértil

4.2.3 Tabnet ANS

O Tabnet ANS é o portal onde a Agência Nacional de Saúde Suplementar (ANS) disponibiliza informações de planos de saúde, beneficiários e provedoras dos mesmos. Lá é utilizado o mesmo sistema do Tabnet DataSUS para disponibilizar os dados da mesma maneira. (SUS, 2023a)

O indicador que utiliza essa base de dados é “proporção de mulheres de 10 a 49 anos usuárias exclusivas do SUS”. Isso pois, basta subtrair a quantidade de beneficiários da população total para saber qual a parcela que são usuárias do SUS.

4.2.4 MS, IBGE, CNES e SIH

No caso das fontes de dados mencionadas a seguir, os indicadores ou foram calculados por outros membros da equipe, ou provieram de instituições. Nesses casos, foram feitos ajustes de formatação dos dados, para atender o formato convencionado. Cada um dos ajustes implementados está listado a seguir:

- O Indicador de sífilis congênita³ já é organizado e calculado pelo MS através do Sistema de Informação de Agravos de Notificação (SINAN)(SUS, 2023e), com isso, foi necessário apenas a recuperação desses dados e então passá-los para o formato convencionado.
- No bloco de indicadores socioeconômicos foi associado o IDH do município (IDH-M) para enriquecer a análise do cenário, importante destacar que a origem dessa informação é o censo do IBGE do ano de 2010 ⁴.
- O Cadastro Nacional de Estabelecimentos de Saúde (CNES) (SUS, 2023b) visa possuir o cadastro de todos os estabelecimentos de saúde brasileiros – sejam públicos, conveniados ou privados. Este foi utilizado para indicadores do bloco “assistência ao parto” relacionados ao deslocamento para o parto.
- Por fim, o Sistema de Informações Hospitalares (SIH) (DATASUS, 2023c) baseia os indicadores relacionados a taxa de aborto e morbidade materna, sendo esse último desenvolvido na etapa 1 quais seriam as variáveis determinantes para cada categoria de morbidade.

³ DVIAHV | Indicadores Sífilis. Disponível em <<http://indicadoressifilis.aids.gov.br/>>. Acessado em: 19 de Setembro 2023.

⁴ IDHM IBGE 2010. Disponível em <<https://www.undp.org/pt/brazil/idhm-munic%C3%ADpios-2010>>. Acessado em: 18 de Setembro 2023.

4.3 AMBIENTE DE COLABORAÇÃO NO PCDaS

O PCDaS providenciou um ambiente moderno e robusto de colaboração através do *JupyterHub* configurado por eles, além de treinamentos em ciência de dados, mineração, análise e visualização de dados. (PCDaS, 2023a) Foram realizadas reuniões mensais para acompanhamento, levantamento de questões e problemas em relação ao projeto e o ambiente disponibilizado.

4.3.1 *JupyterHub*

Após algumas reuniões iniciais de alinhamento com a equipe do PCDaS, foi fornecido para os membros da equipe login e senha para acesso ao JupyterHub. Além disso, foi dada uma introdução às ferramentas através de um *Jupyter Notebook*.⁵

Para esse acesso foi disponibilizado um ambiente de hospedagem do *JupyterHub*, em que as configurações da máquina eram: processador “Intel(R) Xeon(R) CPU E5-2630 v3”, 120 GB de memória RAM. A versão do kernel do Sistema Operacional Linux era 5.4.0 e a distribuição “Ubuntu 20.04.5 LTS (Focal Fossa)”. Este ambiente era compartilhado com outros projetos e parcerias do PCDaS, apesar disso, não foram notados gargalos ou sobrecargas do sistema que pudessem atrasar o desenvolvimento do projeto.

4.3.2 API do *Elasticsearch*

A interação com a API disponibilizada pelo PCDaS para a equipe foi pouco eficiente dada a limitação de dados que são fornecidos em cada pedido da consulta, por isso, foi escolhido o método de consultas utilizando a linguagem SQL (Elasticsearch team, 2023a). A API do *Elasticsearch* é bem documentada, por isso, foi possível contornar o problema mencionado ao realizar um novo pedido da mesma consulta para retornar mais dados (Elasticsearch team, 2023b) e assim obter todos os dados daquela consulta.

Além da distribuição de acesso aos membros da equipe que iriam se debruçar sobre o trabalho da etapa 3, também foi fornecido um token para autorizar as consultas na API. Esse token deve ser repassado em todos os pedidos de consulta, incluindo aquelas onde se solicita mais dados.

4.4 CÁLCULO DOS INDICADORES

Inicialmente, cada indicador era elaborado em um arquivo *Jupyter Notebook* diferente, com o desdobramento do projeto, foi melhorada a forma com que os códigos eram organizados. Na entrega, a organização do trabalho estava como um arquivo *Jupyter Notebook* para cada bloco de indicador e um módulo em *Python*.

⁵ JupyterHub PCDaS. Disponível em <<https://jupyterhub.iciet.fiocruz.br>>. Acessado em: 21 de Setembro 2023.

O trabalho de desenvolvimento em um *Notebook* permite a organização do documento em blocos de código e de texto formatável, possibilitando inclusive a exibição de um resultado logo após a execução de um bloco de código – sendo esse resultado um gráfico, uma tabela ou uma mensagem – é fácil regredir em termos de modularização e padronização de codificação, onde essas metodologias são comuns e incentivadas em projetos de desenvolvimento de sistemas. Nesse sentido, muito do que foi produzido possui repetições de código que poderiam ser abstraídas e modularizadas com funções e laços de repetição.

Apesar da tendência mencionada acima, existiu um esforço de modularizar os procedimentos e torná-los mais reutilizáveis e com funcionalidades destacadas. Assim, no final do projeto, foi criado um módulo de nome “funcoes_comuns.py” onde são definidas 7 funções. Apesar de ser um arquivo de código *Python*, o módulo é facilmente incluído para utilização dentro de um *Jupyter Notebook*.

Os códigos e dados compilados se encontram disponibilizados em um repositório no *GitHub*⁶.

4.4.1 Funções Comuns

Apêndice 6.1, é disponibilizado o módulo na íntegra. O código 6 possui comentários explicativos de seu funcionamento, logo bem documentado para entendimento e uso posterior. As funções de destaque são:

- build_df_by_sql_query
- fetch_all_data_from_response
- fetch_pop_fem_from_tabnet
- get_tb_municipios

As outras funções possuem apenas caráter auxiliar:

- dot_print
- return_sql_query
- rows_and_columns_to_df

A função “fetch_all_data_from_response” foi realizada com o auxílio da documentação do *Elasticsearch* (Elasticsearch team, 2023b) para buscar o restante dos dados solicitados na consulta. Ela é chamada apenas na função “build_df_by_sql_query”, caso exista mais dados a serem retornados. Nessa função, o parâmetro é uma consulta SQL, e seu retorno é um *Dataframe* da biblioteca *Pandas* onde os dados são trabalhados.

⁶ Códigos e dados compilados. Disponível em <<https://github.com/gabrielxfs/maternal-near-miss/>>.

A função “fetch_pop_fem_from_tabnet” utiliza o utilitário Curl⁷ para buscar os dados do Tabnet DataSUS e organizá-los no formato convencionado.

A função “get_tb_municipios” formaliza o formato convencionado a partir de uma tabela contendo todos os municípios e seus respectivos códigos identificáveis nas bases de dados do DataSUS.

4.4.2 Bloco 1

Os indicadores do bloco 1 são relacionados às condições socioeconômicas e de acesso ao serviço de saúde. No código 1, podem ser verificadas as consultas em SQL. No *Jupyter Notebook* do bloco 1 essas consultas são chamadas, os dados são organizados junto com os calculados de outra forma por outros membros e por fim é gerado um arquivo csv com os dados compilados.

4.4.3 Bloco 2

Os indicadores do bloco 2 estão relacionados ao planejamento reprodutivo. No código 2, podem ser verificadas as consultas em SQL. No *Jupyter Notebook* do Bloco 2 são realizadas as consultas, é chamado a função “fetch_pop_fem_from_tabnet”, é reorganizado no formato convencionado os dados feitos por outros membros – provenientes do SIH e Tabnet ANS – e por fim é gerado um arquivo csv com os dados compilados.

4.4.4 Bloco 3

Os indicadores do bloco 3 são relacionados à assistência pré-natal. No código 3, podem ser verificadas as consultas em SQL. No *Jupyter Notebook* do Bloco 3 são realizadas as consultas, é reorganizado no formato convencionado os dados de sífilis congênita – provenientes do MS – e por fim é gerado um arquivo csv com os dados compilados.

4.4.5 Bloco 4

Os indicadores do bloco 4 são relacionados à assistência ao parto. No código 4, podem ser verificadas as consultas em SQL. No *Jupyter Notebook* do Bloco 4 são feitas as consultas e é gerado um arquivo csv com os dados compilados.

4.4.6 Bloco 5

Os indicadores do bloco 5 são relacionados às condições de nascimento. No código 5, podem ser verificadas as consultas em SQL. No *Jupyter Notebook* do Bloco 5 são feitas as consultas e é gerado um arquivo csv com os dados compilados.

⁷ Curl. Disponível em <<https://curl.se/docs/manpage.html>>. Acessado em: 23 de Setembro 2023.

4.4.7 Ajustes do “Deslocamento ao Parto”

Os indicadores de “deslocamento ao parto”, que foram realizados por uma pesquisadora do IBGE, não estavam no formato convencionado. Deste modo foi necessário reorganizar esses dados. Assim foi feito esse ajuste em um *Jupyter Notebook* referenciado no código 7.

5 EXPERIMENTOS E RESULTADOS

Os dados disponibilizados eram objeto de análise para a construção de um painel de vigilância da saúde materna, de forma que fosse adequado às demandas que os gestores de saúde possuem. Nesse sentido, surgiu a necessidade de fazer os primeiros experimentos de formatação das visualizações, e para observar o comportamento dos dados foi escolhido o município do Rio de Janeiro para análise dos indicadores, já que esse era um cenário familiar para as pesquisadoras envolvidas no projeto. É importante destacar que todos os dados apresentados nos gráficos são anônimos e disponíveis publicamente no site do DataSUS, garantindo que não houve quebra de protocolos ao exibi-los ou ao disponibilizá-los por meio do *Tableau Public*. Após a análise e validação dos dados, que se deu por meio de reuniões recorrentes para ajustes e consultas, foi possível repassar os procedimentos para a compilação dos indicadores e a documentação dos mesmos, para a equipe do Observatório Obstétrico.

5.1 EXPERIMENTOS

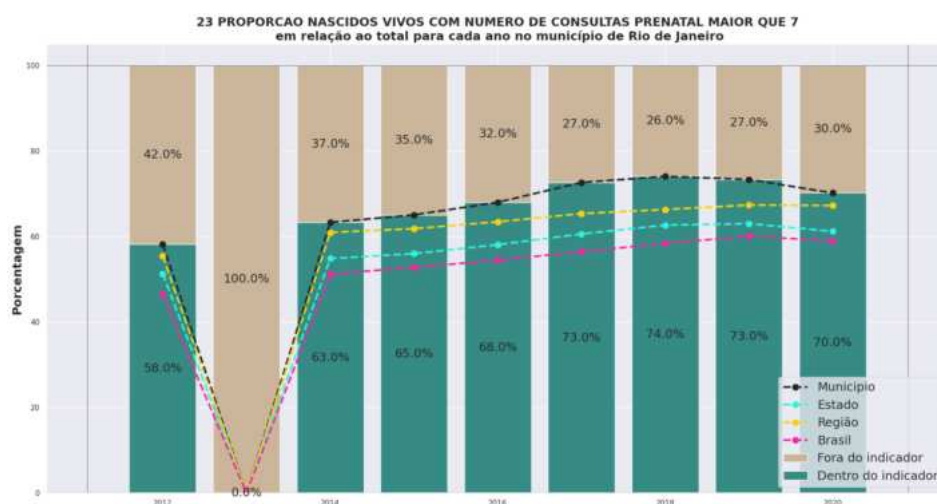
A experimentação com visualizações foi fundamental para promover clareza para as pesquisadoras envolvidas no projeto, de forma que serviu de base para validação acerca da confiabilidade dos dados e acurácia do ETL. Nesta seção serão apresentados os diferentes gráficos desenvolvidos para validação ao longo do projeto.

5.1.1 Gráficos iniciais feitos no *matplotlib*

As experimentações foram realizadas por meio do uso da biblioteca *matplotlib*, onde inicialmente foram explorados gráficos de barras e linhas sobrepostas, para análise ao longo dos anos, onde cada ano possui um valor - geralmente uma razão. A seguir encontram-se alguns dos gráficos experimentados.

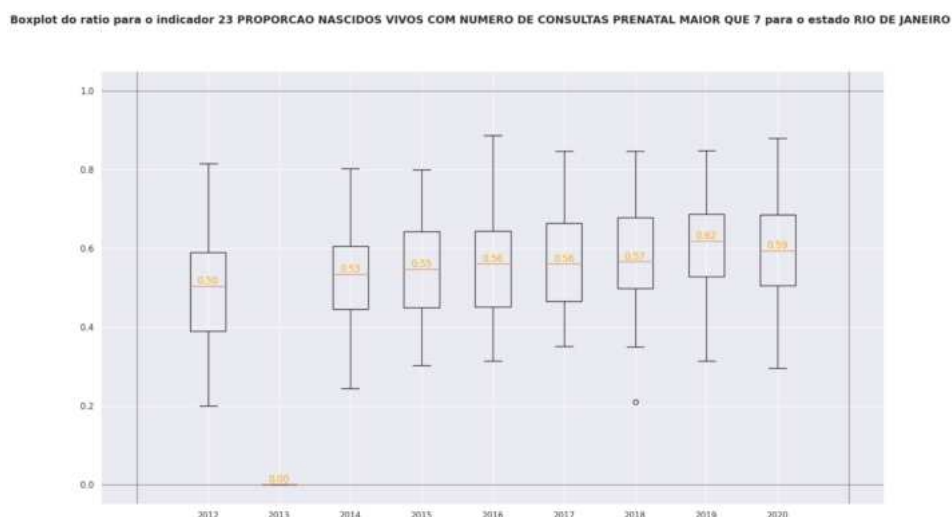
Na Figura 1 é possível observar que no ano de 2013 faltam dados. Isso é um problema na própria base disponibilizada. Além disso, o indicador pareceu coerente de acordo com a avaliação dos pesquisadores.

Figura 1 – Gráficos de Linhas/Barras do Indicador “Consultas pré-natal maior que 7”



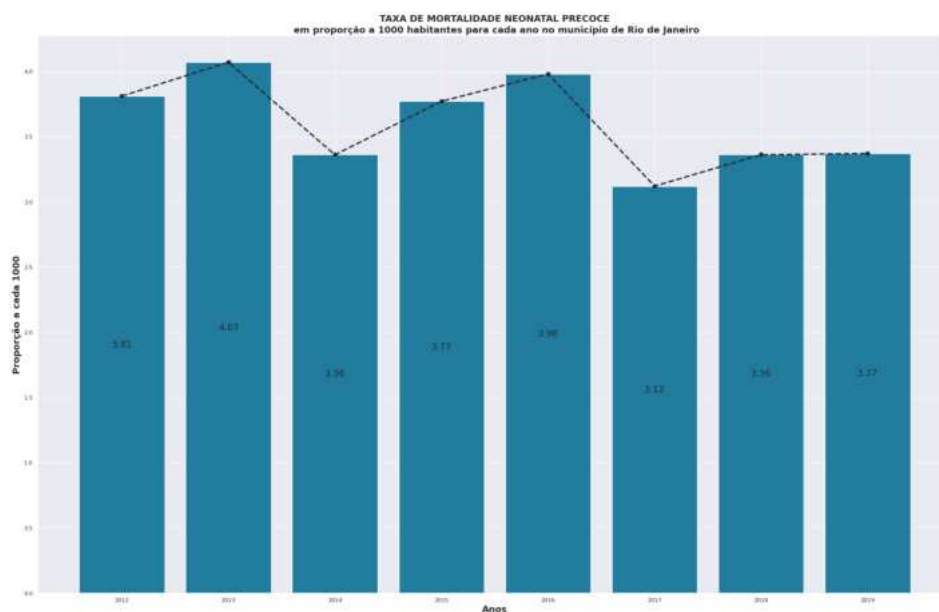
Na Figura 2 é possível observar como os dados são bem distribuídos para esse indicador, sem muitos outliers presentes.

Figura 2 – *Boxplot* do Indicador “Consultas pré-natal maior que 7”



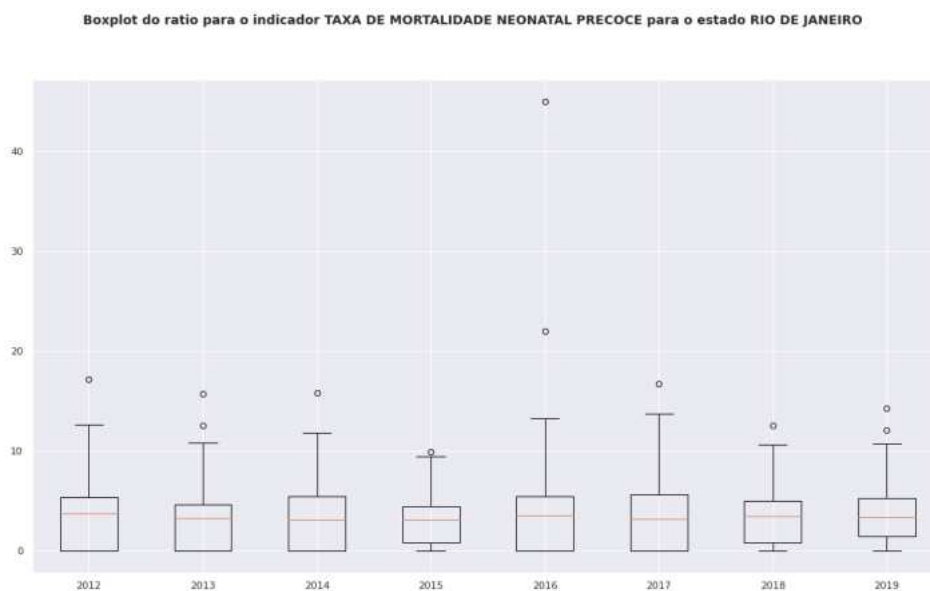
Na figura 3 é possível observar que os dados são bem comportados, porém foi decidido excluir do painel por perceberem discrepâncias de outras análises já conhecidas.

Figura 3 – Gráfico Linhas/Barras do Indicador “Taxa de mortalidade neonatal precoce”



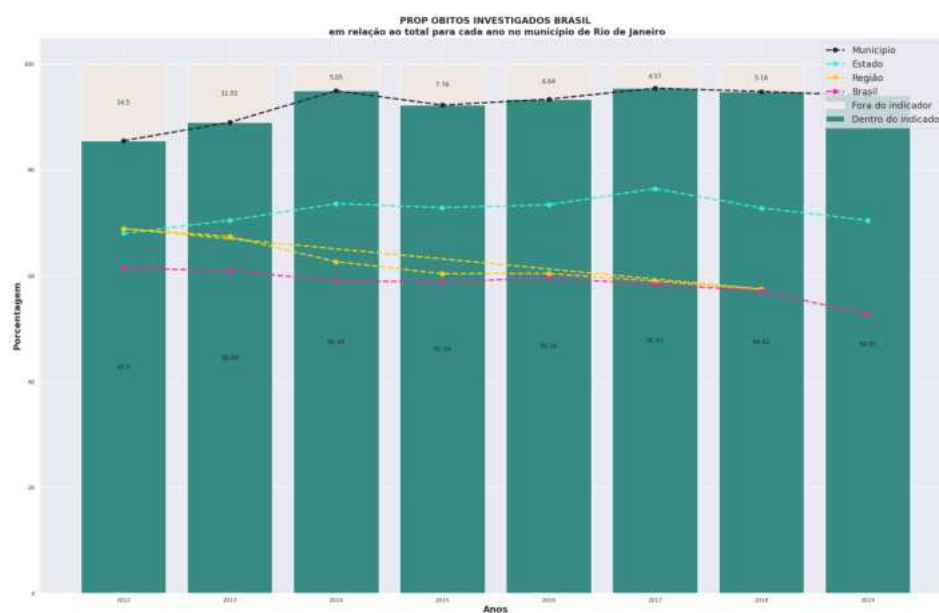
Na figura 4 é possível observar alguns casos de municípios no Estado do Rio de Janeiro onde os valores são destoantes de uma distribuição normal, além desta distribuição estar concentrada em valores menores.

Figura 4 – *Boxplot* do Indicador “Taxa de mortalidade neonatal precoce”



Na figura 5 o gráfico de linhas referente a região não possuía valores em 2019 por falhas no tratamento dos dados faltantes na gama de municípios compilados do sudeste.

Figura 5 – Gráfico Linhas/Barras do Indicador “Proporção de óbitos maternos investigados”



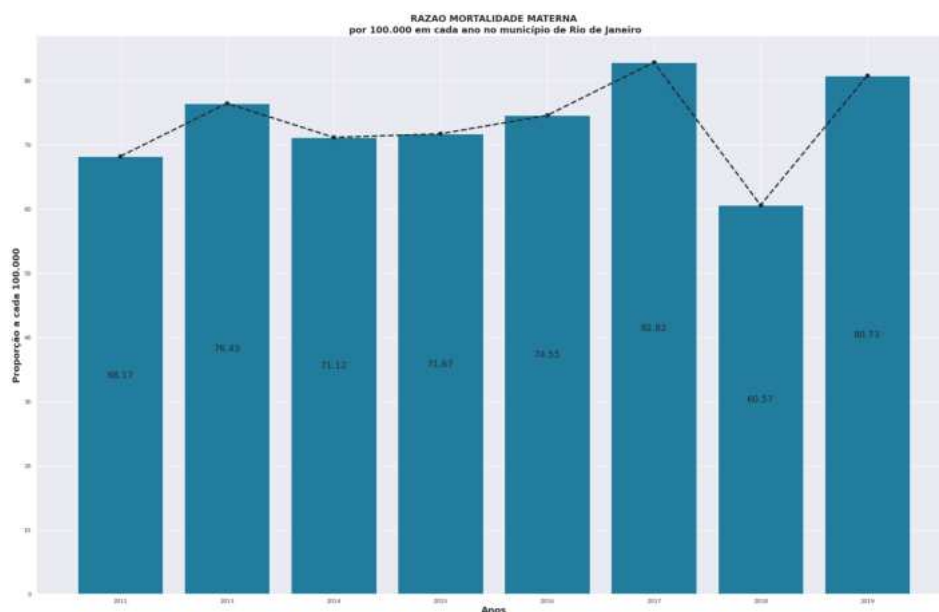
Na figura 6 é possível observar uma boa distribuição, porém alertou as especialistas para averiguar sobre o grau de preenchimento dessa variável – dado que a mediana estava muito abaixo do esperado –, sendo que foi sinalizado a mudança de fonte desses dados, uma vez que nem todos atualizam essa informação no SIM.

Figura 6 – *Boxplot* do Indicador “Proporção de óbitos maternos investigados”



Nas figuras 7 e 9 os dados se comportam bem, porém foi amplamente discutida a correteude em relação ao número.

Figura 7 – Gráfico Linhas/Barras do Indicador “Razão de mortalidade materna”



As figuras 8 e 10 corroboram com a preocupação de analisar esse indicador em municípios pequenos, onde a ocorrência de poucos óbitos provocaria picos na medida desse indicador. Também é possível observar uma distribuição enviesada em valores baixos e com muitos valores grandes, deste modo, sendo uma possível distribuição exponencial, onde a maior parte dos valores são pequenos e uma pequena parcela possui valores muito discrepantes.

Figura 8 – *Boxplot* do Indicador “Razão de mortalidade materna”

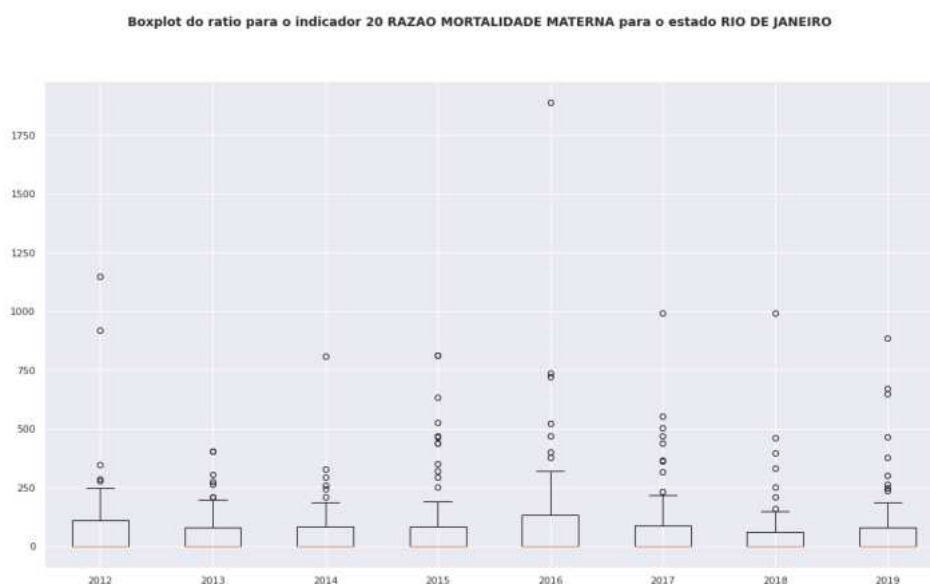


Figura 9 – Gráfico Linhas/Barras do Indicador “Razão de mortalidade materna direta”

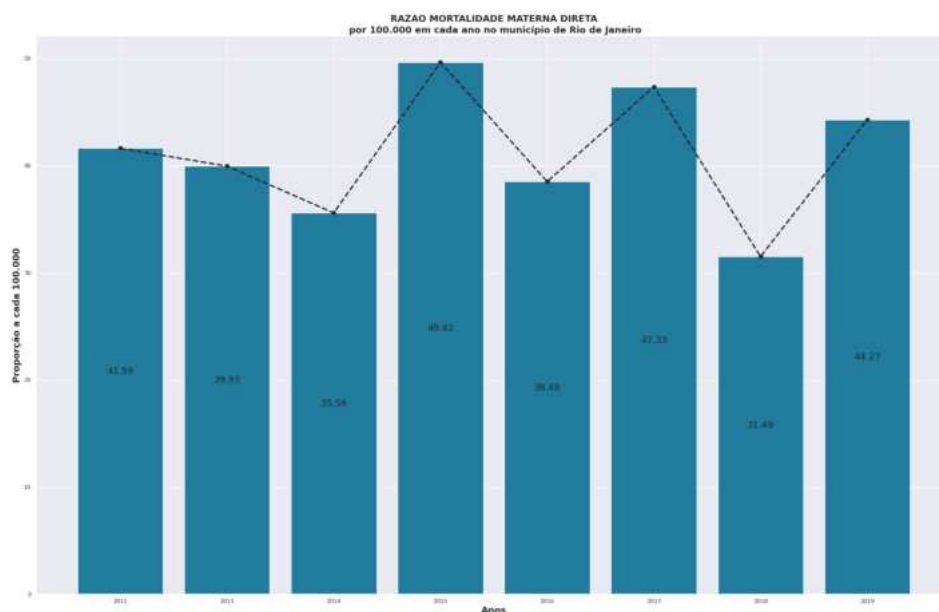
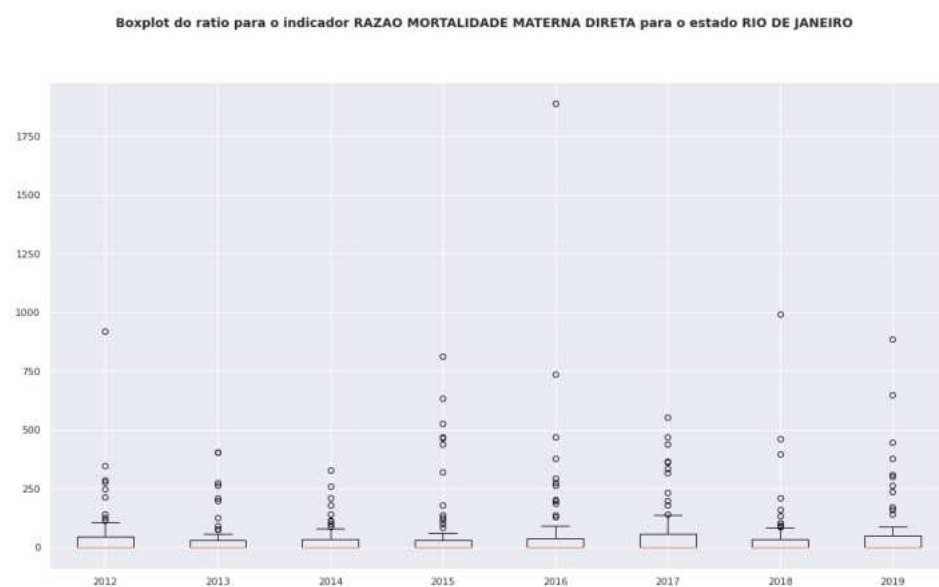
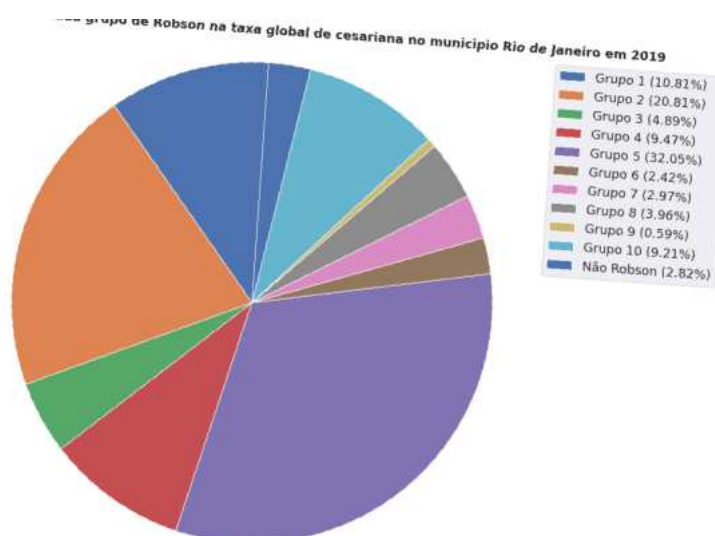


Figura 10 – *Boxplot* do Indicador “Razão de mortalidade materna direta”



Na figura 11 exibe um dos testes de visualização que poderiam ser incluídos no painel, onde seria plausível para os indicadores de grupo de Robson, pois possuem muitas categorias a serem analisadas. Nesse caso, esse modelo de visualização não foi incorporado, visto que a avaliação é de que o gráfico do tipo pizza não faz sentido para análises diretas quando existem muitos setores representados.

Figura 11 – Gráfico de Pizza do Indicador “Grupo de Robson na taxa global de cesariana”



5.1.2 Experimento de visualizações no *Tableau*

Anterior a decisão de como e onde seria hospedado e desenvolvido o painel, foram discutidas junto à equipe – que em parceria – desenvolveu o desenho gráfico do projeto do painel, a *Odd Studio*¹, algumas ferramentas para desenvolvimento, como o *PowerBI*, *R Shiny* e *Tableau*. Com isso, para iniciar os testes, foi recomendado pela *Odd Studio* a experimentação com o *Tableau*, uma ferramenta de visualização de dados poderosa com várias possibilidades de gráficos e integração. Importante ressaltar que foi utilizada a versão gratuita *Tableau Public*, que possui algumas limitações em relação à versão premium.

Dos gráficos 12 a 14 são exibidos alguns dos testes realizados. Gráficos de barra empilhada e mapas foram novas visualizações que a ferramenta possibilitou gerar com mais facilidade em relação à utilização do *matplotlib*. Apesar disso, o desenvolvimento final do painel apresenta apenas gráficos de barras, linhas e quadros informativos.

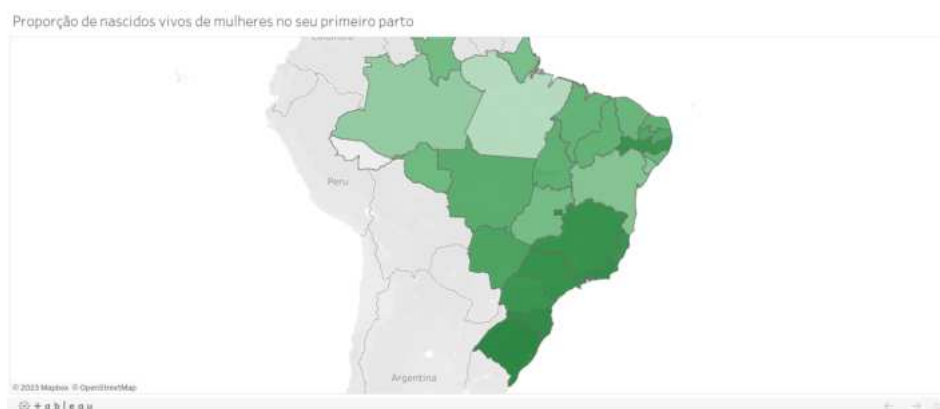
¹ *Odd Studio*. Disponível em <<https://odd.studio/>>. Acessado em: 14 de Setembro 2023.

Figura 12 – Gráfico de barras empilhadas do Indicador “Contribuição do Grupo de Robson para a taxa global de cesariana”



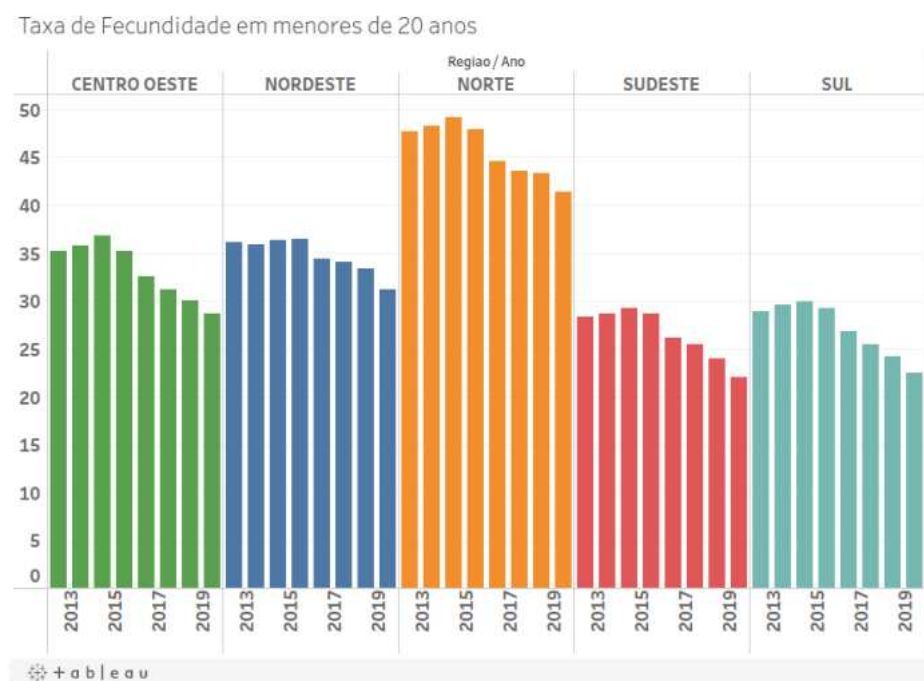
Fonte: <<https://public.tableau.com/app/profile/igor.murras/vizzes>>

Figura 13 – Mapa do Indicador “Proporção de nascidos vivos de mulheres no seu primeiro parto”



Fonte: <<https://public.tableau.com/app/profile/igor.murras/vizzes>>

Figura 14 – Gráfico de barras do Indicador “Taxa de fecundidade em menores de 20 anos”



Fonte: <<https://public.tableau.com/app/profile/igor.murras/vizzes>>

5.2 RESULTADOS

O resultado se resume em o painel ser desenvolvido, em termos de codificação e programação do site, pela equipe do Observatório Obstétrico. Mas antes disso, em diversas reuniões de alinhamento e construção, foi desenvolvido o projeto do painel em parceria com a *Odd Studio*.

5.2.1 Design do painel com a *Odd Studio*

A *Odd Studio* é uma empresa que tem como objetivo transformar dados em soluções através do projeto gráfico. De acordo com o *website* oficial, a missão da empresa é descrita como “Transformamos dados em soluções únicas pelas lentes do *design*.” Desde o início do desenvolvimento do projeto, foram discutidas questões relacionadas às variáveis envolvidas, como seriam apresentadas graficamente, o objetivo de cada visualização, e como essas informações seriam comparadas, seja internamente, seja com alguma referência externa.

Além disso, foi informado diversos exemplos de painel de visualização utilizando o *Tableau Public*, inclusive um relacionado a COVID-19². Também foi repassado, dado um modelo teórico resumido, que um bom painel precisa ter bem consolidado uma relação

² COVID-19: Government measures - worldwide. Disponível em <<https://public.tableau.com/app/profile/visualitics/viz/...>>. Acessado em: 26 de Setembro 2023.

lógica entre os dados, uma hierarquia visual e um objetivo bem definido dado o público alvo e a função que almeja ter.

Importante destacar que as decisões de

6 CONCLUSÃO

O desenvolvimento deste trabalho representou uma jornada de aprendizado e aperfeiçoamento, tanto no âmbito profissional quanto acadêmico. Durante esse período, a aplicação de técnicas de ciência de dados, aliada ao contato direto com profissionais da área da saúde, nos proporcionou uma compreensão mais profunda da importância dos dados como objetos de análise. Além disso, evidenciou como os dados ganham valor quando associados ao conhecimento especializado de profissionais da área.

O potencial de impacto social deste trabalho é significativo, pois o objetivo principal foi desenvolver um painel de vigilância da saúde materna. Este painel depende de bases de dados sólidas e bem estruturadas, que formam o alicerce para qualquer tipo de análise subsequente. Assim, o trabalho não apenas estabelece as bases de dados necessárias, mas também promove a capacidade de realizar análises mais precisas e informadas, tudo isso com a supervisão e crivo de profissionais da saúde.

Além disso, é importante destacar que, em entrevistas realizadas pela equipe da *Odd Studio* com gestores de saúde, foi relatado que muitos municípios pequenos não possuem recursos para construir os indicadores trabalhados no projeto¹. Dado isso, é crucial mencionar que a incompletude ou a qualidade dos dados serve como métrica para avaliar o grau de informatização e digitalização — seja do ponto de vista dos processos e práticas profissionais, seja do ponto de vista dos recursos e infraestrutura — dos diversos pontos de atendimento do SUS nas mais diversas localidades. Esses dados podem ser utilizados para avaliar a qualidade do serviço e determinar o quanto é necessário investir em saúde naquela região (ANDREU-PEREZ et al., 2015).

Assim, as aplicações de análises de dados não são importantes apenas no sentido clínico e específico dos cuidados em saúde, mas também servem para uma análise sociológica, política e econômica das possíveis soluções nesses campos aplicados à saúde, especialmente à saúde pública (ANDREU-PEREZ et al., 2015). Por fim, é importante reforçar que a colaboração interdisciplinar se mostrou extremamente benéfica neste projeto e pode ser ainda mais valiosa em futuros projetos, abrindo novas perspectivas e oportunidades de inovação, o que será discutido na próxima seção.

¹ Entrevistas Painel Vulnerabilidade à Morte Materna. Disponível em <<https://docs.google.com/document/d/1-Dq7miOO4EGhIPIfA7uFqhDwD385eTOzlOum3tO80AA/edit#heading=h>>. Acessado em: 25 de Setembro 2023.

6.1 TRABALHOS FUTUROS

Dada a progressão de trabalhos anteriores (CARVALHO; ALMEIDA, 2017)(SAMPALIO, 2019)(COSTA; PEREIRA, 2021) em torno da: utilização de modelos de banco de dados mais atuais, maturação das técnicas de ETL e aumento do escopo dos dados em termos do conjunto de anos, UFs e bases de dados, este trabalho homologa as projeções dos mesmos, sem necessariamente construir do zero uma base de dados ou replicar um procedimento de ETL, visto que foi utilizada a plataforma PCDaS.

Mencionando os princípios de: reprodutibilidade da ciência e, paradoxalmente, o “Não repita a si mesmo” (*"Don't repeat yourself"* (DRY) (HUNT; THOMAS, 2000)); seria interessante que os procedimentos de ETL das bases de dados do DATASUS - e suas respectivas correlações com outras fontes de dados tal como IBGE e outras - fossem geridos como um projeto de código aberto. Deste modo, pesquisadores de todo Brasil poderiam contribuir com o escopo mais geral de tratamento, transformação, e engenharia de Dados, ao passo que poderiam replicar o projeto em suas respectivas infraestruturas, os adequando e complementando com suas especificidades.

Outros desdobramentos sugeridos pela equipe incluem a expansão do painel para abranger a saúde perinatal, o aumento da granularidade espacial dos dados para níveis intermunicipais – como zonas, bairros e até hospitais – e a ampliação da granularidade temporal para dados mensais, em vez de anuais. Essas melhorias poderiam fortalecer ainda mais o impacto e a aplicabilidade do painel em estudos e práticas de saúde pública.

REFERÊNCIAS

- ANDREU-PEREZ, J. et al. Big Data for Health. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 19, n. 4, p. 1193–1208, jul. 2015.
- BIONETWORKS, Info@sagebase. o. S. **PCDaS. Plataforma de Ciência de Dados aplicada à Saúde**. 2021. [Accessed 06-09-2023]. Disponível em: <https://doi.org/10.7303/syn25882127>.
- CARVALHO, R.; ALMEIDA, I. Ambiente de exploração de dados de saúde usando um Banco de Dados NoSQL. **Universidade Federal do Rio de Janeiro**, Universidade Federal do Rio de Janeiro, ago. 2017.
- CHAMBERLIN, D. D. Early History of SQL. **IEEE Annals of the History of Computing**, IEEE, v. 34, n. 4, p. 78–82, nov. 2012.
- CHANDRA, D. G. BASE analysis of NoSQL database. **Future Generation Computer Systems**, North-Holland, v. 52, p. 13–21, nov. 2015. ISSN 0167-739X.
- COSTA, P. G.; PEREIRA, G. L. A. ELK saúde: ambiente de análise e visualização de dados do SUS utilizando Elasticsearch e Kibana. **Universidade Federal do Rio de Janeiro**, Universidade Federal do Rio de Janeiro, jan. 2021. Disponível em: <https://pantheon.ufrj.br/handle/11422/13689>.
- DATASUS. **Informações de Saúde (TABNET) – DATASUS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/IcT5Z>.
- DATASUS. **SIAB**. 2023. [Accessed 06-09-2023]. Disponível em: <https://web.archive.org/web/20230608062008/http://www2.datasus.gov.br/SIAB/index.php?area=01>.
- DATASUS. **SIHSUS - DATASUS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://web.archive.org/web/20191031112905/http://datasus.saude.gov.br/sistemas-e-aplicativos/hospitales/sihsus>.
- DONOHU, D. 50 Years of Data Science. **Journal of Computational and Graphical Statistics**, Taylor & Francis, out. 2017. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734>.
- Elasticsearch team. **Overview | Elasticsearch Guide [8.11] | Elastic**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/88UK2>.
- Elasticsearch team. **Paginating through a large response | Elasticsearch Guide [8.11] | El. . .** 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/cH6E6>.
- Elasticsearch team. **What is Elasticsearch? | Elasticsearch Guide [8.10] | Elastic**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/SaB1t>.
- HUNT, A.; THOMAS, D. **The pragmatic programmer: from journeyman to master**. [S.l.]: Addison-Wesley Longman Publishing Co., Inc., 2000.

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data**. Hoboken, NJ, USA: John Wiley and Sons, Inc., 2004.

Matplotlib development team. **History — Matplotlib 3.8.1 documentation**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/Bjuvr>.

MCKINNEY, W. pandas: a Foundational Python Library for Data Analysis and Statistics. **Python High Performance Science Computer**, jan. 2011. Disponível em: https://www.researchgate.net/publication/265194455_pandas_a_Foundational_Python_Library_for_Data_Analysis_and_Statistics.

MURDOCH, M. T. B. The Inevitable Application of Big Data to Health Care. **JAMA**, JAMA Network, v. 309, n. 13, p. 1351–1352, abr. 2013. ISSN 0098-7484.

NATIONS, G. A. U. **Resolution adopted by the General Assembly on 6 July 2017**. 2017. https://upload.wikimedia.org/wikipedia/commons/9/9d/A_RES_71_313_E.pdf. [Accessed 06-09-2023].

PCDAS. **Vigilância da mortalidade materna e da morbidade materna grave | PCDaS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/cduuc>.

PCDAS, T. de Governança e Engenharia de D. **Sistema de Informações sobre Nascidos Vivos – SINASC - PCDaS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://pcdas.iciet.fiocruz.br/conjunto-de-dados/sistema-de-informacao-sobre-nascidos-vivos/documentacao>.

PERKEL, J. M. Why Jupyter is data scientists' computational notebook of choice. **Nature**, Nature Publishing Group, v. 563, p. 145–146, out. 2018.

Project Jupyter. **Project Jupyter | JupyterHub**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.ph/bW7Hl>.

Python Software Foundation. **General Python FAQ — Python 3.12.0 documentation**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/LapWL>.

REDDY, M. **API Design for C++**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.

SAMPAIO, R. S. Ambiente de Dados do SIHSUS com MongoDB. **Universidade Federal do Rio de Janeiro**, Universidade Federal do Rio de Janeiro, ago. 2019. Disponível em: <https://pantheon.ufrj.br/handle/11422/11197>.

SAÚDE, S. S. Único de. **Dicionário de Dados SIM: tabela DO**. 2012. <https://svs.aims.gov.br/daent/cgiae/sim/documentacao/dicionario-de-dados-SIM-tabela-DO.pdf>. [Accessed 06-09-2023].

SAÚDE, S. S. Único de. **Dicionário de Dados SINASC: tabela DN**. 2012. https://svs.aims.gov.br/daent/cgiae/sinasc/documentacao/dicionario_de_dados_SINASC_tabela_DN.pdf. [Accessed 06-09-2023].

SILVA, Y. R. d. S. d. Análise da qualidade das informações do Grupo de Robson contidas no SINASC, no município do Rio de Janeiro, 2014 e 2019. **Universidade Federal do Rio de Janeiro**, Universidade Federal do Rio de Janeiro, dez. 2022. Disponível em: <https://pantheon.ufrj.br/handle/11422/20170>.

SUS - Sistema Único de Saúde. **DATASUS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/RTcYU>.

SUS, A. **ANS TABNET**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/iXhmf>.

SUS, M. **cnes**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/3BVS6>.

SUS, S. **Apresentação - SIM - CGIAE - DAENT - SVSA/MS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/JsCt4>.

SUS, S. **Apresentação - SINASC - CGIAE - DAENT - SVSA/MS**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.ph/MBFTx>.

SUS, S. **SINANWEB - Página inicial**. 2023. [Accessed 06-09-2023]. Disponível em: <https://archive.is/GDgCT>.

UNWIN, A. **Why Is Data Visualization Important? What Is Important in Data Visualization**. 2020. [Accessed 06-09-2023]. Disponível em: <https://archive.is/SDiil>.

APÊNDICES

APÊNDICE A – CÓDIGOS EM PYTHON E CONSULTAS SQL

Código 1 – Consultas em SQL aos dados do SINASC referentes ao Bloco 1

```
-- Nascidos vivos de mulheres de 2012 a 2020
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_DE_NASCIDOS_VIVOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres com idade inferior a 20 anos
(gestação na adolescência)
SELECT CODMUNRES, ano_nasc, count(*) as NVM_MENOR_QUE_20_ANOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
IDADEMAE >= 10 AND
IDADEMAE <= 19
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres com idade entre 20 e 34 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_ENTRE_20_E_34_ANOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
IDADEMAE >= 20 AND
IDADEMAE <= 34
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres com idade igual ou superior
a 35 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_MAIOR_QUE_34_ANOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
IDADEMAE >= 35 AND
IDADEMAE <= 55
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres de cor da pele branca
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_COR_DA_PELE_BRANCA
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
```

```

RACACORMAE=1
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres de cor da pele preta
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_COR_DA_PELE_PRETA
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
RACACORMAE=2
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres de cor da pele parda
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_COR_DA_PELE_PARDA
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
RACACORMAE=4
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres de cor da pele amarela
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_COR_DA_PELE_AMARELA
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
RACACORMAE=3
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos de mulheres indígenas
SELECT CODMUNRES, ano_nasc, count(*) as NVM_INDIGENAS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
RACACORMAE=5
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com escolaridade até 4 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_ESCOLARIDADE_ATE_3
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
(ESCMAE=1 OR ESCMAE=2)
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com escolaridade de 4 a 7 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_ESCOLARIDADE_DE_4_A_7
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND

```

```

ano_nasc <= 2020) AND
(ESMAE=3)
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com escolaridade de 8 até 11 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_COM_ESCOLARIDADE_DE_8_A_11
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
(ESMAE=4)
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com escolaridade mais de 11 anos
SELECT CODMUNRES, ano_nasc, count(*) as
      NVM_COM_ESCOLARIDADE_ACIMA_DE_11
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
(ESMAE=5)
GROUP BY CODMUNRES, ano_nasc;

```

Código 2 – Consultas em SQL aos dados do SINASC referentes ao Bloco 2

```

-- Nascidos vivos de mulheres de 2012 a 2020
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_DE_NASCIDOS_VIVOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020
GROUP BY CODMUNRES, ano_nasc;

-- Número de nascidos vivos de mulheres com menos de 20 anos
SELECT CODMUNRES, ano_nasc, count(*) as NVM_MENOR_QUE_20
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
IDADEMAE < 20
GROUP BY CODMUNRES, ano_nasc;

-- Número de mulheres com mais de três partos anteriores
SELECT CODMUNRES, ano_nasc, count(*) as
      MULHERES_COM MAIS DE TRES PARTOS ANTERIORES
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
(QTDPARTNOR > 3
OR (QTDPARTNOR > 2 AND QTDPARTCES > 0)
OR (QTDPARTNOR > 1 AND QTDPARTCES > 1)
OR (QTDPARTNOR > 0 AND QTDPARTCES > 2)

```

```
OR QTDPARTCES > 3)
GROUP BY CODMUNRES, ano_nasc;
```

Código 3 – Consultas em SQL aos dados do SINASC referentes ao Bloco 3

```
-- Nascidos vivos de mulheres de 2012 a 2020
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_DE_NASCIDOS_VIVOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com pelo menos uma consulta de pré-natal
SELECT CODMUNRES, ano_nasc,
count(*) as MULHERES_COM_PELO_MENOS_UMA_CONSULTA_PRENATAL
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
CONSPRENAT >= 1
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de mulheres com início do pré-natal nos três primeiros
meses de gestação
SELECT CODMUNRES, ano_nasc,
count(*) as MULHERES_COM_INICIO_PRECOCE_DO_PRENATAL
FROM "datasus-sinasc"
WHERE (ano_nasc >= 2012 AND
ano_nasc <= 2020) AND
(MESPRENAT=1 OR MESPRENAT=2 OR MESPRENAT=3)
GROUP BY CODMUNRES, ano_nasc;

-- Número de mulheres com mais de sete consultas de pré-natal
SELECT CODMUNRES, ano_nasc,
count(*) as MULHERES_COM MAIS_DE_SETE_CONSULTAS_PRENATAL
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020 AND
CONSPRENAT > 7
GROUP BY CODMUNRES, ano_nasc;
```

Código 4 – Consultas em SQL aos dados do SINASC referentes ao Bloco 4

```
-- Nascidos vivos de mulheres de 2012 a 2020
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_DE_NASCIDOS_VIVOS
FROM "datasus-sinasc"
WHERE ano_nasc >= 2012 AND
ano_nasc <= 2020
GROUP BY CODMUNRES, ano_nasc;
```

```

-- Quantidade de nascimentos por cesariana
SELECT CODMUNRES, ano_nasc, count(*) as MULHERES_COM_PARTO_CESARIANA
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 1 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
      MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_1
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=1
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 2 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
      MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_2
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=2
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 3 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
      MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_3
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=3
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 4 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
      MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_4
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=4
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 5 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as

```

```

        MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_5
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=5
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 6 a 9 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
        MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_6_ao_9
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON>=6 AND TPROBSON <=9
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascidos vivos do grupo 10 de Robson
SELECT CODMUNRES, ano_nasc, count(*) as
        MULHERES_DENTRO_DO_GRUPO_DE_ROBSON_10
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=10
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 1 de Robson para a taxa global de
    cesariana
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_CESARIANA_GRUPO_ROBSON_1
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=1 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 2 de Robson para taxa global de
    cesariana
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_CESARIANA_GRUPO_ROBSON_2
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=2 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 3 de Robson para taxa global de
    cesariana
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_CESARIANA_GRUPO_ROBSON_3
FROM "datasus-sinasc"

```

```

WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=3 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 4 de Robson para taxa global de
cesariana
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_CESARIANA_GRUPO_ROBSON_4
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=4 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 5 de Robson para taxa global de
cesariana
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_CESARIANA_GRUPO_ROBSON_5
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=5 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 6 a 9 de Robson para taxa global de
cesariana
SELECT CODMUNRES, ano_nasc, count(*) as
TOTAL_CESARIANA_GRUPO_ROBSON_6_ao_9
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON >=6 AND TPROBSON <=9 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

-- Contribuição relativa do grupo 10 de Robson para taxa global de
cesariana
SELECT CODMUNRES, ano_nasc, count(*) as
TOTAL_CESARIANA_GRUPO_ROBSON_10
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
TPROBSON=10 AND PARTO=2
GROUP BY CODMUNRES, ano_nasc;

```

Código 5 – Consultas em SQL aos dados do SINASC referentes ao Bloco 5

```

-- Nascidos vivos de mulheres de 2012 a 2020
SELECT CODMUNRES, ano_nasc, count(*) as TOTAL_DE_NASCIDOS_VIVOS
FROM "datasus-sinasc"

```

```

WHERE ano_nasc >=2012 AND
ano_nasc <=2020
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de baixo peso ao nascer
SELECT CODMUNRES, ano_nasc, count(* as NASCIDOS_VIVOS_COM_BAIXO_PESO
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
PESO < 2500
GROUP BY CODMUNRES, ano_nasc;

-- Quantidade de nascimentos prematuros
SELECT CODMUNRES, ano_nasc, count(* as NASCIDOS_VIVOS_PREMATUROS
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
GESTACAO < 5
GROUP BY CODMUNRES, ano_nasc;

-- Número de nascimentos termo precoce
SELECT CODMUNRES, ano_nasc, count(* as NASCIDOS_VIVOS_TERMO_PRECOCE
FROM "datasus-sinasc"
WHERE ano_nasc >=2012 AND
ano_nasc <=2020 AND
SEMAGESTAC >= 37 AND
SEMAGESTAC <= 38
GROUP BY CODMUNRES, ano_nasc;

```

Código 6 – Módulo de Funções Comuns em Python

```

import subprocess
import re
import requests
import pandas as pd
import os
import json
import sys

# endereço do acesso a api do elasticsearch do pcdas cujo
# contem as bases de dados do datasus (sim, sinasc, etc.)
url_base = "https://bigdata-api.fiocruz.br"
# token de acesso da api
token = '' # o token fornecido foi omitido

# função para escrever um ponto a cada vez que é chamada
def dot_print():
    sys.stdout.write('.')

```



```

sys.stdout.flush()

# função que retorna o objeto com o token e a query sql
# para se consultar no elasticsearch
def return_sql_query(query_string):
    ret = {
        "token": {
            "token": token
        },
        "sql": {
            "sql": {
                "query": query_string
            }
        }
    }
    return ret

# Função para pegar o próximo bloco de dados pela api
# através do cursor, ao qual é um token que é repassado
# a cada novo bloco de dados, cessa quando não for
# repassado algum cursor
def fetch_all_data_from_response(res, initial_records_rows):
    # enquanto for repassado algum cursor
    while 'cursor' in res.json():
        res_next = {
            "token": {
                "token": token
            },
            "sql": {
                "sql": {
                    "cursor": res.json()['cursor']
                }
            }
        }
        # requisita o próximo bloco de dados
        res = requests.post(os.path.join(url_base, 'sql_query'),
                             json=res_next)
        # acumula na lista
        initial_records_rows+=res.json()['rows']
        dot_print()
    print('(y)')
    return initial_records_rows

# Função para converter os objetos 'colunas' e 'linhas'
# advindo do elasticsearch para um dataframe pandas
def rows_and_columns_to_df(columns, rows):
    columns_array = [col['name'] for col in columns]

```

```

return pd.DataFrame(rows, columns=columns_array)

# Função para montar um dataframe pandas a partir dos
# dados advindos de acesso do elasticsearch via api
def build_df_by_sql_query(sql_query):
    # Pega o objeto sql a ser passado para a requisição
    sql_t = return_sql_query(sql_query)
    # Faz a primeira requisição dos dados
    resp_sql_t = requests.post(os.path.join(url_base,
        'sql_query'), json=sql_t)
    record_columns=resp_sql_t.json()['columns']
    record_rows=resp_sql_t.json()['rows']
    # Requisita o resto dos dados
    fetch_all_data_from_response(resp_sql_t, record_rows)
    # Informa a quantidade de linhas recuperadas
    print("Record rows length : "+str(len(record_rows)))
    # Converte os dados recuperados para dataframe pandas
    df_t = rows_and_columns_to_df(record_columns, record_rows)
    return df_t

# Função que faz uma requisição ao site do tabnet
# para recuperar a população feminina de todos os
# municípios do Brasil dado um respectivo ano,
# com as opções de 10 a 49 ou 10 a 20
def fetch_pop_fem_from_tabnet(year, faixa):
    if faixa == '20':
        command = "curl 'http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/
            cnv/popsvsbr.def' -X POST --data-raw 'Linha=Munic%EDpio&Coluna=
            Ano&Incremento=Popula%E7%E3o_residente&Arquivos=pop"+year+".dbf
            &SRegi%E3o=TODAS_AS_CATEGORIAS__&pesqmes2=Digite+o+texto+e+ache
            +f%E1cil&SUnidade_da_Federa%E7%E3o=TODAS_AS_CATEGORIAS__&
            pesqmes3=Digite+o+texto+e+ache+f%E1cil&SMunic%EDpio=
            TODAS_AS_CATEGORIAS__&pesqmes4=Digite+o+texto+e+ache+f%E1cil&
            SCapital=TODAS_AS_CATEGORIAS__&pesqmes5=Digite+o+texto+e+ache+f
            %E1cil&SRegi%E3o_de_Sa%FAde_%28CIR%29=TODAS_AS_CATEGORIAS__&
            pesqmes6=Digite+o+texto+e+ache+f%E1cil&SMacrorregi%E3o_de_Sa%
            FAd=TODAS_AS_CATEGORIAS__&pesqmes7=Digite+o+texto+e+ache+f%
            E1cil&SMicrorregi%E3o_IBGE=TODAS_AS_CATEGORIAS__&pesqmes8=
            Digite+o+texto+e+ache+f%E1cil&SRegi%E3o_Metropolitana_-_RIDE=
            TODAS_AS_CATEGORIAS__&pesqmes9=Digite+o+texto+e+ache+f%E1cil&
            STerrit%F3rio_da_Cidadania=TODAS_AS_CATEGORIAS__&pesqmes10=
            Digite+o+texto+e+ache+f%E1cil&SMacrorregi%E3o_PNDR=
            TODAS_AS_CATEGORIAS__&SAmaz%F4nia_Legal=TODAS_AS_CATEGORIAS__&
            SSemi%E1rido=TODAS_AS_CATEGORIAS__&SFaixa_de_Fronteira=
            TODAS_AS_CATEGORIAS__&SZona_de_Fronteira=TODAS_AS_CATEGORIAS__&
            SMunic%EDpio_de_extrema_pobreza=TODAS_AS_CATEGORIAS__&SSexo=2&
            pesqmes17=Digite+o+texto+e+ache+f%E1cil&SFaixa_Et%E1ria_1=3&

```

```

        SFaixa_Et%E1ria_1=4&pesqmes18=Digite+o+texto+e+ache+f%E1cil&
        SFaixa_Et%E1ria_2=TODAS_AS_CATEGORIAS__&pesqmes19=Digite+o+
        texto+e+ache+f%E1cil&SIdade_simples=TODAS_AS_CATEGORIAS__&
        formato=prn&mostre=Mostra'"
elif faixa == '49':
    command = "curl 'http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/
    cnv/popsvsbr.def' -X POST --data-raw 'Linha=Munic%EDpio&Coluna=
    Ano&Incremento=Popula%E7%E3o_residente&Arquivos=pop"+year+".dbf
    &SRegi%E3o=TODAS_AS_CATEGORIAS__&pesqmes2=Digite+o+texto+e+ache
    +f%E1cil&SUnidade_da_Federa%E7%E3o=TODAS_AS_CATEGORIAS__&
    pesqmes3=Digite+o+texto+e+ache+f%E1cil&SMunic%EDpio=
    TODAS_AS_CATEGORIAS__&pesqmes4=Digite+o+texto+e+ache+f%E1cil&
    SCapital=TODAS_AS_CATEGORIAS__&pesqmes5=Digite+o+texto+e+ache+f
    %E1cil&SRegi%E3o_de_Sa%FAde_%28CIR%29=TODAS_AS_CATEGORIAS__&
    pesqmes6=Digite+o+texto+e+ache+f%E1cil&SMacrorregi%E3o_de_Sa%
    FAd=TODAS_AS_CATEGORIAS__&pesqmes7=Digite+o+texto+e+ache+f%
    E1cil&SMicrorregi%E3o_IBGE=TODAS_AS_CATEGORIAS__&pesqmes8=
    Digite+o+texto+e+ache+f%E1cil&SRegi%E3o_Metropolitana_-_RIDE=
    TODAS_AS_CATEGORIAS__&pesqmes9=Digite+o+texto+e+ache+f%E1cil&
    STerrit%F3rio_da_Cidadania=TODAS_AS_CATEGORIAS__&pesqmes10=
    Digite+o+texto+e+ache+f%E1cil&SMacrorregi%E3o_PNDR=
    TODAS_AS_CATEGORIAS__&SAmaz%F4nia_Legal=TODAS_AS_CATEGORIAS__&
    SSemi%E1rido=TODAS_AS_CATEGORIAS__&SFaixa_de_Fronteira=
    TODAS_AS_CATEGORIAS__&SZona_de_Fronteira=TODAS_AS_CATEGORIAS__&
    SMunic%EDpio_de_extrema_pobreza=TODAS_AS_CATEGORIAS__&SSexo=2&
    pesqmes17=Digite+o+texto+e+ache+f%E1cil&SFaixa_Et%E1ria_1=3&
    SFaixa_Et%E1ria_1=4&SFaixa_Et%E1ria_1=5&SFaixa_Et%E1ria_1=6&
    SFaixa_Et%E1ria_1=7&pesqmes18=Digite+o+texto+e+ache+f%E1cil&
    SFaixa_Et%E1ria_2=TODAS_AS_CATEGORIAS__&pesqmes19=Digite+o+
    texto+e+ache+f%E1cil&SIdade_simples=TODAS_AS_CATEGORIAS__&
    formato=prn&mostre=Mostra'"

    # Busca a população feminina no tabnet
    raw_data = subprocess.check_output(command, shell=True)
    # Pega o que está dentro da tag <PRE></PRE>
    pro_data_1 = re.split('PRE>', raw_data.decode("iso-8859-1"))
    if len(pro_data_1) > 1:
        pro_data_2 = pro_data_1[1]
    else:
        pro_data_2 = ''

    # Cria uma lista pela separação de 'Enters'
    pro_data_3 = pro_data_2.split('\n')
    # Considera apenas a partir da segunda linha
    # e desconsidera as três últimas linhas
    pro_data_4 = pro_data_3[2:-3]
    pro_data_5 = []
    line = ''
    for i in pro_data_4:

```

```

line = i
# Remove \r
line = line.split('\r')[0]
# Cada coluna é separável por um "
col = line.split('"')
# Pega o código do município na segunda coluna
cod_mun_res = re.search('[0-9]{6}', col[1])
if cod_mun_res != None:
    col[1] = cod_mun_res.group()
# Volta a linha como texto
line = ''.join(col)
# Onde tem - pode ser considerado como 0
line = line.replace('-', '0')
# Considera a linha como uma lista, antes separada por ';'
line = line.split(";")
# Insere o ano que referente como segunda coluna
line.insert(1, '20'+year)
# Insere a linha como lista, exceto a última coluna
pro_data_5.append(line[:-1])
# na lista a ser devolvida
return pro_data_5

# Função que constroi o dataframe base com a tupla CODMUNRES, ANO
def get_tb_municipios():
    # Anos a serem considerados como base
    ano_df = pd.DataFrame([2012,2013,2014,2015,2016,
        2017,2018,2019,2020], columns= ['ano_nasc'])
    # Tabela com dados compilados dos municípios brasileiros
    fetch_mun_df = pd.read_excel("./tb_municipios.xlsx")
    mun_df = pd.DataFrame([], columns = ['CODIGO_MUNICIPIO'])
    # Código do município de 7 dígitos
    mun_df['CODIGO_MUNICIPIO'] = fetch_mun_df[
        'codigo_municipio_completo']
    b_df = pd.merge(mun_df.assign(key=1),
        ano_df.assign(key=1),on='key').drop('key', axis=1)
    b_df['CODIGO_MUNICIPIO'] = b_df['CODIGO_MUNICIPIO'].astype(str)
    # CODMUNRES possui 6 dígitos, por isso retirar o último dígito
    b_df['CODMUNRES'] = b_df['CODIGO_MUNICIPIO'].replace(
        r'.$', '', regex=True)
    # Ordena o dataframe primeiro por código depois por ano
    b_df = b_df.sort_values(by=['CODMUNRES', 'ano_nasc'])
    return b_df

```

Código 7 – Ajustes dos indicadores de “deslocamento ao parto”

```

get_ipython().run_line_magic('matplotlib', 'inline')
import requests
import pandas as pd

```

```

import os
import json
import sys
import matplotlib.pyplot as plt

# Funções auxiliares
def dot_print():
    sys.stdout.write('.')
    sys.stdout.flush()

# Dataframe base de Código do município e Ano
ano_df = pd.DataFrame([2012,2013,2014,2015,2016,2017,2018,2019,2020],
    columns= ['ANO'])
fetch_mun_df = pd.read_excel("../..//csv_files/tb_municipios.xlsx")
mun_df = pd.DataFrame([], columns = ['CODIGO_MUNICIPIO'])
mun_df['CODIGO_MUNICIPIO'] = fetch_mun_df['codigo_municipio_completo']
base_df = pd.merge(mun_df.assign(key=1), ano_df.assign(key=1), on='key').
    drop('key', axis=1)
base_df['CODIGO_MUNICIPIO'] = base_df['CODIGO_MUNICIPIO'].astype(str)
base_df['CODMUNRES'] = base_df['CODIGO_MUNICIPIO'].replace(r'.$', '',
    regex=True)
base_df = base_df.sort_values(by=['CODMUNRES', 'ANO'])
base_df.head()

# Tratamento das planilhas de deslocamento para o parto
df_mun_array = []
df_uf_array = []
for i in range(9):
    dot_print()
    df_mun_aux = pd.read_excel("./BR"+str(12+i)+"dc_Rosa.ods")
    df_mun_aux.insert(1, 'ANO', [2012+i] * len(df_mun_aux.index), True)
    df_mun_array.append(df_mun_aux)
    dot_print()
    df_uf_aux = pd.read_excel("./Brasil-UFs-12a20_Rosa.ods", sheet_name="
        BR"+str(12+i))
    df_uf_aux.insert(1, 'ANO', [2012+i] * len(df_uf_aux.index), True)
    df_uf_array.append(df_uf_aux)
print(' (+) ')

df_mun = pd.concat(df_mun_array, ignore_index=True)
df_uf = pd.concat(df_uf_array, ignore_index=True)

df_uf2 = df_uf.dropna(subset=['NOME'])
df_uf2 = df_uf2.fillna({'CODUFR': 55})
df_uf2 = df_uf2.fillna(0)
df_uf2 = df_uf2.sort_values(by=['CODUFR', 'ANO'])

```

```

df_uf2 = df_uf2.rename(columns={"CODUFR": "UF"})
#df_uf2.head()
df_uf2.to_csv('./indicadores_bloco4_deslocamento_parto_UF_2012-2020.csv',
              ,index=False)

df_mun2 = df_mun.sort_values(by=['CODMUNRES', 'ANO'])
df_mun2 = df_mun2.iloc[: -7, :]
del df_mun2['NOME']
base_df['CODMUNRES'] = base_df['CODMUNRES'].astype(int)
df_mun2['CODMUNRES'] = df_mun2['CODMUNRES'].astype(int)
df_mun3 = pd.merge(base_df, df_mun2, how='left', on=['CODMUNRES', 'ANO'])
del df_mun3['CODIGO_MUNICIPIO']
df_mun3 = df_mun3.sort_values(by=['CODMUNRES', 'ANO'])
#df_mun3.head()
df_mun3.to_csv('./indicadores_bloco4_deslocamento_parto_municipio_2012-2020.csv', index=False)

df_void = pd.read_excel('./Incompletitude_indicadores_deslocamento_parto.xlsx', sheet_name='BD')

df_void1 = df_void.rename(columns={"ano": "ANO", "uf": "UF", "mun_res": "CODMUNRES"})
df_void1['CODMUNRES'] = df_void1['CODMUNRES'].astype(int)
df_void2 = pd.merge(base_df, df_void1, how='left', on=['CODMUNRES', 'ANO'])
df_void2 = df_void2.sort_values(by=['CODMUNRES', 'ANO'])
df_void2 = df_void2.fillna(0)
del df_void2['CODIGO_MUNICIPIO']

df_void2.to_csv('./Incompletitude_indicadores_deslocamento_parto.csv',
               index=False)

```

ANEXOS

ANEXO A – LISTA DE INDICADORES PRELIMINARES

Tabela 1 – Lista de indicadores preliminares

Indicador	Fonte do dado
Proporção de nascidos vivos de mulheres com idade inferior a 20 anos (gestação na adolescência)	SINASC, período 2012 a 2019.
Proporção de nascidos vivos de mulheres com idade igual ou superior a 35 anos	SINASC, período 2012 a 2019.
Proporção de nascidos vivos de mulheres de cor da pele preta	SINASC, período 2012 a 2019.
Proporção de nascidos vivos de mulheres de cor da pele preta ou parda	SINASC, período 2012 a 2019.
Proporção de nascidos vivos de mulheres indígenas	SINASC, período 2012 a 2019.
Proporção de mulheres que estavam tendo seu primeiro parto	SINASC, período 2012 a 2019.
Proporção de mulheres com mais de 3 partos anteriores	SINASC, período 2012 a 2019.
Proporção de mulheres com baixa escolaridade (até 4ª série)	SINASC, período 2012 a 2019.
Proporção de mulheres com baixa escolaridade (até 7ª série)	SINASC, período 2012 a 2019.
Proporção de gestações múltiplas	SINASC, período 2012 a 2019.
Proporção de mulheres com início precoce do pré-natal (até o 3º mês de gestação)	SINASC, período 2012 a 2019.
Proporção de mulheres com número adequado de consultas de pré-natal segundo idade gestacional no parto	SINASC, período 2012 a 2019.
Proporção de nascimentos por cesariana	SINASC, período 2012 a 2019.
Proporção de nascidos vivos do grupo 2 de Robson	SINASC, período 2012 a 2019.

Proporção de nascidos vivos do grupo 5 de Robson	SINASC, período 2012 a 2019.
Proporção de baixo peso ao nascer	SINASC, período 2012 a 2019.
Proporção de nascimentos prematuros	SINASC, período 2012 a 2019.
Proporção de nascimentos termo precoce	SINASC, período 2012 a 2019.
Razão de Mortalidade Materna	SIM 2012 a 2019; SINASC 2012 a 2019.
Razão de Mortalidade Materna Direta	SIM 2012 a 2019; SINASC 2012 a 2019.
Taxa de mortalidade fetal	SIM 2012 a 2019; SINASC 2012 a 2019.
Taxa de mortalidade neonatal precoce	SIM 2012 a 2019; SINASC 2012 a 2019.
Proporção de óbitos de mulher em idade fértil investigados	SIM.
Proporção de óbitos maternos declarados investigados	SIM.
Incidência de sífilis congênita	SINAN 2012-2019; SINASC 2012-2019.
Taxa de fecundidade total	SINASC 2012 a 2019; projeção população por município (disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?popsvs/cnv/popbr.def)
Taxa de fecundidade específica em mulheres menores de 20 anos	SINASC 2012 a 2019; projeção população por município (disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?popsvs/cnv/popbr.def)
Taxa de fecundidade específica em mulheres de 20 a 34 anos	SINASC 2012 a 2019; projeção população por município (disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?popsvs/cnv/popbr.def)
Taxa de fecundidade específica em mulheres com 35 ou mais anos	SINASC 2012 a 2019; projeção população por município (disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?popsvs/cnv/popbr.def)
Cobertura populacional com equipes de saúde da família	https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCoberturaAB.xhtml

Cobertura populacional com equipes de Atenção básica	https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCoberturaAB.xhtml
Proporção de partos SUS	SINASC 2015-2019; microdados da ANS 2015-2019
Taxa de abortos inseguros por 1000 mulheres em idade fértil	SIH/SUS 2012-2019; Tabnet ANS 2012-2019; População residente de mulheres de 10-49 anos no período 2012-2019 (disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?popsvs/cnv/popbr.def)
Taxa de abortos inseguros por 100 nascidos vivos	SIH/SUS 2012-2019; Tabnet ANS 2012-2019; SINASC 2012-2019

ANEXO B – QUANTIDADE DE MUNICÍPIOS POR FAIXAS DE NASCIDOS VIVOS

Tabela 2 – Quantidade de municípios por faixas de nascidos vivos

Municípios	Faixa de Nascidos Vivos
11	> 20.000
13	10.000 - 20.000
52	5.000 - 9.999
403	1.000 - 4.999
499	500 - 999
4596	< 500

Fonte: Tabnet Datasus ano 2019

ANEXO C – LISTA DE INDICADORES DECIDIDOS

Tabela 3 – Lista de indicadores decididos

Indicador	Fonte de dados
Proporção de nascidos vivos de mulheres com idade inferior a 20 anos (gestação na adolescência)	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com idade de 20 a 34 anos	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com idade de 35 ou mais anos	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres brancas	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres pretas	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres pardas	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres amarelas	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres indígenas	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com menos de 4 anos de estudo	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com 4 a 7 anos de estudo	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com 8 a 11 anos de estudo	SINASC 2012 - 2020
Proporção de nascidos vivos de mulheres com mais de 11 anos de estudo	SINASC 2012 - 2020
Proporção mulheres 10 a 49 anos usuárias exclusivas do SUS	Usuárias planos de saúde: Tabnet ANS (www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_02.def); População feminina de 10 a 49 anos: disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popsvsbr.def

Cobertura populacional com equipes de saúde da família	SIAB 2012-2020 (https://egestorab.saude.gov.br/paginas/acessoPublico/relatorios/relHistoricoCoberturaAB.xhtml)
Taxa de fertilidade específica em mulheres menores de 20 anos	SINASC (nascidos de mulheres < 20 anos), Tabnet DataSUS (estimativa populacional mulheres <20 anos, disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popsvsbr.def)
Proporção de nascidos vivos de mulheres com mais de 3 partos anteriores	SINASC
Taxa de abortos inseguros por 1000 mulheres em idade fértil	SIH (internações por aborto), Tabnet DataSUS (estimativa populacional mulheres de 10 a 49 anos, disponível em http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popsvsbr.def), Mulheres de 10 a 49 anos usuárias de planos de saúde médicos (Tabnet ANS, disponível em www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_02.def)
Taxa de abortos inseguros por 100 nascidos vivos	SIH (número de internações por aborto, CID O03 a O08), SINASC (número total de nascidos vivos), Mulheres de 10 a 49 anos usuárias de planos de saúde médicos (Tabnet ANS, disponível em www.ans.gov.br/anstabnet/cgi-bin/dh?dados/tabnet_02.def)
Cobertura de assistência Pré-natal	SINASC
Proporção de mulheres com início precoce do pré-natal (até o 3º mês de gestação)	SINASC
Proporção de mulheres com mais de sete consultas de pré-natal	SINASC
Incidência de sífilis congênita	SINASC e MS (casos de sífilis congênita, disponíveis em http://indicadoressifilis.aids.gov.br)
Proporção de nascimentos por cesariana	SINASC

Proporção de nascidos vivos do grupo 1 de Robson	SINASC
Proporção de nascidos vivos do grupo 2 de Robson	SINASC
Proporção de nascidos vivos do grupo 3 de Robson	SINASC
Proporção de nascidos vivos do grupo 4 de Robson	SINASC
Proporção de nascidos vivos do grupo 5 de Robson	SINASC
Proporção de nascidos vivos do grupo 6 a 9 de Robson	SINASC
Proporção de nascidos vivos do grupo 10 de Robson	SINASC
Proporção de cesarianas no grupo 1 de Robson	SINASC
Proporção de cesarianas no grupo 2 de Robson	SINASC
Proporção de cesarianas no grupo 3 de Robson	SINASC
Proporção de cesarianas no grupo 4 de Robson	SINASC
Proporção de cesarianas no grupo 5 de Robson	SINASC
Proporção de cesarianas no grupo 6 a 9 de Robson	SINASC
Proporção de cesarianas no grupo 10 de Robson	SINASC
Contribuição relativa do grupo 1 de Robson para a taxa global de cesariana	SINASC
Contribuição relativa do grupo 2 de Robson para taxa global de cesariana	SINASC
Contribuição relativa do grupo 3 de Robson para taxa global de cesariana	SINASC

Contribuição relativa do grupo 4 de Robson para taxa global de cesariana	SINASC
Contribuição relativa do grupo 5 de Robson para taxa global de cesariana	SINASC
Contribuição relativa do grupo 6 a 9 de Robson para taxa global de cesariana	SINASC
Contribuição relativa do grupo 10 de Robson para taxa global de cesariana	SINASC
Proporção de baixo peso ao nascer	SINASC
Proporção de nascimentos prematuros	SINASC
Proporção de nascimentos termo precoce	SINASC