

Trabalho Final- Análise em bases do DATASUS

André Bergmann Lisboa e Lucas Cardoso Azevedo

Descrição da Investigação

Para o trabalho foi escolhida a análise de tipos de parto (Cesáreo e Vaginal) para os anos de 2018 a 2021 nos estados da Região Sul do Brasil (RS, PR e SC), com o objetivo de verificar relações com o perfil das mães e analisar o comportamento e a evolução ao longo do período escolhido. Para isso foi utilizada a base SINASC (Sistema de Informação sobre Nascidos Vivos) presente na biblioteca *pySUS*, que simplificou a obtenção dos dados diretamente em *Python*. A questão escolhida teve como ponto de partida a relação entre parto cesáreo e peso do recém-nascido, conforme sugerido na especificação, porém foram incluídas outras variáveis que possibilitaram a observação de padrões de comportamento e insights interessantes durante a análise exploratória dos dados.

A análise consistiu na obtenção, tratamento, limpeza e transformação em *Python*, utilizando *Jupyter Notebook*, e na elaboração de um dashboard com *Power BI* com as visualizações das relações observadas e insights obtidos. Todo o material está aberto e disponível em:

- https://github.com/LucasAzevedo/analise_partos_datasus

Integração, Limpeza e Transformação

As bibliotecas utilizadas foram: *Pandas* (versão 1.5.3 para compatibilidade com o *pySUS*), *Numpy*, *datetime*, *pathlib* e os módulos *sinasc* e *parquets_to_dataframe* da biblioteca *pySUS* para a obtenção dos dados e importação em *dataframe*.

Na importação dos dados seguiu-se a documentação da biblioteca [1], primeiramente realizando o download dos arquivos *parquet* dos estados e anos escolhidos com o módulo *sinasc* e posteriormente a leitura em *dataframe* com o método *parquets_to_dataframe* e concatenação dos dados.

Na escolha das colunas foi levado em consideração o percentual de valores não nulos da coluna (acima de 98%) bem como a pertinência da informação em relação ao estudo, verificando junto ao dicionário da base SINASC [2]. Desta forma as colunas escolhidas foram:

- CODMUNNASC: Código do município de Nascimento
- IDADEMAE: Idade da mãe
- ESTCIVMAE: Estado Civil da mãe
- GESTACAO: Semanas de gestação
- GRAVIDEZ: Tipo de Gravidez
- PARTO: Tipo de Parto
- CONSULTAS: Número de consultas de pré-natal
- DTNASCMAE: Data de nascimento da mãe
- DTNASC: Data de nascimento do nascido
- HORANASC: Hora de nascimento do nascido
- RACACOR: Tipo de raça e cor do nascido
- PESO: Peso do nascido
- ESCMAE2010: Escolaridade da mãe (padrão 2010)
- RACACORMAE: Tipo de raça e cor da mãe

O tratamento foi realizado em 3 etapas:

1. Transformação das colunas numéricas em inteiro: IDADEMAE e PESO;
2. Mapeamento das colunas categóricas para identificadores textuais: CODMUNNASC, ESTCIVMAE, GESTACAO, GRAVIDEZ, PARTO, CONSULTAS, RACACOR, ESCMAE2010 e RACACORMAE
3. Tratamento e transformação dos casos especiais de data e hora: DTNASCMAE, DTNASC e HORANASC. Criação das colunas ANONASC e MESNASC, com o dado do ano e mês de nascimento, respectivamente.

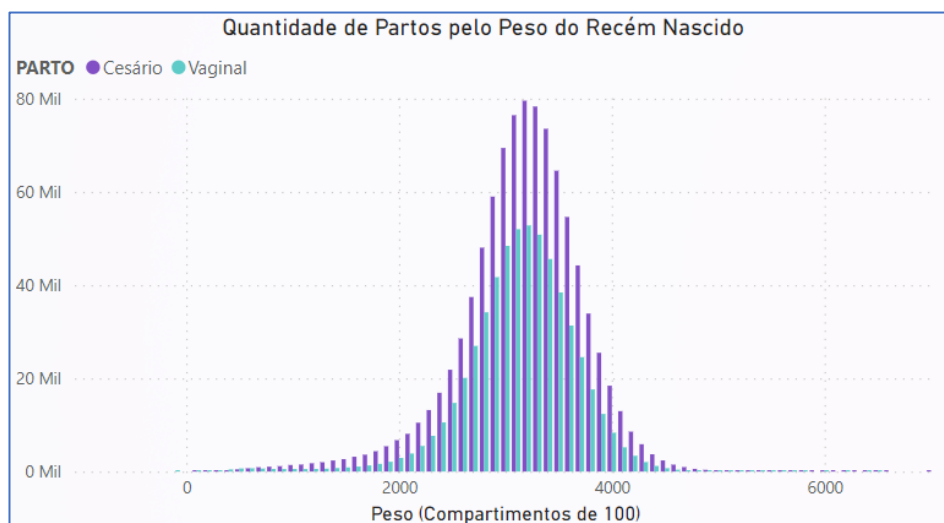
Na etapa 2, foi utilizado como referência o dicionário de dados da tabela SINASC em todas as colunas com exceção da CODMUNNASC, em que utilizou-se da base DTB (Divisão Territorial Brasileira) do IBGE [3], para a obtenção do nome e UF do município através do seu código de 6 dígitos.

Por fim, foi realizado a remoção de registros com tipo de parto diferente de Cesáreo e Vaginal, visto que o estudo trata da comparação entre os dois; de idades acima de 65 anos, identificadas como outliers e não confiáveis; e de registros em cidades fora dos estados alvo do estudo (validação feita após o cruzamento com o DTB do IBGE). Após todos os tratamentos o *dataframe* final resultou em 1.517.140 linhas e 18 colunas.

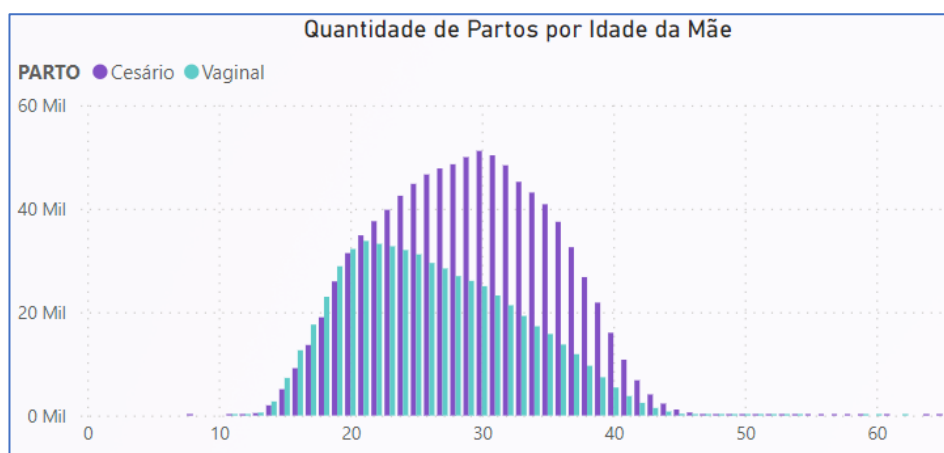
Dashboard e Resultados

Em seguida ao pré-processamento dos dados foi feita a exportação em formato *csv* para que se pudesse importar no *Power BI Desktop* e elaborar as visualizações e análise dos dados. No *Power BI* não foi necessária nenhuma operação nos dados, visto que estes já haviam sido pré-processados e a ferramenta já identifica os tipos de dados na importação. Para as visualizações foram criadas duas páginas, uma com o objetivo de analisar os tipos de parto pelo perfil da mãe, e outra para uma análise evolutiva no período estudado.

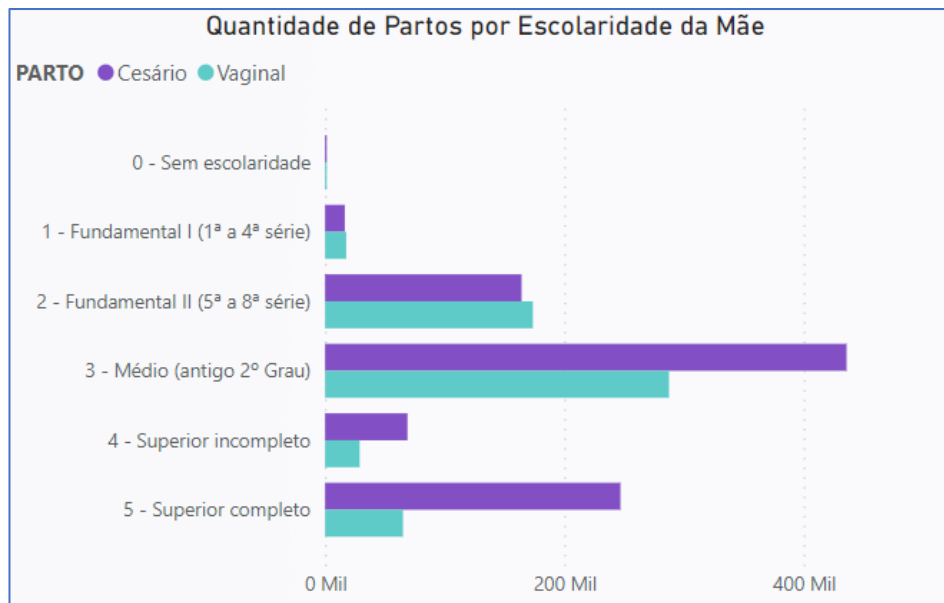
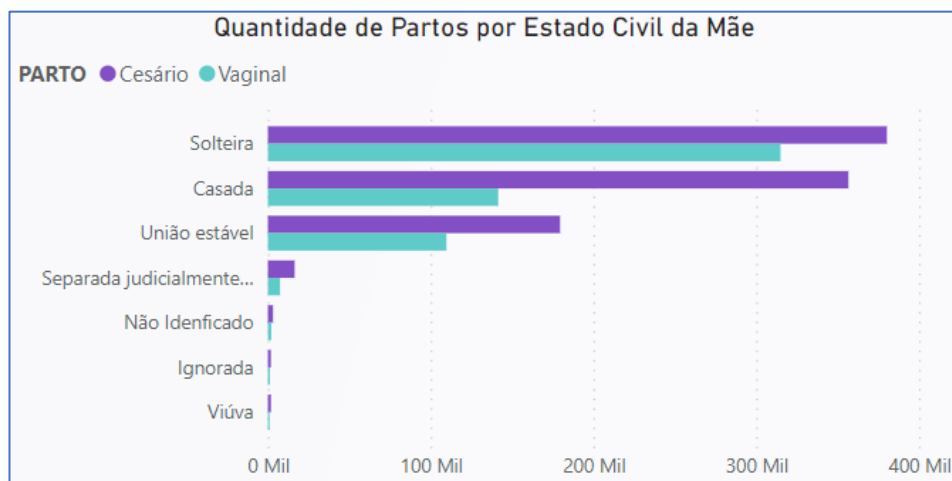
A primeira investigação foi se havia relação entre partos cesáreos e o peso do nascido, e como pode ser visto abaixo, não há. Com o histograma podemos verificar que a distribuição no peso é exatamente a mesma para partos cesáreos e vaginais, apenas alterando pela quantidade total de cada tipo.



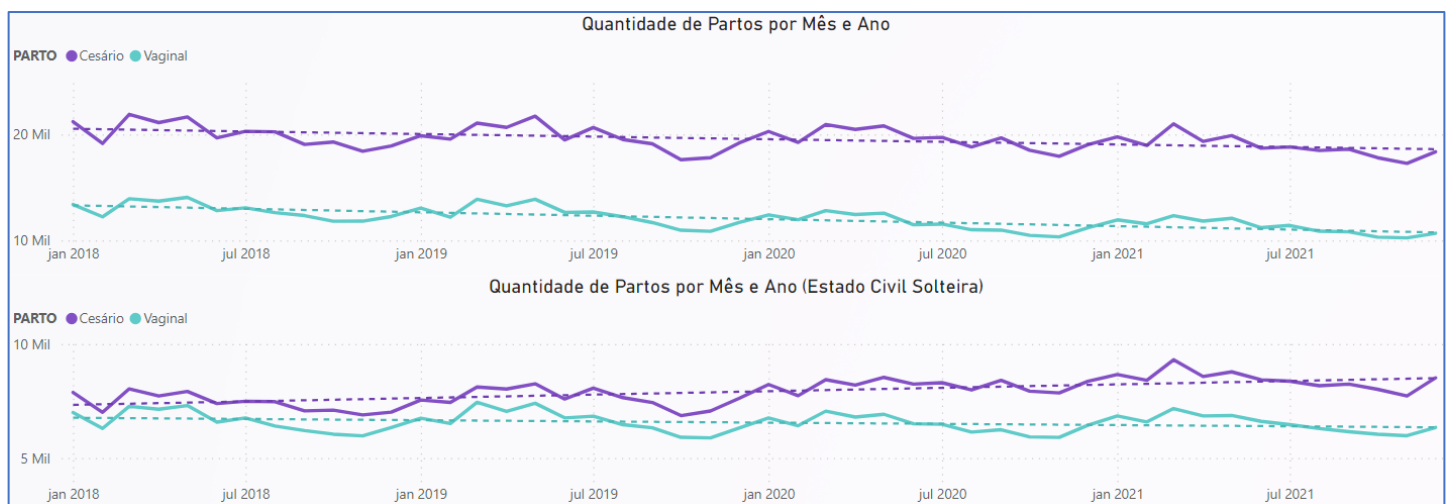
Por outro lado, foi possível avaliar que há uma relação com idade, em que o parto cesáreo tenda a ocorrer mais em idades mais altas enquanto o parto vaginal ocorre mais em idades menores, chegando a ser maior entre 15 e 20 anos.

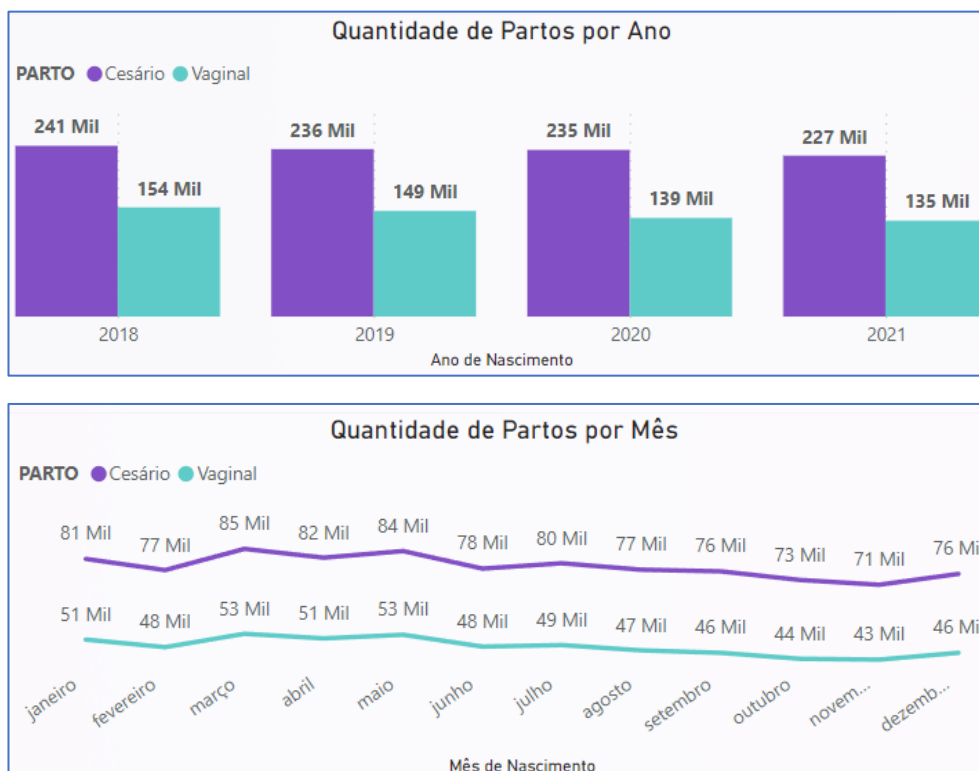


Outros fatores interessantes no comparativo entre os tipos de partos foram vistos quando analisado por Estado Civil e Escolaridade da Mãe. Por Estado Civil fica claro quando observado no gráfico que a maior diferença entre os tipos ocorre em mães casadas e em união estável do que em mãe solteiras, onde a dominância do cesáreo é menor. Por Escolaridade da Mãe, observa-se que na medida que aumenta a escolaridade, o parto cesáreo se distancia mais do vaginal, percentualmente.



Já na análise evolutiva foi observada uma diferença constante ao longo do período, com exceção do caso das mães solteiras que, mesmo sendo o estado civil com a menor diferença entre os tipos de parto, esta diferença vem aumentando no período, conforme realçado abaixo com as linhas de tendência. Além disso, nota-se uma certa sazonalidade nos partos, com uma alta nos meses de março, abril e maio, e uma baixa nos meses de setembro, outubro e novembro.





Conclusões

Com base na investigação realizada sobre os tipos de parto (Cesáreo e Vaginal) nos estados da Região Sul do Brasil no período de 2018 a 2021, diversas conclusões podem ser destacadas.

1. Relação entre Parto Cesáreo e Peso do Recém-Nascido:

Contrariamente à hipótese inicial de que haveria uma relação entre o tipo de parto e o peso do recém-nascido, a análise revelou que não existe uma correlação significativa. A distribuição de pesos para partos cesáreos e vaginais é semelhante, indicando que o peso do bebê não é um fator determinante na escolha do tipo de parto.

2. Influência da Idade da Mãe:

Foi identificada uma relação entre a idade da mãe e o tipo de parto. Cesáreas tendem a ocorrer mais frequentemente em mães de idade mais avançada, enquanto partos vaginais são mais comuns em mães mais jovens, especialmente na faixa etária de 15 a 20 anos.

3. Impacto do Estado Civil e Escolaridade da Mãe:

A análise por estado civil evidenciou que a diferença entre os tipos de parto é mais pronunciada em mães casadas e em união estável do que em mães solteiras. Quanto à escolaridade da mãe, percebeu-se que o percentual de partos cesáreos aumenta à medida que a escolaridade materna cresce.

4. Análise Evolutiva ao Longo do Período:

Ao longo dos anos de 2018 a 2021, observou-se uma diferença consistente nos tipos de parto, com algumas variações notáveis. O aumento da diferença entre os tipos de parto em mães solteiras foi destacado, indicando uma mudança nesse grupo específico ao longo do tempo. Além disso, a análise temporal revelou padrões sazonais nos nascimentos, com picos nos meses de março, abril e maio, e quedas em setembro, outubro e novembro.

5. Ferramentas Utilizadas:

O estudo demonstrou a eficácia do uso de ferramentas como Python, Jupyter Notebook e Power BI para a obtenção, tratamento e visualização de dados. A transparência e compartilhamento do código-fonte no GitHub proporcionam uma base sólida para a replicabilidade e revisão do estudo por outros profissionais e pesquisadores.

Em suma, a análise abrangente dos dados de partos na Região Sul do Brasil revelou não apenas informações sobre a relação entre o tipo de parto e características como idade e escolaridade da mãe, mas também insights valiosos sobre padrões temporais e diferenças evolutivas ao longo dos anos. Essas conclusões podem fornecer uma base para discussões mais aprofundadas e possíveis intervenções na área da saúde materna.

Referências

[1] Documentação PySUS

<https://pysus.readthedocs.io/en/latest/>

[2] Dicionário de dados SINASC

https://svs.aids.gov.br/daent/cgiae/sinasc/documentacao/dicionario_de_dados_SINASC_tabela_DN.pdf

[3] DTB - Divisão Territorial Brasileira (IBGE)

<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/23701-divisao-territorial-brasileira.html?=&t=downloads>