

Trabalho Final - Coleta, preparação e análise de dados

Guilherme Zamin Scherer, Leonardo Fronza de Miranda e Pedro Barcelos
Ciência de Dados e IA — PUCRS

<https://github.com/leomiranda16/cpatf.git>

8 de julho de 2024

Resumo

Este artigo descreve o processo de resolução do trabalho final realizado no 3º semestre do curso de Ciência de Dados e Inteligência Artificial da PUCRS durante a cadeira de Coleta, Preparação e Análise de Dados, que trata da escolha de bases de dados abertos do *DATASUS*, realização de etapas de pré-processamento e criação da visualização dos dados.

A seguir, o relatório descreve as bases de dados escolhidas, técnicas de limpeza e transformação dos dados, métodos de visualização dos mesmos, e, por fim, a análise realizada.

Coleta

Após a análise dos *DataSets* disponíveis no site *DATASUS*, foi pensado nos temas mais interessantes a serem discutidos e analisados com o auxílio das ferramentas disponibilizadas ao longo do semestre. Foi discutido muito ao longo dos meses quais seriam os temas do grupo para este trabalho, dentre eles, os mais notáveis foram: Tuberculose, acidentes auto infligidos na pandemia e após a mesma e discrepância entre os pesos dos fetos nascidos através do processo de cesária ou vaginal. No fim, o tema escolhido foi o último.

Após a escolha do tema, o grupo seguiu adiante para selecionar os dados pertinentes e começar o processo de processamento e limpeza dos dados. É interessante mencionar que todos os dados foram retirados da mesma base de dados, para evitar ambiguidade e discrepância dos dados.

Finalmente, foram baixadas um total de 5 tabelas distintas para a realização da análise e também, foi levado em conta 6 tipos de atributos: idade da mãe (2 intervalos), tempo de gestação do feto, peso do feto, tipo de parto e o local de ocorrência do parto.

Preparação

Após a escolha das bases de dados, foi feita uma breve limpeza nos mesmos. Com a utilização da biblioteca *Pandas*, vista nas aulas da disciplina, foi utilizado o comando *dropna()* a fim de remover linhas onde haviam dados de valor nulo, para não ocorrer conflitos enviando os arquivos para o *PowerBI*.

Após a retirada desses valores *NaN*, abrimos os arquivos no Excel e excluímos a linha ‘Ignorados’, pois, obviamente, eram dados que não nos interessavam. Como os *datasets* tinham uma linha de totais, fizemos alterações nesses valores, pois após a retirada de uma linha, a conta dos totais não fecharia. Para atualizar esses valores, foi utilizada a função *SOMA=*, no próprio *Excel*.

A fim de obter uma análise mais aprofundada sobre o tema em questão, resolvemos concatenar o dataset do tempo de gestação dos fetos com o dataset base, no qual as colunas são apenas cesária e vaginal. Com essa concatenação, conseguimos uma melhor visualização dos dados e um novo parâmetro para ser levado em questão na hora da análise.

Análise

Após a montagem dos *dashboards* utilizando a ferramenta *Power BI*, chegamos a algumas conclusões interessantes sobre os dados. Primeiramente, geramos dois gráficos de proporção, um referente às mulheres mais novas (menores de 10 anos até 29 anos) e o outro às mulheres mais velhas (mais de 29 anos) que tiveram o parto. Analisando essas proporções, vimos claramente que os partos normais eram mais comuns nas mulheres mais novas e os partos por cesárea apareciam mais frequentemente nas mulheres mais velhas (provavelmente por já ser o segundo filho).

Mesmo que esses dados mostrem essas informações, o nosso objetivo era verificar se existe relação entre o tipo de parto e o peso dos bebês, e somente com esses gráficos de proporção, não conseguimos chegar a nenhum tipo de conclusão a respeito da existência de alguma relação entre esses aspectos. Ainda estávamos curiosos, a fim de buscar mais parâmetros que pudessem apontar algum indício para influenciar no peso das crianças, então pensamos que o local do parto pudesse ter alguma relevância para nossa análise. Como a alimentação da mãe afeta diretamente no peso do filho, notamos que partos em aldeias indígenas ou domiciliares poderiam ser um ótimo parâmetro para nos auxiliar na visualização dos dados discrepantes, pois a alimentação e condições de saúde básica seriam divergentes à maioria dos partos hospitalares.

Gerada a visualização do gráfico dos locais de ocorrência, percebemos, novamente, outro parâmetro inconclusivo para o peso dos bebês, as proporções entre as distribuições dos pesos eram muito semelhantes entre ambos os locais analisados.

Por fim, recorremos às informações do tempo de gestação em relação ao peso das crianças. No primeiro momento, não estávamos convictos que essa comparação seria frutífera, porém, fomos surpreendidos com os resultados obtidos. Logo, notamos que os fetos com o desenvolvimento menor que 28 semanas eram extremamente magros, pesando menos de 1500g. Cerca de 200.000 bebês que levaram de 22 a 27 semanas de gestação, se enquadram na categoria extremamente prematuro e têm uma média de 750g, extremamente baixa quando comparada às outras categorias analisadas.

Concluímos, finalmente, que apenas com os dados disponibilizados pelo *DATASUS*, não foi possível encontrar nenhuma relação entre o tipo de parto/idade materna/local de

ocorrência e o peso dos bebês, porém o tempo de gestação é um fator determinante, em alguns casos. Após breves pesquisas sobre o assunto, notamos que os principais aspectos que contribuem para o tema em questão são o estado nutricional pré-gestacional e alimentação materna durante a gravidez, dados que, infelizmente, não são possíveis encontrar na fonte de dados utilizada.

Referências

- [1] Informações de saúde (TABNET) - DATASUS. DATASUS, 2024. Disponível em: <https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>. Acesso em: 18 jun. 2024.
- [2] PANDAS - Python Analysis Library. PANDAS, 2024. Disponível em: <https://pandas.pydata.org/>. Acesso em: 28 jun. 2024.
- [3] MONTENEGRO, Bruna. Power BI: o que é, para que é usado e como funciona?. EBAC, 2023. Disponível em: <https://ebaonline.com.br/blog/o-que-e-power-bi>. Acesso em: 05 jul. 2024.
- [4] Prematuridade - Nota técnica 2019. Sociedade Brasileira de Pediatria, 2019. Disponível em: https://www.sbp.com.br/fileadmin/user_upload/Nota_Tecnica_2019_Prematuridade.pdf. Acesso em: 05 jul. 2024.
- [5] TOURINHO, Amanda Braga; REIS, Lílían Barros De Sousa Moreira. Peso ao Nascer: Uma Abordagem Nutricional. Artigo de Revisão (PDF), 2012. Disponível em: https://bvsmms.saude.gov.br/bvs/periodicos/revista_ESCS_v23_n1_a02_peso_ao_nascer.pdf. Acesso em: 05 jul. 2024.