**Projeto Final: Entrevistando dados**

Integrantes do grupo:

| * Gabriela Bertolo * Melissa Duarte * Michelly Neris * Nicole Lacerda * Paulo Fehlauer |
| --- |

O objetivo do projeto final é demonstrar conhecimentos aplicados de entrevista a bases de dados utilizando planilha eletrônica e/ou SQL. Para isso, você deverá seguir os passos abaixo:

1. Escolha ao menos uma base de dados pública para trabalhar. Justifique em um parágrafo sua escolha, incluindo também informações básicas sobre a base (fonte, metodologia de coleta, data de atualização e limitações mais evidentes).
   1. Escolhemos a [base de dados](https://pcdas.icict.fiocruz.br/conjunto-de-dados/sistema-de-informacao-sobre-nascidos-vivos/) do SINASC - Sistema de Informações sobre Nascidos Vivos, na versão tratada pela Plataforma de Ciência de Dados aplicada à Saúde (PCDaS), da Fiocruz. A escolha se deu pelo interesse do grupo no tema da natalidade e pelo fato de se tratar de uma base já tratada, padronizada e enriquecida pela equipe da Fiocruz. Os dados originais provêm do DATASUS e são obtidos a partir da coleta da “Declaração de Nascidos Vivos” (DN), documento padrão e de preenchimento obrigatório em todo o território nacional. O tratamento feito pelo PCDaS resulta em um dataset anual com todos os registros das declarações de nascidos vivos contidas no SINASC. Os dados disponíveis hoje no sistema compreendem os anos de 1996 a 2021 e as bases são entregues em formato .csv separadas por estado e ano. Neste trabalho, optamos por restringir a análise ao estado de São Paulo e aos anos de 2019 a 2021.
2. Crie um roteiro de entrevista para aplicar a essa base de dados.
   1. Como se distribuem os partos natural e cesáreo ao longo da semana?
   2. Há alguma correlação entre a proporção de cesáreas e os finais de semana?
   3. E entre a proporção de cesáreas e vésperas de feriados?
   4. Como a pandemia de Covid-19 afetou esse cenário?
3. Documente no github as funções e cálculos aplicados para obter as respostas.

EXECUTANDO TUDO EM 'estrutura da tabela'

Na linha 1:

SELECT \*

FROM SP2021

LIMIT 10

Resulto: 10 linhas retornadas em 79 ms

EXECUTANDO TUDO EM 'total nascidos'

Na linha 1:

SELECT DTNASC, COUNT(\*) AS total\_ocorrencias

FROM SP2021

GROUP BY DTNASC;

Resulto: 365 linhas retornadas em 2242 ms

EXECUTANDO TUDO EM 'SQL 3'

Na linha 1:

SELECT PARTO, DTNASC, COUNT(\*) AS total\_ocorrencias

FROM SP2021

GROUP BY PARTO, DTNASC;

Resulto: 798 linhas retornadas em 3170 ms

EXECUTANDO TUDO EM 'SQL 4'

Na linha 1:

SELECT

DTNASC,

SUM(CASE WHEN PARTO = 1 THEN 1 ELSE 0 END) AS Vaginal,

SUM(CASE WHEN PARTO = 2 THEN 1 ELSE 0 END) AS Cesarea,

SUM(CASE WHEN PARTO = 3 THEN 1 ELSE 0 END) AS Outros,

COUNT(\*) AS Total\_Ocorrencias

FROM

SP2021

GROUP BY

DTNASC;

Resulto: 365 linhas retornadas em 2575 ms

1. Indique um caminho inicial de pauta que poderia ser explorado a partir dos insights obtidos.
   1. Observamos uma correlação evidente entre a proporção de cesáreas em relação ao dia da semana. Ao longo dos anos, é consistente a queda na proporção de cesáreas nos sábados e domingos. Em 2021, por exemplo, apenas 2 domingos registraram número de cesáreas maior do que 50%, sendo que a média diária para o mesmo ano é de aproximadamente 57% de cesáreas.
   2. Não foi possível identificar correlação específica entre a proporção de cesáreas e vésperas de feriados. Um possível refinamento dessa pergunta poderia partir da diferenciação entre dados da rede pública e da rede privada.
   3. Tampouco foi possível identificar possíveis efeitos da pandemia sobre essas proporções, visto que as linhas dos 3 anos analisados (2019, 2020 e 2021) são bastante similares.
   4. Especialistas alegam que não é admitido fazer uma cesariana antes das 39 semanas de gestação - quando não tem recomendação médica. Mas o Brasil tem um elevado índice de cesarianas sem indicação. A ideia da nossa pauta seria mostrar quando se concentram essas cesáreas. Segundo médicos que conversamos, na saúde suplementar, principalmente, é muito comum um obstetra fazer 4 ou 5 cesáreas no mesmo dia, porque existe a possibilidade de agendamento. Isso sugere um possível comodismo entre os profissionais de saúde, que não respeitam o tempo natural do trabalho de parto e optam pela facilidade. Com isso, médicos marcam cesáreas antes do período certo e isso pode afetar a saúde do bebê e da mãe.

Indique abaixo o(s) link(s) para a documentação do projeto no github.

| *https://github.com/melissaduarte/mjd\_fundamentoseetica/blob/main/trabalhofinal.md* |
| --- |