

Disciplina: Bootcamp Análise de Sistemas e Banco de Dados;

Professor responsável: Victor H. B. Canto;

Grupo: 9;

Alunos: Diego K. Yoshida¹; Eduardo V. Santos²; Gabriel J. S. Zanata³; Henry C. Dias⁴; Luiz G. H. V. da Cruz⁵; Matheus S. Bitencourt⁶; Thomas H. S. Pinheiro⁷; Vinicius L. Santos⁸.

Desenvolvimento de um painel de BI: estatísticas sobre os casos de suicídio no Brasil

Bauru, 2024

-
- 1 diego.24100039ead@alunos.unisagrado.edu.br
 - 2 eduardo.24100480ead@alunos.unisagrado.edu.br
 - 3 gabriel.23100426ead@alunos.unisagrado.edu.br
 - 4 henry.24200191ead@alunos.unisagrado.edu.br
 - 5 luiz.24200213ead@alunos.unisagrado.edu.br
 - 6 matheus.24200068ead@alunos.unisagrado.edu.br
 - 7 thomas.24100016ead@alunos.unisagrado.edu.br
 - 8 vinicius.24100293ead@alunos.unisagrado.edu.br

1 BASE DE DADOS

A base de dados pública escolhida encontra-se disponível no Instituto de Pesquisa Econômica Aplicada (IPEA): Atlas da Violência

(<https://www.ipea.gov.br/atlasviolencia/filtros-series/16/suicidios>), reunindo dados estatísticos sobre os casos de suicídio no Brasil de 1980 a 2022. Ela está dividida em duas categorias principais: *quantidade* (número de casos) e *taxa* (*casos a cada 100 mil habitantes*). A partir destas duas categorias, são separadas com base na causa: *geral* (todas as causas que levaram ao falecimento) e *por meio de armas de fogo*. Cada uma dessas categorias acompanha suas respectivas divisões demográficas: *homens e mulheres adultos* e *homens e mulheres jovens* (no caso desta base de dados, os *jovens* são agrupados sem distinção de sexo). A partir das divisões apresentadas, cada uma apresenta um agrupamento de tabelas contendo os dados em si, representando um escopo geográfico: *país (Brasil)*, *regiões*, *UF*, *capitais* e *municípios*:

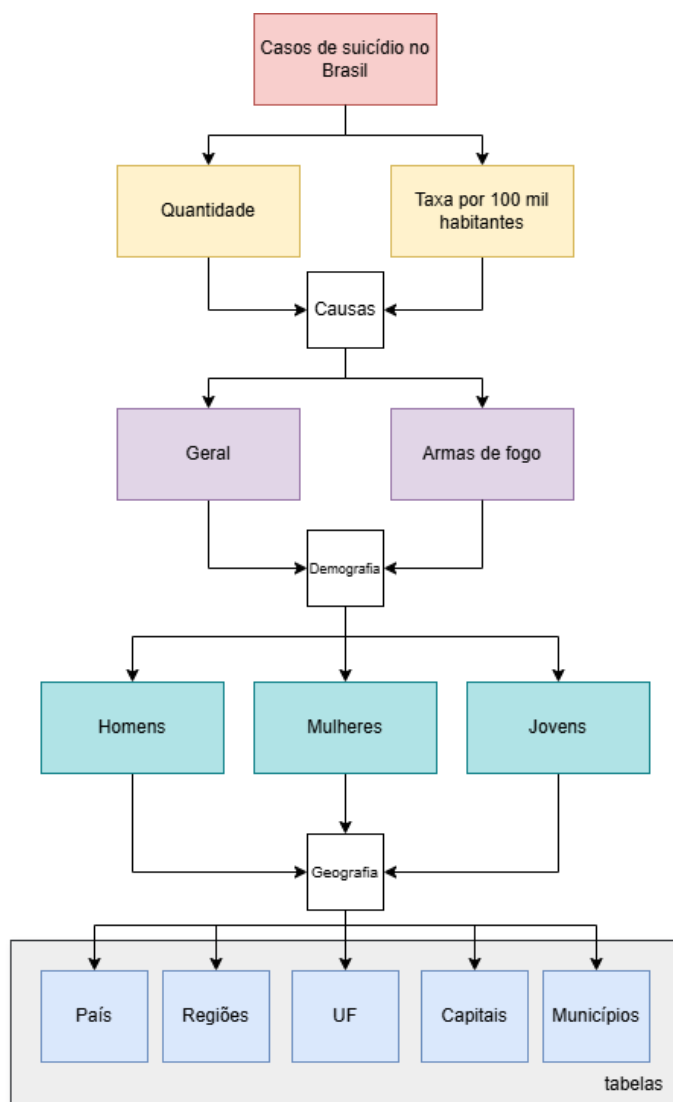


Fig. 1 – Agrupamento por especialização

1.1 DADOS DISPONÍVEIS

Todos os dados disponíveis estão no formato de tabelas em arquivo .csv (*valores separados por vírgula*), e seguem a mesma estrutura de colunas: *cod*, *nome*, *periodo* e *valor*. Apesar da padronização quanto ao nome das colunas nas tabelas, a coluna *valor* pode armazenar dois tipos de dados diferentes (números inteiros ou números com casas decimais) conforme o agrupamento. No caso, as tabelas do agrupamento *QUANTIDADE* utilizam o valor numérico inteiro, enquanto que em *TAXA* utilizam valores com casas decimais.

Nome do campo	Tipo de dado	Descrição
<i>cod</i>	INTEIRO	Identificação numérica única da localidade geográfica
<i>nome</i>	TEXTO	Localidade no escopo geográfico (nome de cidade, UF etc.)
<i>período</i>	INTEIRO	Ano ao qual os dados da linha se referem (de 1980 a 2022)
<i>valor</i> (quantidade)	INTEIRO	Número de casos totais
<i>valor</i> (taxa)	DECIMAL	Número de casos a cada 100 mil habitantes

Tabela 1 – Dicionário de dados representando todas as tabelas existentes, com a variação da coluna “valor”

No total, a base de dados conta com 75 tabelas, que somadas ocupam 77,3 MB de armazenamento.

A	B	C	D	E
<u>cod</u>	nome	<u>período</u>	<u>valor</u>	
3123205	Dores do Indaiá	1989	0	
3123304	Dores do Turvo	1989	0	
3123403	Doresópolis	1989	0	
3123502	Douradoquara	1989	0	
3123528	Durandé	1989	0	
3123601	Elói Mendes	1989	0	
3123700	Engenheiro Caldas	1989	0	
3123809	Engenheiro Navarro	1989	0	
3123858	Entre Folhas	1989	0	
3123908	Entre Rios de Minas	1989	0	
3124005	Ervália	1989	2	

Fig. 2 – Exemplo de visualização de uma das tabelas na ferramenta LibreOffice Calc

1.2 INCONSISTÊNCIAS IDENTIFICADAS

Ao investigarmos os dados contidos nas tabelas após o *download* a partir da fonte oficial, foi possível identificar a ausência de alguns dados. Todas as tabelas do escopo *capitais* encontram-se vazias, sendo necessário extrair os valores a partir das tabelas de *municípios* para completá-las. Além disso, a maior parte das tabelas possui dados no período a partir de 1989, diminuindo a abrangência real da base de dados.