

DATA SCIENCE

Kaique Maia Reis Silva

Descrição do Projeto e das Variáveis

Este projeto acadêmico visa analisar dados de óbitos por doenças respiratórias disponibilizados pelos Cartórios de Registro Civil do Brasil. O objetivo principal é fornecer uma compreensão detalhada e visual das características demográficas e geográficas dos óbitos, com foco nas variáveis de faixa etária, sexo, estados (UF) e tipo de doença. Esta análise é crucial para identificar padrões, tendências e possíveis desigualdades nos óbitos por doenças respiratórias no Brasil.

Descrição das Variáveis:

uf (Unidade da Federação):

- **Tipo:** Object
- **Descrição:** Representa a unidade da federação onde ocorreu o óbito.
- **Número de valores nulos:** 0

local_obito (Local do Óbito):

- **Tipo:** Object
- **Descrição:** Local do óbito.
- **Número de valores nulos:** 0

tipo_doenca (Tipo de Doença):

- **Tipo:** Object
- **Descrição:** Indica o tipo de doença relacionada ao óbito.
- **Número de valores nulos:** 0

faixa_etaria (Faixa Etária):

- **Tipo:** Object
- **Descrição:** Faixa etária do indivíduo falecido.
- **Número de valores nulos:** 0

sexo (Sexo):

- **Tipo:** Object
- **Descrição:** Sexo do indivíduo falecido.
- **Número de valores nulos:** 0

total (Total):

- **Tipo:** Numérica (int64)
- **Descrição:** Número total de óbitos registrados para a combinação específica de variáveis.
- **Número de valores nulos:** 0

Estatísticas Descritivas

total:

- count 11282.000000
- mean 53.074012
- std 227.071283
- min 1.000000
- 25% 1.000000
- 50% 5.000000
- 75% 22.000000
- max 5991.000000

Count (total):

- **11282.000000:** Isso significa que há 11.282 registros de óbitos na base de dados.

Média (mean):

- **53.074012:** Em média, cada registro tem aproximadamente 53 óbitos.

Desvio Padrão (std):

- **227.071283:** O desvio padrão é uma medida da variabilidade dos dados em torno da média. Um valor de 227,071283 indica que há uma grande variação no número de óbitos registrados.

Quartis (25%, 50%, 75%) e min e max:

- **min:** 1.000000, significa que o menor número de óbitos em um registro é 1.
- **25% (primeiro quartil):** 1.000000, significa que 25% dos registros têm 1 óbito ou menos.
- **50% (mediana ou segundo quartil):** 5.000000, significa que 50% dos registros têm 5 óbitos ou menos.
- **75% (terceiro quartil):** 22.000000, significa que 75% dos registros têm 22 óbitos ou menos.
- **max:** 5991.000000, significa que o maior número de óbitos em um registro é 5991.

Descrição das Análises Realizadas

Gráfico de Distribuição de Óbitos por Estado:

SP e MG tem as maiores taxas de óbito isoladamente, com 763 e 696, respectivamente.

BA, CE, GO, MA, MT, PB, PE, PR, RJ, RN e RS tem uma taxa moderada-alta que varia de 407 a 555 óbitos.

AC, AL, AM, AP, DF, ES, MS, PI, RO, RR, SC, SE E TO tem uma taxa baixa-moderada que varia de 125 a 389 óbitos.

O Sudeste e o Nordeste são as regiões mais afetadas, com altas concentrações de óbitos, enquanto as regiões Sul e Centro-Oeste têm números moderados, com variações significativas entre os estados dentro dessas regiões. O Norte é a região menos afetada no geral.

Gráfico de Distribuição de Óbitos por Sexo:

São 5768 óbitos masculinos, 5263 femininos e 251 indeterminados.

Gráfico de Distribuição de Óbitos por Faixa Etária:

A faixa-etária de 40 até 99 anos ocupam as maiores quantidades de óbitos registrados, de 1.035 até 1.329.

A faixa etária de <9, 20-29 e 30-39 tem números moderados de óbitos registrados, de 766 até 899.

As faixas-etárias com menor quantidade de registros são as de 10-19 e >100 (além da N/I), com 547, 596 e 495, respectivamente.

Gráfico de Distribuição de Óbitos por Tipo de Doença:

Insuficiência Respiratória, Pneumonia e Septicemia causam um alto volume de óbitos, por volta de 1.750 a 2.000.

Covid e Srag causam um baixo-moderado volume de óbitos, com cerca de 700.

Outras doenças, assim como Indeterminadas, causam cerca de 3.400 e 1.000 óbitos, respectivamente.

Gráfico de Números de Óbitos por Faixa Etária:

É notável que em praticamente todas as faixas etárias, a ordem de doenças com mais óbitos são: “Outras”, Insuficiência Respiratória ou Pneumonia, Septicemia, “Indeterminada”, Covid e Srag.

Todas as faixas-etárias abaixo de 60-69 tem mais óbitos por Insuficiência Respiratória do que Pneumonia, e todas que vem depois, tem mais mortes por Pneumonia.

Inferências e testes estatísticos realizados

As conclusões finais levando em consideração todos os insights discutidos, são:

Homens ocupam 52% dos óbitos (mulheres 48%). Essa diferença, é influenciada por diversos fatores, como padrões de comportamento, exposição ocupacional e fatores socioeconômicos.

Idosos são mais propensos a falecer por doenças respiratórias, isso pode ser devido a:

- Sistema imunológico enfraquecido.
- Presença de doenças crônicas.
- Redução da capacidade pulmonar.
- Complicações de tratamentos médicos.

A região Sudeste, junto com outros Estados como BA, CE e PE por exemplo, tem os maiores números de óbitos, isso se deve a alta densidade populacional, mas também a fatores como, infraestrutura de saúde e a resposta regional à pandemia.