

UNIVERSIDADE PRESBITERIANA MACKENZIE

**Curso de Tecnologia em Ciência de Dados**

ANNA TERESA SOARES SACCHI - RA: 10441273

GLEIDER MACKEDANZ DE CAMPOS - RA: 10433464

GUSTAVO AZEVEDO GOMYDE - RA: 10424543

KELLY HARO VASCONCELLOS - RA: 10441014

**Título:** Análise Exploratória de Dados do Dataset “Notificação de Suspeita de Covid-19” - Comparação Entre os Estados Mais e Menos Populosos das Regiões Brasileiras.

São Paulo

Novembro de 2024

UNIVERSIDADE PRESBITERIANA MACKENZIE

**Curso de Tecnologia em Ciência de Dados**

ANNA TERESA SOARES SACCHI - RA: 10441273

GLEIDER MACKEDANZ DE CAMPOS - RA: 10433464

GUSTAVO AZEVEDO GOMYDE - RA: 10424543

KELLY HARO VASCONCELLOS - RA: 10441014

**Título:** Análise Exploratória de Dados do Dataset “Notificação de Suspeita de Covid-19” - Comparação Entre os Estados Mais e Menos Populosos das Regiões Brasileiras.

Este estudo investiga a incidência de casos suspeitos de COVID-19 nos estados mais e menos populosos de cada região brasileira, focando em dados da síndrome gripal. A análise compara a evolução da doença, a distribuição da vacinação, as características demográficas dos pacientes, e a relação entre sintomas e óbitos. O estudo também examina a atuação de profissionais da saúde e os casos assintomáticos.

São Paulo

Novembro de 2024

## GLOSSÁRIO

**COVID-19:** Doença causada pelo coronavírus SARS-CoV-2, que se espalhou globalmente em 2020, causando uma pandemia.

**e-SUS Notifica:** Sistema utilizado pelo Ministério da Saúde para notificação de casos suspeitos ou confirmados de síndrome gripal e COVID-19.

**Síndrome gripal:** Conjunto de sintomas relacionados a infecções respiratórias virais, incluindo febre, tosse, dor de garganta e dificuldade respiratória.

**Data Science:** Campo interdisciplinar que utiliza métodos, algoritmos, processos e sistemas para extrair conhecimento e insights de dados.

**IBGE:** Sigla para Instituto Brasileiro de Geografia e Estatística, o órgão oficial de estatísticas do Brasil. Ele fornece dados demográficos, econômicos e geográficos.

**Python:** Linguagem de programação amplamente utilizada em ciência de dados e outras áreas da computação.

**Github:** Plataforma de hospedagem de código, usada para o controle de versão e colaboração em projetos de programação.

**Dataset:** Conjunto de dados organizado em formato tabular (linhas e colunas), onde cada linha representa uma instância (registro) e cada coluna, uma variável (atributo). Datasets são a base para análises em data science e podem conter informações de diversas naturezas.

**Repositório:** É um local centralizado onde dados ou códigos são armazenados e gerenciados, permitindo o versionamento e a colaboração em projetos. Pode ser local ou remoto, acessível através de plataformas como GitHub.

**Machine Learning:** É uma área da inteligência artificial que permite que sistemas aprendam e façam previsões a partir de dados, sem serem explicitamente programados. Utiliza algoritmos para identificar padrões e melhorar o desempenho ao longo do tempo.

**Insights:** São descobertas ou entendimentos profundos obtidos a partir da análise de dados, observações ou experiências, que revelam novas perspectivas e auxiliam na tomada de decisões.

**Dataframe:** É uma estrutura de dados bidimensional usada em bibliotecas como Pandas, que organiza dados em formato de tabela, com linhas e colunas rotuladas, facilitando a manipulação e análise de grandes conjuntos de dados.

## RESUMO

O projeto visa realizar uma análise comparativa da incidência de síndromes gripais, com foco na COVID-19, nas regiões Norte, Sul, Sudeste, Centro-Oeste, Nordeste do Brasil, utilizando dados do sistema e-SUS Notifica. A análise explora variáveis como o estado de notificação, idade, sexo, histórico de vacinação, ocorrência de sintomas em caso de óbito, profissionais da saúde e assintomáticos, com o objetivo de identificar padrões regionais e comportamentais na evolução dos casos durante a pandemia. Com base em um conjunto de dados extensivo e rico, o estudo utiliza técnicas de Ciência de Dados e a linguagem de programação Python para gerar insights, incluindo medidas de posição e dispersão, análise e detecção de anomalias. Além disso, o projeto busca identificar correlações entre variáveis relacionadas ao perfil dos pacientes e sua resposta vacinal. A pesquisa proporciona uma compreensão mais profunda sobre como diferentes estados brasileiros foram afetados pela COVID-19, contribuindo para o aprimoramento de políticas de saúde pública.

## **LISTA DE FIGURAS**

|  |           |
|--|-----------|
| Figura 01 – Distribuição total das vacinações do Covid-19 de 2020 a 2024.                                    | <b>12</b> |
| Figura 02 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados mais populosos.               | <b>12</b> |
| Figura 03 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados menos populosos.              | <b>13</b> |
| Figura 04 – Volumetria total de vacinas aplicadas por laboratório distribuídos por tipo de dose.             | <b>15</b> |
| Figura 05 – Volumetria total, em milhões, de vacinas aplicadas por laboratório.                              | <b>16</b> |
| Figura 06 – Distribuição etária geral dos pacientes notificados com Covid-19 em todos os estados do dataset. | <b>18</b> |
| Figura 07 – Distribuição percentual do sexo (Feminino e Masculino).  | <b>19</b> |
| Figura 08 - Frequência dos sintomas em casos de óbito.   | <b>21</b> |
| Figura 09 - Frequência que profissionais da área da saúde e suas áreas aparecem na base de dados.            | <b>23</b> |
| Figura 10 - Frequência casos assintomático dividido por sexo.  | <b>24</b> |
| Figura 11 - Frequência de casos assintomáticos divididos por Raça/Cor.                                       | <b>25</b> |
| Figura 12 - Frequência de casos assintomáticos divididos por faixa etária.                                   | <b>27</b> |

## **LISTA DE TABELAS**

|   |           |
|---|-----------|
| Tabela 01 - Total de ocorrências de sintomas quando ocorre o óbito do paciente.   | <b>21</b> |
| Tabela 02 - Profissionais das 5 áreas da saúde que aparecem com maior frequência. | <b>22</b> |
| Tabela 03 - Quantidade casos assintomáticos divididos por sexo.                   | <b>24</b> |
| Tabela 04 - Quantidade casos assintomáticos divididos por Raça/Cor.               | <b>25</b> |
| Tabela 05 - Quantidade casos assintomáticos divididos por faixa etária.           | <b>26</b> |

## **SUMÁRIO**

|  |           |
|--|-----------|
| 1. Introdução  | <b>08</b> |
| 2. Objetivo  | <b>08</b> |
| 3. Apresentação da Empresa e Problema de Pesquisa  | <b>08</b> |
| 4. Metadados   | <b>10</b> |
| 5. Análise Exploratória de Dados (AED)   | <b>10</b> |
| 5.1 Análise 01 - Comparação da evolução do Covid-19 entre regiões menos populosas e mais populosas | <b>11</b> |
| 5.2 Análise 02 - Volumetria geral das vacinas aplicadas por laboratório                            | <b>14</b> |
| 5.3 Análise 03 - Comparação entre sexo e idade dos pacientes com Covid-19                          | <b>17</b> |
| 5.4 Análise 04 - Ocorrência de Sintomas em Casos de Óbito  | <b>21</b> |
| 5.5 Análise 05 - Quantidade e área de atuação de profissionais da área da saúde                    | <b>22</b> |
| 5.6 Análise 06 - Casos Assintomáticos  | <b>24</b> |
| 6. Esboço Data Storytelling  | <b>28</b> |
| 7. Apresentação do Projeto – Vídeo   | <b>37</b> |
| 8. Apresentação dos Slides   | <b>37</b> |
| 9. Conclusão   | <b>47</b> |
| 10. Referências  | <b>48</b> |





## 1. INTRODUÇÃO

A pandemia de COVID-19, que teve início no final de 2019, trouxe profundas mudanças sociais, econômicas e de saúde em todo o mundo, exigindo uma resposta rápida e eficaz de governos e instituições de saúde. No Brasil, a análise da incidência de casos suspeitos de COVID-19 revelou disparidades significativas entre as diferentes regiões e estados do país, especialmente quando se considera a população de cada localidade. Este projeto tem como objetivo realizar uma análise exploratória de dados do dataset “Notificação de Suspeita de COVID-19”, com foco na comparação entre estados que representam os maiores e menores índices populacionais em cada uma das cinco regiões brasileiras: Norte, Nordeste, Sudeste, Sul e Centro-Oeste.

Por meio da análise de dados brutos relacionados a casos de síndrome gripal com suspeita de COVID-19, buscamos entender como a distribuição e a evolução da doença se comportam em diferentes contextos demográficos. Ao investigar a relação entre a densidade populacional e a incidência de casos, pretendemos identificar padrões que possam contribuir para a formulação de políticas públicas e estratégias de prevenção mais eficazes. Além disso, a pesquisa abordará a volumetria das vacinas aplicadas no contexto geral, a relação entre sexo e idade dos pacientes em relação aos estados mais populosos em comparação com os menos populosos, o impacto dos sintomas mais encontrados nos óbitos e uma análise nos sintomas assintomáticos, buscando oferecer uma visão abrangente e crítica sobre a pandemia em território brasileiro.

## 2. OBJETIVO DO ESTUDO

O objetivo deste estudo é realizar uma análise comparativa da incidência de suspeita de COVID-19 em estados com diferentes densidades populacionais nas regiões brasileiras, utilizando dados do sistema e-SUS Notifica.

## 3. APRESENTAÇÃO DA EMPRESA E PROBLEMA DE PESQUISA

**Nome da Empresa:** Science for Life SA

**Missão:** Através da análise de dados temos como missão contribuir para gestão de risco de doenças infecciosas, fornecendo dados sobre padrões e tendências de disseminação das doenças.

**Visão:** Buscamos democratizar os dados através de uma plataforma open source visando a melhoria da qualidade de vida da população.

**Segmento de Atuação:** Pesquisa em saúde pública e análise de dados epidemiológicos.

**Market Share:** Atuamos em colaboração com instituições de saúde pública, universidades e organizações não governamentais, focando na análise de dados de saúde em nível regional e nacional.

**Número de Colaboradores:** 04

**Iniciativas em Data Science:** Estamos desenvolvendo estudos para predição de epidemiologias a nível nacional.

**Proposta do que será desenvolvido:** Serão conduzidas 06 análises exploratórias buscando elucidar as seguintes questões:

- Evolução das doenças nas regiões brasileiras ( norte, nordeste, sul, sudeste e centro-oeste, com os estados mais populosos versus estados menos populosos).
- Volumetria geral das vacinações.
- Distribuição e predominância entre o sexo e idade nos estados.
- Ocorrência de Sintomas em Casos de Óbito.
- Quantidade e área de atuação de profissionais da área da saúde.
- Análise nos casos assintomáticos.

Para cada análise, foram aplicados filtros e tratadas as anomalias relevantes ao seu contexto específico.

**Problema de Pesquisa:** A análise busca entender como a incidência de casos de COVID-19 varia entre os estados mais e menos populosos nas diferentes regiões do Brasil, para explorar esse problema buscamos dividir em 04 fases (decomposição), em cada uma das análises foi realizada a identificação de padrões, agrupando dados pelas variáveis pertinentes, após isso realizamos as filtragens ( abstração), por fim o algoritmo resultou na visualização dos dados.

**Pipeline de dados:**

- Coleta dos dados no site DataSus;
- Armazenamento dos CVs no drive da empresa;
- Processamentos dos CVS para gerar única base de dados;
- Processamentos individuais para criação das análises;
- Análises a partir das visualizações de dados;
- Documentação no Github e relatório.

## 4. METADADOS

**Dataset:** Notificação de Suspeita de COVID-19

- **Tipo de Arquivo:** CSV
- **Origem dos Dados:** Sistema e-SUS Notifica, Ministério da Saúde
- **Sensibilidade:** Dados públicos com informações demográficas e de saúde.
- **Validade:** Dados coletados de março de 2020 até 2024.
- **Proprietário dos Dados:** Ministério da Saúde
- **Descrição dos Atributos Utilizados:**
  - Estado
  - Estado IBGE
  - Estado Notificação IBGE
  - Sexo
  - Idade
  - Código Laboratorio Primeira Dose
  - Código Laboratorio Segunda Dose
  - Data Coleta Primeiro Teste
  - Data Coleta Segundo Teste
  - Código Resultado Primeiro Teste
  - Código Resultado Segundo Teste
  - Sintomas
  - Se é um profissional da área da saúde
  - CBO
  - Evolução do caso

## 5. ANÁLISE EXPLORATÓRIA DE DADOS (AED)

**Estrutura do Dataset:**

- Número de linhas: 2.791.570
- Número de colunas existentes: 64
- Número de colunas utilizadas: 15
- Tipos de dados utilizados (variáveis):
  - Estado: Texto
  - Estado IBGE: Texto
  - Estado Notificação IBGE: Texto
  - Sexo: Categórico (Masculino, Feminino ou Indefinido)

- Idade: Número Inteiro
- Código Laboratorio Primeira Dose: Categórico
- Código Laboratorio Segunda Dose: Categórico
- Data Coleta Primeiro Teste: Data e hora
- Data Coleta Segundo Teste: Data e hora
- Código Resultado Primeiro Teste: Número com vírgula (Float)
- Código Resultado Segundo Teste: Número com vírgula (Float)
- Sintomas: Texto
- Se é um profissional da área da saúde: Categórico binário (Sim ou Não)
- CBO: Texto
- Evolução do caso: Categórico (Cura, Ignorado, Cancelado, Óbito ou nulo)

### **5.1. Análise 01 - Comparação da evolução do Covid-19 entre regiões menos populosas e mais populosas**

- Foi identificado neste dataset um total de 15.216 registros de vacinas, referente a 90.803 registros agrupados por data, para cada registro havia pelo menos uma realização de teste de Covid-19, considerando as primeiras duas coletas de teste de Covid-19 que a pessoa realizou. A Figura 01 representa a distribuição da volumetria de vacinas para todos os estados selecionados para o estudo, independente de outras variáveis. Além disso foi explorado a utilização da confirmação de casos de Covid, porém foi descartada, dado que reduzia significativamente a volumetria dos casos a serem analisados.
- Link da análise: [https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/Comparacao\\_COVID\\_regioesBrasileiras.py](https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/Comparacao_COVID_regioesBrasileiras.py)
- Colunas relevantes:
  - Data Coleta Primeiro Teste: Data e hora
  - Data Coleta Segundo Teste: Data e hora
  - Código Resultado Primeiro Teste: Número com vírgula (Float)
  - Código Resultado Segundo Teste: Número com vírgula (Float)

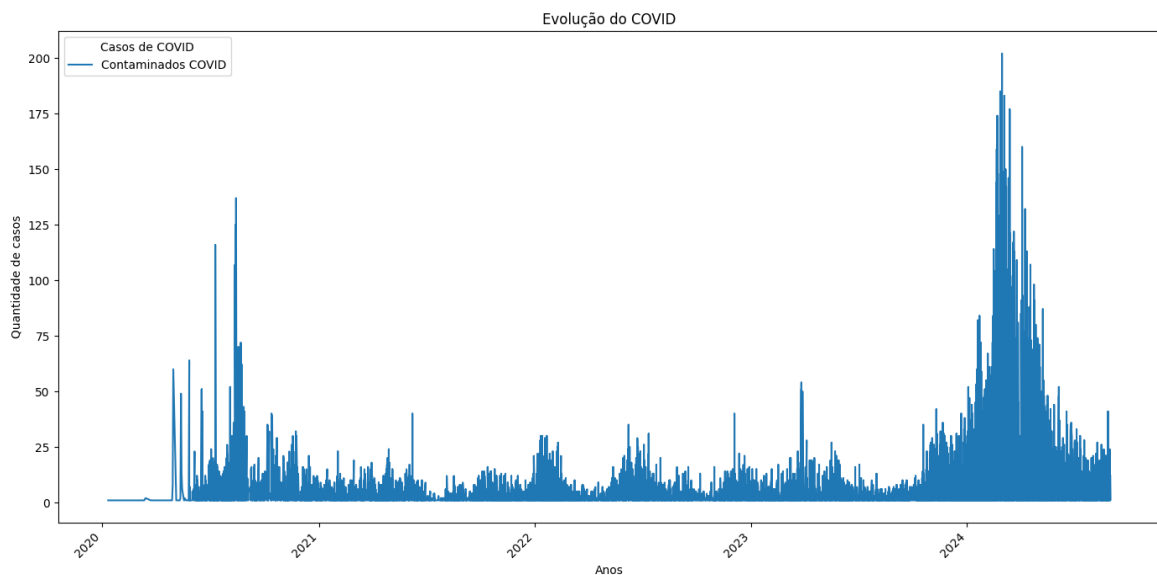


Figura 01 – Distribuição total das vacinações do Covid-19 de 2020 a 2024.

### Medidas de Posição e Dispersão:

A Figura 01 sugere que houve um pico de vacinações no início e no final do período. O maior número de vacinações por dia registrado na base foi de 202 vacinações, enquanto que valor mínimo foi de 0 vacinações ao dia. Em 28% do dataframe, referente a todos os registros em que a contagem era maior que 1, existe 47% da volumetria total de casos (42.757 registros agrupados), assim podemos concluir que houve uma boa distribuição das vacinas ao longo do tempo, por exeção dos picos, inicial e final.

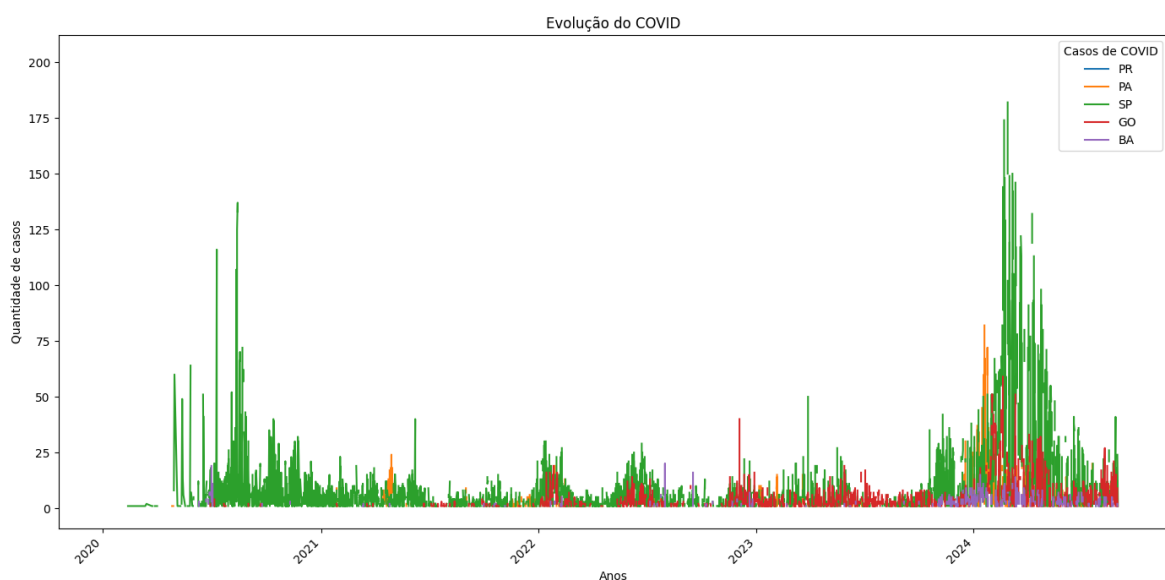


Figura 02 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados mais populosos

Podemos ver que para os estados mais populosos o gráfico segue o mesmo padrão do gráfico total, sendo um subgrupo do primeiro. Em 18% do dataframe, referente a todos os registros em que a contagem era maior que 9, existe 66% da volumetria total de casos (54.128 registros agrupados), assim podemos concluir que houve uma maior concentração das vacinas em certos períodos do tempo, o que pode ter influenciado o perfil do primeiro gráfico, refletindo os picos inicial e final dado que a representatividade deste subgrupo na base total é de 89,9%.

Finalmente, para os estados menos populosos, o gráfico difere do padrão dos outros dois, sendo que por ser o grupo menos populoso a quantidade de registros não agrupados é de 9.176 pessoas. Em 44% do dataframe, referente a todos os registros em que a contagem era maior que 1, existe 77% da volumetria total de casos (7.107 registros agrupados), assim podemos concluir que houve uma concentração em certos períodos, ainda que menos que o do subgrupo anterior.

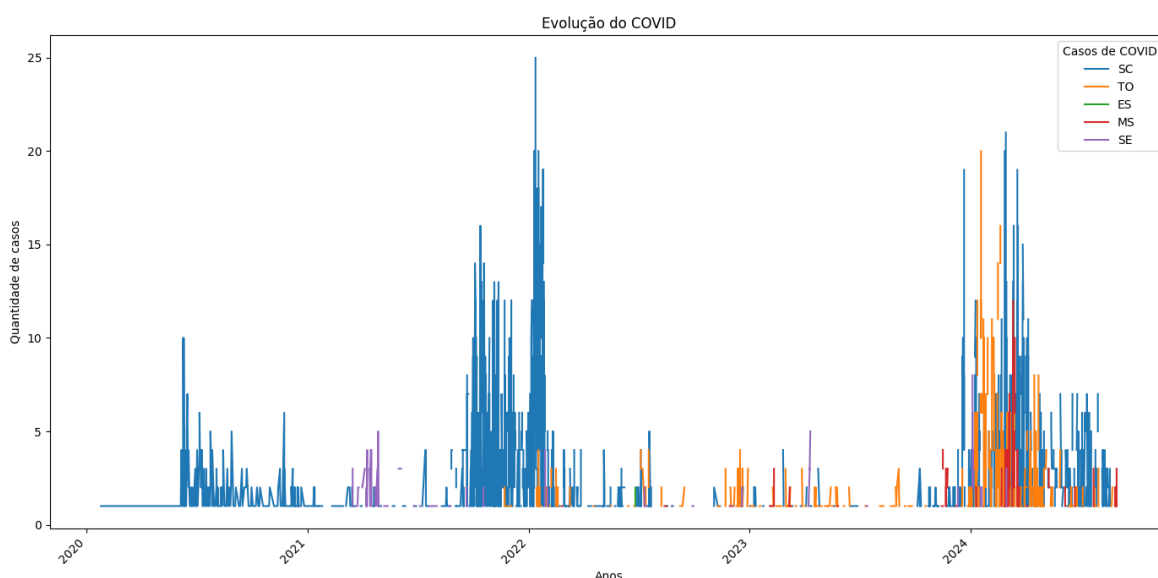


Figura 03 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados menos populosos

#### Distribuição de Variáveis:

- Foram realizadas 81.644 vacinações nos estados PR, SC, PA, TO, SP.
- Foram realizadas 9176 vacinações nos estados ES, GO, MS, BA, SE.
- Existe na base total uma concentração de casos entre os períodos 06/2020 até 01/2021 e 08/2023 até 03/2024, para o grupo menos populoso existe uma 3ª zona de concentração dos casos que ocorre no período de 07/2021 até 01/2022.

### **Correlações:**

É conhecido que houveram diversos picos da doença Covid-19 ao longo de toda a pandemia, e esse gráfico reflete isto parcialmente. Também podemos notar a correlação entre as volumetrias de vacinação entre os grupos mais e menos populosos, no qual a volumetria do primeiro ultrapassa em mais de 8 vezes a do segundo, vale notar que as diferenças devem ser ocasionadas por múltiplos fatores além da própria volumetria, como a organização do sistema de saúde dos municípios além da distribuição da vacina em cada um deles, o que poderia justificar o pico tardio que temos no grupo menos populoso.

### **Anomalias e Valores:**

- **Anomalias:** Dentro da totalidade de casos temos 19,7% nos quais constam apenas 1 vacinação (dias e regiões variadas), o que provavelmente não reflete a realidade.
- **Valores:** As colunas de data foram filtradas para remover casos em que a pessoa não fez o teste de covid (ambas as datas vazias), que impediriam correlacionar os registros da base com a evolução do covid. A pessoa fazer o teste pressupõe que ela teve alguns dos sintomas associados a Covid-19.

## **5.2. Análise 02 - Volumetria geral das vacinas aplicadas por laboratório**

- Foram identificados neste dataset um total de 3.450.409 registros de vacinas aplicadas em diferentes estados brasileiros, com o objetivo de ilustrar a quantidade total de vacinas por dose e a quantidade total de vacinas aplicadas por laboratório.
- Link da análise: <https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/vacinacao.py>
- Colunas relevantes:
  - Código Laboratorio Primeira Dose: De qual laboratório a primeira dose da vacina foi aplicada.
  - Código Laboratorio Segunda Dose: De qual laboratório a segunda dose da vacina foi aplicada.

### **Medidas de Posição e Disposição:**

- A distribuição dos valores em ambos os gráficos dão uma ideia visual da dispersão, mostrando como os números variam entre laboratórios e doses em termos de quantidade de vacinas aplicadas.
- As contagens (frequências) das vacinas aplicadas por laboratório podem ser vistas como uma forma de medida de posição. Por exemplo, o número total de vacinas

aplicadas por cada fabricante é uma posição no conjunto de dados, mostrando onde estão concentradas as aplicações (primeira e segunda doses).

### Distribuição de Variáveis:

- Total de registros da Primeira Dose por laboratório:
  - Astrazeneca: 640.780
  - Biontech: 373
  - Butantan: 619.180
  - Janssen: 1.773
  - Pfizer: 521.396
  - Sinovac: 834
- Total de registros da Segunda Dose por laboratório:
  - Astrazeneca: 579.948
  - Biontech: 117
  - Butantan: 577.685
  - Janssen: 2.980
  - Pfizer: 504.548
  - Sinovac: 795

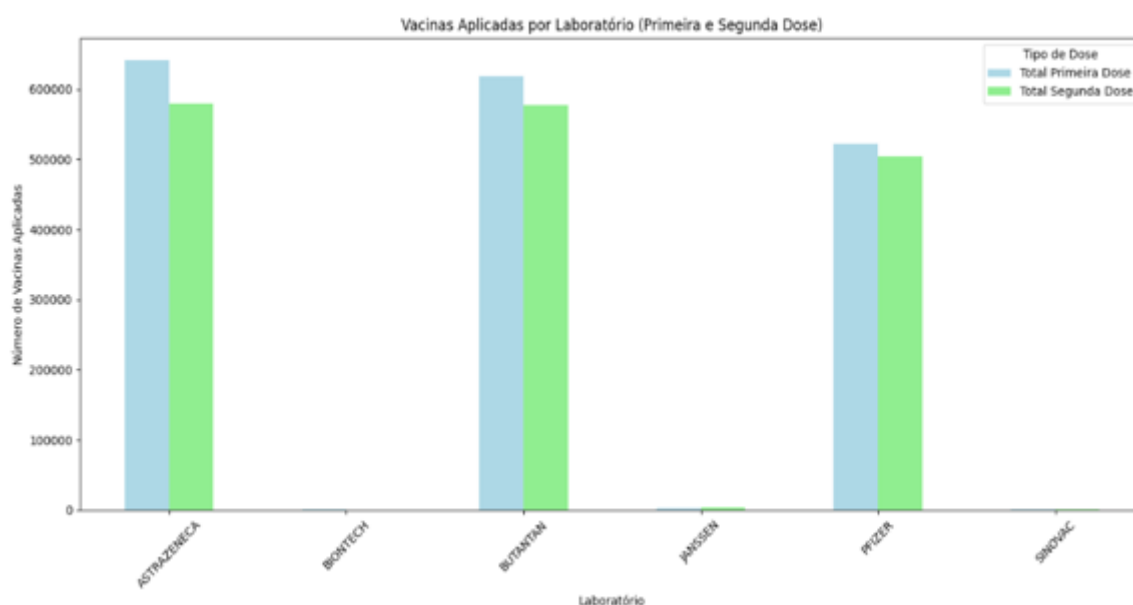


Figura 04 - Volumetria total de vacinas aplicadas por laboratório distribuídos por tipo de dose.



- Total de registros de vacinas aplicadas por laboratório:
  - Astrazeneca: 1.220.728
  - Biontech: 490
  - Butantan: 1.196.865
  - Janssen: 4.753
  - Pfizer: 1.025.944
  - Sinovac: 1.629

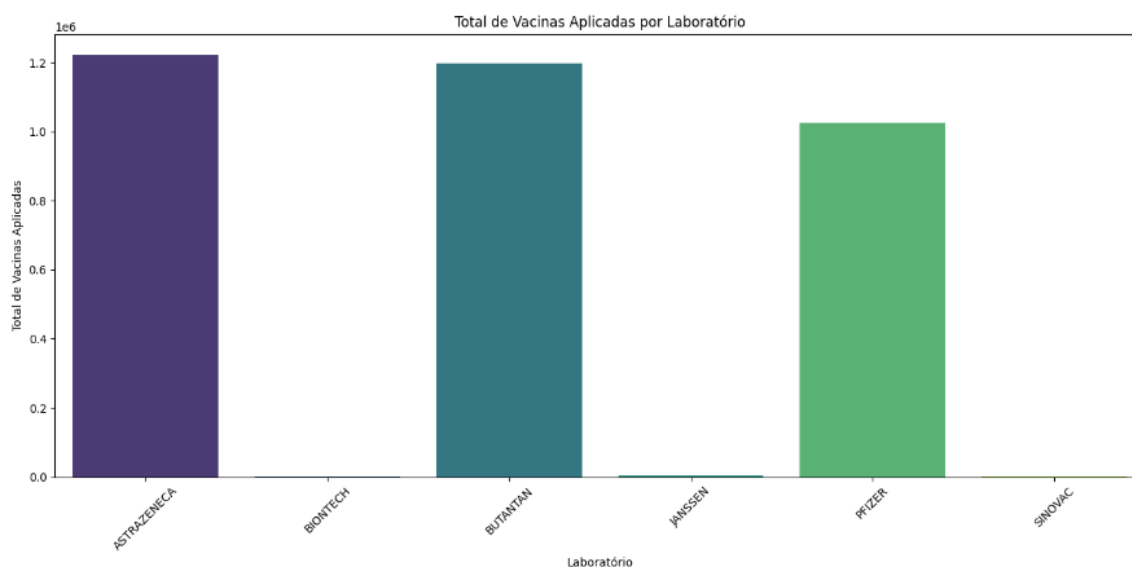


Figura 05 - Volumetria total, em milhões, de vacinas aplicadas por laboratório.

## Correlações:

### Volumetria de vacinas aplicadas por doses:

- **Correlação entre doses:** Nota-se que o volume total de vacinas aplicadas da primeira dose em comparação a segunda de todos os laboratórios são de pouca diferença, pois é indicado ao paciente que para garantir a proteção indicada é essencial seguir a imunização completa.

### Volumetria de vacinas aplicadas por laboratório:

- **Correlação entre laboratórios:** Percebe-se que o volume total de vacinas aplicadas por laboratório deram destaque para 3 laboratórios. São eles:
  - Astrazeneca: Custos mais baixos e facilidade de armazenamento em temperaturas normais de refrigerador facilitaram a distribuição, especialmente em países em desenvolvimento.

- Butantan: A tecnologia tradicional de vacina inativada é bem compreendida e aceita em muitos países. A CoronaVac foi amplamente utilizada nas regiões da América Latina.
- Pfizer: Rápido desenvolvimento e resultados positivos em ensaios clínicos, além de uma boa resposta imunológica. Sua facilidade de produção em larga escala também contribuiu para a rápida distribuição.

#### **Anomalias e Valores:**

- **Anomalias:** Foram identificados valores não padrão em ambas as colunas Código Laboratorio Primeira Dose e Código Laboratorio segunda Dose. Para resolver, fizemos um agrupamento de nomes. Exemplo: (Astrazeneca, Astrazeneca AB, Astrazeneca/Fiocruz, Astrazeneca/Oxford).
- **Valores:** As duas colunas foram filtradas para remover entradas NaN, que se permanecessem impediriam a análise.

### **5.3. Análise 03 - Comparação entre sexo e idade dos pacientes com Covid-19**

- Para esta análise, o dataset foi normalizado, resultando em 2.005.120 registros de notificações de casos de COVID-19 em diferentes estados brasileiros.
- Link da análise: [https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/media\\_sexo\\_idade.py](https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/media_sexo_idade.py)
- Colunas relevantes:
  - Sexo: Gênero do paciente (Feminino, Masculino).
  - Idade: Idade dos pacientes.
  - Estado Notificação IBGE: Código do estado de notificação.

#### **Medidas de Posição e Dispersão:**

- Total de registros: 2.005.120
- Número de registros inválidos: 5 registros com valores nulos, um número inexpressivo em relação à quantidade total.
- Média das Idades por Estado e Sexo:
  - Sul: PR (Feminino: 30 anos, Masculino: 27 anos), SC (Feminino: 30 anos, Masculino: 28 anos)
  - Norte: PA (Feminino: 30 anos, Masculino: 27 anos), TO (Feminino: 32 anos, Masculino: 29 anos)

- Sudeste: SP (Feminino: 30 anos, Masculino: 26 anos), ES (Feminino: 33 anos, Masculino: 31 anos)
- Centro-Oeste: GO (Feminino: 31 anos, Masculino: 27 anos), MS (Feminino: 30 anos, Masculino: 27 anos)
- Nordeste: BA (Feminino: 28 anos, Masculino: 23 anos), SE (Feminino: 28 anos, Masculino: 24 anos)

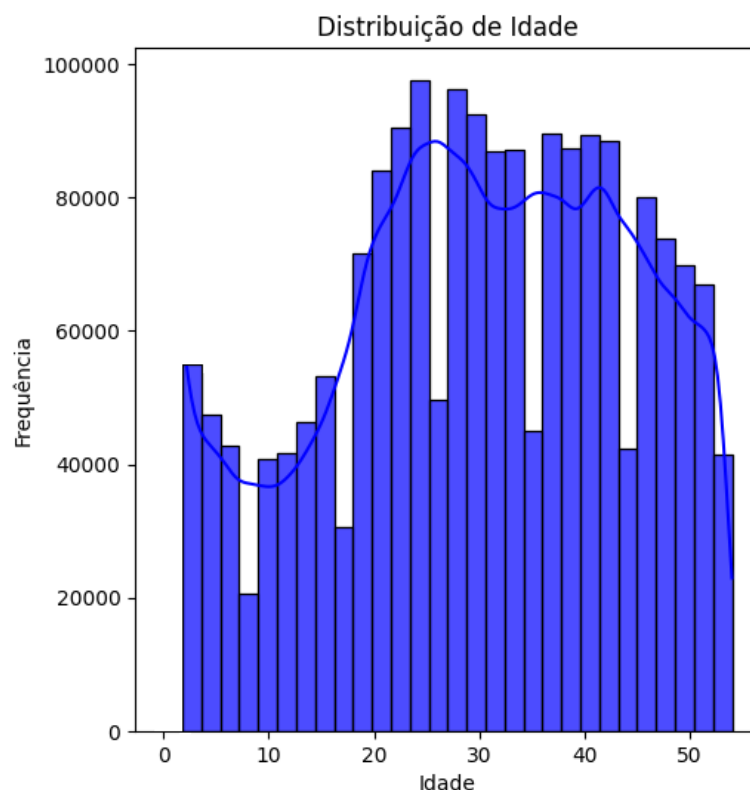


Figura 06 – Distribuição etária geral dos pacientes notificados com Covid-19 em todos os estados do dataset.

#### Distribuição de Variáveis:

- Distribuição por Sexo em cada Estado:
  - Sul: PR: Feminino: 59.9%, Masculino: 40.1%, SC: Feminino: 57.6%, Masculino: 42.4%
  - Norte: PA: Feminino: 58.3%, Masculino: 41.7%, TO: Feminino: 64.3%, Masculino: 35.7%
  - Sudeste: SP: Feminino: 59.0%, Masculino: 41.0%, ES: Feminino: 49.7%, Masculino: 50.3%
  - Centro-Oeste: GO: Feminino: 59.6%, Masculino: 40.4%, MS: Feminino: 62.2%, Masculino: 37.8%

- Nordeste: BA: Feminino: 56.4%, Masculino: 43.6%, SE: Feminino: 56.9%, Masculino: 43.1%

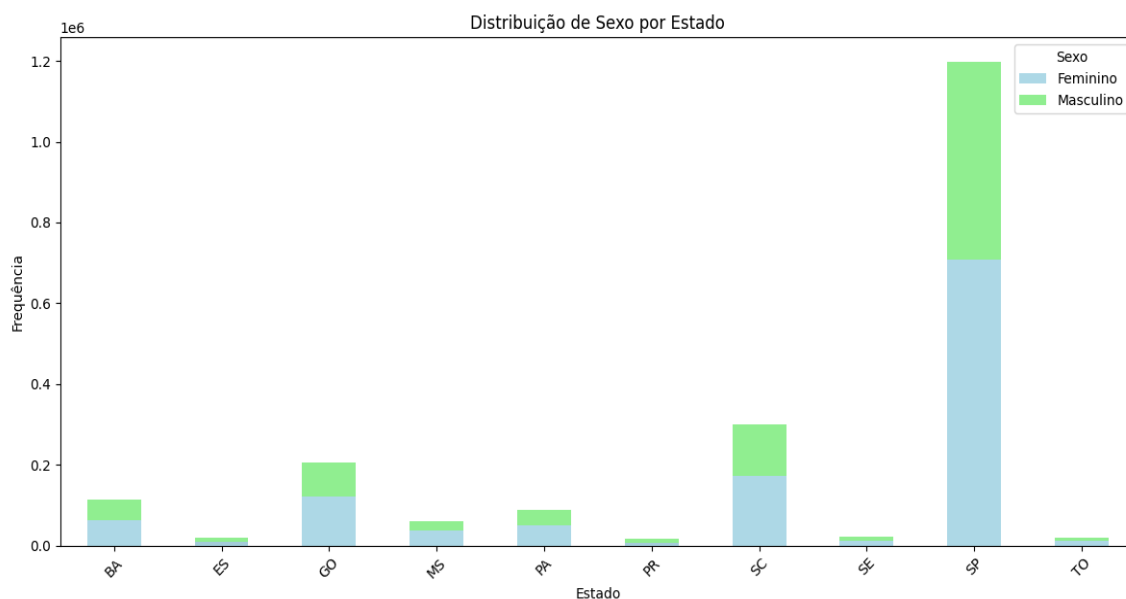


Figura 07 – Distribuição percentual do sexo (Feminino e Masculino) entre os pacientes notificados com Covid-19 em diferentes estados brasileiros.

## Correlações:

### Média de Idade das Pessoas Notificadas:

- **Distribuição Etária:** A média de idade dos notificados por estado variam de 23 a 33 anos, revelando uma predominância entre adultos jovens. Essa faixa etária sugere que a maior interação social, característica desse grupo, pode facilitar a disseminação do vírus, especialmente em ambientes de aglomeração, como no trabalho e em atividades de lazer.
- **Justificativa da Distribuição:** A falta de registros significativos de crianças e idosos pode ser atribuída a diversos fatores: **Assintomaticidade das Crianças:** Pesquisas indicam que crianças e adolescentes frequentemente apresentam a COVID-19 de forma assintomática ou com sintomas leves, resultando em uma subnotificação entre esse grupo. **Sintomas Graves em Adultos e Idosos:** Já os adultos de meia-idade e idosos tendem a desenvolver sintomas mais severos, levando a uma maior hospitalização e, conseqüentemente, à exclusão dessas notificações no dataset, que foca apenas nos casos leves e moderados.

**Predominância do Sexo:** A análise releva uma predominância do sexo feminino em várias regiões. Essa tendência pode ser explicada por diversos fatores. Estudos sugerem que, embora os homens possam ter uma maior propensão a contrair COVID-19, são as

mulheres que frequentemente apresentam formas leves da doença. Como o dataset se concentra em pacientes com sintomas leves, essa diferença é relevante. Além disso, as mulheres tendem a ser mais proativas na busca por assistência médica, enquanto os homens muitas vezes mostram resistência em procurar ajuda. Esse comportamento social pode contribuir para o maior número de notificações entre as mulheres.

#### **Estados Mais Populosos versus Menos Populosos:**

- **Diferença na Notificação de Casos:** Estados como São Paulo, com alta densidade populacional, apresentam maior exposição ao vírus, resultando em um número mais elevado de casos notificados.
- **Infraestrutura de Saúde:** A infraestrutura de saúde varia significativamente entre estados. Estados menos populosos costumam ter menos unidades de saúde, o que pode limitar o acesso e reduzir a disposição das pessoas em buscar atendimento. Em contraste, a diversidade de serviços de saúde em estados mais populosos pode incentivar um comportamento mais proativo na busca por tratamento e, conseqüentemente, aumentar as notificações.

#### **Anomalias e Valores:**

- **Anomalias:** Foram identificados valores não padrão na coluna sexo, os quais foram removidos durante o pré-processamento (exemplos: Ba, Indefinido, Sc, espaços vazios).
- **Valores Perdidos:**
  - Não há valores inválidos na coluna sexo após o pré-processamento.
  - A coluna idade foi filtrada para remover entradas NaN, que se permanecessem impediriam a análise.

#### **5.4 - Análise 04 - Ocorrência de Sintomas em Casos de Óbito**

- Link da análise: [https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos\\_obitos.py](https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos_obitos.py)

#### **Medidas de Posição e Dispersão:**

Do total da base analisada, houve 2667 óbitos.

| Sintomas              | Ocorrências | Porcentagem (%) |
|-----------------------|-------------|-----------------|
| Assintomático         | 1.110       | 23,08           |
| Tosse                 | 841         | 17,49           |
| Dispneia              | 666         | 13,85           |
| Febre                 | 613         | 12,74           |
| Outros                | 587         | 12,20           |
| Coriza                | 340         | 7,07            |
| Dor de Cabeça         | 300         | 6,24            |
| Dor de Garganta       | 273         | 5,67            |
| Distúrbios Gustativos | 41          | 0,85            |
| Distúrbios Olfativos  | 38          | 0,79            |

Tabela 01 - Total de ocorrências de sintomas quando ocorre o óbito do paciente.

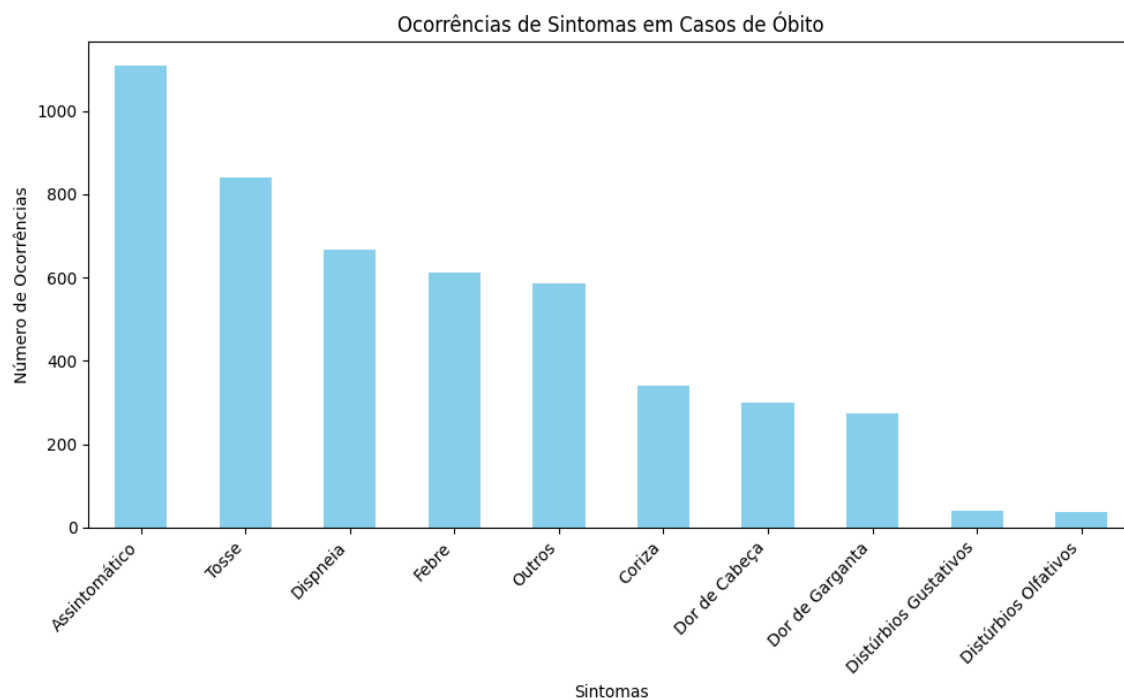


Figura 08 - Frequência dos sintomas em casos de óbito.

#### Correlações:

- **Assintomáticos:** O fato de que o maior número de óbitos ocorreu entre pessoas assintomáticas é um dado preocupante e pode sugerir que muitos desses

indivíduos não buscaram tratamento a tempo ou que subestimaram a gravidade da doença por não apresentarem sintomas.

- **Predominância dos sintomas respiratórios:** Tosse e dispneia, sendo sintomas respiratórios clássicos, estão fortemente associados a casos graves que resultaram em óbito. Isso destaca a necessidade de monitorar pacientes com esses sintomas mais de perto.
- **Sintomas menos frequentes:** Sintomas como distúrbios gustativos e olfativos são relatados com menos frequência nos casos de óbito, o que pode sugerir que esses sintomas estão mais associados a quadros mais leves da doença.

### 5.5 - Análise 05 - Quantidade e área de atuação de profissionais da área da saúde

- Link da análise: [https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos\\_profissionais\\_saude.py](https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos_profissionais_saude.py)

#### Medidas de Posição e Dispersão:

O total possui 51.088 profissionais da área da saúde, desses os que tem a maior frequência:

| Área da saúde (CBO)                 | Ocorrências | Porcentagem (%) |
|-------------------------------------|-------------|-----------------|
| Técnicos e Auxiliares de enfermagem | 15.917      | 31,15           |
| Enfermeiros e afins                 | 8.594       | 16,82           |
| Agentes Comunitários de Saúde       | 2.869       | 5,61            |
| Médico                              | 2.845       | 5,57            |
| Farmacêutico                        | 1.349       | 2,64            |

Tabela 02 - Profissionais das 5 áreas da saúde que aparecem com maior frequência.

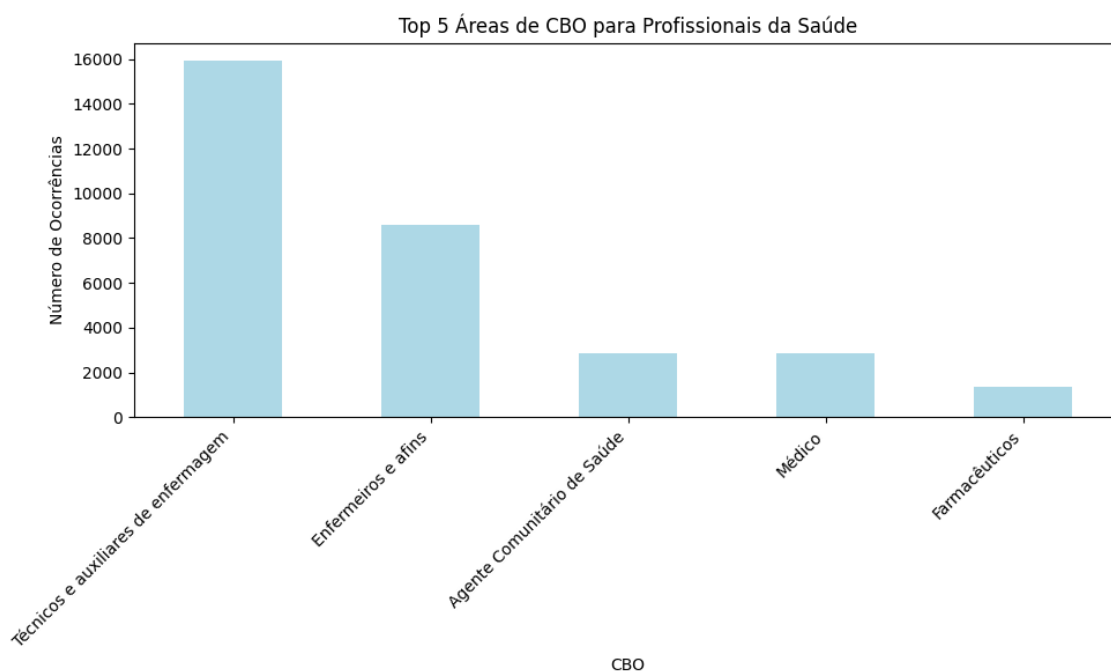


Figura 09 - Frequência que profissionais da área da saúde e suas áreas aparecem na base de dados.

### Correlações:

1. **Técnicos e auxiliares de enfermagem** são os mais expostos, possivelmente devido à sua maior interação e contato próximo com pacientes, especialmente em ambientes hospitalares, o que os coloca em risco elevado de infecção.
2. **Enfermeiros**, como supervisores diretos e cuidadores, também estão altamente expostos e desempenham um papel crucial no combate à pandemia.
3. **Agentes Comunitários de Saúde** têm uma exposição significativa devido à sua presença constante nas comunidades, trabalhando na prevenção e no monitoramento de possíveis casos.
4. **Médicos** possuem uma exposição moderada, com menos ocorrências comparado a técnicos e enfermeiros, possivelmente devido à sua interação mais pontual e menos prolongada com pacientes.
5. **Farmacêuticos** são menos expostos diretamente, mas desempenham funções vitais no fornecimento de medicamentos e EPIs.

Essas conclusões destacam como diferentes categorias de profissionais da saúde foram impactadas de maneira diversa durante a pandemia, dependendo da natureza de suas funções e do nível de contato com pacientes infectados. A análise também evidencia a importância dos EPIs e das medidas de proteção para esses trabalhadores, especialmente para os que têm contato mais direto com os pacientes.



## 5.6 - Análise 06 - Casos Assintomáticos

- Link da análise: [https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos\\_assintomaticos.py](https://github.com/KellyVasconcellos/projeto-aplicado-/blob/main/src/scripts/casos_assintomaticos.py)

Na base de dados possui 514.727 casos em que o sintoma do paciente é assintomático, nesses casos possuem as seguintes divisões:

1. Frequência por sexo:

### Medidas de Posição e Dispersão:

| Sexo       | Quantidade |
|------------|------------|
| Feminino   | 277.866    |
| Masculino  | 236.836    |
| Indefinido | 25         |

Tabela 03 - Quantidade casos assintomáticos divididos por sexo.

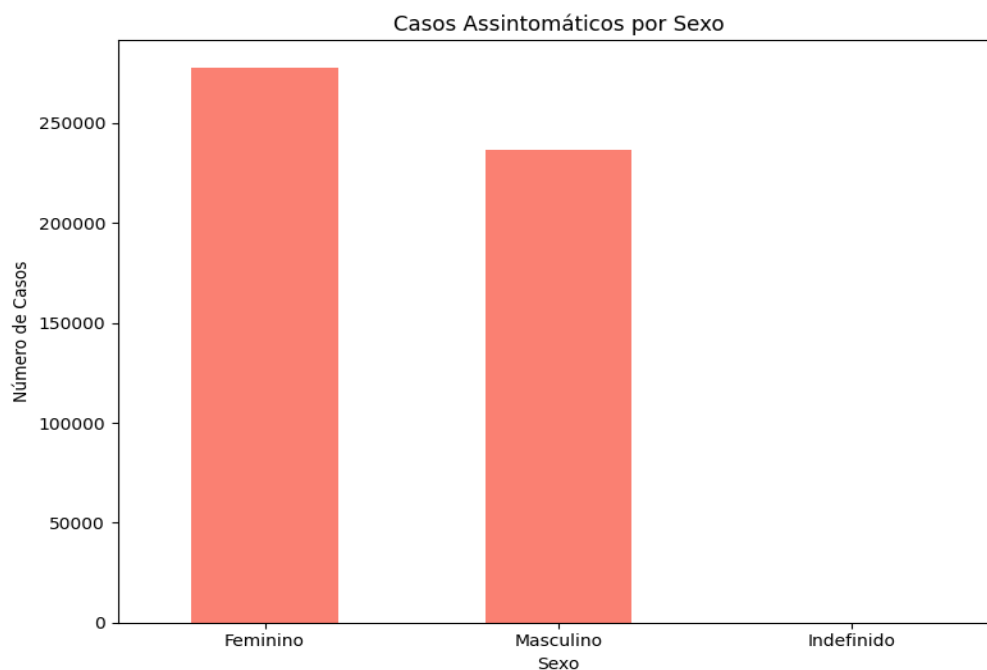


Figura 10 - Frequência casos assintomático dividido por sexo.

### Correlações:

A distribuição por sexo está relativamente equilibrada, com uma leve predominância de casos assintomáticos em pessoas do sexo feminino. A diferença de 54% (feminino) para 46% (masculino) indica uma pequena concentração maior de casos assintomáticos em

mulheres. Isso pode indicar fatores biológicos, comportamentais ou mesmo de testagem diferenciada, mas é necessário mais contexto para afirmar com certeza.

## 2. Frequência de Casos Assintomáticos por Raça/Cor:

### Medidas de Posição e Dispersão:

| Raça/cor | Quantidade |
|----------|------------|
| Branca   | 199.531    |
| Parda    | 152.118    |
| Ignorado | 107.224    |
| Amarela  | 33.445     |
| Preta    | 20.837     |
| Indigena | 1.189      |

Tabela 04 - Quantidade casos assintomáticos divididos por Raça/Cor.

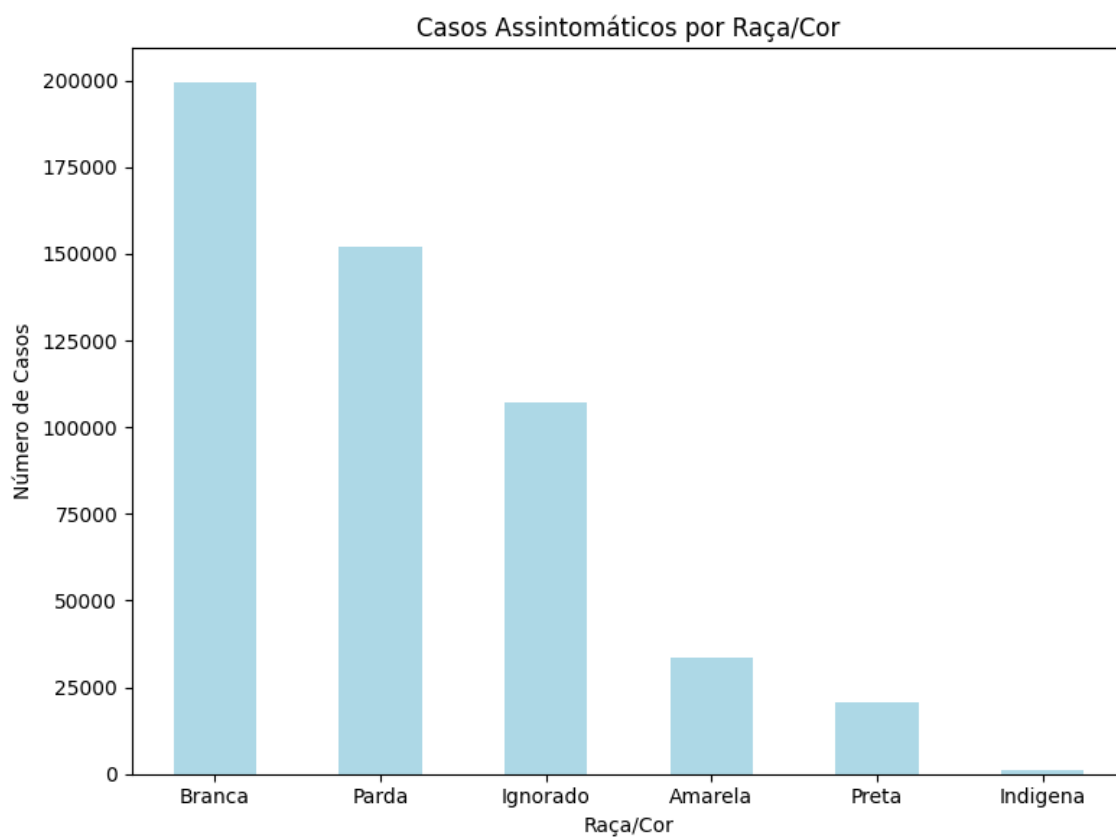


Figura 11 - Frequência de casos assintomáticos divididos por Raça/Cor.

### Correlações:

A maioria dos casos assintomáticos é reportada entre pessoas **Branças** (41%) e **Pardas** (31%). Há também um número significativo de dados com **Raça/Cor Ignorada** (22%), o que pode dificultar uma análise mais precisa. A população **Amarela** e **Preta** apresentam casos assintomáticos em proporções menores, e o número de **Indígenas** é bastante baixo.

Essa distribuição pode refletir desigualdades no acesso aos serviços de saúde, diferenças demográficas das populações afetadas, ou questões relacionadas à testagem em grupos mais vulneráveis.

### 3. Frequência de casos Assintomáticos por faixa etária:

#### Medidas de Posição e Dispersão:

| Faixa etária | Quantidade |
|--------------|------------|
| 19 - 30      | 94.732     |
| 31 - 40      | 82.096     |
| 41 - 50      | 77.928     |
| 0 - 18       | 66.920     |
| 51 - 60      | 21.809     |
| 61 - 70      | 0          |
| 70+          | 0          |

Tabela 05 - Quantidade casos assintomáticos divididos por faixa etária.

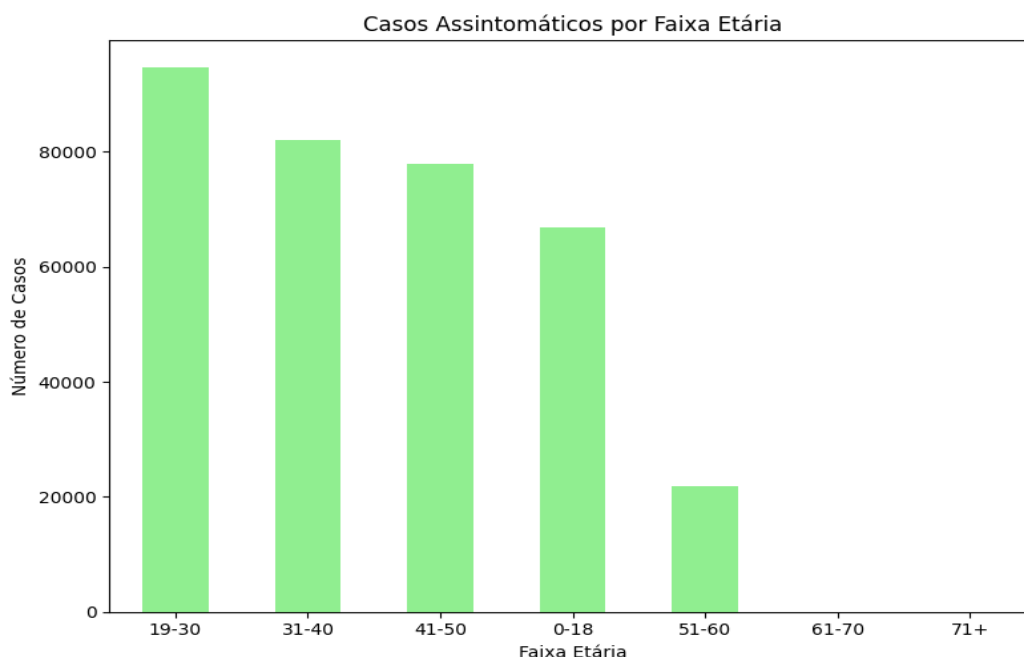


Figura 12 - Frequência de casos assintomáticos divididos por faixa etária.

### Correlações:

A maior concentração de casos assintomáticos está na faixa etária de **19 a 30 anos** (19%), seguida pelas faixas de **31-40** (16%) e **41-50** (15%). **Jovens e adultos até 50 anos** representam a maioria dos casos assintomáticos.

A ausência de casos reportados nas faixas de **61-70 anos** e **71+ anos** pode indicar que, em grupos etários mais avançados, há uma tendência de os casos serem sintomáticos, ou pode refletir falhas de registro/testagem ou dificuldades desses grupos em acessar os serviços de saúde para testagem. Além disso, é possível que idosos, sendo mais vulneráveis, tenham mais probabilidade de apresentar sintomas graves e, portanto, não estarem entre os casos assintomáticos.

## 6. ESBOÇO DATA STORYTELLING

Narrativa para gravação final:

### 1. Introdução (Setup) – 1º Integrante (aprox. 1 min)

**Objetivo:** Apresentar o grupo, o tema do projeto e o contexto do estudo.

- **Abertura:**

- Apresentação dos membros da equipe e RA.
- Contextualização da empresa: Somos a empresa Science for Life SA, uma organização dedicada à pesquisa em saúde pública e análise de dados epidemiológicos. Nossa missão é contribuir para a gestão de risco de doenças infecciosas por meio da análise de dados, e nossa visão é democratizar essas informações através de uma plataforma open source.
- Descrição do projeto: Nosso projeto é focado na análise comparativa da incidência de COVID-19 entre estados mais e menos populosos nas diferentes regiões do Brasil, utilizando dados do sistema e-SUS Notifica.

- **Apresentação do Problema:**

- Problema central: Buscamos entender como a densidade populacional afeta a evolução dos casos de COVID-19. Observamos que a pandemia teve impactos variados entre os estados e, portanto, queremos investigar se os estados mais populosos conseguiram responder de forma mais eficaz em comparação aos menos populosos.
- Importância da pesquisa: A análise é fundamental para identificar padrões que possam informar políticas públicas e estratégias de saúde para futuras crises sanitárias, contribuindo assim para a melhoria da qualidade de vida da população.

- **Protagonista e Contexto:**

- O protagonista da nossa análise é a população brasileira, dividida entre estados com diferentes densidades populacionais, que enfrentaram desafios distintos durante a pandemia.

**Transição para a próxima seção:** "Agora vamos entrar na análise dos dados e mostrar os resultados que encontramos sobre a evolução da COVID-19 nas diferentes regiões."

---

### 2. Conflito (Problema de Pesquisa) – 2º Integrante (aprox. 1 min)

**Objetivo:** Explicar o desafio central, descrevendo as desigualdades entre as regiões.

- **Descrição do Conflito:**

- O conflito que nossa análise aborda é a desigualdade na evolução da COVID-19 entre os estados mais e menos populosos. Observamos que, enquanto estados como **São Paulo (SP)** tiveram um volume muito maior de casos, estados menos populosos, como **Tocantins (TO)** e **Espírito Santo (ES)**, mostraram uma evolução mais lenta e fragmentada da pandemia.

- **Proposta Analítica:**

- Nossa proposta foi realizar uma análise exploratória dos dados de notificação de casos de COVID-19. Usamos seis análises para entender a evolução dos casos, o impacto da vacinação, os sintomas nos óbitos, a demografia dos infectados e a participação dos profissionais de saúde. Isso nos permitiu identificar padrões específicos para cada tipo de estado e apontar soluções que podem ser aplicadas em futuras crises sanitárias.

**Transição para a próxima seção:** “Agora, vamos explorar cada uma das análises detalhadamente.”

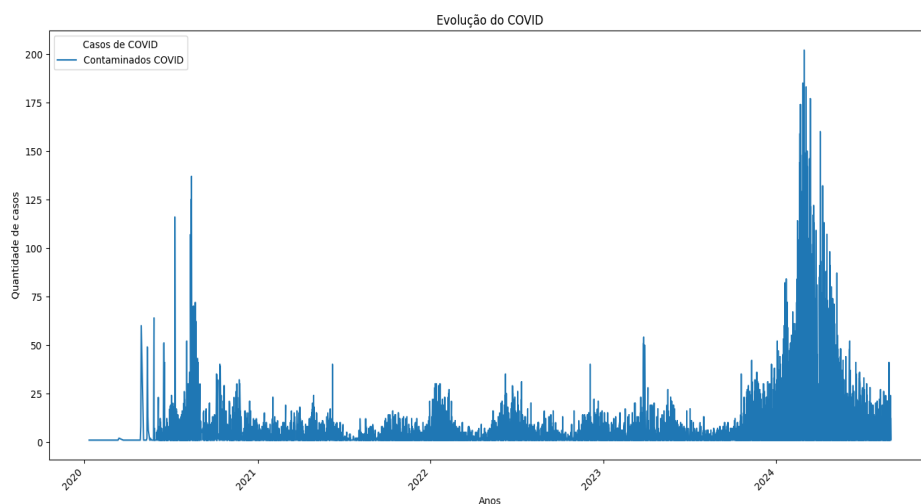
---

### 3. Ponto de Virada (A Análise) – 3º e 4º Integrantes (aprox. 2 min)

**Objetivo:** Apresentar os principais resultados da análise de dados, detalhando cada uma das seis análises realizadas no projeto.

#### Parte 1 - Análises sobre Evolução e Vacinação – 3º Integrante

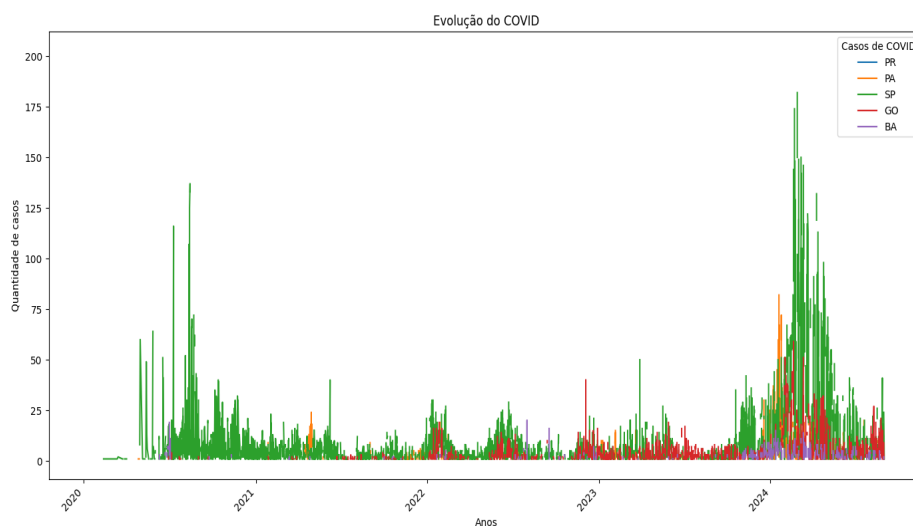
- **Análise 01 (Comparação da Evolução da COVID-19 entre Estados Mais e Menos Populosos):**
  - Nossa primeira análise focou na comparação da evolução da COVID-19 entre estados mais e menos populosos. Como mostrado na **Figura 01**, o estado de **São Paulo (SP)**, que é mais populoso, apresentou um número muito maior de casos e picos durante os períodos de maior transmissão. Já estados menos populosos, como **Tocantins (TO)**, apresentaram uma curva de crescimento mais gradual e com menor número de casos.



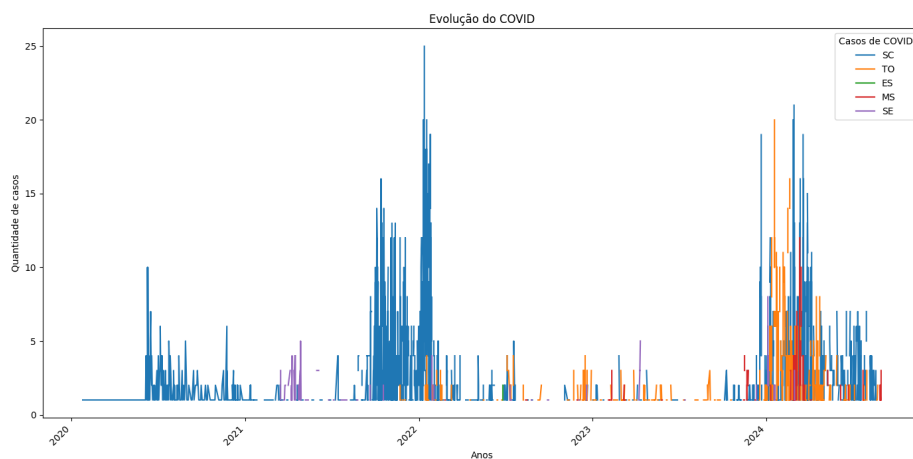
**Figura 01** – Distribuição total das vacinações do Covid-19 de 2020 a 2024.

- **Explicação Gráfica (Figura 02 e 03):**

- A visualização que desenvolvemos mostra claramente essa diferença de evolução. O estado de São Paulo, sendo mais populoso, registrou um maior número de casos e conseguiu distribuir as vacinas de forma mais concentrada em períodos de maior demanda, enquanto estados menos populosos tiveram um ritmo de evolução mais lento.



**Figura 02** – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados mais populosos

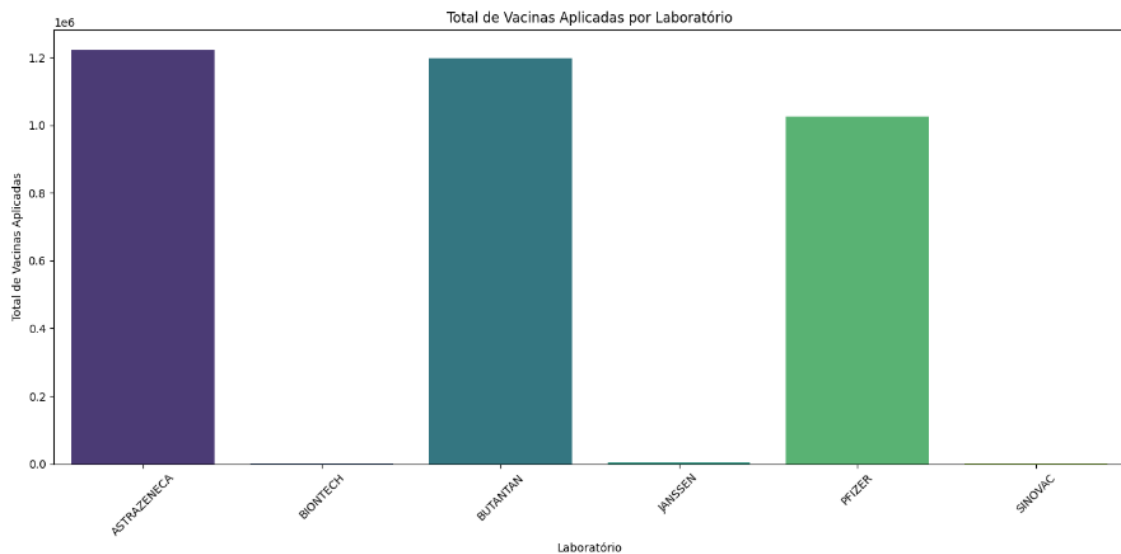


**Figura 03** – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados menos populosos.

- **Análise 02 (Volumetria Geral das Vacinas por Laboratório):**

- Nossa segunda análise focou na volumetria geral das vacinas aplicadas por laboratório. Como mostrado na **Figura 04**, as vacinas da AstraZeneca e do Instituto Butantan foram as mais utilizadas, com a Pfizer também tendo uma

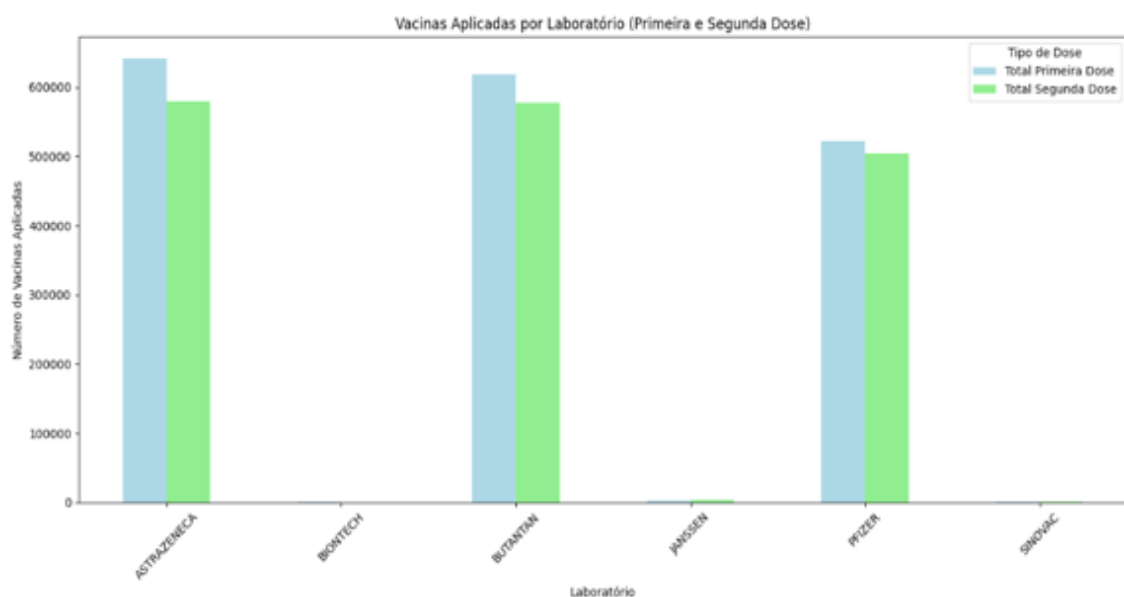
participação significativa. Esses laboratórios lideraram a distribuição de vacinas, contribuindo de forma expressiva para o combate à pandemia.



**Figura 04** - Volumetria total, em milhões, de vacinas aplicadas por laboratório.

○ **Explicação Gráfica (Figura 05):**

- Esses gráficos mostram a distribuição das vacinas por laboratório, com AstraZeneca e Butantan sendo os principais fornecedores. A distribuição consistente entre a primeira e a segunda dose também reflete o sucesso da campanha de vacinação, garantindo que os pacientes completassem o esquema vacinal.



**Figura 05** - Volumetria total de vacinas aplicadas por laboratório distribuídos por tipo de dose.



**Transição para a próxima parte:** “Agora vamos falar sobre a demografia dos infectados e os sintomas nos casos de óbito.”

---

## Parte 2 - Demografia, Sintomas e Impacto nos Profissionais de Saúde – 4º Integrante

- **Análise 03 (Comparação entre Sexo e Idade dos Pacientes):**
  - Nossa terceira análise explorou as diferenças entre os sexos e as faixas etárias dos pacientes com COVID-19. Os resultados mostram que mulheres foram as mais afetadas em termos de notificações, especialmente nos estados como **Tocantins (TO)** e **Mato Grosso do Sul (MS)**, onde representaram mais de 60% dos casos.
  - **Explicação Gráfica (Figura 06 e Figura 07):**
    - O gráfico da **Figura 06** mostra a distribuição etária geral dos pacientes, com uma predominância de jovens adultos. O gráfico da **Figura 07** destaca a predominância feminina em todos os estados analisados.

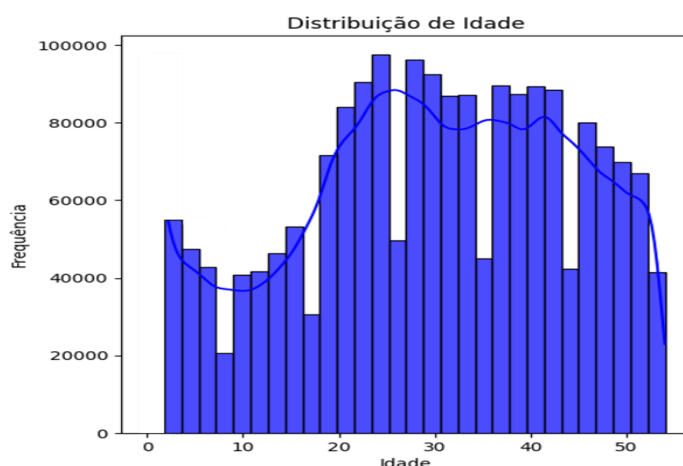
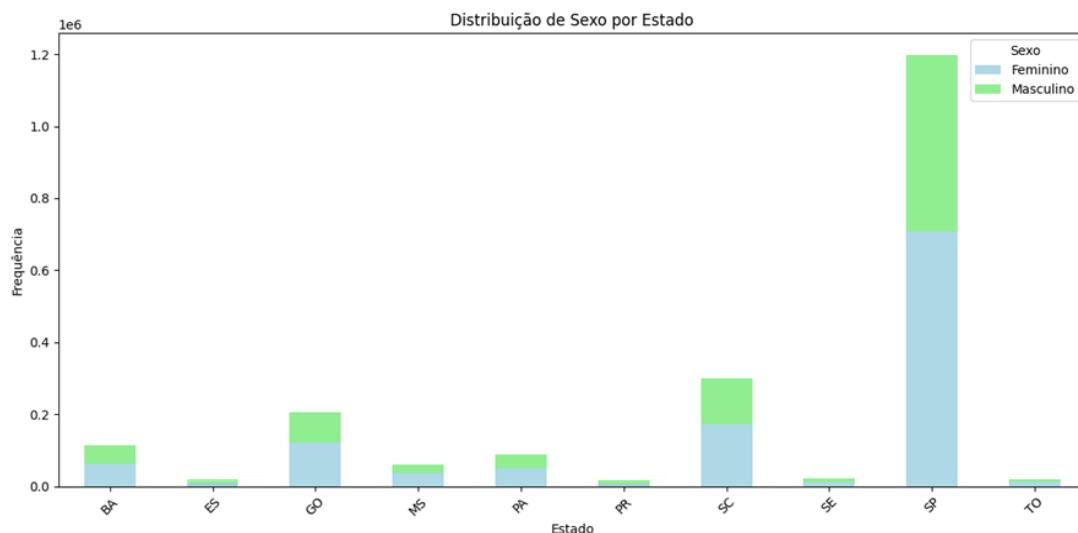


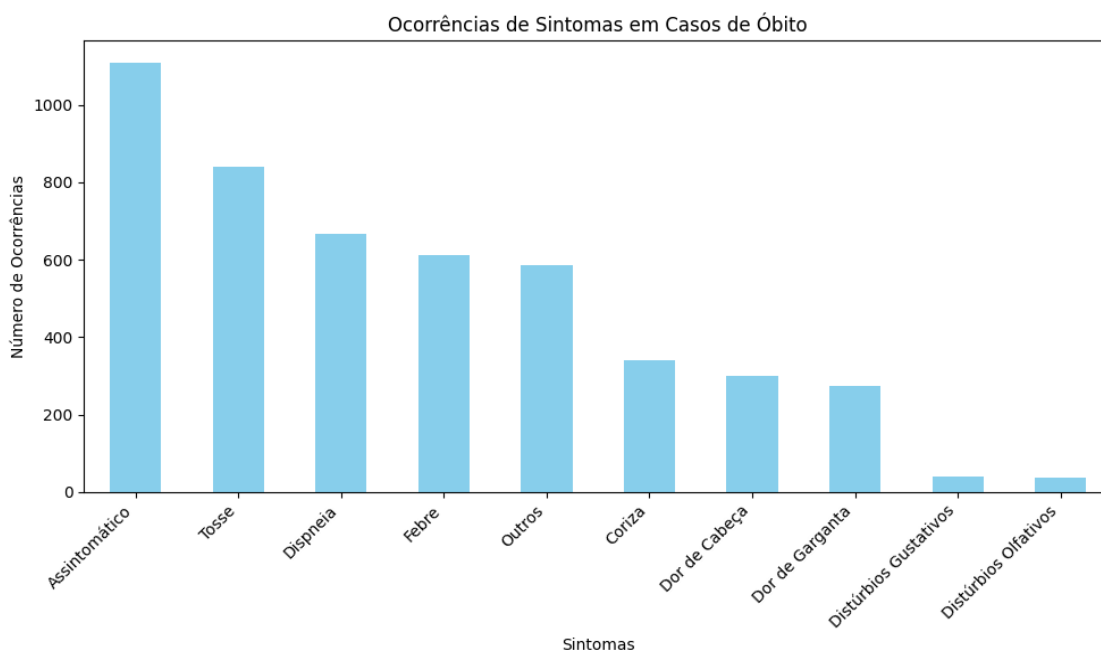
Figura 06 - Distribuição etária geral dos pacientes notificados com Covid-19 em todos os estados do dataset.



**Figura 07** - Distribuição percentual do sexo (Feminino e Masculino) entre os pacientes notificados com Covid-19 em diferentes estados brasileiros.

- **Análise 04 (Ocorrência de Sintomas em Casos de Óbito):**

- Na análise dos sintomas em casos de óbito, observamos que muitos dos óbitos ocorreram entre pessoas assintomáticas, sugerindo que esses indivíduos não procuraram tratamento a tempo. Além disso, sintomas como tosse e dispneia foram os mais comuns entre os óbitos.
- **Explicação Gráfica (Figura 08):**
  - No gráfico da Figura 08 mostra que, apesar de sintomas como tosse e dispneia serem frequentes em óbitos, mais de 23% dos óbitos ocorreram em pacientes que não apresentaram sintomas visíveis.



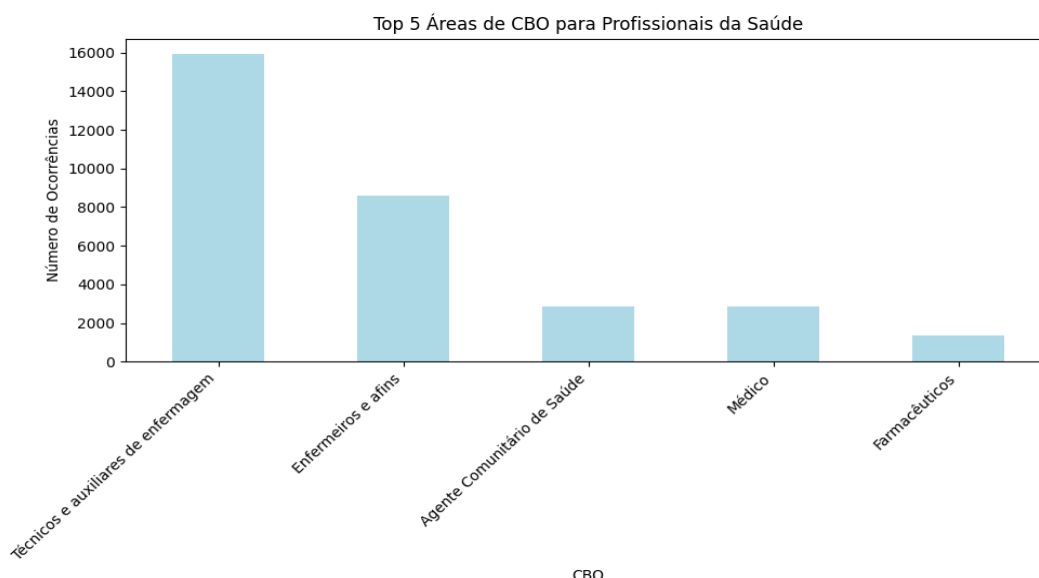
**Figura 08** - Frequência dos sintomas em casos de óbito.

**Transição para a próxima parte:** “Agora vamos analisar o impacto nos profissionais de saúde e os casos assintomáticos.”

### Parte 3 - Profissionais de Saúde e Casos Assintomáticos – 4º Integrante

- **Análise 05 (Profissionais da Saúde):**

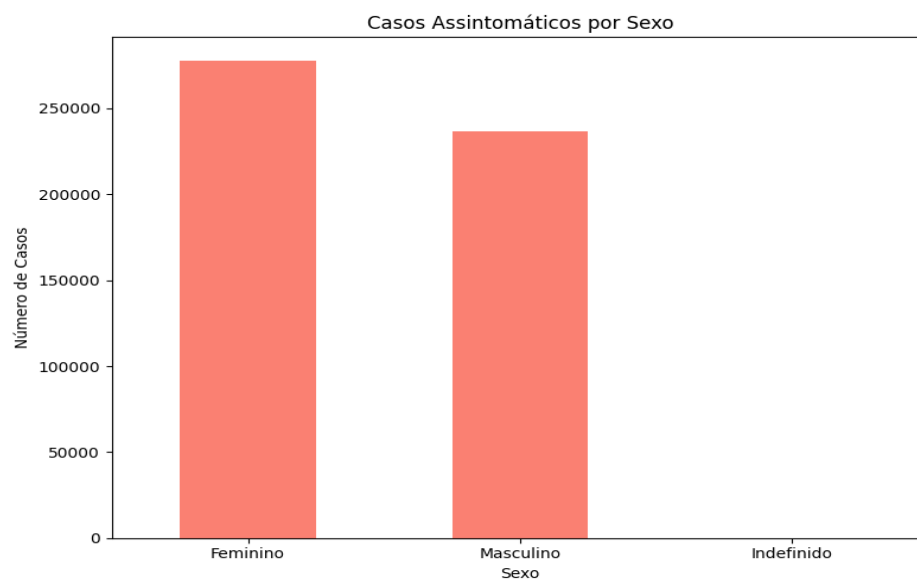
- Na quinta análise, exploramos a quantidade e a área de atuação dos profissionais da saúde infectados. Descobrimos que os técnicos e auxiliares de enfermagem foram os mais impactados, representando mais de 31% dos profissionais da saúde infectados, seguidos por enfermeiros, agentes e médicos.
- **Explicação Gráfica (Figura 08):**
  - A Figura 09 mostra claramente como os profissionais de saúde da linha de frente, como os técnicos de enfermagem, foram os mais expostos e infectados durante a pandemia.



**Figura 09** - Frequência que profissionais da área da saúde e suas áreas aparecem na base de dados.

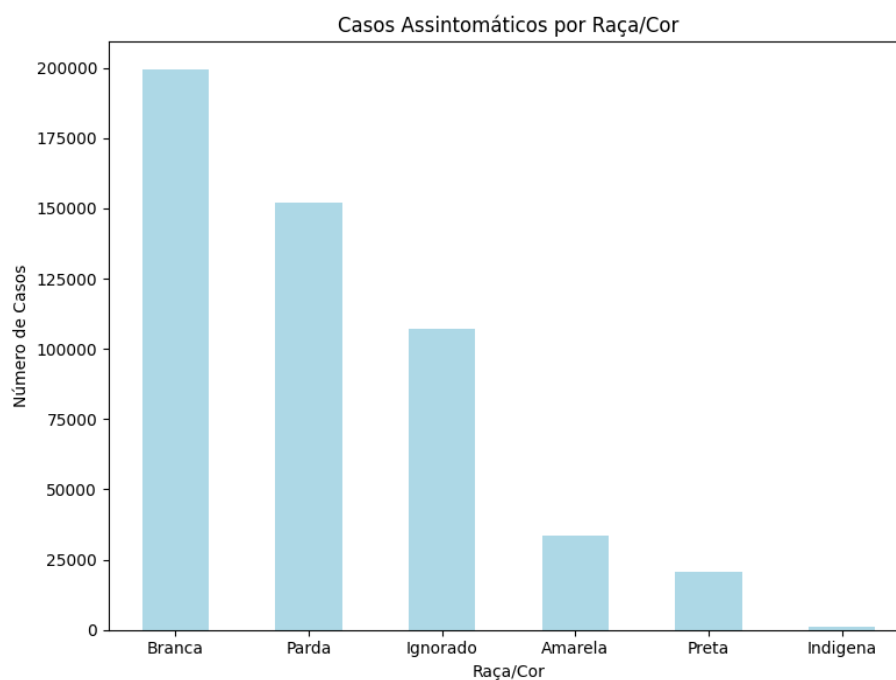
- **Análise 06 (Casos Assintomáticos):**

- Por fim, analisamos os casos assintomáticos. A maioria dos casos assintomáticos foi reportada entre pessoas jovens, com predominância na faixa etária de 19 a 30 anos. Observamos também uma leve predominância de casos assintomáticos em pessoas do sexo feminino.
- **Explicação Gráfica (Figura 10, Figura 11 e Figura 12):**
  - A Figura 10 mostra a diferença de 54% (feminino) para 46% (masculino) indica uma pequena concentração maior de casos assintomáticos em mulheres.



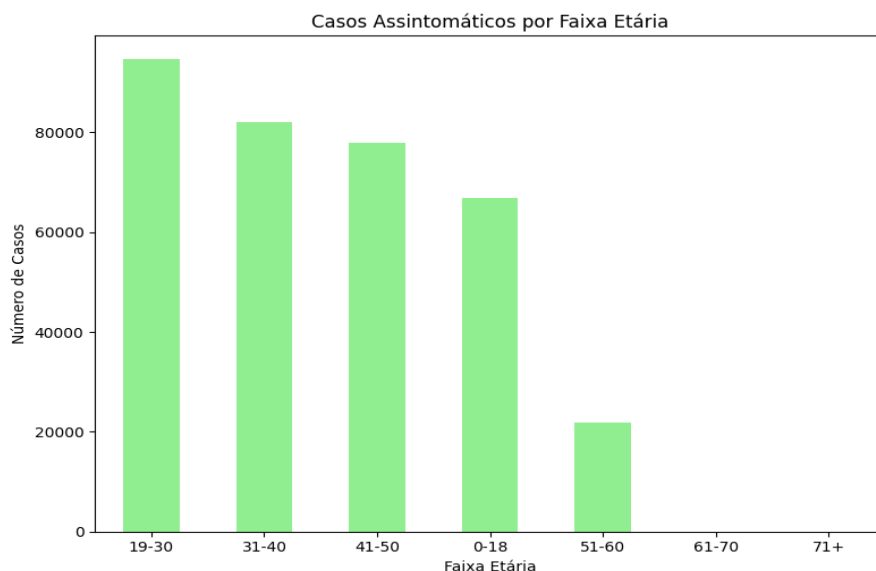
**Figura 10** - Frequência casos assintomático dividido por sexo.

- A Figura 11 mostra que a maioria dos casos assintomáticos é reportada entre pessoas Brancas (41%) e Pardas (31%). Há também um número significativo de dados com Raça/Cor Ignorada (22%), o que pode dificultar uma análise mais precisa. A população Amarela e Preta apresentam casos assintomáticos em proporções menores, e o número de Indígenas é bastante baixo.



**Figura 11** - Frequência de casos assintomáticos divididos por Raça/Cor.

- A Figura 12 mostra a concentração dos casos assintomáticos por faixa etária, evidenciando que os jovens, principalmente entre 19 e 30 anos, foram os mais afetados.



**Figura 12** - Frequência de casos assintomáticos divididos por faixa etária.

**Transição para os resultados pretendidos:** “Agora vamos discutir as principais conclusões e o impacto dessas análises.”

#### 4. Resolução (Resultados Pretendidos e Conclusão) – 2º Integrante (aprox. 1 min)

- **Discussão dos Resultados Pretendidos:**

- Nossas análises sugerem que os estados mais populosos, apesar de enfrentarem uma maior carga de casos, conseguiram distribuir melhor os recursos de saúde e as vacinas. Os estados menos populosos, por outro lado, enfrentaram maiores dificuldades logísticas, resultando em uma resposta mais lenta e menos eficiente.

- **Conclusão:**

- O estudo revela que os dados de entrada são fundamentais para os resultados obtidos nas análises, influenciando diretamente as conclusões. A defasagem das bases utilizadas pela empresa, que contém muitos registros nulos e não padronizados, é um ponto crítico que pode comprometer a validade das análises. A contagem por estado, que não reflete a população total real, limita a precisão das inferências que podem ser feitas.
- Com isso, esperamos que nossos insights possam contribuir para a criação de políticas públicas mais eficazes, que levem em consideração as diferenças regionais no Brasil. Entender como a pandemia impactou

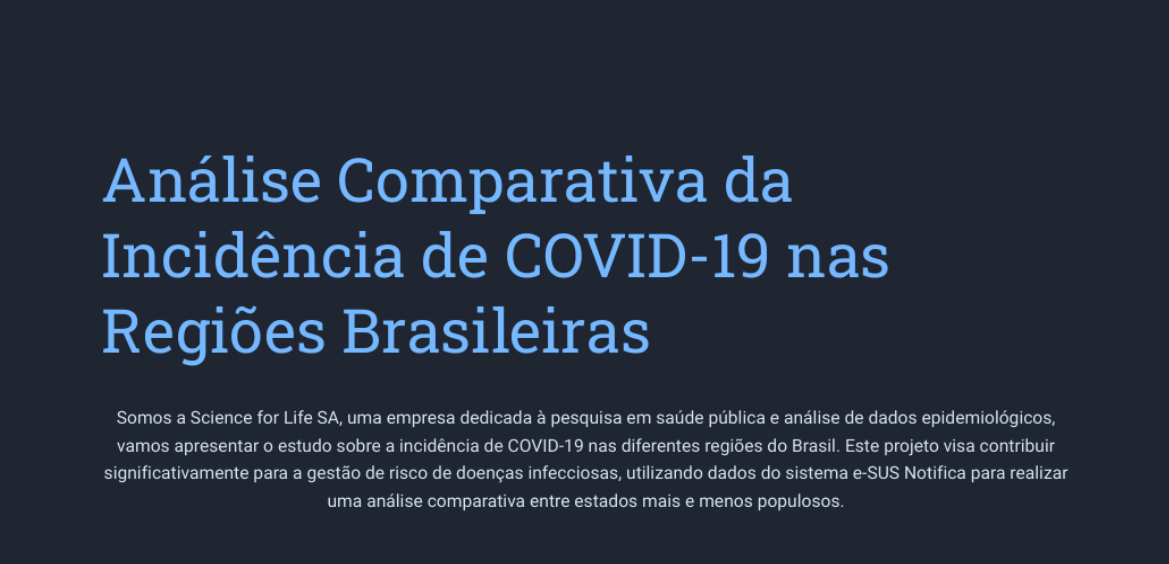
diferentes populações nos ajuda a planejar uma resposta mais eficiente para futuras crises de saúde pública.

**Encerramento:** “Agradecemos pela atenção. Todos os detalhes e códigos do nosso projeto estão disponíveis no GitHub para consulta [link]”.

## 7. APRESENTAÇÃO DO PROJETO:

**VÍDEO:** [https://www.youtube.com/watch?v=ox\\_e85v5HpA](https://www.youtube.com/watch?v=ox_e85v5HpA)

## 8. APRESENTAÇÃO DOS SLIDES:



**Análise Comparativa da Incidência de COVID-19 nas Regiões Brasileiras**

Somos a Science for Life SA, uma empresa dedicada à pesquisa em saúde pública e análise de dados epidemiológicos, vamos apresentar o estudo sobre a incidência de COVID-19 nas diferentes regiões do Brasil. Este projeto visa contribuir significativamente para a gestão de risco de doenças infecciosas, utilizando dados do sistema e-SUS Notifica para realizar uma análise comparativa entre estados mais e menos populosos.



**Nossa Missão**

Democratizar informações através de uma plataforma open source, esta pesquisa é fundamental para identificar padrões que podem informar políticas públicas e estratégias de saúde em futuras crises sanitárias. A análise explorará a evolução da doença, o impacto da vacinação, aspectos demográficos e a participação dos profissionais de saúde na resposta à pandemia.

O protagonista da nossa análise é a população brasileira.

# Desigualdade na Evolução da COVID-19

## Estados Mais Populosos

São Paulo, como exemplo de estado altamente populoso, enfrentou um volume significativamente maior de casos de COVID-19. A densidade populacional e a interconectividade das grandes cidades contribuíram para uma rápida disseminação do vírus. No entanto, esses estados também demonstraram maior capacidade de mobilização de recursos e infraestrutura de saúde.

## Estados Menos Populosos

Tocantins, representando estados menos populosos, experimentou uma evolução mais lenta da pandemia. Fatores como menor densidade populacional e relativo isolamento geográfico podem ter contribuído para uma progressão mais gradual dos casos. Contudo, esses estados enfrentaram desafios únicos, como limitações na infraestrutura de saúde e na distribuição de recursos.



## Proposta Analítica

### Análise Exploratória

Nossa proposta foi realizar uma análise exploratória dos dados de notificação de casos de COVID-19. Usamos seis análises para entender a evolução dos casos, o impacto da vacinação, os sintomas nos óbitos, a demografia dos infectados e a participação dos profissionais de saúde. Isso nos permitiu identificar padrões específicos para cada tipo de estado e apontar soluções que podem ser aplicadas em futuras crises sanitárias.

### Evolução Temporal

Compararemos a evolução da pandemia entre estados com diferentes níveis de população.



## Proposta Analítica

### Impacto da Vacinação

Analisaremos o impacto da vacinação na incidência de COVID-19. Consideraremos o tipo de vacina, a cobertura vacinal por região.

### Análise Demográfica

Investigaremos o impacto da pandemia em diferentes grupos demográficos, como idade, sexo e condições socioeconômicas. Analisaremos a taxa de mortalidade, a gravidade dos sintomas.

## Evolução e Vacinação

### Evolução da COVID-19 entre Estados Mais e Menos Populosos

Nossa primeira análise focou na comparação da evolução da COVID-19 entre estados mais e menos populosos. Como mostrado na **Figura 01**, o estado de **São Paulo (SP)**, que é mais populoso, apresentou um número muito maior de casos e picos durante os períodos de maior transmissão. Já estados menos populosos, como **Tocantins (TO)**, apresentaram uma curva de crescimento mais gradual e com menor número de casos.

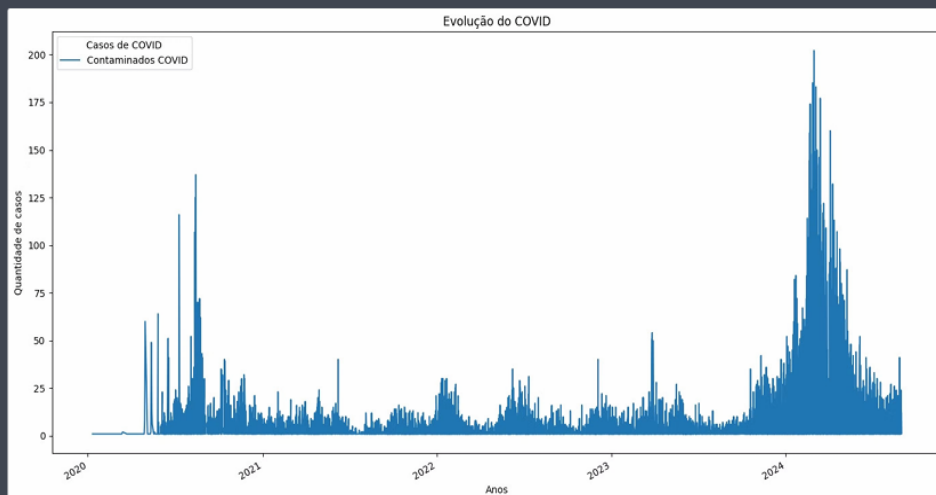


Figura 01 – Distribuição total das vacinações do Covid-19 de 2020 a 2024.

## Evolução e Vacinação

### Distribuição de Vacinas

A visualização que desenvolvemos mostra claramente essa diferença de evolução. O estado de São Paulo, sendo mais populoso, registrou um maior número de casos e conseguiu distribuir as vacinas de forma mais concentrada em períodos de maior demanda, enquanto estados menos populosos tiveram um ritmo de evolução mais lento.

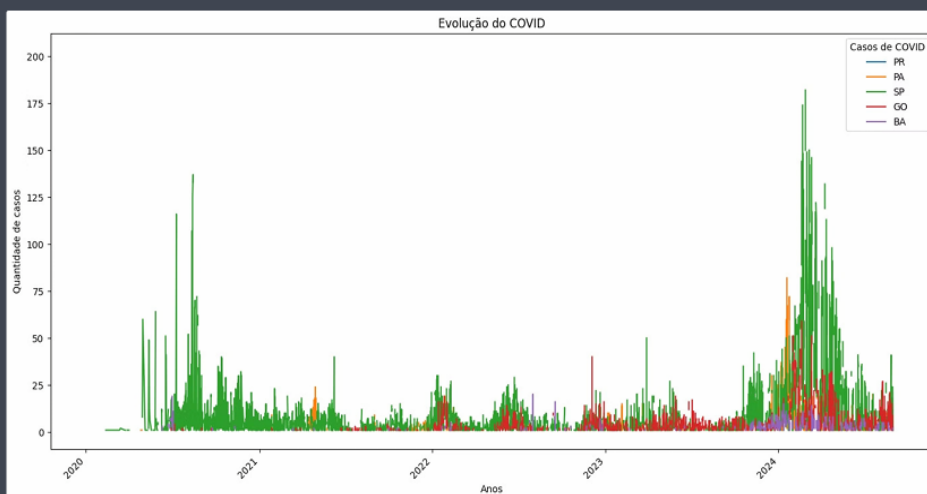


Figura 02 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados mais populosos.



## Evolução e Vacinação

### Distribuição de Vacinas

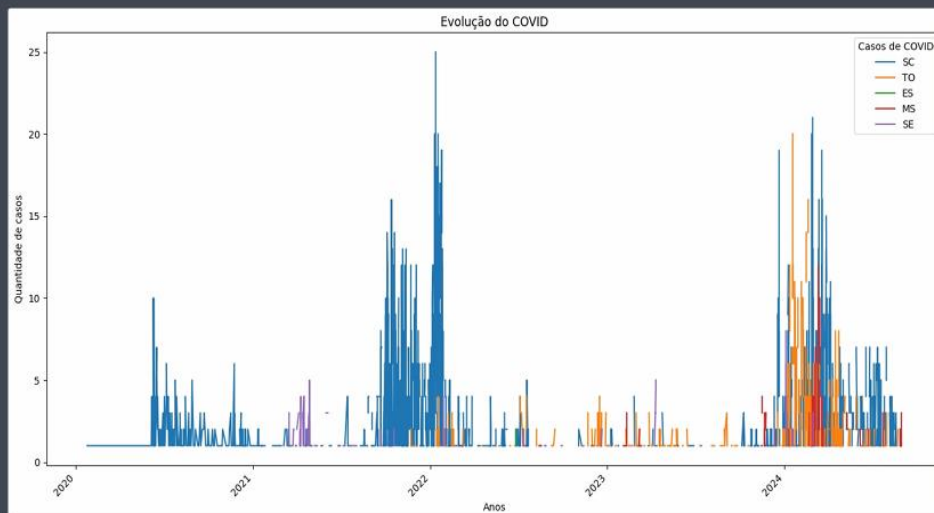


Figura 03 – Distribuição das vacinações do Covid-19 de 2020 a 2024 - 5 estados menos populosos.

## Evolução e Vacinação

### Volumetria Geral das Vacinas por Laboratório

Nossa segunda análise focou na volumetria geral das vacinas aplicadas por laboratório. Como mostrado na **Figura 04**, as vacinas da AstraZeneca e do Instituto Butantan foram as mais utilizadas, com a Pfizer também tendo uma participação significativa. Esses laboratórios lideraram a distribuição de vacinas, contribuindo de forma expressiva para o combate à pandemia.

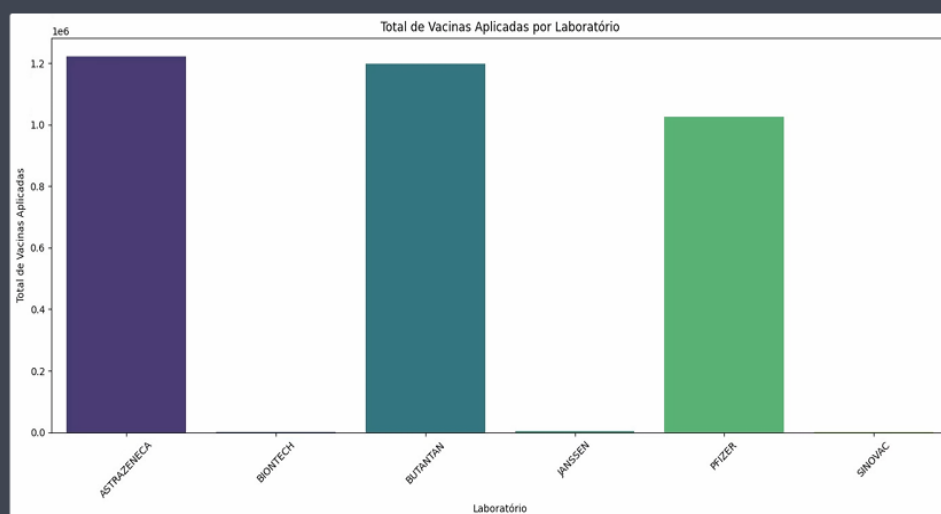


Figura 04 – Volumetria total, em milhões, de vacinas aplicadas por laboratório.

## Evolução e Vacinação

### Volumetria Geral das Vacinas por Laboratório

Esses gráficos mostram a distribuição das vacinas por laboratório, com AstraZeneca e Butantan sendo os principais fornecedores. A distribuição consistente entre a primeira e a segunda dose também reflete o sucesso da campanha de vacinação, garantindo que os pacientes completassem o esquema vacinal.

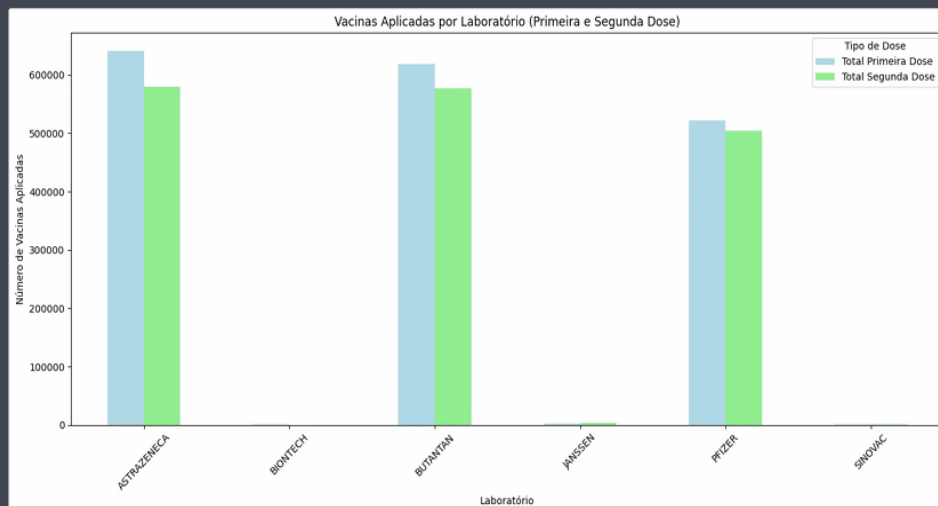


Figura 05 - Volumetria total de vacinas aplicadas por laboratório distribuídos por tipo de dose.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Comparação do Sexo e Idade dos Pacientes

Nossa terceira análise explorou as diferenças entre os sexos e as faixas etárias dos pacientes com COVID-19. Os resultados mostram que mulheres foram as mais afetadas em termos de notificações, especialmente nos estados como Tocantins (TO) e Mato Grosso do Sul (MS), onde representaram mais de 60% dos casos.

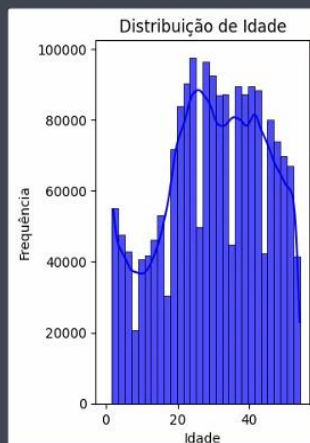
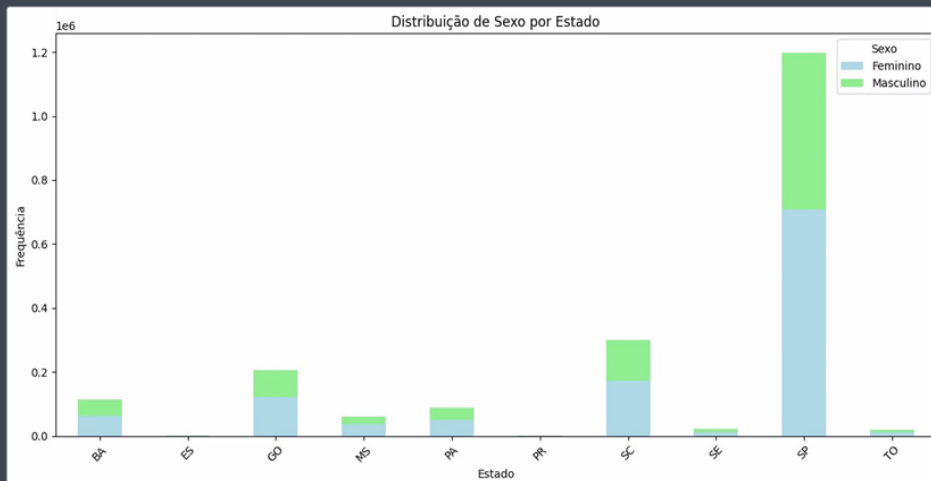


Figura 06 - Distribuição etária geral dos pacientes notificados com Covid-19 em todos os estados do dataset.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Comparação do Sexo e Idade dos Pacientes

O gráfico da **Figura 06** mostra a distribuição etária geral dos pacientes, com uma predominância de jovens adultos. O gráfico da **Figura 07** destaca a predominância feminina em todos os estados analisados.



**Figura 07** - Distribuição percentual do sexo (Feminino e Masculino) entre os pacientes notificados com Covid-19 em diferentes estados brasileiros.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Ocorrência de Sintomas em Casos de Óbito

#### Óbitos entre Assintomáticos

Um número significativo de óbitos ocorreu entre indivíduos inicialmente assintomáticos, destacando a importância do diagnóstico precoce e monitoramento contínuo, mesmo em casos aparentemente leves.

#### Variação de Sintomas

Observou-se que apesar de sintomas como tosse e dispnéia serem frequentes em óbitos, mais de 23% dos óbitos ocorreram em pacientes que não apresentaram sintomas visíveis.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Ocorrência de Sintomas em Casos de Óbito

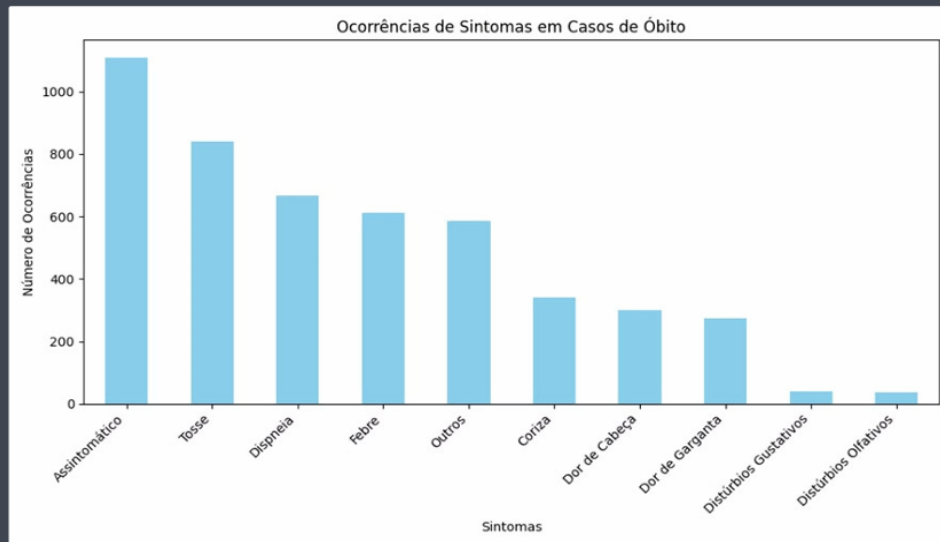


Figura 08 - Frequência dos sintomas em casos de óbito.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Profissionais de Saúde

Na quinta análise, exploramos a quantidade e a área de atuação dos profissionais da saúde infectados. Descobrimos que os técnicos e auxiliares de enfermagem foram os mais impactados, representando mais de 31% dos profissionais da saúde infectados, seguidos por enfermeiros, agentes e médicos.

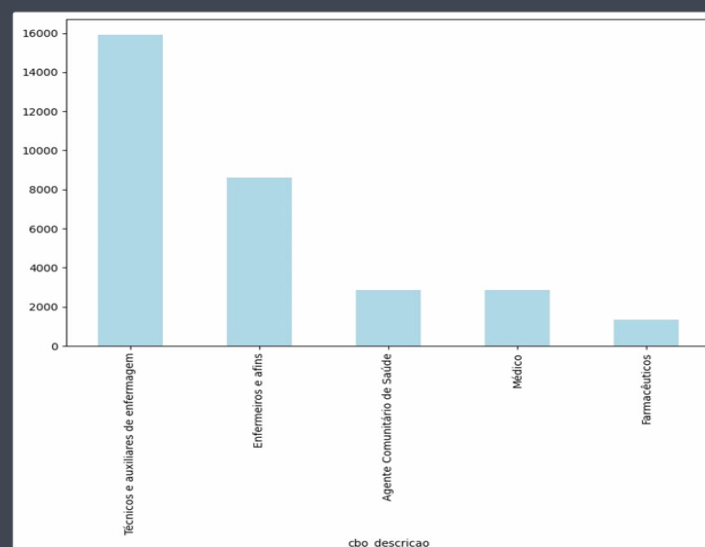


Figura 09 - Mostra claramente como os profissionais de saúde da linha de frente, como os técnicos de enfermagem, foram os mais expostos e infectados durante a pandemia.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Casos Assintomáticos - Predominância em Jovens

A maioria dos casos assintomáticos foi reportada entre jovens, especialmente na faixa etária de 19 a 30 anos. Isso sugere que este grupo pode ter sido um vetor importante na transmissão silenciosa do vírus.

### Gênero e Assintomáticos

Observou-se uma predominância feminina entre os casos assintomáticos, alinhando-se com a tendência geral de maior incidência em mulheres. Isso pode estar relacionado a fatores hormonais ou a diferenças nos padrões de testagem.

### Impacto na Transmissão

A alta proporção de casos assintomáticos destacou a importância de medidas preventivas universais, como o uso de máscaras e distanciamento social, para conter a propagação do vírus mesmo entre indivíduos aparentemente saudáveis.



Made with Gamma

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Casos Assintomáticos

A Figura 10 mostra a diferença de 54% (feminino) para 46% (masculino) indica uma pequena concentração maior de casos assintomáticos em mulheres.

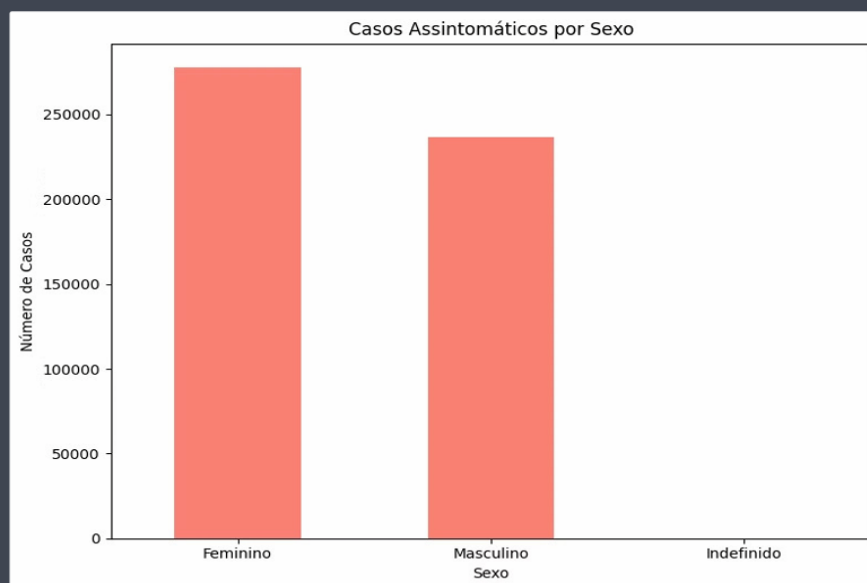


Figura 10 - Frequência casos assintomático dividido por sexo.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Casos Assintomáticos

A Figura 11 mostra que a maioria dos casos assintomáticos é reportada entre pessoas Brancas (41%) e Pardas (31%). Há também um número significativo de dados com Raça/Cor Ignorada (22%), o que pode dificultar uma análise mais precisa. A população Amarela e Preta apresentam casos assintomáticos em proporções menores, e o número de Indígenas é bastante baixo.

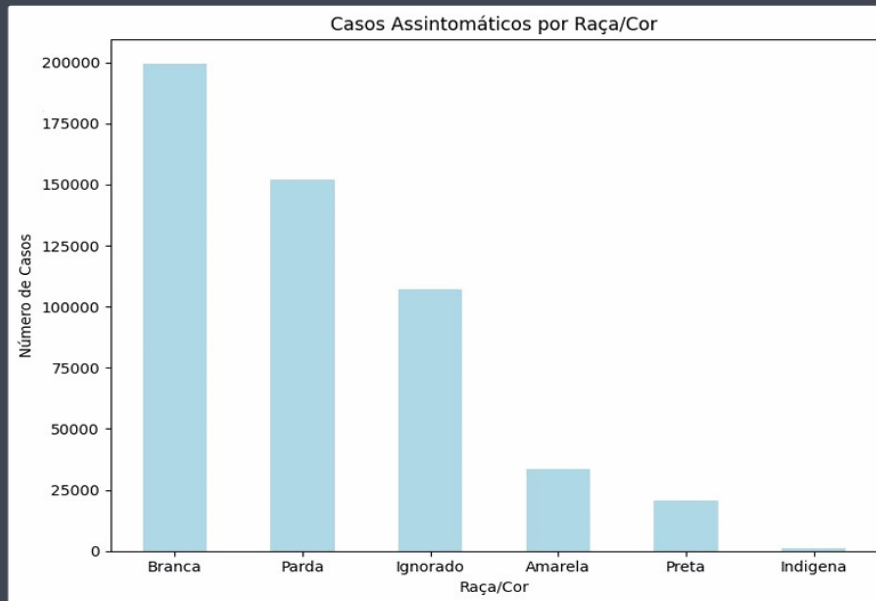


Figura 11 - Frequência de casos assintomáticos divididos por Raça/Cor.

## Demografia, Sintomas e Impacto nos Profissionais de Saúde

### Casos Assintomáticos

A Figura 12 mostra a concentração dos casos assintomáticos por faixa etária, evidenciando que os jovens, principalmente entre 19 e 30 anos, foram os mais afetados.

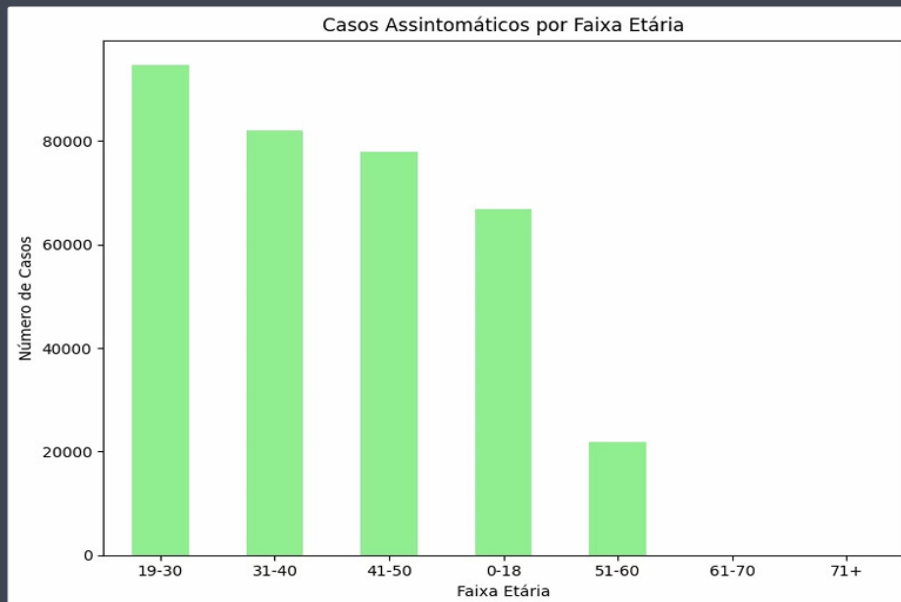


Figura 12 - Frequência de casos assintomáticos divididos por faixa etária.

## Conclusões e Implicações Futuras



### Qualidade dos Dados

A análise ressaltou a importância crítica da qualidade e padronização dos dados de entrada. A defasagem e inconsistências nos dados limitaram a precisão e validade de algumas análises, destacando a necessidade de melhorias nos sistemas de notificação e coleta de dados epidemiológicos.



### Políticas Públicas

Os insights gerados por este estudo têm o potencial de informar políticas públicas mais eficazes e adaptadas às realidades regionais do Brasil. Recomenda-se uma abordagem diferenciada para estados mais e menos populosos, considerando suas características únicas de infraestrutura e demografia.



### Pesquisas Futuras

Este projeto abre caminho para pesquisas futuras mais aprofundadas, incluindo análises de longo prazo sobre o impacto da COVID-19 na saúde pública brasileira e estudos comparativos com outras emergências sanitárias para melhorar a preparação para futuras pandemias.

## **9. CONCLUSÃO**

Com esse estudo, podemos concluir que os dados de entrada são extremamente relevantes para os resultados obtidos nas análises, influenciando diretamente as conclusões a serem tiradas dos mesmos. As bases utilizadas pela empresa estão, notadamente, defasadas, apresentando uma grande quantidade de registros nulos e não padronizados. Além disso, a contagem por estado não reflete a população total real de cada um deles. Entretanto, conseguimos observar dados que refletem a realidade, como a distribuição das síndromes gripais e seus “picos”, além do fato de que mulheres são mais proativas em cuidar de sua saúde. Também verificamos que diferentes categorias de profissionais da saúde foram impactadas de maneiras diversas durante a pandemia, dependendo da natureza de suas funções e do nível de contato com pacientes infectados. Essas conclusões reforçam a necessidade de estudarmos os dados de saúde e disponibilizá-los, para que sejam utilizados para estudos preditivos (ciência de dados e machine learning), possibilitando uma gestão de crise mais assertiva a cada surgimento de novas doenças.



## 10. REFERÊNCIAS

### Link do Vídeo de apresentação:

[https://www.youtube.com/watch?v=ox\\_e85v5HpA](https://www.youtube.com/watch?v=ox_e85v5HpA)

### Link do repositório:

<https://github.com/KellyVasconcellos/projeto-aplicado-l>

### Link do dataset unificado:

[https://drive.google.com/drive/u/2/folders/1L4M5FezfxvhBohmyX8\\_DO7y\\_P0E1f17I](https://drive.google.com/drive/u/2/folders/1L4M5FezfxvhBohmyX8_DO7y_P0E1f17I)

### Link do dataset por estado:

Síndromes gripais leve:

<https://opendatasus.saude.gov.br/dataset/notificacoes-de-sindrome-gripal-leve-2024>

### Links de reportagens que sustentam nossos insights:

**GOVERNO DO ESTADO DE SÃO PAULO.** Saiba quais são os intervalos entre doses das vacinas anticovid. Capital SP, 2021. Disponível em: <https://capital.sp.gov.br/w/noticia/saiba-quais-sao-os-intervalos-entre-doses-das-vacinas-anticovid>. Acesso em: 28 set. 2024.

**SENADO FEDERAL.** Brasil poderia ter sido primeiro do mundo a vacinar, afirma Dimas Covas à CPI. Senado, 27 maio 2021. Disponível em: <https://www12.senado.leg.br/noticias/materias/2021/05/27/brasil-poderia-ter-sido-primeiro-do-mundo-a-vacinar-afirma-dimas-covas-a-cpi>. Acesso em: 28 set. 2024.

**ODIA.** Dados do SUS mostram que as mulheres buscam mais atendimento médico que os homens. O Dia, 2022. Disponível em: <https://odia.ig.com.br/brasil/2022/07/6443730-dados-do-sus-mostram-que-as-mulheres-buscam-mais-atendimento-medico-que-os-homens.html>. Acesso em: 28 set. 2024.

**CNN BRASIL.** Homens contraem e transmitem mais COVID-19 do que as mulheres, diz estudo. CNN Brasil, 2021. Disponível em: <https://www.cnnbrasil.com.br/saude/homens-contraem-e-transmitem-mais-covid-19-do-que-as-mulheres-diz-estudo/>. Acesso em: 28 set. 2024.

**BRASIL DE FATO.** Propagação silenciosa: crianças assintomáticas podem carregar coronavírus por semanas. Brasil de Fato, 31 ago. 2020. Disponível em: <https://www.brasildefato.com.br/2020/08/31/propagacao-silenciosa-criancas-assintomaticas-podem-carregar-coronavirus-por-semanas>. Acesso em: 28 set. 2024.