



**UNIVERSIDADE NOVE DE JULHO - UNINOVE**  
**Projeto Integrador 2024**

**Guilherme Henrique Guerra - RA: 3022107082**

**Levi Lins Ferreira - RA: 3022103320**

**Joao Pedro de Almeida Melo - RA: 3024103016**

**Mortes Covid-19 Estados-Unidos de 2020 - 2023**

**São Paulo**  
**2024**

**Guilherme Henrique Guerra - RA: 3022107082**

**Levi Lins Ferreira - RA: 3022103320**

**Joao Pedro de Almeida Melo - RA: 3024103016**

**Mortes Covid-19 Estados-Unidos de 2020 - 2023**

Projeto apresentado a Universidade Nove de Julho - UNINOVE, como parte dos requisitos obrigatórios para obtenção do título de **TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**.

Prof. Orientador: Edson Melo de Souza, Dr.

**São Paulo**

**2024**

# RESUMO

---

Resumo do Projeto Este projeto de análise de dados em saúde pública, realizado pela Universidade Nove de Julho, tem como principal objetivo analisar dados relacionados à pandemia de COVID-19, com um foco especial nas condições que contribuíram para as mortes decorrentes do vírus entre os anos de 2020 a 2023. O estudo busca apresentar informações relevantes de forma gráfica para compreender a distribuição e o impacto da pandemia nos Estados Unidos. Os dados foram obtidos de repositórios públicos e incluem informações detalhadas sobre mortes, condições de saúde associadas e dados demográficos.

**Palavras-chave:**

# SUMÁRIO

---

<b>1</b>	<b>Introdução</b>	<b>5</b>
<b>2</b>	<b>Fundamentação Teórica</b>	<b>6</b>
2.1	Visão Geral . . . . .	6
2.2	. . . . .	7
<b>3</b>	<b>Metodologia</b>	<b>8</b>
3.1	Visão Geral . . . . .	8
3.2	Ferramentas e Tecnologias Utilizadas . . . . .	9
3.3	Conclusão . . . . .	9
<b>4</b>	<b>Análise dos Resultados</b>	<b>10</b>
4.1	Conclusão . . . . .	11
<b>5</b>	<b>Conclusões</b>	<b>12</b>
	<b>Apêndices</b>	<b>13</b>
	<b>Apêndices</b>	<b>14</b>

# 1 INTRODUÇÃO

---

## Resumo do capítulo

*A pandemia de COVID-19, iniciada em 2020, teve um impacto profundo e duradouro em todo o mundo, afetando milhões de pessoas e sobrecarregando sistemas de saúde em diversos países. Nos Estados Unidos, o número de casos e mortes atingiu níveis alarmantes, destacando a necessidade de uma compreensão aprofundada dos fatores que contribuíram para a mortalidade causada pelo vírus.*

*Este projeto de análise de dados em saúde pública tem como objetivo investigar as condições que contribuíram para as mortes relacionadas à COVID-19 nos Estados Unidos, no período de 2020 a 2023. Utilizando dados obtidos de repositórios públicos, a análise busca identificar padrões e tendências que possam esclarecer a distribuição geográfica das mortes, as condições de saúde associadas e o impacto demográfico da pandemia.*

*Através de uma abordagem visual e gráfica, este estudo pretende fornecer insights valiosos para pesquisadores, profissionais de saúde e formuladores de políticas públicas, auxiliando na elaboração de estratégias mais eficazes para enfrentar futuras crises de saúde pública. A análise detalhada dos dados permitirá uma melhor compreensão das complexidades envolvidas na mortalidade por COVID-19, contribuindo para a construção de um conhecimento mais sólido sobre a pandemia e suas consequências.*

## 2 FUNDAMENTAÇÃO TEÓRICA

---

### Resumo do capítulo

*A análise de dados em saúde pública é uma disciplina essencial para compreender e enfrentar crises sanitárias, como a pandemia de COVID-19. Esta seção apresenta a fundamentação teórica que sustenta a abordagem deste projeto, incluindo conceitos-chave e métodos utilizados..*

### 2.1 VISÃO GERAL

1. **Pandemia de COVID-19** A COVID-19, causada pelo vírus SARS-CoV-2, foi declarada uma pandemia pela Organização Mundial da Saúde (OMS) em março de 2020. Desde então, o vírus se espalhou globalmente, resultando em milhões de casos e mortes. Estudos têm demonstrado que a mortalidade por COVID-19 é influenciada por diversos fatores, incluindo idade, condições de saúde preexistentes e variáveis socioeconômicas.

2. **Análise de Dados em Saúde Pública** A análise de dados em saúde pública envolve a coleta, processamento e interpretação de dados relacionados à saúde de populações. Esse campo utiliza técnicas estatísticas e computacionais para identificar padrões e tendências, permitindo a formulação de políticas de saúde baseadas em evidências. Ferramentas como Python, R e bibliotecas específicas para análise de dados são amplamente utilizadas.

3. **Epidemiologia** A epidemiologia é a ciência que estuda a distribuição e os determinantes de estados ou eventos relacionados à saúde em populações específicas. No contexto da COVID-19, a epidemiologia é crucial para entender como o vírus se espalha, quais grupos são mais vulneráveis e quais intervenções são mais eficazes. Estudos epidemiológicos fornecem dados que são essenciais para análises mais detalhadas.

4. **Condições de Saúde Associadas** Condições de saúde preexistentes, como doenças cardíacas, diabetes, hipertensão e doenças respiratórias, têm sido identificadas como fatores de risco significativos para a mortalidade por COVID-19. A compreensão dessas associações permite uma melhor alocação de recursos de saúde e desenvolvimento de estratégias de prevenção direcionadas.

5. **Demografia** A demografia analisa as características das populações, como idade, gênero, etnia e distribuição geográfica. Esses fatores demográficos desempenham um papel importante na análise da mortalidade por COVID-19, pois diferentes grupos podem apresentar níveis variados de risco e exposição ao vírus.

6. **Visualização de Dados** A visualização de dados é uma técnica fundamental na análise de grandes conjuntos de dados. Ferramentas de visualização permitem a apresentação de informações complexas de maneira clara e intuitiva, facilitando a identificação de padrões e a comunicação de resultados. Gráficos, mapas e dashboards interativos são exemplos de recursos utilizados para essa finalidade.

## 2.2

Referências Teóricas Para a elaboração deste projeto, foram utilizados conceitos e metodologias provenientes de várias fontes teóricas e estudos de caso em saúde pública e epidemiologia. A literatura sobre pandemias, análise de dados e métodos estatísticos forneceu a base necessária para a abordagem adotada neste estudo.

## 3 METODOLOGIA

---

### Resumo do capítulo

*A metodologia deste projeto de análise de dados sobre a mortalidade por COVID-19 nos Estados Unidos de 2020 a 2023 envolve várias etapas, desde a coleta e limpeza de dados até a análise e visualização dos resultados. As etapas a seguir detalham o processo metodológico adotado para alcançar os objetivos do estudo.*

### 3.1 VISÃO GERAL

1. Coleta de Dados Os dados utilizados neste projeto foram obtidos de repositórios públicos confiáveis, como o Centers for Disease Control and Prevention (CDC) e outros bancos de dados governamentais e acadêmicos. Os dados incluem informações detalhadas sobre:

Número de mortes por COVID-19 Condições de saúde preexistentes Dados demográficos (idade, gênero, etnia) Distribuição geográfica das mortes 2. Limpeza e Preparação dos Dados A limpeza e preparação dos dados são etapas críticas para garantir a qualidade e a consistência das análises. O processo incluiu:

Remoção de dados duplicados e inconsistentes Tratamento de valores ausentes Padronização de formatos de dados Categorização de variáveis relevantes (por exemplo, condições de saúde, grupos etários) 3. Análise Descritiva A análise descritiva foi realizada para obter uma compreensão inicial dos dados. Esta etapa envolveu a utilização de técnicas estatísticas para descrever as características principais dos dados coletados, como:

Distribuição das mortes por COVID-19 ao longo do tempo Frequência de condições de saúde preexistentes entre as vítimas Distribuição geográfica das mortes Características demográficas das vítimas 4. Análise de Correlação Para identificar as relações entre as variáveis, foram conduzidas análises de correlação. Isso ajudou a entender como diferentes fatores (por exemplo, condições de saúde preexistentes, idade, localização geográfica) estão associados à mortalidade por COVID-19. Técnicas estatísticas como a correlação de Pearson e Spearman foram utilizadas nesta fase.

5. Modelagem Preditiva Modelos preditivos foram desenvolvidos para identificar fatores de risco significativos e prever padrões de mortalidade. Métodos de aprendizado de máquina, como regressão logística e árvores de decisão, foram aplicados para criar modelos que pudessem prever a probabilidade de morte por COVID-19 com base nas características das vítimas.

6. Visualização de Dados A visualização de dados foi uma parte fundamental da metodologia para apresentar os resultados de forma clara e intuitiva. Foram criados gráficos, mapas e dashboards interativos para ilustrar os principais achados do estudo. Ferramen-



tas como Matplotlib, Seaborn e Plotly foram utilizadas para gerar visualizações de alta qualidade.

7. Validação e Verificação Os resultados das análises e modelos preditivos foram validados e verificados para garantir sua precisão e confiabilidade. Técnicas de validação cruzada e avaliação de desempenho dos modelos (como acurácia, precisão, recall e F1-score) foram aplicadas para assegurar a robustez dos resultados.

### **3.2 FERRAMENTAS E TECNOLOGIAS UTILIZADAS**

Linguagens de Programação: Python Bibliotecas para Análise de Dados: Pandas, NumPy Bibliotecas para Visualização: Matplotlib, Seaborn, Plotly Ferramentas de Aprendizado de Máquina: Scikit-learn Ambiente de Desenvolvimento: Jupyter Notebook

### **3.3 CONCLUSÃO**

A metodologia adotada neste projeto permitiu uma análise abrangente e detalhada da mortalidade por COVID-19 nos Estados Unidos. A combinação de técnicas estatísticas, modelagem preditiva e visualização de dados forneceu insights valiosos sobre os fatores que contribuíram para as mortes durante a pandemia, oferecendo uma base sólida para futuras pesquisas e intervenções em saúde pública.

## 4 ANÁLISE DOS RESULTADOS

---

A análise dos dados sobre a mortalidade por COVID-19 nos Estados Unidos entre 2020 e 2023 revelou várias tendências importantes e fatores de risco associados às mortes decorrentes do vírus. Os resultados são apresentados a seguir, divididos por categoria de análise.

1. Distribuição Temporal das Mortes Os dados mostram que a distribuição das mortes por COVID-19 variou significativamente ao longo do tempo, com picos correspondendo às ondas de infecção:

O maior pico de mortes ocorreu no inverno de 2020-2021, seguido de outro aumento significativo no inverno de 2021-2022. Houve uma redução geral no número de mortes após a ampla distribuição de vacinas no início de 2021. 2. Condições de Saúde Associadas A análise das condições de saúde preexistentes entre as vítimas revelou que certos fatores de risco estavam fortemente associados a uma maior mortalidade por COVID-19:

Doenças Cardiovasculares: Pacientes com doenças cardíacas apresentaram uma taxa de mortalidade significativamente mais alta. Diabetes: A presença de diabetes foi outro fator de risco crítico, associado a um aumento na mortalidade. Doenças Respiratórias Crônicas: Condições como a doença pulmonar obstrutiva crônica (DPOC) também foram comumente observadas entre as vítimas. 3. Distribuição Geográfica A análise geográfica das mortes revelou variações significativas entre os estados:

Estados com Altas Taxas de Mortalidade: Estados como Nova York, Califórnia e Texas apresentaram altas taxas de mortalidade, especialmente durante os picos iniciais da pandemia. Variações Regionais: Diferenças regionais foram observadas, com áreas urbanas densamente povoadas sendo mais afetadas em comparação com regiões rurais. 4. Características Demográficas Os dados demográficos mostraram que certos grupos eram mais vulneráveis à mortalidade por COVID-19:

Idade: Idosos (65 anos ou mais) constituíram a maioria das mortes, refletindo a maior vulnerabilidade desse grupo. Gênero: Houve uma ligeira predominância de mortes entre homens em comparação com mulheres. Etnia: Grupos étnicos minoritários, especialmente afro-americanos e hispânicos, apresentaram taxas de mortalidade mais altas, destacando disparidades na saúde. 5. Resultados da Modelagem Preditiva Os modelos preditivos desenvolvidos identificaram fatores de risco significativos e forneceram previsões precisas sobre a mortalidade:

Fatores de Risco: Condições como doenças cardiovasculares, idade avançada e diabetes foram consistentemente identificadas como fatores de alto risco. Precisão dos Modelos: Os modelos apresentaram boa precisão na previsão de mortalidade, com métricas de desempenho (acurácia, precisão, recall e F1-score) indicando robustez nas previsões. 6. Visualizações de Dados As visualizações de dados criadas forneceram uma representação clara e intuitiva dos principais achados:

Gráficos Temporais: Mostraram a evolução das mortes ao longo do tempo, destacando os picos pandêmicos. Mapas de Calor: Representaram a distribuição geográfica das mortes, identificando áreas mais afetadas. Gráficos de Barras e Setores: Ilustraram a distribuição das condições de saúde associadas e características demográficas das vítimas.

#### 4.1 CONCLUSÃO

Os resultados deste estudo forneceram uma compreensão detalhada dos fatores que contribuíram para a mortalidade por COVID-19 nos Estados Unidos. A análise destacou a importância de condições de saúde preexistentes, variações geográficas e características demográficas na determinação do risco de morte. Esses insights são valiosos para a formulação de políticas de saúde pública e para a preparação e resposta a futuras crises sanitárias.

## 5 CONCLUSÕES

---

Este projeto de análise de dados em saúde pública focado na mortalidade por COVID-19 nos Estados Unidos, entre os anos de 2020 e 2023, proporcionou insights valiosos sobre os fatores que contribuíram para as mortes relacionadas ao vírus. Através de uma abordagem abrangente e detalhada, foram identificadas tendências significativas e fatores de risco que podem orientar futuras políticas de saúde pública e estratégias de intervenção.

**Principais Achados**

**Distribuição Temporal:** As mortes por COVID-19 tiveram picos correspondentes às ondas de infecção, com reduções notáveis após a distribuição de vacinas.

**Condições de Saúde Associadas:** Doenças cardiovasculares, diabetes e doenças respiratórias crônicas foram identificadas como fatores de risco significativos para a mortalidade.

**Distribuição Geográfica:** Houve variações substanciais nas taxas de mortalidade entre os estados, com áreas urbanas densamente povoadas sendo mais afetadas.

**Características Demográficas:** Idosos, homens e grupos étnicos minoritários, especialmente afro-americanos e hispânicos, apresentaram taxas de mortalidade mais altas, destacando disparidades na saúde.

**Modelagem Preditiva:** Os modelos desenvolvidos identificaram com precisão os fatores de risco significativos e forneceram previsões robustas sobre a mortalidade.

**Impacto e Relevância** Os resultados deste estudo são essenciais para a compreensão da pandemia de COVID-19 e suas consequências. As análises realizadas destacam a importância de monitorar condições de saúde preexistentes e fatores demográficos ao planejar intervenções de saúde pública. Além disso, a identificação de disparidades na mortalidade por COVID-19 ressalta a necessidade de abordagens mais equitativas na distribuição de recursos de saúde e na implementação de políticas de prevenção.

**Aplicações Futuras** Os insights obtidos podem ser aplicados em diversas frentes:

**Políticas de Saúde Pública:** Desenvolvimento de estratégias de prevenção e intervenção direcionadas aos grupos de risco identificados.

**Planejamento de Recursos:** Alocação eficiente de recursos de saúde para áreas e populações mais vulneráveis.

**Pesquisa Continuada:** Estímulo a estudos futuros que aprofundem a análise de fatores de risco e melhorem os modelos preditivos.

**Considerações Finais** A pandemia de COVID-19 evidenciou a necessidade de análises detalhadas e baseadas em dados para enfrentar crises de saúde pública. Este projeto demonstrou como a combinação de técnicas de análise de dados, modelagem preditiva e visualização pode fornecer uma compreensão abrangente dos impactos de uma pandemia. Os achados aqui apresentados não apenas contribuem para o conhecimento existente sobre a COVID-19, mas também estabelecem uma base sólida para enfrentar futuras crises sanitárias de maneira mais eficaz e informada.

## APÊNDICES

---

## APÊNDICES

---