



CURSO: TECNOLOGIA EM CIÊNCIA DE DADOS

SEMESTRE: 2º SEMESTRE

COMPONENTE CURRICULAR / TEMA: PROJETO APLICADO I

NOME DO PROFESSOR: EVERTON KNIHS



**Universidade Presbiteriana Mackenzie**

**UNIVERSIDADE PRESBITERIANA MACKENZIE  
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA**

Análise Exploratória das Relações entre Taxa de Mortalidade e Fatores Socioeconômicos no Brasil

**São Paulo  
2023**



**SUMÁRIO**

<b>1</b>	<b>TÍTULO ESCOLHIDO .....</b>	<b>4</b>
<b>2</b>	<b>MEMBROS.....</b>	<b>4</b>
<b>3</b>	<b>CONTEXTO DO ESTUDO.....</b>	<b>5</b>
<b>4</b>	<b>CRONOGRAMA DE ATIVIDADES .....</b>	<b>8</b>
<b>5</b>	<b>REFERÊNCIAS DE AQUISIÇÃO DO DATASET .....</b>	<b>9</b>
<b>6</b>	<b>DESCRIÇÃO DO DATASET .....</b>	<b>10</b>
<b>7</b>	<b>EXPLORAÇÃO DE CORRELAÇÕES RELEVANTES .....</b>	<b>10</b>
<b>8</b>	<b>PROBLEMAS E FENÔMENOS ENCONTRADOS.....</b>	<b>11</b>
<b>8</b>	<b>LINK DO GITHUB .....</b>	<b>12</b>
	<b>REFERÊNCIAS.....</b>	<b>12</b>



**1. TÍTULO ESCOLHIDO:**

Análise Exploratória das Relações entre Taxa de Mortalidade e Fatores Socioeconômicos no Brasil

**2. MEMBROS:**

Nome: Adriano Mamoru Takeshita

TIA: 23022647

E-mail: [10923022647@mackenzista.com.br](mailto:10923022647@mackenzista.com.br)

Nome: Camila Vieira

TIA: 23005432

E-mail: [camila.vieira1@mackenzista.com.br](mailto:camila.vieira1@mackenzista.com.br) / [10923005432@mackenzista.com.br](mailto:10923005432@mackenzista.com.br)

Nome: Gabriel Schonenberger de Campos

TIA: 23011165

E-mail: [10923011165@mackenzista.com.br](mailto:10923011165@mackenzista.com.br)

Nome: Gustavo Santiago Zarpelão

TIA: 23002824

E-mail: [10923002824@mackenzista.com.br](mailto:10923002824@mackenzista.com.br)

Nome: Luís Eduardo Alves de Moura da Silva

TIA: 23009470

E-mail: [10923009470@mackenzista.com.br](mailto:10923009470@mackenzista.com.br)



### 3. CONTEXTO DO ESTUDO:

#### Instituição escolhida:

##### **Brasil como Foco:**

O Brasil, como um dos países mais populosos e diversificados do mundo, apresenta um cenário sociodemográfico complexo e dinâmico. Isso o torna um contexto extremamente relevante para a análise de dados, especialmente quando se trata de temas sensíveis e fundamentais, como a taxa de mortalidade. A rica heterogeneidade regional, as desigualdades socioeconômicas marcantes e os desafios em saúde pública proporcionam uma oportunidade única para investigar profundamente as interações entre fatores sociais, econômicos e demográficos e seu impacto na taxa de mortalidade. Além disso, ao compreender melhor as dinâmicas que influenciam as taxas de mortalidade no Brasil, pode-se contribuir para a formulação de políticas mais eficazes e estratégias de intervenção direcionadas.

##### **IBGE como Fonte de Dados:**

O Instituto Brasileiro de Geografia e Estatística (IBGE) é uma instituição renomada e respeitada, responsável por coletar, analisar e disseminar informações vitais sobre a sociedade brasileira. Sua ampla gama de dados abrange diversos aspectos da vida no Brasil, desde indicadores demográficos e econômicos até informações geográficas. Ao escolher o IBGE como fonte de dados, garante-se acesso a informações confiáveis, consistentes e atualizadas, fundamentais para a condução de uma análise de dados sólida e confiável. Além disso, a transparência metodológica e a abrangência dos metadados fornecidos pelo IBGE asseguram a qualidade e a compreensão dos dados utilizados para a pesquisa.

**Ao unir a rica diversidade do Brasil com a qualidade e a confiabilidade dos dados do IBGE, este projeto tem como objetivo fornecer insights importantes que podem contribuir significativamente para o entendimento das relações complexas entre variáveis sociais, econômicas e de saúde, assim como para o desenvolvimento de estratégias de melhoria e bem-estar da população brasileira.**



## **Premissas do Projeto:**

1. Utilização dos dados fornecidos pelo IBGE como fonte confiável e abrangente de informações demográficas e socioeconômicas.
2. Implementação de técnicas de ciência de dados para analisar e extrair insights das informações relacionadas à taxa de mortalidade.
3. Consideração das variáveis sociais, econômicas e demográficas para uma análise mais completa e precisa.
4. Através da Análise Exploratória, foco na identificação de padrões, correlações e tendências que possam ajudar a compreender os fatores associados à taxa de mortalidade.
5. Fonte de Dados Confiável: Utilização de dados do IBGE como fonte confiável de informações socioeconômicas e demográficas.
6. Foco na Taxa de Mortalidade: Ênfase na análise da taxa de mortalidade no Brasil e suas relações com fatores sociais e econômicos.
7. Pensamento Computacional: Aplicação das premissas de decomposição, reconhecimento de padrões, abstração e algoritmo na análise dos dados.
8. Contribuição para Políticas Públicas: Objetivo de oferecer insights relevantes que possam ser possíveis fontes de informação voltadas ao bem-estar da população.

## **Objetivos e Metas:**

**Objetivo Principal:** Analisar as relações entre a taxa de mortalidade e diversos fatores sociais, econômicos e demográficos no Brasil.



**Metas:** Identificar correlações significativas entre a taxa de mortalidade e outras variáveis; compreender como fatores como renda, acesso a serviços básicos, segurança e saúde estão relacionados à taxa de mortalidade.

## **Metodologia:**

Neste projeto, adotou-se uma abordagem baseada no pensamento computacional e na análise exploratória de dados, utilizando as premissas de decomposição, reconhecimento de padrões, abstração e algoritmo.

O objetivo principal é desvendar insights relevantes sobre as relações entre a taxa de mortalidade no Brasil e os fatores socioeconômicos, aproveitando ao máximo os dados fornecidos pelo IBGE.

## **Decomposição:**

Iniciaremos decompondo o problema em partes menores e mais gerenciáveis. Dividiremos a análise em etapas, considerando diferentes variáveis, como renda, educação, acesso a serviços básicos e segurança. Isso nos permitirá focar em aspectos específicos e compreender melhor como eles contribuem para a taxa de mortalidade.

## **Reconhecimento de Padrões:**

Utilizaremos técnicas de reconhecimento de padrões para identificar correlações preliminares entre variáveis. Também utilizaremos outros estudos que já existem. Através de gráficos, tabelas e métricas descritivas, buscaremos padrões visuais e tendências iniciais nos dados, destacando possíveis conexões entre os fatores sociais e a taxa de mortalidade.

## **Abstração:**

A abstração nos permitirá simplificar a complexidade dos dados, concentrando-nos nos aspectos mais relevantes. Agruparemos variáveis relacionadas e criaremos indicadores compostos que representam características socioeconômicas mais abrangentes. Essa abordagem simplificada facilitará a análise e a interpretação dos resultados.

## **Algoritmo:**

Implementaremos algoritmos de análise exploratória, como cálculos de médias, desvios-padrão e coeficientes de correlação. Isso nos ajudará a quantificar as relações entre as variáveis e a medir a



força das associações encontradas. Representaremos visualmente as tendências identificadas através de Power BI.

Através da combinação dessas premissas do pensamento computacional e técnicas de análise exploratória de dados, exploraremos as nuances das relações entre a taxa de mortalidade e fatores sociais no Brasil. Essa abordagem nos permitirá desenvolver uma compreensão mais profunda das interações subjacentes e fornecer insights relevantes que contribuam para uma análise sólida e informada.

## 4. CRONOGRAMA DE ATIVIDADES

### 1ª ETAPA

1. **1º encontro: Brainstorm (12/08/2023)** – Adriano, Camila, Gabriel, Luis.
2. **Pesquisa de indicadores e possíveis relações (Entre 12/08/2023 e 18/08/2023)** – Adriano, Camila, Gabriel, Luis.
3. **2º encontro: Discussão sobre os indicadores e Busca de Dados (18/08/2023):** Coleta de informações demográficas, socioeconômicas e de saúde do IBGE; Discussão e definição do escopo - Adriano, Camila, Gabriel, Luis.
4. **Downloads das tabelas de indicadores definidos (Semana do dia 20/08)** – Gabriel e Luis.
5. **Criação de Conta no GitHub (Semana do dia 20/08)**
6. **Desenvolvimento do documento da 1ª etapa (De 13/08 a 21/08)** – Camila.
7. **Revisão documento da 1ª etapa (Até 22/08)** - Todos.
8. **Entrega da 1ª etapa (Até 28/08).**

### 2ª, 3ª e 4ª ETAPA – PREVISÃO DO CRONOGRAMA

1. **Desenvolvimento do DER (Entre 29/08 e 01/09)** – Camila.
2. **Pré-processamento e Limpeza de Dados (Entre 02/09 e 08/09):** Tratamento de dados faltantes, duplicados e inconsistências através de Python e PowerQuery(DAX) - Todos.
3. **Análise Exploratória (Até 24/09):** Exploração inicial dos dados para identificar padrões e insights iniciais - Todos.
4. **Entrega da 2ª etapa (25/09).**
5. **Esboço do Storytelling (Entre 26/09 a 05/10)** – Todos.
6. **Modelagem dos dados e Scripts (Entre 05/10 a 12/10)** – Todos.





7. **Visualização de Dados:** Criação de gráficos e visualizações para apresentar os resultados de forma clara e compreensível (PowerBI) **(12/10 a 20/10)** - Todos.
8. **Revisão e entrega no GitHub (Até 30/10)** - Todos.
9. **3ª entrega: Relatório Final (De 13/08 a 30/10):** Documentação completa das análises realizadas e dos insights obtidos - todos.
10. **Gravação da apresentação (Até 15/11) – Todos.**
11. **Entrega da 4ª etapa (20/11).**

## 5. REFERÊNCIAS DE AQUISIÇÃO DO DATASET

O macro dos dados utilizados neste projeto fora obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE), uma fonte confiável de informações demográficas, socioeconômicas e geográficas no Brasil.

Também explorou-se a Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua. Este conjunto de dados oferece uma visão abrangente das condições socioeconômicas da população brasileira, incluindo indicadores como renda, educação, emprego e saúde.

Censo Demográfico 2020: O Censo Demográfico fornece um panorama detalhado da população, habitação e características socioeconômicas do Brasil.

### **Limitações de Uso:**

É importante mencionar que, embora os dados provenham de fontes confiáveis como o IBGE, algumas limitações podem estar presentes. Uma limitação encontrada foi falta de padronização nos dados, que requerem um pré-processamento cuidadoso para garantir que os dados sejam estruturados de forma adequada e coerente para análise.

### **Período de Coleta:**

Os dados utilizados neste projeto foram coletados entre 2014 e 2019 considerando os períodos específicos de realização das pesquisas da PNAD, Censo Demográfico 2020.



## 6. DESCRIÇÃO DO DATASET

O dataset utilizado neste projeto explora informações relevantes sobre a taxa de mortalidade no Brasil, além de dados socioeconômicos que podem influenciar ou serem influenciados por essa métrica crucial de saúde pública. A proposta deste conjunto de dados é investigar as relações complexas entre a taxa de mortalidade e diversos fatores socioeconômicos, permitindo uma compreensão mais profunda dos fenômenos que impactam a saúde e o bem-estar da população brasileira.

### **Conteúdo do Dataset:**

O dataset engloba as variáveis demográficas, econômicas e sociais:

- Indicador 1.2.1 - Proporção da população vivendo abaixo da linha de pobreza nacional, por sexo
- Indicador 1.2.1 - Proporção da população abaixo da linha de pobreza nacional, por grupos de idade
- Indicador 3.9.2 - Taxa de mortalidade atribuída a fontes de água inseguras, saneamento inseguro e falta de higiene, por sexo e grupo de idade
- Indicador 16.1.4 - Proporção da população de 15 anos ou mais de idade que se sente segura quando caminha sozinha na área onde vive durante a noite, por cor ou raça
- Indicador 16.1.1 - Número de vítimas de homicídios intencionais por 100 mil habitantes, por sexo
- Indicador 3.5.2 - Consumo de álcool em litros de álcool puro per capita, por pessoas de 15 anos ou mais de idade
- Indicador 3.6.1 - Taxa de mortalidade por acidentes de trânsito, por sexo e grupos de idade

## 7. EXPLORAÇÃO DE CORRELAÇÕES RELEVANTES

- Taxa de Mortalidade por Álcool & Proporção da População Abaixo da Linha de Pobreza e escolaridade
- Taxa de Mortalidade por Doenças Crônicas & Proporção da População Abaixo da Linha de Pobreza
- Taxa de Homicídios & Sensação de Segurança Noturna
- Taxa de Homicídios & Proporção de Pessoas Abaixo da Linha de Pobreza



- Taxa de homicídio & renda por região
- Taxa de homicídios (apenas sexo feminino) & Sensação de segurança noturna
- Taxa de homicídios (filtrar pretos e pardos) & Proporção da população abaixo da linha da pobreza
- Taxa de mortalidade por falta de saneamento & Proporção da população abaixo da linha da pobreza

## 8. PROBLEMAS E FENÔMENOS ENCONTRADOS

Observou-se possíveis associações significantes entre variáveis socioeconômicas, como renda e educação, e as taxas de mortalidade. Populações com menor renda e níveis educacionais mais baixos tendem a apresentar taxas de mortalidade mais elevadas, demonstrando a influência direta desses fatores na saúde e bem-estar.

*Entendeu-se também, que é possível associar:*

O Impacto da Sensação de Segurança nos Homicídios;

Desigualdades de Renda e Qualidade de Vida: As disparidades de renda também desempenham um papel crucial na qualidade de vida e na saúde;

Correlações entre Qualidade Ambiental e Saúde;

### **Problemas Identificados nesta primeira etapa:**

Durante a exploração dos dados, também foram identificados os seguintes problemas:

Despadronização dos Dados: observou-se que os dados frequentemente não seguem um padrão uniforme, com informações combinadas em uma mesma coluna. Isso demandará um processo de organização e extração cuidadosa dos dados relevantes.

Falta de Consistência Temporal: Em alguns casos, não há o período de coleta específico desejado para determinadas variáveis. Isso restringe a quantidade de indicadores a ser utilizada.



## **9. LINK DO GITHUB**

[https://github.com/camisv/PROJETO\\_APLICADO\\_I](https://github.com/camisv/PROJETO_APLICADO_I)

### **Referências Bibliográficas**

#### **Instituto Brasileiro de Geografia e Estatística (IBGE)**

Instituto Brasileiro de Geografia e Estatística. Dados Demográficos e Socioeconômicos do Brasil. Disponível em: <<https://sidra.ibge.gov.br/acervo#/S/C2/T/QL>>. Acesso em: 18 de agosto de 2023.

#### **Departamento de Informática do Sistema Único de Saúde (DATASUS)**

Departamento de Informática do Sistema Único de Saúde. Dados de Saúde e Indicadores Epidemiológicos. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>>. Acesso em: 24 de agosto de 2023.

#### **Pesquisa Nacional por Amostra de Domicílios (PNAD)**

Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios Contínua. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9127-pesquisa-nacional-por-amostra-de-domicilios.html>>. Acesso em: 24 de agosto de 2023.

#### **Censo Demográfico:**

Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html>>. Acesso em: 24 de agosto de 2023.