

UNIVERSIDADE PRESBITERIANA MACKENZIE

AMANDA SCHWANTES MAIA  
ELAINE THAMIRES PEREIRA  
GIOVANA VIEIRA DA SILVA  
IAGO FONSECA MARINHO PEREIRA  
MURILO GUIMARÃES

ESTUDO DE PADRÕES DE CASOS DE SUICÍDIOS UTILIZANDO  
ANÁLISE EXPLORATÓRIA DE DADOS

São Paulo

2023

AMANDA SCHWANTES MAIA TIA: 10922519963  
ELAINE THAMIREZ PEREIRA TIA 10922517650  
GIOVANA VIEIRA DA SILVA TIA 10922520791  
IAGO FONSECA MARINHO PEREIRA TIA 10922500138  
MURILO GUIMARÃES TIA 10922505458

ESTUDO DE PADRÕES DE CASOS DE SUICÍDIOS UTILIZANDO ANÁLISE  
EXPLORATÓRIA DE DADOS

Trabalho de Conclusão de disciplina  
intitulada Projeto Aplicado I, componente  
do curso Ciência de Dados da  
Universidade Presbiteriana Mackenzie.

Orientador: Prof. Leonardo Massayuki Takuno

São Paulo  
2023

AMANDA SCHWANTES MAIA  
ELAINE THAMIREZ PEREIRA  
GIOVANA VIEIRA DA SILVA  
IAGO FONSECA MARINHO PEREIRA  
MURILO GUIMARÃES

ESTUDO DE PADRÕES DE CASOS DE SUICÍDIOS UTILIZANDO ANÁLISE  
EXPLORATÓRIA DE DADOS

Trabalho de Conclusão de disciplina  
intitulado Projeto Aplicado I, componente  
do curso Ciência de Dados da  
Universidade Presbiteriana Mackenzie.

Aprovado em

BANCA EXAMINADORA

---

## RESUMO

O objetivo desse estudo é trazer análises sobre os índices em casos de suicídios registrados mundialmente pela OMS e suas correlações de acordo com alguns parâmetros apresentados no *dataset*. Utilizando técnicas e algoritmos de análise exploratória de dados para avaliar as tendências e motivações desses casos para que seja possível identificar padrões e evitar ocorrências futuras.

**Palavras-chave:** Suicídio; Análise Exploratória de dados, Algoritmos.

## **ABSTRACT**

The aim of this study is to analyze the numbers and possible motivations in cases of suicides registered worldwide by the WHO and their correlations according to some parameters presented in the dataset. Using exploratory data analysis techniques and algorithms to assess the trends and motivations of these cases so that it is possible to identify patterns and prevent future occurrences.

**Keywords:** Suicides; Exploratory Data Analysis, Algorithms.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Sumarização e Medidas de Centralidade (Média, Mediana e Moda) .....	16
Figura 2 – Informações úteis sobre o <i>dataset</i> (função <i>info()</i> em <i>Python</i> ).....	17
Figura 3 – Medidas de Dispersão para o Atributo “ <i>suicides_no</i> ” .....	17
Figura 4 – Medidas de Dispersão para o Atributo “ <i>HDI for year</i> ” .....	17
Figura 5 – Medidas de Dispersão para o Atributo “ <i>gdp_per_capita (\$)</i> ” .....	18
Figura 6 – Seleção dos Dados sem <i>Outliers</i> (atributo relativo ao PIB per capita)...	18
Figura 7 – Verificação de quantidade de <i>NAs</i> em cada atributo.....	19
Figura 8 – Porcentagem de Valores Nulos por Atributo.....	19

## LISTA DE TABELAS

Tabela 1	Cronograma da primeira etapa.....	10
Tabela 2	Cronograma da segunda etapa.....	11
Tabela 3	Cronograma da terceira etapa.....	12
Tabela 4	Cronograma da terceira etapa.....	13

## SUMÁRIO

1	INTRODUÇÃO.....	8
2	CAPÍTULO 1 – DOS OBJETIVOS DESTE PROJETO.....	10
3	CAPÍTULO 2 – DO CONJUNTO DE DADOS.....	10
4	CAPÍTULO 3 – DO CRONOGRAMA DE ATUAÇÃO.....	11
5	CAPÍTULO 4 – CONTEXTO DO ESTUDO.....	14
	4.1 – APRESENTAÇÃO DA EMPRESA.....	14
	4.2 – METADADOS.....	15
	4.3 – PROBLEMA DO ESTUDO.....	16
	4.4 – ANÁLISE EXPLORATÓRIA DOS DADOS.....	17
	4.4.1 – ANÁLISE GRÁFICA.....	21
7	CONSIDERAÇÕES FINAIS.....	26
8	REFERÊNCIAS BIBLIOGRÁFICAS.....	27
9	GLOSSÁRIO.....	28



## INTRODUÇÃO

O Projeto Aplicado I consiste num estudo prático com dados reais que serão localizados por meio de informações públicas ou privadas de uma ou mais organizações, onde serão realizadas as demandas e probabilidades analíticas dentro dessa empresa. Isso em razão de contemplar um determinado contexto imposto pelo grupo. Ao final da pesquisa, espera-se alcançar dois resultados: a análise exploratória de dados, que é utilizada para analisar e averiguar conjuntos de dados, e uma proposta analítica a ser aplicada a esses dados selecionados.

Do conjunto de dados nomeado como “*Suicide Rates Overview (1985 to 2021)*”, em português “Visão Geral das Taxas de Suicídio (1985 a 2021)”, escolhido pelo grupo compila informações cruzadas, dentre estas, um *dataset* elaborado através de pesquisas realizadas pela Organização Mundial da Saúde (OMS ou *WHO* - sigla em inglês para *World Health Organization*), órgão responsável pela gestão de saúde internacional, outro feito pelo Grupo Banco Mundial (*The World Bank*), instituição financeira internacional, e por fim, um realizado pelo Programa das Nações Unidas para o Desenvolvimento baseado no Índice de Desenvolvimento Humano (IDH), em inglês *Human Development Insights (HDI)*. Assim sendo, a coleção de dados a ser desenvolvida é de caráter público e detém informações entre os anos de 1985 a 2021.

Posto isto, o trabalho acadêmico confeccionado pelos alunos do segundo semestre do curso de Ciência de Dados da Universidade Presbiteriana Mackenzie tem como propósito investigar e analisar os indicadores associados ao crescimento ou declínio das taxas de suicídio em diferentes localidades de esfera global. Ademais, o grupo busca compreender as causas e fatores de risco que contribuem para esse fenômeno através das pesquisas que aplicarão a exploração de dados.

A temática definida pelos alunos trata-se de uma iniciativa de enorme importância social e de grande relevância acadêmica, já que o suicídio é uma questão complexa e multifatorial, que envolve não apenas aspectos individuais, mas também sociais e culturais. Em suma, portanto, após as investigações será possível concluir dados relevantes que possam ser empregados para a prevenção e redução do comportamento suicida em escala global, algo que pode ser vantajoso nas áreas da saúde mental, saúde pública e políticas sociais, bem como auxiliar no planejamento e estratégias hábeis de políticas públicas e intervenções preventivas

## **CAPÍTULO 1 – DOS OBJETIVOS DESTE PROJETO**

Devido ao ambiente acadêmico ser bastante fértil para o desenvolvimento de ideias voltadas ao bem comum e favorecimento social, o objeto de pesquisa foi consentido através da busca desse propósito. Assim, a escolha do tema foi uma das grandes dores sociais contemporâneas, o suicídio. Problema que, gerador de catástrofes familiares, econômicas, culturais e sociais, necessita de estudos aprofundados e ações para a prevenção de novos casos.

Os objetivos que o projeto de análise exploratória se propõe em perfazer, são:

- Distinguir fatores de riscos associados aos suicídios, constatando padrões e possíveis presenças de grupos de alto risco, se baseando na idade, gênero, cultura, entre outros;
- Encontrar correlações entre a taxa de suicídio e outros atributos socioeconômicos, como por exemplo, o PIB per capita e IDH;
- Identificar tendências de suicídio ao longo dos anos, compreendendo se as taxas estão diminuindo, aumentando ou permanecendo constantes em determinado país ou contexto histórico;
- Realizar análises comparativas entre dois ou mais países para encontrar diferenças relevantes;
- Descobrir prováveis mediações preventivas para auxiliar na redução das taxas de suicídio de determinado país ou região.

A definição do problema, o entendimento e preparação dos dados, são etapas primordiais no processo de revelação de *insights*. Portanto, as técnicas e métodos empregados visam apurar os dados e informações na busca de extrair o conhecimento necessário para alcançar os objetivos acima citados.

## **CAPÍTULO 2 – DO CONJUNTO DE DADOS**

O *dataset* escolhido pelo grupo apresenta dados colecionados de outros quatro conjuntos de dados, sendo idealizado para verificar os indícios associados ao crescimento das taxas de suicídio entre diversas localidades e de âmbito global.

Como dito anteriormente, os outros quatro conjuntos utilizados como referência são das instituições *Human Development Insights (HDI)*, *The World Bank* e dois da Organização Mundial da Saúde (OMS). Além disso, o autor do compilado versa que esta análise pode auxiliar no monitoramento e definir estratégias assertivas para a prevenção desses atos que atormentam nações.

Em relação ao arquivo pode-se dizer que contém 12 colunas e 27.820 mil linhas. Quanto aos dados das colunas são intitulados como *Country* (país), *Year* (ano), *Sex* (sexo), *Age* (idade), *Suicides Number* (número de suicídios), *Population* (população), *Suicides/100k pop* (suicídios/100 mil habitantes), *Country-year* (País – Ano), *HDI for year* (IDH do ano), *GDP for year* (PIB do ano), *GDP per capita* (PIB per capita) e *Generation* (geração).

Em síntese, o *dataset* selecionado mostra as características necessárias para o desenvolvimento do Projeto Aplicado I e, ainda, por abordar um tema social importante que raramente discute-se no ambiente de trabalho, o grupo considerou fundamental tratar esse tema no ambiente acadêmico a fim de conscientizar e promover a discussão sobre essa questão relevante.

### CAPÍTULO 3 – CRONOGRAMA DE ATUAÇÃO

<b>Etapa 01</b>	<b>Entrega 26/03/2023</b>	<b>Responsáveis</b>
<b>Organização do Documento:</b> Capa, sumário, lista de figuras, lista de tabelas, introdução e capítulos.	13/03/2023	Amanda, Elaine
<b>Criação do Repositório:</b> Organizado com pastas, possui sumário, Objetivo descrito no documento <i>readme</i> , contém os nomes dos membros	15/03/2023	Iago, Murilo
<b>Cronograma:</b> Atividades, datas, responsabilidades e milestones, em conformidade com os prazos do projeto no componente curricular	16/03/2023	Amanda, Elaine, Giovana, Iago, Murilo
<b>Objetivos do projeto:</b> apresentando a empresa, a área de atuação, as necessidades e o problema a ser trabalhado	17/03/2023	Amanda, Elaine, Giovana, Iago, Murilo
<b>Dataset e Metadados:</b> Metadados, colunas e linhas em quantidade, que podem atender aos objetivos do trabalho	18/03/2023	Iago, Giovana

Tabela 01 – Cronograma da primeira etapa

<b>Etapa 02</b>	<b>Entrega 16/04/2023</b>	<b>Responsáveis</b>
<b>Organização do Documento:</b> Capa, sumário, lista de figuras, lista de tabelas, introdução e capítulos.	Consolidação 31/03/2023	Amanda, Elaine, Giovana
<b>Seção de Proposta Analítica inserida no Documento:</b> Apresentação dos itens conforme apresentado na videoaula, nos slides sobre informações da Organização/Empresa, Problema do Estudo, Metadados e a Proposta do que será desenvolvida e Pipeline de dados para apoiar o processo analítico	Consolidação 05/04/2023	Iago, Murilo, Elaine
<b>Scripts da Análise Exploratória em Python:</b> Os <i>scripts</i> inseridos no <i>GitHub</i> atendendo ao que foi solicitado para compor a Análise Exploratória dos Dados do Projeto. Os <i>scripts</i> são <i>notebooks Jupyter</i> , escritos em <i>Python</i> desenvolvidos com comandos, bibliotecas e linhas de comentários, atendendo a boas práticas de organização do código	Consolidação 07/04/2023	Amanda, Iago, Murilo
<b>Objetivos do projeto:</b> apresentando a empresa, a área de atuação, as necessidades e o problema a ser trabalhado	Consolidação 14/04/2023	Amanda, Elaine
<b>Seção de Análise Exploratória inserida no Documento:</b> Apresentação dos itens alinhados aos <i>scripts</i> que foram inseridos no <i>GitHub</i> . O texto apresenta uma descrição das variáveis, complementando os Metadados que já havia sido apresentado, para cada coluna foram descritos itens como número de exemplares, valor máximo/mínimo, variância, desvio, distribuição, quantidade de <i>NAs</i> , existência de outliers, entre outros. Utilizou-se alguns gráficos para melhor detalhar a amostra	18/03/2023	Iago, Murilo, Giovana.

Tabela 02 – Cronograma da segunda etapa

Etapa 03	Entrega 07/05/2023	Responsáveis
<b>Esboço do Storytelling:</b> Documento descrevendo a narrativa que o grupo pretende fazer para a gravação final. A narrativa traz os seguintes elementos: Apresentação do grupo, Nome do projeto, Empresa/Organização de estudo, área do problema, descrição do problema/gap, proposta analítica, dados disponíveis, análise exploratória, resultados pretendidos	Consolidação 21/04/2023	Amanda, Elaine, Giovana
<b>Scripts da Análise Exploratória em Python:</b> Os <i>scripts</i> foram inseridos no <i>GitHub</i> e atendem ao que foi solicitado para compor a Análise Exploratória dos Dados do Projeto. Os <i>scripts</i> são <i>notebooks Jupyter</i> , escritos em <i>Python</i> e foram desenvolvidos com comandos, bibliotecas e linhas de comentários, atendendo a boas práticas de organização do código	Consolidação 28/04/2023	Iago, Giovana, Elaine
<b>Seção de Análise Exploratória inserida no Documento:</b> A seção foi inserida e apresenta os itens alinhados aos scripts que foram inseridos no <i>GitHub</i> . O texto apresenta uma descrição das variáveis, complementando os Metadados que já havia sido apresentado, para cada coluna foram descritos itens como número de exemplares, valor máximo/mínimo, variância, desvio, distribuição, quantidade de <i>NAs</i> , existência de <i>outliers</i> , entre outros. Utilizou-se alguns gráficos para melhor detalhar a amostra	Consolidação 05/05/2023	Amanda, Iago, Murilo

Tabela 03 – Cronograma da terceira etapa

Etapa 04	Entrega 28/05/2023	Responsáveis
<b>Apresentação gravada:</b> disponibilizada no <i>Youtube</i> , e link no <i>Github</i> . A duração do vídeo entre 5 e 10 minutos e apresenta os seguintes elementos: Apresentação do grupo,	Consolidação 22/05/2023	Iago, Elaine, Giovana

Nome do projeto, Empresa/Organização de estudo, área do problema, descrição do problema/ <i>gap</i> , proposta analítica, dados disponíveis, análise exploratória, resultados Pretendidos		
<b>Documento Final:</b> Encaminhado no moodle e inserido no <i>GitHub</i> . Ele está devidamente organizado e apresenta os elementos que foram solicitados no início do projeto	Consolidação 26/05/2023	Iago, Murilo, Amanda

Tabela 03 – Cronograma da quarta etapa

## CAPÍTULO 4 – CONTEXTO DO ESTUDO

### 4.1 APRESENTAÇÃO DA EMPRESA

O conjunto de dados "*Suicide Rates Overview* (1985 to 2021)", em português "Visão Geral das Taxas de Suicídio (1985 a 2021)", é um compilado de informações cruzadas, das quais, formam um dataset desenvolvido por meio de pesquisas executadas por instituições internacionais e que, ainda, exprimem cunho social, entre elas é a Organização Mundial da Saúde (OMS).

A entidade em questão tem como missão promover a saúde, para alcançar o mais alto nível possível de saúde em escala mundial. Além disso, aspira em um mundo, o qual todas as pessoas podem contar com suporte à saúde de qualidade sem distinção de raça, credo, gênero ou qualquer outra condição. Os pilares em trabalhar com respeito, dignidade humana, equidade, integridade e transparência em suas ações traduzem os valores dessa organização.

A OMS (ou *WHO* em inglês, como já citado neste estudo) atua em diversas áreas da saúde, incluindo na atuação de estudos sobre doenças transmissíveis e não transmissíveis, sobre nutrição, vacinação, na promoção de saúde, preservação do meio ambiente e, destacando o contexto vigente, a saúde mental. Neste quesito, o bem-estar mental é trabalhado fortemente pela agência, buscam monitorar e planejar estratégias voltadas à prevenção do comportamento suicida, pois acreditam que a medida de proteção não é apenas uma questão individual ou familiar, mas algo que beneficia a sociedade, o sistema de saúde e a economia em geral.

Em 2020, a OMS lançou um Atlas de Saúde Mental Global, uma de suas iniciativas na área de Ciência de Dados, com o intuito de aumentar a utilização da

análise de dados e aplicá-la nas tomadas de decisões na prestação de serviços de saúde mental em todo o mundo. Além disso, a organização também tem investido no desenvolvimento de sistemas de vigilância para monitorar o aumento de doenças e outros indicadores de saúde em diferentes países.

Existem excelentes trabalhos a serem destacados em relação aos estudos de prevenção ao suicídio e desenvolvidos pela instituição, entre eles, um publicado no Relatório Mundial sobre a Saúde Mental em 2018, onde se evidenciou a importância de investir em serviços de suporte à saúde mental para fornecer tratamento adequado e acesso à ajuda para todos os que sofrem de transtornos mentais, baseando-se nos direitos humanos e enfatizando a participação de toda a comunidade que convive com esses pacientes. A publicação foi considerada um convite a vários países para voltarem sua atenção para este problema, possibilitando a criação de ações de caráter preventivo idealizados pelos governos, contando inclusive com informações fornecidas diretamente pela organização para colaborar na implementação desses planos.

Ainda sobre trabalhos em destaque, a OMS tem hoje um excelente encargo de treinamento de profissionais da área da saúde que destina melhorar o diagnóstico precoce e tratamento de possíveis pessoas em risco de suicídio.

## 4.2 METADADOS

O dataset que será empregado em nosso estudo é um arquivo CSV de origem aberta, ou seja, está disponível para uso público. As informações contidas variam entre os anos de 1985 a 2021, entretanto durante a análise exploratória será desconsiderado alguns desses períodos. Como dito anteriormente, no capítulo 2, os dados que serão utilizados pertencem a um compilado de outros quatro conjuntos e os proprietários são eles: *Human Development Insights (HDI)*, *The World Bank* e dois da Organização Mundial da Saúde (OMS), porém o grupo teve acesso à coleção através do site *Kaggle*.

Os atributos do conjunto de dados incluem *Country* (país), *Year* (ano), *Sex* (sexo), *Age* (Idade), *Suicides Number* (número de suicídios), *Population* (população), *Suicides/100k pop* (suicídios/100 mil habitantes), *Country-year* (País – Ano), *HDI for year* (IDH do ano), *GDP for year* (PIB do ano), *GDP per capita* (PIB per capita) e *Generation* (geração). As variáveis podem ser consideradas, tendo em vista o tipo,

como numéricas (*int64* e *float64*) e instância de uma classe, no caso os atributos do gênero objeto.

Com as informações descritas acima, será possível analisar as tendências de suicídio em diferentes países e em diferentes perspectivas populacionais ao longo dos anos. Todavia, vale ressaltar que os dados se referem a casos de suicídios reais e, portanto, podem ser sensíveis e exigem sensibilidade e cautela ao serem utilizados pelo grupo. Além disso, é necessário levar em consideração que a base de dados não inclui todos os atributos sobre fatores socioeconômicos, ou ainda culturais e psicológicos, isto é, fatores adversos que podem estar diretamente associados ao comportamento suicida e, conseqüentemente, fatores esses que não cabem neste estudo.

#### **4.3 PROBLEMA DO ESTUDO**

Como já mencionado, o objetivo do estudo é analisar e compreender as tendências e fatores que implicam a incidência de suicídio em diferentes localidades, por meio das variáveis existentes no conjunto de dados atual. Diante disso, o grupo espera obter insights que permitam reconhecer tais aspectos e que, de certa forma, encontre quais atributos tendem a este flagelo que permeia a sociedade.

Na percepção do grupo espera-se explorar as correlações entre as variáveis, como IDH, PIB ou geração, identificar as taxas durante os anos, se houve aumento ou declínio dependendo do período, investigar os números por país ou região geográfica, quais países têm taxas mais elevadas, observar os índices por gênero ou faixa etária, como homens e mulheres ou idades específicas, entre outras relações.

Perante aos dados disponíveis para realizar a pesquisa é notável que alguns deles sejam faltantes, isto é, apresentam-se como valores ausentes, e ainda, outra parte deles apresentam-se como classes que não suportam o momento da análise, em outros termos, não são identificados ou admissíveis pelos algoritmos na hora da exploração. Além disso, por se tratar de um conjunto de dados extenso, pois possui inúmeras linhas, alguns desses dados serão abstraídos para que a sondagem seja apropriada.

Ao aplicar a abordagem do pensamento computacional durante o processo tornará os problemas em soluções, além de desenvolver o pensamento crítico e



criativo. Esse tipo de técnica abrange competências como a decomposição de problema, o reconhecimento de padrões, abstração e algoritmos, perspectivas que serão utilizadas, conforme a explicação no parágrafo acima. Ademais, o *dataset* será dividido em dois conjuntos, o populacional e amostral, pois dessa maneira será possível lidar com a totalidade das observações prováveis ou não, além de que esta distinção é importante, porque indicará a forma de conduzir determinadas análises. Sendo assim, o grupo alcançará características e padrões da população a partir da amostra, sem a necessidade de reunir todos os dados, algo inviável em diversas ocasiões.

#### 4.4 ANÁLISE EXPLORATÓRIA DOS DADOS

Durante a primeira investigação dos dados vigentes, percebeu-se que o conjunto, na realidade, possui 31.756 linhas e 12 colunas. As variáveis são do tipo numérico (*int64* e *float64*) e do tipo instância de uma classe, no caso os atributos do gênero objeto. Após o importe das bibliotecas e a preparação dos dados e, isso inclui a classificação e transformação dos mesmos, foi empregada a função de sumarização estatísticas em *Python*, o *summary()*, para encontrar as medidas de posição, sendo elas a média, a mediana e a moda, conforme a imagem a seguir:

Figura 1 - Sumarização e Medidas de Centralidade (Média, Mediana e Moda)

	suicides_no	population	suicides/100k pop	HDI for year	gdp_per_capita (\$)
<b>count</b>	31756.00	31756.00	31756.00	31756.00	31756.00
<b>mean</b>	229.12	7217454.24	11.72	0.79	17588.95
<b>std</b>	852.36	57993230.36	21.59	0.06	19464.86
<b>min</b>	0.00	278.00	0.00	0.38	251.00
<b>25%</b>	3.00	128821.00	0.37	0.79	3765.00
<b>50%</b>	25.00	546832.50	4.29	0.79	10062.00
<b>75%</b>	123.00	2908763.00	14.56	0.79	25622.00
<b>max</b>	22338.00	1411100000.00	515.09	0.98	126352.00

Fonte: Os autores (2023).

Figura 2 - Informações úteis sobre o *dataset* (função *info()* em *Python*)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31756 entries, 0 to 31755
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype
---  -
0   country              31756 non-null  object
1   year                 31756 non-null  int64
2   sex                  31756 non-null  object
3   age                  31756 non-null  object
4   suicides_no          30556 non-null  float64
5   population            31756 non-null  int64
6   suicides/100k pop    31756 non-null  float64
7   country-year          31756 non-null  object
8   HDI for year          12300 non-null  float64
9   gdp_for_year ($)     31756 non-null  object
10  gdp_per_capita ($)    31756 non-null  float64
11  generation            31756 non-null  object
dtypes: float64(4), int64(2), object(6)
memory usage: 2.9+ MB
```

Fonte: Os autores (2023).

Sabe-se que as medidas de posição e de dispersão representam, de maneira resumida, os dados analisados e seria a etapa antecedente para fazer qualquer dedução estatística, isto é, a partir das informações alcançadas é possível expandir determinadas conclusões e se direcionar à análises mais aprofundadas. Tendo isso em vista, foram utilizadas duas funções para calcular o desvio padrão e a variância, sendo elas: *std()* e *var()* relativo a três atributos tais como o número de suicídios, IDH por ano e o PIB per capita, além de que foi executada a função *quantile()* para conferir o 1º e 3º quartil, como visto abaixo:

Figura 3 - Medidas de Dispersão para o Atributo "*suicides\_no*"

```
DESVIO PADRÃO: 852.3589
VARIÂNCIA: 726515.6245
1ª QUARTIL: 25.0
3ª QUARTIL: 123.0
```

Fonte: Os autores (2023).

Figura 4 - Medidas de Dispersão para o Atributo "*HDI for year*"

```
DESVIO PADRÃO: 0.0607
VARIÂNCIA: 0.0037
1ª QUARTIL: 0.7937
3ª QUARTIL: 0.7937
```

Fonte: Os autores (2023).

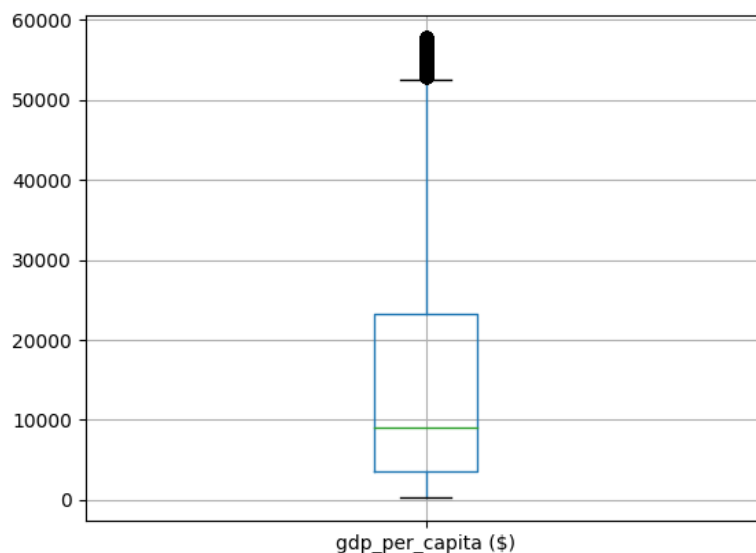
Figura 5 - Medidas de Dispersão para o Atributo “*gdp\_per\_capita* (\$)”

```
DESVIO PADRÃO: 19464.8611  
VARIÂNCIA: 378880819.1826  
1ª QUARTIL: 10062.0  
3ª QUARTIL: 25622.0
```

Fonte: Os autores (2023).

Sobre as correlações, foi utilizada a função *corr()* com propósito de explorar a análise multivariada e, ainda, para a identificação dos *outliers* e do escore padrão foi operada a fórmula (regra) de quartis e o resultado foi apresentado em gráficos do tipo *boxplot*, bem como alguns valores discrepantes foram extraídos para melhor entendimento, de acordo com a representação a seguir:

Figura 6 - Seleção dos Dados sem *Outliers* (atributo relativo ao PIB per capita)



Fonte: Os autores (2023).

Em referência à distribuição e frequência dos dados categóricos, sabe-se que há a mesma proporção de homens e mulheres, assim como em relação à idade entre as categorias de faixa etária, com exceção do intervalo de idades entre 5 a 14

anos que apresenta uma assiduidade menor. Para o atributo correspondente ao ano, pode-se dizer que apresenta bastante variação entre eles, tendo maior frequência entre 2004 a 2009 e com menor no ano de 2016.

Ao verificar o *dataset*, notou-se que inúmeros dados são ausentes, algo que pode prejudicar o estudo, em particular a validade e confiabilidade dos resultados da análise. Diante da situação, foi aplicada a função *isnull()* para essa finalidade e constatou que somente os atributos “*suicides\_no*” e “*HDI for year*” possuem linhas faltantes. À vista disso, esses valores foram substituídos pela mediana e pela média, respectivamente, para não perder uma quantidade significativa de exemplares. Caso não trocasse ou usasse o *dropna()*, perderia um pouco mais de 60% das informações referentes ao IDH, por exemplo.

Figura 7 - Verificação de quantidade de NAs em cada atributo

```
country          0
year             0
sex              0
age              0
suicides_no      1200
population        0
suicides/100k pop 0
country-year      0
HDI for year      19456
gdp_for_year ($)  0
gdp_per_capita ($) 0
generation        0
dtype: int64
```

Fonte: Os autores (2023).

Figura 8 - Porcentagem de Valores Nulos por Atributo

```
HDI for year      61.267162
suicides_no       3.778813
country           0.000000
year              0.000000
sex               0.000000
age               0.000000
population         0.000000
suicides/100k pop 0.000000
country-year      0.000000
gdp_for_year ($)  0.000000
gdp_per_capita ($) 0.000000
generation         0.000000
dtype: float64
```

Vale ressaltar que ao longo da verificação dos dados foi realizada a pipeline, a qual indica a sequência e métodos aplicados no decorrer da análise exploratória. Sendo assim, obteve a seguinte ordenação:

#### 1. Importação de bibliotecas

Neste primeiro momento, importamos as bibliotecas, pacotes de recursos disponíveis em *Python* para a análise exploratória de dados. São essas: *pandas*, *numpy*, *pprint*, *seaborn* e *matplotlib*.

#### 2. Importação e Verificação do *dataset*

Nesta etapa, traz-se o *dataset* completo, sem alterações, para dentro do código.

#### 3. Distribuição de Frequências

A intenção aqui é observar como os atributos categóricos estão espalhados e qual a contagem que cada variável aparece no conjunto de dados.

#### 4. Transformação dos *NAs*

São identificados em quais atributos há valores nulos. Estes são substituídos ou desconsiderados.

#### 5. Preparação dos dados (classificação e transformação de dados)

Para que os dados sejam analisados com acurácia, é necessário estarem nos formatos corretos. Este é o momento em que garantimos que os dados quantitativos sejam mensuráveis.

#### 6. Sumarização e Medidas de Centralidade (Média, Mediana e Moda)

Aqui são identificados os pontos centrais da distribuição dos dados.

#### 7. Medidas de Dispersão (Variância, Desvio padrão, Percentis e Quartis)

Agora, é identificada a distribuição dos dados por algumas medidas diferentes.

#### 8. Escore Padrão e Identificação de *Outliers*

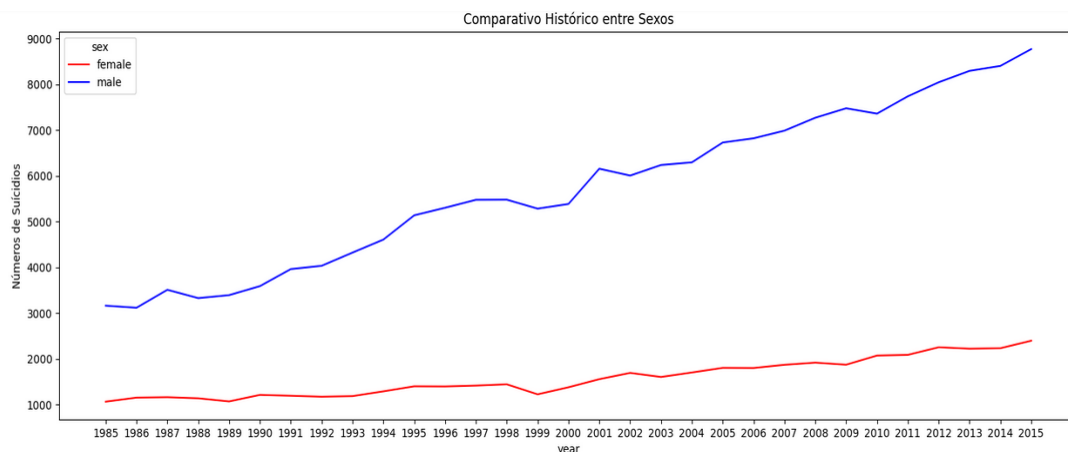
Aqui é definido um parâmetro limite de distância da média para encontrar outliers, valores que divergem de forma extrema dos demais dados no quesito distribuição.

Assim, nesta primeira busca foram detectadas variáveis e informações relevantes em diferentes formatos, o que possibilitará a análise exploratória completa. Entretanto, foram revelados dois atributos nulos, algo que pode impossibilitar determinadas pesquisas futuramente. Portanto, para isso adotar medidas de tratamento será a melhor alternativa. Em suma, essa pesquisa inicial facilitou uma compreensão melhor do que pode ser desenvolvido e quais aspectos devem ser ponderados e estabelecidos para o êxito da análise. Além disso, esse contato inicial nos permite definir quais algoritmos e quais tarefas analíticas devem ser adotadas a partir destes dados.

#### **4.4.1 ANÁLISE GRÁFICA**

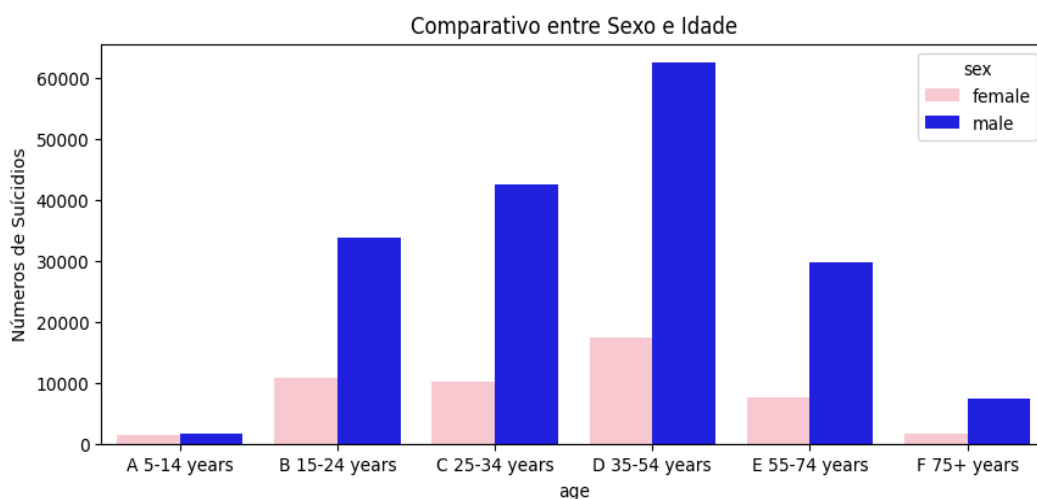
Após a realização da análise exploratória do dataset vigente, observou-se no Brasil um aumento no número de suicídios ao longo dos anos, tendo em vista o aumento populacional. Além disto, foi verificado que os índices mais elevados estão centralizados no sexo masculino e a faixa etária mais afetada é entre 35 a 54 anos, englobando a geração Baby Boomers e a geração X, como demonstram as imagens a seguir:

Comparativo Histórico entre Sexos



Fonte: Os autores (2023).

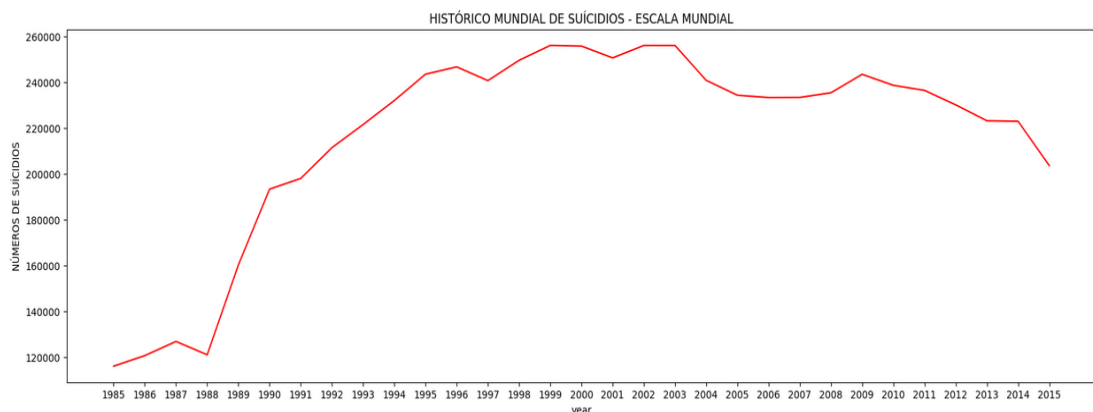
### Comparativo entre Sexo e Idade



Fonte: Os autores (2023).

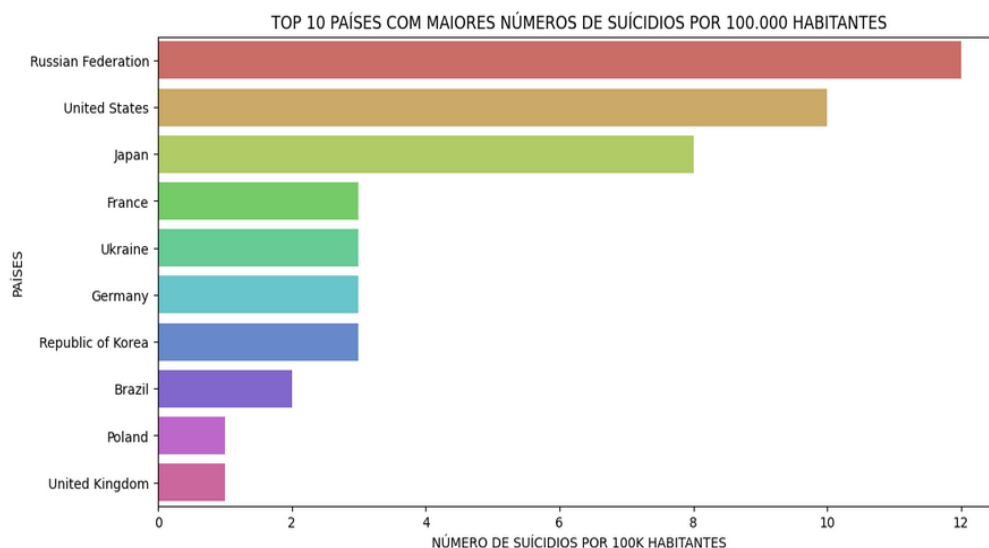
Ao analisar os dados globais, constata-se que a partir do ano de 1988 houve um crescimento significativo no número de suicídio em todo mundo, ademais, demonstra um padrão análogo ao do Brasil de acordo com o sexo e o intervalo de idades. Os países com os maiores índices do comportamento suicida por 100 mil habitantes são, respectivamente, Rússia, Estados Unidos e Japão, conforme os gráficos a seguir:

### Histórico Mundial de Suicídios



Fonte: Os autores (2023).

### Top 10 Países com Maiores Números de Suicídios por 100 mil/habitantes



Fonte: Os autores (2023).

Em frente a essas informações, os resultados nos fornecem uma visão abrangente dos padrões e desafios relacionados ao suicídio, destacando a necessidade de estratégias de prevenção eficazes tanto no Brasil quanto no mundo. Em resumo, os dados mostram que o suicídio é um problema multifatorial que afeta diferentes grupos de pessoas, de idades diferentes e de diferentes contextos econômicos. Acima disso, é importante entender que algumas respostas para esses resultados estão atreladas a alguns motivos históricos. Diante disso, foram considerados os seguintes aspectos:

### **BRASIL (1985-2015):**



### **Comparativo histórico de suicídios:**

- É possível observar através do gráfico de linha, que durante o intervalo de 1985 a 2015, ocorreu um crescimento gradativo nos números de suicídios do Brasil.
- Percebe-se que a partir de 1994, possivelmente pela implementação do Plano Real, ocorreu uma amenização no aumento da taxa de suicídios no Brasil, podendo representar certa esperança da população no plano econômico.

### **Comparativo histórico entre sexo:**

- Em termos absolutos, nota-se que em comparação com sexo masculino, o crescimento dos casos do sexo feminino são relativamente mais constante. A evolução do número de suicídio do sexo masculino é mais acentuada.

### **Comparativo entre Sexo e Idade:**

- Na faixa etária que apresenta as maiores taxas (35-54 anos), existe uma possível relação com as frustrações pessoais, pois culturalmente essa faixa etária considera que o brasileiro deve ser bem sucedido socialmente e financeiramente. Crise dos 30 anos, queda na produção de testosterona, autoconfiança e etc.

### **Comparativo entre Sexo:**

- Como já foi possível observar em outros gráficos, o número de caso de suicídios do sexo masculino é maior que do sexo feminino. Representando 78.37% e 21.63% respectivamente.

## **MUNDO (1985-2015):**

### **Histórico mundial de suicídios:**

- Percebe-se um grave aumento no índice entre os anos de 1988 e 1990. Período histórico marcado por mudanças políticas e sociais que causaram instabilidade e consequências econômicas, portanto podem ter gerado alto nível de estresse social, além de transtornos psicológicos na população. (Ex: Declínio da União Soviética, fim da Guerra Fria, Queda do Muro de Berlim, período de hiperinflação)

### **Histórico mundial de suicídios entre sexos:**

- Estudo da UNB cita que os homens, em geral, fazem uso de práticas mais letais na tentativa de suicídio, sendo um dos possíveis fatores de maior índice entre os homens. ([http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-294X2018000200008](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-294X2018000200008))

#### **Comparativo entre sexo e geração:**

- Boomers e Geração X representam a maior parte da população economicamente ativa do pós-guerra, momento mundial marcado pela altíssima taxa de desemprego.

#### **Comparativo entre sexo:**

- Como já foi possível observar em outros gráficos, o número de caso de suicídios do sexo masculino é maior que do sexo feminino. Representando 76.89% e 23.11% respectivamente.

#### **Ranking 10 países com maior índice:**

- Além das causas populacionais e da Guerra Fria que relacionam Rússia e EUA como países com maiores índices de suicídio, o Japão aparece como terceiro maior. No Japão, existe uma herança histórica do período imperial, relacionada ao Bushido, código de honra dos samurais, no qual existia o ato do suicídio chamado seppuku. O seppuku era considerado um ato de extrema dignidade, realizado por samurais para expiar falhas, restaurar a honra de sua família ou cumprir ordens de superiores. Os casos durante a Segunda Guerra Mundial, nos atos dos pilotos kamikazes, têm profunda relação com a antiga prática. Culturalmente o Japão ainda sofre consequências dessa tradição, no qual pessoas cometem suicídio para restaurar a honra quando consideram que fracassaram em alguma etapa da vida.

## 7. CONSIDERAÇÕES FINAIS

Ao finalizar a análise exploratória de dados sobre as taxas de suicídio, podemos concluir que a investigação e compreensão dos fatores associados a esse fenômeno são de extrema importância social e acadêmica. Através da análise dos dados coletados, foi possível identificar padrões, correlações e tendências relacionadas às taxas de suicídio em diferentes localidades e ao longo do tempo e foi possível identificar fatores de risco associados a idade e gênero além de correlações com indicadores socioeconômicos, como PIB per capita e IDH.

A colaboração entre as instituições internacionais, como a Organização Mundial da Saúde (OMS), o Programa das Nações Unidas para o Desenvolvimento (PNUD) e o Grupo Banco Mundial, foi fundamental para disponibilizar os conjuntos de dados utilizados no projeto. Essa colaboração evidencia a importância de trabalhar de forma integrada na área da saúde mental e na busca por estratégias eficazes de prevenção ao suicídio.

A análise exploratória de dados mostrou-se como uma ferramenta poderosa para a obtenção de insights e conhecimentos que podem contribuir para a formulação de políticas públicas e intervenções preventivas mais efetivas.

Em suma, o estudo contribuiu para o avanço do conhecimento nessa área e demonstrou a importância de investir em pesquisas e ações preventivas, como por exemplo a necessidade de programas mais efetivos de terapias e medicações específicas, apoio social à população, principalmente àquelas em condições de doenças emocionais. É de extrema importância aos governantes saber lidar com esse problema de saúde pública. Esperamos que os resultados obtidos possam servir como base para a implementação de políticas e estratégias efetivas de prevenção do suicídio em diferentes contextos e contribuam para a promoção do bem-estar mental e da saúde global.

## REFERÊNCIAS

[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>

UNITED NATIONS DEVELOPMENT PROGRAM. (2018). Human development index (HDI). Retrieved from <http://hdr.undp.org/en/indicators/137506>

WORLD BANK. (2018). World development indicators: GDP (current US\$) by country:1985 to 2016. Retrieved from <http://databank.worldbank.org/data/source/world-development-indicators#>

World Health Organization. (2018). Suicide prevention. Retrieved from [http://www.who.int/mental\\_health/suicide-prevention/en/](http://www.who.int/mental_health/suicide-prevention/en/)

## GLOSSÁRIO

**País** – local de residência dos envolvidos;

**Ano** – Ano em que foi registrada a ocorrência;

**Sexo** – gênero;

**Idade** – Idade dos envolvidos;

**Número de Suicídios** – quantidade de registros de casos de suicídio;

**População** – Quantidade de pessoas que habitam em determinada região;

**Suicídios 100 mil habitantes** - quantidade de registros de casos de suicídio a cada 100 mil habitantes em uma determinada região;

**País – Ano** – Local de residência em comparação ao ano onde ocorreram os casos de suicídios;

**IDH – Índice de desenvolvimento Humano** – Indicador de nível de desenvolvimento de uma sociedade;

**PIB – Produto Interno Bruto** – Indicador quantitativo de atividade econômica de um determinado país;

**Geração** – Designação dada para diferenciar faixas etárias;