



CURSO: Tecnologia Em Ciência De Dados

POLO DE APOIO PRESENCIAL: Jundiaí e Higienópolis

SEMESTRE: 2/2023

COMPONENTE CURRICULAR / TEMA: **PROJETO APLICADO I {TURMA 02A} 2023/2**

NOME DO GRUPO – OLHAR INFINITO

23003685 – BEATRIZ DE SOUZA FERREIRA - [10923003685@MACKENZISTA.COM.BR](mailto:10923003685@MACKENZISTA.COM.BR)

23006794 – EDUARDO DAVID - [10923009764@MACKENZISTA.COM.BR](mailto:10923009764@MACKENZISTA.COM.BR)

23023708 – GUSTAVO CASTRO SANGALI - [10923023708@MACKENZISTA.COM.BR](mailto:10923023708@MACKENZISTA.COM.BR)

23008385 – JESSICA CLARA - [10923008385@MACKENZISTA.COM.BR](mailto:10923008385@MACKENZISTA.COM.BR)

23006005 – MOISÉS DE LIMA SOUZA - [10923006005@MACKENZISTA.COM.BR](mailto:10923006005@MACKENZISTA.COM.BR)

NOME DO PROFESSOR: **EVERTON KNIHS**

## O que deve conter na entrega da Etapa 1 da Tarefa Aplicando Conhecimento?

- Título do trabalho.
- Membros do grupo (nome completo, TIA e email)
- Contexto do estudo (premissas do projeto, objetivos e metas, cronograma de atividades, pensamento computacional em contextos organizacionais).
- Referências de aquisição do dataset (origem dos dados, limitações de uso e período da coleta).
- Descrição da origem (informações sobre a organização que gerou os dados e o contexto em que os dados foram gerados).
- Descrição do dataset (o que ele contém, qual é a proposta, quais problemas ou fenômenos foram registrados).



## Sumário

<b>1. Título:</b>	<b>5</b>
<b>2. Introdução:</b>	<b>5</b>
<b>3. Objetivos:</b>	<b>5</b>
<b>4. Metas:</b>	<b>5</b>
<b>5. Cronograma:</b>	<b>6</b>
<b>6. Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais:</b>	<b>6</b>
A. Decomposição:	6
B. Reconhecimento de padrões:	6
C. Abstração:	7
D. Design de Algoritmos:	7
<b>7. Organização e o contexto em que os dados foram gerados:</b>	<b>7</b>
<b>8. Referências de aquisição do dataset:</b>	<b>7</b>
<b>9. Dataset e Metadados</b>	<b>7</b>
A. Dataset:	7
B. Descrição do Dataset:	8
C. Metadados:	8
D. Tipo do Arquivo:	9
E. Sensibilidade:	9
F. Proprietário do Dado:	9
G. Restrições de uso:	9
<b>10. Análise Exploratória:</b>	<b>10</b>
A. Descrição das Variáveis:	10
B. Desafios e Limitações:	19
C. Recomendações Preliminares:	19
D. Conclusão da Análise Exploratória:	19
<b>11. Gráficos e Visualizações:</b>	<b>20</b>
A. Gráfico 01: Quantidade de registros agrupados por WHO_REGION:	20
B. Gráfico 02: Histograma da coluna TOTAL_VACCINATION_PER100:	20
C. Gráfico 03: Box Plot da coluna TOTAL_VACCINATION_PER100:	21
D. Gráfico 04: Histograma da coluna PERSONS_VACCINATED_1PLUS_DOSE_PER100:	21
E. Gráfico 05: Box Plot da coluna PERSONS_VACCINATED_1PLUS_DOSE_PER100:	22
F. Gráfico 06: Histograma da coluna PERSONS_LAST_DOSE_PER100:	22
G. Gráfico 07: Box Plot da coluna PERSONS_LAST_DOSE_PER100:	23
H. Gráfico 08: Histograma da coluna NUMBER_VACCINES_TYPE_USED:	23
I. Gráfico 09: Box Plot da coluna NUMBER_VACCINES_TYPE_USED:	24
J. Código Fonte Gráficos:	24
<b>12. Pipeline de Dados:</b>	<b>26</b>
A. Coleta:	26



B. Limpeza: .....	26
C. Análise Exploratória: .....	26
D. Técnicas de Análise: .....	26
E. Visualização: .....	26
F. Relatório Final: .....	26
<b>13. Proposta Analítica: .....</b>	<b>27</b>
A. Apresentação da OMS (Organização Mundial de Saúde): .....	27
<b>14. Storytelling: .....</b>	<b>29</b>
14.1. Apresentação do Grupo .....	29
14.2. Nome do Projeto .....	29
14.3. Empresa/Organização de Estudo .....	29
14.4. Área do Problema .....	29
14.5. Descrição do Problema / Gap .....	29
14.6. Proposta Analítica .....	30
14.7. Dados Disponíveis .....	30
14.8. Análise Exploratória .....	30
14.9. Resultados Pretendidos .....	30
<b>15. Artefatos do Projeto: .....</b>	<b>31</b>
A. Link Github: .....	31
B. Link Projeto: .....	31
C. Link Artefatos do Projeto: .....	31
D. Link Vídeo: .....	31



## Tabelas

<b>Tabela 01</b>	Cronograma .....	6
<b>Tabela 02</b>	DataSet.....	8
<b>Tabela 03</b>	Informação do DataFrame.....	11
<b>Tabela 04</b>	Análise da coluna 'TOTAL_VACCINATIONS'.....	14
<b>Tabela 05</b>	Análise da coluna 'PERSONS_VACCINATED_1PLUS_DOSE'.....	15
<b>Tabela 06</b>	Análise da coluna 'TOTAL_VACCINATIONS_PER100'.....	15
<b>Tabela 07</b>	Análise da coluna 'PERSONS_VACCINATED_1PLUS_DOSE_PER100'.....	16
<b>Tabela 08</b>	Análise da coluna 'PERSONS_LAST_DOSE'.....	16
<b>Tabela 09</b>	Análise da coluna 'PERSONS_LAST_DOSE_PER100'.....	17
<b>Tabela 10</b>	Análise da coluna 'NUMBER_VACCINES_TYPES_USED'.....	18
<b>Tabela 11</b>	Análise da coluna 'PERSONS_BOOSTER_ADD_DOSE'.....	18
<b>Tabela 12</b>	Análise da coluna 'PERSONS_BOOSTER_ADD_DOSE_PER100'.....	18



## 1. Título:

Vacinação contra a COVID-19: Uma Análise Abrangência e Adesão à Vacinação

## 2. Introdução

A COVID-19, desde seu surgimento, evoluiu em termos de mutações, disseminação e impacto, necessitando de um combate efetivo e uma rápida resposta para conter as consequências da sua disseminação na população mundial.

Para conter a doença, foram desenvolvidas vacinas que desde 2021 passaram a ser disponibilizadas ao redor do mundo para todas as pessoas.

Neste projeto, utilizando pensamento computacional e análise exploratória, vamos analisar e entender a abrangência e adesão à vacinação entre os países ao redor do mundo.

## 3. Objetivos:

Entender o cenário de vacinação contra a COVID-19 nos 229 países que tiveram dados disponibilizados pela OMS (Organização Mundial de Saúde);

Aplicar técnicas de pensamento computacional e análise exploratória com o intuito de analisar dados de vacinação dos países;

Avaliar a abrangência e a adesão à vacinação, e com base nas análises, apontar os países que podem melhorar neste processo;

## 4. Metas:

- Desenvolvimento de uma análise exploratória dos dados de vacinação contra COVID-19;
- Produção de um relatório com recomendações para organizações com base nos resultados;
- Conclusão de uma revisão literária atualizada sobre a vacinação contra a COVID-19;



## 5. Cronograma:

<https://github.com/meddavid/Mackenzie/issues/7#issuecomment-1692416452>

Input	EDT	Nome da tarefa	Duração	Trabalho	Início	Término	Nomes dos recursos	IIRR	TTR
0%	1	PROJETO - Atualização Epidemiológica da COVID-19	75 dias?	0 hrs	Ter 08/08/23	Seg 20/11/23		Ter 08/08/23	ND
0%	1.1	FASE I – PREPARAÇÃO DE DATASET	15 dias?	0 hrs	Ter 08/08/23	Seg 28/08/23		Ter 08/08/23	ND
0%	1.1.1	Artefatos do projeto	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23		Qua 23/08/23	ND
100%	1.1.1.1	Link GITHUB	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Eduardo	Qua 23/08/23	ND
100%	1.1.1.2	Dataset	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Eduardo	Qua 23/08/23	ND
0%	1.1.2	Contexto do Estudo	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23			ND ND
100%	1.1.2.1	Premissas do Projeto	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.2.2	Objetivos	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.2.3	Metas	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.2.4	Cronograma de Atividades	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.2.5	Pensamento Computacional em contextos organizacionais	1 dia	0 hrs	Ter 08/08/23	Ter 08/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.1.3	Referências de Aquisição do Dataset	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23			ND ND
100%	1.1.3.1	Origem dos Dados	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.3.2	Limitação de Uso	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.3.3	Período de Coleta	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.1.4	Descrição da Origem	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23			ND ND
100%	1.1.4.1	Informações sobre a Organização que gerou os dados	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.4.2	Contexto em que os dados foram gerados	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.1.5	Descrição do Dataset	1 dia?	0 hrs	Qua 23/08/23	Qua 23/08/23			ND ND
100%	1.1.5.1	Conteúdo	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.5.2	Proposta	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
100%	1.1.5.3	Registro de Problemas ou Fenômenos	1 dia	0 hrs	Qua 23/08/23	Qua 23/08/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.1.6	Entrega Moodle	1 dia	0 hrs	Seg 28/08/23	Seg 28/08/23	Eduardo		ND ND
0%	1.2	FASE II - TBD	15 dias	0 hrs	Ter 05/09/23	Seg 25/09/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.2.1	BB Professor	1 dia	0 hrs	Ter 05/09/23	Ter 05/09/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.2.2	Atividades - TBD	1 dia	0 hrs	Ter 19/09/23	Ter 19/09/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.2.3	BB Professor	1 dia	0 hrs	Ter 19/09/23	Ter 19/09/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.2.4	Entrega Moodle	1 dia	0 hrs	Seg 25/09/23	Seg 25/09/23	Eduardo		ND ND
0%	1.3	FASE III - TBD	15 dias	0 hrs	Ter 10/10/23	Seg 30/10/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.3.1	BB Professor	1 dia	0 hrs	Ter 10/10/23	Ter 10/10/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.3.2	Atividades - TBD	1 dia	0 hrs	Ter 24/10/23	Ter 24/10/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.3.3	BB Professor	1 dia	0 hrs	Seg 30/10/23	Seg 30/10/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.3.4	Entrega Moodle	1 dia	0 hrs	Seg 30/10/23	Seg 30/10/23	Eduardo		ND ND
0%	1.4	FASE IV - TBD	10 dias	0 hrs	Ter 07/11/23	Seg 20/11/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.4.1	BB Professor	1 dia	0 hrs	Ter 07/11/23	Ter 07/11/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.4.2	Atividades - TBD	1 dia	0 hrs	Seg 20/11/23	Seg 20/11/23	Beatriz;Eduardo;Jessica;Moisés;Vanessa		ND ND
0%	1.4.3	Entrega Moodle	1 dia	0 hrs	Seg 20/11/23	Seg 20/11/23	Eduardo		ND ND

Tabela 1

## 6. Fluxo Baseado em Pensamento Computacional em Contextos Organizacionais:

### A. Decomposição:

Dividir o problema em partes menores: quantidade de países que reportam oficialmente os números de vacinação, adesão à primeira dose da vacina por país e por continente, adesão à segunda ou mais doses, relação entre os tipos de vacinas disponíveis versus número de pessoas vacinadas por país, entre outras.

### B. Reconhecimento de padrões:

Analisar tendências da vacinação: velocidade do início de vacinação por país e continente, relação entre os tipos de vacinas disponíveis versus número de pessoas vacinadas, entre outros.





## **C. Abstração:**

Construir uma análise exploratória sobre a vacinação contra COVID-19: Utilizar dados e pesquisas atuais para criar análise.

## **D. Design de Algoritmos:**

Criar um relatório para tomada de decisão: Com base nas análises, produzir relatório com recomendações para organizações.

## **7. Organização e o contexto em que os dados foram gerados:**

A OMS é a Organização Mundial de Saúde, um organismo internacional ligado ao Sistema ONU que tem por objetivo promover o acesso à saúde de qualidade a todos os povos do mundo. A OMS foi criada em 1948 e tem sua sede em Genebra, Suíça. A OMS coordena o trabalho internacional de saúde por meio da colaboração com países, organizações internacionais, sociedade civil, fundações, academia e instituições de pesquisa.

A OMS tem desempenhado um papel importante na resposta à pandemia de COVID-19, que foi declarada como uma Emergência de Saúde Pública de Importância Internacional em 30 de janeiro de 2020 e como uma pandemia em 11 de março de 2020. A OMS tem fornecido orientações, assistência técnica, recursos e informações sobre a doença, sua prevenção, seu tratamento e sua vacinação.

Mesmo após a pandemia, a OMS continuou com a atualização das informações sobre a COVID-19 em mais de 200 países onde está presente, inclusive atualizando os dados de vacinação que serão utilizados neste projeto para entendimento do cenário de vacinação entre os países.

## **8. Referências de aquisição do dataset:**

Os dados têm origem no site da Organização Mundial de Saúde que traz atualizações de 229 países sobre a vacinação contra COVID-19, que são reunidos a partir de inúmeras fontes, incluindo relatórios diretos dos Estados-Membros, análises da OMS de dados oficiais disponíveis ao público ou dados recolhidos e publicados por sites de terceiros, como o Our World in Data.

O período dos dados é de 22 de julho de 2020 a 20 de agosto de 2023.

## **9. Dataset e Metadados**

### **A. Dataset:**

Fonte: <https://covid19.who.int/data>

Link para download: <https://covid19.who.int/who-data/vaccination-data.csv>



## B. Descrição do Dataset:

O Dataset contém informações sobre a vacinação de países por região. As regiões divididas com base na classificação da Organização Mundial da Saúde (OMS) são: Região Africana (AFRO), Região das Américas (AMRO), Região do Sudeste Asiático (SEARO), Região Europeia (EURO), Região do Mediterrâneo Oriental (EMRO), Região do Pacífico Ocidental (WPRO).

O arquivo contém dados como o total de vacinação em cada país, pessoas que tomaram mais de uma dose da vacina, quantidade de tipos de vacinas usadas em cada país, quantidade de dose de reforço e data da primeira vacinação. Com isso conseguimos analisar quais regiões tiveram menos ou mais vacinas, quantas pessoas se vacinaram e etc.

A partir disso podemos entender a abrangência e adesão à vacinação, que causa um impacto positivo ainda que não prevenindo 100% a disseminação do vírus, mas modera casos mais graves e mortes por Covid-19.

## C. Metadados:

Colunas	Descrição
COUNTRY	País, território, área
ISO3	Código de país ISO Alpha-3
WHO_REGION	Escritórios regionais da OMS: Os Estados-Membros da OMS estão agrupados em seis regiões da OMS: Escritório Regional para África (AFRO), Escritório Regional para as Américas (AMRO), Escritório Regional para o Sudeste Asiático (SEARO), Escritório Regional para a Europa (EURO), Escritório Regional para o Mediterrâneo Oriental (EMRO) e Escritório Regional para o Pacífico Ocidental (WPRO).
DATA_SOURCE	Indica a fonte dos dados: - RELATÓRIO: Dados comunicados pelos Estados-Membros ou provenientes de relatórios oficiais - OWID: Dados provenientes do Our World in Data: <a href="https://ourworldindata.org/covid-vaccinations">https://ourworldindata.org/covid-vaccinations</a>
DATE_UPDATED	Data da última atualização
TOTAL_VACCINATIONS	Doses cumulativas totais de vacina administradas
PERSONS_VACCINATED_1PLUS_DOSE	Número acumulado de pessoas vacinadas com pelo menos uma dose
TOTAL_VACCINATIONS_PER100	Doses cumulativas totais de vacinas administradas por 100 habitantes
PERSONS_VACCINATED_1PLUS_DOSE_PER100	Cumulativo de pessoas vacinadas com pelo menos uma dose por 100 habitantes
PERSONS_LAST_DOSE	Número acumulado de pessoas vacinadas com uma série primária completa
PERSONS_LAST_DOSE_PER100	Número acumulado de pessoas vacinadas com uma série primária completa por 100 habitantes
VACCINES_USED	Nome abreviado combinado da vacina: "Empresa – Nome do produto"
FIRST_VACCINE_DATE	Data das primeiras vacinações. Equivalente à data de início/lançamento da primeira vacina administrada num país.
NUMBER_VACCINES_TYPES_USED	Número de tipos de vacinas utilizadas por país, território, área.
PERSONS_BOOSTER_ADD_DOSE	Número cumulativo de pessoas vacinadas com pelo menos uma dose de reforço ou dose adicional.
PERSONS_BOOSTER_ADD_DOSE_PER100	Número cumulativo de pessoas vacinadas com pelo menos uma dose de reforço ou dose adicional por 100 habitantes.

Tabela 2





**D. Tipo do Arquivo:**

O arquivo está no formato CSV.

**E. Sensibilidade:**

O arquivo não possui dados sensíveis e está compatível com a LGPD.

**F. Proprietário do Dado:**

O arquivo é de domínio público e foi disponibilizado pela OMS.

**G. Restrições de uso:**

A OMS não restringiu sua utilização.



## 10. Análise Exploratória:

A etapa de análise exploratória desempenha um papel fundamental na compreensão aprofundada do dataset disponível. Esta seção visa enriquecer o conhecimento sobre as variáveis contidas no conjunto de dados, oferecendo uma perspectiva detalhada que complementa as informações já fornecidas nos metadados iniciais..

### A. Descrição das Variáveis:

Nesta parte, serão apresentados detalhes abrangentes sobre as variáveis presentes no dataset. Isso incluirá informações como o número total de registros para cada variável, os valores máximo e mínimo, a variância e o desvio padrão. Além disso, serão exploradas as distribuições das variáveis, destacando se elas seguem padrões conhecidos, como uma distribuição normal ou binomial. Também será avaliada a presença de dados ausentes e a identificação de valores discrepantes (outliers) que possam influenciar análises subsequentes.

### 01. Arquivo: Não aplicado

**Caminho:** Não aplicado

#### **Desenvolvimento:**

- pip install pandas
- python.exe -m pip install --upgrade pip
- pip install matplotlib
- pip install seaborn

#### **Retorno:**

- Successfully installed numpy-1.26.0 pandas-2.1.1 python-dateutil-2.8.2 pytz-2023.3.post1 six-1.16.0 tzdata-2023.3
- Installing collected packages: pyparsing, pillow, packaging, kiwisolver, fonttools, cyclor, contourpy, matplotlib
- Successfully installed pip-23.2.1



## 02. Arquivo: open\_dataset.py

**Caminho:** .git\Projeto\_aplicado\_I\99. Artefatos\02. scripts\_python

### Desenvolvimento:

```
import pandas as pd

# Carregar o arquivo CSV em um DataFrame
url =
"https://raw.githubusercontent.com/meddavid/Mackenzie-
Projeto-Aplicado-
I/adf5948da5b4b0f8fc2bee7fa4dad0ac60f23102/99.%20Artefatos
/01.%20Dataset/vaccination-data.csv"
df = pd.read_csv(url)

# Mostra um resumo do DataFrame (tipos de dados, não
nulos, etc.)
df.info()
#print(df)
```

### Retorno:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 229 entries, 0 to 228

Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	----
0	COUNTRY	229 non-null	object
1	ISO3	229 non-null	object
2	WHO_REGION	229 non-null	object
3	DATA_SOURCE	229 non-null	object
4	DATE_UPDATED	229 non-null	object
5	TOTAL_VACCINATIONS	228 non-null	float64
6	PERSONS_VACCINATED_1PLUS_DOSE	229 non-null	int64
7	TOTAL_VACCINATIONS_PER100	224 non-null	float64
8	PERSONS_VACCINATED_1PLUS_DOSE_PER100	229 non-null	float64
9	PERSONS_LAST_DOSE	229 non-null	int64
10	PERSONS_LAST_DOSE_PER100	229 non-null	float64
11	VACCINES_USED	0 non-null	float64
12	FIRST_VACCINE_DATE	207 non-null	object
13	NUMBER_VACCINES_TYPES_USED	225 non-null	float64
14	PERSONS_BOOSTER_ADD_DOSE	213 non-null	float64
15	PERSONS_BOOSTER_ADD_DOSE_PER100	213 non-null	float64

dtypes: float64(8), int64(2), object(6)

Tabela 3

### Descrição:

Dataframe:

Número de Registros: 229

Número de colunas: 16



## 03. Arquivos: Analytics.py

### Desenvolvimento:

```
import pandas as pd

# Carregar o DataFrame do github

url =
"https://raw.githubusercontent.com/meddavid/Mackenzie-
Projeto-Aplicado-
I/adf5948da5b4b0f8fc2bee7fa4dad0ac60f23102/99.%20Artefatos/01.%20Dataset/vaccination-data.csv"
dados = pd.read_csv(url)

# Analisando Quantidade de Registros
num_registros = len(dados)
print(f"#### - Quantidade de registros:
{num_registros}\n")

# Analisando Quantidade de registros agrupados por
WHO_REGION
grupos = dados.groupby('WHO_REGION')
num_registros_por_grupo = grupos.size()
print("#### - Quantidade de registros agrupados por
WHO_REGION:")
print(num_registros_por_grupo)
print("\nGrupos formados:")
print(list(grupos.groups.keys()))
print()

# Analisando colunas numéricas
colunas_numericas = [
    'TOTAL_VACCINATIONS',
    'PERSONS_VACCINATED_1PLUS_DOSE',
    'TOTAL_VACCINATIONS_PER100',
    'PERSONS_VACCINATED_1PLUS_DOSE_PER100',
    'PERSONS_LAST_DOSE',
    'PERSONS_LAST_DOSE_PER100',
    'NUMBER_VACCINES_TYPES_USED',
    'PERSONS_BOOSTER_ADD_DOSE',
    'PERSONS_BOOSTER_ADD_DOSE_PER100'
]

for coluna in colunas_numericas:
    print(f"Análise da coluna '{coluna}':")
```



```
# Número de registros
num_registros_coluna = dados[coluna].count()
print(f"Número de registros: {num_registros_coluna}")

# Valor Máximo
valor_maximo = dados[coluna].max()
print(f"Valor Máximo: {valor_maximo:.2f}")

# Valor Mínimo
valor_minimo = dados[coluna].min()
print(f"Valor Mínimo: {valor_minimo:.2f}")

# Variância
variancia = dados[coluna].var()
print(f"Variância: {variancia:.2f}")

# Desvio Padrão
desvio_padrao = dados[coluna].std()
print(f"Desvio Padrão: {desvio_padrao:.2f}")

# Distribuição
distribuicao = dados[coluna].describe().apply(lambda
x: f'{x:.2f}')
print(f"Distribuição:\n{distribuicao}")

# Quantidade de NAs (dados faltantes)
quantidade_nas = dados[coluna].isna().sum()
print(f"Quantidade de NAs (dados faltantes):
{quantidade_nas:.2f}")

# Identificar outliers
Q1 = dados[coluna].quantile(0.25)
Q3 = dados[coluna].quantile(0.75)
IQR = Q3 - Q1
limite_inferior = Q1 - 1.5 * IQR
limite_superior = Q3 + 1.5 * IQR
outliers = dados[(dados[coluna] < limite_inferior) |
(dados[coluna] > limite_superior)]
print(f"Quantidade de outliers: {len(outliers)}")

print("\n")
```



## Retorno:

- Quantidade de registros: 229
- Quantidade de registros agrupados por WHO\_REGION:

### WHO\_REGION

AFRO 48  
AMRO 53  
EMRO 22  
EURO 60  
OTHER 1  
SEARO 10  
WPRO 35  
dtype: int64

- Grupos formados:  
['AFRO', 'AMRO', 'EMRO', 'EURO', 'OTHER', 'SEARO', 'WPRO']
- Análise da coluna 'TOTAL\_VACCINATIONS':

Métrica		Resultado
Número de registros		228
Valor Máximo		3515872818.00
Valor Mínimo		117.00
Variância		80599531141893616.00
Desvio Padrão		283900565.59
Distribuição		
	count :	228.00
	mean:	59204257.11
	std:	283900565.59
	min:	117.00
	25%:	473303.50
	50%:	4705035.00
	75%:	23357295.25
	max:	3515872818.00
<b>Name: TOTAL_VACCINATIONS, dtype: object</b>		
Quantidade de NAs (dados faltantes):		1
Quantidade de outliers:		36

Tabela 4





- Análise da coluna 'PERSONS\_VACCINATED\_1PLUS\_DOSE':

Métrica		Resultado
Número de registros		229
Valor Máximo		1317994878.00
Valor Mínimo		0.00
Variância		12966877158649760.00
Desvio Padrão		113872196.60
Distribuição		
	count :	229
	mean:	24411463.54
	std:	113872196.60
	min:	0.00
	25%:	191403.00
	50%:	2740227.00
	75%:	10884714.00
	max:	1317994878.00
<b>Name: PERSONS_VACCINATED_1PLUS_DOSE, dtype: object</b>		
Quantidade de NAs (dados faltantes):		0
Quantidade de outliers:		33

Tabela 5

- Análise da coluna 'TOTAL\_VACCINATIONS\_PER100':

Métrica		Resultado
Número de registros		224
Valor Máximo		469.78
Valor Mínimo		0.34
Variância		7079.81
Desvio Padrão		84.14
Distribuição		
	count :	224.00
	mean:	156.77
	std:	84.14
	min:	0.34
	25%:	84.74
	50%:	161.05
	75%:	222.96
	max:	469.78
<b>Name: TOTAL_VACCINATIONS_PER100, dtype: object</b>		
Quantidade de NAs (dados faltantes):		5
Quantidade de outliers:		1

Tabela 6



- **Análise da coluna 'PERSONS\_VACCINATED\_1PLUS\_DOSE\_PER100':**

Métrica		Resultado
Número de registros		229
Valor Máximo		<b>163.19</b>
Valor Mínimo		<b>0.00</b>
Variância		<b>639.80</b>
Desvio Padrão		<b>25.29</b>
Distribuição		
	count :	<b>229.00</b>
	mean:	<b>63.85</b>
	std:	<b>25.29</b>
	min:	<b>0.00</b>
	25%:	<b>45.45</b>
	50%:	<b>67.79</b>
	75%:	<b>82.60</b>
	max:	<b>163.19</b>
<b>Name: PERSONS_VACCINATED_1PLUS_DOSE_PER100, dtype: object</b>		
Quantidade de NAs (dados faltantes):		0
Quantidade de outliers:		1

*Tabela 7*

- **Análise da coluna 'PERSONS\_LAST\_DOSE':**

Métrica		Resultado
Número de registros		229
Valor Máximo		1284413713.00
Valor Mínimo		0.00
Variância		11753776143573370.00
Desvio Padrão		108414833.60
Distribuição		
	count :	229
	mean:	22490382.03
	std:	108414833.60
	min:	0.00
	25%:	184801.00
	50%:	2484985.00
	75%:	9209276.00
	max:	1284413713.00
<b>Name: PERSONS_LAST_DOSE, dtype: object</b>		
Quantidade de NAs (dados faltantes):		0
Quantidade de outliers:		34

*Tabela 8*



- Análise da coluna 'PERSONS\_LAST\_DOSE\_PER100':

Métrica		Resultado
Número de registros		229
Valor Máximo		163.19
Valor Mínimo		0.00
Variância		642.65
Desvio Padrão		25.35
Distribuição		
	count :	229
	mean:	58.88
	std:	25.35
	min:	0.00
	25%:	39.05
	50%:	63.59
	75%:	78.35
	max:	163.19
<b>Name: PERSONS_LAST_DOSE_PER100, dtype: object</b>		
Quantidade de NAs (dados faltantes):		0.00
Quantidade de outliers:		1

Tabela 9

- Análise da coluna 'NUMBER\_VACCINES\_TYPES\_USED':

Métrica		Resultado
Número de registros		225
Valor Máximo		12.00
Valor Mínimo		1.00
Variância		8.49
Desvio Padrão		2.91
Distribuição		
	count :	225
	mean:	4.98
	std:	2.91
	min:	1.00
	25%:	3.00
	50%:	4.00
	75%:	7.00
	max:	12.00
<b>Name: NUMBER_VACCINES_TYPES_USED, dtype: object</b>		
Quantidade de NAs (dados faltantes):		4
Quantidade de outliers:		0

Tabela 10



- Análise da coluna 'PERSONS\_BOOSTER\_ADD\_DOSE':

Métrica		Resultado
Número de registros		213
Valor Máximo		833820382.00
Valor Mínimo		0.00
Variância		3713688762559778.50
Desvio Padrão		60940042.36
Distribuição		
	count :	213
	mean:	11635797.94
	std:	60940042.36
	min:	0.00
	25%:	45446.00
	50%:	640419.00
	75%:	4474108.00
	max:	833820382.00
<b>Name: PERSONS_BOOSTER_ADD_DOSE, dtype: object</b>		
Quantidade de NAs (dados faltantes):		16.00
Quantidade de outliers:		32

Tabela 11

- Análise da coluna 'PERSONS\_BOOSTER\_ADD\_DOSE\_PER100':

Métrica		Resultado
Número de registros		213
Valor Máximo		145.94
Valor Mínimo		0.00
Variância		637.30
Desvio Padrão		25.24
Distribuição		
	count :	213
	mean:	32.25
	std:	25.24
	min:	0.00
	25%:	8.33
	50%:	30.84
	75%:	52.40
	max:	145.94
<b>Name: PERSONS_BOOSTER_ADD_DOSE_PER100, dtype: object</b>		
Quantidade de NAs (dados faltantes):		16.00
Quantidade de outliers:		1

Tabela 12



## **B. Desafios e Limitações:**

É fundamental reconhecer e discutir quaisquer desafios e limitações encontrados durante a análise exploratória. Isso pode incluir a identificação de dados faltantes em grande quantidade, a presença de outliers que exigirão considerações adicionais e quaisquer variáveis que não sigam uma distribuição esperada. Essa seção abordará as complexidades encontradas e fornecerá um contexto importante para decisões posteriores no processo de análise.

## **C. Recomendações Preliminares:**

Com base nas descobertas da análise exploratória, serão apresentadas recomendações preliminares. Isso pode envolver estratégias para lidar com dados faltantes, abordagens para tratar outliers e considerações sobre transformações de dados. As recomendações oferecerão diretrizes iniciais para orientar a próxima fase da análise, aproveitando os insights obtidos durante a exploração detalhada.

## **D. Conclusão da Análise Exploratória:**

A análise exploratória é um estágio crítico para a compreensão aprofundada do dataset, envolvendo a descrição minuciosa das variáveis, o uso de visualizações para destacar padrões, o reconhecimento de desafios e limitações, e a formulação de recomendações iniciais para abordar questões identificadas. Isso proporciona uma base sólida para análises subsequentes e tomadas de decisão informadas..



## 11. Gráficos e Visualizações:

Esta parte é dedicada à criação e apresentação de gráficos e visualizações que ajudarão na compreensão das características das variáveis. Isso inclui gráficos de barra, histogramas, box Plots e outras representações visuais adequadas à natureza dos dados.

**Caminho:** .git\Projeto\_aplicado\_I\99. Artefatos\03. GRAFICOS

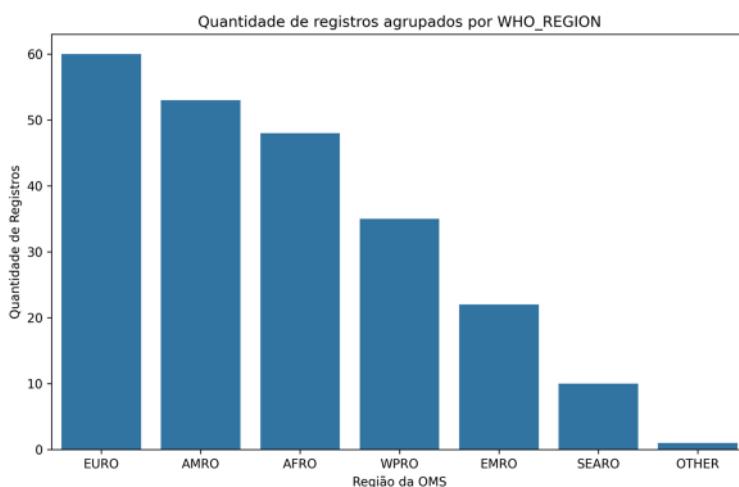
### A. Gráfico 01: Quantidade de registros agrupados por WHO\_REGION

**Tipo do Gráfico:** Barras

**Escala do Gráfico:** 0 a 60

**Eixo Y:** Quantidade de Registros por região

**Eixo X:** Regiões da OMS



### B. Gráfico 02: Histograma da coluna TOTAL\_VACCINATION\_PER100

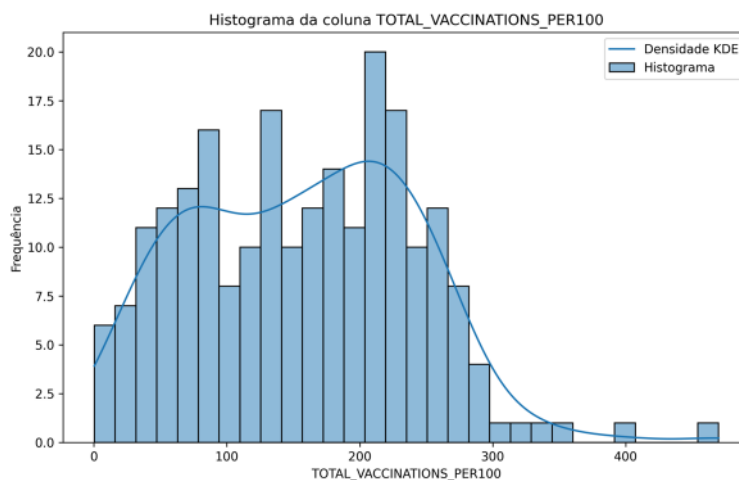
**Tipo do Gráfico:** Histograma

**Escala do Gráfico:** Y: 0.0 a 20.0 X:0 A 400

**Eixo Y:** Frequência

**Eixo X:** TOTAL\_VACCINATIONS\_PER100

**Medidas:** Densidade KDE / Histograma







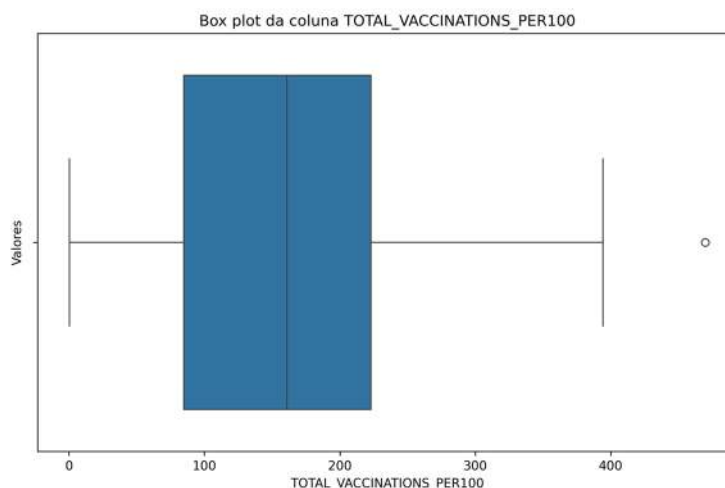
C. **Gráfico 03:** Box Plot da coluna TOTAL\_VACCINATION\_PER100

**Tipo do Gráfico:** Box plot

**Escala do Gráfico:** X:0 A 400

**Eixo Y:** Valores

**Eixo X:** TOTAL\_VACCINATIONS\_PER100



D. **Gráfico 04:** Histograma da coluna PERSONS\_VACCINATED\_1PLUS\_DOSE\_PER100

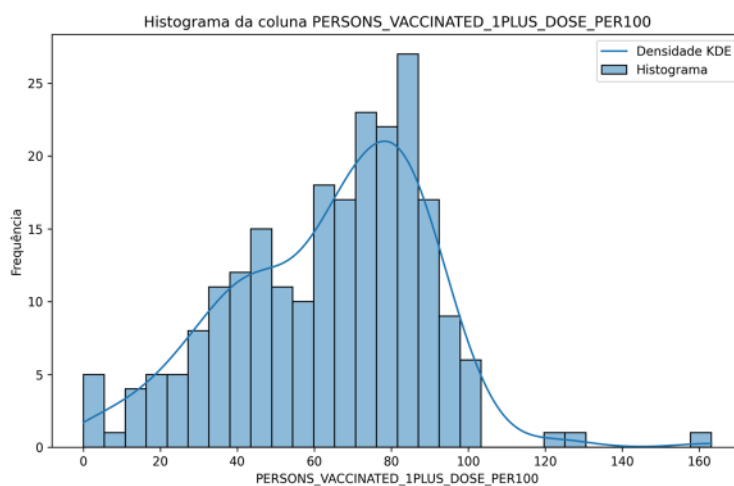
**Tipo do Gráfico:** Histograma

**Escala do Gráfico:** Y: 0 a 25 X: 0 A 160

**Eixo Y:** Frequência

**Eixo X:** PERSONS\_VACCINATED\_1PLUS\_DOSE\_PER100

**Medidas:** Densidade KDE / Histograma





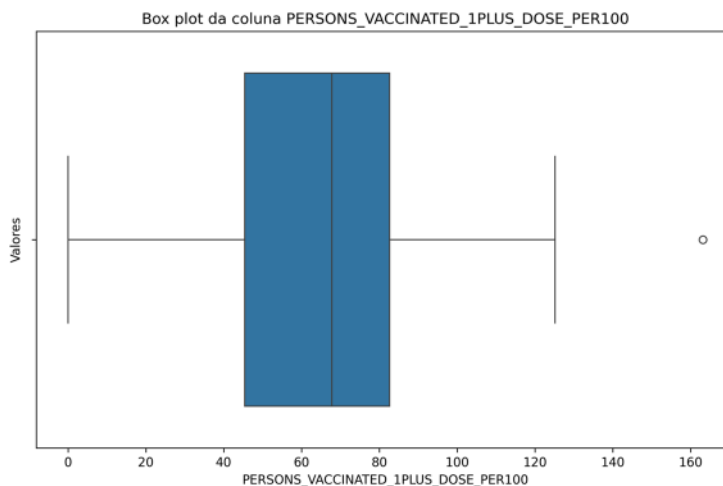
E. **Gráfico 05:** Box Plot da coluna PERSONS\_VACCINATED\_1PLUS\_DOSE\_PER100

**Tipo do Gráfico:** Box plot

**Escala do Gráfico:** X:0 A 400

**Eixo Y:** Valores

**Eixo X:** PERSONS\_VACCINATED\_1PLUS\_DOSE\_PER100



F. **Gráfico 06:** Histograma da coluna PERSONS\_LAST\_DOSE\_PER100

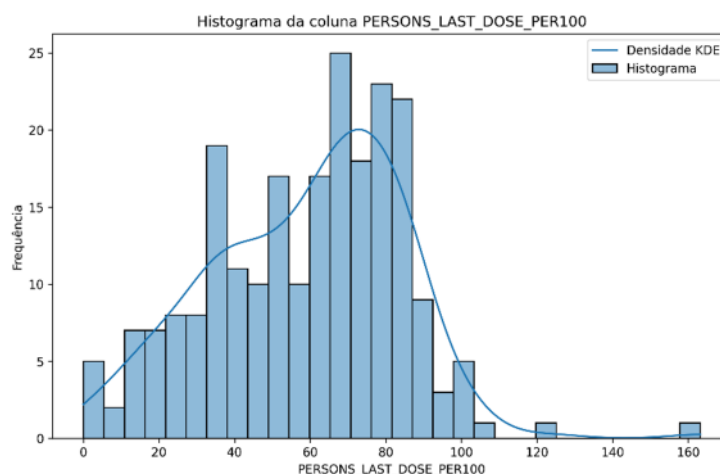
**Tipo do Gráfico:** Histograma

**Escala do Gráfico:** Y: 0 a 25 X: 0 A 160

**Eixo Y:** Frequência

**Eixo X:** PERSONS\_LAST\_DOSE\_PER100

**Medidas:** Densidade KDE / Histograma





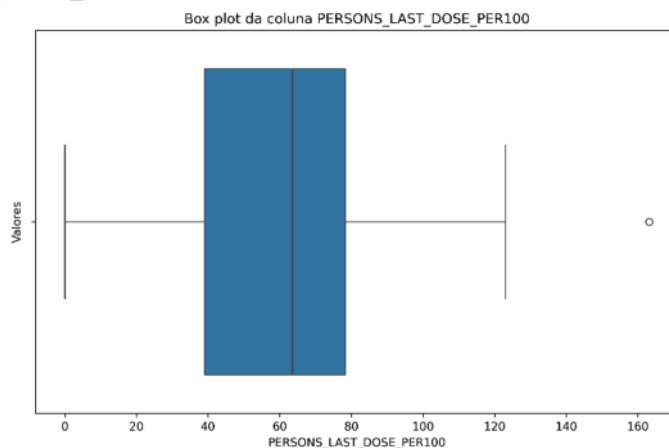
## G. **Gráfico 07:** Box Plot da coluna PERSONS\_LAST\_DOSE\_PER100

**Tipo do Gráfico:** Box plot

**Escala do Gráfico:** X:0 A 160

**Eixo Y:** Valores

**Eixo X:** PERSONS\_LAST\_DOSE\_PER100



## H. **Gráfico 08:** Histograma da coluna NUMBER\_VACCINES\_TYPE\_USED

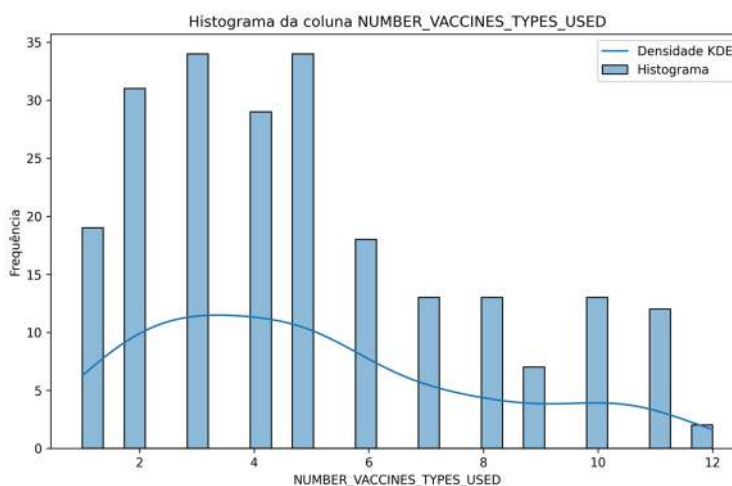
**Tipo do Gráfico:** Histograma

**Escala do Gráfico:** Y: 0 a 35 X: 0 A 12

**Eixo Y:** Frequência

**Eixo X:** NUMBER\_VACCINES\_TYPE\_USED

**Medidas:** Densidade KDE / Histograma





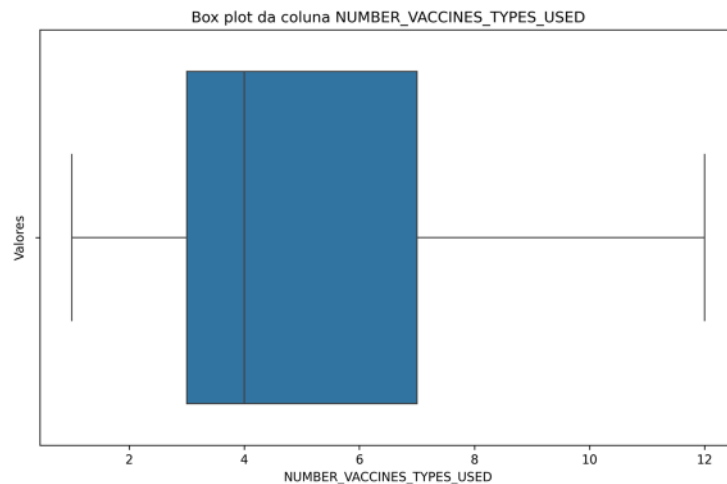
## I. Gráfico 09: Box Plot da coluna NUMBER\_VACCINES\_TYPE\_USED

Tipo do Gráfico: Box Plot

Escala do Gráfico: X:0 A 12

Eixo Y: Valores

Eixo X: NUMBER\_VACCINES\_TYPE\_USED



## J. Código Fonte Gráficos:

### 01. Arquivo: Graphics\_OF.py

Caminho: .git\Projeto\_aplicado\_I\99. Artefatos\02. scripts\_python

### Desenvolvimento:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Carregar o DataFrame a partir de uma URL do github
url =
"https://raw.githubusercontent.com/meddavid/Mackenzie-
Projeto-Aplicado-
I/adf5948da5b4b0f8fc2bee7fa4dad0ac60f23102/99.%20Artefat
os/01.%20Dataset/vaccination-data.csv"
dados = pd.read_csv(url)

# Gráfico de barras para mostrar a quantidade de
registros por WHO_REGION
plt.figure(figsize=(10, 6)) # Define o tamanho da
figura
sns.countplot(data=dados, x='WHO_REGION') # Cria um
gráfico de barras usando Seaborn
plt.title('Quantidade de registros agrupados por
WHO_REGION') # Adiciona um título ao gráfico
plt.xlabel('Região da OMS') # Rótulo do eixo X
plt.ylabel('Quantidade de Registros') # Rótulo do eixo
Y
```



```
plt.savefig('WHO_REGION_bar_plot.png', dpi=300) # Salva
o gráfico como .png com alta resolução
plt.close() # Fecha a figura atual

# Lista das colunas numéricas que serão analisadas
colunas_numericas = [
    'TOTAL_VACCINATIONS_PER100',
    'PERSONS_VACCINATED_1PLUS_DOSE_PER100',
    'PERSONS_LAST_DOSE_PER100',
    'NUMBER_VACCINES_TYPES_USED'
]

# Loop para gerar histogramas e box plots para cada
coluna numérica
for coluna in colunas_numericas:
    # Histograma
    plt.figure(figsize=(10, 6))
    sns.histplot(dados[coluna], bins=30, kde=True) #
Gera histograma com densidade KDE
    plt.title(f'Histograma da coluna {coluna}') #
Adiciona um título ao histograma
    plt.xlabel(coluna) # Rótulo do eixo X
    plt.ylabel('Frequência') # Rótulo do eixo Y
    plt.legend(labels=['Densidade KDE', 'Histograma'])
# Adiciona uma legenda
    plt.savefig(f'{coluna}_histogram.png', dpi=300) #
Salva o histograma como .png
    plt.close() # Fecha a figura atual

    # Boxplot
    plt.figure(figsize=(10, 6))
    sns.boxplot(x=dados[coluna]) # Gera box plot usando
Seaborn
    plt.title(f'Box plot da coluna {coluna}') #
Adiciona um título ao box plot
    plt.xlabel(coluna) # Rótulo do eixo X
    plt.ylabel('Valores') # Rótulo do eixo Y
    plt.savefig(f'{coluna}_boxplot.png', dpi=300) #
Salva o box plot como .png
    plt.close() # Fecha a figura atual
```



## **12. Pipeline de Dados:**

### **A. Coleta:**

- Processo de obtenção dos dados da OMS.

### **B. Limpeza:**

- Remoção de dados faltantes ou inconsistências;
- Transformação de variáveis.

### **C. Análise Exploratória:**

- Estatísticas descritivas (médias, medianas, desvios padrão, etc.);
- Visualizações gráficas (gráficos de barra, mapas de calor, etc.).

### **D. Técnicas de Análise:**

- Algoritmos e ferramentas a serem utilizados (e.g., Python, R);
- Métodos de análise (regressão, clustering, análise de séries temporais).mapas de calo

### **E. Visualização:**

- Ferramentas e abordagens para visualizar os resultados (e.g., Tableau, Power BI, gráficos em Python/R).

### **F. Relatório Final:**

Compilação dos resultados e recomendações em um formato acessível e compreensível





## 13. Proposta Analítica:

Nesta proposta, vamos mergulhar no mundo da vacinação contra a COVID-19 em 229 países, utilizando dados fornecidos pela OMS. Nosso principal objetivo é descobrir como os países avançaram em seus programas de vacinação e identificar aqueles que ainda têm espaço para melhorias.

Usaremos técnicas básicas de análise para avaliar esses dados e criar representações visuais que tornem as informações mais acessíveis e fáceis de entender.

Esperamos que nosso estudo ofereça uma compreensão aprofundada sobre a origem, evolução e repercussões da pandemia na saúde global e na vacinação. Almejamos destacar as conquistas alcançadas e identificar áreas que necessitam de atenção e esforço adicional para combater efetivamente a pandemia.

### A. Apresentação da OMS (*Organização Mundial de Saúde*):

#### Breve histórico:

- Estabelecida em 1948, a OMS rapidamente emergiu como a autoridade central em questões de saúde global. Desde sua fundação, a organização tem ampliado sua influência e capacidade, liderando iniciativas internacionais para melhorar a saúde e prevenir enfermidades. Através das décadas, a OMS tem sido instrumental em enfrentar diversas crises de saúde e em lançar campanhas para erradicar doenças, definindo, assim, padrões e diretrizes que têm sido pilares na saúde mundial.

#### Missão e objetivos:

- A missão fundamental da OMS é agir como a guardiã da saúde em escala global, orientando e consolidando esforços de nações ao redor do mundo para superar desafios na área da saúde. A OMS almeja objetivos claros e nobres: assegurar acesso universal a serviços de saúde de excelência, salvaguardar indivíduos de emergências sanitárias e fomentar o bem-estar em todas as etapas da vida humana.

#### Importância global:

- A OMS tem se destacado como um pilar durante crises de saúde, incluindo os surtos de Ebola e H1N1, onde forneceu diretrizes cruciais e mobilizou recursos indispensáveis. Ela estabelece parcerias sólidas com governos, ONGs e o setor



privado, potencializando esforços conjuntos para avançar em iniciativas de saúde global. Atualmente, sua relevância é ainda mais evidente na batalha contra a COVID-19, onde a OMS desempenha um papel central desde a investigação científica até a distribuição de tratamentos e imunizantes.

### **Estrutura e funcionamento:**

- A OMS opera através de uma estrutura organizada, abrangendo escritórios regionais em diversos continentes e mantendo sua sede central em Genebra. Esta configuração assegura sua atuação e influência em escala global. Dois órgãos proeminentes dentro da OMS são a Assembleia Mundial da Saúde e o Conselho Executivo. Ambos são vitais na formulação de políticas e estabelecimento de diretrizes, garantindo que a entidade mantenha um alinhamento consistente com seus objetivos estratégicos.

### **Iniciativas relativas à COVID-19:**

- Desde os primeiros sinais da pandemia, a OMS tem sido uma referência em diretrizes e tem alocado recursos para combater a COVID-19, apoiando países em suas ações. A OMS tem sido essencial não só na orientação, mas também na pesquisa e distribuição de vacinas contra a COVID-19. Projetos como o COVAX, sob sua liderança, são fundamentais para garantir uma distribuição justa das vacinas pelo mundo.



## **14. Storytelling:**

O storytelling é uma das ferramentas mais poderosas que temos para comunicar e conectar com os outros. É através da narrativa que conseguimos trazer significado, contexto e humanidade aos dados, fatos e estatísticas. Neste contexto, a pandemia da COVID-19, uma das crises mais desafiadoras da nossa era, requer uma narrativa cuidadosa e precisa, para capturar sua amplitude e impacto. A seguir, delineamos o esboço do storytelling que guiará nossa discussão sobre a COVID-19, garantindo que abordemos os pontos cruciais de forma sistemática.

### **14.1. Apresentação do Grupo**

- 14.1.1. Introdução aos Membros da Equipe
- 14.1.2. Breve Histórico e Experiências Anteriores
- 14.1.3. Motivação para Escolher o Tema

### **14.2. Nome do Projeto**

- 14.2.1. Justificativa para o Nome Escolhido
- 14.2.2. Significado e Impacto

### **14.3. Empresa/Organização de Estudo**

- 14.3.1. Histórico da Empresa/Organização
- 14.3.2. Papel da Empresa/Organização na Pandemia
- 14.3.3. Importância e Contribuições para o Combate à COVID-19

### **14.4. Área do Problema**

- 14.4.1. Contextualização da COVID-19 Globalmente
- 14.4.2. Desafios Enfrentados por Empresas e Organizações
- 14.4.3. Impacto Socioeconômico e na Saúde Pública

### **14.5. Descrição do Problema / Gap**

- 14.5.1. Detalhamento da Situação Atual
- 14.5.2. Consequências do Problema
- 14.5.3. Gaps e Áreas de Oportunidade Identificadas



#### **14.6. Proposta Analítica**

- 14.6.1. Metodologia Adotada
- 14.6.2. Objetivos da Análise
- 14.6.3. Ferramentas e Técnicas Utilizadas

#### **14.7. Dados Disponíveis**

- 14.7.1. Fonte de Dados
- 14.7.2. Qualidade e Confiabilidade dos Dados
- 14.7.3. Limitações e Desafios na Coleta de Dados

#### **14.8. Análise Exploratória**

- 14.8.1. Primeiras Impressões e Descobertas
- 14.8.2. Visualizações e Gráficos Relevantes
- 14.8.3. Correlações e Tendências Identificadas

#### **14.9. Resultados Pretendidos**

- 14.9.1. Impacto Esperado da Análise
- 14.9.2. Contribuições para a Empresa/Organização e Sociedade
- 14.9.3. Passos Futuros e Recomendações



## 15. Artefatos do Projeto:

### A. Link Github:

<https://github.com/meddavid/Mackenzie.git>

### B. Link Projeto:

<https://github.com/users/meddavid/projects/1>

### C. Link Artefatos do Projeto:

Scripts	<a href="https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/02.%20scripts_python">https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/02.%20scripts_python</a>
Dataset	<a href="https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/01.%20Dataset">https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/1bcfc57e6e34b1c4621dc602b1d10a616addbc4a/99.%20Artefatos/01.%20Dataset</a>

### D. Link Vídeo:

Youtube: <https://youtu.be/lhsijv1mmJY>

GitHub: <https://github.com/meddavid/Mackenzie-Projeto-Aplicado-I/tree/87dcbcb7f2a264084ae69125900049f95be5f58/99.%20Artefatos/04.%20Video>