



UNIVERSIDADE PRESBITERIANA MACKENZIE
CURSO DE GRADUAÇÃO EM CIÊNCIA DE DADOS

PROJETO APLICADO I

CASOS COVID-19: dados estatísticos da OMS

São Paulo

2024

SUMÁRIO

1	INTRODUÇÃO.....	05
2	SOBRE A ORGANIZAÇÃO.....	05
2.1	ÁREA DE ATUAÇÃO.....	05
2.1.1	Necessidades e problemas.....	06
3	APRESENTAÇÃO DOS DADOS.....	07
3.1	ANÁLISE EXPLORATÓRIA.....	09
3.1.1	Resultados.....	12
4	GITHUB.....	14
	CRONOGRAMA.....	15
	MEMBROS DO PROJETO.....	16

LISTA DE FIGURAS

Figura 1 -	Casos recentes de COVID-19 relatados à OMS	06
Figura 2 -	Total de casos de COVID-19 notificados à OMS	07
Figura 3 –	Dados de casos e mortes de COVID-19	08
Figura 4 -	Cronograma	15

LISTA DE GRÁFICOS

Gráfico 1 - Média de novos casos de COVID-19 no país	12
Gráfico 2 - Tendência temporal de casos cumulativos de COVID-19	13
Gráfico 3 - Coeficiente de variação dos casos de COVID-19 por país	13

1 INTRODUÇÃO

Em um mundo cada vez mais orientado por dados, compreender e analisar informações relevantes se torna crucial para o sucesso de empresas e organizações de diversos setores. Nesse contexto, o desenvolvimento de estudos práticos que explorem dados reais de uma empresa torna-se uma oportunidade valiosa para aplicar conhecimentos teóricos em um contexto real e promover aprendizado significativo.

O presente projeto tem como objetivo principal realizar um estudo prático sobre os dados disponibilizados pela Organização Mundial da Saúde (OMS) sobre os casos atualizados de COVID-19, com o intuito de explorar e analisar informações que possam gerar insights para a compreensão do funcionamento e desempenho do negócio.

2 SOBRE A ORGANIZAÇÃO

A Organização mundial da Saúde (OMS) é uma agência especializada das Nações Unidas dedicada a promover a saúde global, prevenir doenças e melhorar as condições de vida das pessoas em todo o mundo; composta por 194 Estados Membros, a OMS desempenha papel fundamental na coordenação de esforços internacionais para combater a pandemia da COVID-19 desde seu surgimento.

2.1 ÁREA DE ATUAÇÃO

Este projeto concentra-se na análise exploratória dos dados diários de casos e mortes por COVID-19 relatados à OMS. O objetivo é examinar a evolução temporal da pandemia, identificar padrões, tendências e insights relevantes que possam contribuir para uma compreensão mais profunda da disseminação da doença e seus impactos.

2.1.1 Necessidades e problemas

- 1. Identificação de Tendências Temporais:** A análise exploratória de dados permitirá identificar tendências temporais na incidência de casos diários e mortes por COVID-19 ao longo do tempo, incluindo variações geográficas na propagação da doença entre diferentes países e regiões.
- 2. Avaliação da Carga da Doença:** Será possível avaliar a magnitude do impacto da COVID-19 em diferentes países e regiões, examinando o número total de casos e mortes relatados e comparando esses dados entre diferentes áreas geográficas.
- 3. Análise de Disparidades e Desigualdades:** A análise dos dados permitirá identificar disparidades e desigualdades na distribuição da COVID-19, incluindo diferenças no acesso aos serviços de saúde, respostas governamentais e impacto socioeconômico da pandemia.
- 4. Avaliação da Efetividade de Medidas de Controle:** Será possível avaliar a efetividade das medidas de controle implementadas pelos países para conter a propagação da COVID-19, identificando estratégias bem-sucedidas e áreas que requerem maior atenção.

Figura 1 – Casos recentes de COVID-19 relatados à OMS (semanalmente)

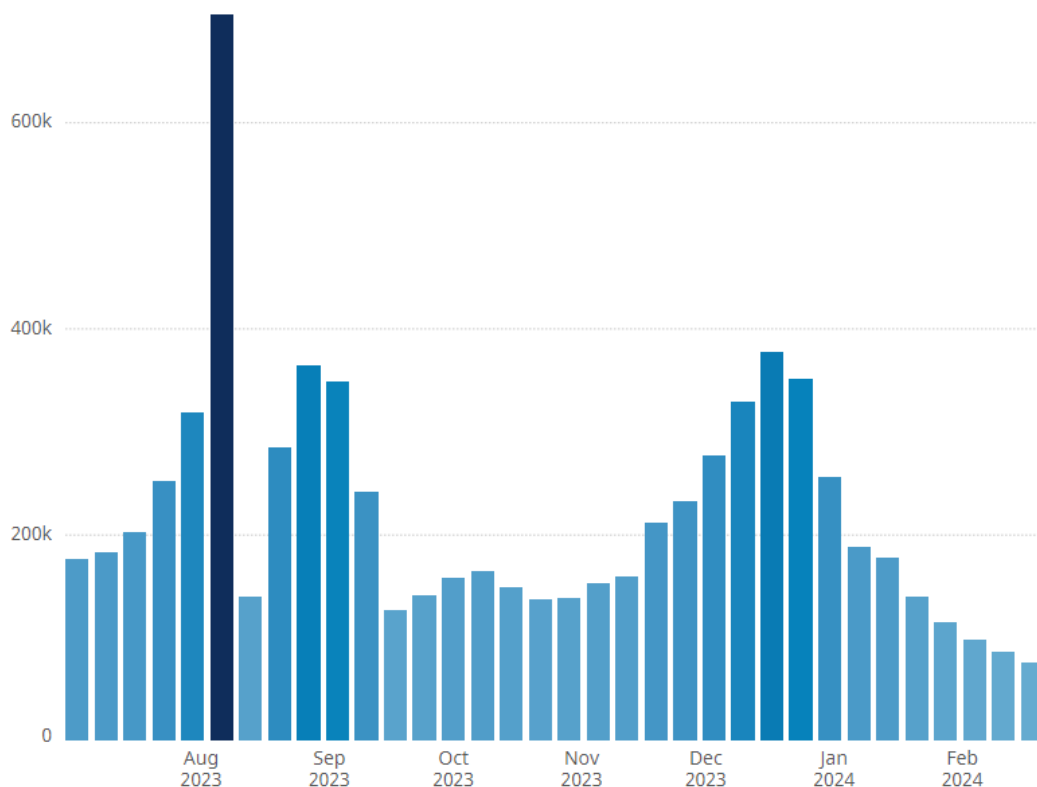
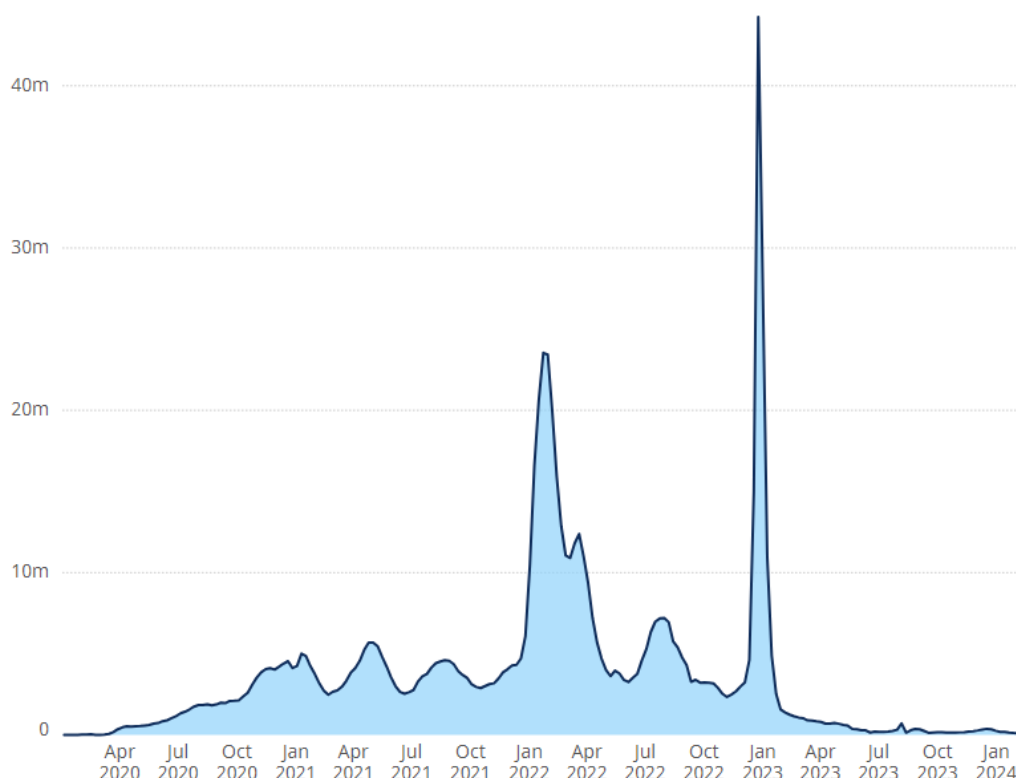


Figura 2 – Total de casos de COVID-19 notificados à OMS (semanalmente)



Fonte: Organização Mundial da Saúde

3 APRESENTAÇÃO DOS DADOS

O painel de controle do coronavírus (COVID-19) da OMS apresenta contagens semanais oficiais de casos, mortes e utilização de vacinas de COVID-19 relatados por países, territórios e áreas. Através destes dados, serão fornecidos recursos de visualização, disseminação e exploração de dados atualizados frequentemente.

Em relação às fontes, os dados de casos e mortes de COVID-19 são recolhidos através de comunicações oficiais ao abrigo do Regulamento Sanitário Internacional (RSI, 2005), complementadas pela monitorização dos sites oficiais dos ministérios da saúde; já os dados de vacinação contra a COVID-19 são recolhidos através de sistemas específicos dos escritórios regionais da OMS para monitorizar a adesão à vacinação contra a COVID-19. Estes dados são reunidos a partir de

inúmeras fontes, incluindo relatórios diretos dos Estados-Membros e análises da OMS de dados oficiais disponíveis ao público.

Para realizar uma análise exploratória nestes dados, alguns passos serão seguidos:

1. Limpeza e preparação dos dados:

- Tratar valores ausentes, dados duplicados e outliers;

2. Exploração Inicial dos dados:

- Visualizar as primeiras linhas dos dados para entender a estrutura;
- Utilizar funções como 'summary()' para obter uma visão geral estatística dos dados.

3. Análise temporal:

- Calcular estatísticas resumidas para períodos específicos, como médias semanais ou mensais.

4. Relacionamentos entre variáveis:

- Explorar correlações entre variáveis, como número de casos e número de mortes, utilizando gráficos de dispersão.

5. Análise por subgrupo:

- Analisar os dados por grupos demográficos, regiões geográficas, faixas etárias (para identificar padrões e diferenças).

6. Comunicação dos resultados:

- Compartilhar visualizações e insights derivados da análise exploratória de dados de forma clara e compreensível;
- Utilizar gráficos elaborados para comunicar resultados de maneira eficaz.

Figura 3 – Dados de casos e mortes de COVID-19

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
	<chr>	<chr>	<chr>	<chr>	<int>	<int>	<int>	<int>
1	18/02/24	AF	Afghanistan	EMRO	211	231990	NA	7982
2	18/02/24	AL	Albania	EURO	NA	334863	NA	3605
3	18/02/24	DZ	Algeria	AFRO	NA	272010	NA	6881
4	18/02/24	AS	American Samoa	WPRO	NA	8359	NA	34
5	18/02/24	AD	Andorra	EURO	NA	48015	NA	159
6	18/02/24	AO	Angola	AFRO	2	107325	NA	1937

3.1 ANÁLISE EXPLORATÓRIA

Nesta seção será apresentada uma análise exploratória de dados, complementando os dados apresentados anteriormente, examinando detalhadamente cada variável em nosso conjunto de dados; serão fornecidos insights iniciais sobre a natureza e a qualidade dos dados disponíveis.

- Dados de casos e mortes de COVID-19

Número de colunas e tipos de dados

```
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date_reported          51840 non-null  object
1   Country_code            51624 non-null  object
2   Country                 51840 non-null  object
3   WHO_region              47952 non-null  object
4   New_cases               36519 non-null  float64
5   Cumulative_cases        51840 non-null  int64
6   New_deaths              24217 non-null  float64
7   Cumulative_deaths       51840 non-null  int64
dtypes: float64(2), int64(2), object(4)
memory usage: 3.2+ MB
```

Contagem de valores, média, desvio padrão, valor mínimo, quartis e valor máximo

	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
count	3.651900e+04	5.184000e+04	24217.00000	5.184000e+04
mean	2.121360e+04	1.626154e+06	290.43358	1.892637e+04
std	2.798800e+05	7.272572e+06	1232.76986	7.860001e+04
min	-6.507900e+04	0.000000e+00	-3432.00000	0.000000e+00
25%	5.200000e+01	2.943000e+03	4.00000	2.100000e+01
50%	4.760000e+02	3.904600e+04	22.00000	4.150000e+02
75%	4.549000e+03	4.769692e+05	110.00000	6.249000e+03
max	4.047548e+07	1.034368e+08	47687.00000	1.177223e+06

Uma compreensão mais profunda das variáveis foi fornecida pela análise exploratória. Identificamos valores extremos e a presença de valores ausentes que podem afetar nossa análise. Estes dados nos ajudarão a tomar decisões sobre o próximo passo em nossa análise de dados.

O mesmo procedimento foi aplicado na análise dos dados de vacinação a seguir:

- Dados de vacinação

Número de colunas e tipos de dados

```
Data columns (total 16 columns):
#      Column                                     Non-Null Count  Dtype
---  -
0     COUNTRY                                     210 non-null    object
1     ISO3                                       210 non-null    object
2     WHO_REGION                                206 non-null    object
3     DATA_SOURCE                             210 non-null    object
4     DATE_UPDATED                             203 non-null    object
5     TOTAL_VACCINATIONS                       204 non-null    float64
6     PERSONS_VACCINATED_1PLUS_DOSE            204 non-null    float64
7     TOTAL_VACCINATIONS_PER100                202 non-null    float64
8     PERSONS_VACCINATED_1PLUS_DOSE_PER100     202 non-null    float64
9     PERSONS_LAST_DOSE                        204 non-null    float64
10    PERSONS_LAST_DOSE_PER100                 202 non-null    float64
11    VACCINES_USED                             0 non-null      float64
12    FIRST_VACCINE_DATE                       196 non-null    object
13    NUMBER_VACCINES_TYPES_USED               0 non-null      float64
14    PERSONS_BOOSTER_ADD_DOSE                 190 non-null    float64
15    PERSONS_BOOSTER_ADD_DOSE_PER100          179 non-null    float64
dtypes: float64(10), object(6)
memory usage: 26.4+ KB
None
```

Contagem de valores, média, desvio padrão, valor mínimo, quartis e valor máximo

	TOTAL_VACCINATIONS	PERSONS_VACCINATED_1PLUS_DOSE	\
count	2.040000e+02	2.040000e+02	
mean	6.667101e+07	2.746332e+07	
std	3.000893e+08	1.204356e+08	
min	4.619000e+03	1.638000e+03	
25%	9.638368e+05	5.154345e+05	
50%	7.751394e+06	3.954866e+06	
75%	2.742228e+07	1.402173e+07	
max	3.516881e+09	1.318027e+09	

	TOTAL_VACCINATIONS_PER100	PERSONS_VACCINATED_1PLUS_DOSE_PER100	\
count	202.000000	202.000000	
mean	152.762376	62.801980	
std	84.654363	23.943557	
min	4.000000	4.000000	
25%	81.250000	45.250000	
50%	154.000000	67.000000	
75%	220.000000	83.000000	
max	470.000000	100.000000	

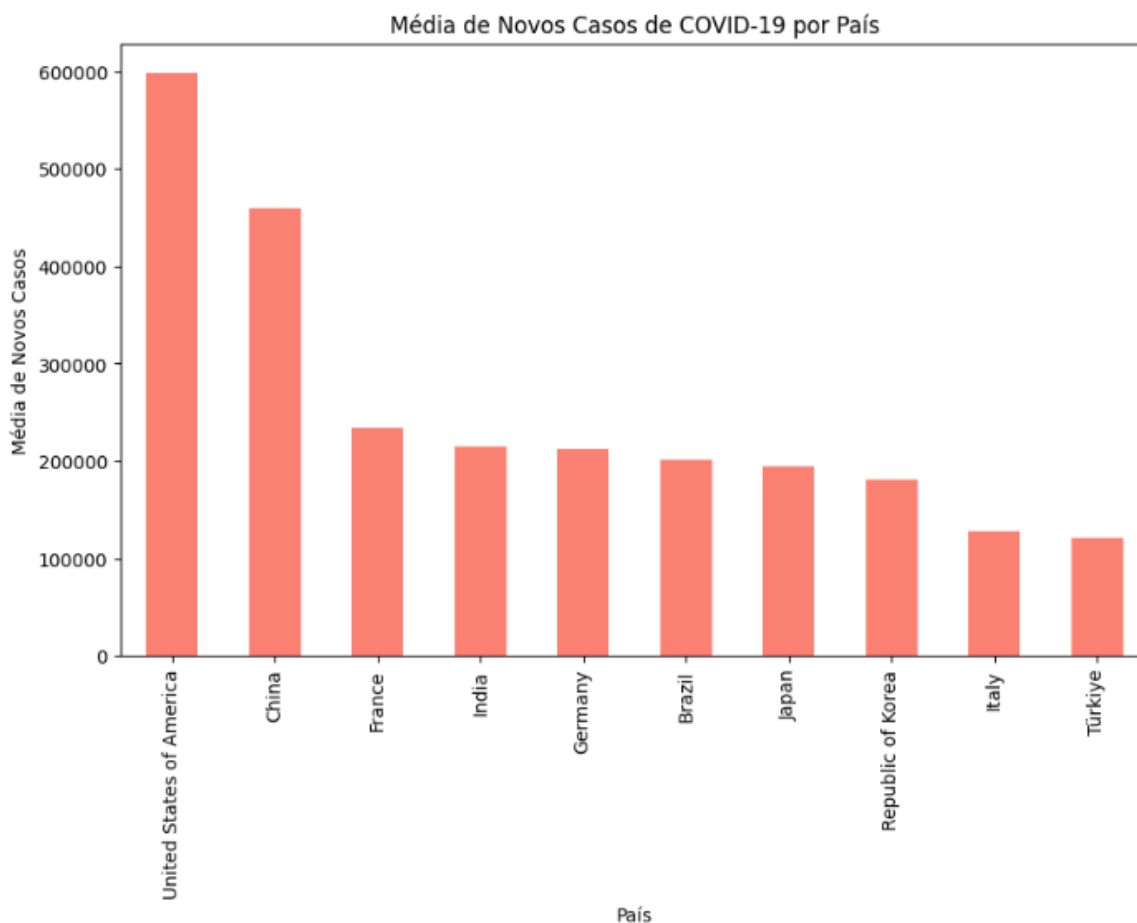
	PERSONS_LAST_DOSE	PERSONS_LAST_DOSE_PER100	VACCINES_USED	\
count	2.040000e+02	202.000000	0.0	
mean	2.531342e+07	57.866337	NaN	
std	1.146467e+08	24.131830	NaN	
min	1.635000e+03	3.000000	NaN	
25%	4.462615e+05	39.250000	NaN	
50%	3.591682e+06	63.000000	NaN	
75%	1.222551e+07	77.750000	NaN	
max	1.284480e+09	100.000000	NaN	

	NUMBER_VACCINES_TYPES_USED	PERSONS_BOOSTER_ADD_DOSE	\
count	0.0	1.900000e+02	
mean	NaN	1.312432e+07	
std	NaN	6.444804e+07	
min	NaN	3.140000e+02	
25%	NaN	8.050650e+04	
50%	NaN	8.676245e+05	
75%	NaN	5.374504e+06	
max	NaN	8.340601e+08	

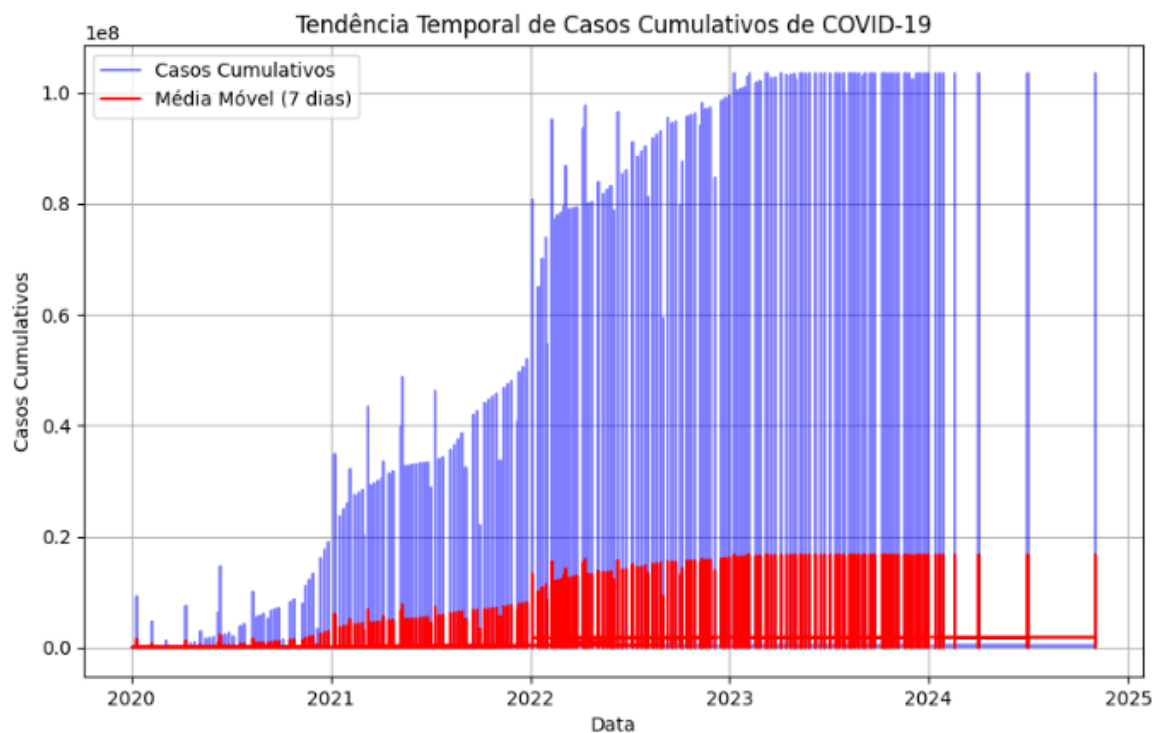
	PERSONS_BOOSTER_ADD_DOSE_PER100
count	179.000000
mean	32.681564
std	23.457690
min	1.000000
25%	9.500000
50%	31.000000
75%	54.500000
max	83.000000

3.1.1 Resultados

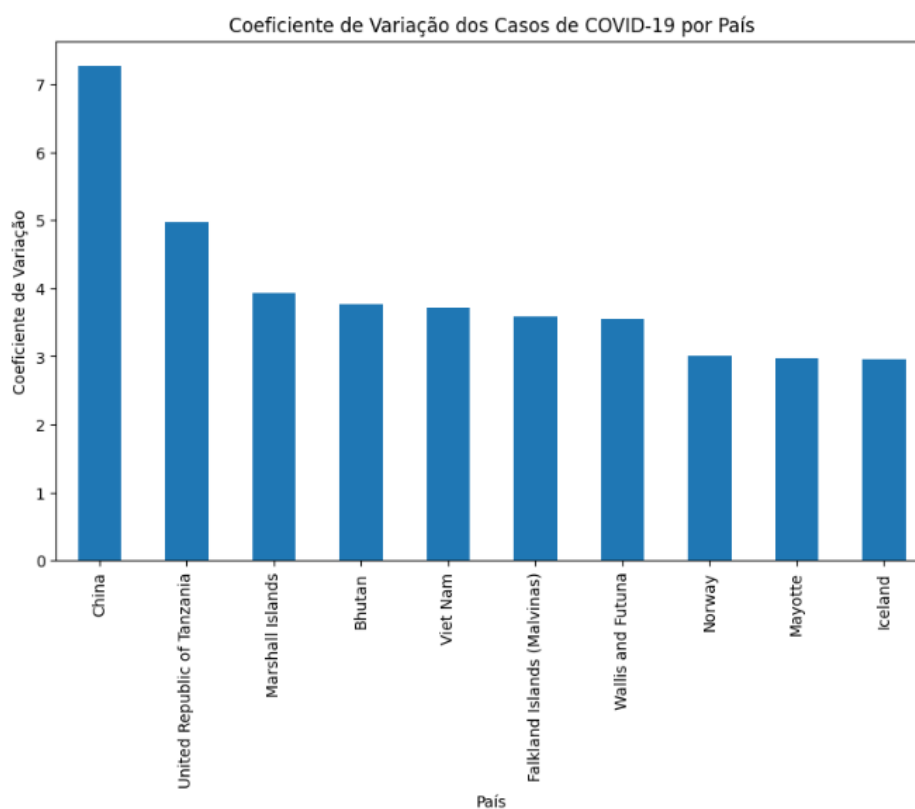
A partir da análise exploratória dos dados, foi possível fazer uma avaliação abrangente do impacto da pandemia em diferentes países, por exemplo, ao analisar a média de novos casos de COVID-19 por país, foi observada a dinâmica da transmissão da doença, identificando áreas de maior e menor impacto.



Para a análise da identificação de tendências temporais, foi utilizado uma média móvel com uma janela de 7 dias para suavizar os dados de casos cumulativos de COVID-19, a fim de destacar tendências de longo prazo e reduzir o ruído nos dados; isto permitiu visualizar padrões mais claros e identificar tendências significativas na evolução da pandemia ao longo do tempo.



O coeficiente de variação foi utilizado para fazer uma análise das disparidades e desigualdades na distribuição de casos de COVID-19, permitindo identificar desigualdades substanciais na distribuição de casos de COVID-19 entre diferentes regiões e grupos populacionais.



4 GITHUB

Acesso ao repositório do projeto:

<https://github.com/GrupoProjetoAplicado/Projeto/tree/main>

CRONOGRAMA

MEMBROS

Erick da Silva Isidoro

Gabriel Oliveira

Giovanna Sobral da Silva

Kelly Graziely dos Santos Pena