**Integrantes:**

Dawter Breno Hilario Caetano de Castro – 819155125

Giovanna Butini Pompeu – 8222243827

Gustavo Henrique Lucchi Pires – 819227510

Henrique de Carvalho Santos – 819230798

Matheus Lobo dos Santos – 82120742

1. **Contexto e Objetivo**

A COVID-19 é uma doença infecciosa causada pelo Coronavírus. É transmitida principalmente por meio de gotículas de pessoas infectadas. A gravidade dos sintomas varia muito de pessoa para pessoa. Até o presente momento, observa-se que cerca de 80% dos casos confirmados são assintomáticos e rápidos. A maioria das pessoas que se encaixam nesse grupo, se recupera sem nenhuma sequela. No entanto, 20% das pessoas terão infecções graves e precisarão de oxigênio ou auxílio médico.

Em 31 de dezembro de 2019, a Organização Mundial da Saúde (OMS) foi alertada sobre vários casos de pneumonia na cidade de Wuhan na China. Tratava-se de um novo tipo de Coronavírus que não havia sido identificado antes em seres humanos.

Uma semana depois, em 7 de janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de Coronavírus. Logo o novo SARS-CoV2 se espalhou rapidamente pelo mundo. Em 30 de janeiro de 2020, a OMS declarou que o surto do novo Coronavírus constituía em uma Emergência de Saúde Pública de importância internacional. Essa decisão buscou aprimorar a coordenação, a cooperação e a solidariedade global para interromper a propagação do vírus.

Atualmente existem quatro vacinas contra a doença que já receberam autorização da Agência Nacional de Vigilância Sanitária (Anvisa) para uso no Brasil: CoronaVac, AstraZeneca, Pfizer e Janssen. Além disso, em maio de 2023, a OMS decretou o fim da emergência pandêmica.

Neste estudo, o grupo irá analisar os dados para entender os padrões relacionados à doença. O foco será responder as seguintes perguntas:

Quais os principais fatores que implicaram para a rápida propagação da doença?

Existe algum padrão entre pacientes com a doença agravada? Ou ainda, existe algum padrão entre pacientes que chegaram à óbito?

Tendo em vista que a vacina não é obrigatória, qual o perfil do público vacinado? Existe algum padrão? Se sim, isso fica ainda mais visível após a primeira dose?

1. **Fonte de Dados**

Utilizaremos conjuntos de dados históricos do ano de 2020 até 2022 sobre os casos de Covid-19 nos estados brasileiros. Os conjuntos de dados são encontrados no OpenDataSUS, site de dados do Ministério da Saúde e Brasil.IO, site de repositórios de dados públicos.

<https://opendatasus.saude.gov.br/organization/ministerio-da-saude>

https://brasil.io/dataset/covid19/files/

1. **Método de Análise**

Através de modelagens estatísticas, vamos realizar uma análise exploratória para entender o motivo das questões citadas na seção 1. O objetivo é identificar padrões e descobrir a relação entre os dados.

1. **Ferramentas**

Para chegarmos ao propósito final, iremos utilizar a linguagem Python para leitura, tratamento e análise exploratória dos dados na plataforma Jupyter Notebook.

1. **Metodologia**

O processo do trabalho será realizado perante as seguintes etapas:

1. **Preparação do ambiente**

Antes de carregar os dados brutos para utilização, iremos preparar o ambiente importando as bibliotecas necessárias para auxiliar na construção do projeto, como por exemplo, a biblioteca Pandas para a análise de dados e a biblioteca Matplotlib para visualização dos dados.

1. **Carregamento/Leitura dos Dados**

A seguir, iremos realizar o carregamento dos datasets (formato .csv) para começar a transformação.

1. **Transformação/Limpeza dos Dados**

Para conseguirmos utilizar na análise os dados extraídos, iremos realizar uma limpeza/transformação nos dados como:

- Remoções de colunas desnecessárias para o objetivo final.

- Transformação dos dados ausentes para o melhor formato em relação à análise.

- Normalização das colunas, como por exemplo, em dois conjuntos temos colunas referentes ao sexo dos indivíduos, iremos padronizar para estarem descritos da mesma forma.

- Ajustar o tipo (string, integer, etc) de alguns campos.

1. **Análise dos Dados**

Após a limpeza dos dados, a etapa será utilizar modelagens estatísticas para respondermos as questões da seção 1.

1. **Visualização dos Dados**

Com a etapa 4 realizada, o objetivo será estruturar as análises e montar visualizações para a apresentação final com os insights obtidos ao longo do estudo.

1. **Dicionário de Dados**

O dicionário de dados está especificado no documento Excel anexado juntamente. Além das colunas citadas, iremos utilizar algumas bases, encontradas no site ‘Brasil.IO’, sumarizadas e agregadas com o objetivo de contribuição para nossa análise.