# **Descrição do Projeto e das variáveis**

## **Projeto**

Este projeto visa analisar todos os bancos de dados disponíveis no site da Transparência do Registro Civil, focando em informações relacionadas à COVID-19. O objetivo é explorar e entender as tendências e padrões dentro dos dados, realizando análises descritivas e inferenciais para fornecer insights significativos sobre o impacto da pandemia. Traremos gráficos para mostrar e relacionar os atributos dos dados.

Além disso, é importante ressaltar que o foco desse projeto será para mortes que tiveram como causador o COVID, tendo em vista que nesse banco de dados há mais causas de morte além da COVID, como pneumonia.

Vale ressaltar que alguns valores serão aproximados, usando as casas decimais como referência, onde aqueles que tiverem a primeira casa sendo maior ou igual a 5, o valor do número da unidade será arredondado para 1 valor acima.Essa aproximação será feita apenas a título de explicação, todos os cálculos foram feitos com os valores reais para maior precisão.

## **Variáveis**

Os bancos de dados contém as mesmas variáveis relacionadas a registros de óbitos. Abaixo, descrevemos as variáveis encontradas, incluindo o seu tipo, a presença de valores nulos e a descrição dos atributos.

### **UF:**

* Tipo: String
* Distribuição: Todas as UF do Brasil.

### **Tipo de doença:**

* Tipo: String
* Distribuição: Inclui causas relacionadas à COVID-19 e outras causas diversas.

### **Local do óbito:**

* Tipo: String
* Distribuição: Locais onde ocorreram os óbitos.

### **Faixa etária:**

* Tipo: String
* Distribuição: Intervalos de idade, variando de <9 até >100.

### **Sexo:**

* Tipo: String
* Distribuição: Aproximadamente equilibrando entre masculino e feminino.

### **Total:**

* Tipo: Numérico (inteiro)
* Distribuição: O total de óbitos relativos às variáveis anteriores.

## **Valores nulos**

Com base no retorno de comandos executados para todos os bancos, é possível verificar que há valores nulos em alguns deles, e sempre em uma mesma variavel, tendo a incidencia de valores nulos na variavel sexo, sendo 3 em 2022, 3 em 2021, 1 em 2020 e 6 em 2019.

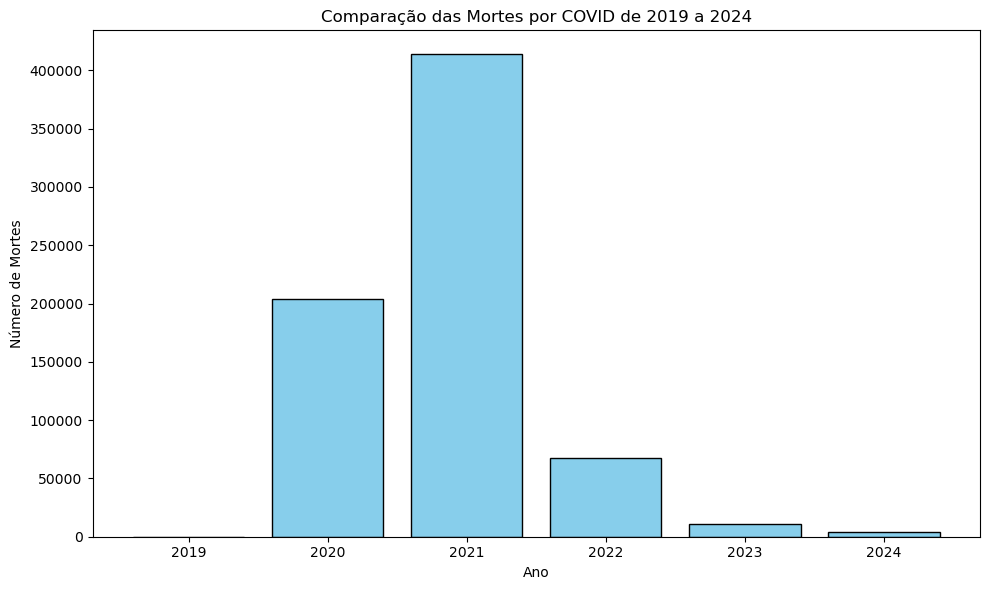
# **Estatísticas Descritivas**

Usando média e mediana, conseguimos algumas informações importantes, como por exemplo, a relação de mortes por COVID em relação aos estados do Brasil:

## **Média de mortes por COVID**

Calculando a soma das mortes por COVID de todos os anos, temos o valor de 699271, que dividido pelo número de estados (27), chegamos à conclusão de que houve em média, aproximadamente 25899 mortes por estado no Brasil.

Já calculando a média por anos, tivemos 116545 mortos por COVID por ano.

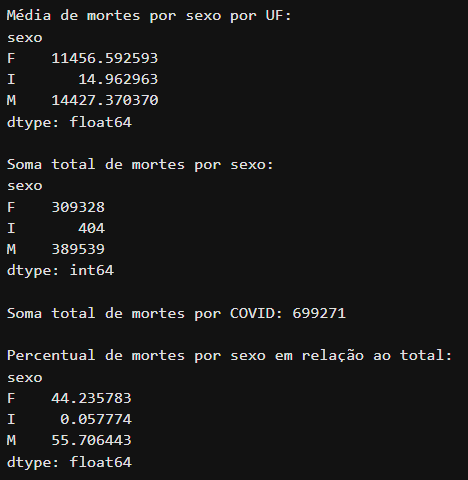


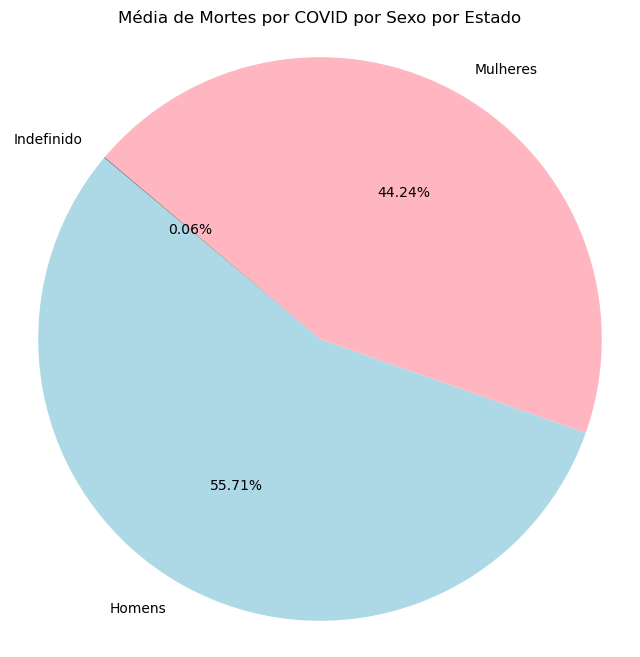
## **Sexo**

Ainda levando em consideração as mortes por UF, foi possível fazer um levantamento de quantos homens e quantas mulheres morreram em cada estado em média. Faz-se necessário ressaltar que devido a presença de 13 valores nulos, esses valores são desconsiderados para essa análise, uma vez que sua baixa incidência não envolve a análise como um todo.

Em média, morreram aproximadamente 14427 homens por Unidade Federativa no país, o que representa cerca 56% da média total de 25899 pessoas. Já em relação às mulheres, pode ser constatado que morreram de COVID, em média, 11457 pessoas por Unidade Federativa, o que equivale a 44%.

Quanto a pessoas de sexo indefinido, foi percebido que há uma quantidade de 0.06%. O que representa um valor pequeno em relação ao total.

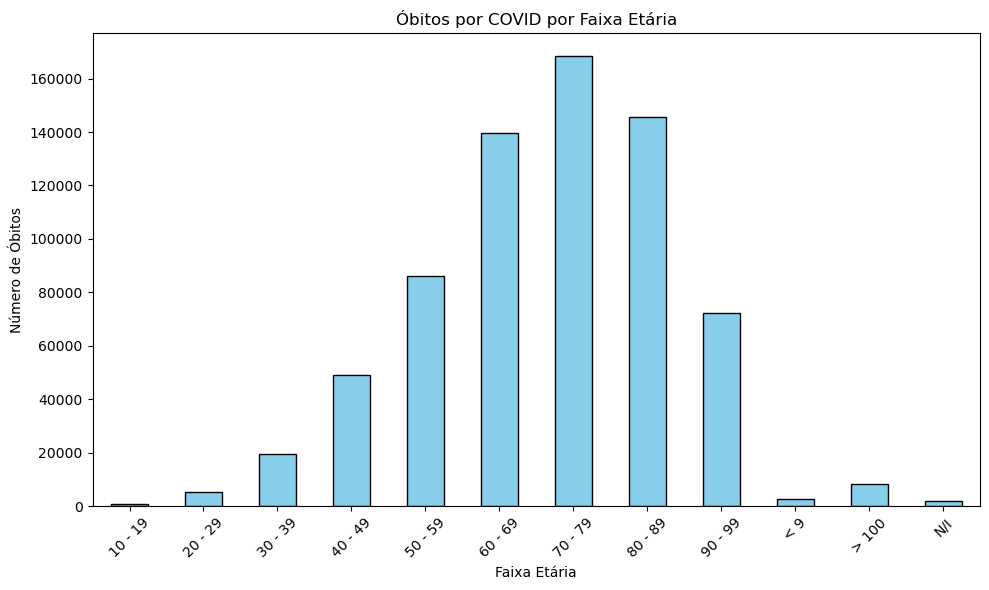




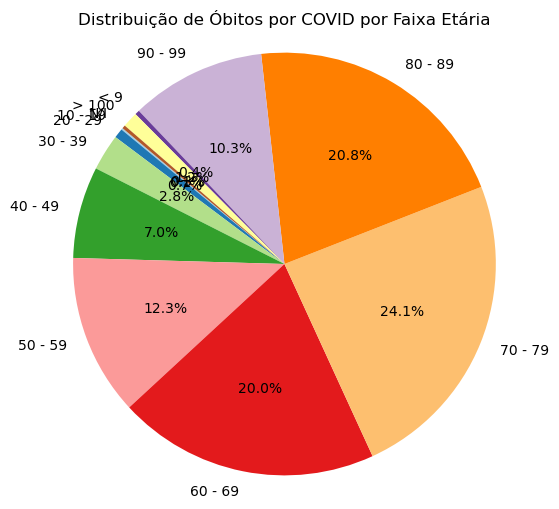
Com o grafico acima, podemos visualisar melhor os dados citados anteriormente, alem de perceber melhor quão pouca é a quantidade de sexo indefinido presente no banco de dados, uma vez que mal da pra visualisar no grafico. Além disso, é perceptível que, devido a média de mortes por homens por UF ser de 14427, enquanto a de mulheres é de 11457, há uma diferença no gráfico que pende mais para o lado dos homens.

## **Faixa Etária**

Como não temos a idade exata dos óbitos, analisamos a quantidade de óbitos por faixa etária. Ao analisar, foi possível concluir que a maior parte dos óbitos, aconteceu na faixa etária de 70 - 79 anos de idade. Essa faixa representa 24.10% do total de óbitos no Brasil. A seguir, há gráficos para ilustrar melhor esses dados.



Com esse gráfico, é colocado em evidência a quantidade de mortes para cada faixa etária, sendo entre 80 e 89 a que possui a maior quantidade de óbitos.

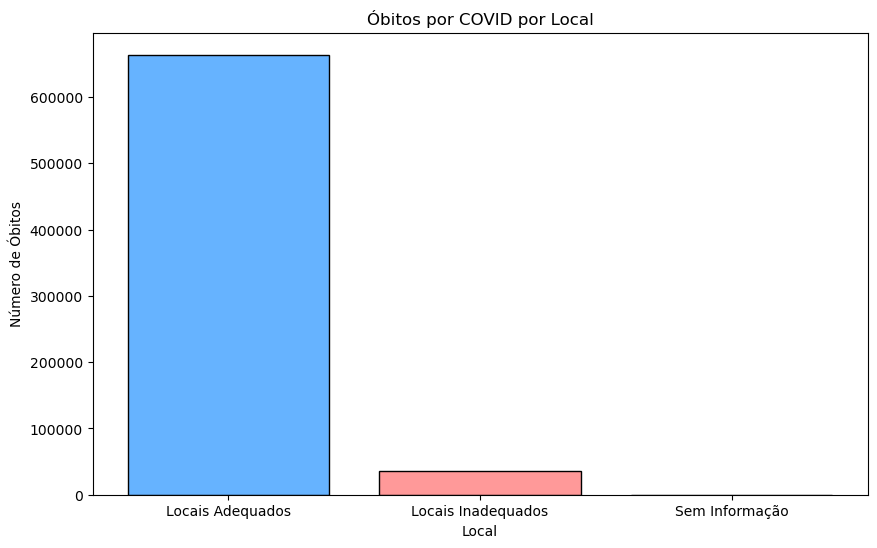


Por fim, com esse gráfico, pode-se ver a distribuição desses óbitos, mas tendo a porcentagem como referência.

## **Locais dos óbitos**

Para essa análise, é levado em consideração que hospitais, postos de saúde e outros serviços de saúde são tidos como locais adequados para essas mortes, enquanto qualquer outro lugar que não se encaixe nesses parâmetros são tidos como inadequados.

Utilizando de algumas linhas de comandos, percebesse que da quantidade total de 699271 óbitos, 663968 ocorreram em locais adequados, o que representa aproximadamente 94.95% da quantidade total. Sendo assim, outros 35303 óbitos ocorreram em locais inadequados, representando cerca de 5.05% do total, e outros 0 não há informações sobre o local dos óbitos, sendo estes 0% dos óbitos.

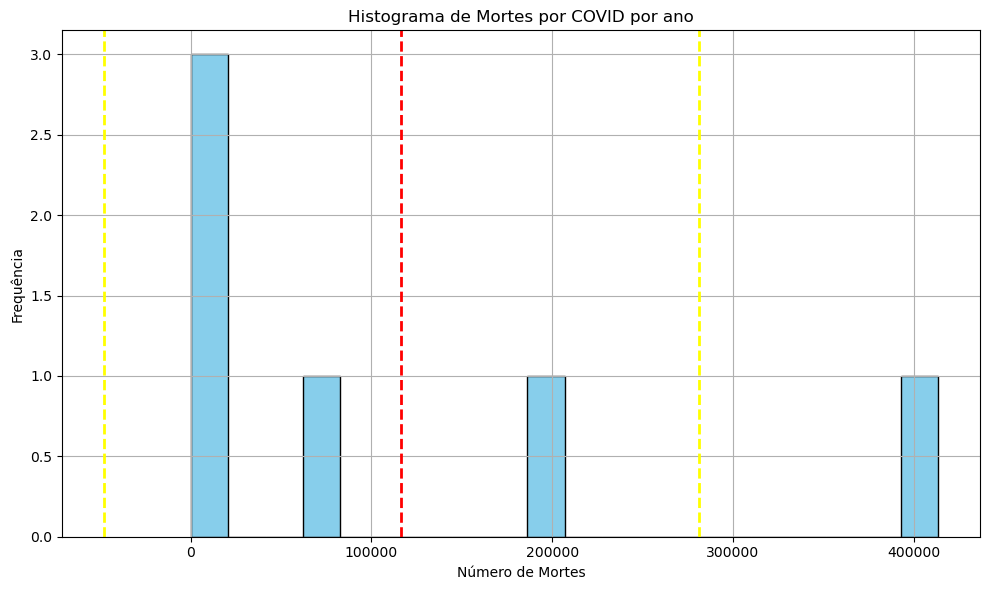


Por fim, com esse gráfico, percebe-se que houve mais mortes em locais adequados, do que em locais inadequados, e não houve nenhuma morte em locais desconhecidos.

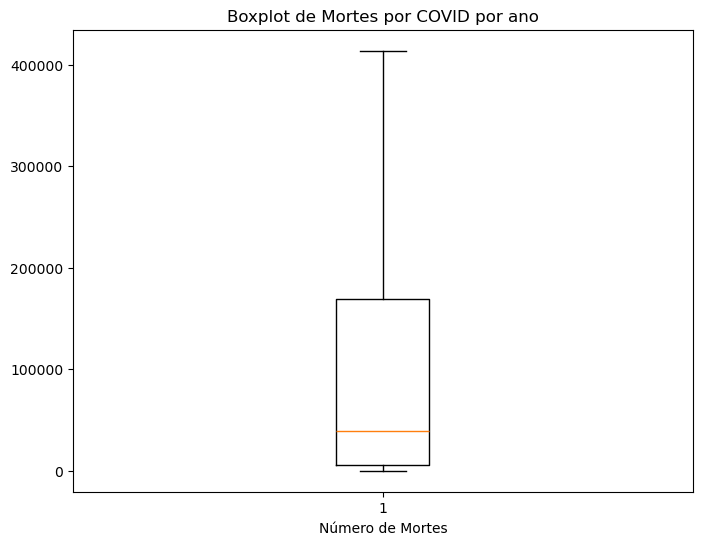
## **Desvio Padrão e Coeficiente de Variação (Mortes por COVID por uf)**

Calculando o desvio padrão do número de mortes por estado, que resultou em aproximadamente 164762.47. A partir disso, podemos calcular também o CV (coeficiente de variação), que resultou em 141.37%, indicando que podem existir outliers, uma vez que foi uma alta taxa.

## **Identificação de outliers (mortes por estado)**



Neste gráfico temos algumas informações que facilitam a identificação de outliers. A linha vermelha no centro mostra a média de óbitos por estado, enquanto as linhas amarelas indicam o desvio padrão, o que facilita a observação de outliers. Podemos ver em nosso histograma que embora tenham anos com números de mortes muito elevados, e que se distanciam muito da média, a maior parte dos anos não teve uma contagem de óbitos tão alta, mostrando assim que existe um pico evidente de morte entre os anos analisados.



Por fim, esse é um boxplot para ilustrar melhor a presença de outliers, por meio da representação de pontos além dos limites dos indicados, sendo assim, revelando que não há outliers, embora haja uma alteração alta no coeficiente de variação.

# **Conclusão**

Com todos os dados mostrados até agora, pode-se perceber o impacto dos óbitos no Brasil. Além disso, podemos chegar às seguintes conclusões.

## **Mortes por COVID**

A partir dos dados e da nossa análise, podemos concluir que quanto mais populoso o estado, mais óbitos há neles. Apesar da média por estado ser de 25899, sendo um valor já alto, há uma variação muito grande, o que não torna a média um valor tão aproximado para todos os estados. Um exemplo disso é quando se observa o Roraima, com 1975 óbitos, e São Paulo, com 190130, que são, respectivamente, o estado com menos e mais óbitos, que além de muito distante entre eles, são muito distantes da média.

Além disso, percebe-se também que do total de 699271 óbitos por COVID, tiveram uma média de 116545 por ano

## **Faixa etária**

A partir da análise da faixa etária, é perceptível que os óbitos por faixa etária, indicam que idades mais próximas a 70 tem uma maior incidência de óbitos, o que é algo natural, tendo um pico na faixa entre 70 e 79. Além disso, percebe-se que os idosos são a maioria, sendo que somente a faixa etária que teve o maior número de óbitos representou 24.10% do total de óbitos, e a porcentagem de pessoas idosas ( na faixa de 60 anos ou mais) é de 57.04%, sendo assim há uma distribuição mais tendenciosa para a parte dos idosos, uma vez que eles representam mais da metade do quantidade total de óbitos.

## **Locais adequados para óbitos**

Sobre isso pode-se concluir que a maioria dos óbitos ocorreram em locais que foram julgados adequados foi de 94.95%, e outros 5.05% são de locais inadequados para lidar com esses óbitos, e não há nenhum óbito em local não informado. Ao perceber isso, podemos tirar como conclusão que a maior parte da população está consciente e procurando os devidos locais para serem tratados de suas enfermidades, uma vez que os casos mais graves que resultam em morte tem uma concentração elevada para locais adequados.

## **Outliers e variação**

A análise de outliers mostra-se importante uma vez que é por ela que vemos a quantidade de variação do gráfico e o quão assertivas são as médias e medianas calculadas. Como por exemplo a análise de outliers para a quantidade de óbitos por ano, uma vez que foi evidenciado por meio desta que embora haja uma grande variação entre esses estados, por exemplo o estado com maior incidência de óbitos possui 96 vezes mais óbitos que o estado com menos óbitos, além da diferença de uma vezes e meia para a média de óbitos por ano, não há a presença de outliers.