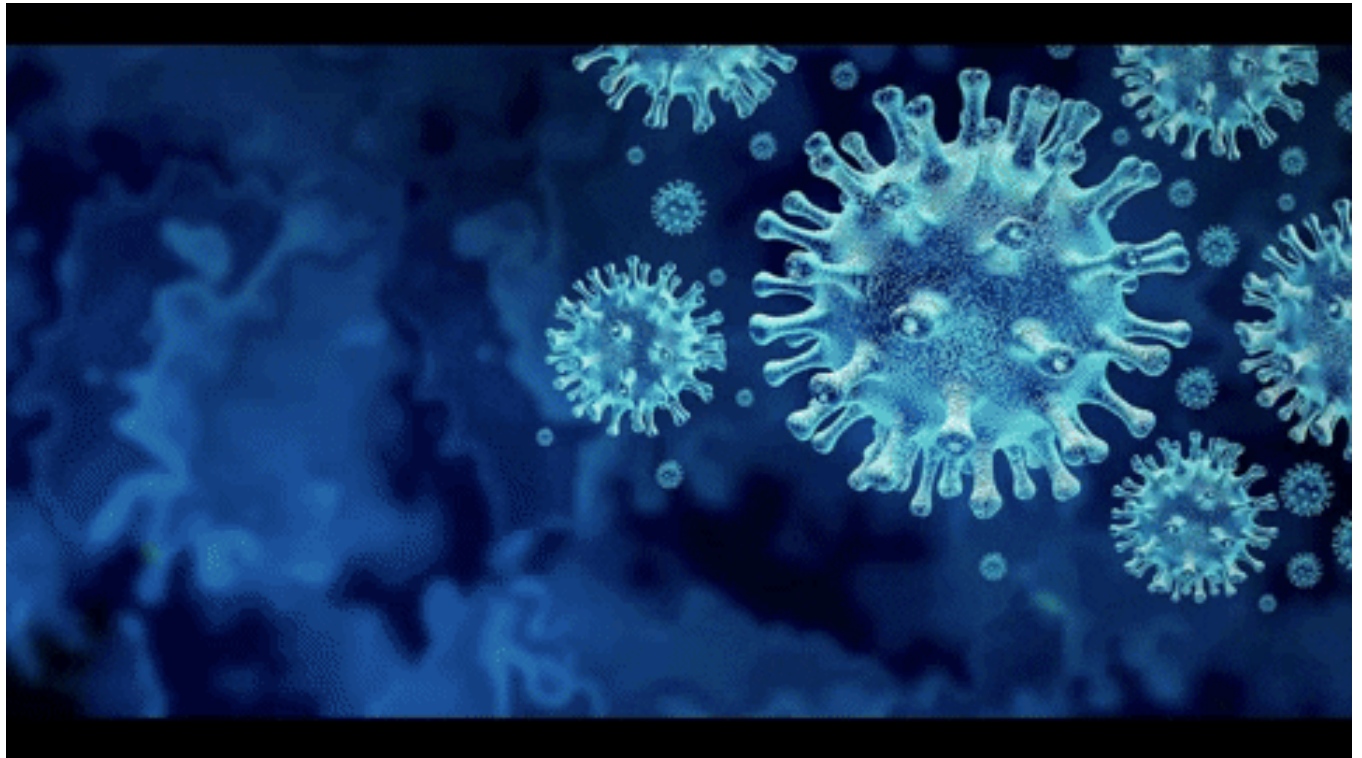


Análise de Dados do COVID-19 em Santa Catarina

 [gustavoquadra](#)

 gglquadra@gmail.com



Olá! Meu nome é Gustavo Luiz de Quadra e seja muito bem-vindo ao meu projeto referente ao **Módulo 02 - Visualização de Dados com Seaborn e Matplotlib** do Bootcamp de Data Science Aplicada da [Alura](#)!

Caso você tenha interesse em verificar o que foi aprendido nesse segundo módulo, podes clicar [aqui](#) para acessar as minhas anotações das aulas.

Sumário do Projeto

1. Introdução
2. Definição do Problema de Negócio
3. Preparação dos Dados
4. Análise Exploratória dos Dados
5. Criação de Hipóteses
6. Conclusão
7. Agradecimentos

1. Introdução

O **COVID-19** é a doença causada por uma nova espécie de coronavírus, denominada **SARS-CoV-2** e pertence a uma família de vírus que já circulava no Brasil antes da pandemia e era responsável por grande parte dos resfriados comuns. A origem do vírus ainda está sendo estudada, mas é muito provável que ela teve origem em **Wuhan, China** através de morcegos no ano de 2019.

O COVID-19 possui **sintomas leves** na maioria dos casos (tosse, dor de garganta, coriza), mas **pode ser extremamente agressiva** (falta de ar) à uma minoria. Costuma ser mais **perigoso para idosos com mais de 60 anos e pessoas com doenças pré-existentes**.

Para a prevenção do COVID-19, é importante mantermos as **mãos higienizadas com álcool em gel, utilizarmos máscaras e evitarmos aglomerações**. E também, aliada as medidas de prevenção, felizmente, hoje (04/06/2021), **estamos em campanha de vacinação no mundo inteiro para vencermos essa pandemia!**

2. Definição do Problema de Negócio

2.1 Objetivo do Projeto

O objetivo principal do projeto será **realizar uma análise de dados** referente à diversas informações (idade, município, sexo, sintomas...) **de pacientes que foram confirmados com COVID-19 no estado de Santa Catarina no Brasil**, buscando responder questionamentos e formular hipóteses para os nossos dados.

2.2 Dados Utilizados

A tabela de dados de pacientes confirmados com COVID-19 foi extraída do [Portal de Dados Abertos do Estado de Santa Catarina](#) em **04/06/2021** e sofreram um processo de manipulação para utilizarmos as seguintes informações :

Dicionário dos Dados :

- recuperados : Indicação de que o paciente foi recuperado
- data_inicio_sintomas : Data do início dos sintomas
- sintomas : Sintomas do paciente
- sexo : Sexo biológico do paciente
- municipio : Município de residência do paciente
- obito : Indicação de que o paciente veio a óbito
- idade : Idade do Paciente

OBS : Devido a limitação de armazenamento (25MB) de arquivos no GitHub, realizei a separação da tabela de dados em 5 arquivos menores que serão unidos na preparação dos dados.

3. Preparação dos Dados

Para essa primeira etapa, iremos realizar a importação dos pacotes e dos dados que serão utilizados em nossa análise.

Os dados estão armazenados no meu repositório do GitHub e podem ser acessados clicando [aqui](#).

3.1 Importação dos Pacotes

Para a análise, iremos realizar a importação dos seguintes pacotes :

- **pandas(pd) 1.1.5** : Pacote para carregamento e manipulação dos dados.
- **seaborn(sns) 0.11.1** : Pacote para visualização e criação de gráficos.
- **matplotlib.pyplot(plt) 3.2.2** : Pacote para visualização e criação de gráficos.

3.2 Importação dos Dados

Os dados foram separados em 5 arquivos na minha máquina utilizando o comando `split -l 195665 Todos-Confirmados-SC.csv` (cada um com 195.665 linhas), com os seguintes nomes :

- **Todos-Confirmados-SC-1.csv** (possui os nomes das colunas)
- **Todos-Confirmados-SC-2.csv**
- **Todos-Confirmados-SC-3.csv**
- **Todos-Confirmados-SC-4.csv**
- **Todos-Confirmados-SC-5.csv**

Nesse primeiro momento, iremos realizar o carregamento desses dados e unir os mesmos em apenas uma tabela.

Após realizarmos a união das nossas tabelas de dados, podemos verificar que a tabela final `dados_SC` possui **978.324 linhas** e **7 colunas**.

Abaixo, também podemos verificar as 5 primeiras linhas de nossa tabela.

A tabela de dados possui 978324 linhas e 7 colunas.

Out[4]:

	recuperados	data_inicio_sintomas	sintomas	sexo	municipio	obito	idade
0	Sim	2020-11-21	Nao_Informado	Feminino	Florianopolis	Nao	66.0
1	Sim	2021-02-22	TOSSE, DOR DE GARGANTA, FEBRE	Feminino	Sao Jose	Nao	29.0
2	Sim	2021-03-10	TOSSE, DISPNEIA, DOR DE GARGANTA, FEBRE	Masculino	Joinville	Nao	24.0
3	Sim	2021-01-03	Nao_Informado	Masculino	Lages	Nao	37.0
4	Sim	2020-12-09	Nao_Informado	Feminino	Sao Jose	Nao	57.0

3.3 Criação de Colunas

Para auxiliar em nossas análises, iremos realizar a criação de 2 novas colunas `mes_inicio_sintomas` e `ano_inicio_sintomas` referentes ao **mês** e **ano** dos primeiros sintomas.

Outra alteração que realizaremos será a divisão das idades dos pacientes que foram confirmados com covid em grupos de idade, excluindo assim a coluna `idade` para trabalharmos apenas com os grupos criados, sendo eles :

- Menor que 20
- 21-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71-80
- Maior que 81

Após a realização dessa nova adição, podemos demonstrar novamente como a nossa tabela se encontra após as alterações.

Podemos verificar novamente as 5 primeiras linhas para validarmos se a inserção das novas colunas `mes_inicio_sintomas`, `ano_inicio_sintomas` e `idade_categoria` ocorreram corretamente.

Out[7]:

	recuperados	data_inicio_sintomas	mes_inicio_sintomas	ano_inicio_sintomas	sintomas	sexo	muni
0	Sim	2020-11-21	11	2020	Nao_Informado	Feminino	Florianc

	recuperados	data_inicio_sintomas	mes_inicio_sintomas	ano_inicio_sintomas	sintomas	sexo	muni
1	Sim	2021-02-22	2	2021	TOSSE, DOR DE GARGANTA, FEBRE	Feminino	Sac
2	Sim	2021-03-10	3	2021	TOSSE, DISPNEIA, DOR DE GARGANTA, FEBRE	Masculino	Joi
3	Sim	2021-01-03	1	2021	Nao_Informado	Masculino	I
4	Sim	2020-12-09	12	2020	Nao_Informado	Feminino	Sac

3.4 Criação de Tabelas

Também iremos realizar criações de uma tabela referente à quantidade de cada sintoma para os que foram divulgados (muitos possuem valores **Nao_Informado** que realizarei à remoção para a criação dessa tabela)

Para a criação dessa tabela, iremos realizar a divisão dos dados em sintomas de pacientes que não foram à óbitos e sintomas de pacientes que foram à óbitos.

Out[11]:

	Quantidade_Sem_Obito	Sintoma_Sem_Obito_Por_Pessoa	Quantidade_Com_Obito	Sintoma_Com_Obito_P
Tosse	434560	65.54	10531	
Febre	315201	47.54	8295	
Dor De Garganta	309571	46.69	3440	
Dispneia	125342	18.90	12916	
Mialgia	121707	18.35	1167	
Cefaleia	89884	13.56	627	
Diarreia	62078	9.36	1986	
Dor No Corpo	53060	8.00	452	
Coriza	47788	7.21	342	
Cansaco	22229	3.35	320	
Congestao Nasal	8632	1.30	26	

4. Análise Exploratória dos Dados

Finalizada a etapa de importação dos dados e a criação de novas tabelas ou colunas, podemos partir para a análise exploratória de nossas tabelas!

A **Análise Exploratória** é uma etapa onde buscamos **entender a disposição e a características dos dados** de cada variável (coluna) de nossa tabela de dados.

Conforme podemos analisar abaixo, possuímos diferentes tipos de dados para as colunas, sendo eles :

- object : Valores reconhecidos como texto

- datetime : Valores reconhecidos como data
- int64 e float64 : Valores numérico inteiros e decimais
- category : Valores categóricos

```
Out[12]: recuperados          object
data_inicio_sintomas      datetime64[ns]
mes_inicio_sintomas        int64
ano_inicio_sintomas        int64
sintomas                   object
sexo                       object
municipio                  object
obito                      object
idade_categoria            category
dtype: object
```

4.1 Análise Exploratória Variáveis Categóricas

Para essa primeira parte, iremos realizar a verificação das variáveis que podem assumir um range fixo de valores (por exemplo, sim ou não, masculino ou feminino).

As variáveis que se enquadram nessa categoria são : recuperados , sexo , municipio , obito e idade_categoria .

Iremos realizar primeiro uma análise em quais são os valores e a quantidade de valores únicos que cada variável categórica pode assumir em nossa tabela.

Podemos verificar que as variáveis recuperados , sexo , obito e idade_categoria possuem valores muito bem definidos, ao contrário de municipio que possui **296** valores diferentes.

```
Valores da coluna recuperados : ['Sim' 'Nao']
Quantidade de valores que a coluna recuperados pode assumir : 2
```

```
Valores da coluna sexo : ['Feminino' 'Masculino' 'NAO INFORMADO']
Quantidade de valores que a coluna sexo pode assumir : 3
```

```
Valores da coluna municipio : ['Florianopolis' 'Sao Jose' 'Joinville' 'Lages' 'Quilombo' 'Palho
ca'
'Forquilha' 'Braco Do Norte' 'Alfredo Wagner' 'Cacador' 'Vargem'
'Balneario Camboriu' 'Joacaba' 'Blumenau' 'Icara' 'Xanxere' 'Chapeco'
'Rio Das Antas' 'Navegantes' 'Tangara' 'Rio Dos Cedros' 'Fraiburgo'
'Campos Novos' 'Praia Grande' 'Ermo' 'Concordia' 'Itajai'
'Faxinal Dos Guedes' 'Jaragua Do Sul' 'Armazem' 'Videira' 'Irani'
'Corupa' 'Criciuma' 'Uniao Do Oeste' 'Itapema' 'Biguacu' 'Tubarao'
'Timbe Do Sul' 'Araquari' 'Camboriu' 'Curitibanos' 'Indaial'
'Trombudo Central' 'Itapoa' 'Laguna' 'Seara' 'Nova Trento' 'Mafra'
'Sao Bento Do Sul' 'Trevizo' 'Governador Celso Ramos' 'Ouro'
'Santo Amaro Da Imperatriz' 'Balneario Picarras' 'Capinzal' 'Porto Uniao'
'Ararangua' 'Pescaria Brava' 'Sao Francisco Do Sul' 'Morro Grande'
'Tres Barras' 'Bom Retiro' 'Rio Do Sul' "Herval D'Oeste" 'Monte Carlo'
'Itaipopolis' 'Brusque' 'Celso Ramos' 'Tijucas' 'Schroeder'
'Otacilio Costa' 'Ituporanga' 'Presidente Getulio' 'Jacinto Machado'
'Penha' 'Abelardo Luz' 'Turvo' 'Chapadao Do Lageado'
'Balneario Barra Do Sul' 'Ponte Alta Do Norte' 'Timbo' 'Anchieta'
'Guaramirim' 'Canoinhas' 'Sao Bonifacio' 'Irineopolis' 'Canelinha'
'Sao Joao Batista' 'Passo De Torres' 'Vargem Bonita' 'Barra Velha'
'Ponte Alta' 'Rio Negrinho' 'Papanduva' 'Urussanga' 'Garopaba'
'Pinhalzinho' 'Cocal Do Sul' 'Massaranduba' 'Pomerode' 'Botuvera'
'Correia Pinto' 'Sao Joaquim' 'Sao Joao Do Sul' 'Laurentino'
'Nova Veneza' 'Santa Rosa Do Sul' 'Paulo Lopes'
'Balneario Arroio Do Silva' 'Gravatal' 'Nova Erechim' 'Descanso'
'Santa Rosa De Lima' 'Lontras' 'Grafo Para' 'Orleans' 'Imbituba'
'Porto Belo' 'Vidal Ramos' 'Paraiso' 'Maravilha' 'Santa Cecilia'
'Sombrio' 'Treze Tilias' 'Sao Joao Do Itaperiu' 'Guabiruba' 'Meleiro'
'Antonio Carlos' 'Sao Martinho' 'Xaxim' 'Apiuna' 'Itapiranga' 'Caibi'
'Anita Garibaldi' 'Ouro Verde' 'Bom Jesus' 'Palmitos' 'Balneario Gaivota'
'Entre Rios' 'Angelina' 'Presidente Castello Branco' 'Rio Do Campo'
'Dona Emma' 'Taio' 'Rancho Queimado' 'Santa Terezinha Do Progresso'
'Capivari De Baixo' 'Sao Jose Do Cerrito' 'Ibirama' 'Imbuia' 'Xavantina'
'Rodeio' 'Gaspar' 'Morro Da Fumaca' 'Rio Do Oeste' 'Ipumirim' 'Jaguaruna'
```

'Sao Carlos' 'Tunapolis' 'Palma Sola' 'Sao Miguel Do Oeste'
'Coronel Freitas' 'Planalto Alegre' 'Campo Alegre' 'Guatambu'
'Cordilheira Alta' 'Caxambu Do Sul' 'Sao Ludgero' 'Irati'
'Dionisio Cerqueira' 'Santa Terezinha' 'Lajeado Grande' 'Maracaja'
'Garuva' 'Pedras Grandes' 'Lauro Muller' 'Monte Castelo' 'Treze De Maio'
'Ipuacu' 'Lebon Regis' 'Luiz Alves' 'Matos Costa' 'Bombinhas' 'Sangao'
'Jabora' 'Campo Ere' 'Iomere' 'Ervail Velho' 'Lindolia Do Sul' 'Saudades'
'Ibicare' 'Pouso Redondo' 'Catanduvas' 'Bom Jardim Da Serra'
'Sao Domingos' 'Benedito Novo' 'Macieira' 'Agua Doce' 'Vargeao' 'Galvao'
'Sao Bernardino' 'Jose Boiteux' 'Sao Lourenco Do Oeste' 'Mondai'
'Salto Veloso' 'Ita' 'Romelandia' 'Arabuta' 'Cerro Negro' 'Modelo'
'Jupia' 'Aurora' 'Iraceminha' 'Aguas De Chapeco' 'Ponte Serrada' 'Zortea'
'Ibiam' 'Riqueza' 'Sideropolis' 'Peritiba' 'Sao Joao Do Oeste' 'Luzerna'
'Sul Brasil' 'Ilhota' 'Belmonte' 'Arroio Trinta' 'Lacerdopolis'
'Balneario Rincão' 'Aguas Mornas' 'Agrolandia' 'Doutor Pedrinho'
'Campo Belo Do Sul' 'Brunopolis' 'Cunha Pora' 'Alto Bela Vista'
'Barra Bonita' 'Sao Miguel Da Boa Vista' 'Piratuba' 'Vitor Meireles'
'Princesa' 'Sao Jose Do Cedro' 'Tigrinhos' 'Rio Fortuna'
'Coronel Martins' 'Major Vieira' 'Guaraciaba' 'Pinheiro Preto'
'Timbo Grande' 'Aguas Frias' 'Ipora Do Oeste' 'Bom Jesus Do Oeste'
'Formosa Do Sul' 'Guarujá Do Sul' 'Passos Maia' 'Ascurra' 'Cunhatai'
'Paial' 'Anitapolis' 'Atalanta' 'Agronomica' 'Urubici' 'Ipira'
'Bandeirante' 'Painel' 'Mirim Doce' 'Saltinho' 'Arvoredo'
'Braco Do Trombudo' 'Petrolandia' 'Imarui' 'Nova Itaberaba'
'Sao Pedro De Alcantara' 'Marema' 'Salete' 'Santiago Do Sul'
'Leoberto Leal' 'Bela Vista Do Toldo' 'Jardinopolis' 'Witmarsum'
'Major Gercino' 'Frei Rogerio' 'Serra Alta' 'Novo Horizonte'
'Abdon Batista' 'Rio Rufino' 'Flor Do Sertao' 'Urupema' 'Capao Alto'
'Palmeira' 'Bocaina Do Sul' 'Santa Helena' 'Sao Cristovao Do Sul'
'Calmon' 'Presidente Nereu' 'Indaial ']

Quantidade de valores que a coluna municipio pode assumir : 296

Valores da coluna obito : ['Nao' 'Sim']

Quantidade de valores que a coluna obito pode assumir : 2

Valores da coluna idade_categoria : ['61-70', '21-30', '31-40', '51-60', '41-50', '71-80', 'Maior que 81', 'Menos que 20', NaN]

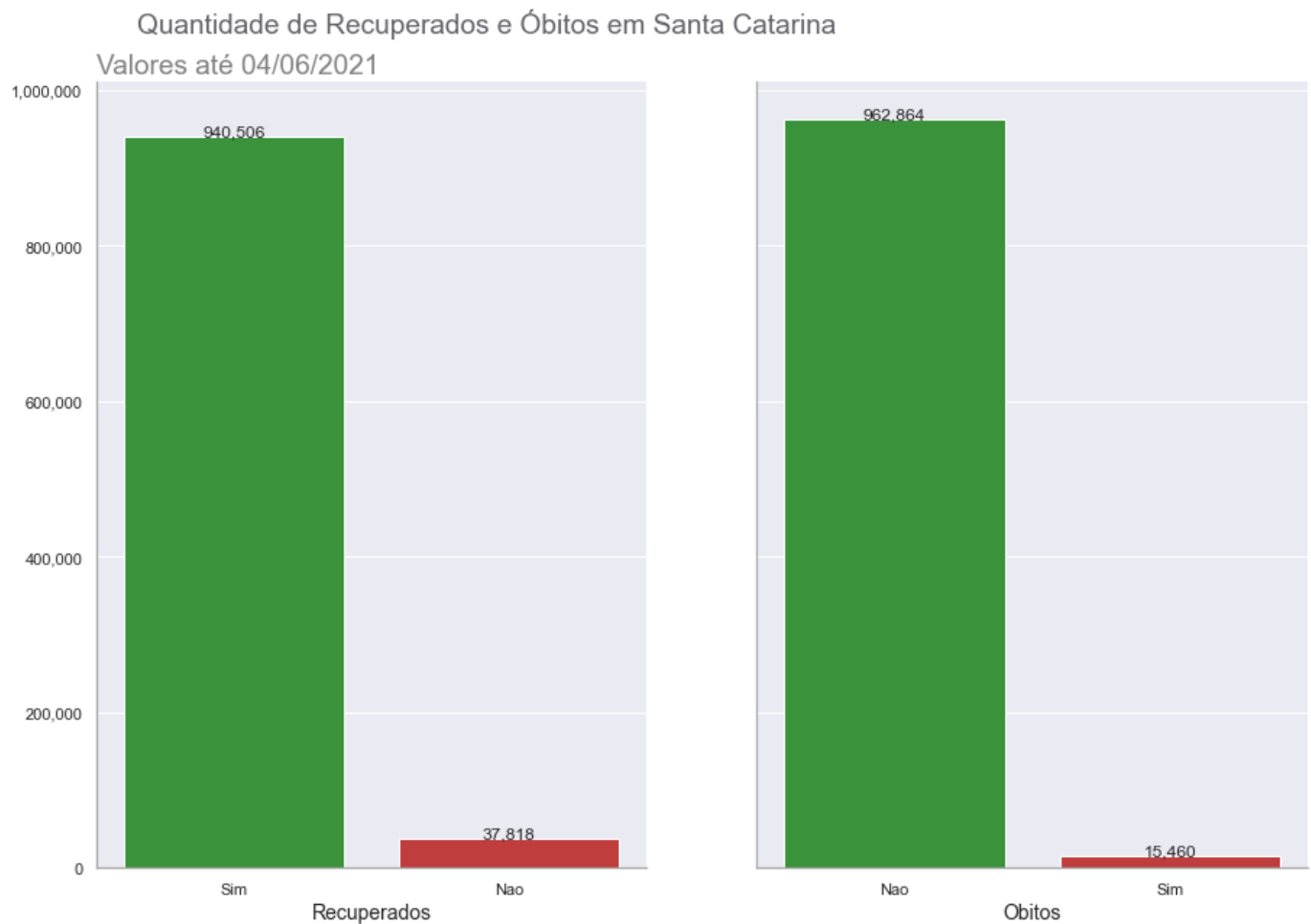
Categories (8, object): ['Menos que 20' < '21-30' < '31-40' < '41-50' < '51-60' < '61-70' < '71-80' < 'Maior que 81']

Quantidade de valores que a coluna idade_categoria pode assumir : 8

4.1.1 Recuperados e Óbitos

Primeiro, vamos analisar as colunas recuperados e obitos da nossa tabela, onde podemos verificar que **FELIZMENTE** possuímos uma disparidade entre os valores positivos (se recuperar ou não ir à óbito, em verde) com os valores negativos (não se recuperar ou ir à óbito, em vermelho).

Podemos identificar também que **22.358** (diferença entre não recuperados e pessoas que foram à óbito) pacientes até o dia **04/06/2021** podem estar lutando contra o COVID-19.



4.1.2 Sexo Biológico e Idades

Após verificarmos as variáveis de recuperados e obito, vamos verificar como estão distribuídas as variáveis sexo e idade_categoria.

Como principal objetivo em nossa análise, iremos focar nos seguintes pontos:

- **Demonstrar os valores absolutos e proporcionais de Total de Casos, Recuperados e Óbitos por categoria de idade.**
- **Demonstrar os valores absolutos de Total de Casos, Recuperados e Óbitos por sexo biológico.**
- **Demonstrar os valores absolutos de Total de Casos, Recuperados e Óbitos por sexo biológico e categoria de idade.**

Para auxiliar em nossa visualização, iremos realizar a criação de uma tabela que indica o tipo (Total / Recuperado / Obito), o valor absoluto e a proporção para cada categoria da nossa coluna Idade_Categoria.

Out[16]:

	Idade_Categoria	Tipo	Quantidade	Quantidade_Prop
0	Menos que 20	Total	106218	100.0
1	21-30	Total	202710	100.0
2	31-40	Total	230542	100.0
3	41-50	Total	177778	100.0
4	51-60	Total	135131	100.0

Demonstrar os valores absolutos e proporcionais de Total de Casos, Recuperados e Óbitos por categoria de idade.

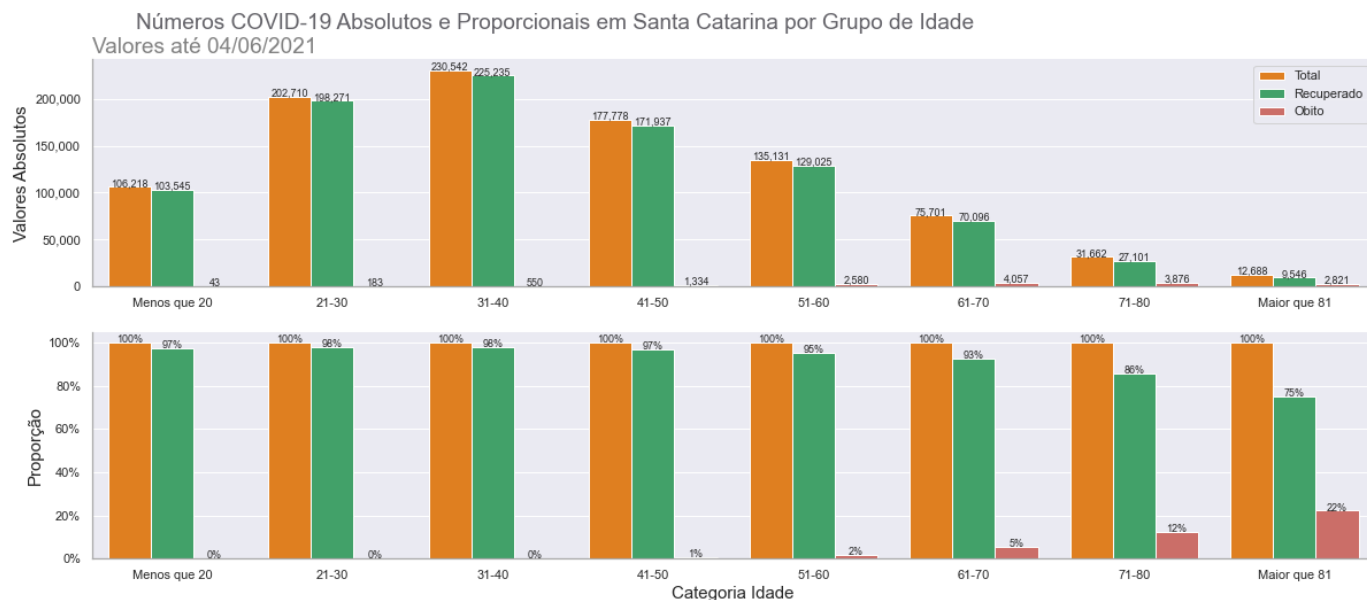
Após realizarmos a criação da tabela, podemos verificar com os gráficos abaixo, como estão distribuídas os números absolutos e as proporções por idade de:

- **Total** : Quantidade do grupo que testou positivo para COVID-19.
- **Recuperado** : Quantidade do grupo que se recuperou do COVID-19.
- **Óbito** : Quantidade do grupo que veio à óbito após COVID-19.

Analisando o gráfico, podemos verificar uma maior quantidade de testagem de positivos entre os grupos **"Menos que 20 até 51-60"** que acabam passando de **100.000 positivos**.

É interessante notarmos ao verificar o gráfico de proporções como os grupos de **"71-80" (12% dos que testaram positivo foram à óbito)** e **"Maior que 81" (22% dos que testaram positivo foram à óbito)** possuem uma taxa de óbito muito maior e consequentemente, menor de recuperados.

O que comprova o quão severo é o COVID-19 para os grupos de idosos.



Demonstrar os valores absolutos de Total de Casos, Recuperados e Óbitos por sexo biológico.

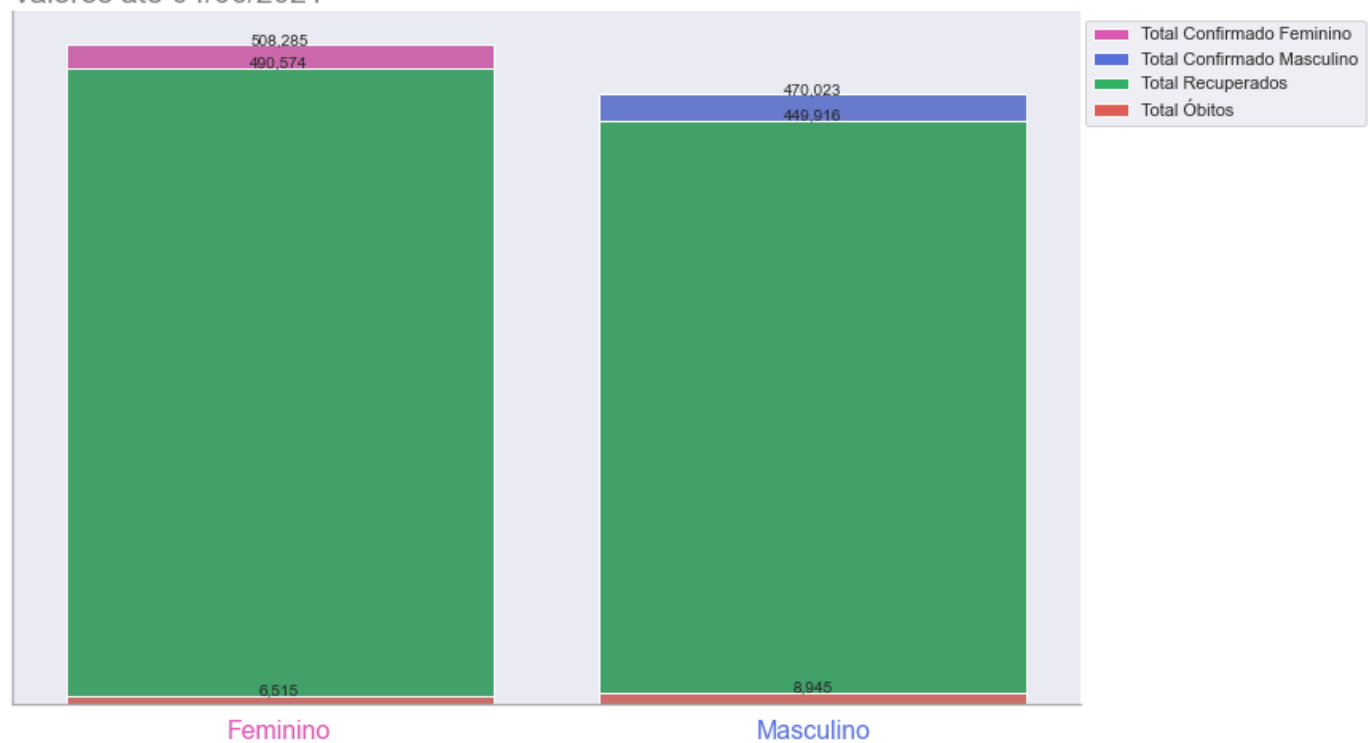
Após a verificação da variável `idade` vamos identificar como está distribuída a quantidade de valores absolutos em nossa variável `sexo`.

Como podemos verificar abaixo, possuímos **mais casos confirmados em mulheres (508.285 contra 470.023 em homens)** e **490.574 (96,51% do total de confirmados em mulheres) se recuperaram** do COVID-19 contra **449.916 homens recuperados (95,72% do total)**.

Mesmo com valores menores de casos confirmados para **homens**, podemos verificar no estado de Santa Catarina que é o **sexo que possui mais mortes em valores absolutos (8.945 contra 6.515 do sexo feminino)** e proporcional (**1,9% do total** contra, aproximadamente, **1,3% do total do sexo feminino**).

Números Covid-19 em Santa Catarina por Sexo Biológico

Valores até 04/06/2021



Demonstrar os valores absolutos de Total de Casos, Recuperados e Óbitos por sexo biológico e categoria de idade.

Para isso, irei realizar a criação de uma tabela que separa os valores totais, recuperados e óbitos por sexo e a categoria da idade.

Out[19]:

	Feminino_Total	Masculino_Total	Feminino_Recuperado	Masculino_Recuperado	Feminino_Obito
idade_categoria					
Menos que 20	53631	52585	52257	51286	22
21-30	106076	96628	103841	94424	93
31-40	119563	110976	117025	108207	215
41-50	92618	85159	89836	82100	523
51-60	70852	64277	68084	60939	1025
61-70	38585	37115	36272	33823	1566
71-80	16561	15100	14564	12536	1646
Maior que 81	7479	5209	5848	3698	1419

Após a criação dessa tabela, iremos realizar a divisão da tabela em outras 2 tabelas :

- **sexo_idade_feminino = Total, Recuperados e Óbitos feminino**
- **sexo_idade_masculino = Total, Recuperados e Óbitos masculino**

Out[21]:

	idade_categoria	Tipo	Quantidade
0	Menos que 20	Total	53631
1	21-30	Total	106076
2	31-40	Total	119563
3	41-50	Total	92618
4	51-60	Total	70852

Out[22]:

	idade_categoria	Tipo	Quantidade
--	-----------------	------	------------

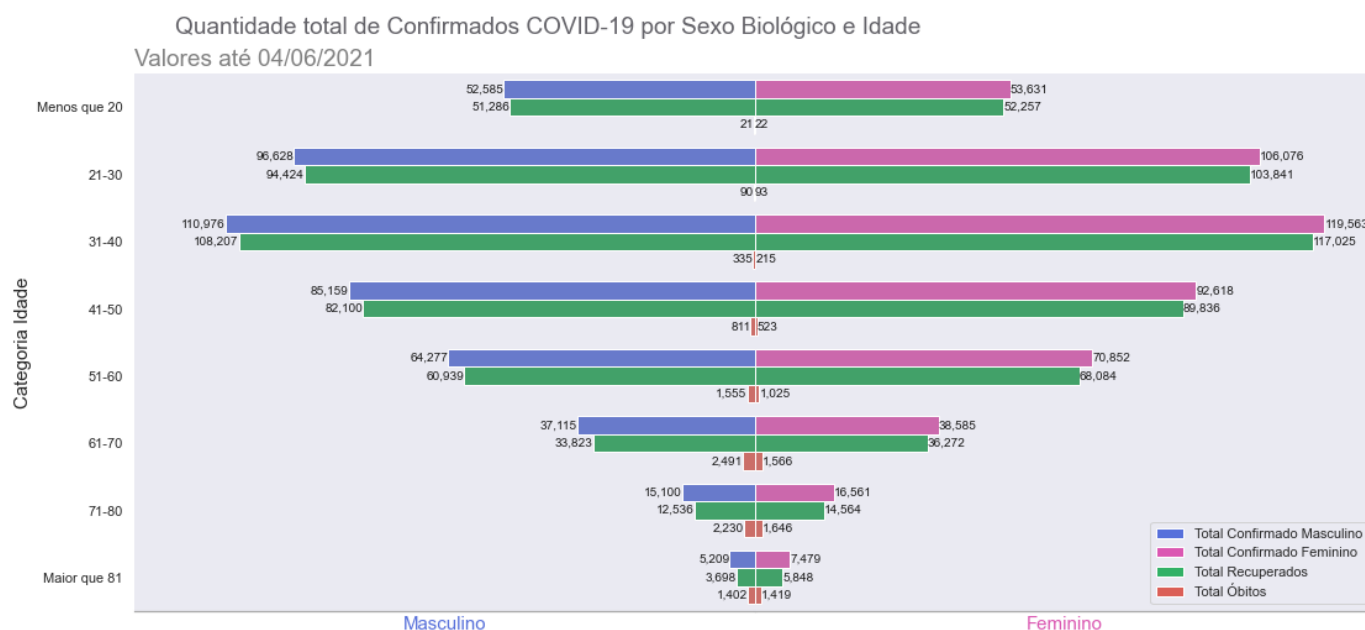
0	Menos que 20	Total	52585
1	21-30	Total	96628
2	31-40	Total	110976
3	41-50	Total	85159
4	51-60	Total	64277

Com base nessa tabela, conseguimos realizar a criação de um gráfico que demonstra como estão separados os valores absolutos de Totais Confirmados, Recuperados e Óbitos por Covid-19 em cada grupo de idade por sexo biológico.

É interessante notarmos como os valores de **confirmados e recuperados são muito altos** para grupos entre **"Menos que 20"** e **"41-50"**, visto que a defesa imunológica nessa idade é mais alta quando comparada com idades avançadas.

A partir de **51-60**, conseguimos notar tanto para Masculino e Feminino que começa a ocorrer **mais de 1.000 óbitos e menos confirmados e recuperados**, o que significa que a **taxa de mortalidade é muito maior para esses grupos**.

Analisando o gráfico, conseguimos notar a **maior taxa de mortalidade** para **homens acima de 81 anos**, com **5.209 confirmados e 1.402 óbitos (26% do total foram à óbito)**.



4.1.3 Municípios

E por fim, iremos realizar a análise da nossa variável "município" que corresponde ao município de Santa Catarina que teve o caso confirmado, para essa análise iremos focar nos seguintes pontos :

- **Demonstrar os valores absolutos do Total de Casos Confirmados, Recuperados e Óbitos nos municípios de Santa Catarina (ordenar por maior valor Total e por Óbitos).**
- **Demonstrar os valores proporcionais de Óbitos nos municípios de Santa Catarina (ordenar por maior valor proporcional de Óbitos e verificar a proporção dos municípios vistos anteriormente)**

OBS : Visto que as cidades **Flor do Sertão**, **Indaial** e **Novo Horizonte** não possuem óbitos nos registros, iremos realizar a **remoção** das mesmas.

Para auxiliar em nossa análise, realizei a criação de uma tabela com os valores de cada município com os valores de Total de Casos Confirmados, Recuperados e Óbitos, como podemos verificar abaixo na representação das 5 primeiras linhas :

Out[27]:

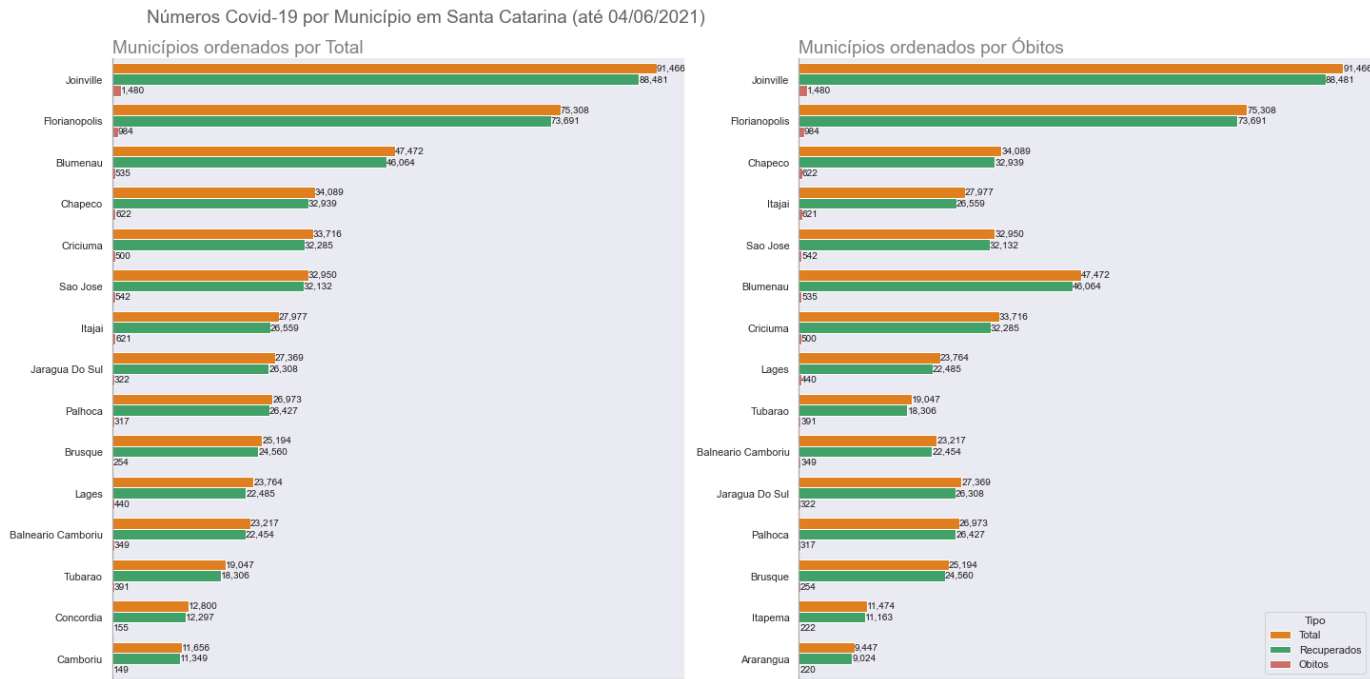
	Município	Tipo	Quantidade
0	Abdon Batista	Total	372
1	Abelardo Luz	Total	1389
2	Agrolândia	Total	1351
3	Agronômica	Total	595
4	Água Doce	Total	1156

Demonstrar os valores absolutos do Total de Casos Confirmados, Recuperados e Óbitos nos municípios de Santa Catarina (ordenar por maior valor Total e por Óbitos).

Conforme podemos verificar abaixo, possuímos as seguintes cidades no top 10 (ordenados por total e obitos) :

Ordenado total	Ordenado por óbito
Joinville	Joinville
Florianópolis	Florianópolis
Blumenau	Chapeco
Chapeco	Itajaí
Criciúma	Sao José
Sao Jose	Blumenau
Itajaí	Criciúma
Jaragua do Sul	Lages
Palhoca	Tubarão
Brusque	Balneário Camboriu

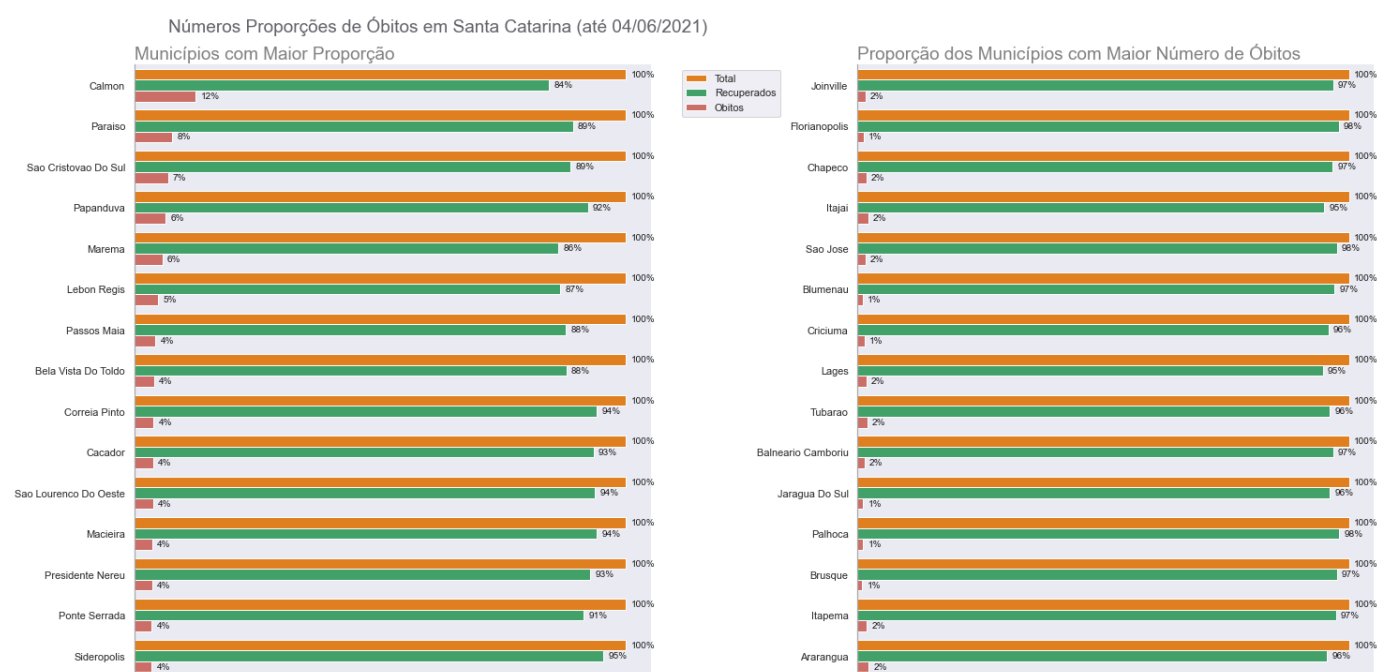
É importante notar que as cidades com maior casos de COVID-19 possuem a maior quantidade de óbitos em valores absolutos, alterando algumas posições na tabela.



Demonstrar os valores proporcionais de Óbitos nos municípios de Santa Catarina (ordenar por maior valor proporcional de Óbitos e verificar a proporção dos municípios vistos anteriormente)

Após a criação da nossa tabela, também iremos adicionar os valores proporcionais por município, onde os tipos Total recebem o valor de 100%.

É importante notarmos que as cidades que possuem o maior número de casos e maior número de óbitos não necessariamente possuem a maior quantidade proporcional de óbitos por total.



4.2 Análise Exploratória Datas

Para essa segunda etapa de exploração, iremos nos concentrar nas variáveis relacionadas as datas, que são `data_inicio_sintomas`, `mes_inicio_sintomas` e `ano_inicio_sintomas`.

Para traçarmos um objetivo nessa análise, iremos focar nos seguintes pontos :

- **A evolução dos casos, recuperados e óbitos durante o tempo.**
- **Demonstrar a evolução da soma acumulada.**
- **Verificar em quais meses do ano de 2020 tiveram mais casos absolutos e proporcionais.**

Para auxiliar em nossa análise, iremos realizar a criação de uma tabela que realiza o agrupamento dos valores por data, indo de 25/02/2020 à 30/04/2021.

Out[31]:

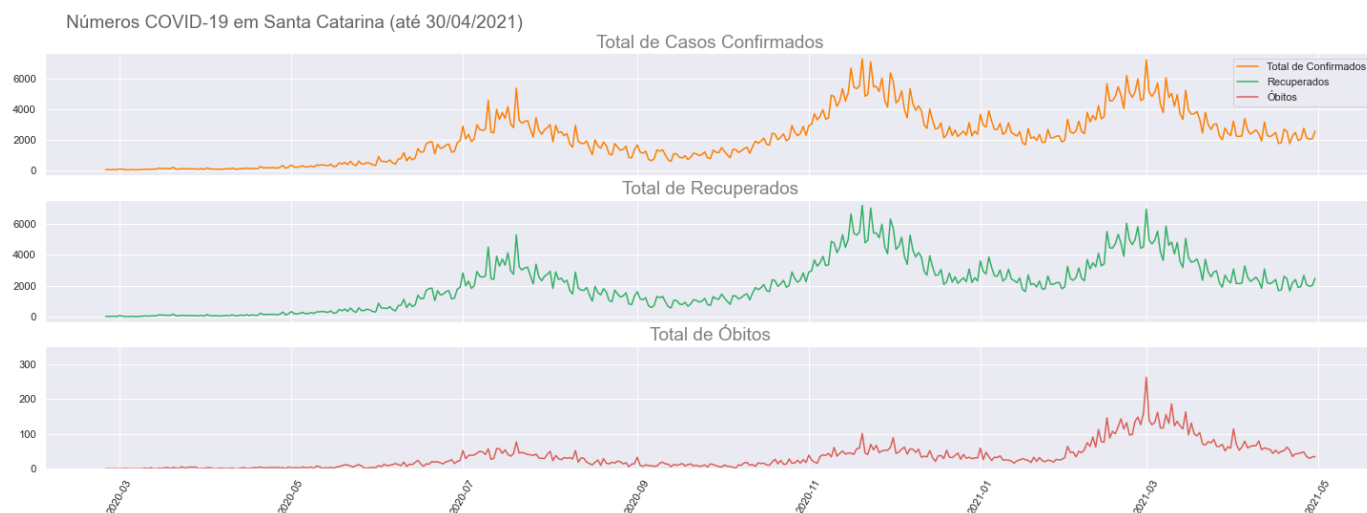
	Total	Recuperados	Obitos
data_inicio_sintomas			
2020-02-25	13	13	0
2020-02-26	18	18	0
2020-02-27	14	14	0
2020-02-28	21	21	0
2020-02-29	7	7	0
...
2021-04-26	2744	2665	48
2021-04-27	2106	2055	35
2021-04-28	2022	1973	30
2021-04-29	2040	1991	34
2021-04-30	2546	2474	35

431 rows × 3 columns

Evolução dos casos, recuperados e óbitos durante o tempo

Conforme podemos verificar abaixo, notamos que os nossos valores de Recuperados acompanha o valor de Total de Confirmados.

Também podemos notar como os gráficos se comportam de maneira muito semelhante se modificando em picos de altas (07/2020, 11/2020 e 03/2021) e baixas.



Demonstrar a evolução da soma acumulada



Verificar em quais meses do ano de 2020 tiveram mais casos absolutos e proporcionais

Para responder essa questão, iremos realizar a criação de uma tabela que realiza o agrupamento dos dados por mês e também demonstrar a proporção, sendo Total o valor 100% de cada mês.

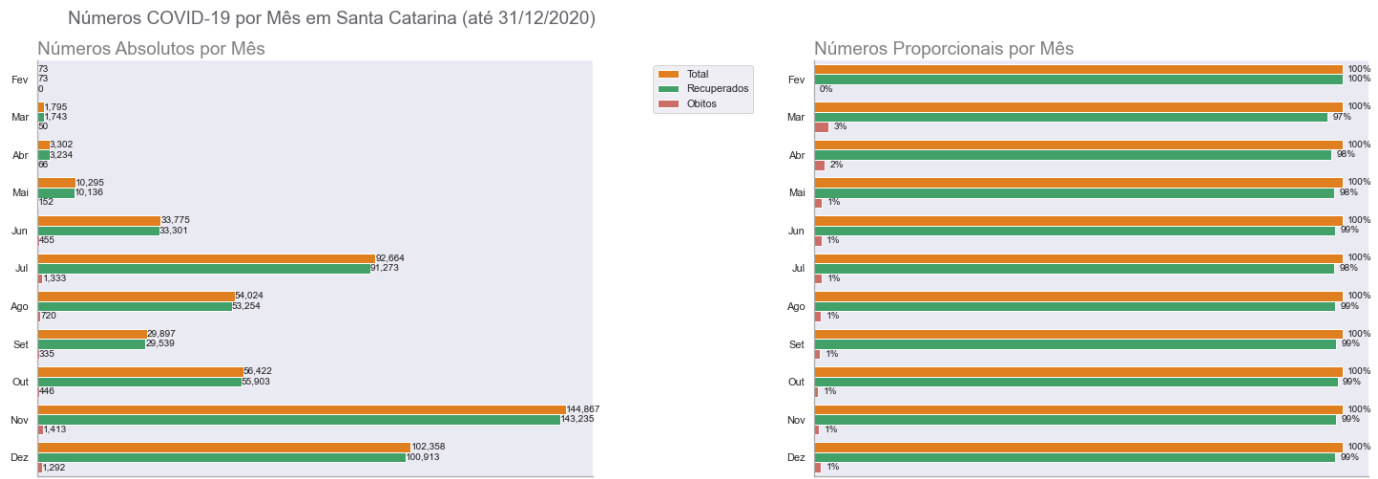
Out[34]:

	mes_inicio_sintomas	Tipo	Quantidade	Quantidade_Prop
0	2	Total	73	100.0
1	3	Total	1795	100.0
2	4	Total	3302	100.0
3	5	Total	10295	100.0
4	6	Total	33775	100.0

Conforme podemos verificar no gráfico abaixo, podemos identificar que em **Novembro ocorreu o pico de casos absolutos confirmados (144.867) e de óbitos (1.413).**

Podemos verificar que a **maior diferença de valores de casos confirmados foi do mês de Outubro (56.422) para Novembro (144.867),** onde passamos do dobro.

Também é importante destacar a **constância dos óbitos indicarem 1% do Total de Casos Confirmados por mês.**



4.3 Análise dos Sintomas

Finalizada a exploração da nossa primeira tabela, iremos realizar a exploração da nossa tabela de sintomas criada anteriormente.

Para a nossa tabela de sintomas, possuímos alguns termos médicos, como **Dispneia**, **Mialgia** e **Cefaleia**, que são entendidos, respectivamente, como **Falta de Ar**, **Dor Muscular** e **Dor de Cabeça**.

Ao criar a nossa tabela, realizamos a divisão da quantidade de cada sintoma em pacientes que se recuperaram e que foram à óbito.

Como objetivo principal da nossa análise para a tabela de sintomas, vamos **verificar os valores absolutos e a % dos sintomas de Total de Confirmados, Recuperados e Óbitos.**

Out[37]:

	Quantidade_Sem_Obito	Sintoma_Sem_Obito_Por_Pessoa	Quantidade_Com_Obito	Sintoma_Com_Obito_P
Tosse	434560	65.54	10531	
Febre	315201	47.54	8295	
Dor De Garganta	309571	46.69	3440	
Dispneia	125342	18.90	12916	
Mialgia	121707	18.35	1167	
Cefaleia	89884	13.56	627	
Diarreia	62078	9.36	1986	
Dor No Corpo	53060	8.00	452	
Coriza	47788	7.21	342	
Cansaco	22229	3.35	320	
Congestao Nasal	8632	1.30	26	

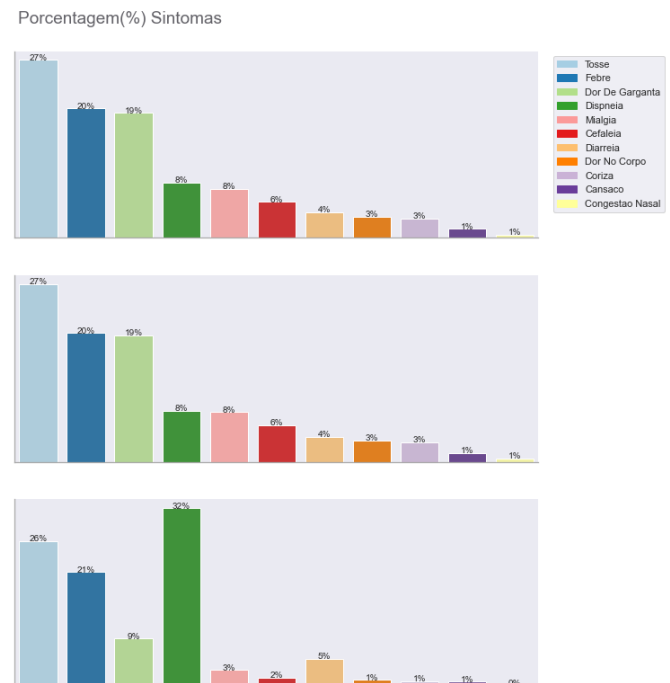
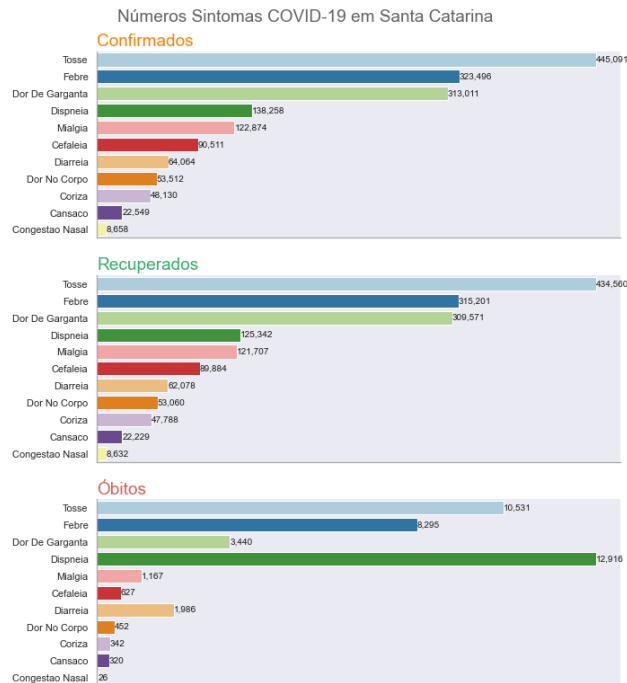
Valores absolutos e a % dos sintomas de Total de Confirmados, Recuperados e Óbitos

É importante notarmos ao realizar a verificação do gráfico, que a ordem dos sintomas e a porcentagem(%), seguem o mesmo padrão para o número de Total de Confirmados e Recuperados do COVID-19.

Ao realizar a verificação dos sintomas em pessoas que foram à óbitos, podemos já verificar um aumento expressivo na porcentagem(%) para **Dispneia(Falta de Ar)**, considerado um dos sintomas graves do COVID-19.

É importante também notarmos que ocorre uma diminuição % (19% para 9%) de Dor de Garganta, um dos sintomas considerados menos comuns e leves.

Para os sintomas de casos totais, podemos verificar que tosse e febre, os sintomas mais comuns da COVID-19 (já comprovados), aparecem, respectivamente, em 1º e 2º lugar dos sintomas mais sentidos.



5. Criação de Hipóteses

Ufa! Finalmente realizada essa intensa etapa de análise exploratória dos nossos dados, podemos realizar a formulação de alguns questionamentos para serem respondidos.

Para essa etapa, formulei os seguintes questionamentos :

- Os municípios que possuem maior número de confirmados também possuem o maior número de óbitos?
- Em 2020, no período do inverno possuímos maior número de confirmados e óbitos devido à baixa imunidade?
- É possível identificar uma maior proporção de óbitos em algum sexo biológico específico?
- Podemos identificar uma maior proporção de óbitos por confirmados em idosos?
- Os pacientes que foram à óbito possuíam sintomas diferentes dos que se recuperaram?

5.1 Os municípios que possuem maior número de confirmados também possuem o maior número de óbitos?

Realizando a análise dos valores absolutos por município de Total de Casos Confirmados e Óbitos por COVID-19, temos como resposta que **sim, municípios que possuem maior número de confirmados possuem o maior número de óbitos.**

Ao realizar a comparação, podemos verificar que 7 municípios estão presentes em ambas listas ordenadas.

Ordem Total de Casos Confirmados
Joinville
Florianopolis
Blumenau

Ordem por Óbitos
Joinville
Florianopolis
Chapeco

Chapeco
Criciúma
São José
Itajaí
Jaraguá do Sul
Palhoca
Brusque

Itajaí
São José
Blumenau
Criciúma
Lages
Tubarão
Balneário Camboriú

Porém, é **importante notarmos que ao realizar a comparação por Total de Casos Confirmados e Óbitos por Total de Casos Confirmados (%) vemos que essa premissa não é verdadeira** e possuímos os seguintes municípios com a maior porcentagem(%) :

Ordem de maior % de Óbitos por Total de Confirmados
Calmon
Paraíso
São Cristóvão do Sul
Papanduva
Marema
Lebon Régis
Passos Maia
Bela Vista do Toldo
Correia Pinto
Cacador
São Lourenço do Oeste
Macieira
Presidente Nereu
Ponte Serrada
Siderópolis

5.2 Em 2020, no período do inverno possuímos maior número de confirmados e óbitos devido à baixa imunidade?

Após a análise exploratória dos dados separados por mês, temos como resposta que **não possuímos uma alta no inverno ou em uma estação específica.**

Conseguimos constatar ao analisar os dados que os meses de **Julho, Novembro e Dezembro** correspondem aos meses que **mais tiveram casos confirmados e óbitos.**

Ordem Total de Casos Confirmados
Novembro
Dezembro
Julho
Outubro
Agosto
Junho
Setembro
Maio
Abril
Março
Fevereiro

Ordem por Óbitos
Novembro
Julho
Dezembro
Agosto
Junho
Outubro
Setembro
Maio
Abril
Março
Fevereiro

A explicação para esses números se baseiam no conceito de "**ondas**", um conceito constantemente utilizado para **contextualizar o comportamento da pandemia**. Segundo o pesquisador José Eduardo Levi, da Universidade de São Paulo (USP), o **surgimento** de uma onda pode ser caracterizado quando ocorre um **aumento progressivo de casos**. E após esse movimento de crescente, o **movimento natural é que a escalada nos índices atinja um pico, permaneça estabilizado por um tempo e então comece a descer, finalizando assim uma onda.**

Para o estado de Santa Catarina a **primeira onda teria sido em Março**, quando foram notificados os primeiros casos. **A segunda em Julho**, quando o quadro da doença se agravou no Estado. **E a terceira durante o mês de Outubro**.

Fonte : [26/10/2020 - NDMais - Santa Catarina enfrenta nova onda da COVID-19](#)

5.3 É possível identificar uma maior proporção de óbitos em algum sexo biológico específico?

Após verificação dos valores por Sexo Biológico, chegamos a conclusão que é possível identificar uma maior proporção de óbitos para indivíduos do sexo masculino.

Verificando os casos, identificamos **mais casos confirmados em mulheres (508.285 contra 470.023 em homens) e 490.574 (96,51% do total de confirmados em mulheres) se recuperaram** do COVID-19 contra **449.916 homens recuperados (95,72% do total)**.

Mesmo com valores menores de casos confirmados para **homens**, podemos verificar no estado de Santa Catarina que é o **sexo que possui mais mortes em valores absolutos (8.945 contra 6.515 do sexo feminino)** e proporcional (**1,9% do total** contra, aproximadamente, **1,3% do total do sexo feminino**).

Caso você tenha interesse, pode verificar essa [matéria da UOL](#) referente à maior mortalidade masculina decorrente do COVID-19.

5.4 Podemos identificar uma maior proporção de óbitos em idosos?

Sim! Ao realizarmos a análise dos dados referente à idade, conseguimos verificar um **aumento considerável na porcentagem (%) de pacientes que foram à óbito para os grupos maiores que 60 anos**, sendo :

- 61-70 anos (5% do Total de Positivos foram à óbito)
- 71-80 anos (12% do Total de Positivos foram à óbito)
- Maior que 81 (22% do Total de Positivos foram à óbito)

A explicação para essa **maior proporção de óbitos em idosos** se deve a **maior fragilidade do sistema imunológico** (menos anticorpos no organismo), que traz uma **dificuldade em enfrentar diversos tipos de doenças**.

Somado à essas características naturais temos o fato de que os pulmões e as mucosas dos idosos se tornam mais frágeis e mais suscetíveis a doenças virais (como o COVID-19).

Também sabemos que o **público dessa faixa etária vai ao hospital com mais frequência, o que aumenta o tempo de exposição a ambientes com micro-organismos e possuem um número maior de comorbidades** (diabetes, hipertensão, doenças cardíacas ou pulmonares), aumentando assim a chance de possuir um quadro grave se forem contaminados.

Fonte : [Medprev - Por que o organismo dos idosos é mais suscetível à Covid-19 e outras doenças?](#)

5.5 Os pacientes que foram à óbito possuíam sintomas diferentes dos que se recuperaram?

Os pacientes **não possuíam sintomas diferentes**, mas **possuíam uma porcentagem(%) maior para alguns sintomas!**

Realizando a análise da tabela dos nossos sintomas, verificamos que ocorre um **grande aumento na porcentagem(%) para Dispneia (Falta de Ar)**, um dos sintomas graves causados pelo COVID-19, aos pacientes que foram a óbito.

É importante também notarmos que ocorre uma **diminuição na porcentagem (19% para 9%) de Dor de Garganta**, um dos sintomas considerados menos comuns e leves.

6. Conclusão

Esse projeto foi muito interessante para a visualização de como está sendo o período de pandemia em meu estado (Santa Catarina) além de desenvolver novas habilidades para a área de análise de dados.

Com a Análise Exploratória, conseguimos identificar as chamadas "ondas" (movimento de crescente, estabilização e diminuição dos números da pandemia) e verificamos que o Total de Confirmados, Recuperados e Óbitos seguem o mesmo padrão.

Também foi interessante validarmos que o público masculino e grupos acima de 60 anos possuem % maiores para a taxa de óbito por COVID-19.

Verificamos também que ocorre um aumento considerado na % de Dispneia (Falta de Ar) para os pacientes que foram à óbito, sendo o sintoma mais recorrente.