

Nome: _____ Matrícula: _____
Data: __/__/__

PROJETO DE DATA SCIENCE CORONAVIRUS COVID-19

“A pandemia de COVID-19 no Brasil teve início em 26 de fevereiro de 2020, após um homem de 61 anos de São Paulo que retornou da Itália testou positivo para a SARS-CoV-2. Desde então, em 7 de maio de 2020, confirmaram-se 125 218 casos, a maior parte deles no estado de São Paulo, causando 8 536 mortes. A transmissão comunitária foi confirmada para todo o território nacional.”

Fonte: https://pt.wikipedia.org/wiki/Pandemia_de_COVID-19_no_Brasil em 07/05/2020.

DATASET

Vários *datasets* estão disponíveis na internet para análise dos dados do avanço da epidemia.

O *dataset* anexado ao exercício foi extraído no site <https://brasil.io/dataset/covid19/caso/> no dia 07/05/2020. Sugiro para a execução do exercício, sempre baixar o *dataset* mais recente.

ATIVIDADES

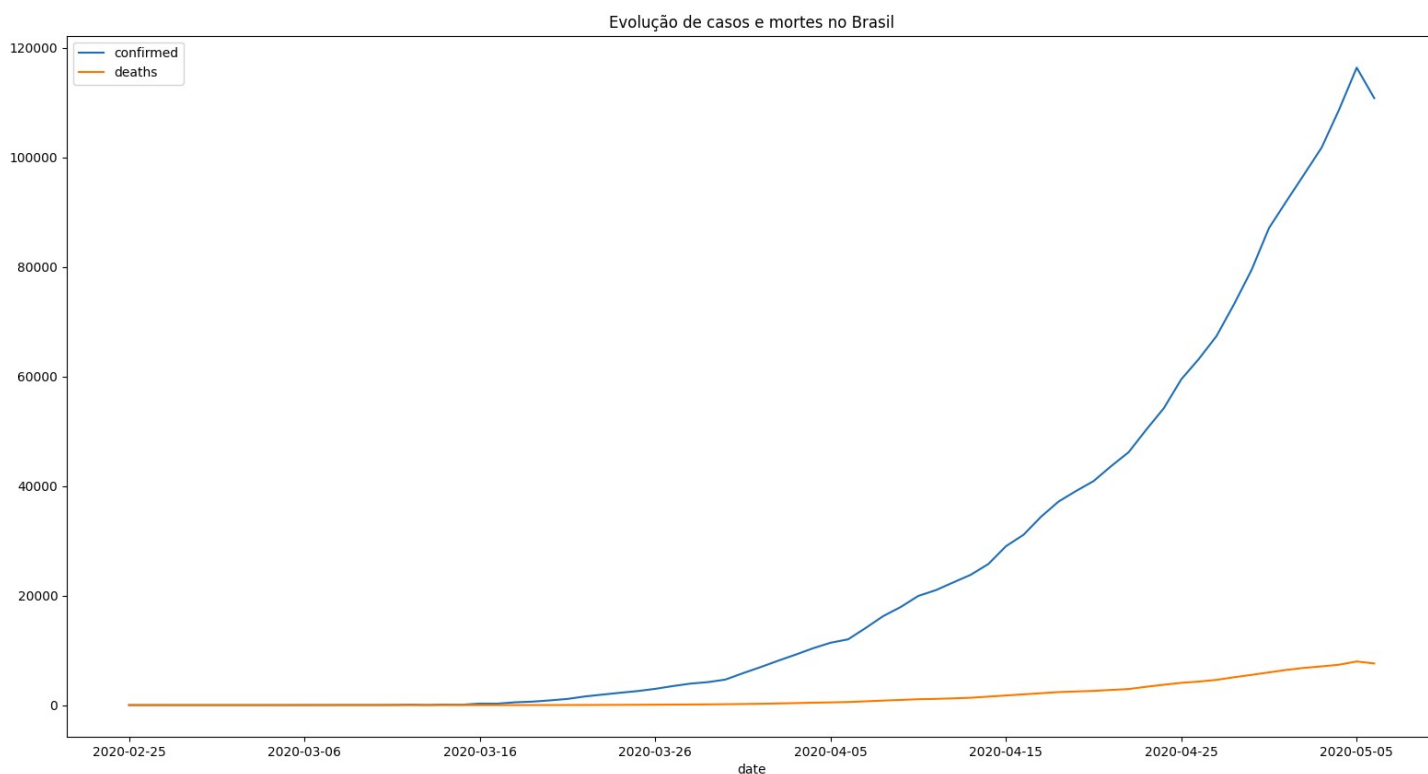
Elabore um programa em *Python* com menus interativos para as seguintes situações:

1 – Elaboração de gráficos e imagens para visualização dos dados:

Dataset: <covid19_20200507.csv>

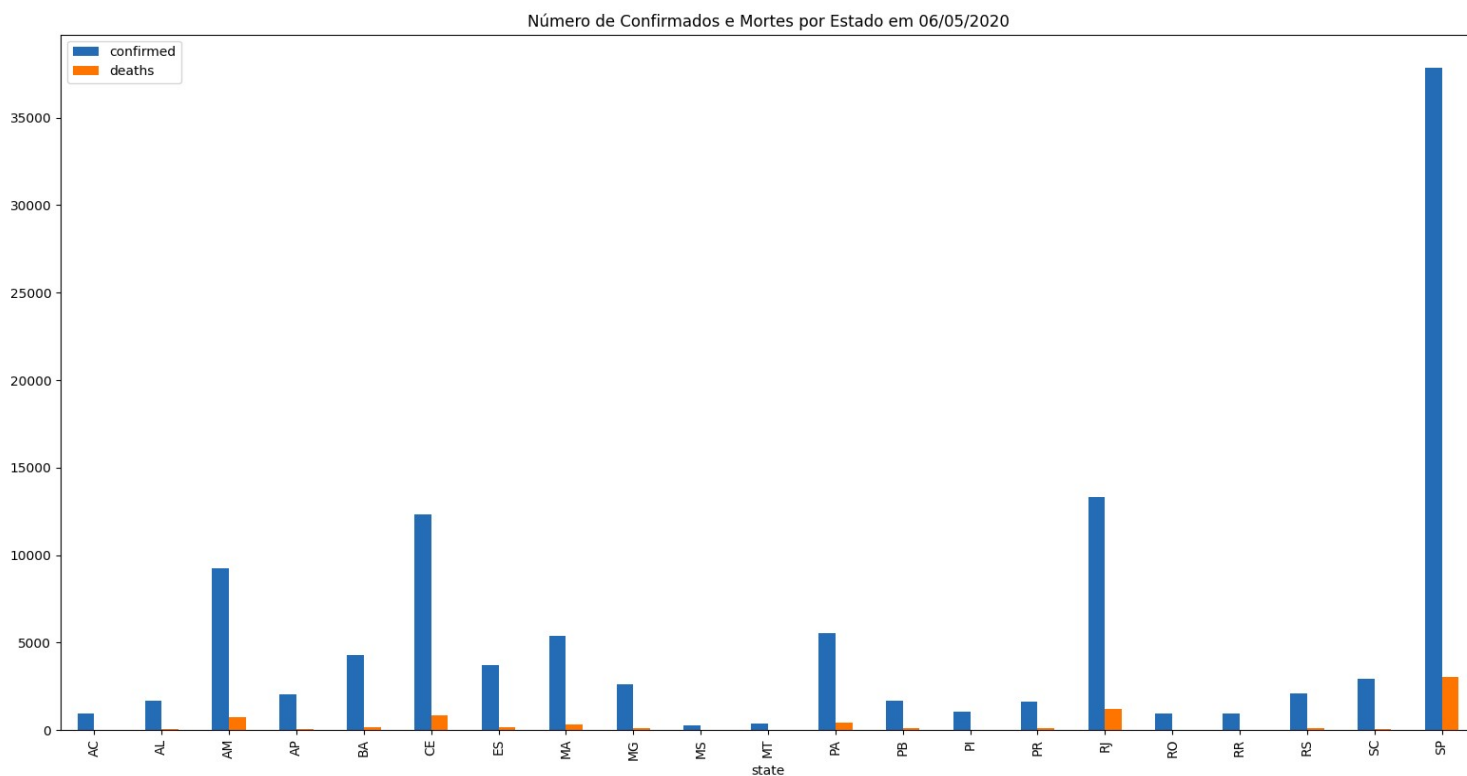
Granularidade: Brasil

A) Gráfico de linha de evolução de casos e mortes no Brasil. Dica: Deve-se somar os casos dos estados agrupados por data. O gráfico deve ficar assim:



Para os próximos, deve-se extrair os dados pela última data mais atualizada.

B) Gráfico de barras para Número de Confirmados e Mortes por Estado em <data_mais_recente>. O gráfico deve ficar assim:



Imagens - Nuvem de Palavras – Word Clouds

“Uma nuvem de tags (nuvem de palavras) é uma representação visual inovadora de dados de texto, normalmente usada para representar metadados de palavras-chave (tags) em sites ou para visualizar texto em formato livre. Tags são geralmente palavras únicas e a importância de cada tag é mostrada com o tamanho ou a cor da fonte. Esse formato é útil para perceber rapidamente os termos mais importantes para determinar sua importância relativa.”

C) Transforme o gráfico de barras acima em uma Word Clouds por estado. Utilize a informação do número de mortes para atribuir a relevância à TAG. A imagem deve ficar assim:

***Os estados que aparecem “maiores” possuem maior número de mortes pela doença na referida data.**

Estados com maior número de mortes por COVID-19. Em 06/05/2020



D) Crie agora uma nova nuvem de palavras por estado. Utilize a informação do número de casos por cada 100k habitantes(*incidência*). A imagem deve ficar assim:

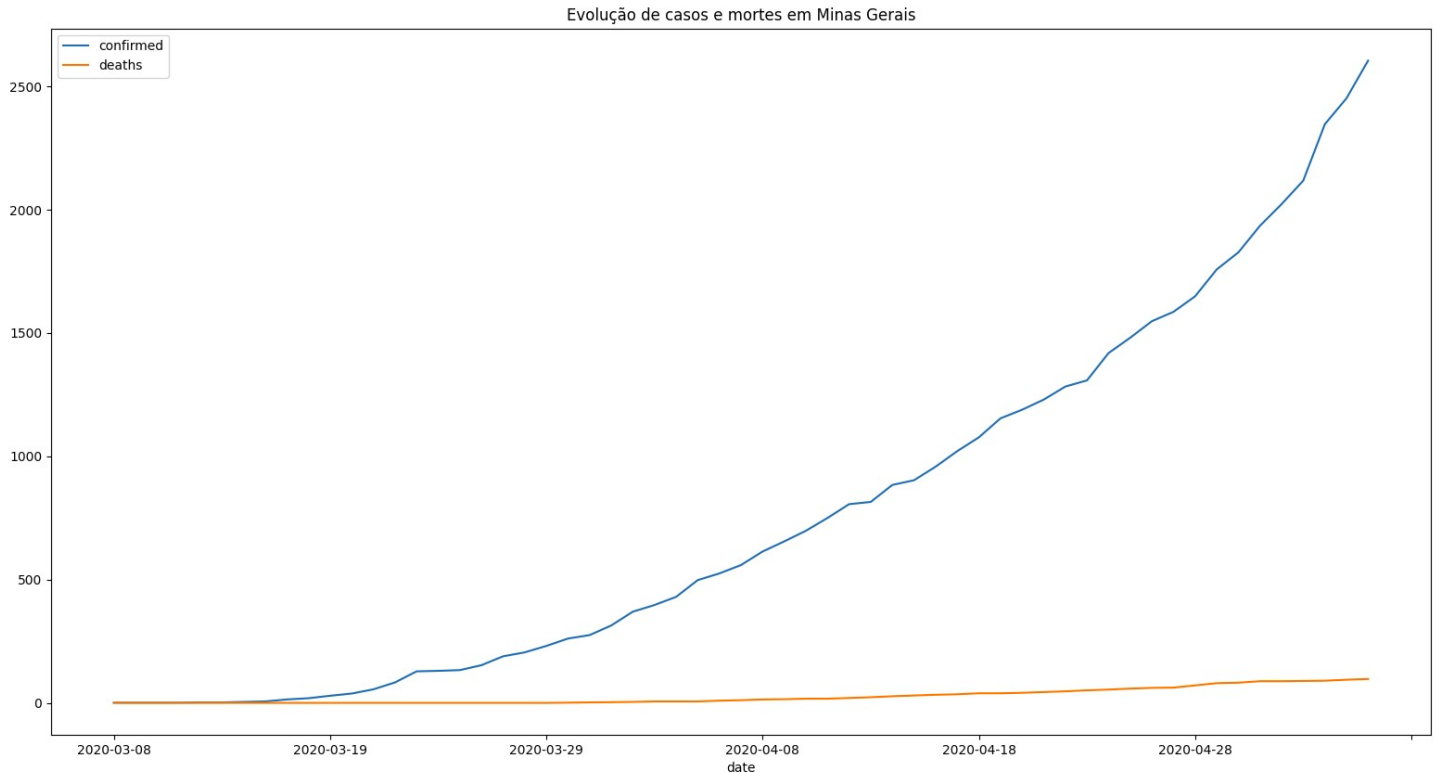
****Os estados que aparecem “maiores” possuem maior incidência(nº de casos por cada 100k habitantes) da doença na referida data.***

Estados com maior incidência(CONFIRMADOS por 100 mil habitantes) de COVID-19. Em 06/05/2020

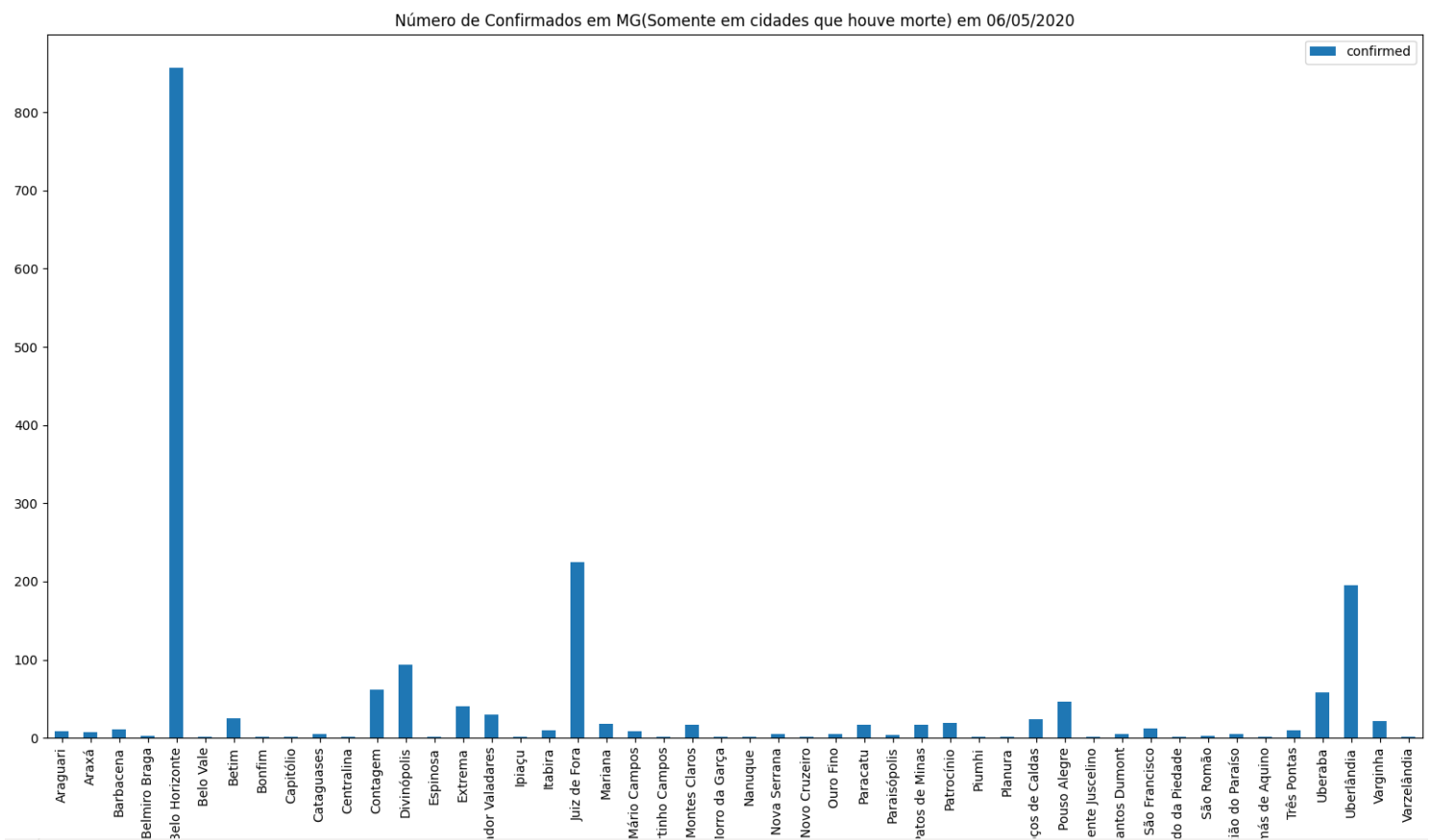


Granularidade: Estadual (Minas Gerais)

A) Gráfico de linha de evolução de casos e mortes no Em Minas Gerais. O gráfico deve ficar assim:



B) Gráfico de barras para Número de Confirmados em MG(Somente em cidades que houve morte). Para <data_mais_recente>. O gráfico deve ficar assim:



C) O gráfico da letra B não ficou legal. Então transforme-o em uma Nuvem de Palavras por Cidade. A cidade que aparecer maior é que contém maior número de casos. Liste somente as cidades em que houve morte. A imagem deve ficar assim:

Cidades de Minas Gerais com maiores números de casos e ocorrências de mortes por COVID-19. Em 06/05/2020

