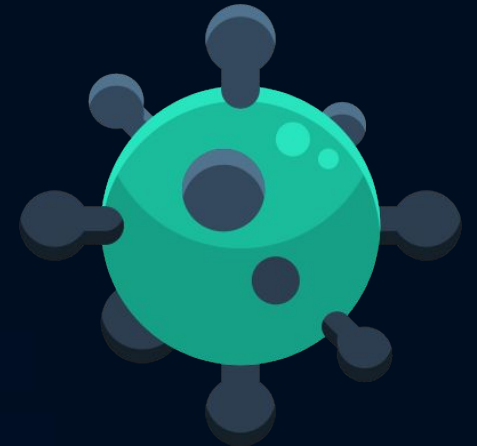


Análise da comunicação das prefeituras do Brasil e a relação com o aumento da COVID-19

IN DATA WE TRUST

Equipe

- Osmary Camila Bortoncello Guber
 - Mestrado em Computação Aplicada - UTFPR
- Jackson Cardoso
 - Mestrado em Computação Aplicada - UTFPR



Contexto

- Pandemia decorrente da COVID-19;
- Prefeituras de Municípios Brasileiros, foco nas capitais dos estados;
- Ciência de dados aplicada à COVID-19.



Objetivos

- Obter dados dos perfis da rede social Twitter das top 100 cidades com maior número de óbitos decorrentes da COVID-19;
- Comparar com o número de casos de COVID-19 por município/UF;
- Verificar se a comunicação feita pelas prefeituras têm alguma relação com o número de casos da COVID-19.

Descrição dos dados estudados

- ID do tweet; data da criação; texto; métricas públicas: quantidade de like; quantidade de retweet; quantidade de reply.
- número de casos confirmados; quantidade de óbitos; % letalidade; número de casos confirmados por 100 mil habitantes; quantidade de óbitos por 100 mil habitantes.
- população estimada; município; UF; região; capital/cidade.

Perguntas de pesquisa

- Existe alguma relação no número de casos da COVID-19 que sejam decorrentes da comunicação das prefeituras?
- Existe uma diferença de mortalidade de COVID-19 em cidades que se comunicam mais no Twitter do que aquelas que não se comunicam?
- Prefeituras que se comunicam bem no Twitter tem maior probabilidade de diminuição nos casos da COVID-19?

Fontes de dados

- Dados principais obtidos da rede social Twitter:
 - Limite de 3200 tweets por perfil de prefeitura, do tweet mais novo para o antigo;
 - 72 mil tweets após a limpeza dos dados.
- Dados base dos boletins de COVID-19 obtidos por meio do portal Brasil.io:
 - Top 100 municípios com maior número acumulado de óbitos;
 - Destas 100 prefeituras, apenas 90 possuem perfil no twitter.
- Dados auxiliares obtidos por meio do portal do IBGE (Instituto Brasileiro de Geografia e Estatística).

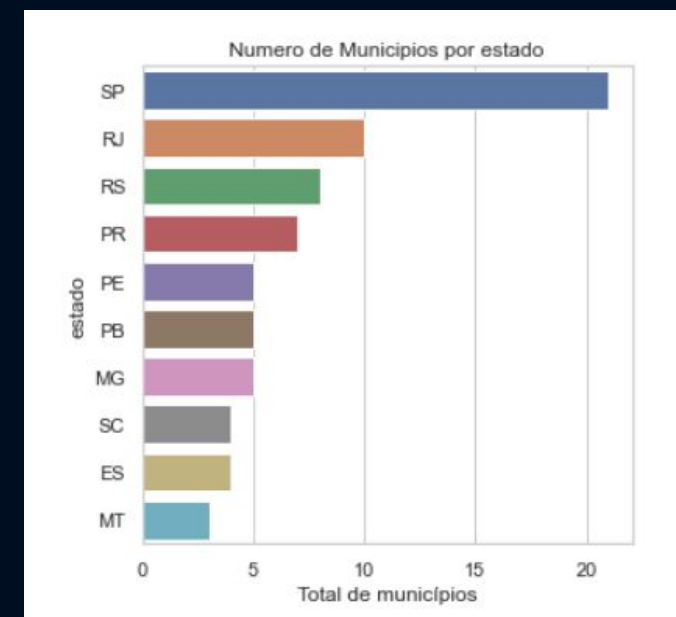
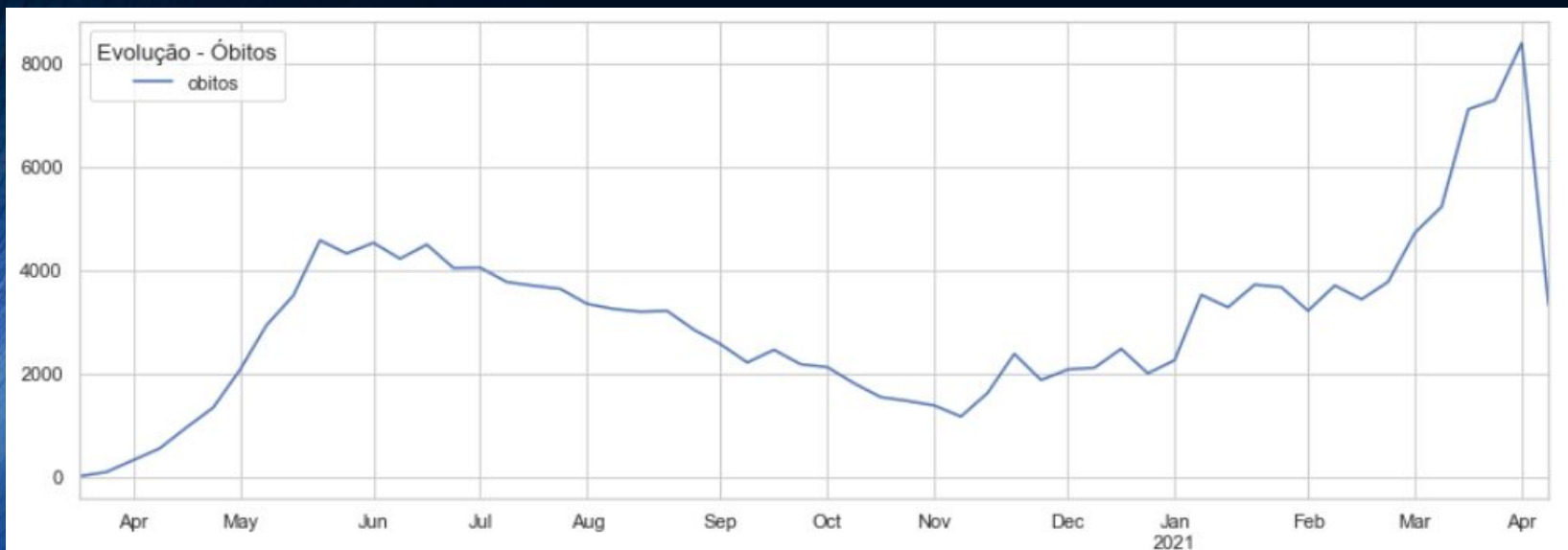
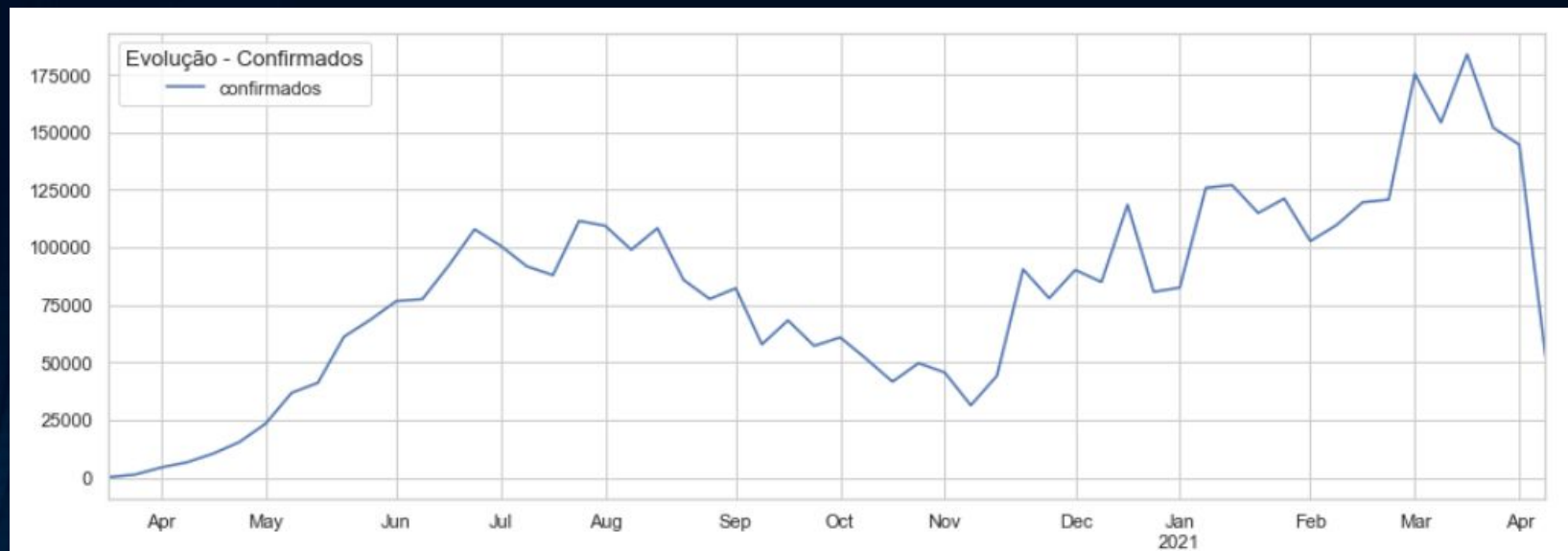
Processamentos executados até o momento

- Extração de dados da rede social Twitter;
- Extração de dados do portal Brasil.io;
- Tratamento dos conjuntos de dados gerados;
- Pré-processamento dos dados do Twitter;
- Plotagem dos gráficos utilizando python via Google Colab.

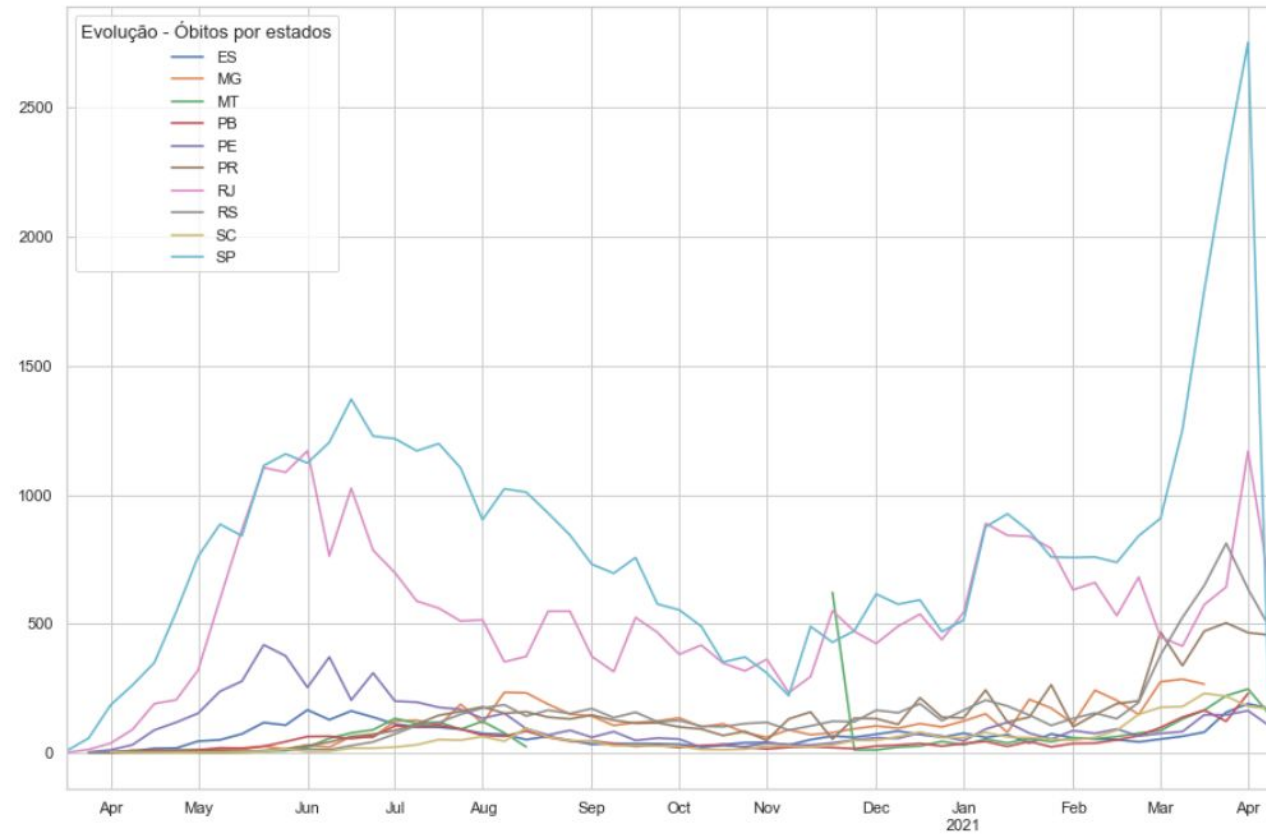
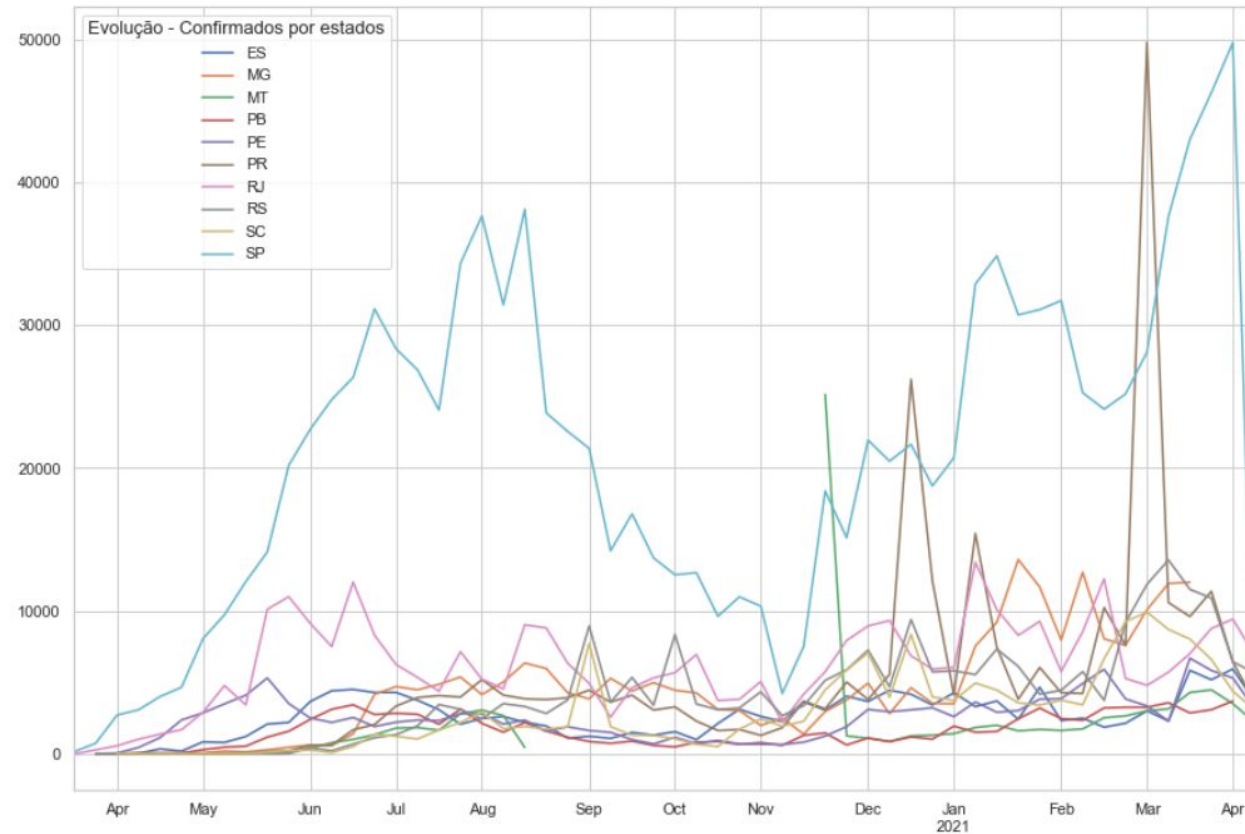
Resultados mais importantes



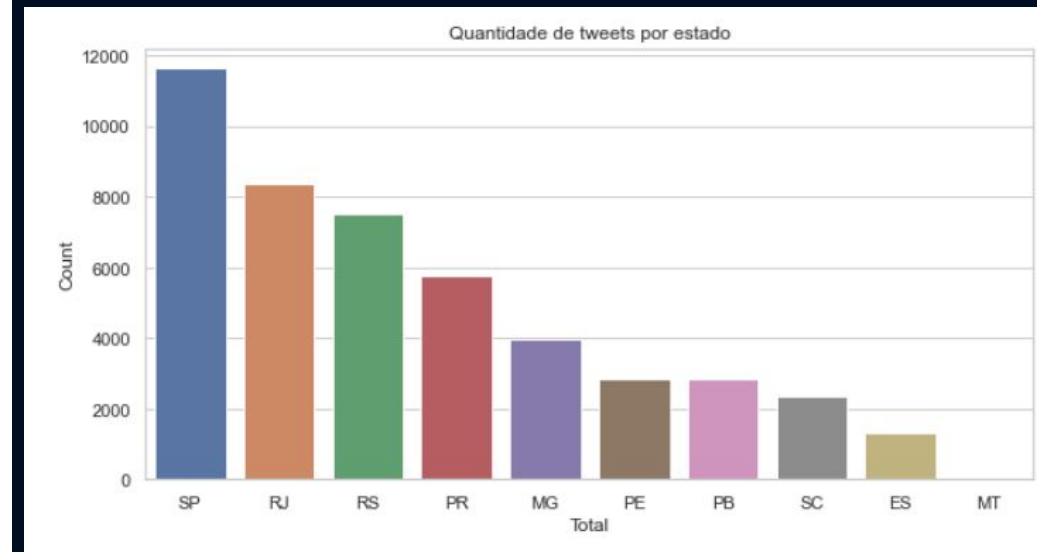
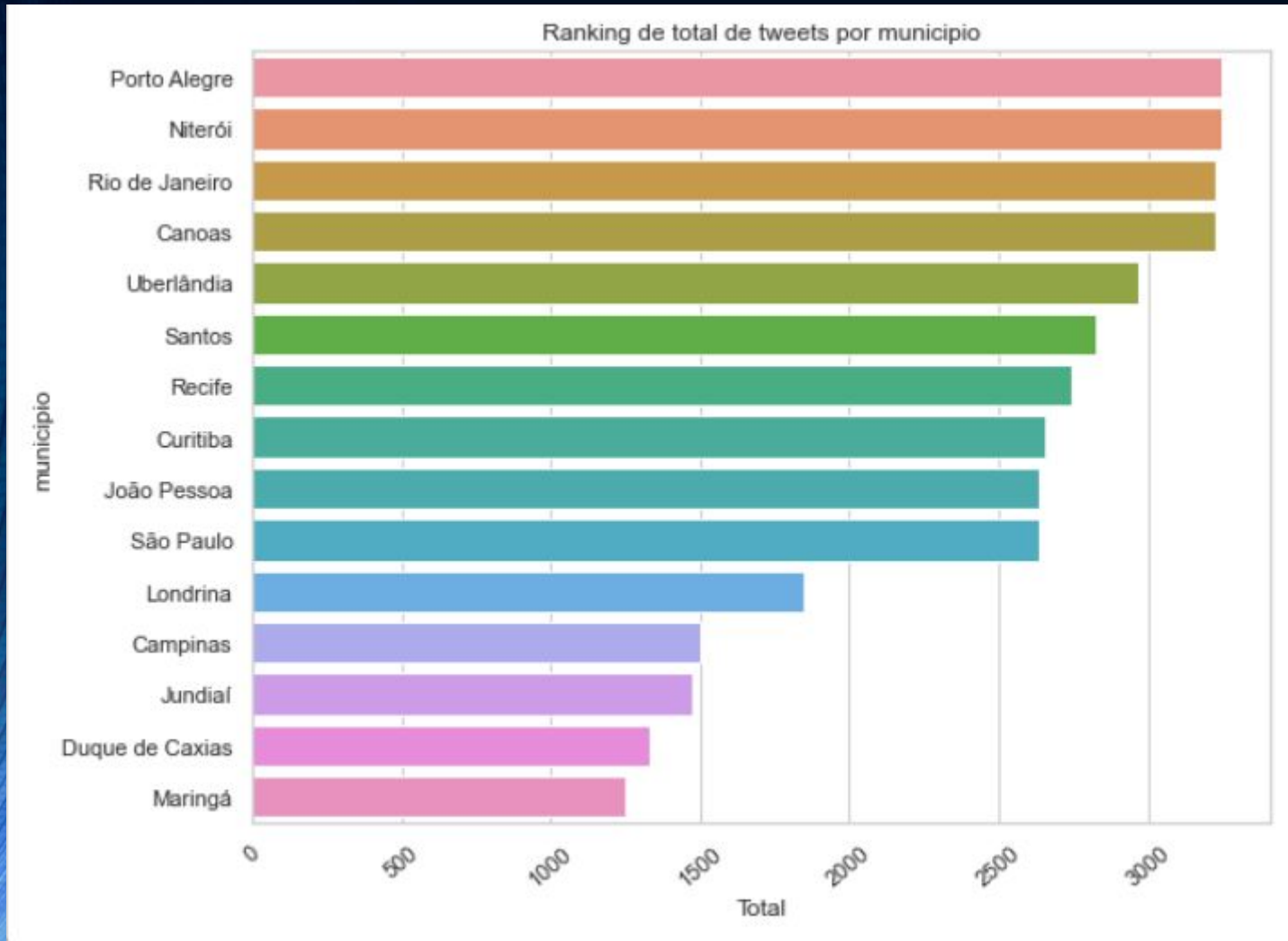
Evolução de casos confirmados X óbitos da COVID-19



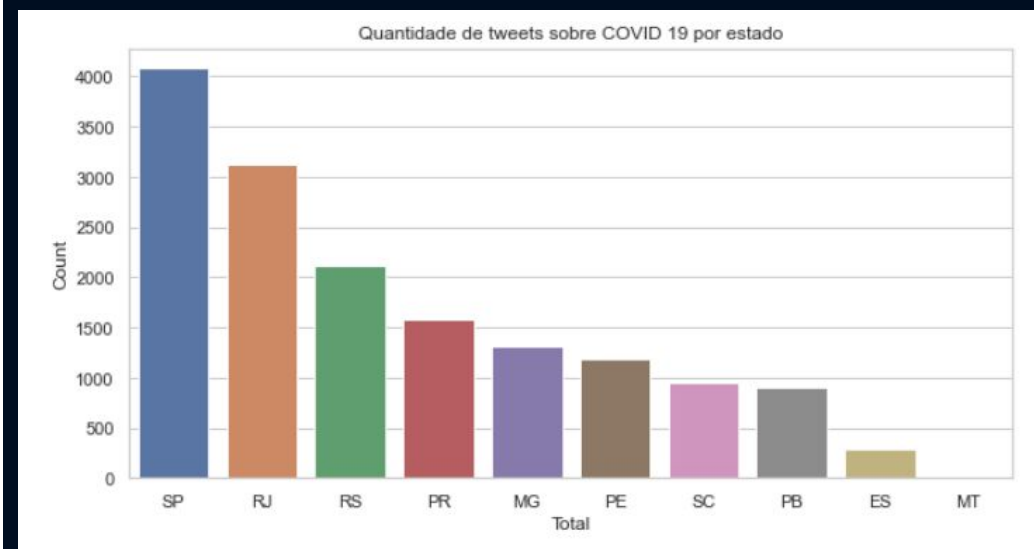
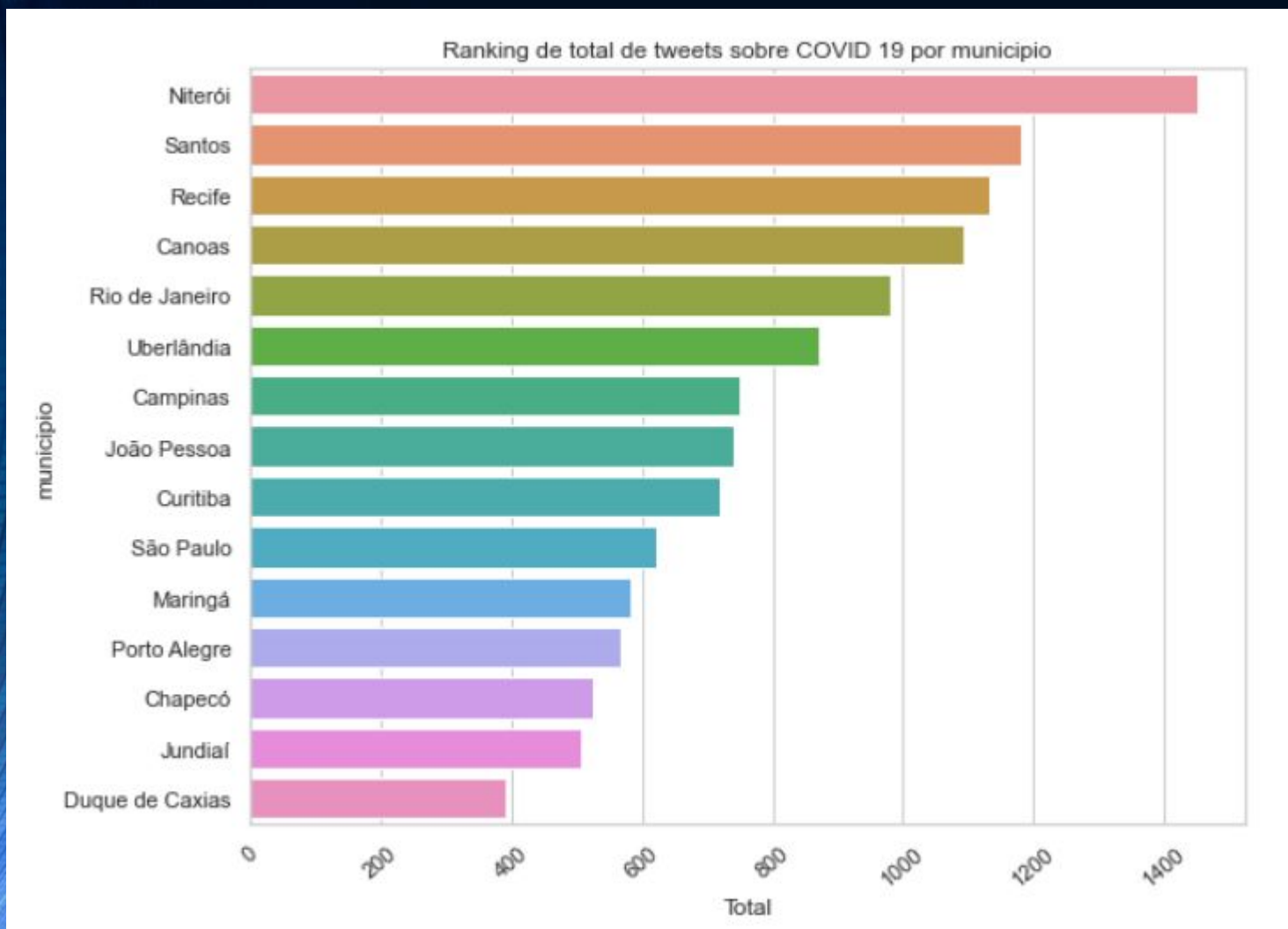
Evolução de casos confirmados X óbitos da COVID-19 por estado



Top municípios e estados com mais Tweets em geral

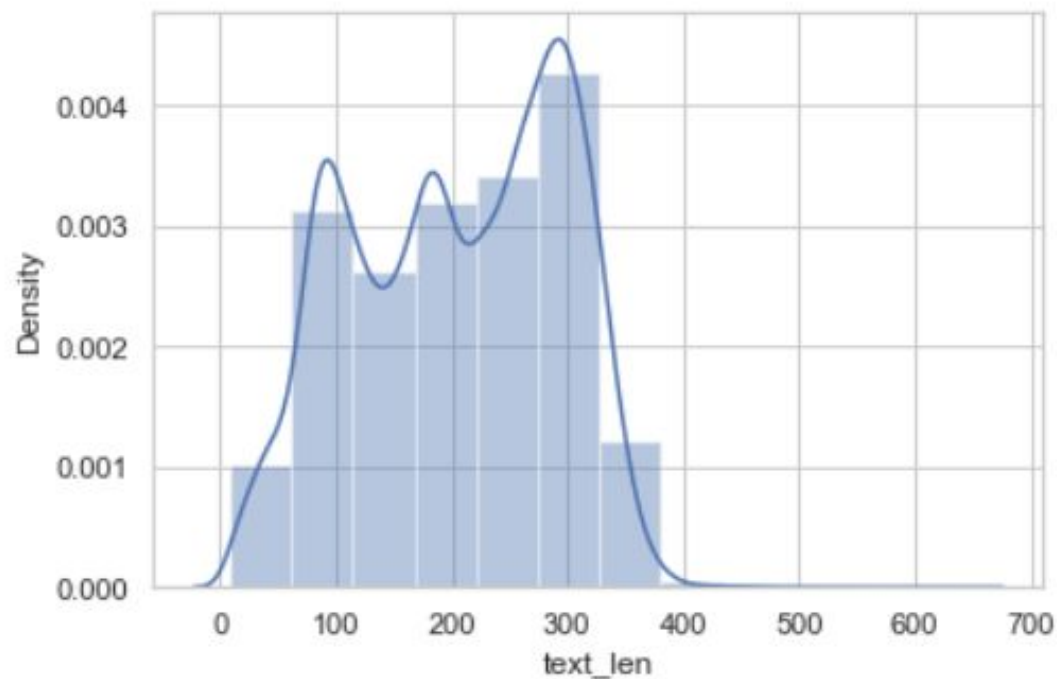


Top municípios com mais Tweets sobre COVID-19

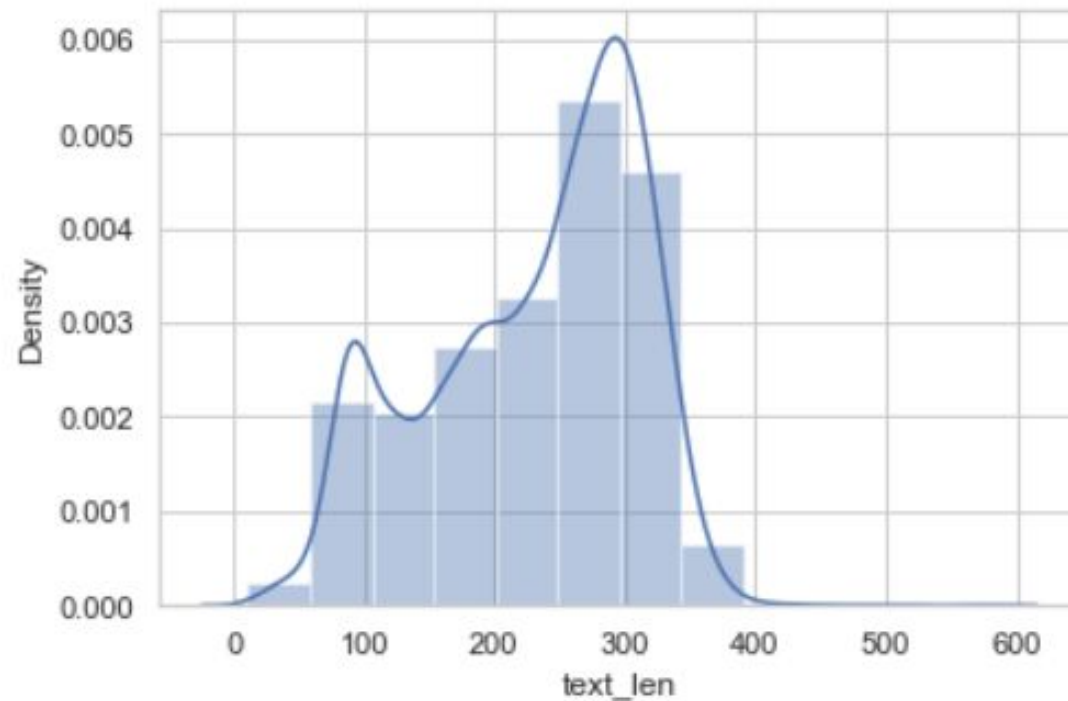


Distribuição de tamanho dos tweets

Tweets em geral

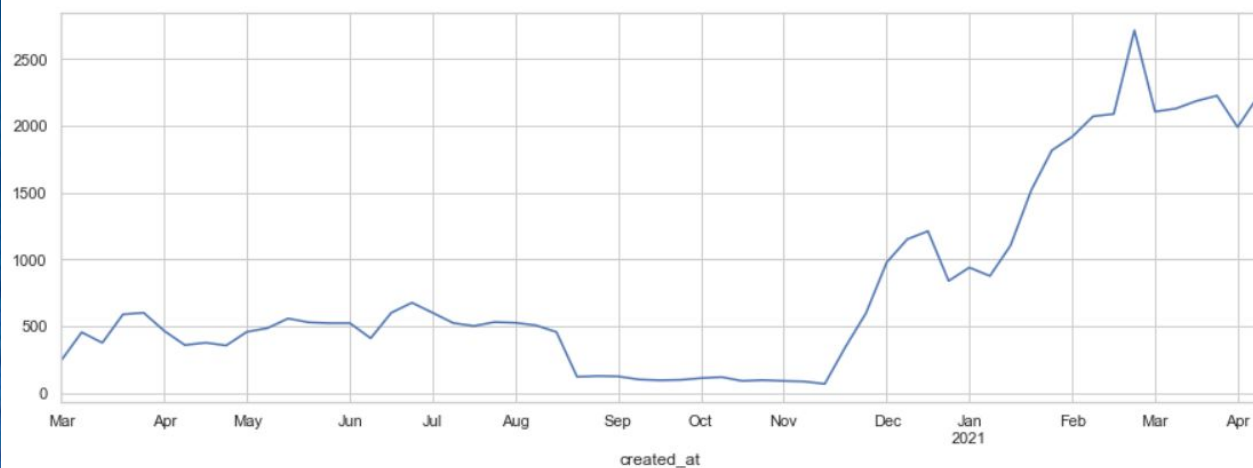


Tweets com termos relacionados a covid

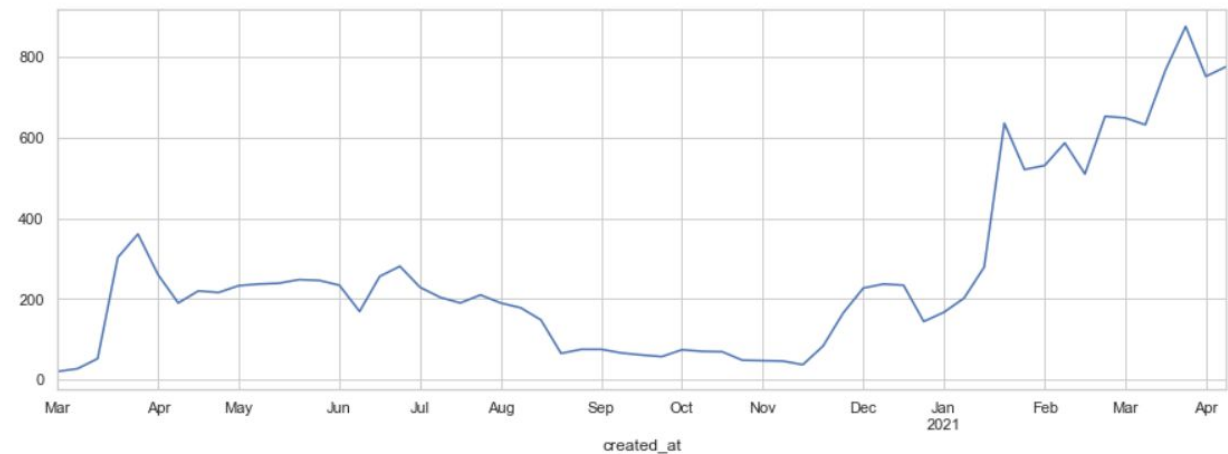


Quantidade de tweets das prefeituras no período da covid

Tweets em geral

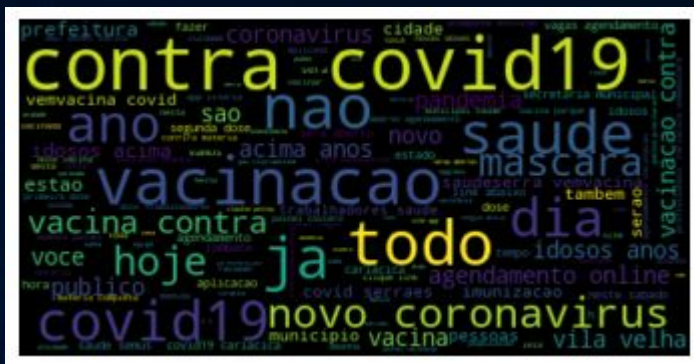


Tweets com termos relacionados a covid

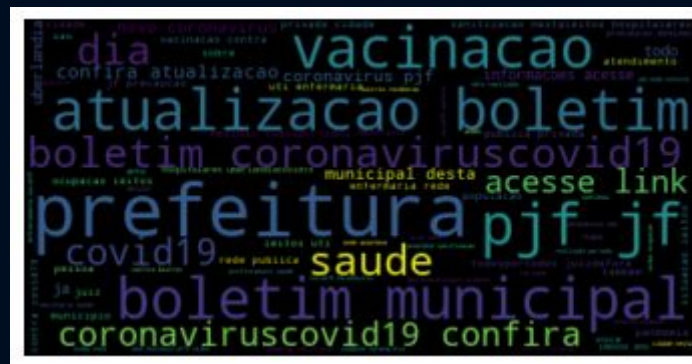


Distribuição dos termos sobre a COVID-19 mais utilizados nos Tweets por UF

ES



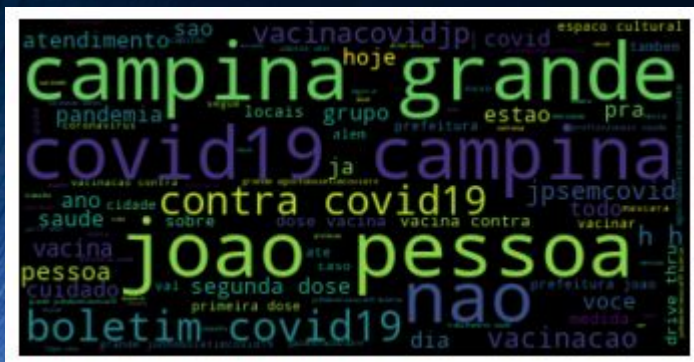
MG



MT



PB



PE

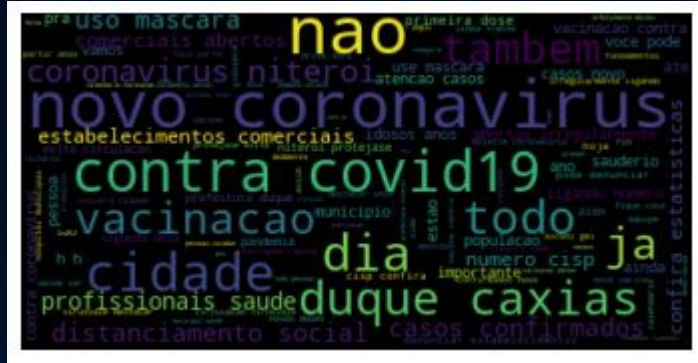


PR

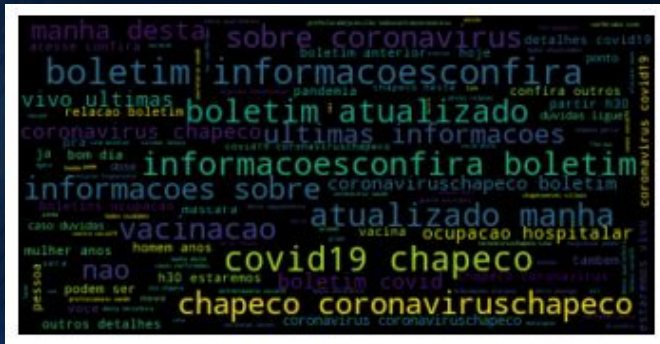


Distribuição dos termos sobre a COVID-19 mais utilizados nos Tweets por UF

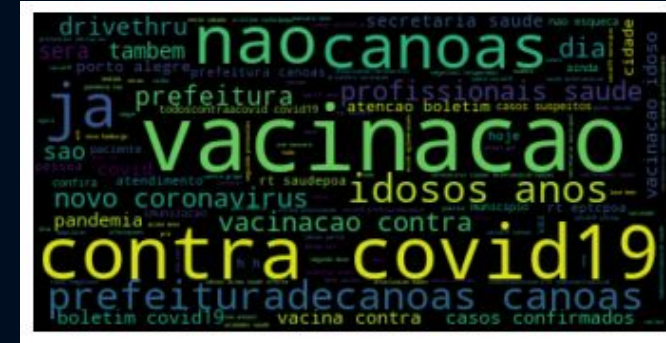
RJ



SC



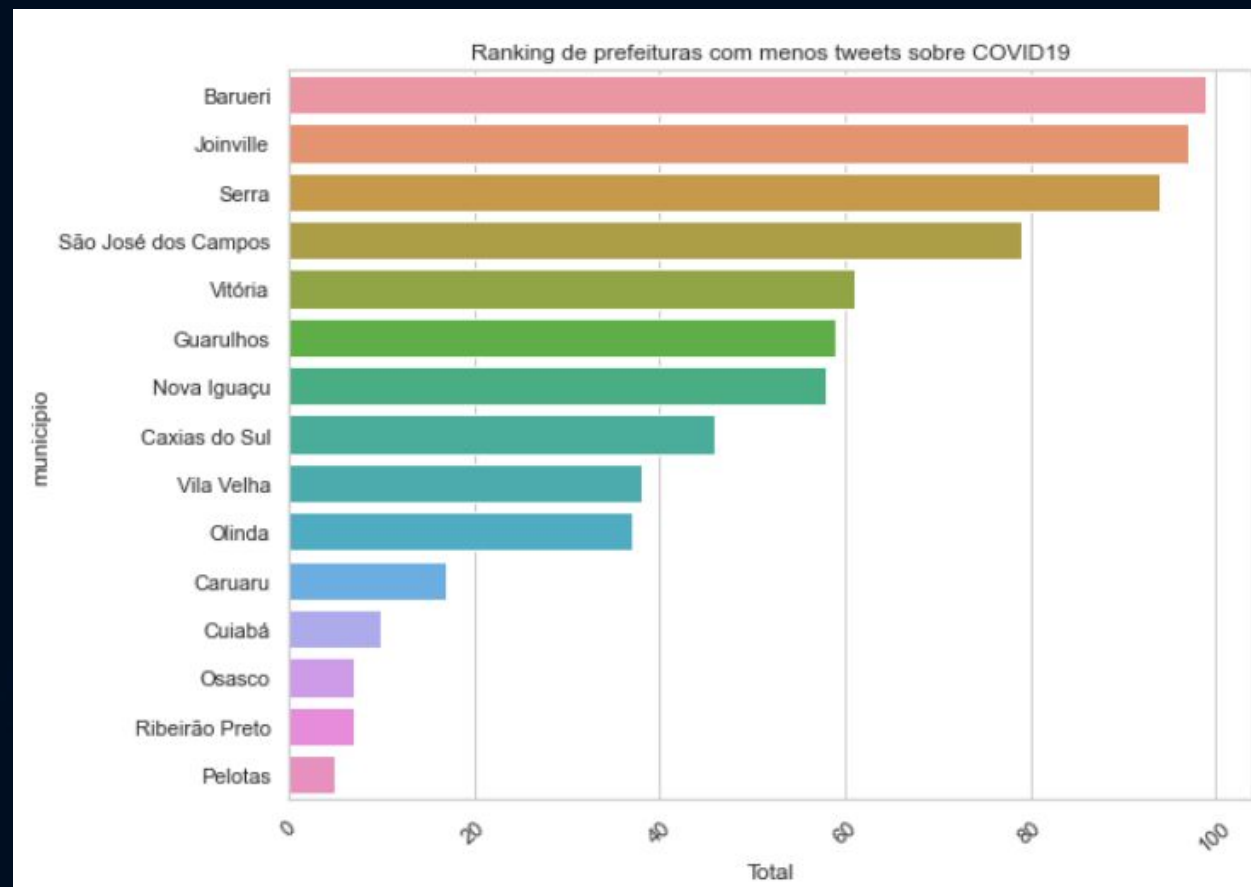
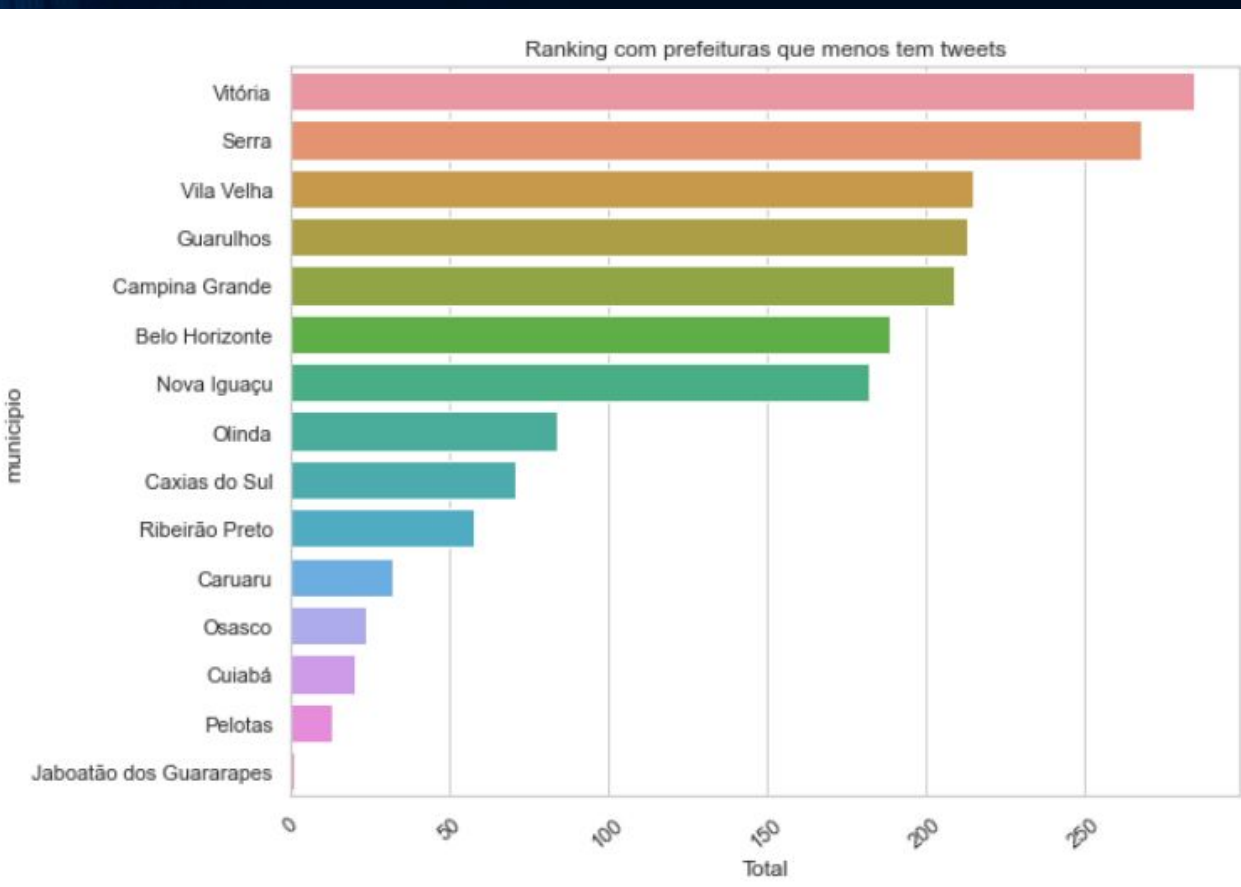
RS



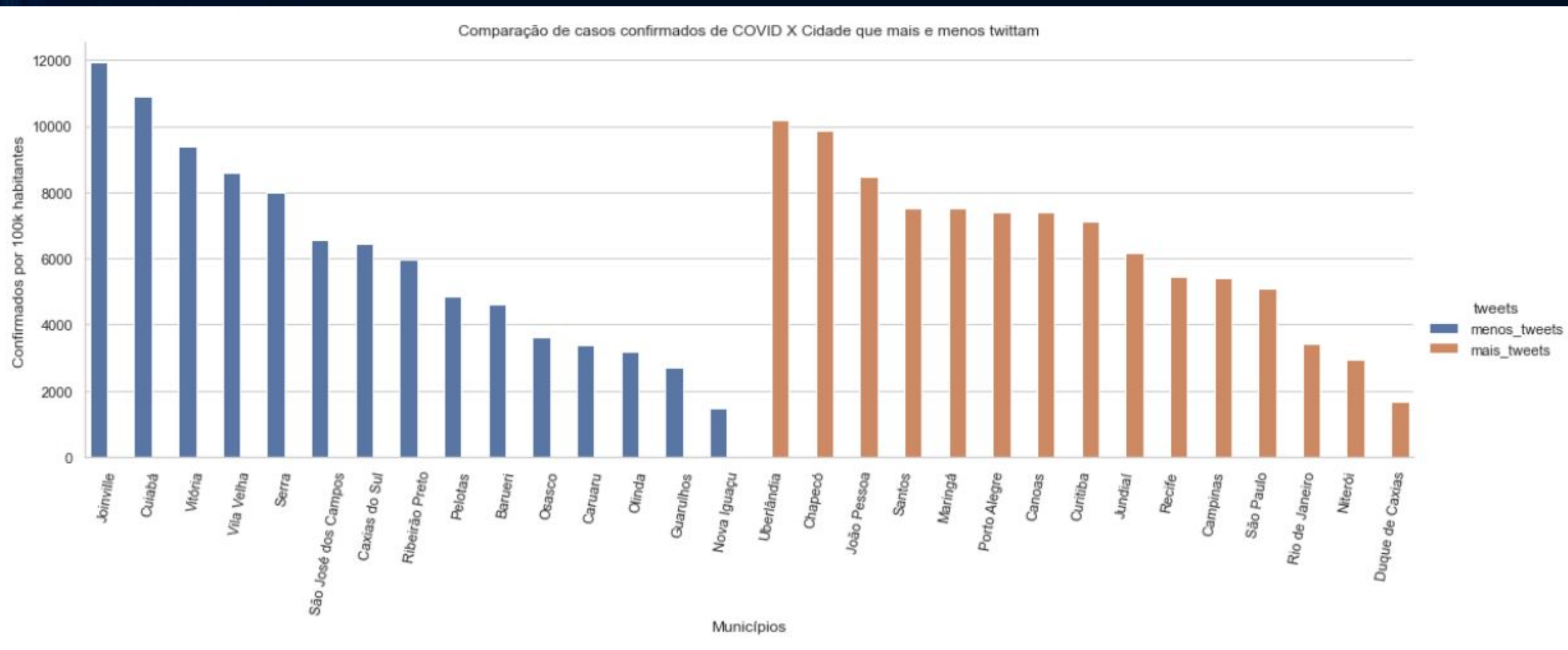
SP



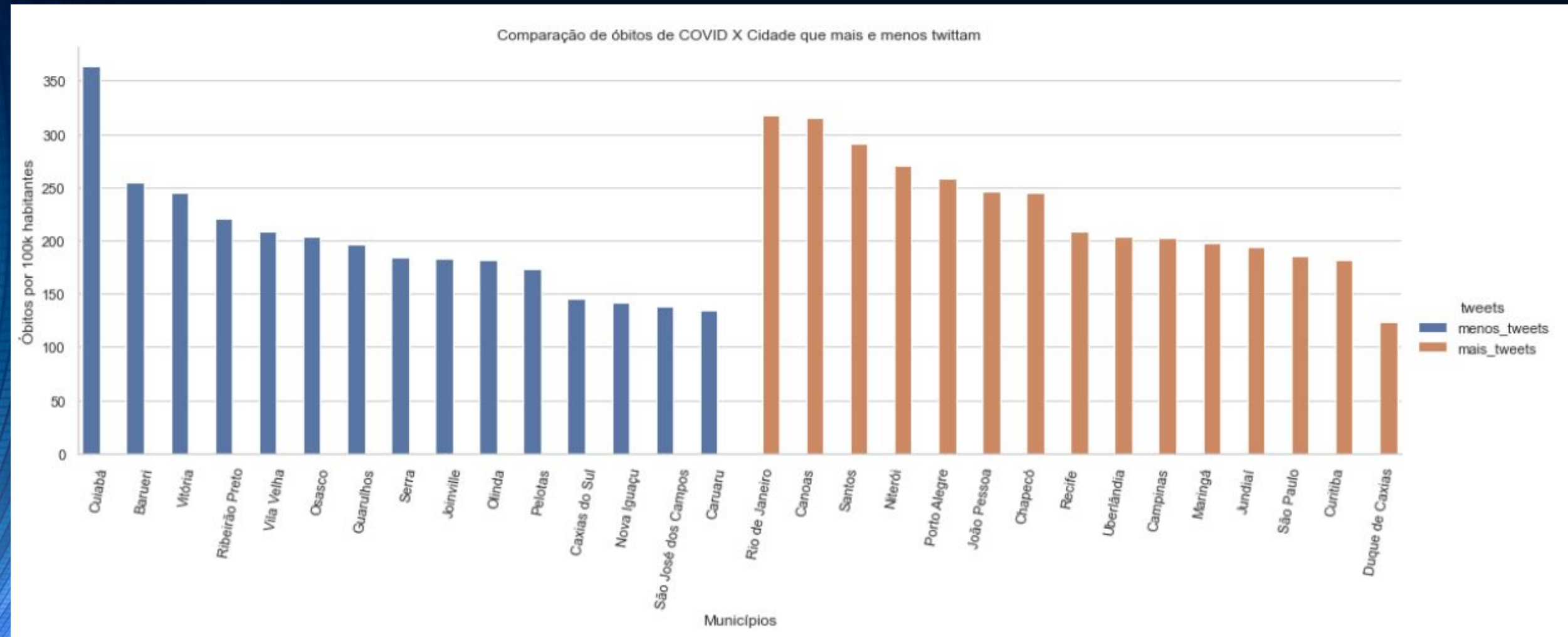
Top municípios com menos Tweets



Comparação das prefeituras que menos twittam sobre COVID-19 com as que mais twittam



Comparação das prefeituras que menos twittam sobre COVID-19 com as que mais twittam



Próximos passos

- Implementação do classificador de polaridades em português;
- Análise e visualização temporal dos dados de texto do Twitter;
- Análise exploratória a fim de testar as hipóteses e responder as perguntas da pesquisa;
- Implementação de estudo com as cidades mais populosas dos estados brasileiros além das capitais.

Desafios

- Correlação dos tweets com os números da COVID-19 inerentes à pesquisa;
- Complexidades nos processamentos dos dados de textos do Twitter (Análise de sentimentos, Topic Modelling e lematização);
- Aplicação de modelo específico para análise de dados de texto.

Obrigado!

