

Atividade 02 - Análise Exploratória

Nesta etapa do trabalho a equipe desenvolveu uma análise exploratória sobre os dados e esta análise servirá de guia para as próximas etapas.

Especificação

Relatório: Os resultados da análise exploratória estão descritos neste relatório conforme as seções a seguir:

- **Tema:** Análise da comunicação das prefeituras do Brasil e a relação com o aumento da COVID-19.
 - Descrição: O presente trabalho pretende extrair dados dos perfis das prefeituras do Brasil na rede social Twitter, focando nas capitais e comparar com o número de casos por município, analisando se a comunicação feita pelas prefeituras têm alguma relação no aumento de casos da COVID-19.
- **Equipe:**
 - Nome: *InDataWeTrust*.
 - Integrantes:
 - Osmary Camila Bortoncello Guber (Mestrado Computação, UTFPR)
 - Jackson Cardoso (Mestrado Computação, UTFPR)
- **Obtenção e processamento de dados:**
 - Extração de dados de texto da rede social Twitter dos perfis das prefeituras analisadas, aplicando limpeza, transformação e organização;
 - Coleta de dados de população estimada residente nos municípios brasileiros por meio do portal do IBGE, aplicando limpeza e organização de dados;
 - Extração de dados do portal Brasil.io dos municípios com maior número de óbitos de COVID-19, aplicando limpeza, transformação e organização.
- **Cobertura e distribuição dos dados:**

Nesta seção apresenta-se algumas visualizações que foram implementadas a fim de atender parte das especificações do tema de pesquisa.

Figura 1 - Evolução de casos confirmados de Covid-19 - Geral



Fonte: Autor

Na Figura 1 apresenta-se a distribuição dos dados obtidos do portal Brasil.io, exibindo o número de casos confirmados no decorrer dos meses desde abril de 2020.

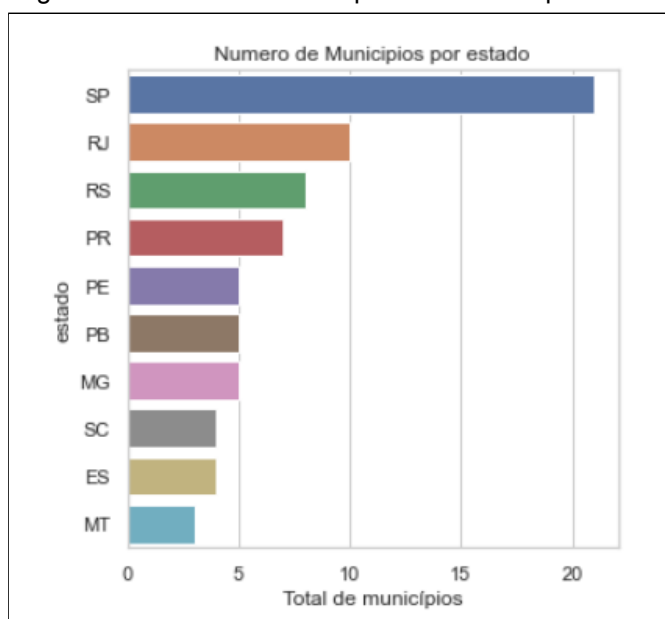
Figura 2 - Evolução de óbitos de Covid-19 - Geral



Fonte: Autor

Na Figura 2 apresenta-se a distribuição dos dados obtidos do portal Brasil.io, exibindo o número de óbitos no decorrer dos meses desde abril de 2020.

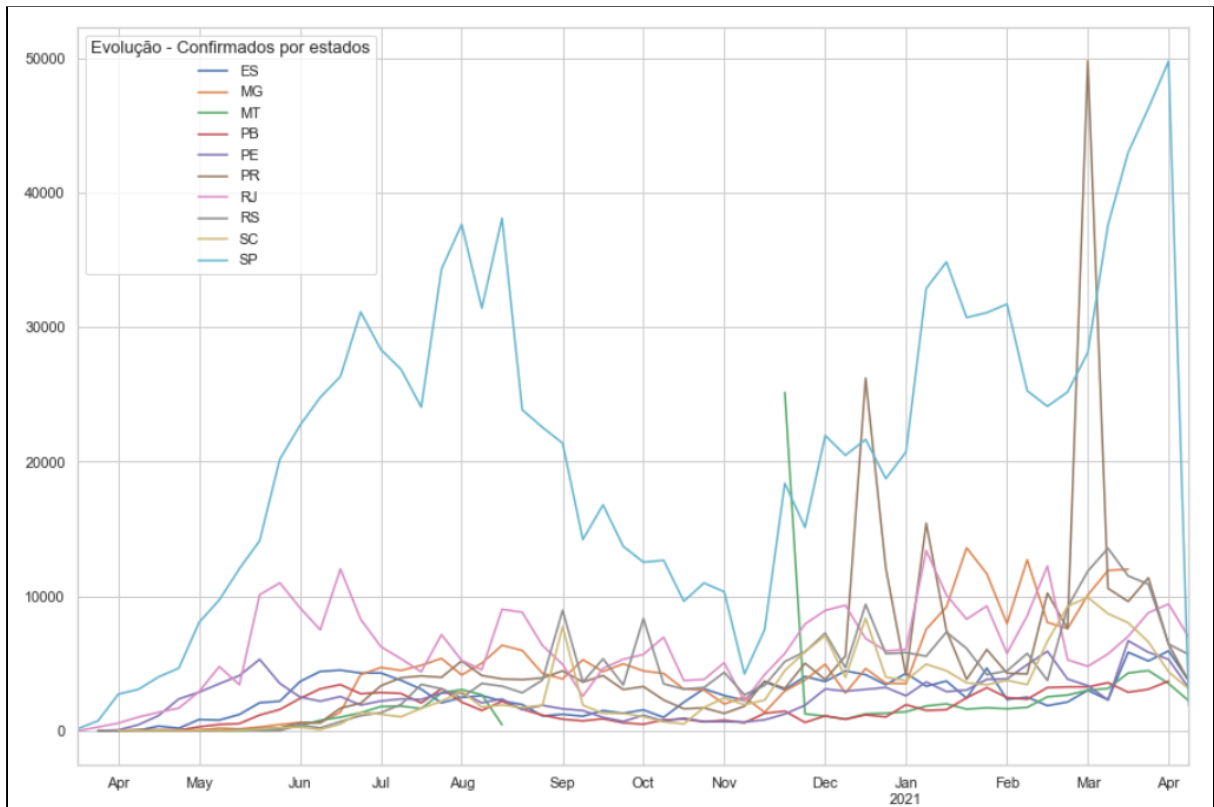
Figura 3 - Número de municípios analisados por estado



Fonte: Autor

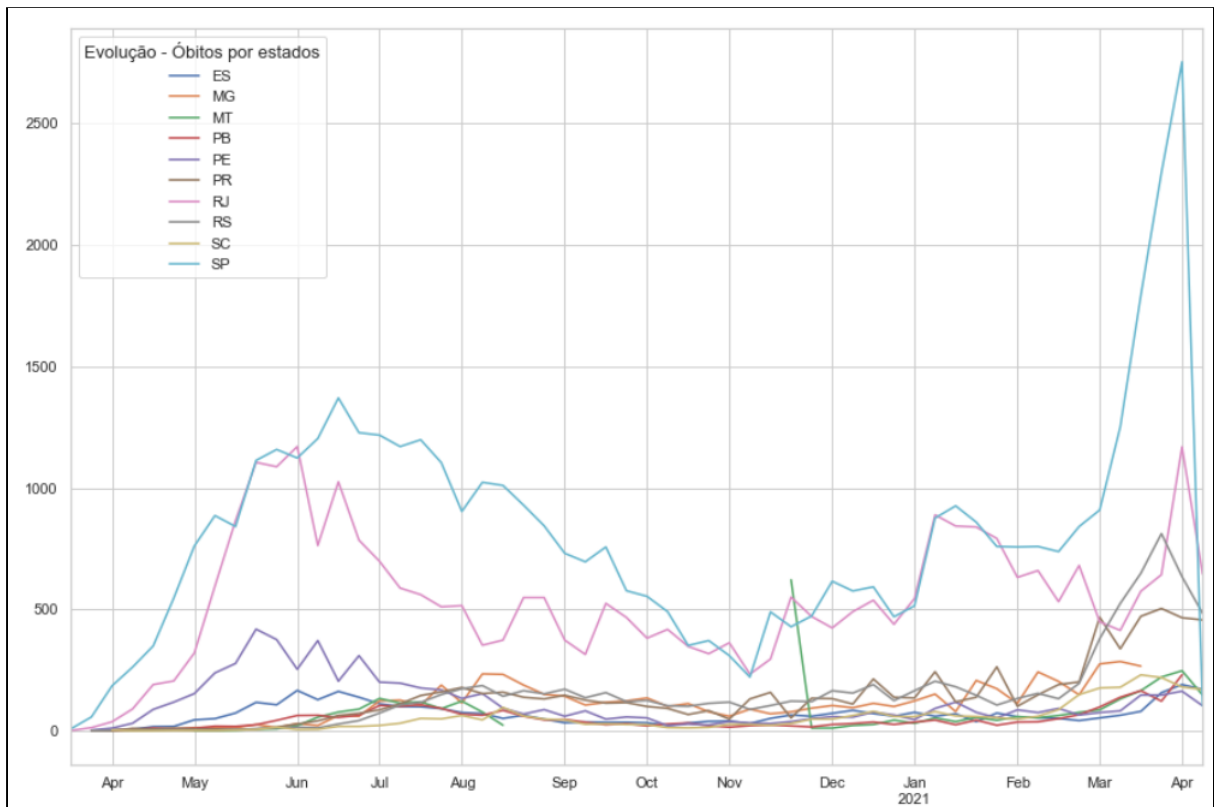
Na Figura 3 é apresentada a distribuição do número de 100 municípios analisados por estado brasileiro, estes municípios têm o maior número de óbitos acumulados conforme os boletins da COVID-19.

Figura 4 - Evolução de casos confirmados por estado



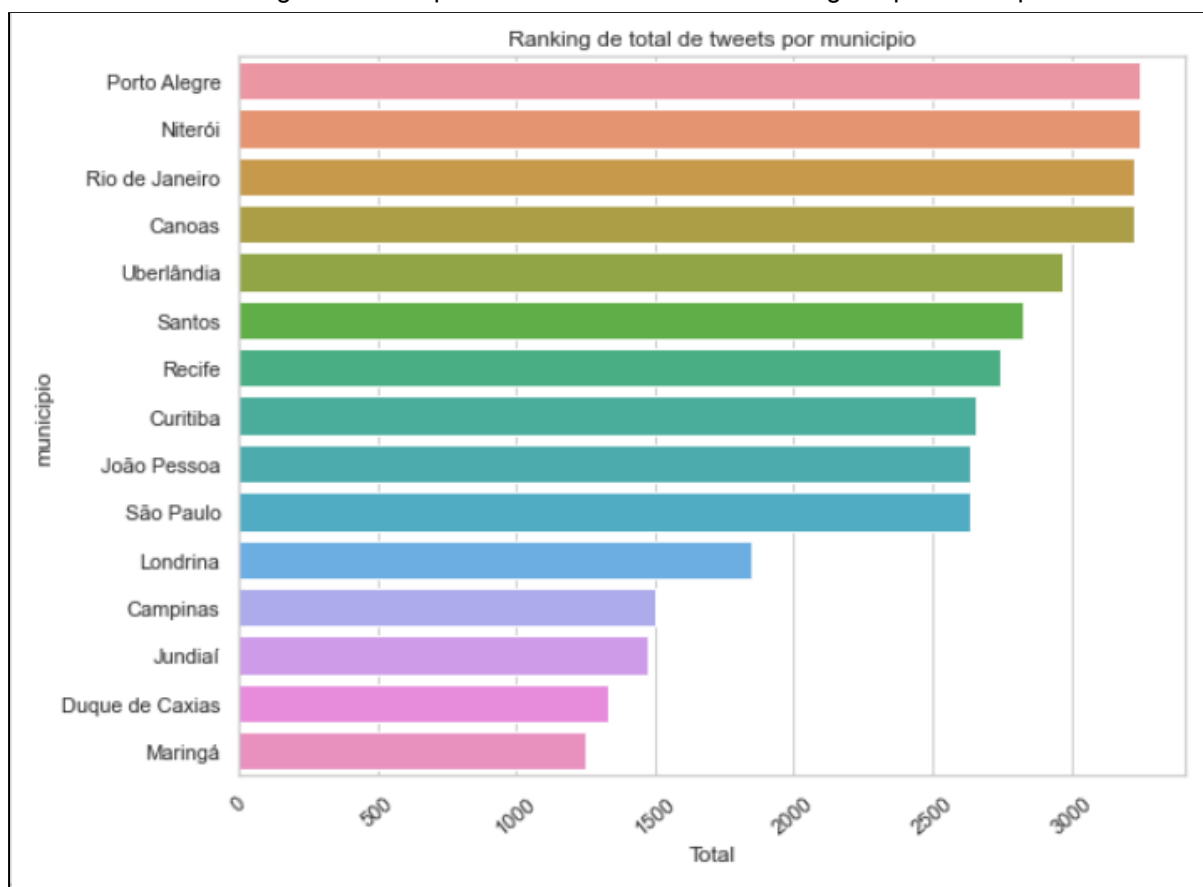
Fonte: Autor

Figura 5 - Número de óbitos por estado



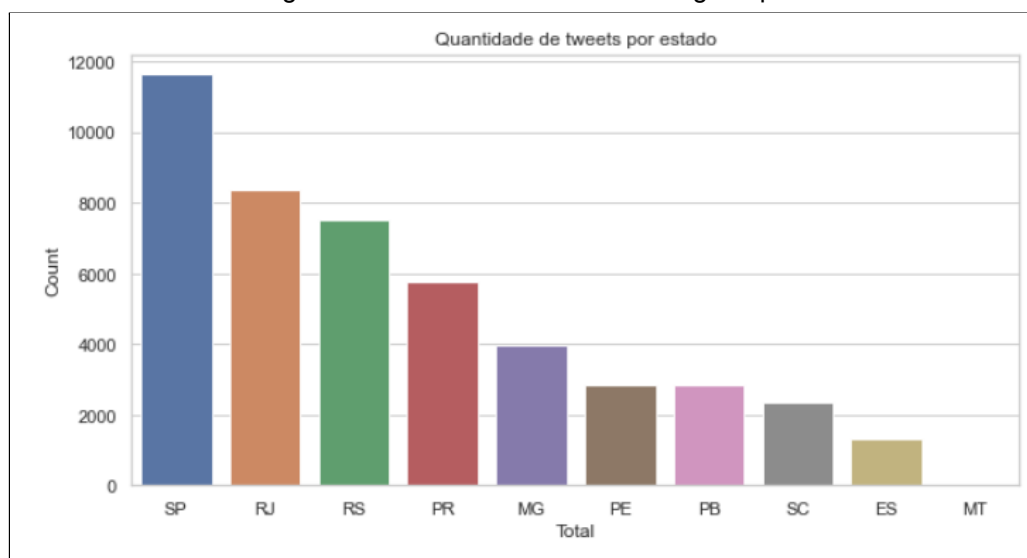
Fonte: Autor

Figura 6 - Ranqueamento do total de tweets em geral por município



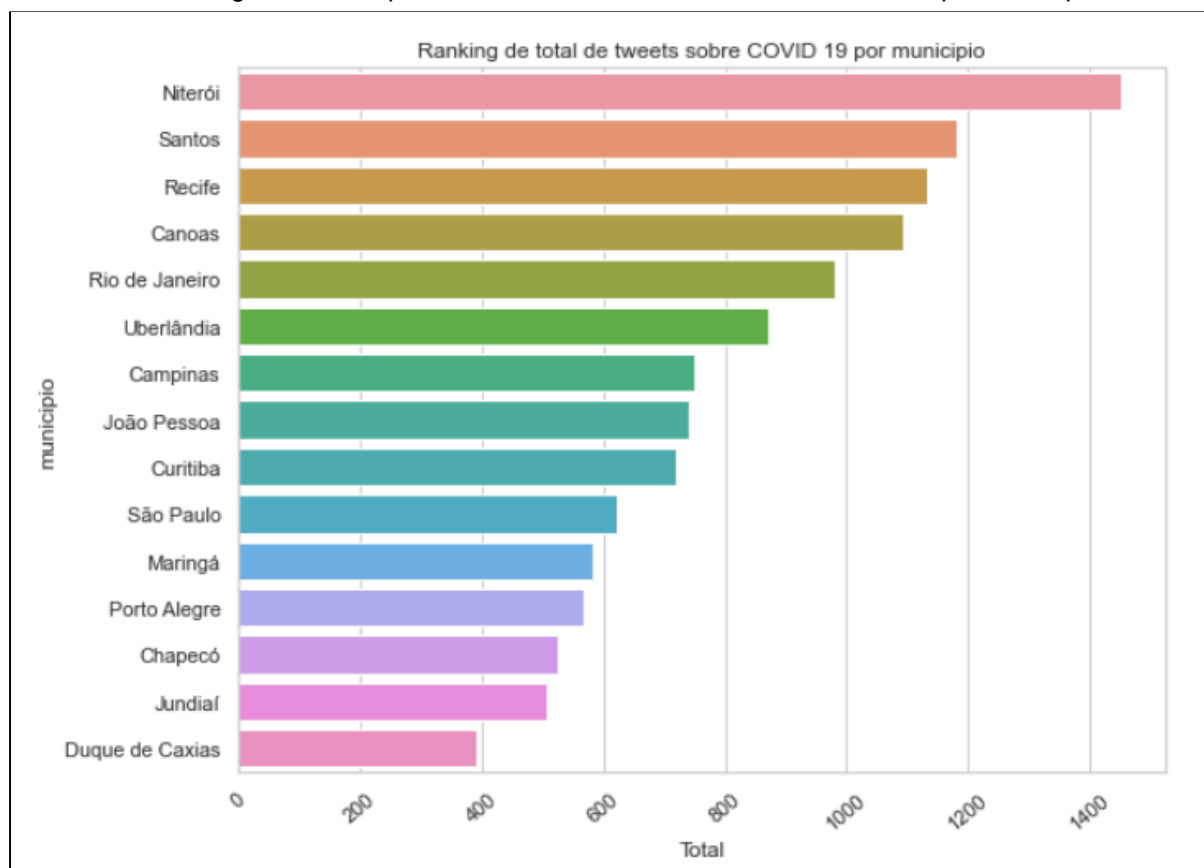
Fonte: Autor

Figura 7 - Quantidade de tweets em geral por estado



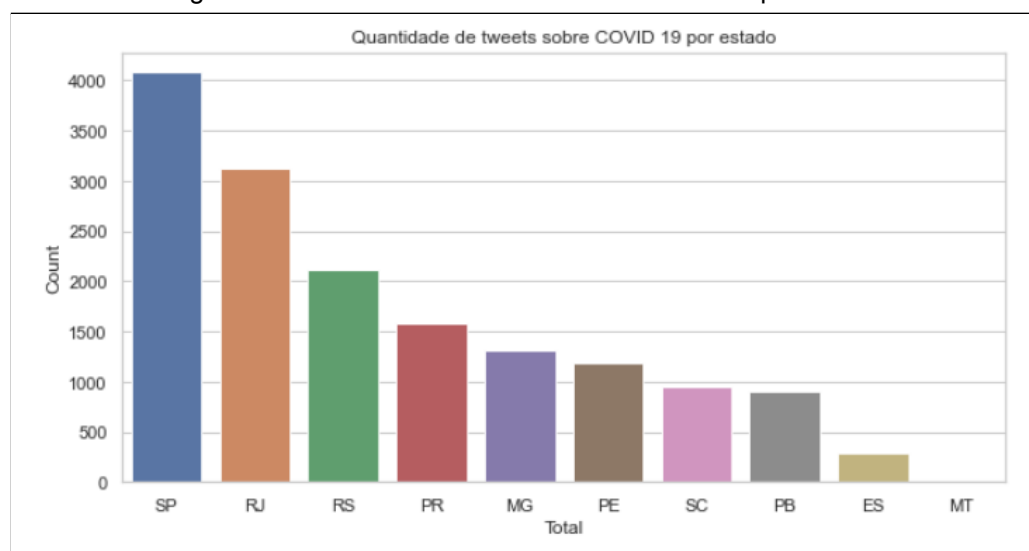
Fonte: Autor

Figura 7 - Ranqueamento do total de tweets sobre COVID-19 por município



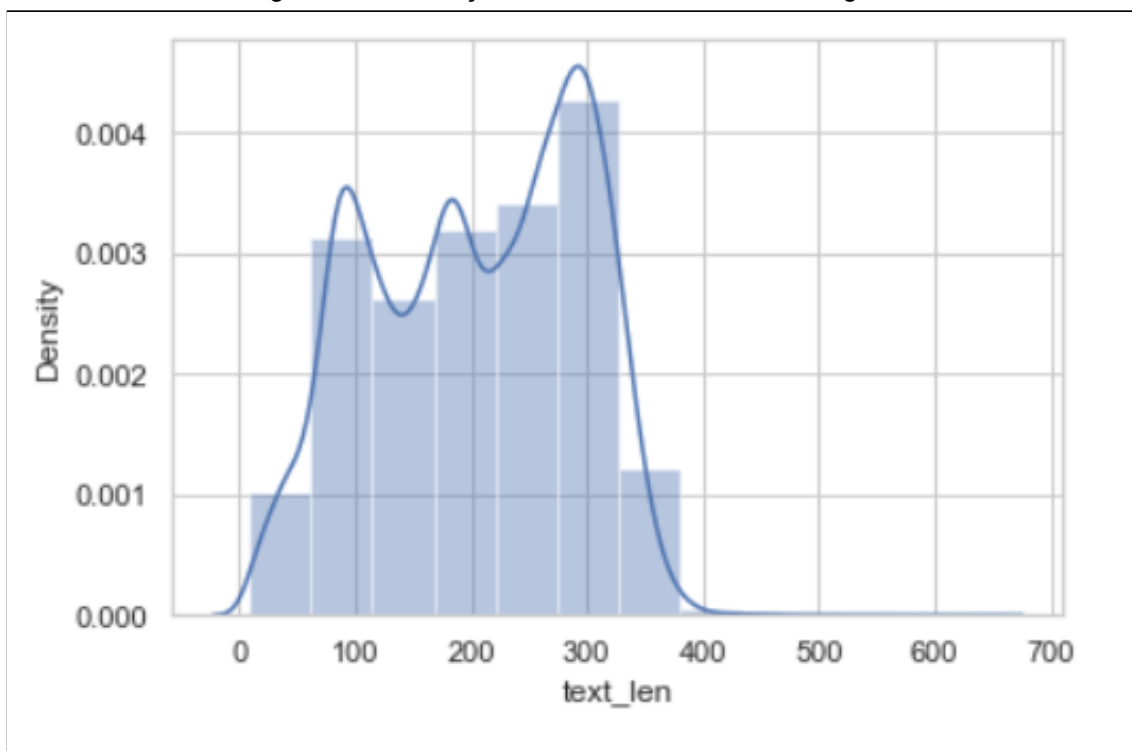
Fonte: Autor

Figura 8 - Quantidade de tweets sobre COVID-19 por estado



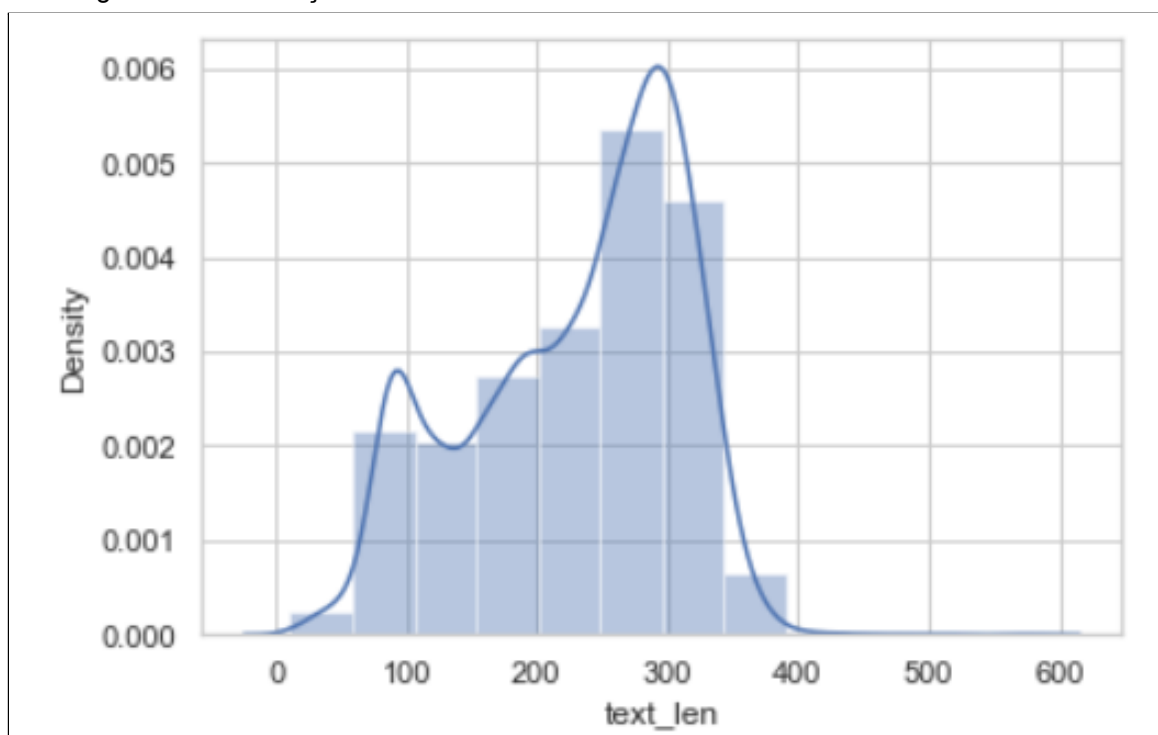
Fonte: Autor

Figura 9 - Distribuição do tamanho dos tweets em geral



Fonte: Autor

Figura 10 - Distribuição do tamanho dos tweets de termos relacionados com COVID-19



Fonte: Autor

The chart displays the frequency of 'created_at' records over a period from March to April. The y-axis represents the count, ranging from 0 to 2500 in increments of 500. The x-axis shows months from Mar to Apr. The data shows a relatively stable period from March to November with counts between 100 and 700. A sharp increase begins in December, peaking at approximately 2700 in late February/early March, before dropping to around 2000 and then rising again towards the end of April.

Month	Approximate Count
Mar	300
Apr	400
May	500
Jun	500
Jul	600
Aug	500
Sep	100
Oct	100
Nov	100
Dec	1000
Jan 2021	900
Feb	1800
Mar	2700
Apr	2000

The chart displays the daily count of new COVID-19 cases in the United States. The x-axis, labeled 'created_at', spans from March 2020 to April 2021. The y-axis represents the number of cases, ranging from 0 to 800. The data shows a major peak in early April 2020, followed by a period of relative stability with minor fluctuations. A significant resurgence begins in January 2021, leading to a second, higher peak in early April 2021.

Date (created_at)	Number of New Cases
Mar 2020	~20
Apr 2020	~360
May 2020	~220
Jun 2020	~240
Jul 2020	~280
Aug 2020	~200
Sep 2020	~80
Oct 2020	~70
Nov 2020	~50
Dec 2020	~230
Jan 2021	~150
Feb 2021	~520
Mar 2021	~650
Apr 2021	~880

[illegible]

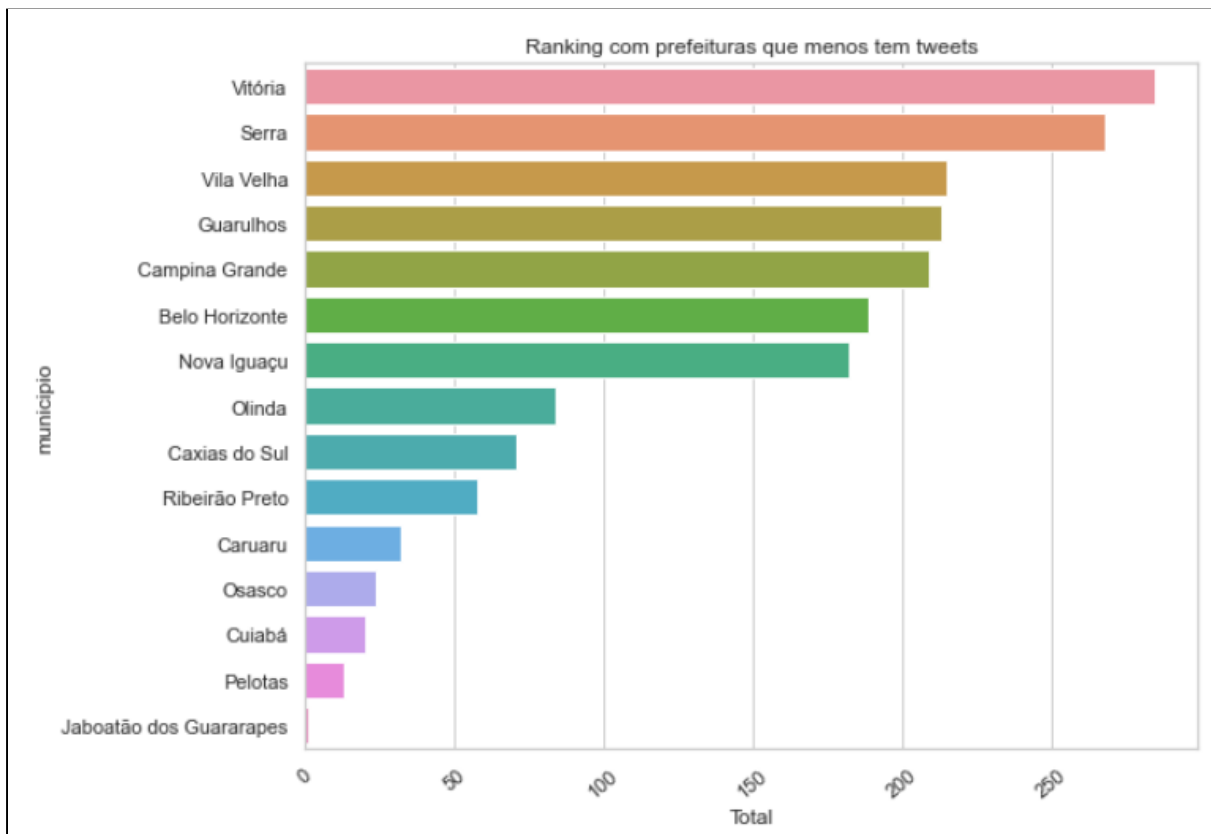
Fonte: Autor

Figura 14 - Nuvens dos termos sobre a COVID-19 mais utilizados nos Tweets por UF - 2



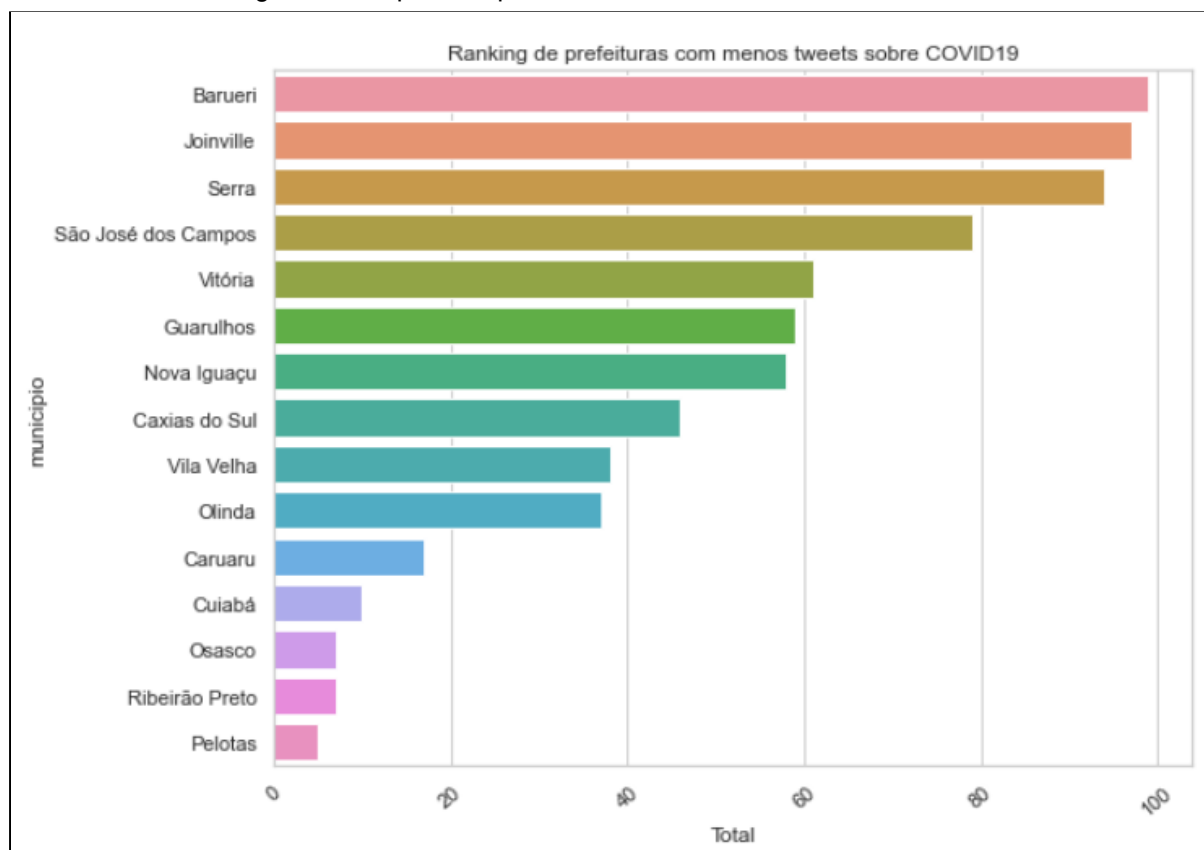
Fonte: Autor

Figura 15 - Top municípios com menos tweets em geral



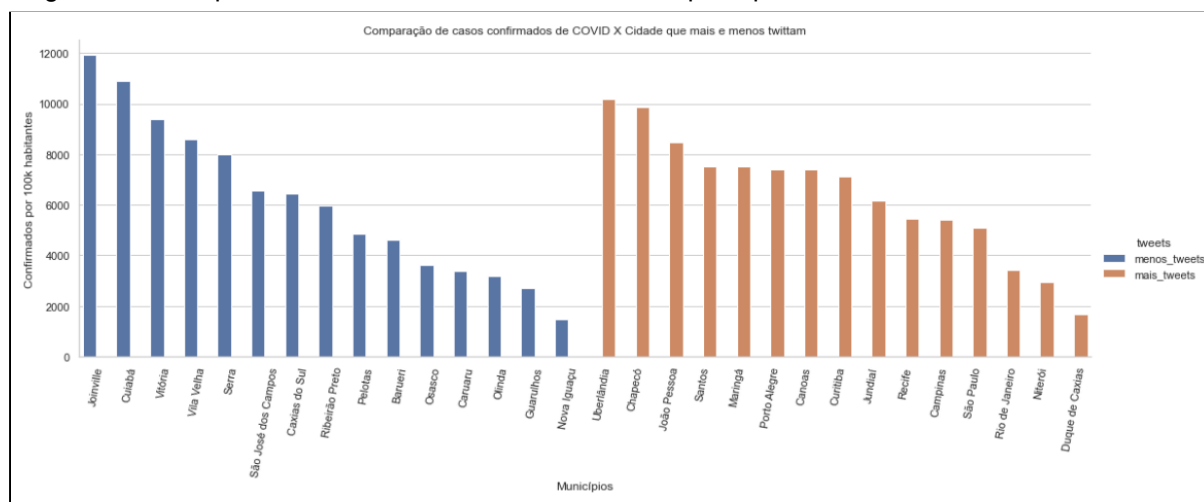
Fonte: Autor

Figura 16 - Top municípios com menos tweets sobre COVID-19



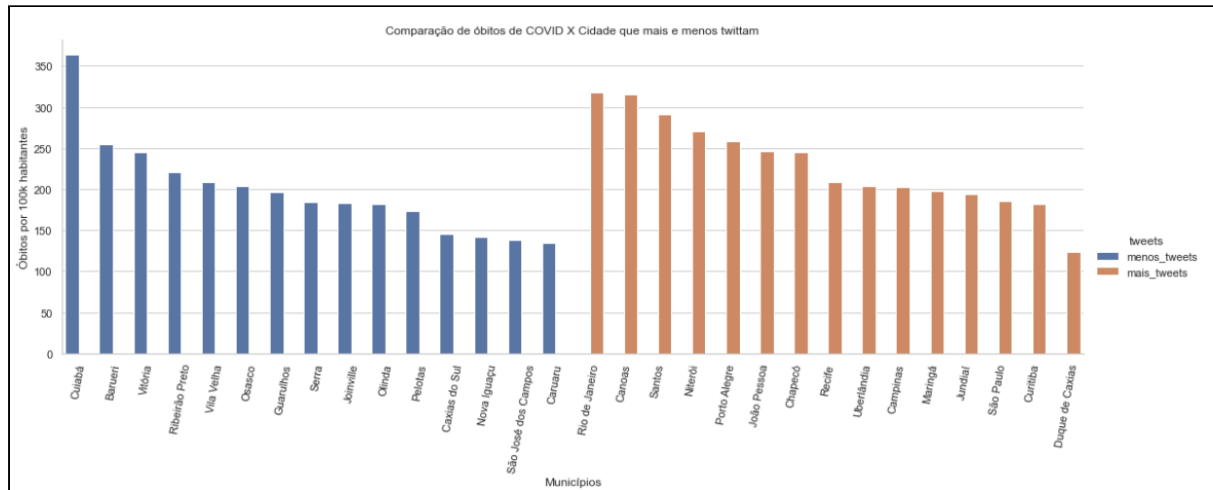
Fonte: Autor

Figura 17 - Comparativo de casos confirmados X municípios que mais/menos tweets de COVID-19



Fonte: Autor

Figura 18 - Comparativo de óbitos X municípios que mais e menos twittam sobre COVID-19



Fonte: Autor

- **Perguntas de pesquisa e explorações iniciais:**

- 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 de COVID-19?

- **Discussão e próximos passos:**

Conforme descobertas realizadas na análise exploratória, os dados são completos e adequados para responder as perguntas elencadas de pesquisa atendendo o tema proposto e permitindo o aprofundamento.

Os próximos passos são:

- Implementação de 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 de estados brasileiros além de capitais.