Brazilian discussion about COVID-19 lockdown policies on Twitter

Fernando Xavier[1], Gustavo Rick Amaral[2], Antonio Mauro Saraiva[1]

[1] Polytechnic School – University of São Paulo (USP)
Av. Prof. Luciano Gualberto, 380 – 05508-010 – São Paulo – SP – Brazil

[2] Faculty of Exact Sciences and Technology – Pontifical Catholic University of São Paulo (PUC-SP)
Rua Monte Alegre, 984 – 05014-901 – São Paulo – SP – Brazil

{fxavier,saraiva}@usp.br, gustrick@gmail.com

Abstract. *The COVID-19 pandemic affected all countries worldwide, causing big changes in people's routines due to public policies for disease spreading control. Among the most impacting measures were social distancing policies and lockdown, leading to an intense discussion by the population. To describe this discussion in Brazil, this research applied data science and natural language methods to analyze posts on Twitter. It processed more than 12.9 million tweets between 2020 and 2021, and the results highlighted the main topics discussed by Brazilian Twitter users, such as the ideological-political component. The approach employed in this research proved to help extract valuable information in massive data mass.*

1. Introduction

In late 2019, the world noticed the initial outbreak of a viral infectious disease in China named COVID-19. This disease, caused by the coronavirus SARS-COV-2, rapidly spread around the world and, in March 2020, was declared a pandemic by the World Health Organization (WHO) [Cucinotta and Vanelli 2020]. Until this time, there are 548,792,366 million cases reported, causing 6,338,435 deaths globally [Johns Hopkins University 2022].

In the early stages of the pandemic, while a vaccine was not developed and the researchers did not sufficiently know the disease, the first policies used to control the disease spreading were related to containing the direct transmission between people. Because COVID-19 had similarities with other coronavirus diseases [Cevik et al. 2021], the main initial measures adopted were mask obligatory use, social distancing, limiting people's movement and orientation on disinfection of surfaces, and body cleaning. Due to its high impact on people's routines, measures to restrict the movement were the theme of several discussions. The most restrictive level policy (lockdown) was unpopular and provoked many protests on streets around the world.

People also used social media platforms to express their opinions about lockdown policies. For public health decision-making by health authorities, the information from social media can help assess and measure the impact of the policies on popular opinion and improve communication strategies. Moreover, extracting information from these data can support campaigns against misinformation, which is a big problem for public health due to its risks to population well-being.

This research aims to analyze the Brazilian's opinions on Twitter regarding lockdown policies during the COVID-19 pandemic. More than 12.9 million tweets citing lockdown policies were collected for this research. Text analysis was conducted to analyze discourse about lockdown policies and extract the main topics considering specific subjects and temporal aspects.

It is expected that this work can contribute to identifying the main topics of discussion of the Brazilian population on one of the most important measures to combat infectious diseases. In addition, it is expected that health management departments in decision-making processes can incorporate the method adopted in this research.

This article is organized as follows: the next section contains an analysis of research related to text mining in the health area, while the third section describes the five-steps method adopted. The results of each step are shown and discussed in Section Four, and the final considerations, including main findings, limitations, and future research, are described in Section Five.

2. Context

The COVID-19 pandemic caused unprecedented changes in people's lifestyle to control the disease dissemination. While vaccines were being developed, several countries defined social distancing policies at different levels. India adopted a national lockdown between March and April 2020 [Mahato et al. 2020], while the first COVID-19 epicenter, Wuhan (China), also adopted a lockdown at the outbreak beginning [Lau et al. 2020]. Italy was the first COVID-19 epicenter in Europe, imposing a lockdown in March 2020; findings suggest that these measures successfully controlled the disease transmission [Guzzetta et al. 2020].

In Brazil, there was not a national lockdown policy, but some locations adopted similar measures to promote some level of social distancing. For example, the São Paulo and Rio de Janeiro states ordered a partial lockdown, mainly in the pandemic beginning, with schools and universities closed, no public events, and restrictions on some economic activities [Nakada and Urban 2020][Dantas et al. 2020]. Besides slowing down the disease spread, these locations experienced other benefits like improving air quality.

Despite benefits and the need to control de COVID-19 spread, the isolation policies were not unanimous worldwide. These policies impacted the economy of the whole world. A study about India's economy estimates a possible loss of around 10–31% of its Gross Domestic Product due to the pandemic [Kanitkar 2020]. There were protests against lockdown in many locations worldwide, such as the United States, Serbia, and Germany [Jørgensen et al. 2020].

Although many people went to the streets to express their opinions about the lockdown and other isolation policies, social media platforms were heavily used to discuss COVID-19, including lockdown measures. A study aimed to analyze the Indians' feelings on Twitter during the government's second and third national lockdown. They found changes from a positive view in the second to a negative one in the third lockdown [Chehal et al. 2021].

3. Method adopted

This research adopted a data science cycle, with well-defined activities from the study planning to text analysis, as shown in Figure 1. The first activity consists of defining the research objectives and which analysis will be necessary to answer the research questions. However, other analyses can be specified during the study conduction according to the preliminary results obtained.

In the first step of this cycle, the study is planned, including the necessary data sources, objectives, methods, and what products must be generated. Products can be many forms of data visualization, such as charts, maps, or tables. In natural language processing projects, the generation of cloud tags summarizes the main topics in a set of documents. After data collection, an exploratory data analysis can be conducted to extract information about the dataset, such as stats related to the records. This step can help identify requirements for the next phases, such as preprocessing techniques and the computing infrastructure needed.

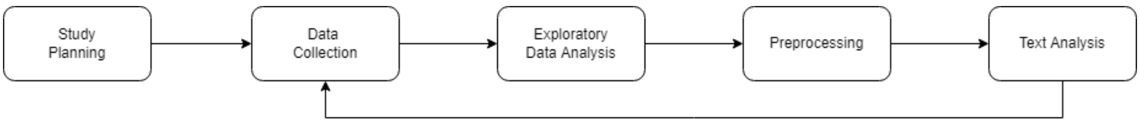

Figure 1. Research steps

Once the data sources are defined, the next step is to collect the data, sometimes in an automated way, such as a script. When the data sources are social media platforms, it is common to collect data using an application programming interface (API) or web scraping. Although not mandatory, this data can be stored in plain files or a database for further analysis. It is important to note that different studies can use the same database and keeping data can be a better project decision to guarantee research traceability. Another factor in favor of the data storage strategy is that some API platforms can return different results if the request is made in different periods.

After data collection, the next step contains activities to prepare data for the analysis step. Depending on the research objective, this preprocessing step includes tasks to select necessary features, remove noise and define strategies for the missing data problem. Considering data related to social media, it can be necessary to remove certain kinds of data, such as images or links. However, not all this data can be classified as noise because this is highly dependent on the research objectives. Even in natural language processing projects, some of this data can represent valuable information for analysis. An example is emojis, a class of images that can be used to define a mood about a text, especially in social media, where informal language and text length platforms' limitations are very common.

Different methods, such as traditional statistical analysis and machine learning, can be applied for the text analysis step. The choice of methods depends on many factors, such as problem type and data characteristics. It is also common to use combined methods, using the best of each. In text analysis research, it is common to use natural language processing combined with machine learning algorithms. This research adopted three strategies: analysis using the most common hashtags, bigrams extraction, and topic modeling. For all strategies, results were compared over time.

4. Results and discussion

4.1. Step 1: Study planning

The main objective of this research is to analyze the Brazilian' discussion about lockdown policies on the social media platform Twitter. To develop this study, the main objective was divided into two research questions:

- Question 1: What are the main general topics around the lockdown discussion?

- Question 2: Are there temporal variations in the discussion topics?

Because the two research questions need the same data source, the activities in the data collection step were executed once. Python and Linux scripts were used for this step to collect and select the required data.

4.2. Step 2: Data Collection

Python scripts were developed to collect data through Twitter's API using search keywords related to COVID-19 between April 2020 and May 2021. Because the objective is only to analyze opinions in the Portuguese language and some words are also used in other languages, the language parameter was included in the API requests. Table 1 contains the original keywords used in API requests and their English translation.

Table 1. Keywords used to search tweets

Type	Original Keyword	Keyword in English
COVID-19 names	Covid coronavirus corona	Covid coronavirus corona
Policies	Lockdown isolamento quarentena	Lockdown isolation quarantine

In addition to the tweet text, Twitter's API returns all information about a tweet, such as date, language, and user information. For this research, it was decided not to collect retweets without quotes, and only two fields (data of tweet and text) were stored in CSV files for further analysis.

Table 2. Keywords used to search tweets

Original Keyword	Keyword in English
lockdown	lockdown
quarentena	quarantine
isolamento	isolation
isolada	isolated
isolado	isolated

The last activity in the data collection step was the selection of only tweets related to the lockdown policies. This step was necessary because some search keywords do not necessarily return tweets about the lockdown but other COVID-19 subjects. Then, a Linux script was used to search only lines containing the words related to the lockdown policies in each CSV file, as described in Table 2.

4.3. Step 3: Exploratory data analysis

Using parameters described previously, 12,497,029 tweets were selected between April 2020 and April 2021. As shown in Figure 2, it is important to note that the discussion around isolation policies decreased over time. This fact can be related to many reasons. At the pandemic's beginning, there was little information about COVID-19 because the research about it was relatively new. Moreover, the unique measures known were those related to controlling the dissemination similar to other coronavirus diseases, such as isolation policies, mask-wearing, and cleaning.

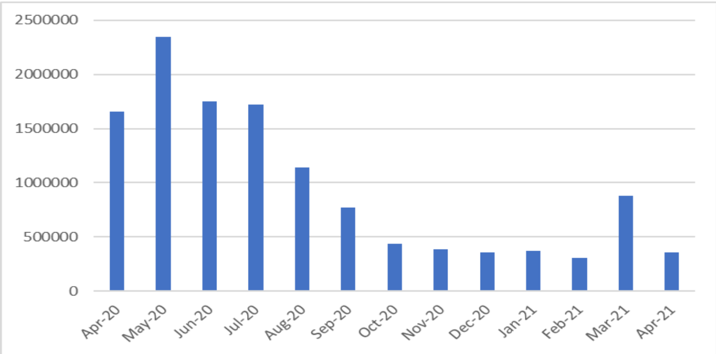

Figure 2. Quantity of tweets related to lockdown policies

Although there was no national isolation policy, many Brazilian cities adopted measures to control the people's movement. However, due to the economic impact of these measures, many cities started to ease the movement restrictions when the total number of cases decreased in Brazil. With cases rising after the Christmas season, new isolation measures were adopted in many locations in Brazil. The Sao Paulo state, for example, decreed a set of restrictions on movement on March 2021. This may be a reason for the rise in this period, as shown in Figure 2.

Additionally, the top 10 hashtags used in all the periods were selected. As described in Table 3, the main hashtags were related to the pandemic and entertainment. At the beginning of the pandemic in Brazil, a large part of the population practiced isolation measures, and promoting online events, such as artist lives, was common. Although it did not appear among the top 10 hashtags, the political aspect was very present. Among the top 100 hashtags used in the period, there were 21,439 mentions of the president of Brazil in tweets related to isolation policies.

4.4. Step 4: Preprocessing

As the objective was to analyze only textual components, other elements of the tweets were removed, such as links, emojis, and numbers. As tweets may contain mentions of other users, these elements were also removed from the text. For some of the analyzes described below, part-of-speech tagging was performed to identify and select the words needed for each analysis.

Table 3. Main hashtags between April 2020 and April 2021

Hashtag	Count
quarentena	102575
livelocalmariliamendonca	53406
covid	51527
lockdown	25581
coronavirus	24096
bbb	22074
fiqueemcasa	19259
melim	16700
ficaemcasa	13561
pandemia	13109

4.5. Step 5: Text Analysis

In this step, some techniques were adopted to describe the content of the tweets, such as bigram extraction, hashtags, and topic modeling. To analyze the complete content, tweets from the entire period were selected. However, given the dynamics of the pandemic in Brazil, specific periods (semesters or months) were selected for temporal evaluation of the discussion on social isolation measures.

4.5.1 Hashtags

To identify the main topics and to compare possible temporal changes in the discussion, the most frequent hashtags were identified for three months: April and October 2020 and April 2021. Table 4 lists the top 5 hashtags by month, and how each semester differs from others in pandemic history is clear.

Some aspects can be observed by analyzing the top 5 listed in Table 4. First, the list of the five most cited hashtags contains three groups of subjects: pandemic, quarantine, and entertainment. In all three months, there were mentions of Brazilian TV shows in tweets about the lockdown. Especially in the first months of the pandemic, online events were common, and many promoted isolation policies with the hashtag #ficaemcasa (#stayathome). In addition, an increase in mentions of the word lockdown can be observed, which did not appear in the top 5 of the first month and started to appear more in the following months.

The first months of 2020 were the pandemic beginning, which can explain why most top 5 hashtags were related to the coronavirus disease or the lockdown subject. The hashtag *quarantine* (quarentena in Portuguese) had almost three times more mentions than the *covid* hashtag when considering only tweets with words used as filters. The other two hashtags were related to the campaign 'stay-at-home' promoted by governments, artists, and other influencers.

The data about the three periods can illustrate, in some way, how the lockdown measures were adopted in Brazil during the pandemic. While many locations adopted isolation measures in the pandemic beginning, the lower cases and pressure made by economic sectors left many governments to ease the lockdown measures in the second semester of 2020. With the growth in cases in the first months of 2021, some locations returned to lockdown measures adoption.

Table 4. Main hashtags per period

Apr-2020	Count	Oct-2020	Count	Apr-2021	Count
#quarentena	20314	#quarentena	2639	#bbb	2564
#covid	7980	#afazenda	2532	#lockdown	2440
#bbb	7135	#covid	1934	#covid	2085
#coronavirus	4830	#lockdown	847	#quarentena	1709
#ficaemcasa	4182	#coronavirus	808	#coronavirus	730

4.5.2 Bigrams

Sometimes, the frequency of words alone may not describe a discussion on an issue enough. Thus, using bigrams or trigrams can be helpful in identifying the context in which a given word is used. Then, using the package NLKT, the main bigrams per semester were extracted, and the five most frequent were presented in Table 5.

In the general analysis of the bigrams for the word "isolation" ("isolamento", in portuguese) and "quarantine" ("quarentena", in portuguese), it was perceived that the occurrences of the expression "this isolation" and "this quarantine" – used to refer to isolation or quarantine as a specific event in time – diminished immensely over time. In the first semester of 2020, there were 5,374 occurrences of the bigram "this isolation" and 387,643 occurrences of the bigram "this quarantine". In the second semester of the same year, the occurrences of "this quarantine" dropped to 287,344 (diminished by 25,87%) and the occurrences of "this isolation" to 4,939 (decreased by 8,09%). In the first semester of 2021, there were 34,457 occurrences of "this quarantine" (diminished by 88,01%) and 1,926 of "this isolation" (reduced by 61%). In the whole period (2020.1 to 2021.1), the occurrences of "this quarantine" diminished by 91,11% and "this isolation" by 64,11%.

Table 5. Main bigrams per semester

1s2020	Count	2s2020	Count	1s2021	Count
Nessa quarentena	387643	Nessa quarentena	287344	Isolamento social	65942
Quarentena acabar	149665	Quarentena pra	110285	Todo mundo	39721
Quarentena pra	128913	Todo mundo	99801	Nessa quarentena	34457
Isolamento social	127233	Isolamento social	94645	Contra lockdown	24846
Essa quarentena	113562	Antes quarentena	76115	Ano passado	22938

The interpretation is that these numbers represent an important change in the discourses and debates on Twitter about the restrictive measures for Covid-19 in Brazil. The heated political debate about this topic did not seem to cease for the whole period. So, the hypothesis is that, over the three semesters, a terminological substitution took place. The center of gravity of the discussions has shifted. At the end of 2020, when Brazil was heading to the "third covid wave", the debate began to gravitate around a new term: "lockdown". The analysis of the bigrams for the word "lockdown" indicates the increase in the occurrences of the bigrams that refer to "lockdown" as an event or that refer to the position concerning the topic (the restrictive measures). For example, the expression "[to] make [the] lockdown" ("fazer [o] lockdown", in portuguese) ou "against [the] lockdown" ("contra [o] lockdown", in portuguese) has increased from 3,654 occurrences in 2020.2 to 16,012 in 2021.1 in the first case and from 3,182 to 24,846 in the second case. It was an increment of 338,20% in the occurrences of the expression "[to] make [the] lockdown" and an increment of 680,83% in the occurrences of the expression "against [the] lockdown".

So, the hypothesis is that, over the three semesters, the more generic words, "isolation" and "quarantine", seem to have been gradually replaced by the more technical term, "lockdown". However, this term was only superficially technical. The debate about the types of isolation (the variations and gradation of the restrictive measures) was reproduced with the new term (although the use of the word "lockdown" to refer to "types of isolation" is inconsistent with its technical definition).

The analysis of the bigrams for the word "isolation" indicates that, in the first semester of 2020, this term was used to discuss types of isolation. In this period, the bigram "horizontal isolation" ("isolamento horizontal") occurred 3,891 times, "total isolation" ("isolamento total") occurred 5,469 times, and vertical isolation" ("isolamento vertical") occurred 7,204 times. These bigrams make up 5,84% of the total of isolation bigrams. In the second semester of 2020, the occurrence of this "type of isolation" category dropped to 1,27%. The "vertical isolation" bigram, for example, occurred only 2,967 times.

In comparison with the isolation bigrams, the analysis for the word "lockdown" indicates that, in the first semester of 2021, there were more "types of lockdown" being discussed: "total lockdown" ("lockdown total"), 7,193 times; "real lockdown" (lockdown [de] verdade), 5,068 times; "full lockdown" (pleno lockdown), 2,277 times; "severe lockdown" ("lockdown severo"), 2,077; "general lockdown" ("lockdown geral"), 1,362 times; "rigorous lockdown" ("lockdown rigoroso"), 1,182 times.

4.5.3 Topic Modelling

Since subjects can vary significantly over a semester, three specific months were selected for topic modeling: April and October 2020, in addition to April 2021. The topic modeling method adopted in this study was the Latent Dirichlet Allocation (LDA), a probabilistic model of how a document could be related to the topics and was previously employed in another study with Twitter data [Zou and Song 2016].

After extracting the 50 most important words of each topic in each period, the topics were classified into the following categories: 1) "report about personal matters (in the context of the pandemic)", 2) "general commentaries about the pandemic", 3) "political debate (about the pandemic)" and 4) "entertainment".

The first category – "report about personal matters" – is composed of words related to primary groups (family, friends, work colleagues) and connected to personal activities of daily living (dressing, feeding, mobility, and other associated tasks). In this category, people tweeted about how they have suffered during the pandemic with the loss of socialization, the new routine, etc. For example, people tweeted about the creative ways they found to cut their hair.

While the first category comprises commentaries about personal matters related to the pandemic, the second category is "general commentaries about the pandemic". The second category is composed of words that express feelings and opinions on the pandemic and associated themes.

The political opinions and views were separated into a third category in which the names of politicians appeared more frequently. The main words of the third category refer to political agents and positions as well as to central terms of the political debate like "economy", "death", "masks", "lockdown", "vaccine", "treatment", "isolation", etc.

Commentaries about TV shows, YouTube concerts, and artists compose the fourth category – "entertainment". In April/2021, an interesting phenomenon was observed in the topic categorized under the label "entertainment": political thematization interfered with entertainment themes. Some of the political and ideological discussion on lockdown and other restrictive measures appeared mingled with commentaries about the TV show "Big Brother Brazil" because one of the participants declared political support to one of the main political parties in Brazil and expressed opinions about the restrictive measures adopted by some states' governors.

5. Final Considerations

This research aimed to analyze Brazilian users' discussion about lockdown policies on Twitter. Three studies were conducted considering the whole period, user types, and the temporal aspect.

Regarding research question 1, the main topics were entertainment (TV shows), political discussion, general commentaries about the pandemic, and personal stories. Results show that the debate on these topics varied over time (research question 2), caused by the events at the moment (in the case of entertainment topic) and the pandemic status concerning the advance of vaccination and reduction of cases.

Results suggested that specific analysis can be more helpful than considering the whole period and users' data. The use of information extraction techniques in large volumes of text proved to be very useful in identifying the main topics discussed on Twitter. Public Health departments can adopt these techniques to monitor people's opinions of adopted measures. Furthermore, it is important that the political aspect was linked to the discussions on isolation measures. This may have negatively impacted the acceptance of the measures, and these impacts may be the subject of future studies.

Regarding limitations, it is important to know that Twitter users do not represent the population. This platform's users differ from other social media, such as Instagram and Facebook. Another limitation is that Twitter is less used in Brazil compared to countries like England and the United States. Moreover, it is important to analyze the influences caused by high movements not necessarily related to the subject of interest. This is the case of the increased engagement generated by artists or fans.

Further research can analyze the role of other kinds of data in describing the mood of a subject. For example, extracting meaning from emojis can reveal more information about a post. Moreover, additional user information, such as description, age, and location, can be used to extract more insights.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Cevik, M., Tate, M., Lloyd, O., et al. (2021). SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, v. 2, n. 1, p. e13–e22.

Chehal, D., Gupta, P. and Gulati, P. (2021). COVID-19 pandemic lockdown: An emotional health perspective of Indians on Twitter. *International Journal of Social Psychiatry*, v. 67, n. 1, p. 64–72.

Cucinotta, D. and Vanelli, M. (2020). WHO Declares COVID-19 a Pandemic. *Acta Biomed*, v. 91, p. 157–160.

Dantas, G., Siciliano, B., França, B. B., Da Silva, C. M. and Arbilla, G. (2020). The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil. *Science of the total environment*, v. 729, p. 139085.

Guzzetta, G., Riccardo, F., Marziano, V., et al. (2020). The impact of a nation-wide lockdown on COVID-19 transmissibility in Italy. *arXiv preprint arXiv:2004.12338*,

Johns Hopkins University (2022). COVID-19 Dashboard. https://coronavirus.jhu.edu/map.html, [accessed on Jul 3].

Jørgensen, F. J., Bor, A., Lindholt, M. F. and Petersen, M. B. (2020). Lockdown evaluations during the first wave of the COVID-19 pandemic. *Aarhus: Aarhus University. doi*, v. 10.

Kanitkar, T. (2020). The COVID-19 lockdown in India: Impacts on the economy and the power sector. *Global transitions*, v. 2, p. 150–156.

Lau, H., Khosrawipour, V., Kocbach, P., et al. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *Journal of travel medicine*,

Mahato, S., Pal, S. and Ghosh, K. G. (2020). Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. *Science of The Total Environment*, v. 730, p. 139086.

Nakada, L. Y. K. and Urban, R. C. (2020). COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil. *Science of the Total Environment*, v. 730, p. 139087.

Zou, L. and Song, W. W. (2016). LDA-TM: A two-step approach to Twitter topic data clustering. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. . IEEE. http://ieeexplore.ieee.org/document/7529581/.