Analysis of COVID-19 Pandemic Related Tweets

10867436

10886338

10835879

10760993

5th December 2023

Table of Content

- ▶ Introduction
- ▶ Description of dataset
- ▶ Information about the users
- ▶ Analysis of Tweets Pattern
- ▶ Analysis of Word Usage in Tweets
- ▶ Analysis of Users Sentiment in Tweets
- ▶ Analysis with T-test
- ▶ Limitations of the dataset
- ▶ Conclusion & Recommendations
- ▶ References

Introduction

As the world grappled with the challenges posed by the pandemic, social media platforms became a significant avenues for individuals to share their thoughts, concerns, and experiences.

These tweets were analysed to provides insights into the sentiments expressed during the unprecedented times of the COVID-19 pandemic.
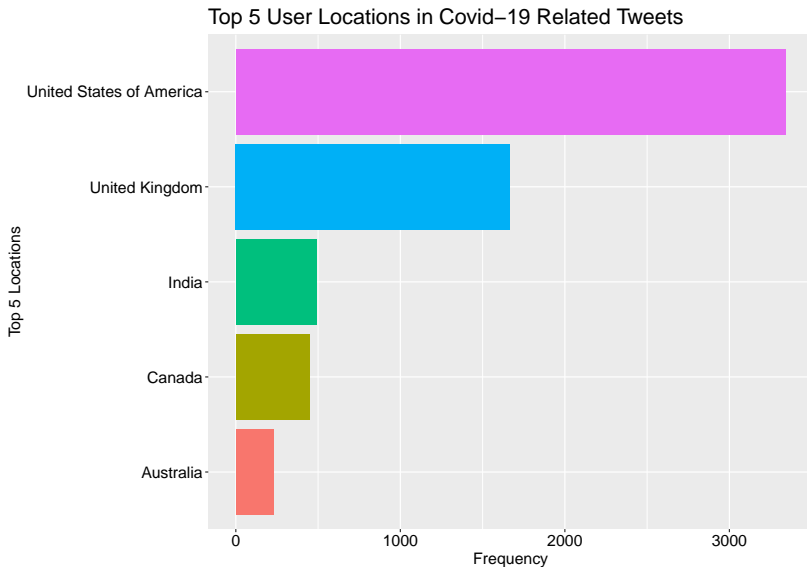
Description of Dataset

The dataset is made up of tweets that were extracted from Twitter now called X.

The dataset contains tweets about the Covid-19 pandemic between the period of March and April 2020 from 41,157 users.
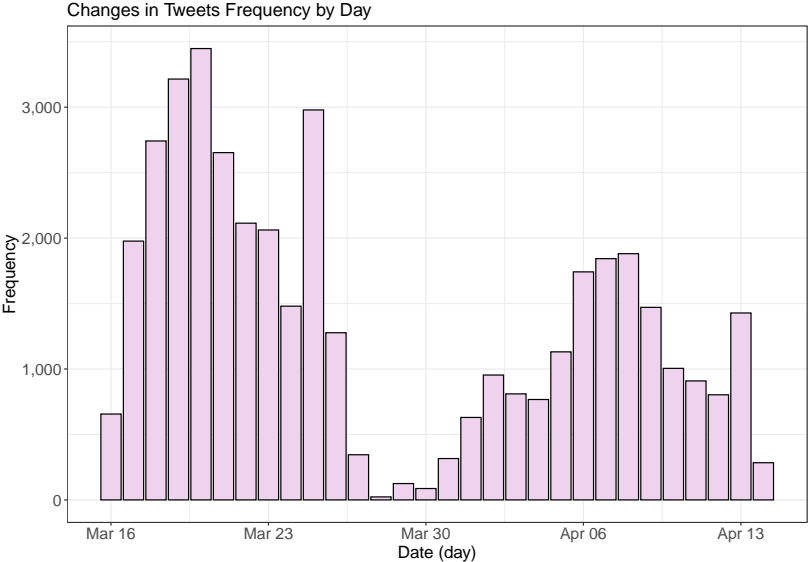
It contains information used to identify the users such as their locations, and the time of tweet and their original tweets.

Information about users

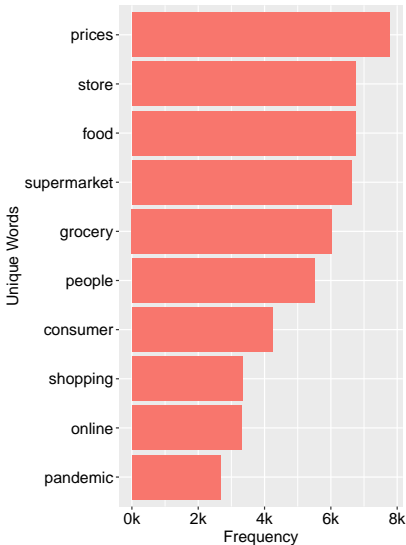A total of 41,157 users from 11,925 locations contributed to tweets in the dataset.

Top 5 User Locations in Covid−19 Related Tweets

Analysis of Tweets Pattern

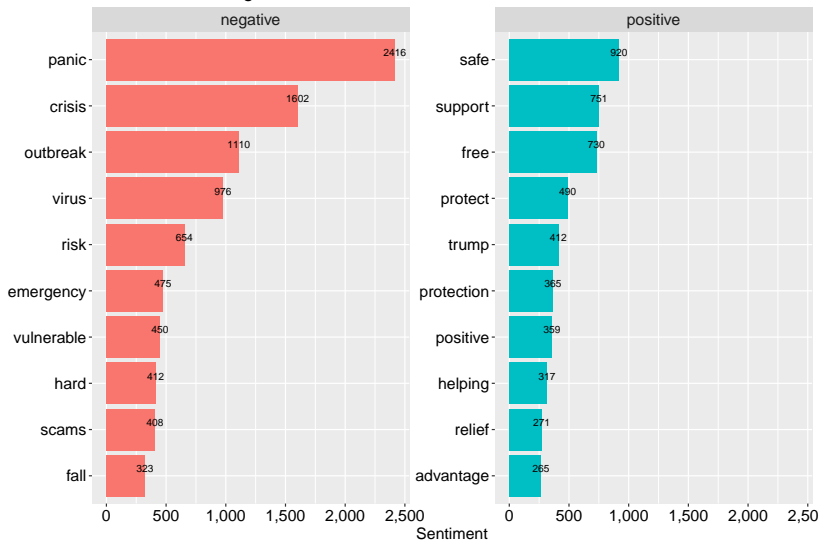Changes in Tweets Frequency by Day

Analysis of Word Usage in Tweets

Most frequently used words in Covid–19 related tweets
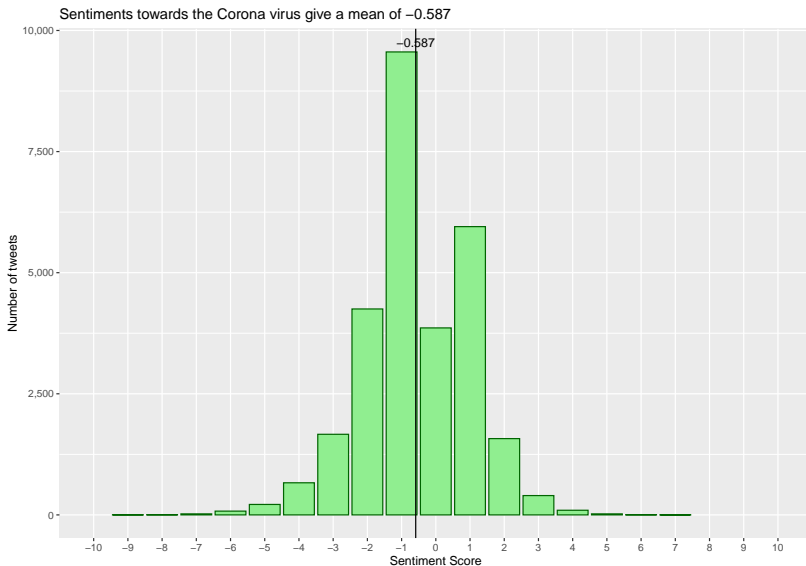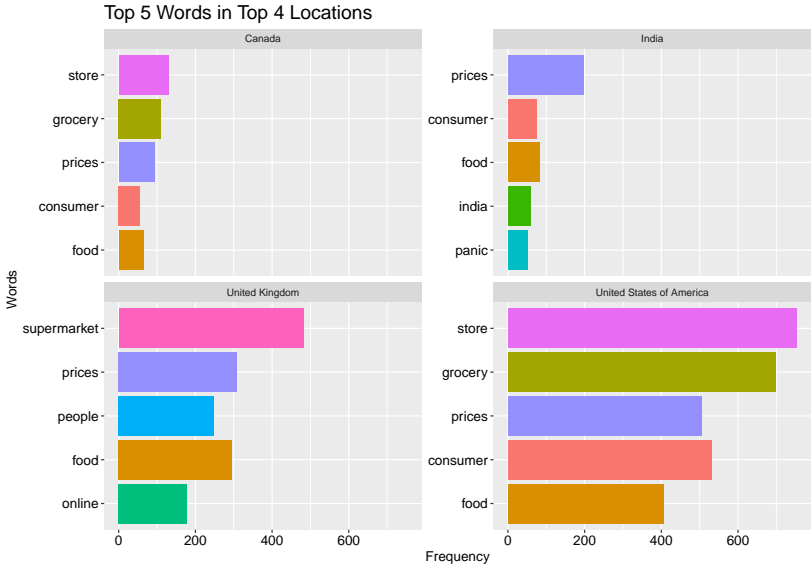
Analysis of Users Sentiment in Tweets

Most common Negative and Positive words in Covid−19 Related Tweets
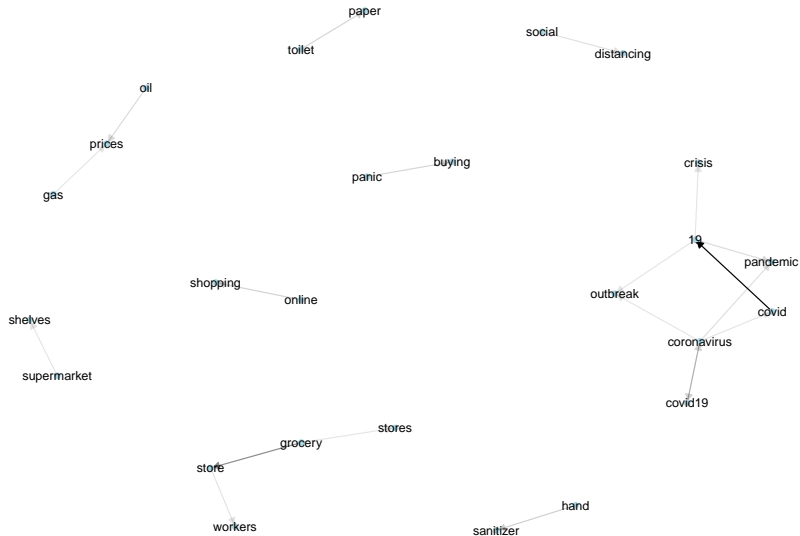
Analysis of Users Sentiment in Tweets (cont.)

Sentiments towards the Corona virus give a mean of −0.587

Analysis of Users Sentiment in Tweets (cont.)

Top 5 Words in Top 4 Locations

Bigram Analysis of Tweets Text

Analysis with T-test

The Independent T-test is a parametric test that is used to find the mean difference in a quantitative variable across a categorical(dichotomous) variable i.e a variable with only two levels or categories. It also follows some other assumptions that data from the two groups must be more than thirty each and the responses must have been randomly selected.

Considering the fact that our data set contains quite a large sample size, there were no quantitative variables in the data set, likewise any categorical(dichotomous) variable with two possible outcome with which we could compare mean differences. Also the data weren't randomly selected and as such a T-test analysis was not applicable in this analysis.

Limitations of the dataset

▶ The dataset was drawn from twitter that does not properly represent the general population.

▶ The temporal breadth of the dataset might not accurately reflect the emotion across a longer time frame, and it might be especially impacted by certain events that occur only in March and April 2020.

Conclusion & Recommendations

During the early phases of the pandemic, people expressed shock, worry, and uncertainty as evidenced in the sentiment analysis score of -0.6. They also expressed concerns about the virus's capacity to spread, the lack of sufficient information, and its possible effects on daily life and health. A dip in number of tweets were noticed between March 28th & 30th 2022; this may have been due to information overload, or distraction by other events which temporarily diverted focus.

In light of these findings, The following recommendations were proposed

- ▶ The government should set up clear and open channels of communication to stop the spread of false information and calm public anxiety.
- ▶ Results showed the often-used negative terms by people were "panic" and "crisis." They should have confidence in the competence of authorities to handle emergencies, and ensure resources are well distributed.

References

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems, 1695. https://igraph.org

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. Journal of Statistical Software, 25(5), 1-54. https://www.jstatsoft.org/article/view/v025i05

Le Pennec E, Slowikowski K (2023). *ggwordcloud: A Word Cloud Geom for 'ggplot2'*. Rpackage version 0.6.1, https://CRAN.R-project.org/package=ggwordcloud

Silge, J., & Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. Journal of Open Source Software, 1(3), 37. https://joss.theoj.org/papers/10.21105/joss.00037

Wickham, H. (2019). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.3.1. https://CRAN.R-project.org/package=tidyverse