# Corona Virus Outbreak

# What do people think about it?

Created By Ong Wei En Team TP-1

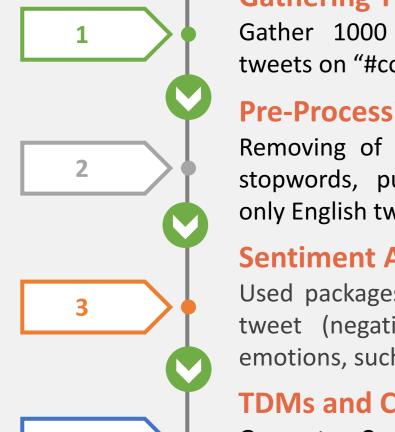
### **INTRODUCTION & OBJECTIVE**

Although half a year has past since the coronavirus first started, more and more new cases are still being discovered everyday. As a result, more and more countries are going under lockdown resulting in the daily lives of people changing, from working from home instead of the office, to studying at home instead of in school. Additionally, this ban has resulted in people being "locked up" at home.

The aim of this project is to use the R language to find out the frequently discussed terms and topics amongst tweeters during this pandemic and how they feel towards it.

### **Approach**

Data for this project were obtained from twitter using the hashtags "#covid" and "#coronavirus". The approach taken is summarised below:



### **Gathering Tweets**

Gather 1000 tweets on "#coronavirus" and 1000 tweets on "#covid"

### **Pre-Processing**

Removing of noisy and irrelevant data such as URLs, stopwords, punctuations, special characters; Keeping only English tweets

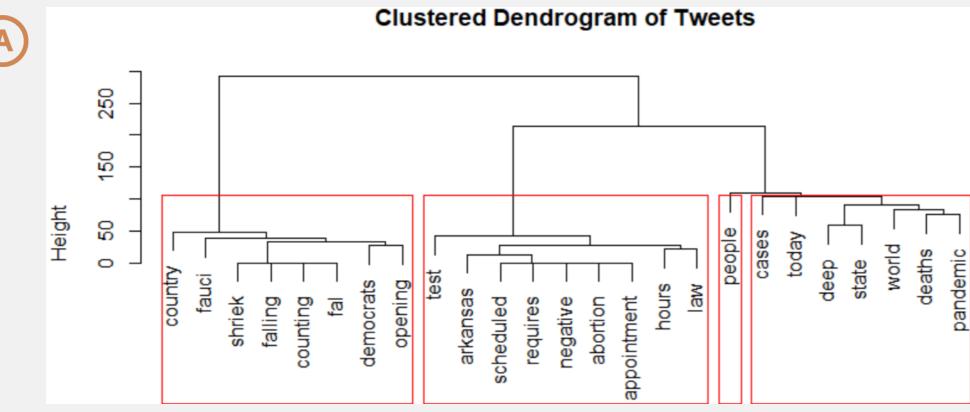
### **Sentiment Analysis**

Used packages, such as the syuzhet package, to score each tweet (negative, positive, neutral) and provide detailed emotions, such as trust, using the lexicon-based approach.

### **TDMs and Corpus and analysis**

Generate Corpus and Term Document Matrix and perform further data cleaning in order to create the word cloud and dendrogram

## **Findings**



"death" and "pandemic" are closely related to each other in the tweets **Observations:** relating to the corona virus (they are equidistant from "world" as they are in the same cluster before joining "world")

People do not like this pandemic as there are many deaths

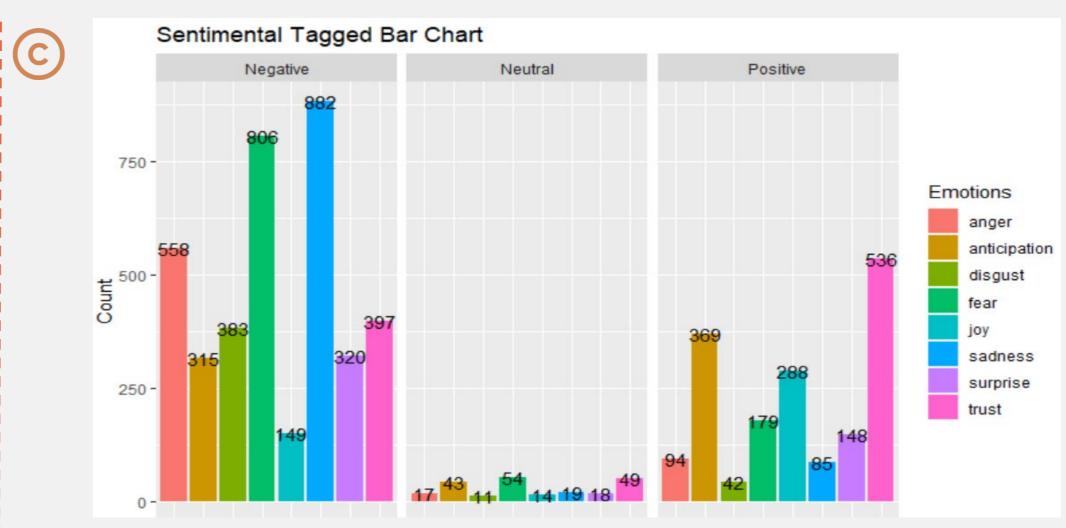
# state stay hours

A word cloud revealed that the most discussed terms amongst tweeters were:

- cases: 173 instances
- country: 167 instances
- fauci: 161 instances

democrats: 155 instances opening: 154 instances

# Temasek



### **Top 3 Emotions**

### **Negative Sentiment**

- 44.1% of tweets conveyed sadness
- 40.3% of tweets conveyed fear
- 27.9% of tweets conveyed anger

# 2.7% of tweets conveyed

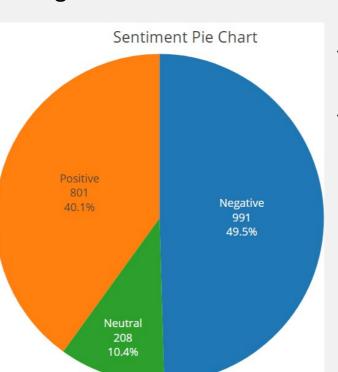
**Neutral Sentiment** 

- fear
- 2.4% of tweets conveyed trust
- 2.1% of tweets conveyed anticipation

### **Positive Sentiment**

26.8% of tweets conveyed trust

- 18.4% of tweets conveyed anticipation
- 14.4% of tweets conveyed joy



The overall percentage of negative sentiments was 49.5%.

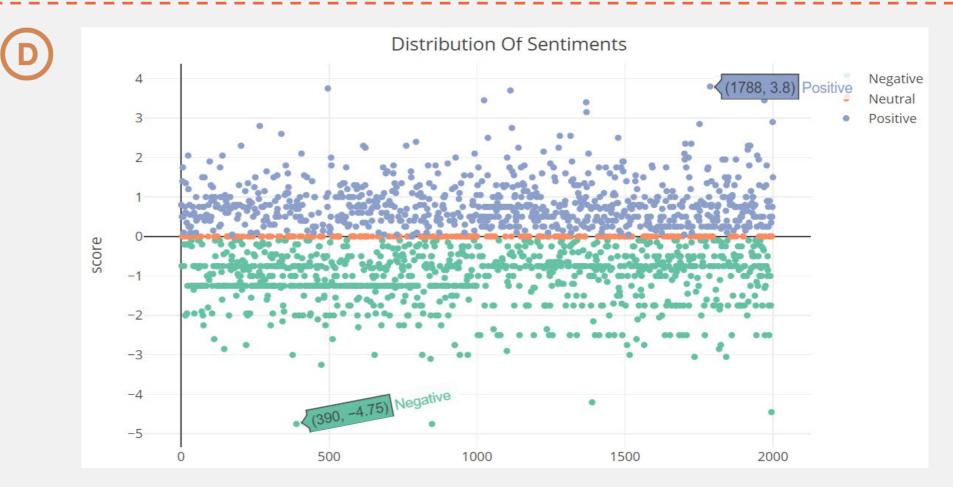
2 people disliked the coronavirus

The overall percentage of positive sentiments was 40.1%.

people were positive towards the coronavirus

The overall percentage of neutral sentiments was 10.4%.

people were neutral towards the coronavirus



### **Observations:**

- There are more negative sentiments than positive and neutral sentiments towards the corona virus
- The highest sentiment value is 3.8 🙂
- The lowest sentiment value is -4.75 (2)
- The degree of happiness is much lower than the unhappiness

### **Future work**

- Need to collect more data and over a longer period
- Plot word clouds using n-grams to provide a context to what the word is referring to.
- Better libraries needed to cater to different languages to encompass the sentiments of more people.
- Use libraries that can cater to real-time streaming data so that we can get live updates about what people are thinking

### Limitations

- The space provided is not big enough for more drill downs to be performed
- Since the data was only scraped on 12 May 2020, the sample size was too small to reflect the overall sentiment of the people and the topics that they are talking about.
- This analysis is conducted on tweets which contains irrelevant words, causing our data to be unclean.
- Sentiment analysis of tweets is challenging because of the limited contextual information that they contain.
- This analysis was only conducted on English tweets due to the libraries being able to cater to the English language