**Reddit vs Facebook: Comparing sentiment towards tightening of COVID-19 restrictions in Singapore across social media platforms**

**Nicole Neo**

Quantitative Methods in Social Science

Master of Arts Thesis, 2022

Columbia University

**Table of Contents**

[Abstract 3](#_Toc110534242)

[Introduction 3](#_Toc110534243)

[Data 9](#_Toc110534244)

[Methods 10](#_Toc110534245)

[Results 13](#_Toc110534246)

[Discussion 13](#_Toc110534247)

[Conclusion 14](#_Toc110534248)

[Appendices 15](#_Toc110534249)

[Bibliography/References 15](#_Toc110534250)

# Abstract

(100-150 words)

# Introduction

**Background**

2021 started on a good note for Singaporeans. Having just entered a ‘Phase 3’ of re-opening that saw substantial loosening of restrictions on group gatherings and events, with cases remaining at extremely low levels, it seemed as though the COVID-19 pandemic was coming to an end. However, that optimism was quickly shattered on May 4 2021. In response to a spike in infections, the government reimposed ‘Phase 2’ restrictions that limited gatherings to groups of 5 and mandated the closing of indoor gyms and fitness studios. What followed was a repetitive cycle of tightening and loosening restrictions over the next 6 months as the government struggled to cope with new waves of infection brought on by the Delta and Omicron variants. The frequently changing restrictions invited heated discussions among Singaporeans that spilled online as well, with comments on posts about the restrictions sometimes running into the hundreds and even thousands on social media platforms like Facebook and Reddit. Such discussion was often polarized, with some Singaporeans applauding the government’s cautiousness amidst rising infection and hospitalization rates, while others lamented the continued cap on social activities despite high vaccination rates.

A survey by the Institute of Policy Studies on Singaporeans’ sentiment towards COVID-19 restrictions revealed that the discourse could be divided along age lines, as age had a statistically significant effect on Singaporeans’ attitude towards the restrictions. Younger Singaporeans, whose mental and social wellbeing have suffered greater disruption during the pandemic compared to their older counterparts, were more willing to support the easing of restrictions.[[1]](#footnote-1) Given that the demographics of Singaporean users of Reddit and Facebook also differ by age (55.8% of Reddit users were aged 18-34 compared to only 43.8% of Facebook users being in that age group as of December 2020)[[2]](#footnote-2), it is consequently plausible that the sentiment of Reddit and Facebook users towards COVID-19 restrictions will also differ.

This thesis therefore seeks to validate whether there is a statistically significant difference in such sentiment between users of Reddit and Facebook. These platforms were chosen as they share strong similarities as platforms with discussion-based features allowing for more continued interaction through longer-form text compared to other platforms like Instagram (more visual media focused) and Twitter (shorter-form text interaction). This study will use state-of-the-art sentiment analysis models to assign a sentiment score for each comment, then conduct two-sample hypothesis testing to determine if Facebook and Reddit sentiment towards the tightening of COVID-19 restrictions in Singapore is statistically different. It will therefore produce not only behavioral and policy insights into Singapore’s COVID-19 response, but also new applications and methodological insights into sentiment analysis and Natural Language Processing (NLP) in general.

**Previous work**

There is currently no study that compares sentiment towards COVID-19 restrictions across Reddit and Facebook, but there have been similar studies of sentiment towards other COVID-19 related topics. These studies may not address my research question directly, but it provides a good understanding of the existing methodology and of how sentiment differs across other platforms and towards other COVID-19 topics.

The only cross-platform sentiment analysis study of COVID-19 content compares sentiment towards COVID-19 vaccines across Facebook comments and tweets in the United Kingdom and the United States. It found that Twitter sentiment was more negative than Facebook’s, with the proportion of negative sentiment being almost twice that on Twitter compared to Facebook.[[3]](#footnote-3) The study also suggested that the difference in sentiment could reflect differences in user demographics across the two platforms,[[4]](#footnote-4) which gives us some reason to expect that Reddit and Facebook sentiment towards COVID-19 restrictions could also differ as a result of varying demographics among the users. For sentiment analysis, this study used an ensemble model of lexicon-based methods VADER and TextBlob and a pre-trained Bidirectional Encoder Representations from Transformers (BERT) model. TextBlob had marginally better performance than VADER, and the lexicon-based methods had higher accuracy in detecting positive sentiment while BERT had higher accuracy in detecting neutral and negative sentiment.[[5]](#footnote-5)

Other studies that analyzed sentiment towards other COVID-19 related content examined only a single social media platform instead of comparing across platforms, but involved similar sentiment analysis techniques. One such study analyzed sentiment of comments on Facebook posts made by Public Health Authorities (PHAs) of Singapore, the UK, and the US about COVID-19. It found that while majority of the comments were negative, Singapore’s PHA still had the most favorable sentiment among its page comments, followed by the US and the UK.[[6]](#footnote-6) A temporal trend analysis of the analysis also revealed that over time from mid-February to early March 2020, the comments became more positive and supportive of the PHAs’ responses to the pandemic.[[7]](#footnote-7) This study hence reveals the value of conducting temporal analysis in uncovering additional insights from the data.

There is also some literature on the application of sentiment analysis methods to text containing Singaporean English (hereafter ‘Singlish’). Singlish is an English-based creole spoken by Singaporeans that contains elements of different languages such as the Hokkien, Teochew and Cantonese dialects spoken by the Chinese as well as Malay and Tamil. Singlish, being an informal language, is not currently recognized by sentiment analysis models like VADER and TextBlob, but could be problematic as Singaporeans tend to use a lot of Singlish on Reddit and Facebook, and Singlish contains many sentiment-bearing expressions that would otherwise go undetected by any sentiment analyzer. For example, among the top Singlish tokens in my comment data are expressions like ‘heng’ (fortunately), ‘haiz’ (a sigh), ‘ccb’ (a vulgarity) and ‘jialat’ (seriously bad). The only solution attempted for lexicon-based sentiment analysis thus far has been a manually constructed Singlish dictionary of 1024 terms from the Dictionary of Singlish and Singapore English[[8]](#footnote-8), the Coxford Singlish Dictionary[[9]](#footnote-9) and a Wikipedia list of Singlish vocabulary[[10]](#footnote-10). However, the researchers who used this method cautioned against it since Singlish is not a full language but just a variant of English, and there are many Singlish expressions that cannot be translated exactly or easily into English.[[11]](#footnote-11)

Considering the gaps in the literature, this thesis can add value by being the first quantitative study that carefully examines sentiments on Reddit and Facebook about the tightening of COVID-19 restrictions in Singapore. The results of the study will give us a deeper understanding of how platforms with different user demographics may (or may not) present different sentiments towards the same topic, and how said sentiment has changed over the course of the pandemic. This study will also serve as a test bed for NLP practices as it will apply and evaluate the performance of current state-of-the-art sentiment analysis models on Singlish text and on out-of-training domains (VADER was initially trained on tweets, opinion editorial articles, technological product reviews and movie reviews data[[12]](#footnote-12) but not on Reddit and Facebook data pertaining to COVID-19 restrictions).

**Scope of this thesis**

Before introducing the data and methodology used for this study, I will first clarify its scope. Firstly, this study will only compare sentiment towards tightening of restrictions in *Singapore* specifically rather than globally. While I cannot ensure that all comments will only discuss Singapore’s restrictions, I can increase the relevance of comments to the localized scope by examining comments on posts from the r/singapore subreddit and from Singaporean Facebook pages.

Secondly, this study will investigate sentiment towards the *tightening* of COVID-19 restrictions in Singapore, and not the loosening of restrictions. This restricted scope was adopted to ensure interpretability of the sentiment analysis results. By assuming that comments are responding to posts about the tightening of restrictions, positive-scoring comments can be interpreted as being supportive of tighter restrictions while negative-scoring comments can be interpreted to be opposing the tightening; conversely, if we assume comments to be responding to restrictions in general, it would be less clear whether a positive comment is supportive or unsupportive of the restrictions. While it would be difficult to ensure that all comments are talking about the tightening of restrictions, I can do so on a best efforts basis by only extracting comments from posts that are about tightening and not loosening of restrictions.

Thirdly, the temporal scope of this thesis will be May to October 2021. I choose to examine this period because restrictions then were the most varied; they were also most controversial among Singaporeans since restrictions were extended for a long period despite improving hospitalization and vaccination rates. A full timeline of the restrictions from May to October 2021 can be found in the Appendix. The limitations of the Pushshift API I used to extract Reddit comments also meant that data after November 2021 could not be extracted.

Lastly, this study only examines the quantitative differences in sentiment between Reddit and Facebook comments, and does not intend to investigate any underlying factors that might be responsible for this difference. The citation of other work above that discusses age as a possible determiner of differing sentiment towards restrictions only serves to explain the motivation behind my research question; is not meant to shape the interpretation of the results.

# Data

For my research, my data will comprise selected comments from Reddit and Facebook from May to October 2021. To ensure as much as possible that the difference in sentiment polarity between Reddit and Facebook comments is due to the platform alone, the content that comments are responding to needs to be as similar as possible. As such, I will only extract comments on posts with identical content about the Singapore government’s tightening of COVID-19 restrictions.

While Singapore’s Ministry of Health (MOH) only has a Facebook page and no official presence on Reddit, the restrictions it announces are always reported by Channel NewsAsia (CNA) and Straits Times (ST). CNA and ST are the primary news channels of Singapore, and all their articles about MOH’s announcement of new restrictions are posted on their respective official Facebook pages. On the r/Singapore subreddit, Redditors will repost the link to the same CNA and ST articles about the announcements without adding any further commentary in the post. Hence, it would be adequate to compare the comments on such posts as they contain the same content across both platforms.

To identify the relevant posts to scrape comments from, I first used the Pushshift API[[13]](#footnote-13) to search for all Reddit posts from May to October 2021 that contained keywords like ‘covid’, ‘phase’, ‘measures’ and ‘suspended’, and which also contained links to CNA and ST articles. From the results, I manually selected posts of links to articles that only announced the *tightening* (and not loosening) of COVID-19 restrictions in Singapore, making sure to exclude links about travel restrictions, border restrictions, vaccinations and commentaries about the restrictions that went beyond the scope of this thesis. Following this, I searched the official CNA and ST Facebook pages for posts of the same articles as those in the selected Reddit posts. I then used the facebook-scraper Python package[[14]](#footnote-14) to extract comments on those posts. The final dataset contains a total of 15 917 comments (9112 from Reddit, 6607 from Facebook) scraped from 34 Reddit posts and 28 Facebook posts (there were more Reddit posts than Facebook posts because some articles were posted multiple times on Reddit). These comments span the time period from May 4, 2021 with the introduction of ‘Phase 2’ restrictions which reduced group gathering sizes from 8 to 5, to October 20, 2021 when ‘Stabilization Phase’ restrictions which limited dining-in to groups of 2 were extended for another month.

# Methodology

**Research design**

The research question for this thesis is: How does the sentiment towards the tightening of COVID-19 restrictions in Singapore vary across Reddit and Facebook? I hypothesize that sentiment on Reddit and Facebook differ, with Reddit sentiment being more negative than Facebook. In this study, the target population would be all Reddit and Facebook content (posts, comments, etc.) about the tightening of COVID-19 restrictions in Singapore. However, obtaining a representative, random and probability-based sample of this population is highly challenging in the context of social media. Social media platforms often greatly restrict web scraping of data due to privacy issues

Therefore thi is only edescriptive. NO HYPOTHESIS TEST (Random data generation is therefore a necessary condition for a meaningful application of standard errors and p-values. When data do not satisfy this probabilistic requirement, p-values are essentially uninterpretable.)

As such, the primary quantity of interest in this study would be the difference in the mean polarity scores of Reddit and Facebook comments. Each comment will be scored by sentiment analysis models on a range from -1 (extremely negative) to 1 (extremely positive), from which we can then compute the mean score within each platform’s comment pool.

My research process comprises the following steps:

1. Obtaining word frequencies: Tokenizing the comment data and taking word frequency counts of the tokens to identify any tokens that need further pre-processing.
2. Text pre-processing: Treating abbreviations and Singlish phrases, then cleaning and tokenizing text to prepare the comment data for sentiment analysis.
3. Sentiment analysis: Applying sentiment analysis models (VADER and TextBlob) to assign sentiment polarity score to each comment.
4. Model evaluation: Quantitatively and qualitatively evaluating the performance of both sentiment analysis models using word clouds and comparison with human gold-standard scoring.
5. Hypothesis testing: Conducting two-sample test (Mann-Whitney U test) to validate hypothesis.
6. Temporal analysis: Plotting sentiment scores over time to understand how Reddit and Facebook sentiment changes over time.

**Word frequencies**

The main purpose of taking word frequency counts of the entire comment corpora is to determine if any of the more frequently occurring tokens in the text are abbreviations or Singlish phrases that need further expanding or translating in the pre-processing stage. To ensure that we count meaningful words, we first need to pre-process the text by removing features like hyperlinks, digits and punctuation; this was done with the help of the regex library in Python.[[15]](#footnote-15) The text was then be converted to lowercase to ensure that words with different capitalizations (e.g. ‘NO’ and ‘no’) would be counted as the same word when taking word frequency counts later. Stop words – common words in the English language that have little informational meaning (e.g. ‘can’, ‘be’, ‘I’) – were also be removed using the Natural Language Toolkit (NLTK) package in Python.[[16]](#footnote-16)

After cleaning the text, text tokenization was performed using NLTK’s word tokenizer function. Tokenization is the process of dividing a larger body of text into smaller parts called tokens. It goes beyond separating text by whitespace into words, even separating contractions (e.g. splitting ‘it’s’ into ‘it’ and ‘s’) and hashtags.[[17]](#footnote-17) A total of 209 387 tokens were obtained from the text.

From these tokens, I first identified the top 1000 tokens by their frequency count. Obtaining the word frequency count is a way to understand the key words in the corpus that could be highly used. The list of top occurring words can also serve as a lexicon for stop words in later stages of analysis, since there may be domain-specific words or common so frequently repeated in the text that they are less meaningful or informational. The top 20 tokens are listed in Figure 1 below.

|  |  |
| --- | --- |
| **Top 20** | |
| **Token** | **Frequency** |
| people | 2230 |
| covid | 1503 |
| like | 1417 |
| still | 1249 |
| cases | 1188 |
| vaccinated | 1165 |
| get | 1159 |
| go | 1107 |
| one | 991 |
| even | 961 |
| time | 888 |
| us | 853 |
| need | 842 |
| think | 828 |
| back | 824 |
| singapore | 822 |
| also | 810 |
| going | 767 |
| open | 731 |
| government | 698 |

**Figure 1:** Top 20 most occurring tokens in the comment corpus and their frequency counts.

From Figure 1, we see that there are some domain-specific key words like ‘covid’, ‘cases’, ‘vaccinated’, ‘singapore’ and ‘government’. These key words show that the extracted comments are discussing the COVID-19 situation and the government’s responses to the pandemic in Singapore, hence reflecting that we have successfully extracted comments relevant to our research question. The list also contains less meaningful words that could be considered as additional stop words should we need to produce more meaningful and sentiment-bearing output later.

Following that, I filtered through all tokens in the text to obtain ‘non-English’ tokens (i.e. tokens that are not in NLTK’s English dictionary), then selected the top 1000 non-English tokens. Tokens outside of NLTK’s English dictionary are likely to be Singlish expressions, Internet slang or abbreviations. Such ‘non-English’ tokens and abbreviations need to be replaced with English words from the dictionary, otherwise they may not be captured during sentiment analysis as I will be using lexicon-based sentiment analyzers that are designed to recognize English words. The top 20 non-English tokens, their meaning and their frequency counts are listed in Figure 2 below.

|  |  |  |
| --- | --- | --- |
| **Top 20**  **(non-English)** | | |
| **Token** | **Meaning** | **Frequency** |
| cb | Circuit Breaker (the name of Singapore’s lockdown measures in April-June 2020), or ‘Chee Bai’ (a Singlish vulgarity that means cunt and is equivalent in severity to the F word) | 453 |
| ktv | Karaoke | 382 |
| govt | Government | 372 |
| lol | Laugh out loud | 297 |
| ppl | People | 238 |
| gov | Government | 223 |
| etc | Et cetera | 220 |
| wfh | Work From Home | 166 |
| hbl | Home-Based Learning | 155 |
| others | Others | 143 |
| moh | Ministry of Health | 138 |
| pls | Please | 138 |
| mrt | Mass Rapid Transit (the name of Singapore’s subway system) | 134 |
| mtf | Ministry Task Force | 112 |
| ndp | National Day Parade | 103 |
| hari | Day in Malay, likely part of the phrase Hari Raya which is another name for the Eid festival | 102 |
| vax | Vaccine | 100 |
| mmtf | Multi-Ministry Task Force | 99 |
| oyk | Ong Ye Kung (Singapore’s Minister of Health) | 99 |
| reddit | Reddit | 85 |

**Figure 2:** Top 20 most occurring non-English tokens in the comment corpus and their frequency counts.

Among the top 20 non-English tokens are a mix of Singlish words, Internet slang and abbreviations for various government policies and bodies in Singapore. These tokens have a rather high frequency count and contain meaningful domain-specific information or sentiment, so it is important to replace them with an equivalent word or phrase in English so that the sentiment analyzers can capture them. I manually constructed a dictionary for this purpose on a best efforts basis – tokens were only replaced if they could easily be replaced with a word or short English phrase, and tokens with a frequency count of less than 10 were not replaced. Some exceptional steps were also taken for the following reasons:

* The phrase ‘Hari Raya’ could not be replaced with a meaningful English phrase as it was the name of a celebration (even its equivalent name of ‘Eid’ is also non-English).
* The abbreviation ‘ICU’ stands for Intensive Care Unit, and the word care contains positive sentiment that would be captured by sentiment analyzers, even if ‘ICU’ does not inherently bear positive sentiment. To prevent confusion, ‘ICU’ was therefore not expanded.
* As seen in Figure 2, ‘CB’ was used in the comment corpora both as a reference to the lockdown measures and as a vulgarity bearing highly negative sentiment. As it would be impossible to systematically determine which meaning the token had in each comment, this token was removed entirely from the corpora altogether to prevent confusion.
* Common Internet slang expressions like ‘lol’, ‘wtf’, ‘lmao’ were not replaced as one of the sentiment analyzers used in this research, VADER, is able to recognize and capture sentiment in such expressions, if any.

**Text pre-processing**

Unlike other common NLP tasks, text pre-processing is not necessary for sentiment analysis. Many conventional pre-processing steps such as the removal of digits, punctuation, and emoticons, the removal of stop words and stemming or lemmatizing words could potentially hamper the accuracy of sentiment analysis and hence should not always be adopted for such a task.[[18]](#footnote-18) This is because punctuation (especially exclamation and question marks), emoticons (e.g. :D or :-O), capitalization (e.g. ‘GREAT’ vs ‘great’) and stop words (which include degree modifiers or negation words like ‘not’, ‘very’ and ‘but’) all affect the polarity and intensity of sentiment in the text, and removing these features will erode these nuances. Lemmatization also ignores sentiment intensity in words with the same base roots (e.g. ‘good’ and ‘better’).[[19]](#footnote-19) Consequently, the same sentence can receive drastically different scores with and without pre-processing, so pre-processing should be applied only if absolutely necessary.[[20]](#footnote-20)

Additionally, pre-processing of text before sentiment analysis is not always necessary because the sentiment analyzers are designed to handle the aforementioned features when scoring text. Both TextBlob and VADER can process negation words like ‘but’ and ‘not’,[[21]](#footnote-21) while VADER also considers punctuation, capitalization, degree modifiers, Western-style emoticons, sentiment-related acronyms (e.g. ‘lol’, ‘wtf’) and common sentiment-bearing Internet slang (e.g. ‘nah’ and ‘meh’).[[22]](#footnote-22) To preserve as many sentiment-bearing features as possible in the text, I have decided to do minimal pre-processing on the comment corpora. I only removed hyperlinks, hashtags and line breaks in the text to increase machine readability, then replaced Singlish text and abbreviations using the manually constructed dictionary mentioned above. No punctuation, digits and stop words were removed, no lemmatization or stemming was applied, and capitalization was also preserved.

**Sentiment analysis**

After some minimal pre-processing, sentiment analysis was conducted on the comment corpora. There are two types of sentiment analysis – supervised and unsupervised. Supervised sentiment analysis involves training machine learning models on data that is pre-annotated with sentiment labels (e.g. ‘positive’, ‘negative’ or ‘neutral’), so that they can classify the sentiment of new text data. Whereas unsupervised sentiment analysis broadly involves using sentiment lexicons (lists of words and phrases that are positive or negative)[[23]](#footnote-23) to obtain the polarity of individual parts of the text, then aggregating the scores into an overall score that determines the text’s polarity.[[24]](#footnote-24) Unlike supervised sentiment analysis, unsupervised sentiment analysis does not require the data to be pre-labelled. As my comment corpora has not been manually annotated with sentiment labels, unsupervised lexicon-based sentiment analysis would be suitable for this study.

Among the many lexicon-based sentiment analysis tools used in NLP studies, SentiWordNet, VADER and TextBlob are more commonly used. SentiWordNet comprises 147306 synsets, each of which has a positive, negative or neutral score on a scale of 0.0 to 1.0. However, there are several key limitations of SentiWordNet. Synset scores are determined by a mix of semi-supervised algorithms instead of by humans, so the scoring has not been validated by gold-standard sources. Most of the synsets also do not have positive or negative polarity, greatly reducing the utility of SentiWordNet in producing output with meaningful sentiment scores. Additionally, SentiWordNet has not been designed to account for sentiment-bearing features in the context of short-form social media text.[[25]](#footnote-25) The limited utility of SentiWordNet is reflected in its relative performance to other sentiment analysis models like VADER and TextBlob. In a study that compared the performance of NLTK (which contains SentiWordNet resources), TextBlob and VADER on a sentiment analysis task of sentence-level snippets from Rotten Tomatoes movie review data, NLTK had the lowest F1 score of 57%, while TextBlob and VADER had much higher scores of 79% and 82% respectively.[[26]](#footnote-26) With this in mind, this study will hence only use VADER and TextBlob to analyze the sentiment of the comment data.

VADER and TextBlob have similar scoring scales for sentiment polarity. TextBlob returns scores for both subjectivity (i.e. whether something is an opinion or a fact) and polarity (i.e. positivity/negativity and intensity of sentiment)[[27]](#footnote-27), but for the purposes of this study only the polarity score will be obtained and the subjectivity score will be ignored. TextBlob’s polarity score ranges from -1.0 to 1.0, where -1.0 indicates extreme negativity and 1.0 indicates extreme positivity.[[28]](#footnote-28) VADER returns a normalized, weighted compound score; it is a sum of valence scores of each sentiment-bearing feature in the input text as per the lexicon, with that sum normalized to be between -1.0 (extreme negative) and 1.0 (extreme positive).[[29]](#footnote-29)

However, VADER and TextBlob do differ in the types of sentiment-bearing features they account for during sentiment analysis. According to TextBlob documentation on GitHub, TextBlob primarily uses the WordNet lexicon and does not consider the valence of non-textual features like emoticons or punctuation.[[30]](#footnote-30) Contrastingly, VADER incorporates not just sentiment-bearing words but also the sentiment of punctuation, capitalization, Western-style emoticons, acronyms (e.g. ‘lol’, ‘wtf’), and common Internet slang (e.g. ‘nah’ and ‘meh’).[[31]](#footnote-31) These characteristics make VADER particularly accustomed to sentiment in online microblog-type contexts,[[32]](#footnote-32) and possibly explains its superior performance to TextBlob in the aforementioned study by Bonta et al. on Rotten Tomatoes movie reviews data which also contain features similar to social media data.[[33]](#footnote-33) Despite the differences in TextBlob and VADER, I will still use both tools to score the comment data in this study as both are the two best performing sentiment analysis tools for social media text, and there has been limited literature comparing the performance of both models on social media text. I will also compare the polarity score output from TextBlob and VADER to evaluate the performance of these models on the sentiment analysis task.

**Hypothesis testing**

After obtaining polarity scores for each comment using the respective TextBlob and VADER Python packages, hypothesis testing will then be conducted to determine if there was a difference in polarity scores on average between Reddit and Facebook comments. In hypothesis testing, it is crucial to determine

The most appropriate test for the data in this study is the Welch t-test, a two-sample test which unlike the typical Student’s t-test does not require the sample variances to be equal. It is advised to always use the Welch t-test when comparing the mean or median for unrelated samples, as unequal sample variances can greatly affect the reliability of the Student’s t-test.[[34]](#footnote-34) It is also not crucial to verify the equality of sample variances when determining whether to use the Welch or Student’s t-test, since Welch test returns the same result as the Student’s test when sample sizes are large and when sample sizes and variances are equal.[[35]](#footnote-35) As such, to ensure the reliability of the hypothesis test, I will use the Welch t-test.

The Welch t-test assumes that both samples are normally distributed. However, with sufficiently large sample sizes (larger than 30 or 40), whether the normality assumption is violated or not is not important; even if the samples are not normally distributed, a parametric test like the Welch t-test can still be used. This is because as per the Central Limit Theorem, in sufficiently large sample sizes the sampling distribution tends towards normality regardless of the underlying distribution of the data.[[36]](#footnote-36) Nonetheless, as shown in Figures 3 and 4 below, the distributions of the VADER and TextBlob scores of Reddit and Facebook comment samples resemble normal distributions, hence the assumption of normality is fulfilled.

**Figure 3 – VADER score**

**Figure 4 – TextBlob score**

As the Welch t-test in as independent samples t-test, it also requires that observations are unrelated within and between samples. To ensure this,

A two-sided Welch t-test will be conducted twice on the samples using the SciPy Python package[[37]](#footnote-37) – once using VADER polarity scores and another time using TextBlob scores. In these tests, the null hypothesis () is that there is no difference in mean polarity score between Reddit and Facebook samples, while the alternative hypothesis () is that the means are unequal. The resulting p-value, the smallest significance level at which we can reject the null hypothesis assuming is true, will be used to determine whether we can reject the null hypothesis.

**Temporal analysis**

Following hypothesis testing, an additional temporal analysis will be conducted to better understand the patterns in the comment data and consequently, to provide greater insights into the results of the hypothesis test. The date when each comment was posted was first extracted from its timestamp using Python’s datetime package. For each comment sample, I then grouped the data by date and obtained the mean VADER or TextBlob score for each date. This data was then used to generate time series line plots of Reddit and Facebook comment sentiment scores over time, with sentiment score (on a scale of -1.0 to 1.0) on the y-axis, date on the x-axis, and a different colored line for Reddit and Facebook. A total of two such plots – one for VADER scores and one for TextBlob scores – will be generated.

# Results

* Results of statistical test
* Figures from the time series plot
* Interpret the results (call attention to aspects that provide new perspective, are policy-relevant or unexpected/particularly striking)

Link back to hypotheses

* Fail to reject hypothesis
* Vary over time – sometimes FB more positive than Reddit, sometimes the other way round! FB generally more positive at start of pandemic and less at end, vs Reddit which was the reverse, in the middle they were convergent.

# Discussion

* Evaluate performance of VADER vs TextBlob
  + Word cloud more generally
    - Remove stop words and domain-specific words to generate meaningful word clouds
  + Inspect top negative and positive comments
  + Consider also that we see evidence of failure to detect sarcasm and other misclassifications from the output displayed above

Limitations in data/methodology

* Limitation in obtaining responses exactly pertaining to tightening of restrictions – even when you get posts about tightening of restrictions, the comments to those posts can be about other restrictions, vaccines, what’s happening in other countries, or replying to other comments
* Limitation in VADER/TextBlob esp when we compare to human annotator
  + VADER/TextBlob trained on movie reviews and tweets about a wide range topics, not specifically on FB/Reddit comment data about COVID-19, and **not as exposed to Singlish**
  + Comments were scored as a whole and not on sentence basis, and comments sometimes consisted of multiple sentences with opposing sentiments
* Limited pool of comments – limited because of timeframe and also because of scraper limitations (PushShift cannot extract past Nov 2021, facebook-scraper hard to scrape comment replies)
* Research limited also because cannot explore possible reasons for the changes in sentiment

# Conclusion

Summary

Implications of results for research question

Possible extensions

# Appendices

# Bibliography/References

Bonta, Venkateswarlu, Nandhini Kumaresh and N. Janardhan. “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis.” *Asian Journal of Computer Science* 8 no. S2 (2019): 1-6. Accessed July 2, 2022. <https://doi.org/10.51983/ajcst-2019.8.S2.2037>.

Emblem, Holly. “Are you Scared, VADER? Understanding how NLP Pre-Processing Impacts VADER Scoring.” July 18 2021. Accessed July 7, 2022. <https://towardsdatascience.com/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadbc91d>.

Ghasemi, Asghar and Saleh Zahediasl. “Normality Tests for Statistical Analysis: A Guide for Non-Statisticians.” *International Journal of Endocrinology and Metabolism* 10 no. 2 (2012): 486-289. Accessed July 28, 2022. <https://doi.org/10.5812/ijem.3505>.

Glen, Stephanie. “Welch’s Test for Unequal Variances.” Statistics How To. Accessed July 15, 2022. <https://www.statisticshowto.com/welchs-test-for-unequal-variances/>.

Goh, Colin and Y.Y. Woo, eds. *The Coxford Singlish Dictionary.* Singapore: Angsana Books, 2001.

Hogenboom, Alexander, Daniella Bal, Flavius Fransincar, Malissa Bal, Franciska de Jong, Uzay Kaymak. “Exploiting emoticons in sentiment analysis.” Paper presented at the 28th Annual Association for Computing Machinery Symposium on Applied Computing, March 2013. <https://doi.org/10.1145/2480362.2480498>.

Hussain, Amir, Ahsen Tahir, Zain Hussain, Zakariya Sheik, Mandar Gogate, Kia Dashtipour, Azhar Ali and Aziz Sheikh. “Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study.” *Journal of Medical Internet Research* 23, no. 4 (2021): 1-10. Accessed September 23, 2021. <https://www.jmir.org/2021/4/e26627/>.

Hutto, C.J. and Eric Gilbert. “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.” *Proceedings of the International AAAI Conference on Web and Social Media* 8 no. 1 (2014): 216-225. Accessed July 2, 2022. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.

Jurafksy, Daniel and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third Edition draft).* Harlow: Pearson Publishing, 2020.

Kemp, Simon. “Digital in Singapore: All the Statistics You Need in 2021.” DataReportal. February 10, 2021. <https://datareportal.com/reports/digital-2021-singapore>.

Lee, Jack. “A Dictionary of Singlish and Singapore English.” Acessed June 20, 2022. <http://www.mysmu.edu/faculty/jacklee/>.

Lo, Siaw Ling, Erick Cambria, Raymond Chiong, David Cornforth. “A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection.” *Knowledge-Based Systems* 105 (2016): 236-247. Accessed June 23, 2022. <https://doi.org/10.1016/j.knosys.2016.04.024>.

Mathews, Mathew, Syafiq Suhaini, Mark Hou and Fiona Phoa. “Living with COVID-19 in Singapore: Attitudes, challenges and the way ahead.” Working Paper No. 43, Institute of Policy Studies, 2021.

Raamkumar, Aravind Sesagiri, Soon Guan Tan and Hwee Lin Wee. “Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison.” *Journal of Medical Internet Research* 22, no. 5 (2020): 1-12. Accessed September 27, 2021. <https://www.jmir.org/2020/5/e19334/PDF>.

Rusticus, Shayna A. and Chris Y. Lovato. “Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study.” *Practical Assessment, Research, and Evaluation* 19 (2014): 1-10. Accessed July 20, 2022. <https://doi.org/10.7275/4s9m-4e81>.

Ruxton, Graeme D. “The unequal variance t-test is an underused alternative Student’s t-test and the Mann-Whitney U test.” *Behavioral Ecology* 17 no. 4 (July/August 2006): 688-690. Accessed July 28, 2022. <https://doi.org/10.1093/beheco/ark016>.

Wikipedia contributors. “Singlish vocabulary." *Wikipedia, The Free Encyclopedia.* Accessed June 20, 2022. <https://en.wikipedia.org/w/index.php?title=Singlish_vocabulary&oldid=1102023299>.

Zhou, Min. “StopWords and Lexicon Normalization in Sentiment Analysis.” July 22 2019. Accessed July 7, 2022. <https://medium.com/data-science-blogs/stopwords-and-lexicon-normalization-for-sentiment-analysis-f9f10f0d4108>.

1. Mathew Mathews et al., “Living with COVID-19 in Singapore: Attitudes, challenges and the way ahead” (Working Paper No. 43, Institute of Policy Studies, December 2021), 53-56. [↑](#footnote-ref-1)
2. Simon Kemp, “Digital in Singapore: All the Statistics You Need in 2021”, DataReportal, February 10, 2021, <https://datareportal.com/reports/digital-2021-singapore>. [↑](#footnote-ref-2)
3. Amir Hussain et al., “Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study,” *Journal of Medical Internet Research* 23, no. 4 (2021): 5-6, accessed September 28 2021, <https://www.jmir.org/2021/4/e26627/>. [↑](#footnote-ref-3)
4. Hussain et al, “Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines”, 6*.* [↑](#footnote-ref-4)
5. *Ibid.,* 2. [↑](#footnote-ref-5)
6. Aravind Sesagiri Raamkumar, Soon Guan Tan and Hwee Lin Wee, “Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison,” *Journal of Medical Internet Research* 22, no. 5 (2020): 6, accessed September 27, 2021, <https://www.jmir.org/2020/5/e19334/PDF>. [↑](#footnote-ref-6)
7. Raamkumar, Tan and Wee, “Measuring the Outreach Efforts of Public Health,” 6-8. [↑](#footnote-ref-7)
8. Jack Lee, “A Dictionary of Singlish and Singapore English,” accessed June 20, 2022, <http://www.mysmu.edu/faculty/jacklee/>. [↑](#footnote-ref-8)
9. Colin Goh and Y.Y. Woo, eds., *The Coxford Singlish Dictionary* (Singapore: Angsana Books, 2001). [↑](#footnote-ref-9)
10. Wikipedia contributors, "Singlish vocabulary," *Wikipedia, The Free Encyclopedia,* accessed June 20, 2022, <https://en.wikipedia.org/w/index.php?title=Singlish_vocabulary&oldid=1102023299>. [↑](#footnote-ref-10)
11. Siaw Ling Lo et al., “A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection,” *Knowledge-Based Systems* 105 (2016): 245, accessed June 23, 2022, <https://doi.org/10.1016/j.knosys.2016.04.024>. [↑](#footnote-ref-11)
12. C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” *Proceedings of the International AAAI Conference on Web and Social Media* 8 no. 1 (2014): 222, accessed July 2, 2022, <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>. [↑](#footnote-ref-12)
13. The Pushshift API provides capabilities to query an archive of all Reddit comments and submissions (posts). Full documentation can be found at <https://github.com/pushshift/api>. [↑](#footnote-ref-13)
14. facebook-scraper is a Python package that enables scraping of comments, reactions and other data from public Facebook pages without an API key. Full documentation can be found at <https://github.com/kevinzg/facebook-scraper>. [↑](#footnote-ref-14)
15. regex is a Python package that enables replacement of substrings of text within a textual body using pattern matching. More documentation can be found at <https://pypi.org/project/regex/>. [↑](#footnote-ref-15)
16. NLTK is a package that provides comprehensive functionalities for NLP in Python; it supports text classification, tokenization, stemming, parts-of-speech tagging, etc. Full documentation can be found at <https://www.nltk.org/api/nltk.html>. [↑](#footnote-ref-16)
17. Daniel Jurafksy and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third Edition draft)* (Harlow: Pearson Publishing, 2020), 2. [↑](#footnote-ref-17)
18. Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019, accessed July 7, 2022, <https://medium.com/data-science-blogs/stopwords-and-lexicon-normalization-for-sentiment-analysis-f9f10f0d4108>. [↑](#footnote-ref-18)
19. Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019. [↑](#footnote-ref-19)
20. Holly Emblem, “Are you Scared, VADER? Understanding how NLP Pre-Processing Impacts VADER Scoring,” July 18 2021, accessed July 7, 2022, <https://towardsdatascience.com/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadbc91d>. [↑](#footnote-ref-20)
21. Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019. [↑](#footnote-ref-21)
22. C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” 221. [↑](#footnote-ref-22)
23. Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” *Asian Journal of Computer Science* 8 no. S2 (2019): 2, accessed July 2, 2022, <https://doi.org/10.51983/ajcst-2019.8.S2.2037>. [↑](#footnote-ref-23)
24. Alexander Hogenboom et al., “Exploiting emoticons in sentiment analysis,” (paper presented at the 28th Annual Association for Computing Machinery Symposium on Applied Computing, March 2013), <https://doi.org/10.1145/2480362.2480498>. [↑](#footnote-ref-24)
25. C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” 218. [↑](#footnote-ref-25)
26. Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” 5. [↑](#footnote-ref-26)
27. Siaw Ling Lo et al., “A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection,”236. [↑](#footnote-ref-27)
28. Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” 4. [↑](#footnote-ref-28)
29. *Ibid.* [↑](#footnote-ref-29)
30. Full documentation for TextBlob can be found at <https://github.com/sloria/TextBlob> and <https://textblob.readthedocs.io/en/dev/>. [↑](#footnote-ref-30)
31. C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” 221. [↑](#footnote-ref-31)
32. *Ibid.*, 225. [↑](#footnote-ref-32)
33. Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” 5. [↑](#footnote-ref-33)
34. Graeme D. Ruxton, “The unequal variance t-test is an underused alternative Student’s t-test and the Mann-Whitney U test,” *Behavioral Ecology* 17 no. 4 (July/August 2006): 688, accessed July 28, 2022, <https://doi.org/10.1093/beheco/ark016>. [↑](#footnote-ref-34)
35. Stephanie Glen, “Welch’s Test for Unequal Variances,” Statistics How To, accessed July 28, 2022, <https://www.statisticshowto.com/welchs-test-for-unequal-variances/>. [↑](#footnote-ref-35)
36. Asghar Ghasemi and Saleh Zahediasl, “Normality Tests for Statistical Analysis: A Guide for Non-Statisticians,” *International Journal of Endocrinology and Metabolism* 10 no. 2 (2012): 486, accessed July 28, 2022, <https://doi.org/10.5812/ijem.3505>. [↑](#footnote-ref-36)
37. SciPy is a Python package containing a stats module that performs statistical functions for probability distributions, summary statistics, statistical tests and more. Full documentation of SciPy’s stats module can be found at <https://docs.scipy.org/doc/scipy/reference/stats.html>. [↑](#footnote-ref-37)