ARCADA

Usage pattern and mental health analysis on Finnish speaking Twitter users

An analysis on mental health over the Covid-19 period

Bao Cai

Master's Thesis
Master of Engineering - Big Data Analytics
April 10, 2022

MASTER'S THESIS	
Arcada University of Applied Sciences	
Degree Programme:	Master of Engineering - Big Data Analytics
Identification number:	8662
Author:	Bao Cai
Title:	Usage pattern and mental health analysis on Finnish speaking Twitter users An analysis on mental health over the Covid-19 period
Supervisor (Arcada):	Leonardo Espinosa Leal
Commissioned by:	

Abstract:
Social media has been seen under negative light as it is correlated to mental health problems. However, it was also acknowledge as an alternative way to collect data and conduct studies and services on mental health. This study collected tweets from Twitter that contains certain keywords, deemed to be related to mental health issues in the Finnish language over the years 2020 and 2021. The tweets were then translated to sentimental gauges, negative, neutral, and positive, and analysed to find usage patterns of the users. The results showed some faint trends of the negative sentimental gauge following the Covid-19 trends and some evidence suggesting that the tweets, or the users by extension, are more likely to be negative toward the end of the week, as well as, later on in the day. The findings from this study are packaged with data volume, minority targeting, and study design limitation. From future studies perspective, this study left a lot of room for improvements as well as a tool for data collection to support in filling those gaps.

Keywords:	Twitter, tweets, depression, mental health, social media, Finnish, sentiment analysis
Number of pages:	48
Language:	English
Date of acceptance:	08.04.2022

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

AFK	Away From Keyboard
API	Application Programming Interface
BERT	Bi-directional Encoder Representations from Transformers
GUI	Graphical User Interface
IRL	In Real Life
NLP	Natural Language Processing
PII	Personal Identifiable Information

1 INTRODUCTION

1.1 Preface

Human beings are social animal (Aristotle 328 BC). A myriad of human traits are tuned by evolution to support this social aspect of human. Social media, either intentionally or accidentally, stands in the way of social interaction, ironically. Before social distancing was relevant due to the pandemic started in early 2020, social medias distance people from each other by hiding everyone behind a screen. We are at the point where terms such as IRL and AFK are common. As Baym (2010) remarked, "How can we be present yet also absent?". Various studies aimed to prove the correlation between social media usage and mental health problems, to name a few Vannucci et al. (2017), Lin et al. (2016), Twenge et al. (2017). But social media can also be used in favour of mental health.

Mental health supporting apps (Strauss et al. 2022) and online treatments (Andersson 2022) are becoming more present. Researchers started to realise the amount of data generated by social media can be used for studies. Concerns were discussed (Lahoud et al. 2022) as well as frameworks (Wongkoblap et al. 2022) to help guide future researches in collecting the right data. More importantly, social media brings about the momentary aspect of mental states that was not possible before (Verhagen et al. 2022).

1.2 Objectives

Twitter was chosen to be the social media platform for this thesis. The scope was tighten further into the Finnish speaking population. The overall objectives are to, first, collect a large amount of tweets from the target population within the 2 years 2020 and 2021. This is done via the official Twitter API (Search Tweets 2022) and no PII data was collected and stored for this study. The second is to apply an NLP model to the data set to generalise and abstract away the tweets into comparable metrics. This part is facilitated by an open-source model, published via the Huggingface, transformer (2022) package and specifically tuned to the Finnish language. The third, and last, is to analyse the metrics to find some usage patterns from the tweets collected. This part would look at year on year trends of the data as well as some more specific usage time frame, such as time of day.

2 RELATED WORK

This section will go over some related study to the topic discussed in the introduction section. Firstly, social media and its possible correlation with mental health issues. Secondly, recognising the impact it might have on mental health, it have more than just bad news to offer. This part will discuss the role of social media as an enabler for mental health studies and treatments. And lastly, since this study frame itself on the Finnish population, the majority, if not all, of the data would be in the Finnish language. This last part will review the studies done on the Finnish language in terms of sentiment analysis.

2.1 Social media in relation to mental health

Human race has evolved, in many different ways, into a "social animal" (Aristotle 328 BC). Waller, Cray & Burrows (2008) found that facial muscles of human were designed, or rather evolved, to communicate the fundamental emotions. Human's hearing capability was specifically adapted to the sound they make (Pickles 1982). Rizzolatti & Craighero (2004) described a "Mirror-Neuron System" assisting human, and some other species such as monkey, in survival by aiding the process of understanding and imitating the signals others give off. Social media disregards a significant chunk of these evolution traits by hiding the social animal behind a screen and keyboard.

Vannucci, Flannery & Ohannessian (2017)'s research shows a positive relationship between anxiety and the usage of social media. Lin, Sidani, Shensa, Radovic, Miller, Colditz, Hoffman, Giles & Primack (2016) demonstrated a similar results, but, with depression instead of anxiety, across three different metrics, total time per day, visits per week, and global frequency scale. New media, as Twenge, Joiner, Rogers & Martin (2017) put it, has a positive correlation with depression and suicidal rate. Although the study covered more than just social media inside new media, for example, internet in general and television use were included, the results suggested a positive trend between screen time and having symptoms of depressions. With that said, there were limitations and none of the results concluded social media causes mental health issues.

Some researches suggested that the relationship between mental health issues and social media usage had more nuances and, more often, a bi-directional relationship. It is possible that social media usage is driven by the mental health issue and not the other way around (Young & Rogers 1998). The way social media omitted the usage of non-verbal communications might attract people with depression due to less anxiety inducing factors. A more specific study by Bevan et al. (2014) regarding sharing life events on Facebook yielded mixed results. The study, focusing on a single social media platform, found that the results actually is dependent on the way social media was used, in this case is the content of the posts on Facebook. In general, sharing positive news has no indication that the user might be suffering from mental health issues, sharing negative news or not sharing at all, fearing for ones' own safety, has negative correlation to quality of life and positive correlation to stress level. In another similar angle, passive usage of social media such as viewing but not posting, following, especially following strangers in great amount, has destructive effects on mental health (Lup et al. 2015). Chen & Lee (2013) argued that interaction with social media correlate with mental distress through a intermediate layer, in addition to the direct route, of communication overload and/or low self-esteem due to being exposed to a plethora of overly positive personas. A separate study on adolescents in the Philippines by Labrague (2014) suggested that intensiveness of social media usage is not enough to signify mental states such as depression. Furthermore, depression usually goes with other psychiatric disorders, and ones' behaviour and usage pattern are complex so that the conclusion is not as definite as other known clinical problems.

Lin et al. (2016), in recognition of this nuance while discussing the results, suggested the bi-directional nature should be an opportunity to use social media as a medium to study about depression. In supporting of this optimistic view, Krausz (2017) remarked that the accessibility of social media could be used to study and provide education and mental health cares that cater better to how the new generation learn and interact.

2.2 Social media as medium to study mental health

The paradigm has shifted and scientific studies have new toys to play with. A plethora of applications can be found online with relative ease, Strauss, Zhang, Jarrett, Patterson & Van Ameringen (2022). The unfortunate problem with this medium is, however, the reliability of researches backing these applications, security and privacy concerns, and the inaccessibility of Mental Health evaluation frameworks used in some of the few reliable ones, barred behind tangible and non-tangible barrier of entry. On another view, therapy sessions can now be conducted online via a screen (Andersson 2022) under various forms, ranging from no instruction from therapists such as guiding texts, audios, and videos, all the way to video communications. The research has shown little to no difference between online therapy and the traditional way, face-to-face, for mild to moderate cases, with some extra mindfulness about the aforementioned security and privacy concerns.

Circling back to social media, Shabbir, Gabarron, Lau & Househ (2016) introduced a wide variety of "dimensions", as they put it, that social media can be used for health care in general. These dimensions include important information sharing, especially during a potential epidemic, or an actual pandemic happening at the time this study was conducted, interaction and engagement with patients, patients empowerment, etc. And most importantly, they recognised that the amount of data that was generated by social media every minute, even, could and should be used for health care delivery, pass the conceptual stage and the surveillance uses.

Lahoud et al. (2022) recently discussed about the use of internet for data collection. Reliability was one of the issue that researchers were concerned about, very early on in the history. The concerns were often about the bias of the data collected, the generalisation capability from the results, fraud, and the results would be different from traditional methods. In summary, there were 2 actual concerns, fraud and the fear that data collected from the internet is not the same as data collected the traditional way. In fact, this can be seen more as a fear of changes rather than the problem with the platform itself because the concerns mentioned were not original. These concerns existed before and will continue to exist in every data collection method. Lahoud et al. went on and discussed the benefits versus the costs of using the internet to collect data instead. Much like any technology, old or new, there are benefits of adopting internet for data collection, namely the vanish of typical overhead costs, especially travelling cost, and anonymity. And there are also drawback, again, in terms of overhead costs, such as adopting and maintenance cost. The problems essentially boiled down to whether there is enough expertise around to facilitate the data collection, instead of the way the data is collected.

And despite the concerns for the internet, and social media by extension, they were used in various studies. In facts, the matter was considered to great extends that there can be up to 13 different type of social media platforms, some of them are blogs, micro-blogs, collaborative projects, etc. (Wongkoblap et al. 2022). And that is only one of the methods to classify social media platforms. Then they went on a great length to discuss various data collection ways that can be applied to different social media platforms, different processing steps and/or machine learning algorithms that can be used to enrich the collected data. The researchers mentioned a workflow that very closely aligned with what was done for this paper. In summary, tweets were collected from Twitter using search API based on some predefined regular expression, and then get annotated and filtered, before being used for depression analysis. Toward the end, they also proposed a framework how social media platforms can be integrated with other organisations such as government and health organisation to create a safe and efficient channel to study and provide aid and protection to the users in need.

On another perspective, Verhagen et al. (2022) talked about timing, the momentary nature of social media data and how this is very valuable to psychological assessments. Traditionally, psychological treatments were developed and used based on probability that were generalised to a large group of similar individuals. Although this is great for general studies, it missed the individuality of each person. On an even deeper level, people experience emotions only momentarily, hence emotional or mental state, as in temporary and subject to changes. And traditionally, it is very difficult to capture the momentary nature of emotions. More often than not, the data collected were reflections of previous emotions or mental states, and context can be lost. Social media, or internet as a broader platform, were designed to satisfy high availability level. With a careful design, data collected from social media platform can provide a better reflection on the momentary nature of mental states as well as the context which gave rise to the emotions.

2.3 Sentiment analysis on Finnish language

There were a lot of works already done for the Finnish language despite being an extremely difficult language to get around. One of them is the FinnSentiment study by Lindén et al. (2020). The authors realised in their search for previous works that these was a lack of a large enough data set in Finnish for sentiment analysis. Therefore, the study provided a data set of 27000 annotated sentences, made available through University of Helsinki's NLP project on Github (XED 2022).

Aside from the data set, other studies also utilised the data set to certain extends. University of Helsinki also has their own application built on top of the Open Substitles data set, called Sentimentator (Sentimentor 2022). And an independent Master's Thesis was done using the same data set (Kajava 2022). But most notably is the adapted model built on top of the data collected by finnSentiment study and made available through the Huggingface's transformer open source package (Hauhio 2021).

3 RESEARCH METHODOLOGY

On high level, this section talks about the technologies facilitated this study. It goes through the source of the data, what kind of data was collected and how, with possible challenges, how it was processed, at raw, which pre-analysis model was applied on top of the raw data, and what analysis to be performed. A complimentary section of experiments goes to more details on how the raw data was filtered and analysed. The steps are summarised as figure 1 below.

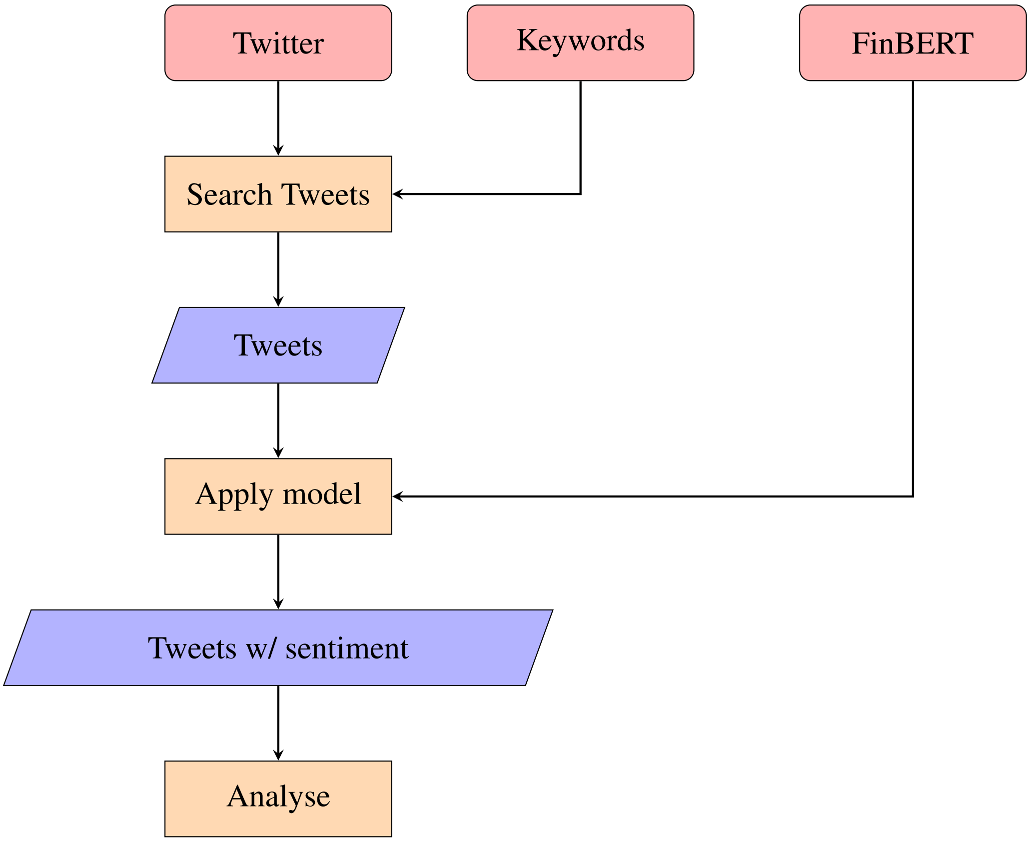

Figure 1. Overall study steps

3.1 Which social media platform?

Twitter was chosen to be the scope platform for this thesis. It was possible to choose more platform, namely Linkedin and Facebook, for data collection. However, while there was no particular reason to choose Twitter in the first place, including data from other platform poses some problems to the data set.

First, the format of the data would be different. While, for example, Facebook allows long texts and inclusion of plethora of other medias, Twitter is limited to just texts and minimal media that can be represented as text (emoji). Second, purpose of usage of these platform are, at the least, slightly different. Twitter is often associated with a blog (Shabbir et al. 2016). Linkedin and Facebook on the other hand are, although different in formality level, a networking site. This would leads to wild and unnecessary, differences in the nature of the data collected, which would not benefit the later analysis. This also raised the last issue, which is demographic. Users dictate contents and contents attract users. The demographic of Twitter's users would be different from that of, probably not so much from Facebook but, Linkedin.

For, the least, these reasons, Twitter was chosen to be the sole source of data for this study. With that said, this does not rule out completely the possibility of using other platforms in support or in addition to this study. This will be discussed further in the last section.

3.2 What data was collected?

Tweets were the primary data collected from Twitter. The tweets were searched based on a set of selected words, case insensitive, see appendix A, which were identified to be closely related to depression (Tana et al. 2018). In the study, Tana et al. (2018) monitored closely the following 6 chosen words, "masennus", "masennusoireet", "masennustesti", "masennus testi" (they meant depression, depression symptom, depression test, and depression test in English respectively), "depression", and "depression test". The rest of the words were found related by Google Search. One thing to note is that, searching for tweets also results in retweets and replies. All of these data points are associated with a separate tweet id, a conversation id, and a unique id that was used to identify unique texts.

Aside from the text and the ids associated to each data point, timestamp when the data point was created, type of the data point, and the author id were also recorded. No personal identifiable information (PII) was collected and this study was not planned out to use PII in first place. Although with the right search, the texts could trace back to their author, the text is, first and foremost, stored in a secured machine with limited access, and the analysis done was on the time pattern of the tweets (this will be clarify in the following subsections) rather on individual tweets. The text would also be used as plain text as input for a machine learning model to gauge the sentiment, however, it would not require any one person to peer at the texts. More details about the implementation will be available at another subsection of this methodologies discussion.

Time frame of the data collections was from early 2020 to the end of 2021. Initially, only 2021 period was chosen, fearing the limitation of the API. However, the window was extended to cover 2020, seeing it also coincide with the Covid-19 pandemic period. This would frame the scope of this paper on the pandemic period. On one hand, more analysis could be done on the pattern of the tweets as 2 years, with similarity in global events, of data was collected. On the other hand, it limited how this study could be generalised for a broader time period.

Finnish was the language of interest and only tweets written in Finnish was scraped off Twitter. Initially, to limit the scope further, only tweets with the geographic location of Finland was collected, from now will be referred to as the Finn set. For good measure, however, a separate set of data was collected alongside the Finland set. This set of data still homed in on the Finnish tweets but its geographic location was expanded to cover all locations. From now on, this second data set will be referred to as the Full set. It is worth noting that, even though the language and geographic location were set to Finnish and Finland respectively, some of the keywords were English, simply because it is not uncommon to mix languages, especially with scientific vocabulary involved. Furthermore, fixing the tweets metadata on Finnish language does not guarantee that the data set consists of all the tweets written in Finnish. The reason for this is the metadata of the tweets is dictated by the language setting of the users and the users are free to choose the appropriate language themselves. This caveat is also true for geographic location setting.

3.3 How the data was collected?

Apache Airflow, version 2.1.4, was used to schedule collections of data. A custom integration, operator in Airflow language, was implemented to establish the connections with Twitter API. More specifically the Search Tweets endpoint of the Twitter API.

Airflow was chosen primarily for its scheduling benefit. The data would be collected specify the scope of search. Some of them are, the important ones, a language of choice, a country, an end time of the period, a specific endpoint (recent with one week maximum look back or all for full database search, only available for Twitter Academic (2022)), expansions (for more information), which fields of users and/or tweets to retrieve. The operator will construct one or several queries base on the provided options and execute them. The end result is the same as providing a ready made query.

Other environmental setup is available in the repository (Cai 2021) mentioned above. Airflow would need to be setup locally and a dag, the pipeline itself, would need to be constructed. They, too, are available in the repository. It is recommended to use a GUI for ease of use to trigger, re-run, and debug the pipeline. Once the pipeline is setup, assuming the machine hosting it is kept running all the time, the pipeline should fetch the data once every week as described in the previous subsection.

3.4 How the data was processed?

As mentioned in the previous subsection, regarding which data was collected. The tweets themselves are not of interest for this study. Instead, the main focus was on the frequency and the time pattern of the tweets for each specific users. That leaves the text with no actual usage, at first.

Gauging the text to categories would make them useful even without the actual text. However, try reading the texts is simply not wise. At the very least, 2 years worth of text is not something one can read and draw conclusion. And on the other hand, giving the text out for other Finnish native speaker to label, similar to what was done on FinnSentiment study (Lindén et al. 2020) raises questions about bias that needs to be controlled, as well as the problem of privacy. At the very least, it would still require one or more person to manually look at the texts, which defeats, partially, the purpose of the automated data scraping pipeline described above. Fortunately, an independent study/project was done by Hauhio (2021), utilising the previously mentioned FinnSentiment study, produced a Bert model to classify whether a given text is negative, neutral, or positive. The model was made available as a part of the transformer package of Huggingface co (Huggingface, transformer 2022).

The model was then included in a second pipeline, which then took every single files produced from the first pipeline, and applied the model to the text, then re-saved it. This transformed the data set from having the text as part of its columns to having the positiveness of the text instead. For each of the given texts, the Bert model outputs an array of 3 floating points probabilities, add up to 1, negative, neutral, and positive indicator. The higher the probabilities, the higher likelihood of the text to be classified as negative, neutral, or positive. For the sake of simplicity, the sentiment of a tweet will be determined by the highest probability of the 3.

3.5 What analysis to perform?

Analysis performed would mostly be time series analysis. Utilising the probabilities outputted from the Bert model as gauge, negative, neutral, and positive, the gauges and simple counts of the tweets would be analysed against the weeks and months over the 2 years period. In addition, year-on-year comparison of the gauges and counts against day of week and time of day were also in the interest of this study. On top of these analysis, it is also important to reflect upon how the trends are different between the Finn set and the Full set. Then, the sentiment gauges trends would be compared against the corona waves or the relevant social situation at certain times, i.e. holiday period.

4 EXPERIMENTS

This section is a complimentary to the previous methodology section which talks mainly about the technology and on high level about the analysis. This section goes into the discovery phase and actions done on the data set, as well as the analysis. The results are presented and discussed in the following sections.

4.1 Model application

Apache Spark was first used for applying the Bert model onto the text scrapped from Twitter. A critical fault occurred with this approach. Spark is supposed to be used for extremely large data sets, ranging at the gigabytes territory.

This study, at first, looked at Twitter for a potential big data source. This was supported by the reported 353 millions users that Twitter could reach with advertisement, the users growth rate exceeding 8% (Kemp 2020), and the massive volume of 340 millions reported, albeit quite old, tweets a day (Twitter 2012). The actual volume, however, is abysmal. The Finn set only accumulated 1603 tweets over the course of 2 years and the Full set, doing a little bit better, only reached 96258 tweets. The situation is actually predictable as there are roughly 5 millions native speaker in Finland according to Finland (2022) in addition to the problems with tweets metadata setting mentioned in the previous section. Adding more injures to injuries, the chosen keywords were quite, rightly, specific to capture correctly the depression related tweets. As a side effects, this also limited greatly how much tweets can be gathered. However, this should not be seen as a downside, since the goal of the study is to home in to depression related tweets and not all the tweets. These attributes of the data set resulted the relatively small size of the data set.

Why is this a problem? In order to process a huge set of data, Spark utilises distributed computation, meaning it wants to set up a large amount of nodes to process the data in chunks. Essentially it is a trade-off between cost and processing time. When the data is barely 100000 tweets, it is wasteful, at the very least, to setup a cluster and process the data with it, overlooking already the fact that the model is ready made and the only heavy processing that needs to be done is to apply the model. In the end, the data was processed with standard Python, and the multiprocessing library instead (Multiprocessing).

4.2 Exploration

Before looking at the sentiment data, let us evaluate how the search keywords were presented in the collected data. This will only look at the Full set instead of both of the set to have a more substantial volume of tweets.

Figure 2 shows 2 of the original 6 keywords that yielded 0 tweets. This might not be a surprise considering when talking about mental health, "diagnose" could be used in place of "test". In addition, there are more than 1 way to express depression test in the Finnish language, as the keywords also have 2 of them, and this might be the more academic one, or a lesser known one.

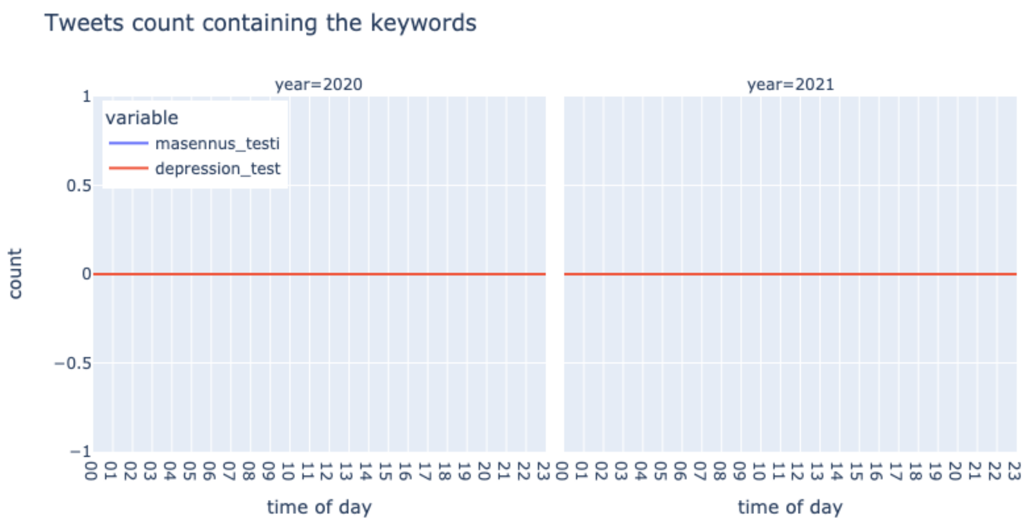

Figure 2. Tweets count of the 2 least presence of the 6 original keywords

Similar to above, these 2 keywords also did not show a lot of traffic in the tweets, figure 3. It does seem like just "masennus" or "depression" is easier to refer to in a tweets (tweets do have a limit in number of characters), figure 4. However, it is very difficult to draw any conclusion or even speculation at this stage alone as this might very well be the hours which the users are most active during the day.

Weekly aggregated view, figure 5, shows a downward trends in volume of the 2 keywords mentioned above. In contrast to the common downward trend, year 2021 shows a rise in "depression" usage on Saturday. The outlook from year to year is that there is a decrease

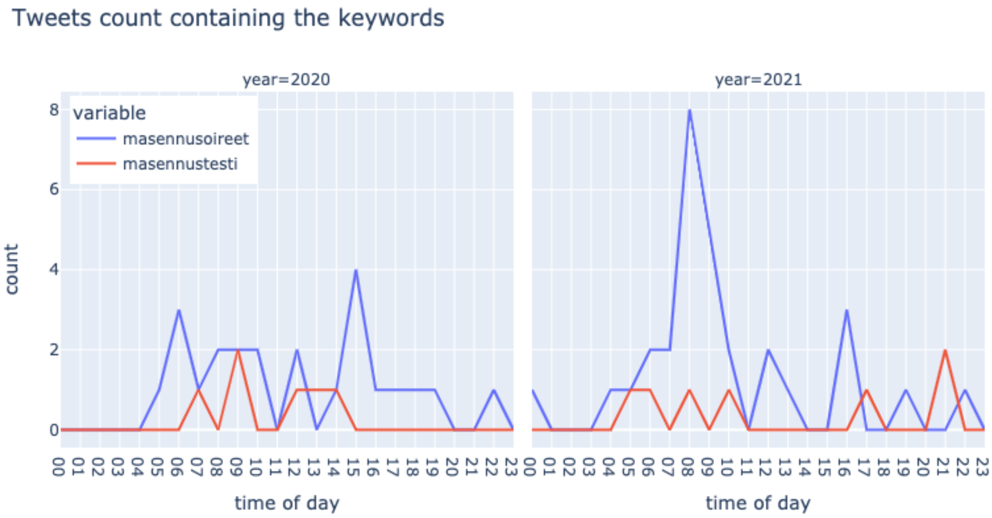

Figure 3. Tweets count of the 2 minor presence of the 6 original keywords

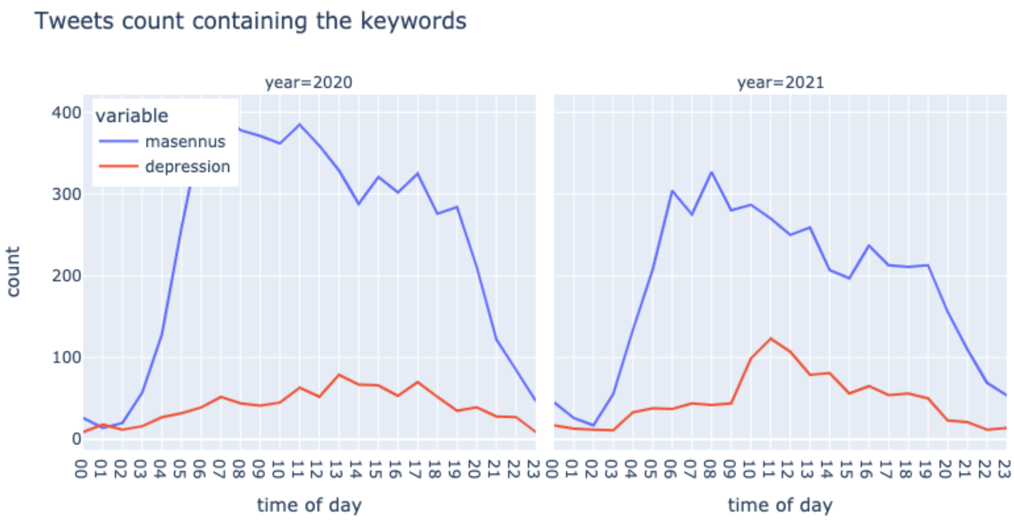

Figure 4. Tweets count of the 2 most dominant presence of the 6 original keywords

in volume of the 2 dominant words. Figure 6 also supports this. It, however, also shows some abnormalities in the peaks on April and December 2021.

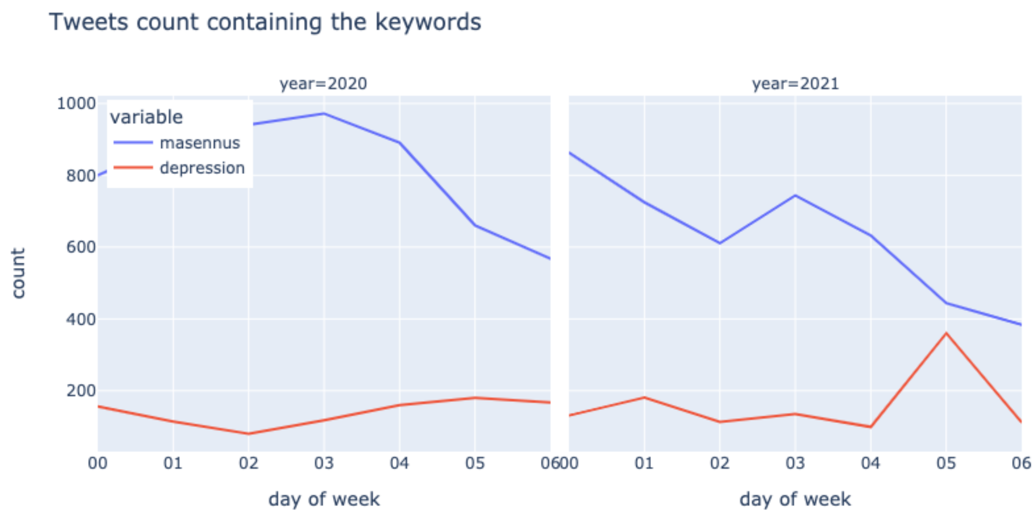

Figure 5. Tweets count of the 2 minor present of the 6 original keywords

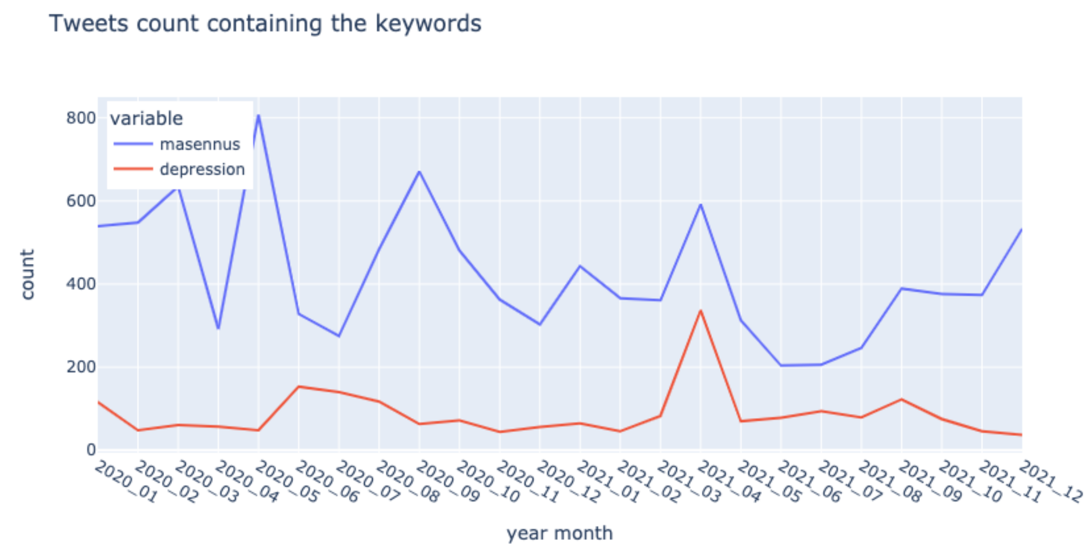

Figure 6. Tweets count of the 2 most dominant present of the 6 original keywords

A brief look at the data on summary table 1 showed that only a small portion of the, already minimal, Finn data set is clearly negative tweets. The majority of the tweets were neutral in sentimental sense.

This is a trait that the Full set also shares. Table 2 shows only 10% of the whole set are negative tweets. A huge portion of the set is neutral. The rest of the set is positive. The

	counts	shares
negative	213	13.3%
neutral	1227	76.5%
positive	163	10.2%

Table 1. Finn set sentiment summary

portions are roughly similar between the 2 sets.

	counts	shares
negative	9611	10.0%
neutral	79851	83.0%
positive	6796	7.0%

Table 2. Full set sentiment summary

But let us also look at a different angle of the data. While the Finn set has a total of 1603 tweets, they only came from 650 unique users. This number for the Full set is 22331 users over the total of 96258 tweets. This translates to, on average, a user only posts once (Finn set) or twice (Full set) a year. While this might be the truth for some or even a lot of cases, it does not allow the data sets to reflect a usage pattern.

Figure 7 is a "small" snapshot of the distribution of the number of tweets contributed by each account in the Finn set. The ironic "small" here was directed at the main part of the distribution as it capped out at less than 4 tweets per account. To put the number into perspectives, if a user only posts on Christmas Eve and New Year Eve every year, said user would already be considered as an outlier of this distribution. The same can be seen on the distribution of the Full set, figure 8.

Although, this following action on the data sets will greatly reduce the volume of the data further, it is necessary because the users who only contributed less than 4 tweets a year, in a sense, did not contribute to the trend or usage pattern of the users in interest. It is, therefore, necessary to select only the frequent users for the later analysis.

Defining how frequent should a user be to be called a frequent user is subjective. Let us have a look at the upper end of the Finn set distribution. Table 3 shows the remaining 10 accounts which contributed 15 or more tweets in the course of 2 years. Judging from the

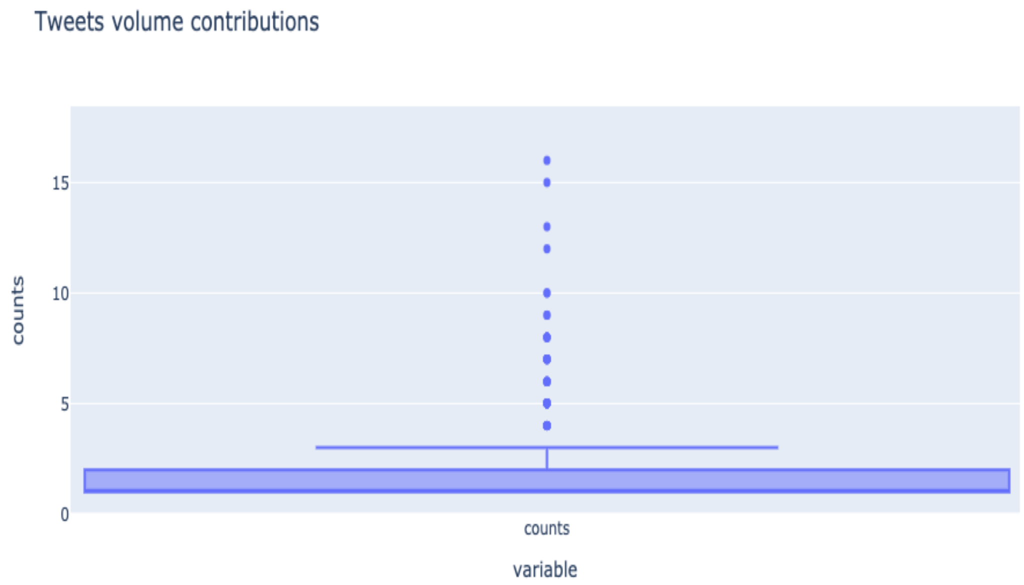

Figure 7. Contributions to the tweet counts per account on the Finn set

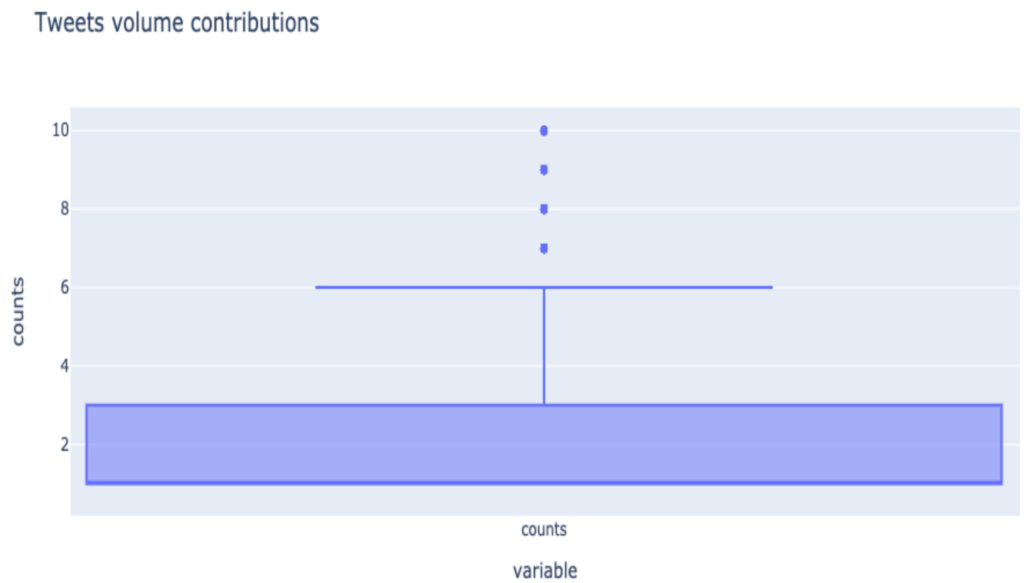

Figure 8. Contributions to the tweet counts per account on the Full set

distribution, once a day is certain not a serviceable threshold to determine which user is a frequent user, not to mention even more frequent users. Weekly is a fair frequency, then again, the Finn set would be reduced to a single user usage pattern (104 or above tweets over the course of 2 years). Taking monthly as the threshold for determine frequent user would retain the 8 most "active" users from the Finn set. These 8 users contributed 441 tweets to the Finn set, or 27% of the whole data set. This number would be 531 users contributed 37396 tweets, or almost 40%, to the Full set.

user	counts
dummy a	171
dummy b	62
dummy c	43
dummy d	42
dummy e	40
dummy f	31
dummy g	26
dummy h	26
dummy i	16
dummy j	15

Table 3. Finn set distribution beyond 15 tweets

With this filtering in place, the shares of the 3 sentiments remain, relatively, the same. Table 4 shows the new counts and shares of the tweets for the Finn set.

	counts	shares
negative	47	10.7%
neutral	361	81.9%
positive	33	7.4%

Table 4. Finn set sentiment adjusted summary

Similarly, the shares in the Full set also changed only so slightly. By doing this filtering, the data sets were pruned of the actual outliers, the infrequent users. At the same time, the data sets are still representative of the actual population since the sentiment shares did not change by a significant amount. And speaking of outliers, it is worth having a second look at the distribution of the Full set.

Due to the small size of the Finn set, the number of users left in the adjusted Finn set is only 8 users. Therefore the distribution is nothing more than the placements of each of

	counts	shares
negative	2350	6.3%
neutral	32214	86.1%
positive	2832	7.6%

Table 5. Full set sentiment adjusted summary

the 8 users on a number line. The adjusted Full set distribution improved a lot, in contrast, figure 9. This is not, however, the full view of the adjusted Full set. The adjusted Full set still ranges from 25, all the way, to 1665 tweets per user. The effective range of the distribution is at 113 tweets per user, the upper fence. Everything comes after that is considered as "outliers" in figure 9. With that being said, the outliers still represent the frequent users and their usage patterns are worth taking into account, going forward.

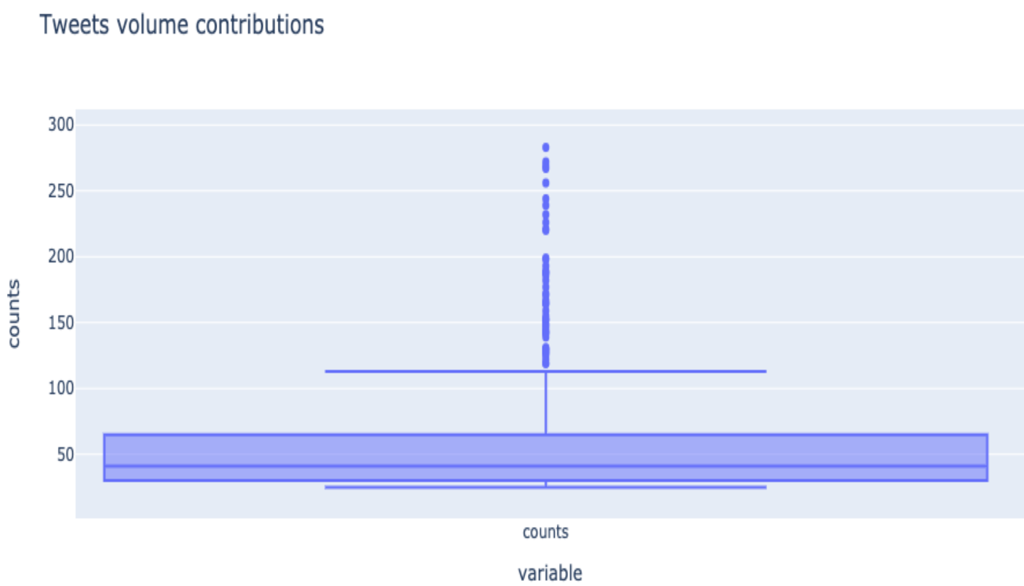

Figure 9. Contributions to the tweet counts per account on the adjusted Full set

4.3 Main analysis

With pre-processing out of the way, the main analysis will look at the tweet gauges and the counts that contribute to the gauges over different time windows.

The gauges in interest were the negative and positive gauges of each of the tweets. The reason why neutral was left out of this part was the exceptional volumes of them. The counts will illustrate this overwhelming volume by also showing the counts of the neutral tweets, either together with negative and positive counts or separately, which ever is more appropriate. The reason again is the exceptional volume of them might overshadow the insight from the negative and positive counts.

As mentioned above, this analysis looked at different time window to discern a usage pattern or something resemblance of a usage pattern from the tweets. The time window will be from more granularity (gauges per week) to less granularity (gauges per month), and from a continuous timeline to period abstractions, day of week and time of day. Daily gauges was left out simply because the volume of the data set did not allow that granular analysis to be carried out. Moreover, a daily trend might just show random up and down trends through out the data set since human beings, as similar to other animals living on Earth, are governed by the circadian rhythms (Kreitzman & Foster 2021). Circadian rhythms are embedded elements in the genome that help control, or rather signal, internal systems to work together in synced, often referred to also as biological clocks. This rhythm for human varies from 22 hours to about 25 hours, average out at about 24 hours and 10 minutes. A daily window look at the data set might be fragmented due to the individuality of each circadian cycle. Instead, looking at the cycles as a collective is more fruitful.

In additional to all the above, the 2 sets of data would be compared to each other. Mainly the period abstractions would be compared to each other. The continuous timeline was a tad difficult to reflect upon for the Finn set due to the "incredible" sample size of 8 users left from the elimination round. The continuous timeline from the Full set would be more valuable from the standalone data set perspective, as well as for when this data set was compared to the overall World Coronavirus situation.

5 RESULTS

This section presents the results of the analysis. First, there will be the outlook of the analysis in form of charts, discussions, and comparisons solely on the data sets themselves. Then the results will pivot into comparisons with external data, specifically the Corona virus trends.

5.1 Small disclaimer

After having gone through the analysis and acquired the feel of the data overall, it is necessary to lay out a disclaimer early on, before talking about the results. These point would be addressed again in details in the conclusions part, but, much like credits need to be given when credits due, disclaimers are only useful when placed at the right place. The keywords were the search terms collected from Google Trends which have significant volumes at certain times. Tana et al. (2018) said it best, why people are using such keywords are not the best reflection of what is happening inside their heads. This study utilising similar keywords on a different medium, predictably, suffered the same limitation. There is no way to really know if these users are experiencing depression or any other mental health. All that this study knows is the users using these keywords have the tendency to be more or less negative at certain time window.

5.2 Usage pattern

5.2.1 Weekly

Figure 10 shows the gauges over each weeks of the 2 years period of the Full set. Overall, year 2020 only has 3 noticeable peaks where the tweets are much more negative than normal. These peaks are at week 30, 41, and 48, corresponding to the summer, autumn, and slightly before Christmas time, beginning of December. However, year 2021 were significantly more negative, beating some previous year's peaks. But what was more troubling is the rise of overall negativity. Periods that were mostly positive before are now negative or, at best, alternating between the sentiments.

Looking at figure 11, the volumes of the negative peaks were not significantly higher than the positive ones. Figure 12 gives a bit more clue to why the peaks can happen. There

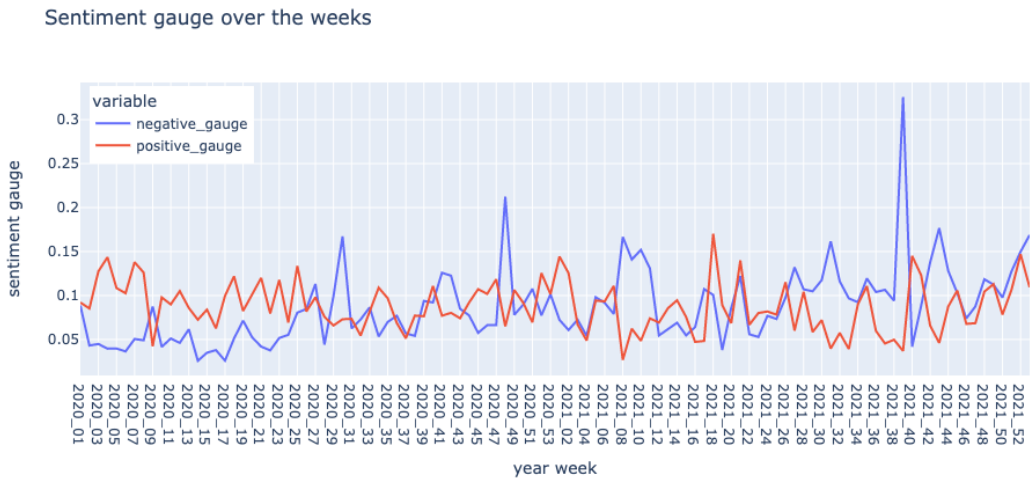

Figure 10. Sentiment gauges over the weeks of the Full set

are much more neutral tweets than any other two, and even combining them usually will not help. At the weeks where the negative peaks happened, there were dips in the volume of the neutral tweets. The humongous amount of neutral tweets, relatively to the other 2 sentiments, normally diluted the negative tweets. All except for one, week 11 of year 2021, the number of neutral tweets increased more than 4 folds from the previous week (from 68 tweets to 286 tweets) and there were no noticeable increase in the volume of negative tweets to match the trend. And yet, the negative gauge at that week remained, relatively, the same as the previous week. Individually, the tweets must have been extremely negative to be able to resist the similar dilution from the neutral tweets as before. What happened during that week?

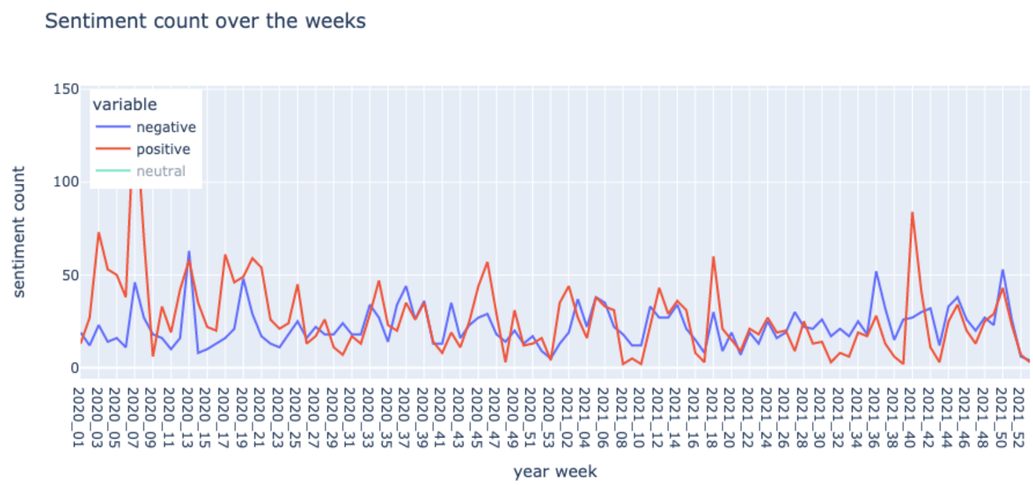

Figure 11. Sentiment counts over the weeks of the Full set

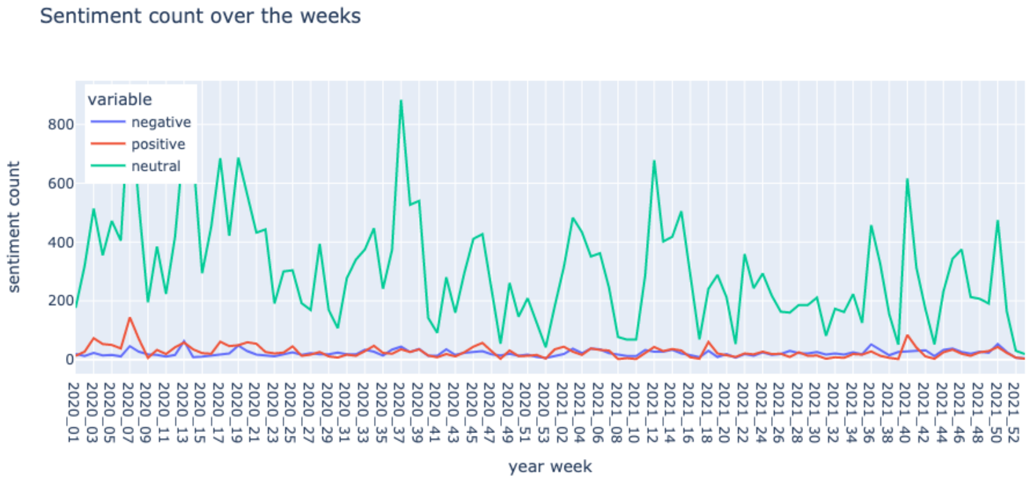

Figure 12. Sentiment counts over the weeks of the Full set, including neutral tweets

Switching gear before heading toward a different time window, figure 13 shows the counts of the tweets in the Finn set over the weeks. Unknowingly, the Full set was collected precisely to prepare for this situation where the data in the Finn set is not sufficient in itself to perform certain analysis. There is, largely, no trends in this data set when it comes to weekly window. The filter applied before analysis on both of the data sets was to filter the data set into a representative enough sample of the population. It, however, was merciful enough and only required, generally, each user to post at least once a month, which is much less granular than weekly. This is why it is not a good idea to include this Finn set in the overall timeline analysis. It would be missing also from the monthly time window.

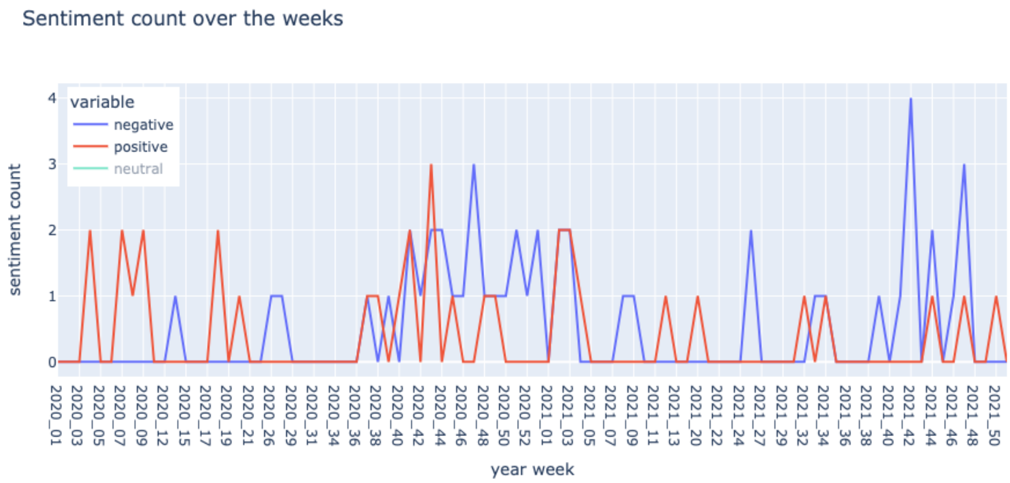

Figure 13. Sentiment counts over the weeks of the Finn set

5.2.2 Monthly

Figure 14 shows the gauges for the Full set, aggregated by months. Similar trends can be discerned, year 2021 was significantly more negative than 2020, and over several months of summer as well. Though the most interesting insight from this is rather the 3 peaks from the weekly window diagram are still, almost, peaks in monthly view. June 2020 and October 2020 were the months where the overall gauges were negative. Reflecting back to the weekly view, June 2020 especially, only had 1 peak week where negativity overtook the scene, and, yet, that became the overall sentiment of the month. There is not much that the counts, figure 15, can offer this time as its shape more or less mimicked the shape of the gauges, or rather the other way around.

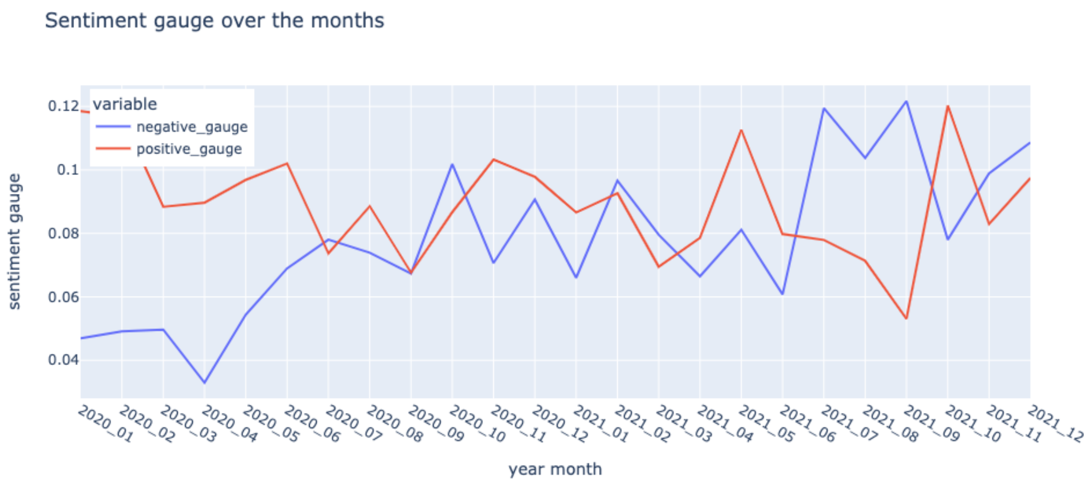

Figure 14. Sentiment gauges over the months of the Full set

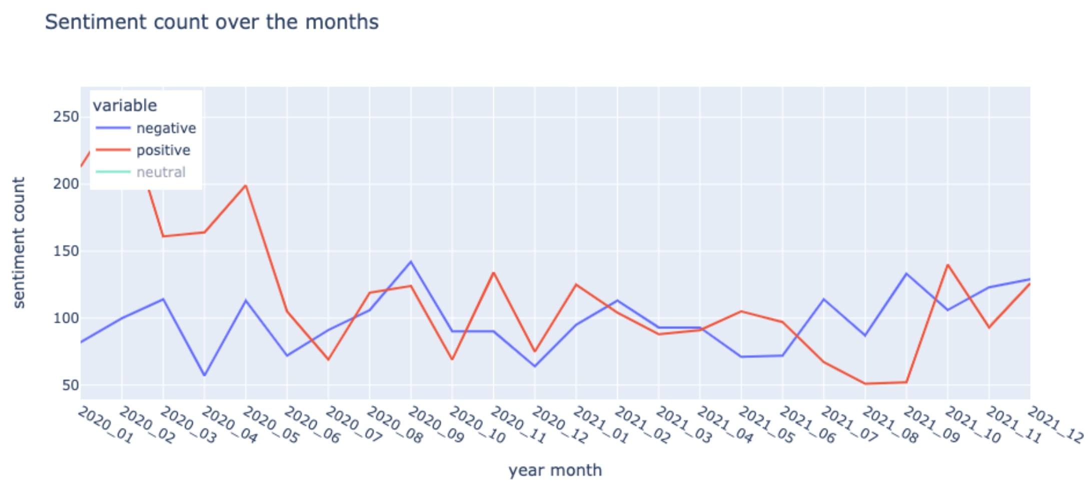

Figure 15. Sentiment counts over the months of the Full set

5.2.3 Day of Week

The x-axis on this part of the analysis will denote the day of the week by numbers from 0 to 6, corresponding to Monday through to Sunday.

Figure 16 and figure 17 show the gauges and the counts respectively of the tweets in Finn set over each day of the week for each year. Notice that the users were not overly negative on every Thursday of the week on year 2021. The reality is that the data set is just too small at some specific data points. Overall, the negative tweets from the users shifted from mainly on Monday and Tuesday, at the beginning of the week, on 2020, to the rest of the week instead. The general consensus seems to be positive on Saturday, the only exception. Though keep in mind that this is a sample of 8 users and there were only 21 tweets sent on Saturday in total (over 2 years).

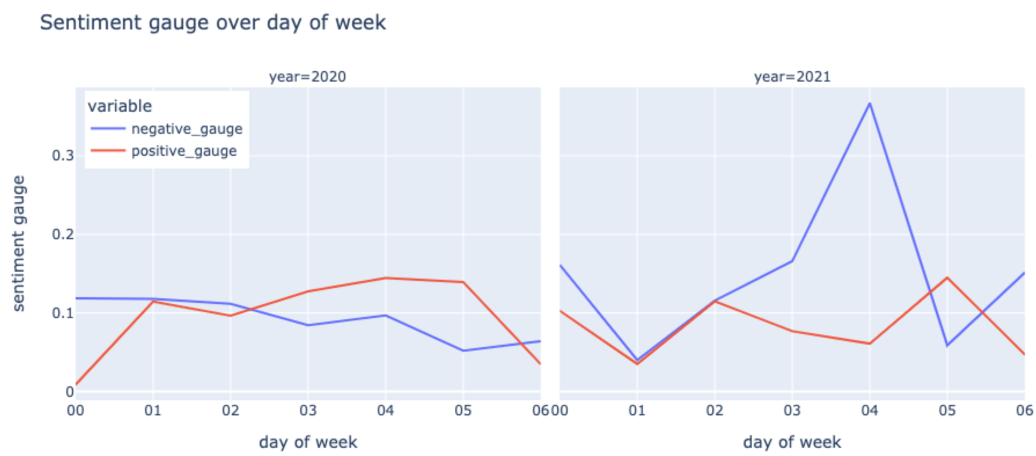

Figure 16. Sentiment gauges over the Days of Week of the Finn set

The Full set supports part of the trends displayed by the Finn set. User behaviour changed over the year, from being mostly positive through out the week to majority of the days were negative. During the year 2021, the users were largely negative over the week, except for Sunday and Tuesday.

Interestingly, while the shapes of the negative and positive tweets differ by a large amount over the 2 years, figure 19, the shape of neutral tweets virtually stayed the same, figure 20. This indicates that the usage pattern of the neutral tweets were not affected by year specific external factors. The year change only scaled the volume of the neutral tweets down by

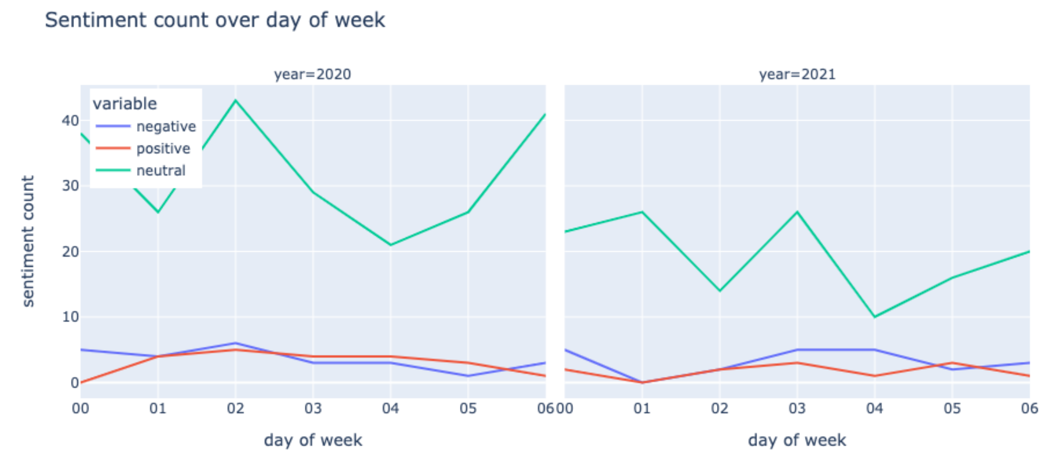

Figure 17. Sentiment counts over the days of week of the Finn set

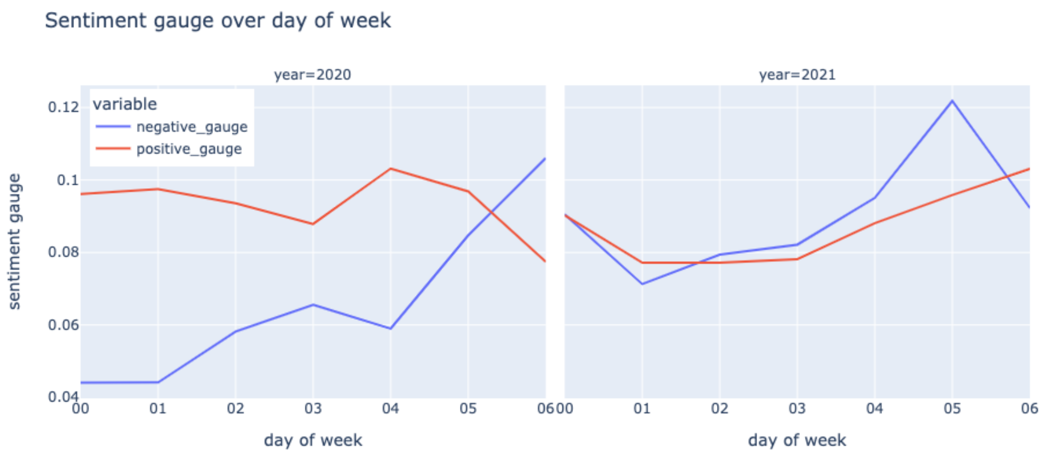

Figure 18. Sentiment gauges over the days of week of the Full set

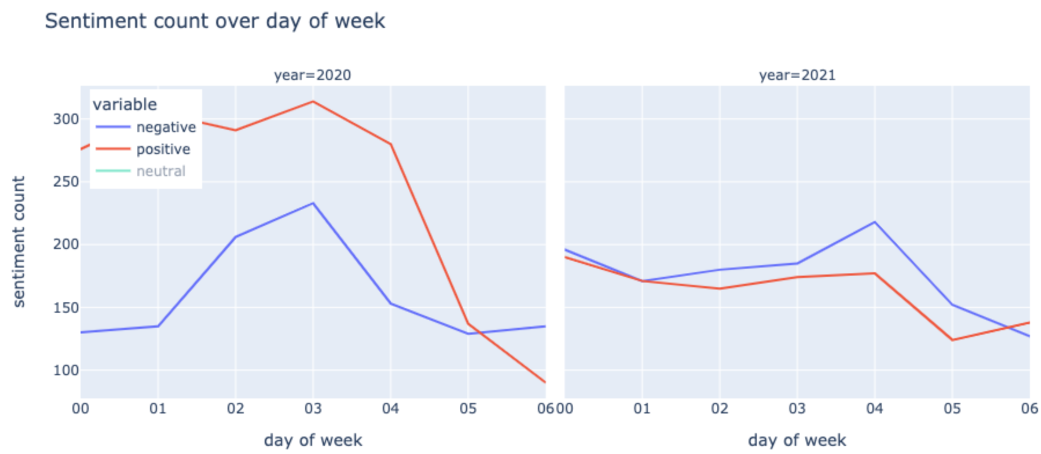

Figure 19. Sentiment counts over the days of week of the Full set

roughly 30%. And while on the subject volume, although having the usage trend changed when the year passed, negative tweets volume remained, approximately, as it was. This might suggest that, these users potentially changed their behaviour due to being exposed to different external factor. Their mental outlooks, however, persisted.

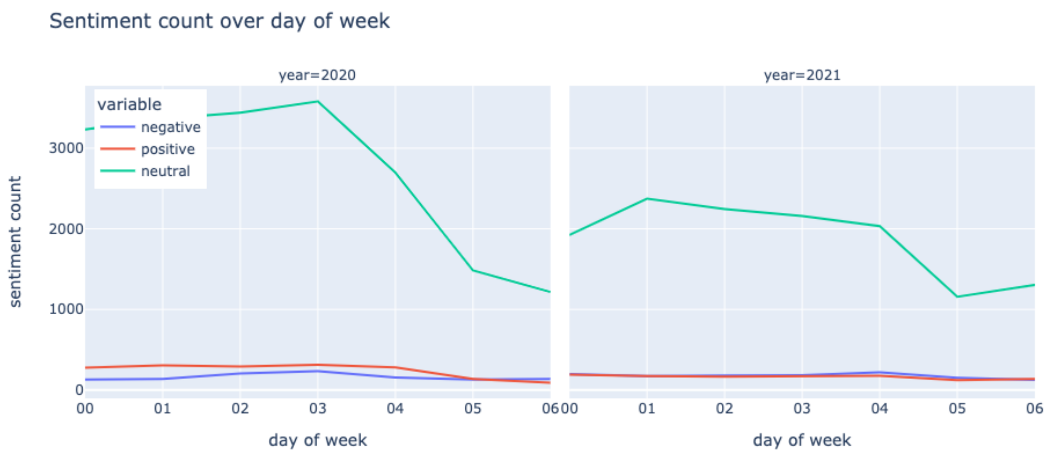

Figure 20. Sentiment counts over the days of week of the Full set, including neutral tweets

5.2.4 Time of Day

Lastly, this is the time of day when the tweets are posted. Figure 21 shows the gauges for the Finn set. This is again a difficult analysis to be made on the Finn set as the filter applied allowed only users with more than 24 tweets to remain. There were still a quite significant amount of the tweets left, 441 tweets, but there were only 8 users in total. With only this many users, although they contributed a large, enough, amount of tweets, there are only so much varieties in terms of usage patterns that they can offer. A few hours of day only have 1 tweets through out the 2 years period, figure 22. There are too many instances where there were no tweets for at least one of the sentiments which rendered the analysis extremely difficult. For instance, on the 15th hours of the day on 2021, all the tweets were considered to be neutral. However, looking solely at the sentiment gauges, the impression is that the sentiment at the 15th hours of the day is relatively negative. This could be caused by the tweets being slightly more negative than normal while still being classified as neutral. It is better to not draw any conclusion regarding hourly behaviour base on the Finn set. At this point, the Finn set serves as a learning experience instead for when new data needs to be acquired.

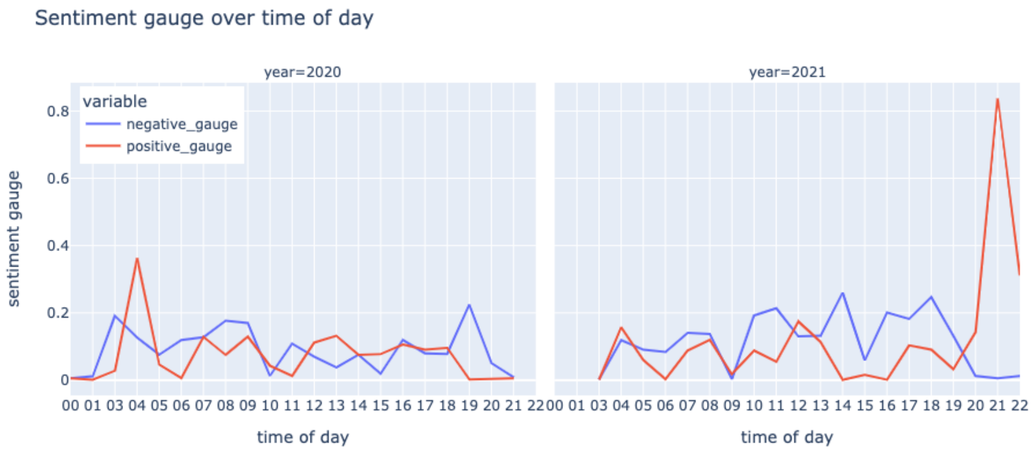

Figure 21. Sentiment gauges over the time of day of the Finn set

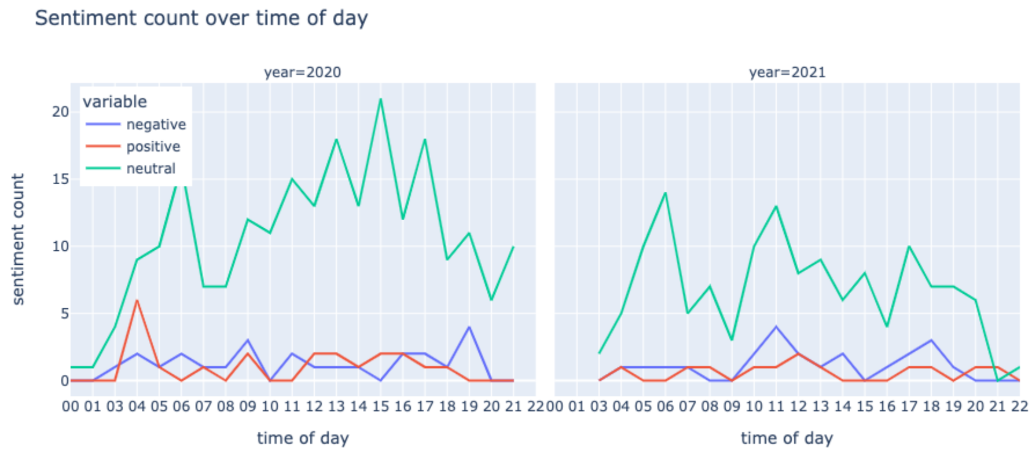

Figure 22. Sentiment counts over the time of day of the Finn set

Switching to the Full set instead, the usage trends are much clearer here. Positivity through out the day do not seem to change much, figure 23. There are some fluctuations at the 2 ends of the cycle, however, that might just be because of the low volume of the tweets at the beginning and end of day. Negativity, on the other hand, is showing a clear hyperbole shape pattern. The volume of the tweets shows an opposite trends for the negative tweets instead, figure 24. This volume of negative tweets was neutralised by the huge amount of neutral tweets through out the most active hours of the day. Similar to the day of week view, the time of day view also shows the same usage volume trend for neutral tweets over the 2 years, figure 25. The decrease in volume of neutral tweets on 2021 also contributes to the rise in negativity earlier on a day than before, despite the negative tweets volume did not change much between the years.

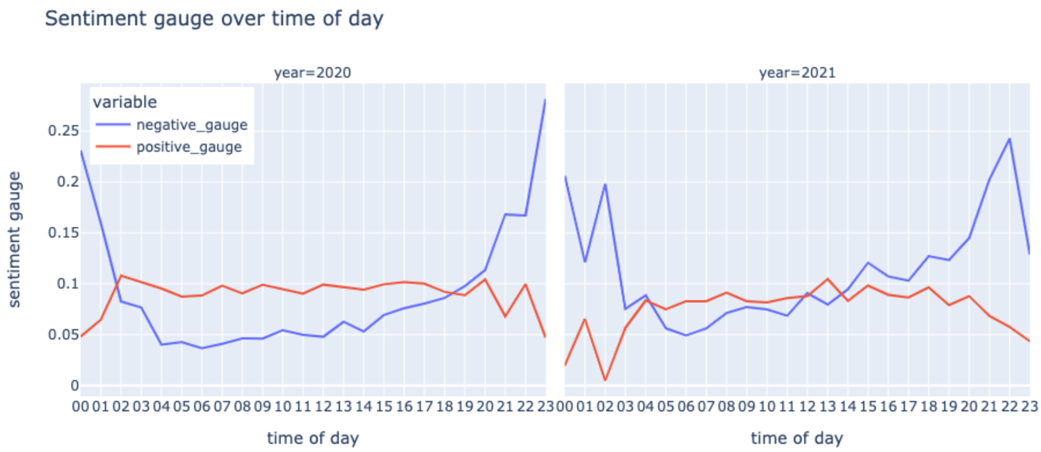

Figure 23. Sentiment gauges over the time of day of the Full set

5.3 Discussions

5.3.1 Monthly trends and Covid-19

Looking back at the trends of the gauges on the Full set, there were some, albeit small, peaks where the sentiment is clearly negative overall. These peaks matches the significant holidays of the year, namely the summer holiday and Christmas holiday. In supporting of this view, figure 26 shows the negative gauge of the Full set plotted together with the number of new Covid cases in the World. Covid is one of the short version used to refer to the SARS-CoV-2 virus that causes the pandemic from late 2019 and early 2020 (Coronavirus disease COVID-19). At the point of this study, the pandemic is still very

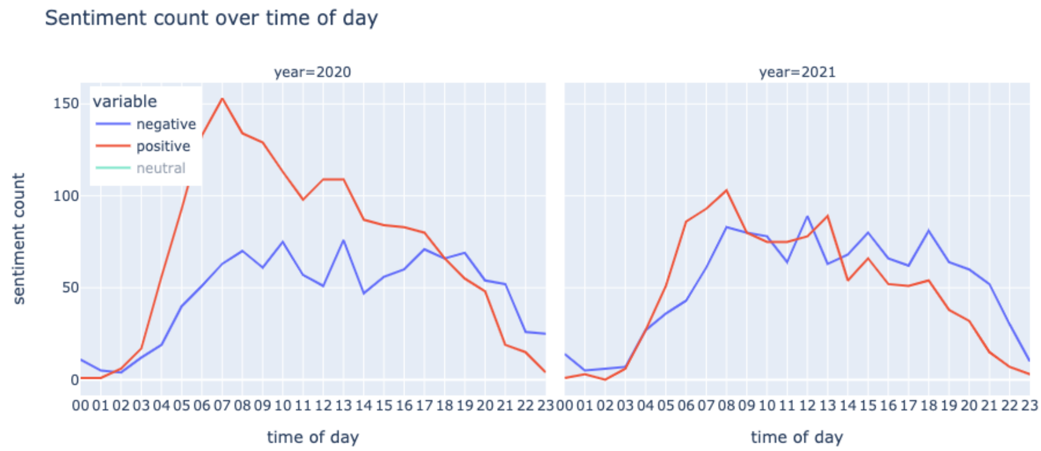

Figure 24. Sentiment counts over the time of day of the Full set

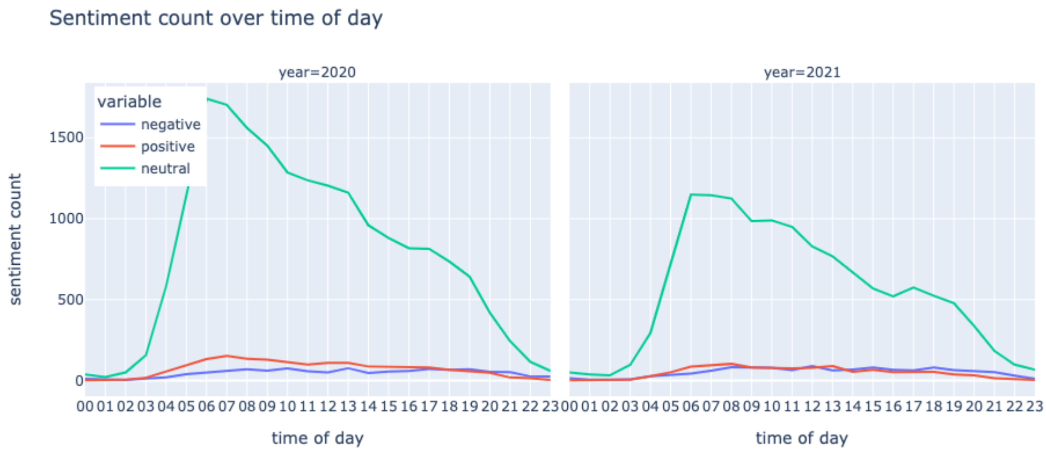

Figure 25. Sentiment counts over the time of day of the Full set, including neutral tweets

much relevant, and the World is slowly accepting a new "normal". During these holiday periods, the number of new Covid cases surged and maintained the peaks for, often, more than just a month. The negative gauge seemed to move, although not fully, with the number of new cases found.

Python's package Scipy (SciPy 2022) was used to calculate the correlation between the 2 variables. The null hypothesis set for this case would be that the negativity of the tweets is not correlated to the number of new Covid cases. From the Scipy package, Pearson correlation coefficient was used, for when the distributions are relatively linear, and Spearman correlation coefficient was also used, in case the distributes are not linear. From figure 26, the distributions of both metrics are slightly linear and on the positive coefficient, but it cannot hurt to consider non-linear as well. Pearson's method outputted a correlation coefficient of 0.657 and p-value of 0.00049, and Spearman's method outputted a correlation coefficient of 0.616 and p-value of 0.0014. Such tiny p-values suggested that there is enough evidence to reject the null hypothesis. Which means there is a positive correlation between the number of new Covid cases and the negative gauge of the tweets. This is not to say that the new Covid cases caused the tweets to be more negative, however, only that they follows the same trend.

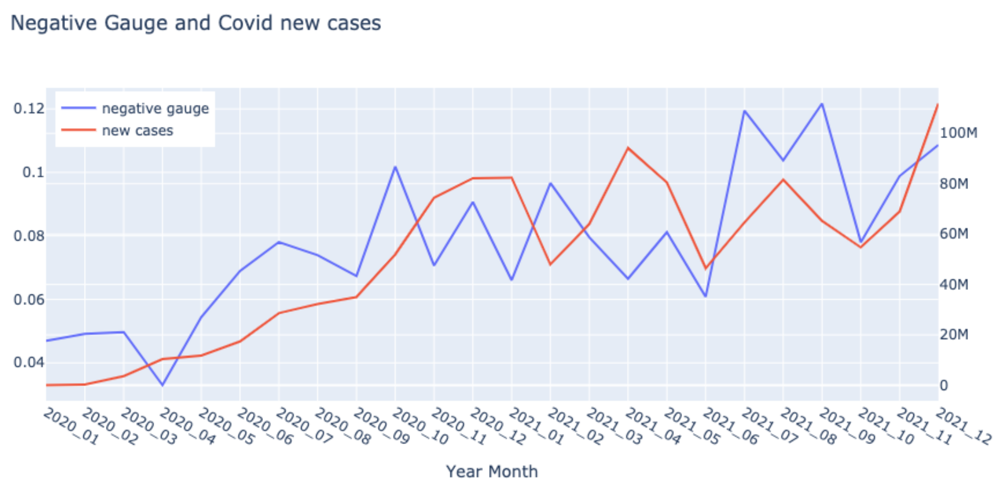

Figure 26. Negative gauge versus Covid 19 new cases monthly

5.3.2 Some what consistent tweets pattern

Although this study set out to study the behaviour of those that might have mental disorder, more specifically depression, the consistency of the neutral tweets pattern is nevertheless interesting. Putting aside the Finn set, the Full set suggested a consistent pattern over, not only, the day, but also, the days of the week, figure 25 and 20. Keep in mind that this consistency happened across 2 different views despite having little to no connections between weekly and daily behaviour.

One explanation might be that this is how the usage pattern of the mentally healthy users look like. Mentally healthy people might be less affected by external factors and keep their usage patterns across the years. This hypothesis is a bit problematic because figure 24 also shows a very consistent counts of tweets for negative and positive tweets. Although, this consistency only appears on the time of day view and not on days of week view for positive and negative tweets.

Another way to look at this is that the usage patterns are actually consistent for the time of day distributions. This would fit very well with the explanation earlier regarding the circadian rhythm that all human beings are subjected to (Kreitzman & Foster 2021). This means that the pattern is suggestive and can be used for signalling and identify, for example, users who post very negative tweets at very late hours in the day. Although, keep in mind that the limitation still stands, posting negative tweets does not necessary mean the users are suffering some form of mental disorder.

5.3.3 Speculations on usage patterns

This part of the text is an attempt to reason the shape of the time of day negative usage pattern.

In 2018, Shin & Kim did a research to study the relationship between having consistent good night sleeps and life satisfaction. They found that those who have high quality sleeps are more likely to have higher life satisfaction. Though it is worth mentioning that the relationship could be bidirectional, as a healthy mind can lead to a better sleep. A year after, Zhao et al. observed that individuals, especially younger individuals, with insomnia symptoms or having shorter sleep duration, have lower level of happiness. In 2021, Coiro et al. reported a relationship between mental health symptoms and sleep quality, in addition to Covid. These studies illustrated that sleep quality is highly important to happiness level.

Steering back to figure 23, 24, and 25, the negative gauge of the tweets are higher late at night. Aside from the high level of negativity, there is a decrease in volume of positive and neutral tweets, which removed the "balancer" on the data set and drove the gauge higher. Similar to the previous studies, this can be interpreted as healthier minds were mostly sound asleep and the few individuals who were still awake are more likely to have lower level of happiness. This lead to the overall higher negativity observed at late hours. With that said, this is not fully backed by the data as the amount of negative tweets also dropped to close-zero, although it never reached zero like positive tweet counts.

6 CONCLUSIONS

6.1 Main findings

Through out this study, there are 3 mains learning points. The first and second are about usage patterns and the factors affecting them. The third, and last, point is regarding data collection.

First, there was not a clear usage pattern for negative tweets that can be discerned from the data when aggregated per week, even from the Full set. There was, however, a somewhat recognisable tendency of the tweets to be more negative over the holiday periods, especially when looking at the monthly aggregation view. Moreover, the overall negativity seemed to rise toward the end of the data period. This trend is in positive correlation with the number of new cases of Covid, which is also rising in the 2 years period.

Second, negative tweets are more visible at later time of the day. This is likely to not only caused by the negativity of the tweets themselves but also the decrease in amount of the other 2 sentiments. Interpretation on this is that healthier minds tend to have better sleep quality as well, which includes having an appropriate sleep duration.

And last, data collection in general is a very important factor for a project. Even more so in this case, considering the fact that this study were targeted at a rather small population, Finnish speaking (and living in Finland). Therefore it is fortunate that the data was collected upfront and adjustment was made at an appropriate time to collect a larger set of data, the Full set. Moreover, the data collection of this study was implemented so that it favours re-usability and empowers future studies; when the time came to collect more data for this study, the initial designed proved its worth by cutting the additional data collection time to almost nothing, aside from processing time.

6.2 Limitations

There are several limitations to this study. First of all, there is no way to determine whether the tweets with these negative sentiment are from individuals who are suffering from mental health issues. Even though the sentiment can be negative, the contexts are missing from the label. It might help if the tweets were gone through one by one as human emotions are complex and there are nuances that cannot be explained by just labelling them as negative, neutral, or positive. However, the contexts for such complex emotions are still missing, regardless of how the tweets were processed. Due to this specific limitation, the results presented a version of this limitation as the disclaimer and merely talked about the gauges of the tweets, rather than making conclusions on the mental states of the users.

Second, data collection had a major flaw of targeting a minority. Finnish is not the most spoken language in the World and filtering them further with keywords only scaled the size of the data set down even further. Due to this, the primary data set, the Finn set, the set that this study was set out to study were demoted into a secondary set instead. Moreover, even with the Full set standing in as the primary data set, the Full set is still a representation of a minority. The data was still small with the 2 years period. If this study was done again but with a much wider range of years, then the volume might be sufficient, if the target is still the Finnish speaking users. Although, widening the time frame which the data is collected means introducing more unknown variables into the mix as yearly situation might affect the sentiment of the tweets. An alternative would be to study a larger population, i.e. the Nordic countries instead of only the Finnish speaking population.

The last major limitation of this study is the way the analysis was carried out. Mental health is a rather personal and specific subject. Analysing it from a collective point of view is not a very good way to approach this issue. Rather, only if the data volume is large enough, and if time allows, analysis should be done on individual user level. With this alternative approach, the conclusion regarding usage pattern can have more certainty. Going one step further, because there are actually ways to identify the real people behind these accounts, if permission from them are acquired, actual information regarding their mental health can potentially be a data point.

6.3 Re-usability

Aside from the analysis, the largest asset that this study produced has to be the repository which was used to collect the data (Cai 2021). As described in the text, there were unexpected turns happened during the analysis that rendered the Finn set rather obsolete. With this repository, an extremely fast turn around was achieved and analysis was not delayed for long. This repository enables also future studies to just grab the data, from Twitter, and start analysis immediately instead of spending time figuring out the twists and turns of the Twitter API, assuming there is sufficient access to the API.

6.4 Future Studies

Future studies can start by solving the limitations of this study. The problem of uncertainty on the mental health issue can only be solved by identifying the users, which is the most difficult option but this would bring the most value. The problem of data volume can be solved by expanding the data set to either a wider time frame or a wider demographic. And lastly, making the analysis more personal or specific would provide much more value in terms of usage patterns.

Aside from the limitations, there are a few other studies that can be done based on this one. One of them is, this study's "side effect", the repository, can be used to build any Twitter data set as long as it can be described in terms of keywords, language, geolocation, and time frame. Any studies that are interested in the Twitter texts for NLP purpose can reuse this study to its advantage.

Another one is the fact, not limitation because this was never the objective of this study, that this study only viewed a part of the Twitter events, tweets. There are a lot more to Twitter than just tweets, namely liking, sharing, and tagging. And these are only the active side of Twitter. What would be the relationship between mental health and passive usage of Twitter, i.e. merely scrolling through the feed and consume the contents?

REFERENCES

Andersson, Gerhard. 2022, *Internet-based psychotherapies*, pp. 377–394.

Aristotle. 328 BC, *Political Book 1*, chap. 1253a.

Baym, Nancy. 2010, *Personal Connections in the Digital Age*.

Bevan, Jennifer; Gomez, Ruth & Sparks, Lisa. 2014, Disclosures about important life events on Facebook: Relationships with stress and quality of life, *Computers in Human Behavior*, vol. 39, , p. 246–253.

Cai, Bao. 2021, *Tweet Search Repository*. Available: https://github.com/BaoCaiH/tweet_search.

Chen, Wenhong & Lee, Kye-Hyoung. 2013, Sharing, Liking, Commenting, and Distressed? The Pathway Between Facebook Interaction and Psychological Distress, *Cyberpsychology, behavior and social networking*, vol. 16, .

Coiro, Mary; Asraf, Kfir; Tzischinsky, Orna; Hadar-Shoval, Dorit; Haddad, Lubna & Wolfson, Amy. 2021, Sleep quality and COVID-19-related stress in relation to mental health symptoms among Israeli and U.S. adults, *Sleep Health*, vol. 7, .

2022, *Coronavirus disease (COVID-19): How is it transmitted?* Last accessed on 29 March 2022. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted.

Finland, Statistics. 2022, *Population and Society*. Last accessed on 28 February 2022. Available: https://www.stat.fi/tup/suoluk/suoluk_vaesto_en.html.

Hauhio, Iikka. 2021, *FinBERT fine-tuned with the FinnSentiment dataset*. Last accessed on 19 January 2022. Available: https://huggingface.co/fergusq/finbert-finnsentiment.

2022, *Huggingface, transformer*. Last accessed on 18 January 2022. Available: https://github.com/huggingface/transformers.

Kajava, Kaisla. 2022, *Cross-lingual Sentiment Preservation in Binary and Multidimensional Classification*. Last accessed on 20 January 2022. Available: https://github.com/cynarr/sentiment-analysis.

Kemp, Simon. 2020, *DIGITAL 2020: OCTOBER GLOBAL STATSHOT*. Last accessed on 17 March 2022. Available: https://datareportal.com/reports/digital-2020-october-global-statshot.

Krausz, Michael. 2017, Social Media and E-mental Health, *European Psychiatry*, vol. 41, , p. S7.

Kreitzman, Leon & Foster, Russell. 2021, *Very Short Introduction to Circadian Rhythms*, 26-36 p..

Labrague, Leodoro. 2014, Facebook use and adolescents' emotional states of depression, anxiety, and stress, *Health Science Journal*, vol. 8, , pp. 80 – 89.

Lahoud, Ashley; Gladstone, Theresa; Clark, Sydney & Flessner, Christopher. 2022, *Collecting data from Internet (and other platform) users for mental health research*, pp. 63–80.

Lin, Liu; Sidani, Jaime; Shensa, Ariel; Radovic, Ana; Miller, Elizabeth; Colditz, Jason; Hoffman, Beth; Giles, Leila & Primack, Brian. 2016, Association between Social Media Use and Depression among U.S. Young Adults, *Depression and anxiety*, vol. 33, .

Lindén, Krister; Jauhiainen, Tommi & Hardwick, Sam. 2020, *FinnSentiment - A Finnish Social Media Corpus for Sentiment Polarity Annotation*. Available: https://arxiv.org/pdf/2012.02613.pdf.

Lup, Katerina; Trub, Leora & Rosenthal, Lisa. 2015, Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed, *Cyberpsychology, behavior and social networking*, vol. 18, , pp. 247–252.

Multiprocessing. Last accessed on 20 March 2022. Available: https://docs.python.org/3/library/multiprocessing.html.

Pickles, James O. 1982, An Introduction to the Physiology of Hearing.

Rizzolatti, Giacomo & Craighero, Laila. 2004, The Mirror-Neuron System, *Annual review of neuroscience*, vol. 27, , pp. 169–92.

2022, *SciPy*. Last accessed on 29 March 2022. Available: https://scipy.org/.

2022, *Search Tweets*. Last accessed on 23 January 2022. Available: https://developer.twitter.com/en/docs/twitter-api/tweets/search/api-reference.

2022, *Sentimentor*. Last accessed on 20 January 2022. Available: https://github.com/Helsinki-NLP/sentimentator.

Shabbir, Syed Abdul; Gabarron, Elia; Lau, Annie & Househ, Mowafa. 2016, An Introduction to Participatory Health Through Social Media, pp. 1–9.

Shin, Ji-eun & Kim, Jung. 2018, How a Good Sleep Predicts Life Satisfaction: The Role of Zero-Sum Beliefs About Happiness, *Frontiers in Psychology*, vol. 9, .

Strauss, John; Zhang, Jasmine; Jarrett, Madeleine; Patterson, Beth & Van Ameringen, Michael. 2022, *Apps for mental health*, pp. 395–433.

Tana, Jonas; Kettunen, Jyrki; Eirola, Emil & Paakkonen, Heikki. 2018, Diurnal Variations of Depression-Related Health Information Seeking: Case Study in Finland Using Google Trends Data, *JMIR Mental Health*, vol. 5, , p. e43.

Twenge, Jean; Joiner, Thomas; Rogers, Megan & Martin, Gabrielle. 2017, Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time, *Clinical Psychological Science*, vol. 6, , p. 216770261772337.

Twitter. 2012, *Twitter turns six*. Last accessed on 28 February 2022. Available: https://blog.twitter.com/official/en_us/a/2012/twitter-turns-six.html#.

2022, *Twitter Academic*. Last accessed on 17 December 2021. Available: https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api#v2-access-level.

Vannucci, Anna; Flannery, Kaitlin & Ohannessian, Christine. 2017, Social Media Use and Anxiety in Emerging Adults, *Journal of Affective Disorders*, vol. 207, , pp. 163–166.

Verhagen, Simone; Os, Jim & Delespaul, Philippe. 2022, *Ecological momentary assessment and other digital technologies for capturing daily life in mental health*, pp. 81–108.

Waller, Bridget; Cray, James & Burrows, Anne. 2008, Selection for Universal Facial Emotion, *Emotion (Washington, D.C.)*, vol. 8, , pp. 435–9.

Wongkoblap, Akkapon; Vadillo, Miguel & Curcin, Vasa. 2022, *Social media big data analysis for mental health research*, pp. 109–143.

2022, *XED*. Last accessed on 20 January 2022. Available: https://github.com/Helsinki-NLP/XED.

Young, K.S. & Rogers, R.C. 1998, The relationship between depression and Internet addiction, *Cyberpsychology and Behavior*, vol. 1, , pp. 25–28.

Zhao, Sheng; Wang, Man Ping; Viswanath, Kasisomayajula; Lai, Agnes; Fong, Daniel; Lin, Chia-Chin; Chan, Sophia & Lam, Tai. 2019, Short Sleep Duration and Insomnia Symptoms were Associated with Lower Happiness Levels in Chinese Adults in Hong Kong, *International Journal of Environmental Research and Public Health*, vol. 16, , p. 2079.

APPENDIX A

List of keywords

masennus

masennus oireet

masennustesti

masennus testi

synnytyksen jälkeinen masennus

masennus hoito

ahdistus

keskivaikea masennus

psykoottinen masennus

lapsen masennus

vakava masennus

depression

vaikea masennus

nuorten masennus

raskaus masennus

väsymys

lievä masennus

masennuslääkkeet

mielenterveys

nuoren masennus

masennus blogi

itsemurha

masennus keskustelu

psykoosi

masennuksen hoito

masennus itsehoito

krooninen masennus

kaksisuuntainen mielialahäiriö

depression test

depression symptoms

manic depression

postpartum depression

crippling depression

clinical depression

high functioning depression