# Insights Extraction

## Main Questions

1. What is the trend? What can we tell from it?
2. Why did the trend occur? Is it possible to infer from data? – where relevant

## Insights

Generally, we observe that frequency of post relevant to mental health have increased significantly with covid, as seen from the upwards trend in *figure 1*. From *figure 2*, post frequency peaked in second quarter of the year for both 2020 and 2021, during April and May respectively.

Chart, histogram

Description automatically generatedChart, line chart

Description automatically generated

Figure 1 Overall Reddit Post Frequency against Time Figure 2 Monthly Reddit Post Frequency

Interestingly, these peaks align with the drastically increasing new community cases each week in April 2020 and May 2021, as shown in *figure 3*. That said, such similarity only suggest correlation not causation, but it hints that the covid situation is influential on redditors’ behaviors and mental state.

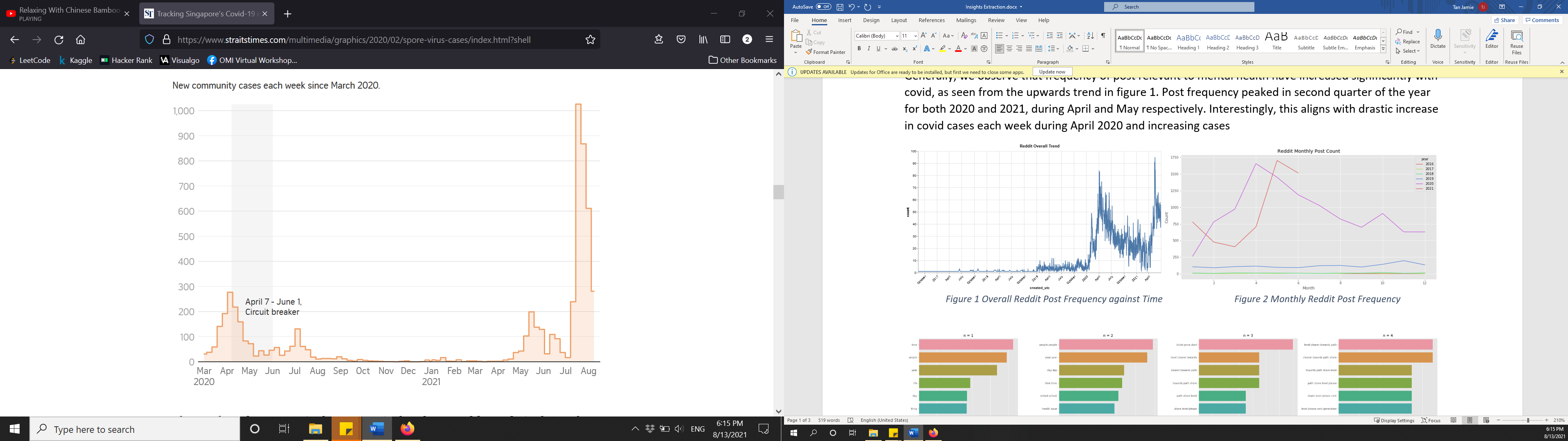


Figure 3 New community cases each week since March 2020 (source: [Straits Times](https://www.straitstimes.com/multimedia/graphics/2020/02/spore-virus-cases/index.html?shell))

Borrowing *figure 1* illustrating *frequency of post against time*, the number of relevant posts in mid-covid is significantly larger than pre-covid period. It is thus no surprise that the values on axes of ngrams representing mid-covid *(figure 5)* timeframe are of greater order.

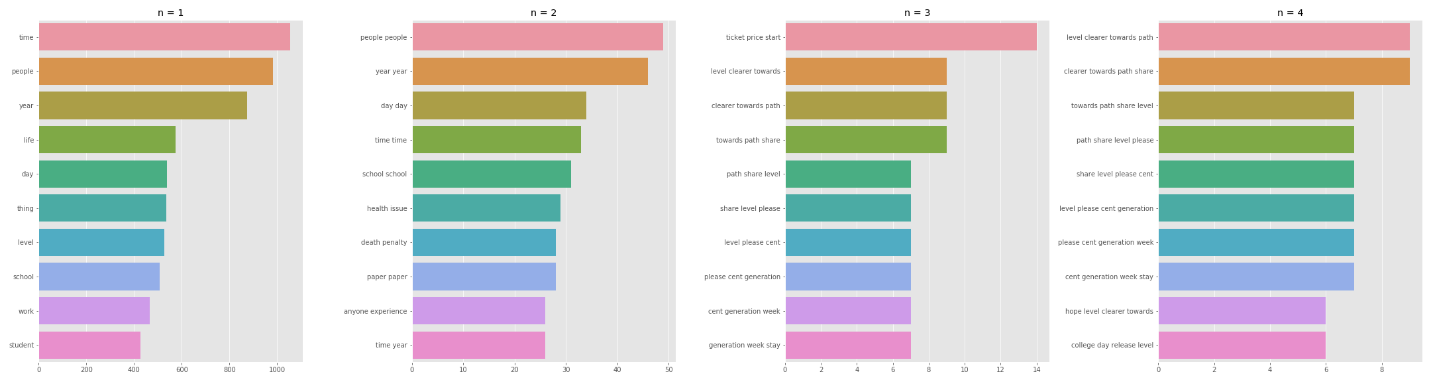


Figure 4 Pre-covid Ngrams

Chart

Description automatically generated

Figure 5 Mid-covid Ngrams

In general, pre and mid covid *1-gram* are similar *(figure 4 & 5)*. However, some new addition to the list during the pandemic include “covid” and “help”. Further probing using [LDA](C://Users/20jam/Documents/GitHub/omdena-singapore-covid-health/src/tasks/task-3-data-exploration-and-analysis/reddit/lda_help.html) raise topics like work and school to associate with “help”. It aligns with the increase in frequency of “school”, “student” and “work” during covid. These suggest that the mental states of working adults and schooling learners are most affected by covid. This is also supported by *4-gram* (mid-covid) that contain terms mostly related to examination, which is a stark contrast to that of pre-covid.

Unsurprisingly, covid-19 is highly discussed, as inferred from “circuit breaker”, “covid case” and “community case” in *figure 5 2-gram*. “Part time” also entered the top-10 used phrase as depicted in *figure 5 2-gram* and “part time job” in *figure 5 3-gram*. Interestingly, the sentiment related to these 2 phrases is highly polarized. Redditors are generally very negative or very positive about it, although most seem positive as shown in *figure 7*. Furthermore, covid may have catalyzed more exchange within the community, as inferred from the addition of “share experience”. These interactions could account for the increasingly positive sentiments despite covid, as observed in *figure 6*.

Chart

Description automatically generated Chart, histogram

Description automatically generated

Figure Overall Reddit Sentiment across Time Figure 7 Sentiment Distribution

\*leaving word cloud out because it is not very informative