

Report on Analysis of Tweets Related to Covid-19

1. Data Set

The data set was prepared by crawling tweets from Twitter containing keywords like - "#Covid_19", "#coronavirusindia", "#CoronaChainScare", "#StayHomeStaySafe", "#CoronavirusPandemic", "#CoronaVirusUpdate". The tweets corpus consisted of 41,402 tweets which were crawled from 27th February, 2020 to 30th March, 2020. The data set consisted of the following fields of information- tweet content, creation date, screen name and retweet count.

2. Week-wise distribution of Tweets

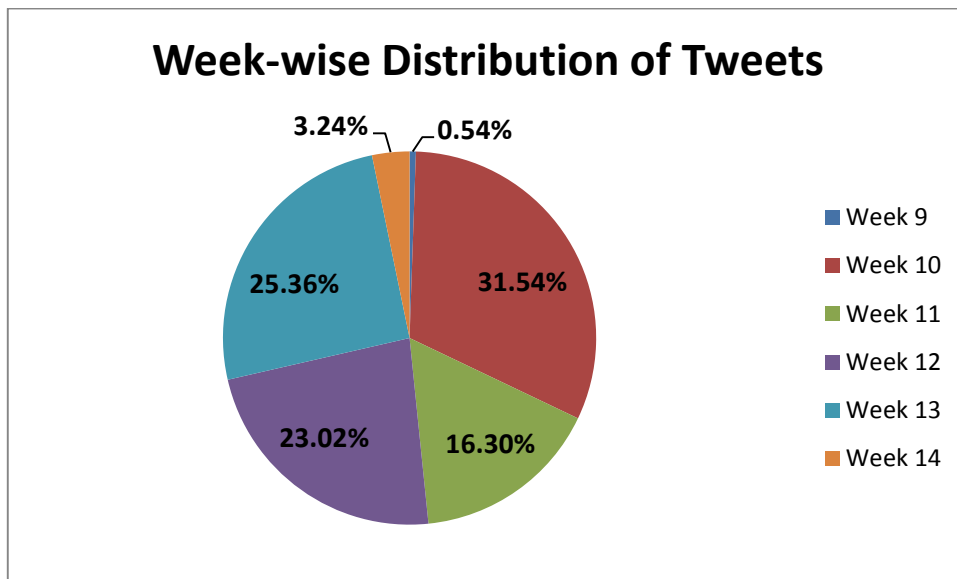


Fig 1. Week-wise distribution of Tweets

The results are based upon the dataset containing tweets from 27th February, 2020 (week 9) to 30th March, 2020 (week 14). The low percentage of tweets in week 9 can be attributed to the fact that the crawler started crawling tweets from 27th February, 2020 which was the second last day of week and therefore very few tweets could be crawled. Similarly, 30th March, 2020 was the second day of week 14 so the data of week 14 is provisional.

3.2 Word Cloud Representation upon tweets by @PMOIndia



Fig. 3: Word Cloud Representation of Tweets

From the word cloud representation in Fig. 3, it can be deduced that the most frequent words apart from “covid” and “coronavirus” are “mannkibaat”, “lockdown”, “quarantine”, “donation” and so on.

3.4 Word Cloud Representation upon tweets by @BorisJohnson



Fig. 5: Word Cloud Representation of Tweets

From the word cloud representation in Fig. 5, it can be deduced that the most frequent words are “press”, “conference”, “downingstreet”, “chancellor”, “protect” and so on.

3.5 Word Cloud Representation upon tweets by @WHO



Fig. 6: Word Cloud Representation of Tweets

From the word cloud representation in Fig. 6, it can be deduced that the most frequent words are “press”, “conference”, “downingstreet”, “chancellor”, “protect” and so on.

4. Week-Wise Frequency Variation of Certain Terms

4.1 Frequency Variation of Most Occurring Terms

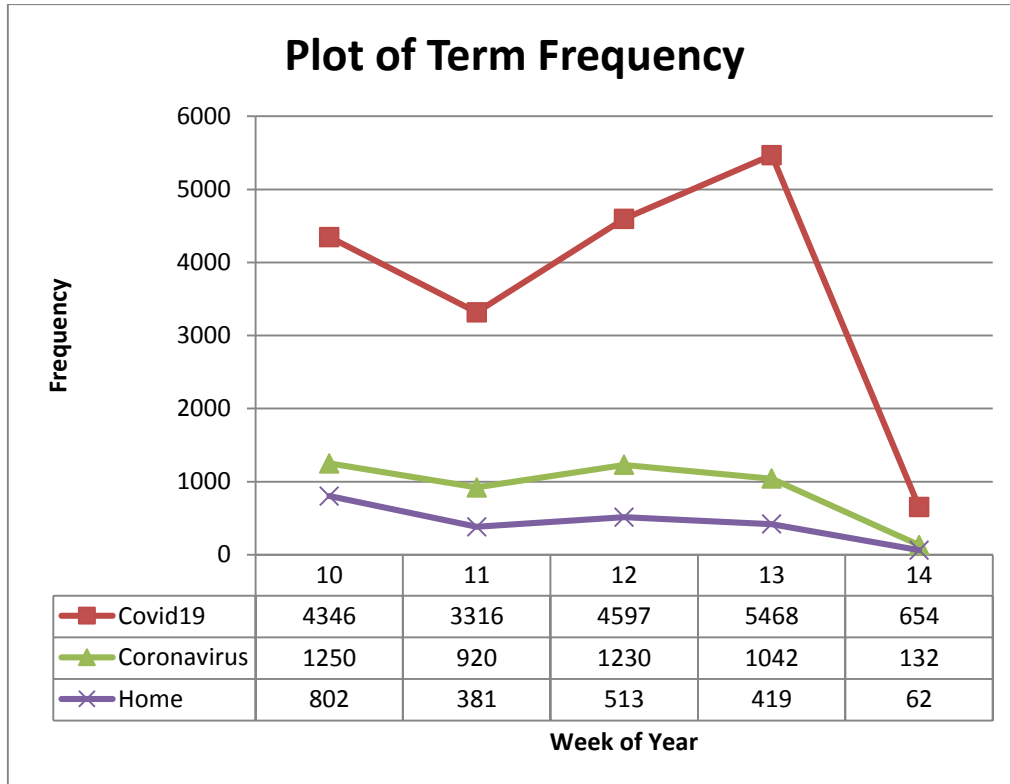


Fig. 7: Frequency Variation of Most Occurring Terms

In Fig. 7, the variation of frequency of most occurring terms like “covid19”, “coronavirus” and “home” has been presented. It has been observed that “covid19” is most frequent followed by “coronavirus” and “home”. For “covid19” there has been a rising trend except for week 11 and week 14. Also, both terms “coronavirus” and “home” follow a similar trend of frequency variation.

4.2 Frequency variation of Relevant Terms

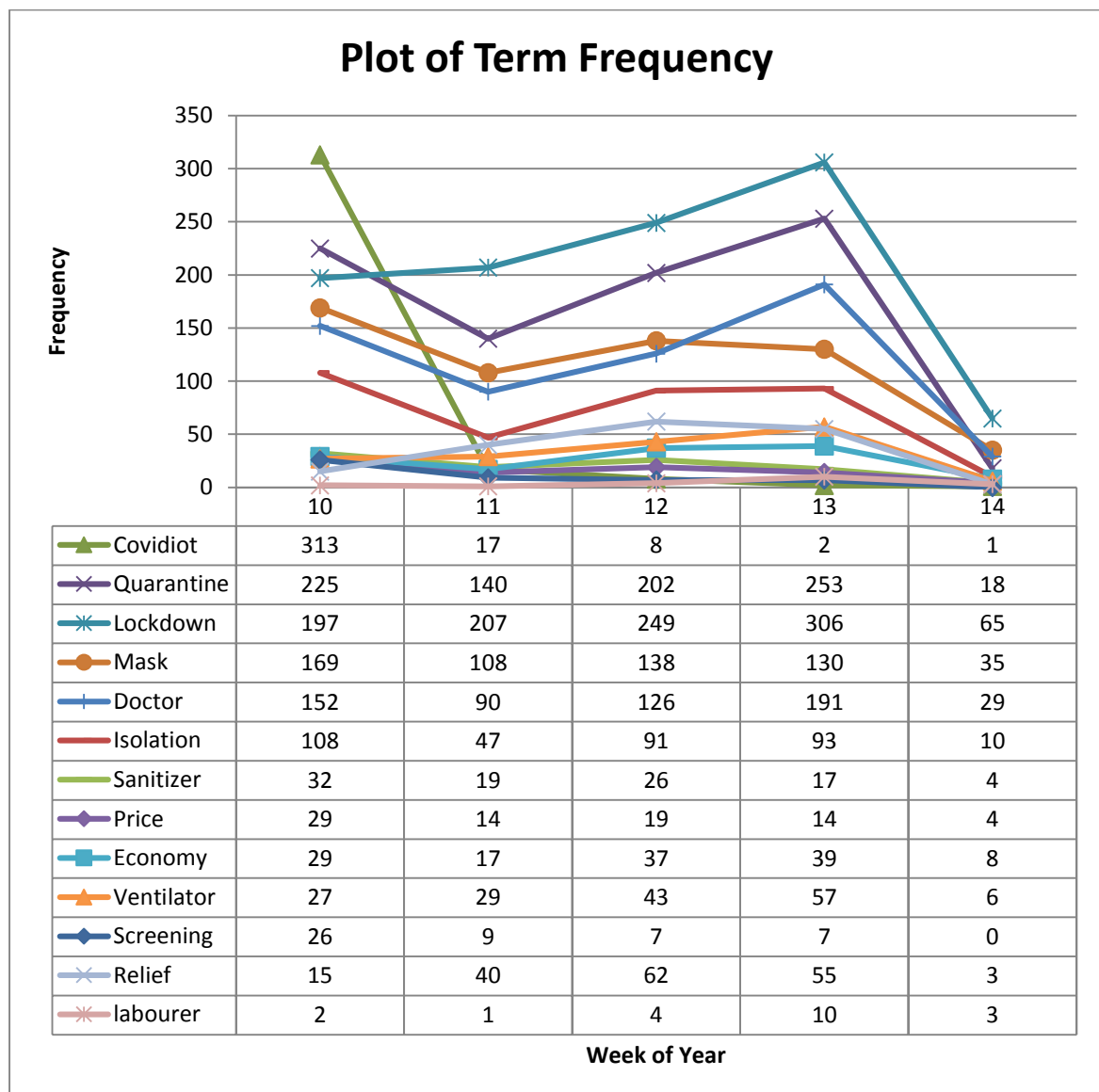


Fig8. Frequency variation of Relevant Terms

In Fig. 8, the week-wise frequency variation of some relevant terms has been presented. It is observed that the most frequent terms are “lockdown”, “quarantine”, “mask”, “doctor” and “isolation”. The frequency of term “covidiot” was highest during week 10 but declined drastically over the following weeks. Also, it can be seen that for all the terms, the overall trend for frequency variation is similar.

5. Week-Wise Sentiment Analysis

To understand the sentiment of people regarding the Covid-19 outbreak, the sentiment analysis upon the tweets have been performed using Valence Aware Dictionary and sEntiment Reasoner (VADER). The sentiment analysis has been performed upon the overall data set and also upon tweets specific to certain profiles of interest like PMO India, RBI, CDC, US President Donald Trump, UK Prime Minister Boris Johnson and WHO. The tweets have been categorized as positive, negative, emotional and factual.

5.1 Week-Wise Sentiment Analysis upon Entire Data Set

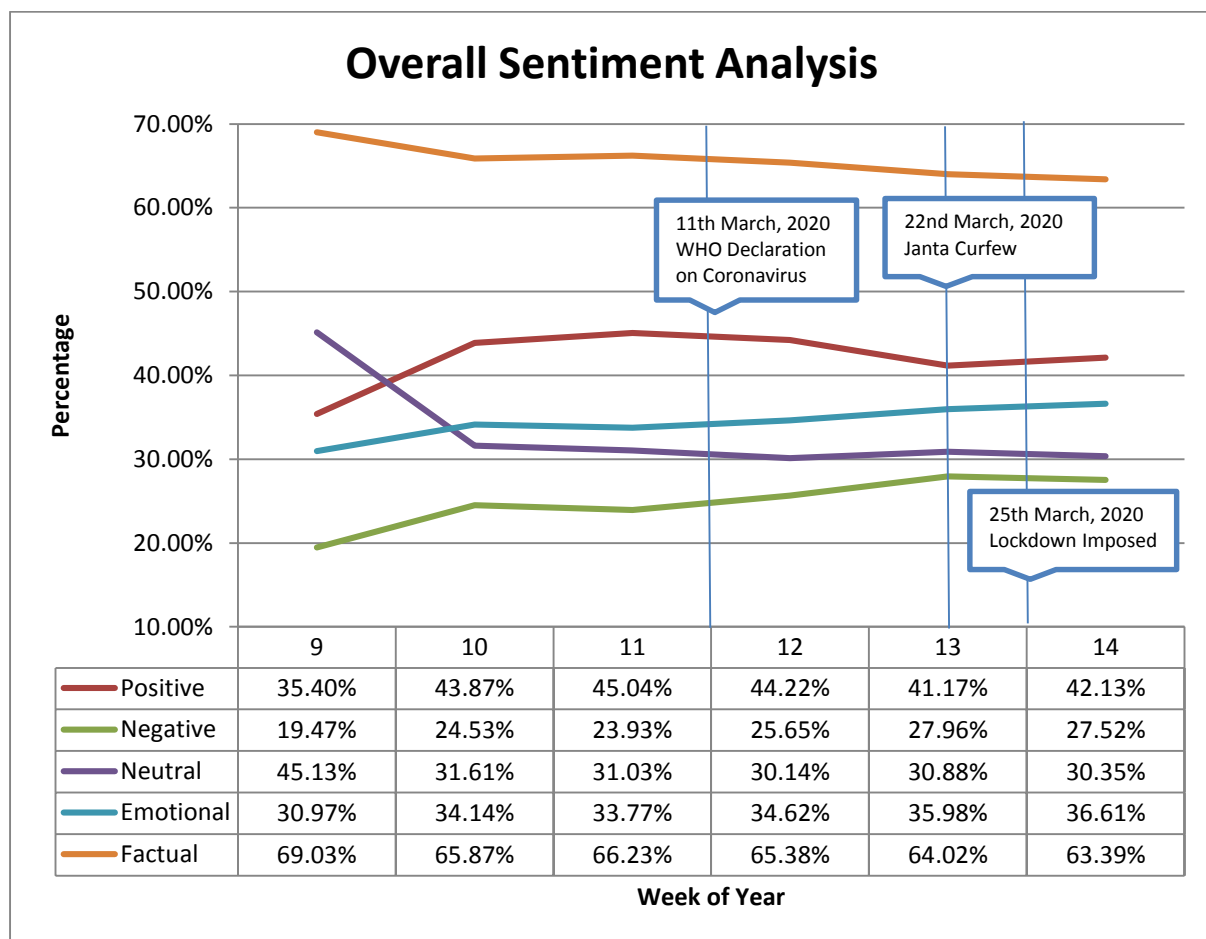


Fig 9. Week-Wise Sentiment Analysis upon Entire Data Set

In Fig. 9, the overall number of factual tweets is the highest, followed by positive and emotional tweets. It can be observed that between week 9 and 10 there has been a sharp rise in the number of positive, negative and emotional tweets, while there has been a sharp decline in the number of neutral and factual tweets. Although, positive tweets are much higher than negative tweets but from week 11 onwards there has been a slight fall in

percentage of positive tweets while starting from week 11, the percentage of negative tweets is on the rise. Timelines for significant events has also been provided in the plot to aid in the analysis of results.

5.2 Plot of Positive Tweets by Users

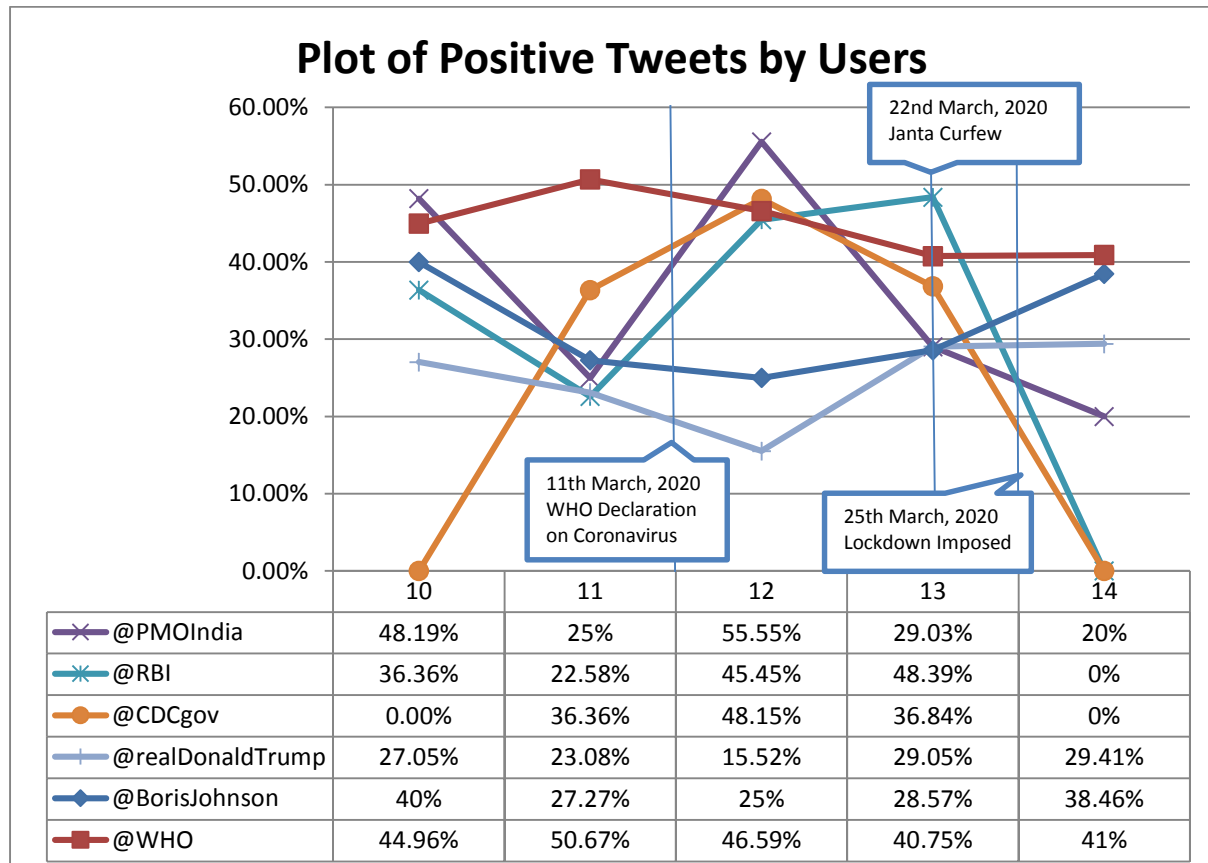


Fig. 10 Plot of Positive Tweets by Users

In Fig. 10, most of the positive tweets can be attributed to “@WHO”, “@PMOIndia” and “@RBI” while least number of positive tweets belongs to “@realDonaldTrump”. Timelines for significant events has also been provided in the plot to aid in the analysis of results.

5.3 Plot of Negative Tweets by Users

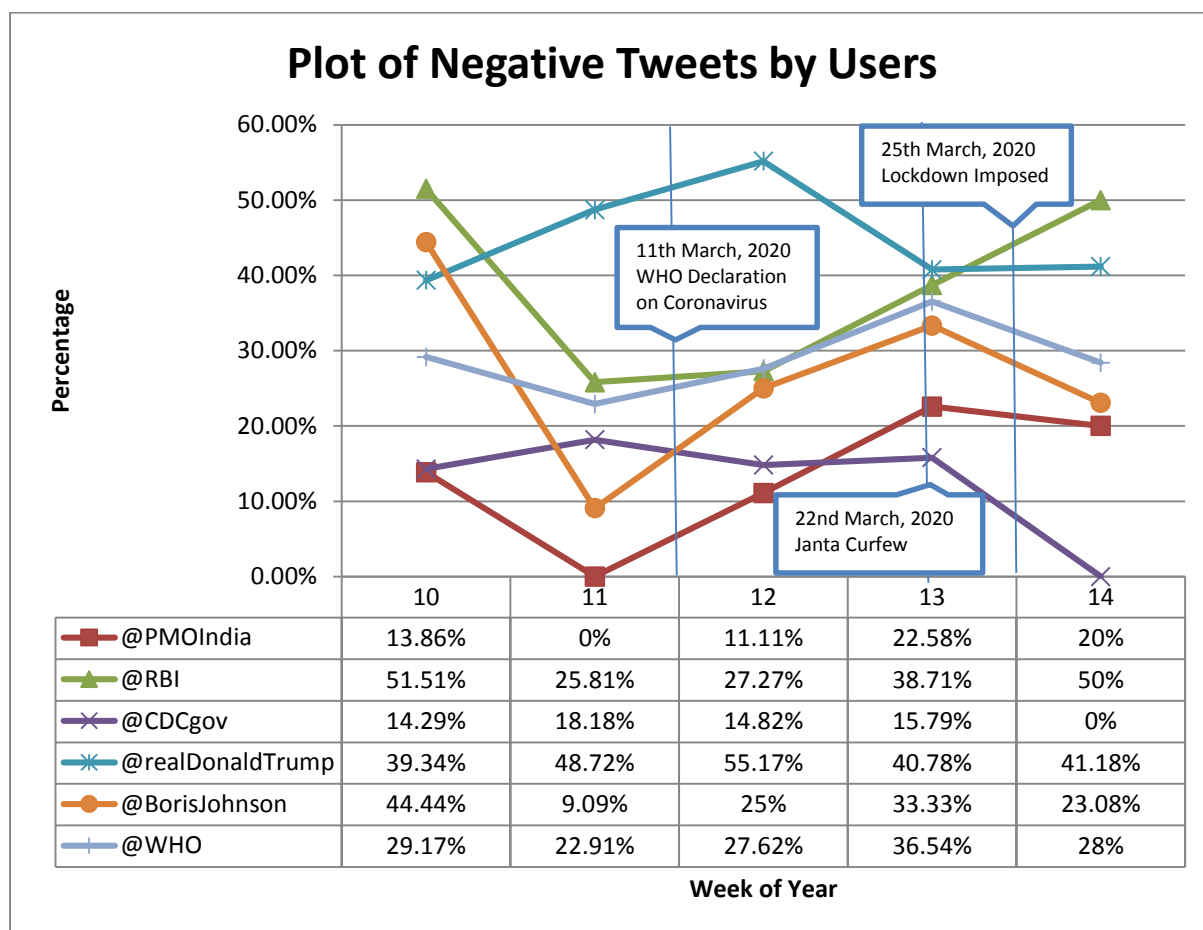


Fig. 11 Plot of Negative Tweets by Users

In Fig. 11, most of the negative tweets can be attributed to “@realDonaldTrump”, “@RBI” and “@WHO” while the least number of negative tweets belong to “@CDCgov” and “@PMOIndia”. Timelines for significant events has also been provided in the plot to aid in the analysis of results.

From Fig. 10 and Fig. 11, it can be concluded that the overall number of positive tweets is greater than the number of negative tweets. Furthermore, most of the positive tweets are attributed to “@WHO” while least number of positive tweets belong to “@realDonaldTrump”. For negative tweets, most of the negative tweets are attributed to “@realDonaldTrump” while the least number of negative tweets belong to “@CDCgov” and “@PMOIndia”.

5.4 Plot of Factual Tweets by Users

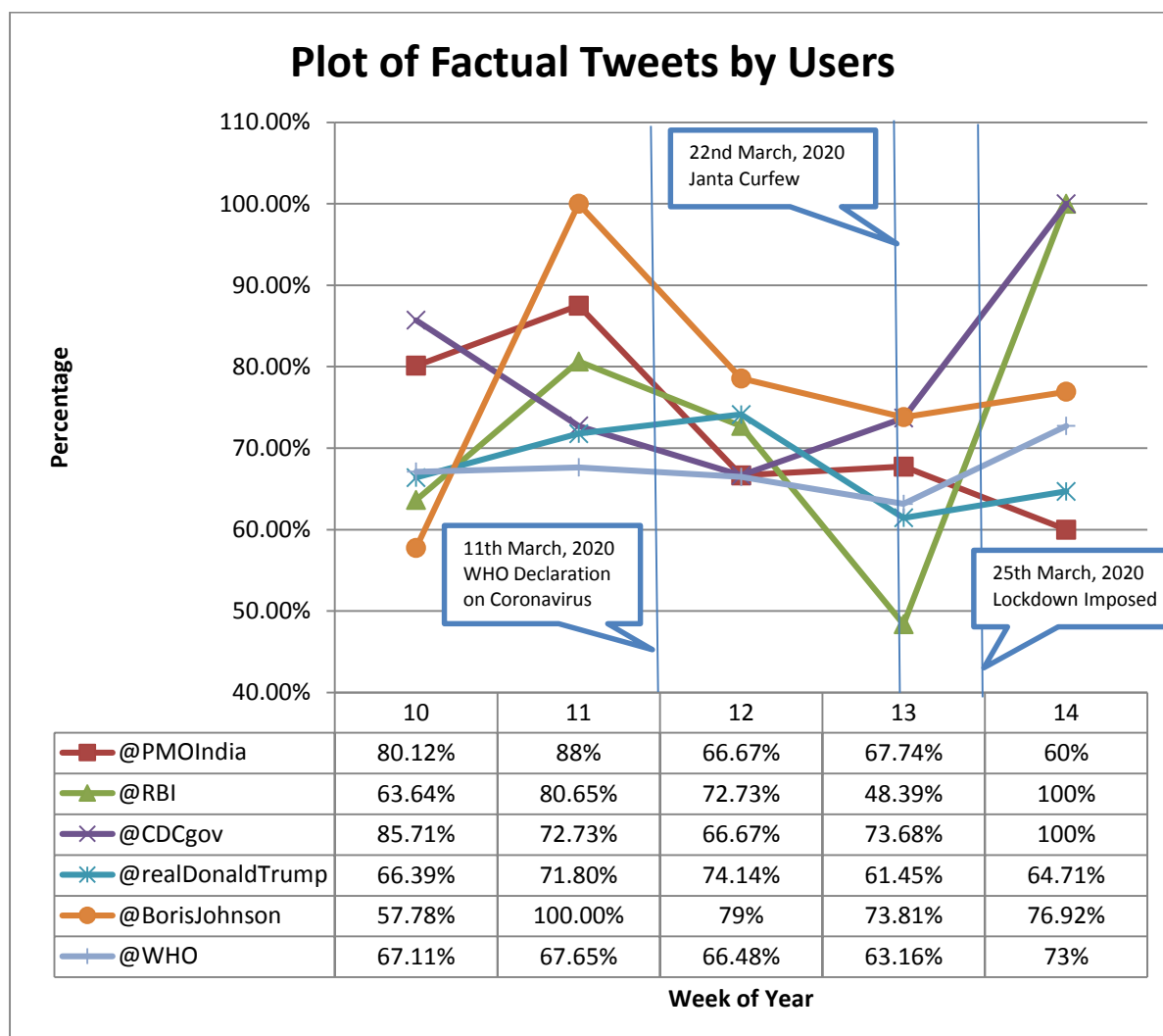


Fig. 12 Plot of Factual Tweets by Users

In Fig. 12, most of the factual tweets can be attributed to “@BorisJohnson” and “@PMOIndia” while least number of factual tweets can be attributed to “@WHO”. Timelines for significant events has also been provided in the plot to aid in the analysis of results.

5.5 Plot of Emotional Tweets by Users

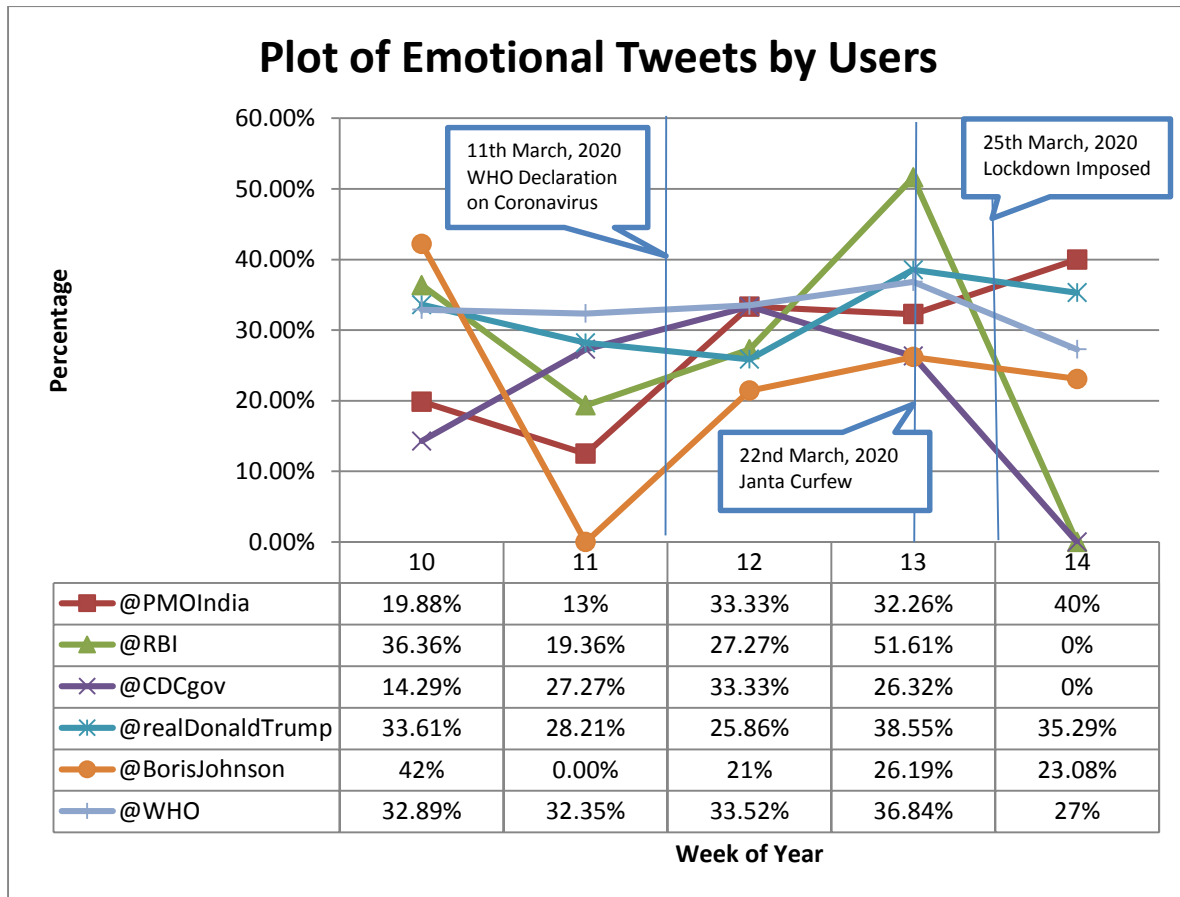


Fig. 13 Plot of Emotional Tweets by Users

In Fig. 13, most of the emotional tweets can be attributed to “@WHO” and “@realDonaldTrump” while least number of emotional tweets can be attributed to “@BorisJohnson”. Timelines for significant events has also been provided in the plot to aid in the analysis of results.