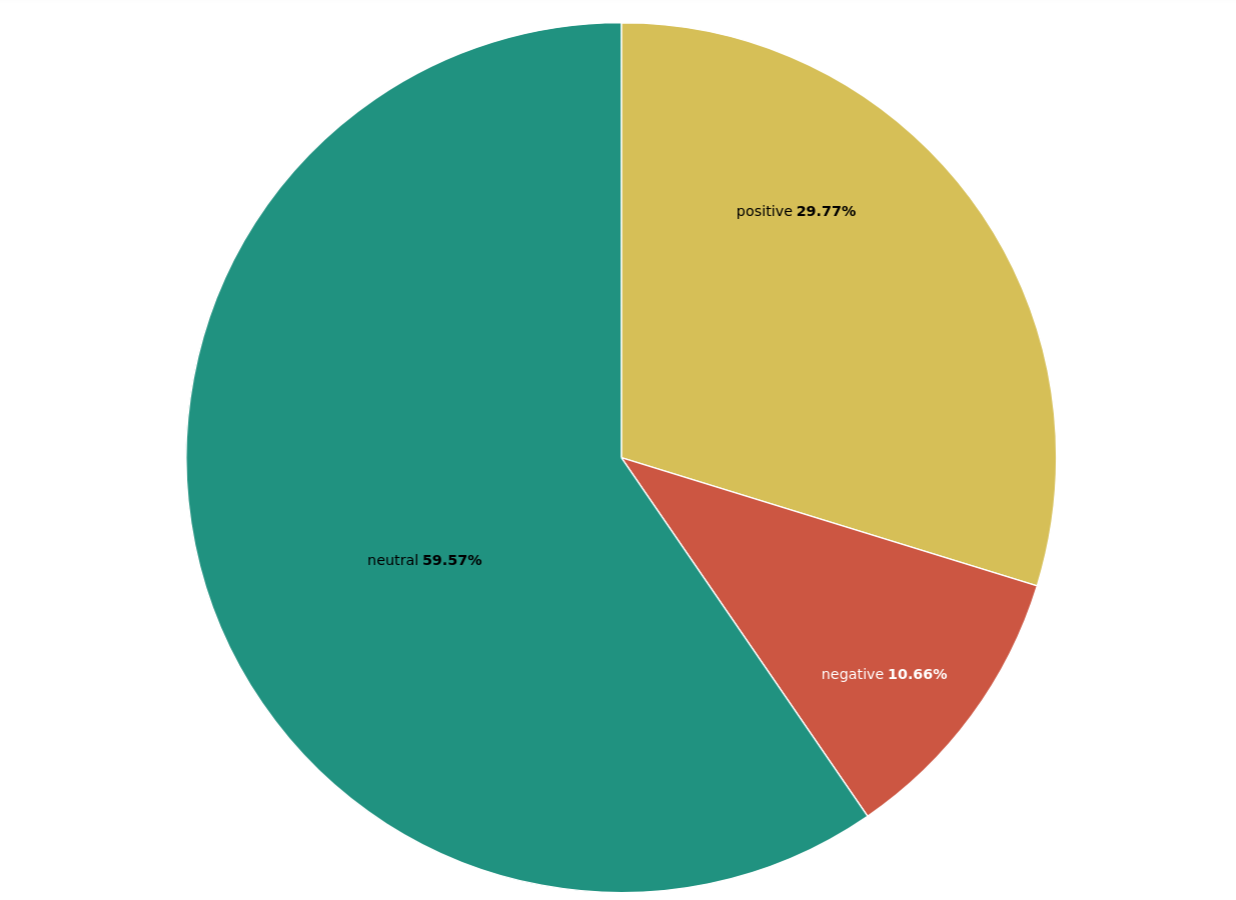
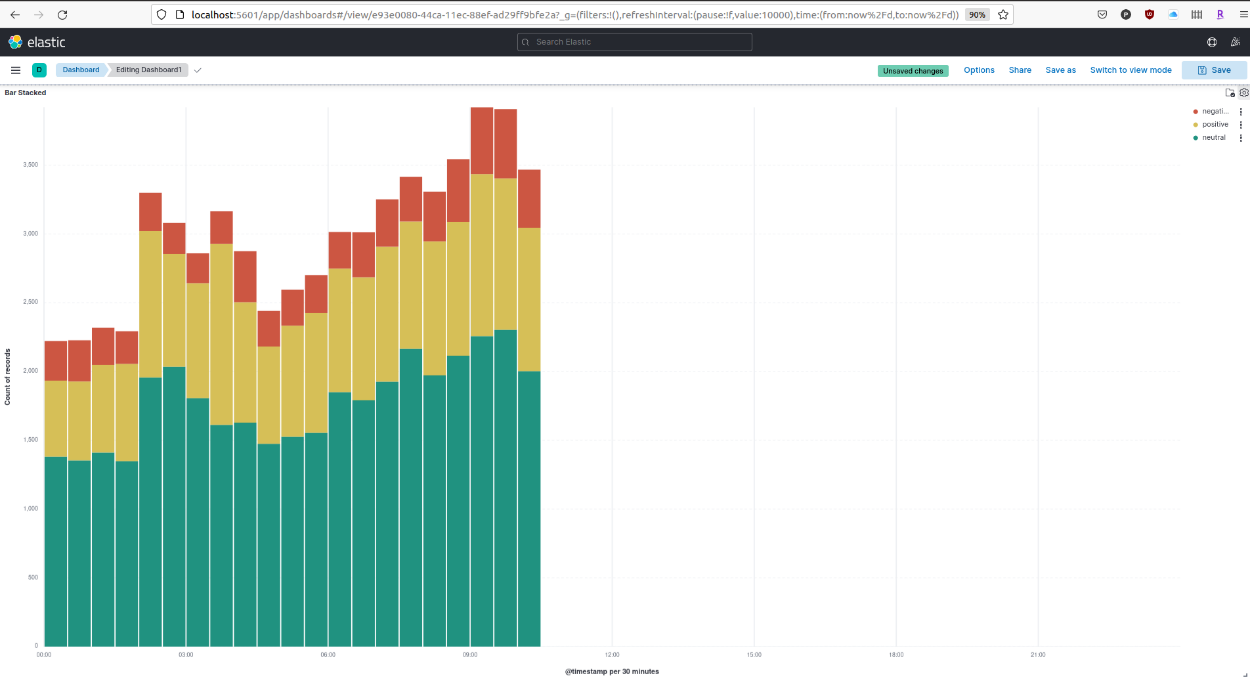
# Spark Streaming with Twitter and Kafka

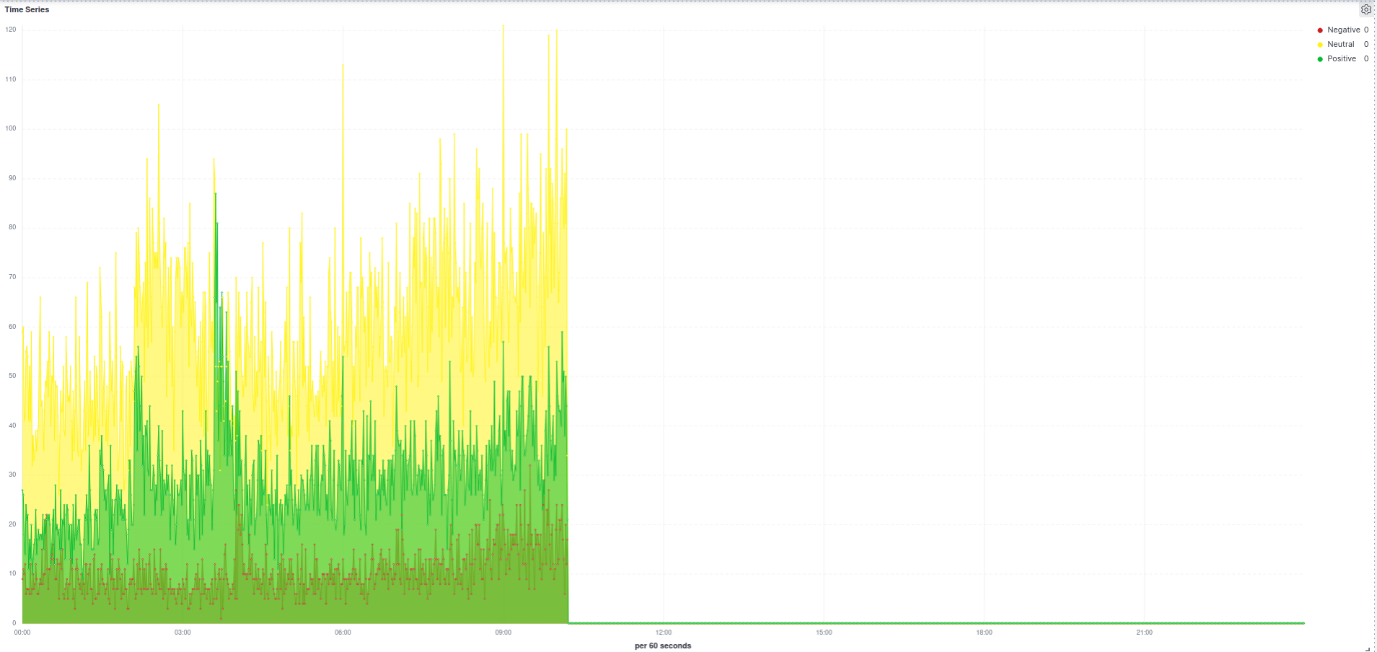
The program was run for roughly, a little over 9 hours to analyze and visualize the sentiment of the tweets related to the search term “covid-19”.

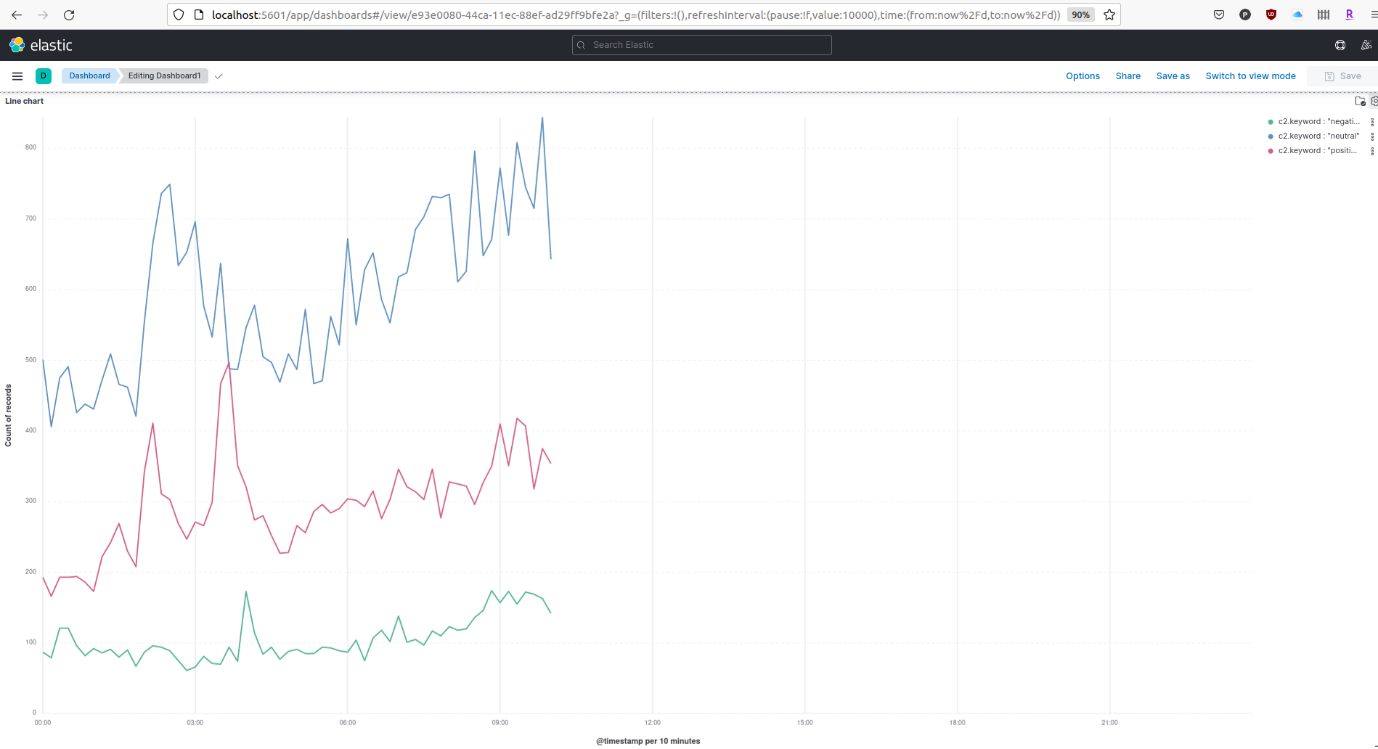
Over the period of 9 hours, here is the distribution of the tweet sentiments:



As we can obtain from the Pie chart, most of the tweet regarding our search term is neutral at 59.57% of the total tweet. Then we have positive at 29.77% and negative at 10.66%







From the Times Series and Lines Chart over Time plots, we can see that approaching 9AM CST on a Sunday, the number of tweets is increasing. This gives up the indication that most of the tweeters are in the US timezone as they get more active in the morning. We can also see that most of the tweets’ sentiment are neutral regarding the search term of “covid-19”. However, around 3AM CST, there is a big surge in positive tweets and smaller surge in negative tweets while the neutral tweets dropped. From this, we can conclude the sentiment of “covid-19” on Twitter is very different based on the time of activity. During the observation time frame, we can identify the group of people who tweet from 12:00AM to 06:00AM who has either very positive or negative sentiment about covid. Meanwhile, the group of people who’s active from 06:00AM CST is more neutral.