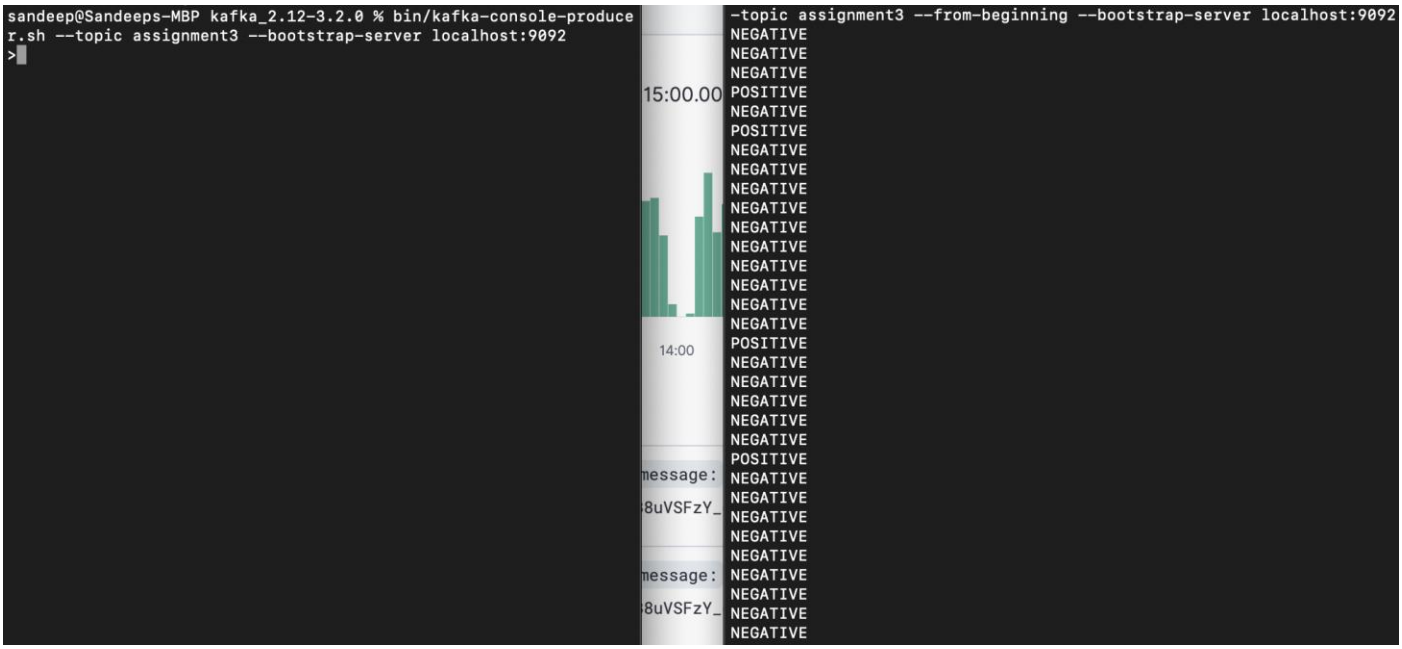
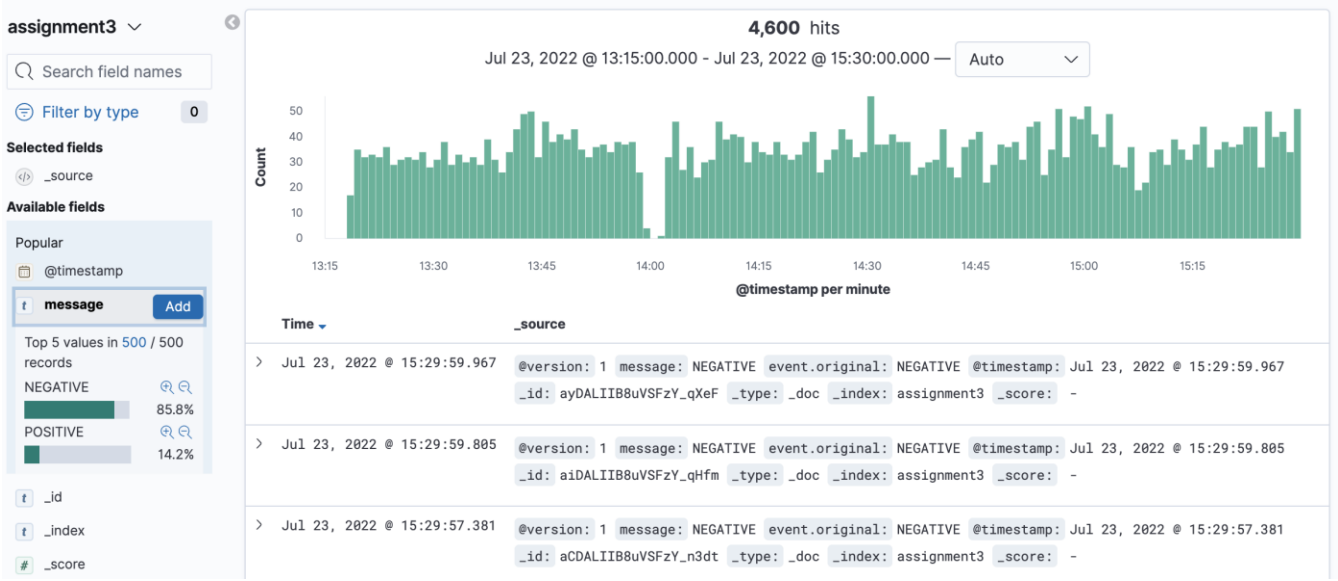
1. Spark Streaming with Twitter and Kafka

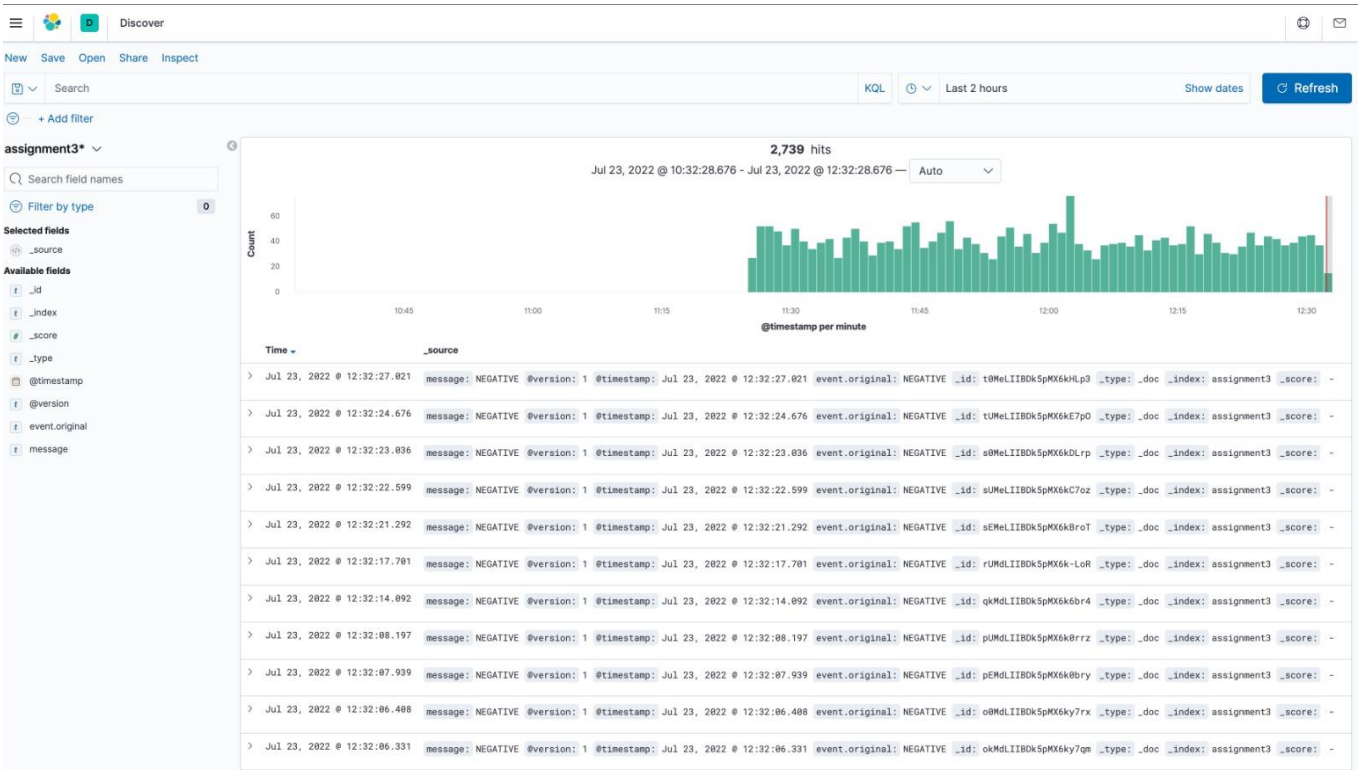
The motive of this assignment is to perform sentiment analysis and then classifying the tweets from Twitter into Positive and Negative.
#covid19 is the is the filter that has been used for this assignment.
Output from the terminal based on sentiment analysis:



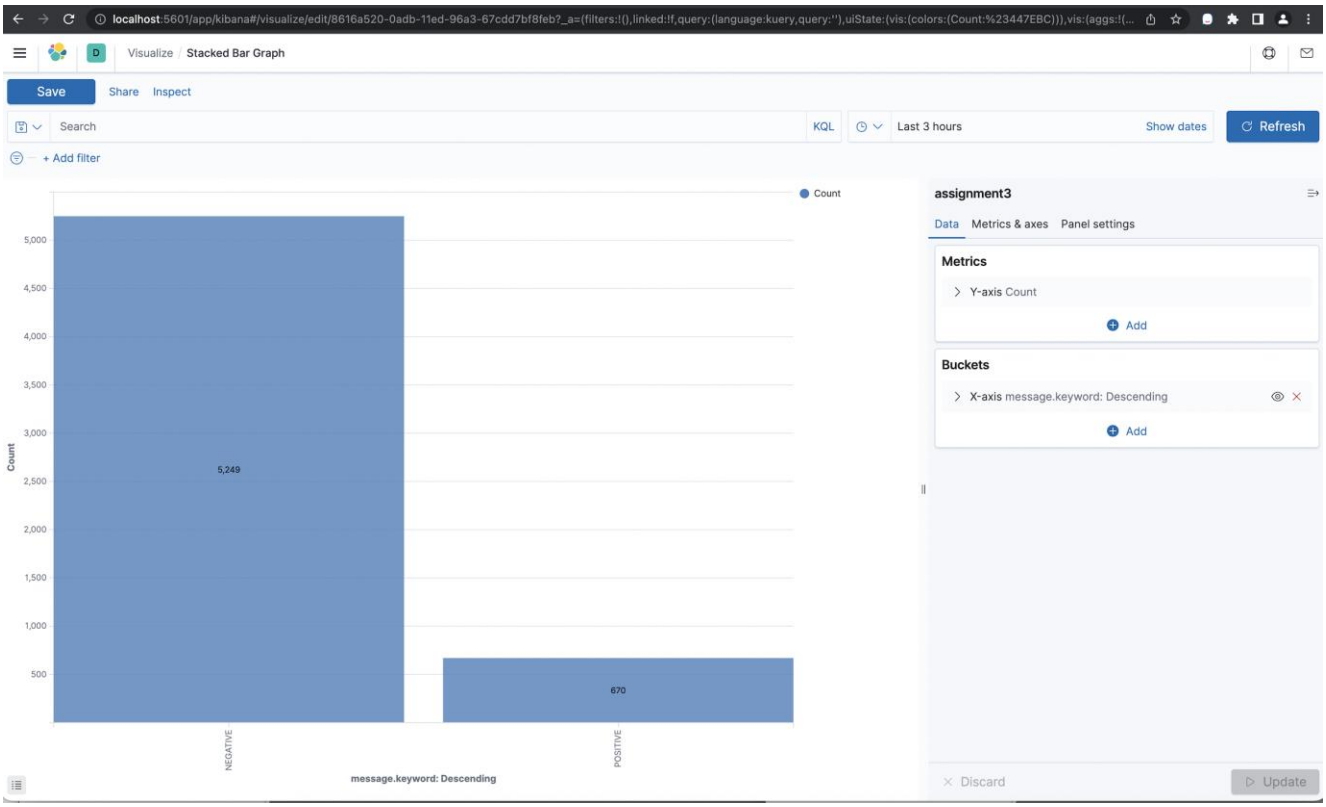
After performing the sentiment analysis, visualized the covid19 tweets using ELK Stack.

Below is the output of my visualization of the data from the ELK stack:

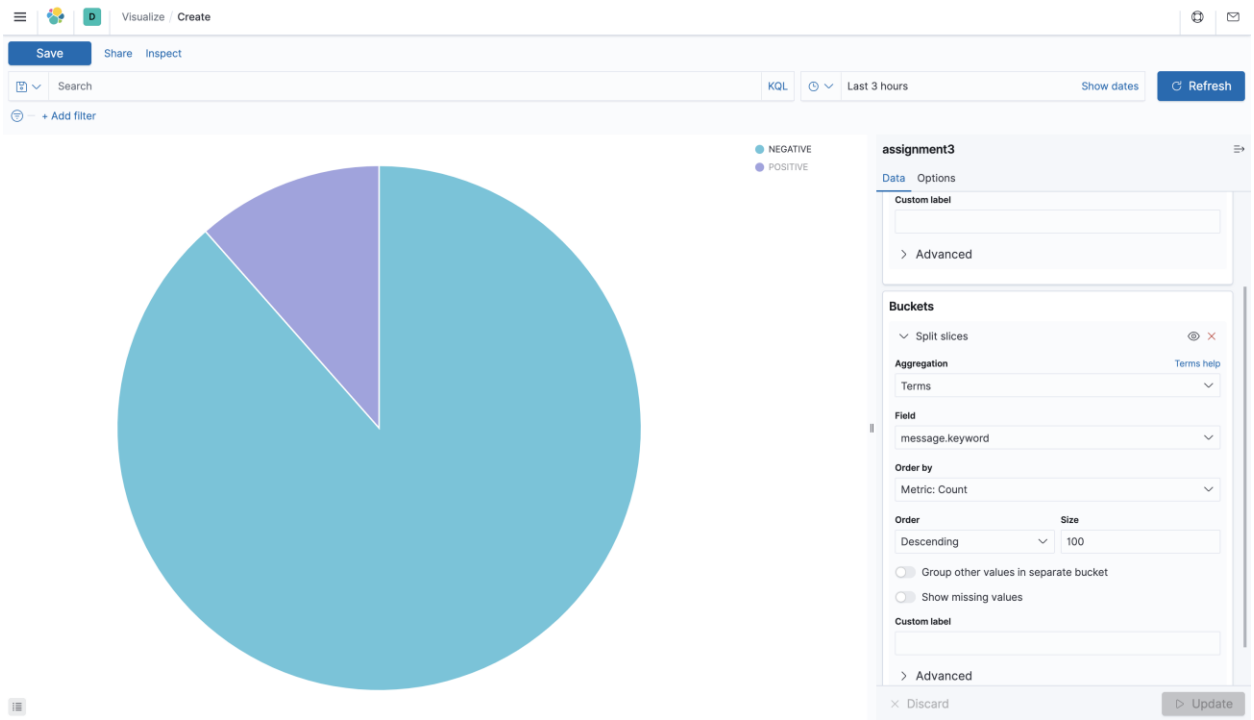


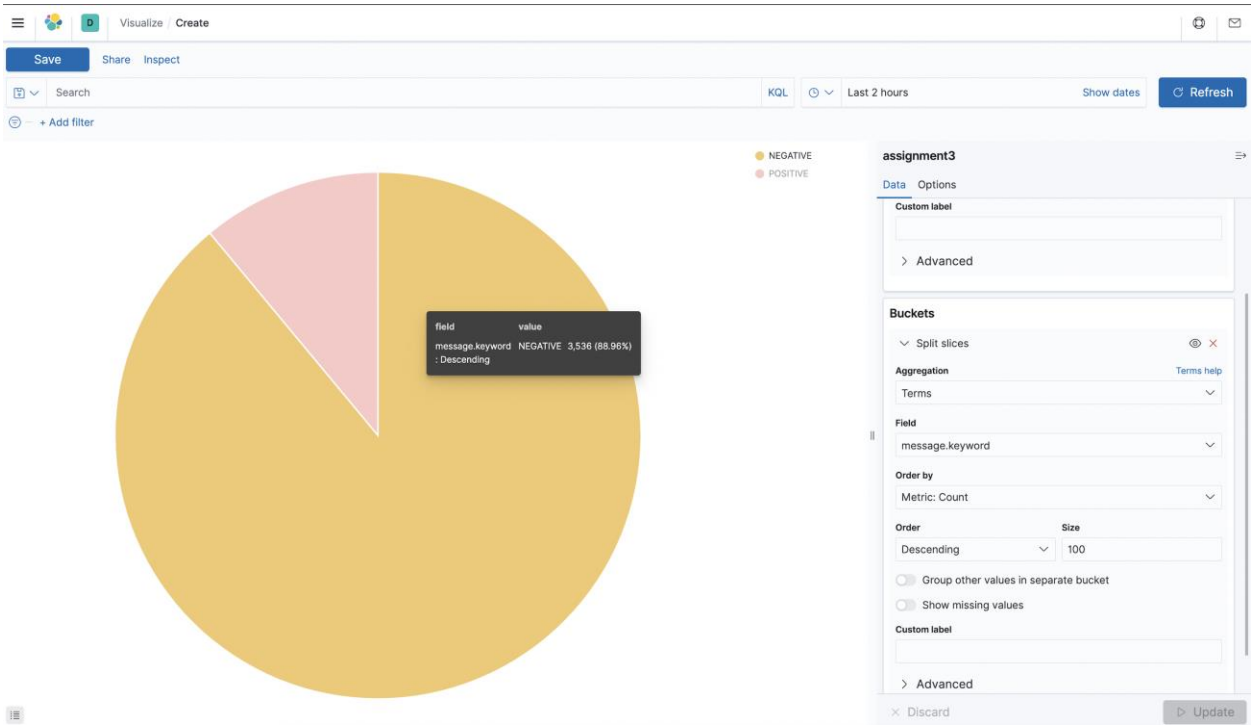


The below graphs represent the categorization of the overall tweets obtained from Twitter using Kafka for a time interval of 2-3 hours:



The variation of the positive and negative tweets over a period of time in the form of pie chart:



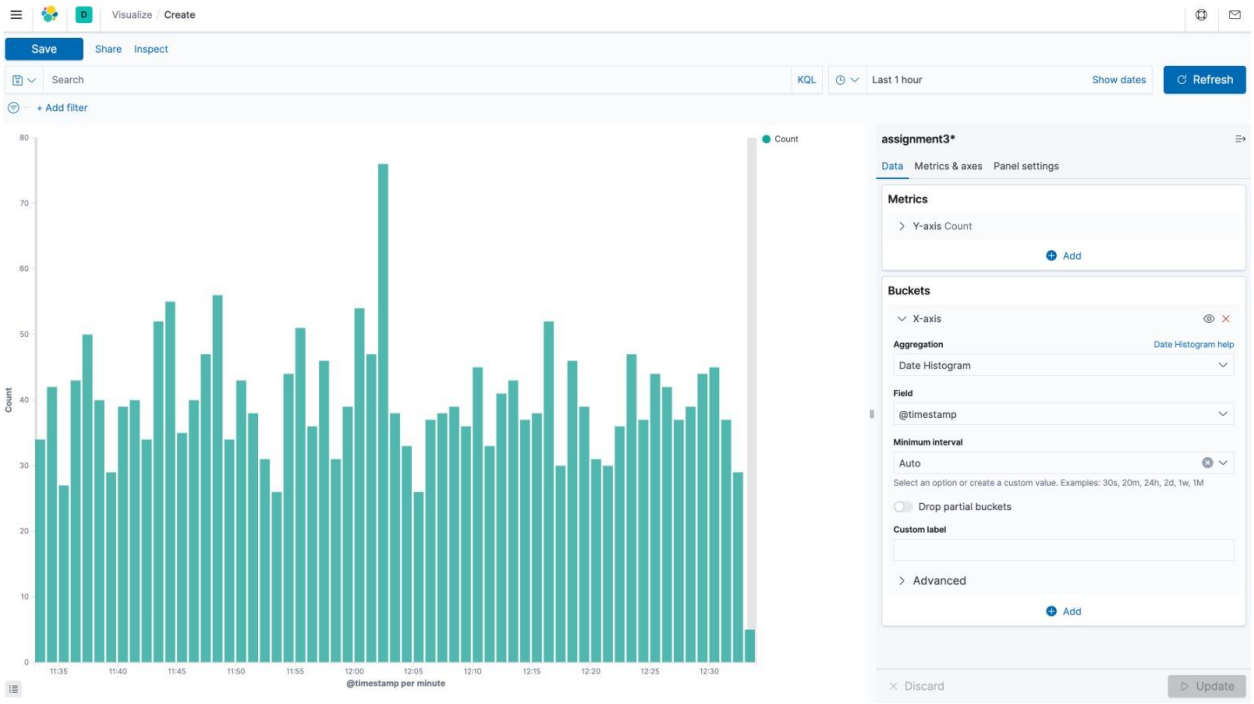


As per the analysis done, there are 89.370% of negative tweets and 10.630% of positive tweets. The ratio of negative to positive tweets is ~9:1.

From the above observations, We can see that most of the tweets related to covid19 have negative sentiment.

Visualization of sentiment against time interval of a day:

The bar graph for covid19 tweets time series as per count of positive and negative tweets is:



Insights:

The number of negative tweets was substantial at the start of the time interval, but this just gradually decreased and had timely peaks.

Number of positive tweets stayed unchanged with only a few dips during the time period of analysis.

There exists huge dissatisfaction with the keyword covid19, and very few people have expressed positive thoughts about it.