# Spark Streaming with Twitter and Kafka

- We get tweets involving these 6 topics. We have separated the tweets into 3 sentiments: POSITIVE, NEUTRAL, AND NEGATIVE.
 The 6 topics(filters) are: - nfl, nba, ps5, Pandemic, Weather, and Music
- The predicted sentiment is appended to the tweet and the entire thing is sent to Kafka broker. Using ELK stack (Elastic Search, Kibana and Logstash), a pipeline is created the between Kafka broker and Elasticsearch. Then visualization for the 6 topics is done using Kibana.

NUMBER OF POSITIVE, NEUTRAL AND NEGATIVE TWEETS

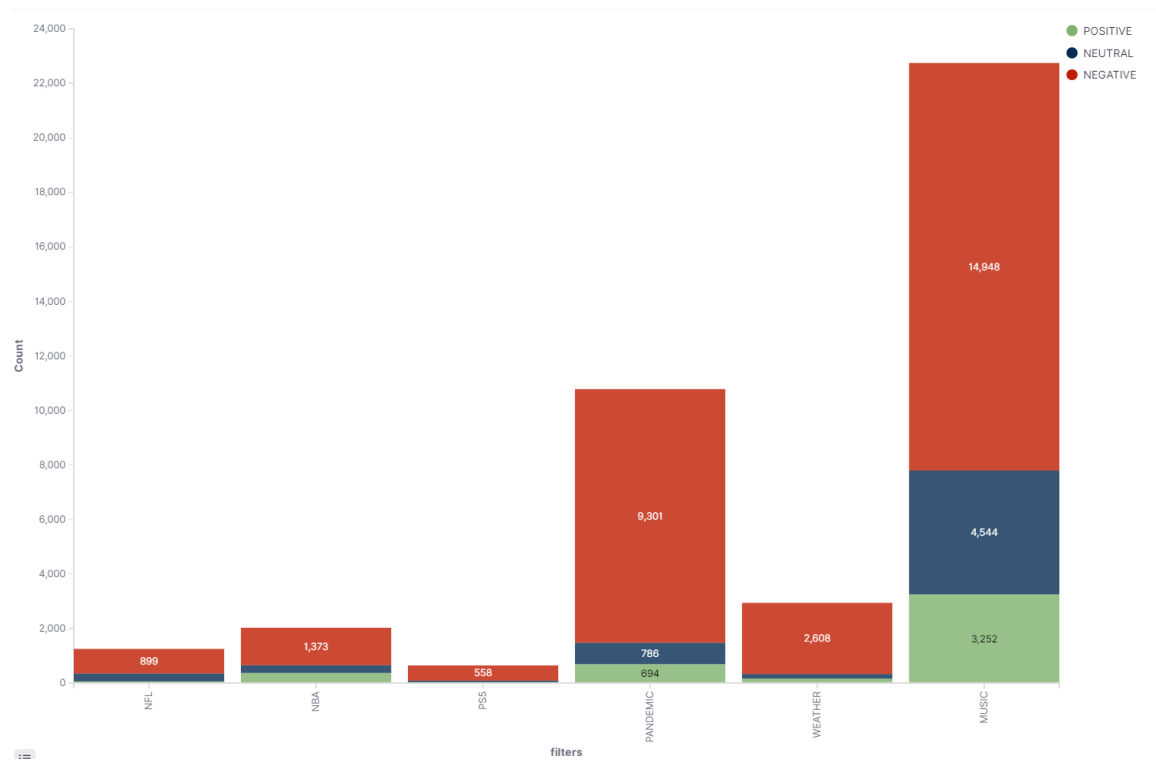Positive Tweets, Neutral Tweets and Negative Tweets
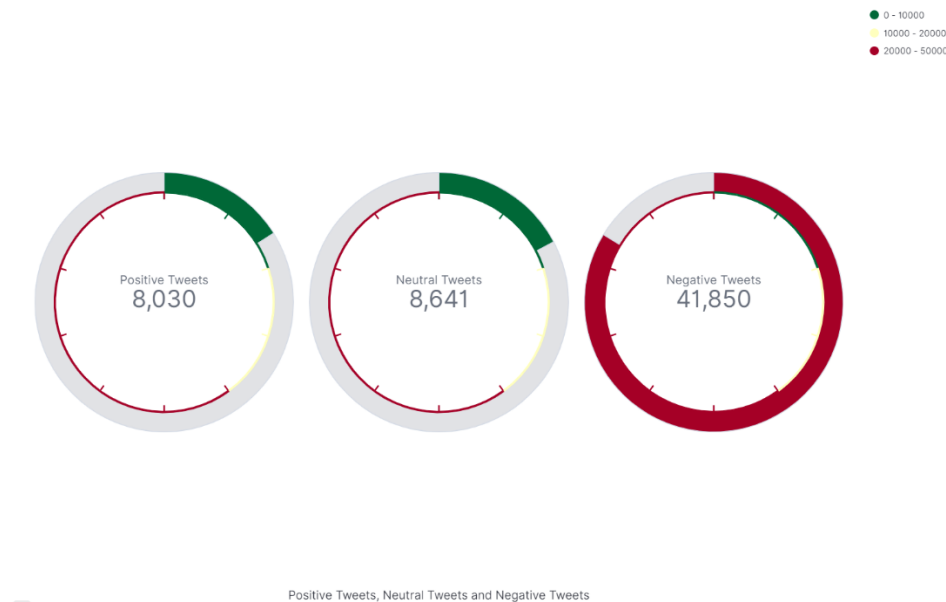
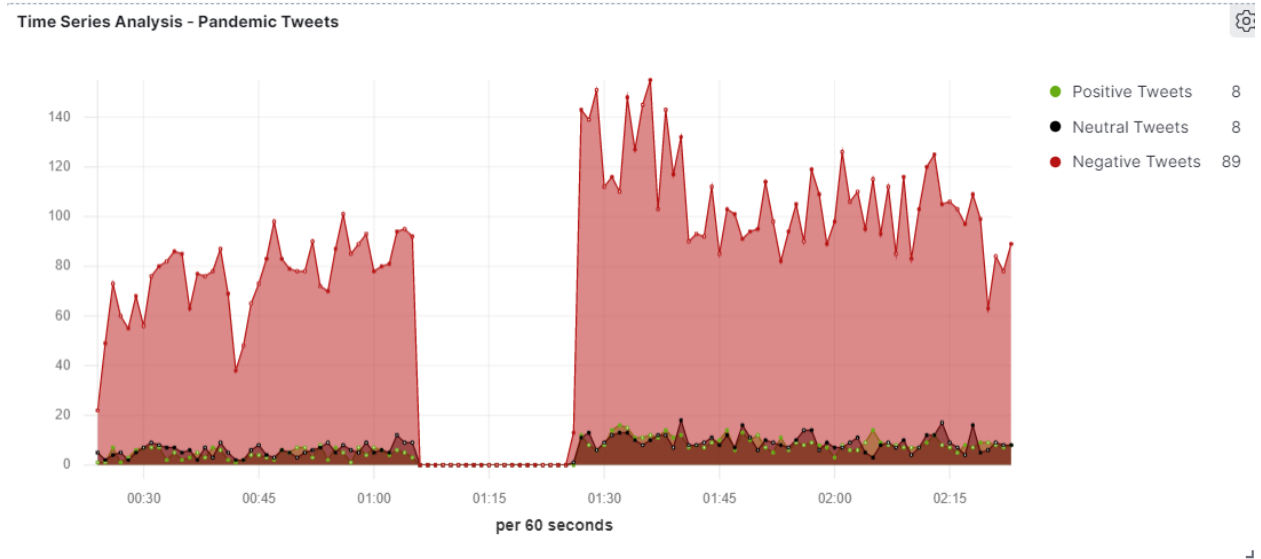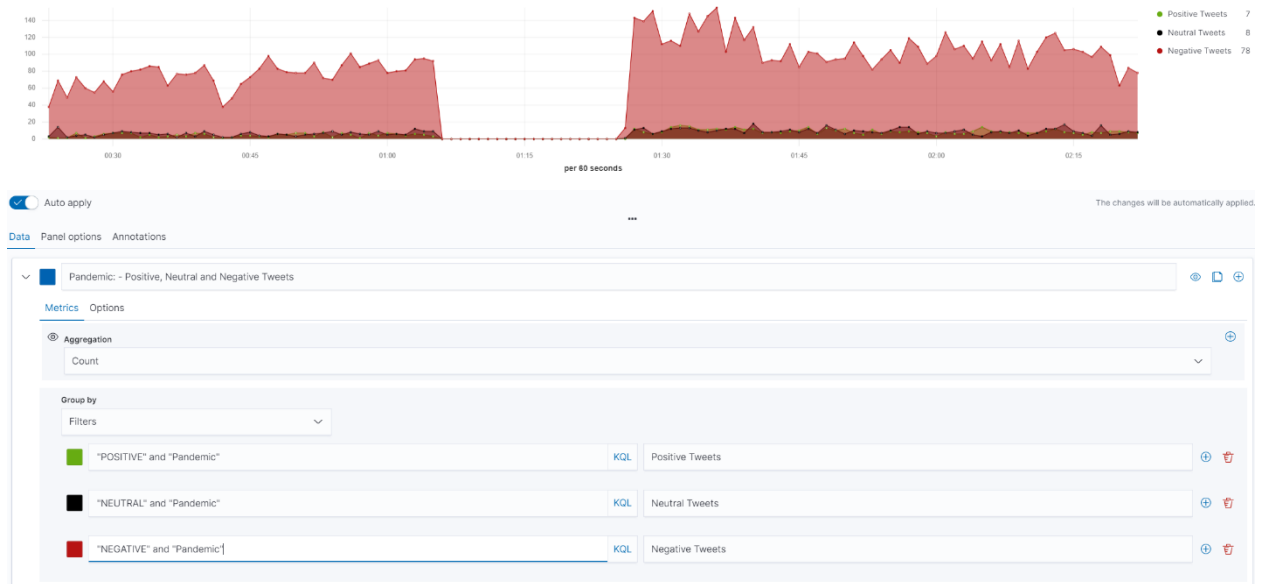## NUMBER OF TWEETS PER TOPIC
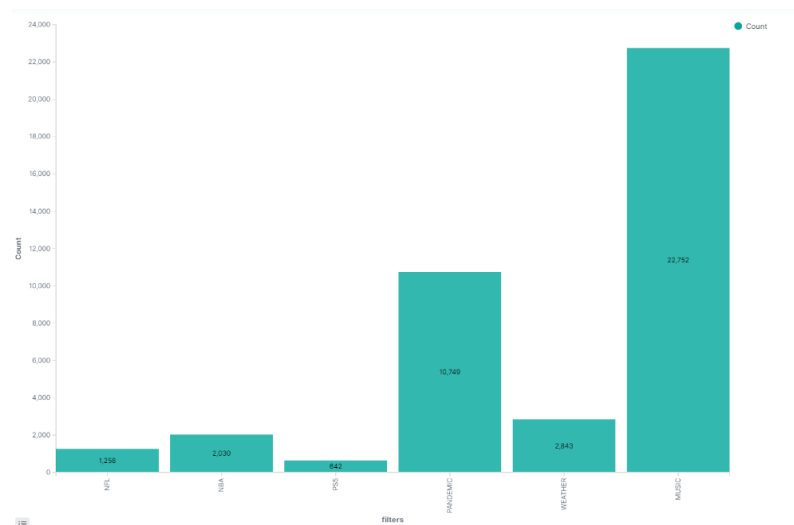
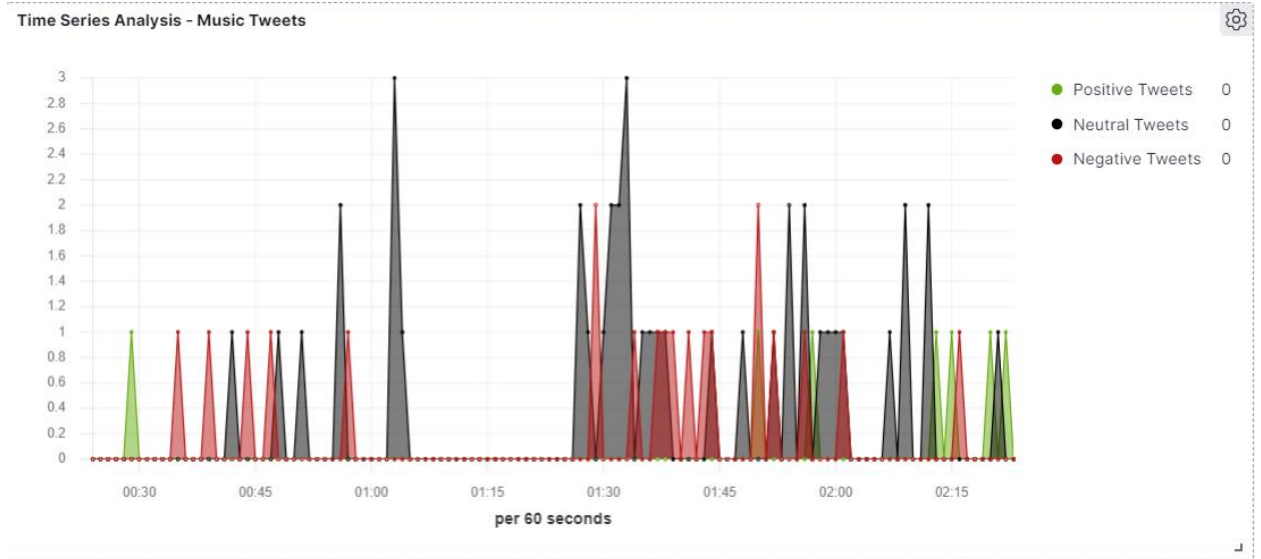## PERCENTAGE OF TWEETS PER TOPIC
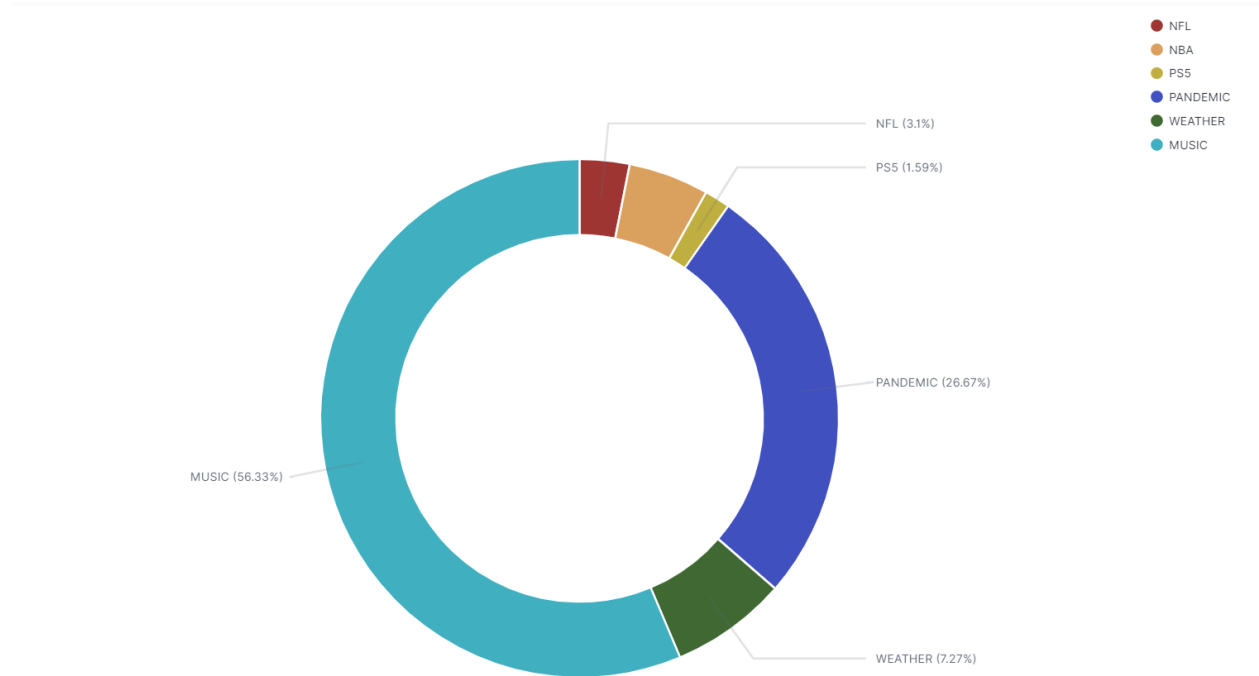
POSITIVE, NEUTRAL, AND NEGATIVE TWEETS PER TOPIC

TIME SERIES ANALYSIS – PANDEMIC TWEETS

Time Series Analysis - Pandemic Tweets

● Positive Tweets	8
● Neutral Tweets	8
● Negative Tweets	89

per 60 seconds

TIME SERIES ANALYSIS – MUSIC TWEETS

● Positive Tweets 0
● Neutral Tweets 0
● Negative Tweets 0

per 60 seconds

⬤ Auto apply

The changes will be automatically applied

•••

Data   Panel options   Annotations

∨ ▮ Pandemic: - Positive, Neutral and Negative Tweets   👁 ⧉ ⊕

Metrics   Options

👁 Aggregation   ⊕

Count   ∨

Group by

Filters   ∨

▮ "POSITIVE" and "Music"   KQL Positive Tweets   ⊕ 🗑

▮ "NEUTRAL" and "Music"   KQL Neutral Tweets   ⊕ 🗑

▮ "NEGATIVE" and "Music"   KQL Negative Tweets   ⊕ 🗑

Time Series Analysis - Music Tweets   ⚙

● Positive Tweets   0
● Neutral Tweets   0
● Negative Tweets   0

per 60 seconds

FULL VISUALIZATION DASHBOARD

Tweet Sentiment- Positive, Neutral and Negative

- ● 0 - 10000
- ● 10000 - 20000
- ● 20000 - 50000

7,969   8,461   41,064

Positive Tweets, Neutral Tweets and Negative Tweets

Tweet count for each topic

● Count

filters	Count
NFL	1,296
NBA	2,061
PS5	
PANDEMIC	10,360
WEATHER	2,900
MUSIC	22,076

Pie Chart - Percentage of tweets for each topic

- ● NFL
- ● NBA
- ● PS5
- ● PANDEMIC
- ● WEATHER
- ● MUSIC

NFL (3.3%)
PANDEMIC (26.36%)
WEATHER (7.38%)
MUSIC (56.16%)

Positive, Neutral and Negative tweet counts for each topic

- ● POSITIVE
- ● NEUTRAL
- ● NEGATIVE

filters	
NFL	
NBA	1,391
PS5	
PANDEMIC	8,987
WEATHER	2,581
MUSIC	14,554 / 4,359 / 3,174

Time Series Analysis - Pandemic Tweets

- ● Positive Tweets 9
- ● Neutral Tweets   8
- ● Negative Tweets 93

per 60 seconds

Time Series Analysis - Music Tweets

- ● Positive Tweets 0
- ● Neutral Tweets   0
- ● Negative Tweets 0

per 60 seconds

ANALYSIS

The 6 topics of tweets for which analysis was done: - nfl, nba, ps5, Pandemic, Weather, Music
Sentiment analysis was done using Stanford NLP.

1) Looking at the tweets counts we can say that for the 6 given topics most of the tweets are negative, whereas positive and neutral tweets are less and comparable in count.
2) Music is the most tweeted about subject (more than 50 percent), whereas pandemic tweets are second in count due to current events.
3) For the 6 given topics we can state conclusively that the sentiment is mostly NEGATIVE.
4) When we look at the timeseries for Pandemic tweets we can say that the sentiment is highly NEGATIVE.
5) When we look at the timeseries for Music tweets, even though most of the sentiment for Music tweets are NEGATIVE, it has a higher percentage of POSITIVE and NEUTRAL tweets compared to Pandemic tweets.