Sentiment Analysis of tweets
Topic: lemonade

A Spark Streaming application has been created using Scala that continuously reads data from Twitter about a topic.

A stream of status updates is obtained using TwitterUtils.

The sentiments associated with the statuses are obtained using Stanford CoreNLP library.

The sentiment data is then sent to ElasticSearch for analysis.

To exchange data between the Spark application and ElasticSearch, Kafka is used as the broker.

To ingest data from the Kafka server and map it to the required ElasticSearch server, LogStash has been used.

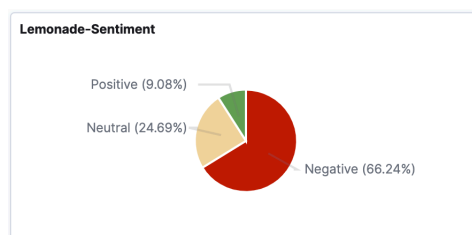Finally, the sentiments are visualized in Kibana.

Here, Twitter statuses have been obtained using keyword 'lemonade' as it is one of the trending topics during the arrival of spring. The sentiments of the same have been obtained and visualized to find what percentage of the tweets have 'positive', 'negative', and 'neutral' sentiments.

Pie-chart snapshots in Kibana showing how the sentiment for 'lemonade' varied over time have been saved.
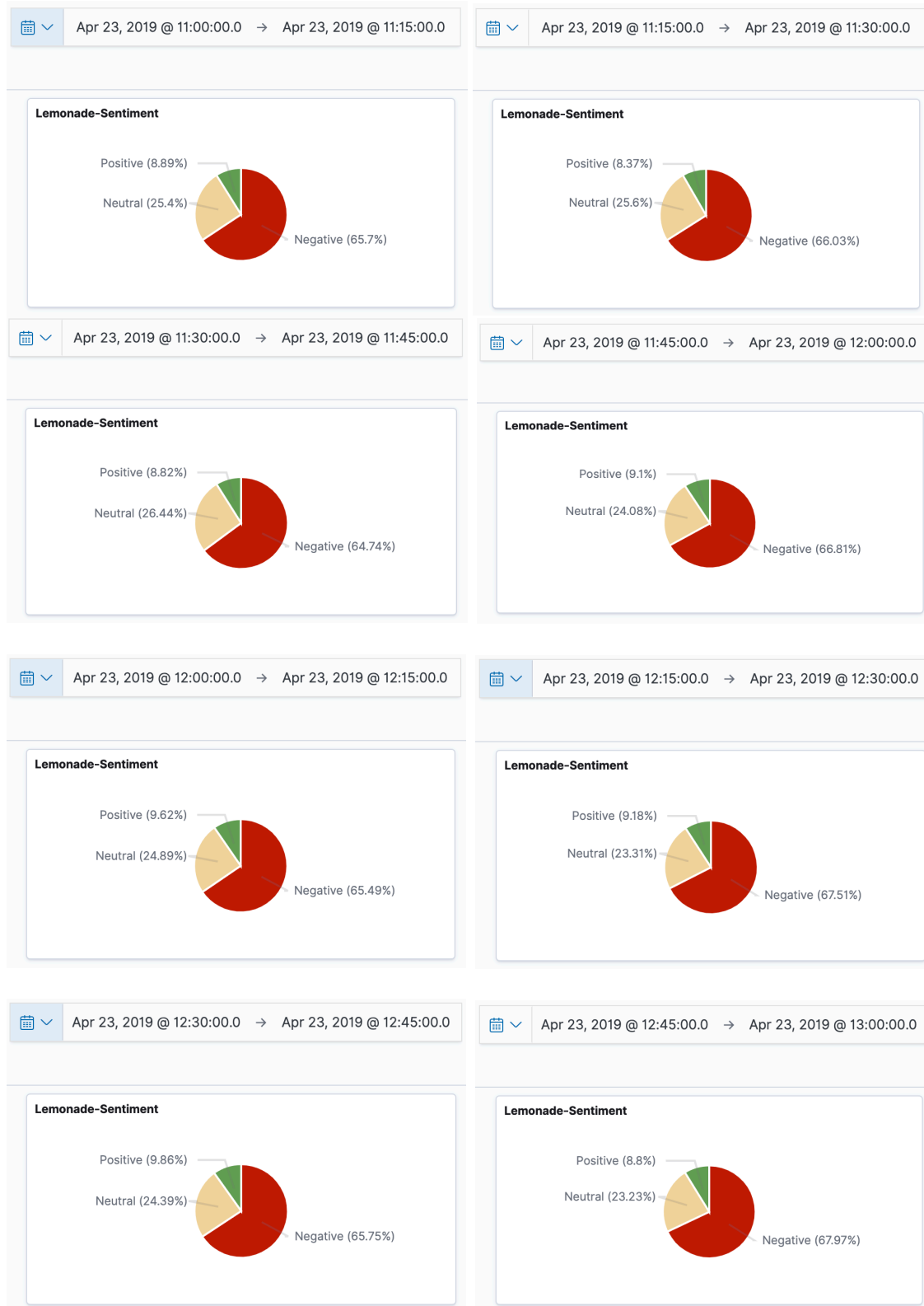
The hourly tweet count involving 'lemonade' on Apr 23, 2019 from 1100 to 1300 hrs. is as below:
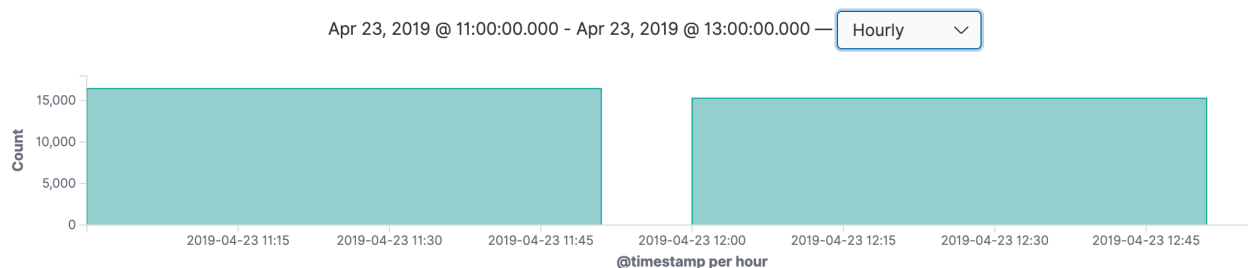
Overall sentiment classification during the time is as below:

The variation of the sentiments by 15 min windows is as below:

| 📅 ⌄ | Apr 23, 2019 @ 11:00:00.0 → Apr 23, 2019 @ 11:15:00.0 |

Lemonade-Sentiment

Positive (8.89%)
Neutral (25.4%)
Negative (65.7%)

| 📅 ⌄ | Apr 23, 2019 @ 11:15:00.0 → Apr 23, 2019 @ 11:30:00.0 |

Lemonade-Sentiment

Positive (8.37%)
Neutral (25.6%)
Negative (66.03%)

| 📅 ⌄ | Apr 23, 2019 @ 11:30:00.0 → Apr 23, 2019 @ 11:45:00.0 |

Lemonade-Sentiment

Positive (8.82%)
Neutral (26.44%)
Negative (64.74%)

| 📅 ⌄ | Apr 23, 2019 @ 11:45:00.0 → Apr 23, 2019 @ 12:00:00.0 |

Lemonade-Sentiment

Positive (9.1%)
Neutral (24.08%)
Negative (66.81%)

| 📅 ⌄ | Apr 23, 2019 @ 12:00:00.0 → Apr 23, 2019 @ 12:15:00.0 |

Lemonade-Sentiment

Positive (9.62%)
Neutral (24.89%)
Negative (65.49%)

| 📅 ⌄ | Apr 23, 2019 @ 12:15:00.0 → Apr 23, 2019 @ 12:30:00.0 |

Lemonade-Sentiment

Positive (9.18%)
Neutral (23.31%)
Negative (67.51%)

| 📅 ⌄ | Apr 23, 2019 @ 12:30:00.0 → Apr 23, 2019 @ 12:45:00.0 |

Lemonade-Sentiment

Positive (9.86%)
Neutral (24.39%)
Negative (65.75%)

| 📅 ⌄ | Apr 23, 2019 @ 12:45:00.0 → Apr 23, 2019 @ 13:00:00.0 |

Lemonade-Sentiment

Positive (8.8%)
Neutral (23.23%)
Negative (67.97%)

We can conclude that the sentiments do not vary much and the majority of tweets containing 'lemonade' are associated with negative sentiments.