Assignment 4

Ishaq Khan

INFS 6353

11/21/2023

Keyword #1: NBA

The keyword that was selected for the first set was “NBA”.

A screenshot of a computer

Description automatically generatedA group of colorful lines

Description automatically generated with medium confidence

This keyword will be collected from X(Twitter) to perform the following:

1. The keyword that I decided to use was “NBA”. The rectangular word cloud for this tweet set has topics related to the NBA. You can see that the larger words in the the word cloud represent the more frequently used words in tweets from the dataset that was collected.

A close-up of words

Description automatically generated

1. The word cloud above from the keyword “NBA” was then converted using the twitter logo as a mask to fit the words into the shape of the bird.

A close-up of words

Description automatically generated

1. The ten most frequently used terms in the data set can be identified using the word cloud. They were:
   1. NBA
   2. https
   3. StateFarm
   4. Attention
   5. Game
   6. Join
   7. Money
   8. Team
   9. Pick
   10. Play
2. Sentiment Analysis

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

The following are tweets after being filtered and applying the NLTK sentiment analyzer. The polarity ranges from -1 to 1 and it tell us overall whether a tweet is negative, neutral, or positive. The higher the number in that specific column(positive, negative, and neutral) the higher the sentiment it is to that specific aspect.

The sentiment score is determined by the value of the polarity. If the polarity is a negative number or a “0”, then the sentiment will have a value of “0”. If the polarity is positive, then the sentiment will have a value of “1”.

Upon further analysis, we find that a majority of the polarity score do not have a negative value. Most of them are either a 0.000 or higher. As a result, the sentiment score will be a 0 for some tweets but mainly due to the fact that the polarity was a 0 and not mainly from a negative value. There are a lot of positive polarity results which means the sentiment score will be a 1. Because the topic was “NBA”, there isn’t really any negative sentiment that can be found in there unless something negative related to the league occurs. Possibly, if a player got injured or there is a strike, then you could expect more negative polarity’s and negative sentiments.

Keyword #2: Health

The keyword that was selected for the first set was “Health”.

A screen shot of a computer

Description automatically generatedA close up of a screen

Description automatically generated

This keyword will be collected from X(Twitter) to perform the following:

1. The keyword that I decided to use was “Health”. The rectangular word cloud for this tweet set has topics related to the Health. You can see that the larger words in the word cloud represent the more frequently used words in tweets from the dataset that was collected.

A close-up of words

Description automatically generated

1. The word cloud above from the keyword “Health” was then converted using the twitter logo as a mask to fit the words into the shape of the bird.

A close-up of words

Description automatically generated

1. The ten most frequently used terms in the data set can be identified using the word cloud. They were:
   1. health
   2. https
   3. Mental health
   4. Discover
   5. Health care
   6. Wellness
   7. priority
   8. world
   9. responsibility
   10. Provinces
2. Sentiment Analysis

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

The following are tweets after being filtered and applying the NLTK sentiment analyzer. The polarity ranges from -1 to 1 and it tell us overall whether a tweet is negative, neutral, or positive. The higher the number in that specific column(positive, negative, and neutral) the higher the sentiment it is to that specific aspect.

The sentiment score is determined by the value of the polarity. If the polarity is a negative number or a “0”, then the sentiment will have a value of “0”. If the polarity is positive, then the sentiment will have a value of “1”.

Upon further analysis, we find that the tends to be quite a few negative polarity results which results in those tweets having a sentiment score of 0. There are some polarity score in the positive range which results in those tweet’s sentiment score to be 1. However, since the topic is “Health”, it is expected to be more sentiment scores of 0 due to recent health events from the previous years.