Date: 05/28/2021

Analysis of sentiment

Via twitter tweets

Author : Nasif Hossain

Twitter is a platform where the opinions of the general people can be seen pretty well. Hence we can collect those tweets in mass amounts while analyzing the sentiment that they give off. We can compare our data to specific parts of the country and have a general idea of what the opinion of people in each city are.

The year 2020 through 2021 has been tough due to the COVID-19 pandemic. But the US is giving out vaccines for everyone in mass amounts at the time of writing which is a relief for most people as we slowly come to an end of this deadly pandemic. However, there are people who claim that the vaccine is not FDA approved, it is untested, which might bring out a lot of side effects in the future. In this project, we are going to explore the sentiments of people tweeting about the vaccine and their opinions on it.

The project is structured in three steps, searching of the tweets using the twitter developer API, analyzing the tweets against a dictionary of word - sentiment values and giving the tweet a total sentiment value and finally, using the data results we get and coming to a conclusion with data as evidence.

Searching tweets

The tweets we will explore are based on geo-location and the parameters are Northern and Southern USA. The SearchTweets class collects the tweet using the collectTweets method which gathers the tweets based on geo-location and puts them in an ArrayList using a SeachTweets method, while also saving the tweets as status objects to a serialized outfile. There is a load method which is called in the save method, where its job is to read the serialized file and put all the status tweet objects from file to the tweets ArrayList. Immediately afterwards, a deleteDuplicate method is called which deletes all the duplicate tweets in the tweets ArrayList. This method only deletes duplicates but not

retweets of the same tweet because we would still count the sentiment value of the retweet as it is from a different person. The queries we used to explore vaccine related tweets were, ‘vaccine’, ‘#vaccinated’, ‘#vaccine’, ‘vaccinated’, ‘fda-approved’. First we search tweets for Southern USA, we take all the cities with latitude lower than ‘38’ and around a 20 Km radius of the city and latitude higher than ‘40’ for Northern USA. In conclusion, the main job of this SeachTweets class is to collect the tweets and store them in a file while also getting rid of the duplicates.

Analyzing Tweets

Tweets are analyzed based on the total sentiment value of each word in a tweet. We make an AnalyzeTweets class which handles the job of analyzing the tweets and gives various data based on the sentiment values. There is a loadHash method which loads a file, which consists of sentiment based words and their sentiment value, into a HashMap. We implement a HashMap here because we want fast access as we are going to work with a lot of words in the tweets and to check the words of each tweet to the words in the HashMap is only O(1) so we get fast and easy access to the dictionary of words-sentiment. We load the Status tweets in an arrayList because we have to access all the elements in the arrayList so the runtime is O(n) for iterating through the arrayList. In the sValueOfTweet method, we calculate the mean sentiment value of tweets which is their total sentiment value divided by the number of sentiment words in the tweets. We assign the status tweets as ‘key’ and average sentiment value as ‘value’ and store it in a HashMap. To solve the issue of negation, e.g “This does not give happiness” might come out as a positive sentence, so we multiply the value of the next couple of words with -1 which would give off a negative sentiment value. Afterwards comes the data based off of the vast number of tweets and their sentiment value. We calculate the mean sentiment value of all the tweets, their standard deviation, their highest and lowest values and finally their percentage of positive and negative tweets in the class also using different methods.

Results and Conclusion

Northern USA

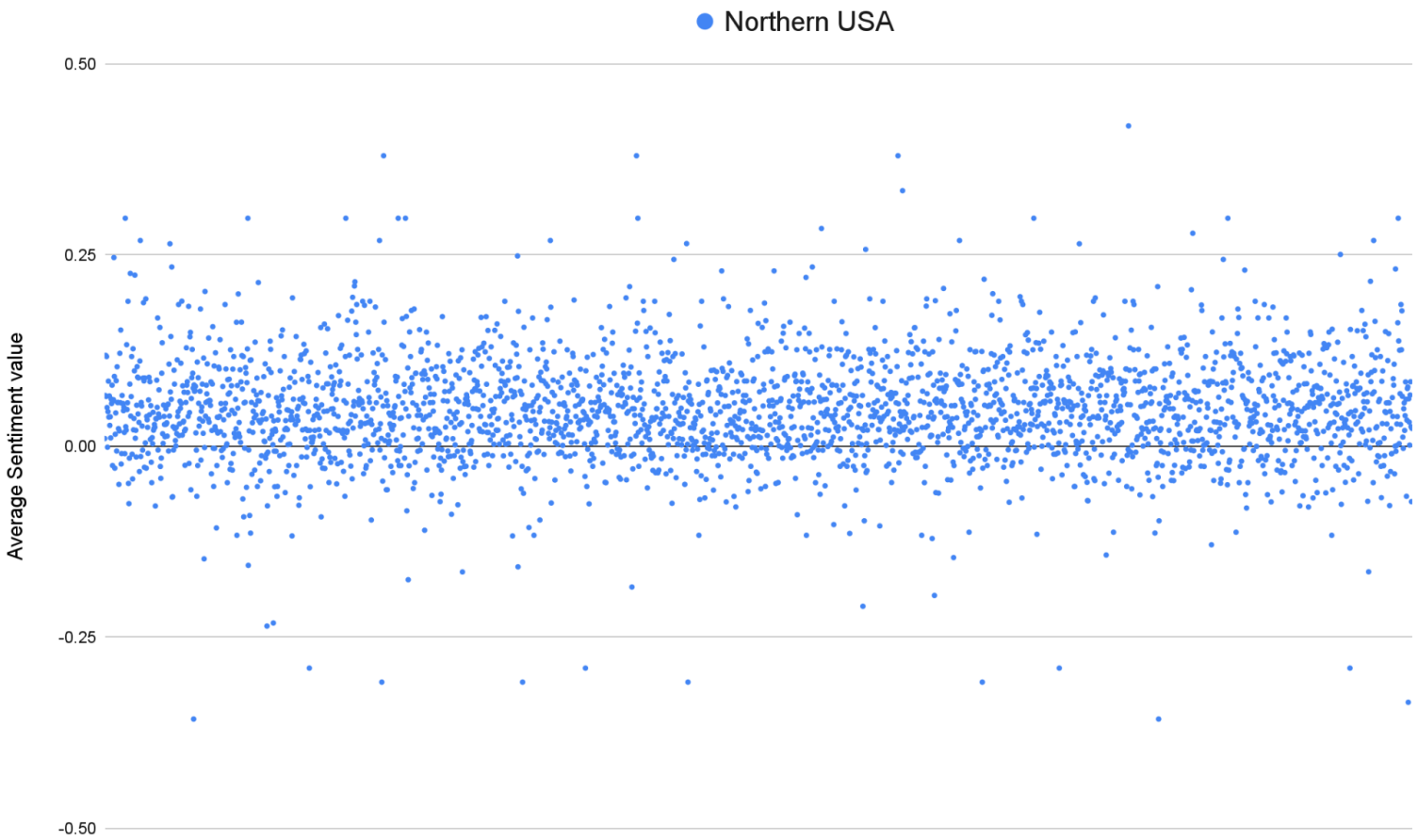
Mean sentiment value = 0.04425399116176371

Standard deviation of sentiment value = 0.06763214544649551

Percentage of positive tweets = 74.43682664054847%

Percentage of negative tweets = 19.262161279791055%

Graph of their sentiment values :



Southern USA

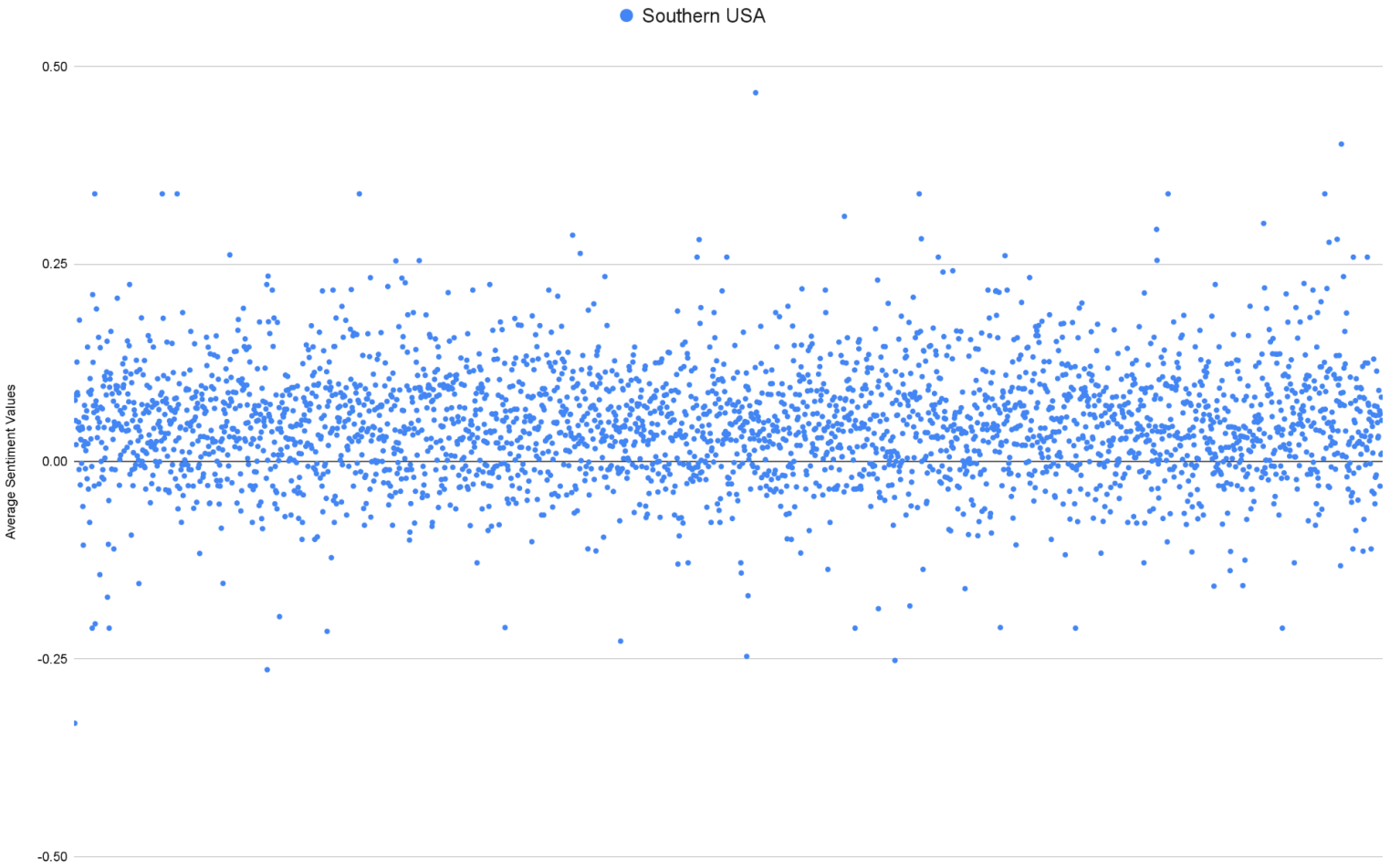
Mean sentiment value = 0.04366987410913319

Standard deviation of sentiment value = 0.0723133556805614

Percentage of positive tweets = 69.8441247002398%

Percentage of negative tweets = 22.48201438848921%

Graph of their sentiment values :



By close comparison of the two graphs and the measure of the percentages of each location, we can tell that the people in the Northern USA have a positive response to the vaccine compared to the people in the Southern USA. Additionally, this claim is supported by the evidence that Northern USA's mean sentiment value is higher than Southern USA's mean sentiment value. Hence we suspect from the evidence of the data that Northern people are more keen and positive on getting vaccinated and Southern people are skeptic of the vaccine.