Proposal for ADS final project

1. Introduction

Twitter is a popular APP among people of different age groups. When people write tweets, they convey emotion by selecting positive or negative words and arrange them in certain order to make sentences. With Natural Language Processing, we can turn tweets into numeric numbers by evaluating their sentiment, which includes how strong their emotion is and whether it’s positive or negative.

1. Statement of the problem

In large cities like New York City, the average stress level is higher than in other places. When people are stressed they have less control of their emotion and sway between the two extremes of excessively positive and excessively negative.

The practical meaning of doing this project is to find a pattern of sentiments of citizens in New York City, both in different districts and in different time periods. Then we can look into time when sentiment is most negative and area where people are more disturbed to figure out what is making them so.

1. Methods

First we request key and password of Twitter API to access tweets. We plan to use tweets sent from places in New York City over the past three years.

Next we use an emotion dictionary which contains a list of words that convey positive emotion and a list of words that convey negative emotion to evaluate the sentiment of each tweet by finding out how many words in the tweets are in the emotion dictionary and compute the overall sentiment by adding weight to each word and summing them up.

Then we visualize the sentiment data using GeoJSON structure.