1. **Project Overview**

The data source that I used was Twitter and the technique I used to analyze/process it was a hybrid of the Twython search and the natural language processing toolkit. On a coding ability level, I wished to learn how to secure API tokens from websites such as Twitter and get a better understanding of the techniques that computer scientists use to analyze speech. On a personal level, I wanted to settle once and for all with a friend whether Marvel or DC is better.

1. **Implementation**

To make clear, the overall idea was to take the positive processed natural language, subtract the negatively processed natural language, and see which comic book company was left with the highest results in relation to how much people like their movies. This requires receiving statuses from Twitter view importing Twython from twython as well as gaining access to the Sentiment Intensity Analyzer via importing SentimentIntensityAnalyzer from nltk.sentiment.vader. I retained the examples from the assignment htm file for reference as to how to properly format the code to get the proper results.

The results were to be produced via a function so that it could be called with the input of two lists of the top five movies from each comic book company, giving both companies a fair shot at winning through the results. Two dictionaries were prepared, avg\_score\_1 and avg\_score\_2 which would resultedly show the average response to each movies, to compare what each other were getting and at some times, what the patriots were receiving via Twitter through the natural language processing toolkit. As to save time for computer processing, the count was limited to 50 statuses, making the responses evaluated fairly recent and subject to change.

The main process of producing results was through for loops, first looping through a list of the two lists of movies, one for each comic book company. The next for loop, looped through each movie, and then was followed by another for loop, going through each status. This seemed to me to be the most efficient way of processing the statuses from each movie from each different list. If the movie was from Marvel, it would affect Marvel’s total score and if the movie was from DC, it would affect DC’s total score. This was done through if statements and then for loops, looping through the dictionaries produced when scores are created, and then creating new dictionaries that would allow for the totaling of scores without affecting the original dictionaries. Lastly, the scores were then averaged, thus through more loops that divided the total scores by how many statuses were counted, which was 50 per movie for 5 movies. An alternative that was considered was creating multiple functions, such as creating a function for the looping, creating another one for the creation of the total scores, and another function for averaging scores and then calling them all inside one other function. Because only one task was being completed and there would be no need to call any function within multiple other functions, the choice was the create one function instead.

1. **Results** [~2-3 paragraphs + figures/examples] Present what you accomplished:

What has been accomplished is an evaluation of what the general public (aka Twitter Users) feel about each comic book company’s top 5 movies at this present moment in time. We get to see the average feeling of negativity, neutrality, positivity, or compound feelings that these twitter users may have in relation to these movies. These come in the form of dictionaries that clearly present all the information to be clearly seen. The negative reaction is then subtracted from the positive reaction to determine which company has overall the better results. What has been discovered through the natural language processing of text analysis is that people generally enjoy the top 5 Marvel movies more than they tend to like the top 5 DC movies.

What has been most interesting about the results are the overwhelming feeling of neutrality that people possess in relation to these movies. It is as though these movies are so ingrained in our culture that people are becoming more and more indifferent towards the ideas of superhero movies in general and that tastes may change in the coming years as a result of this neutrality and this indifference. Not only do we get insight as to the general feelings of Twitter users about these movie, but we get insight into their psychological and sociological behavior and the emotional reactions that are triggered by superhero movies in this day and age where the market is saturated with them.

1. **Reflection**

From a process point of view, because I recently had a conversation with a friend who is obsessed with the DC cinematic universe and my preference is more towards the Marvel cinematic universe, the idea generation was simple. I knew exactly what I wanted to do the second that I realized what the resources available to me could do. The project was not appropriately scoped, because of other priorities, I did not have much time to complete it before the deadline. Going forward, I could generate market data towards an idea or something that the public has been exposed to because I could see what their reactions look like through the natural language processing toolkit. What I wish I knew was how much time I would have to complete this as well as possibly wish I knew some people in our class. Because I had no teammates to work with, the work was not divided.