Twitter Sentiment Analysis:
How does Twitter feel about the Asian population during the COVID-19 era?

The rise of COVID-19 has posed issues for every population of people in every aspect of life. We wanted to give insight on sentiment towards the Asian population.

CPSC 353: Data Communications and Computer Networks
Team - Brent Jang & Alberto Chi Kan Ng
Professor - Michael Fahy

Date: Jan 2020 - April 2020
Programs: Python, Tableau, Twitter API, GitHub
Location: Orange, CA

01 — Introduction

Background
Wuhan, China has been pinpointed as the epicenter for the disease now known as COVID-19. As the knowledge became public, there was an increase in violence towards Asian people around the world, regardless of their cultural background. With this project we hope to gain insight and spread knowledge about this global anti-Asian sentiment and realize the severity of this issue.

Our Solution:
Our solution is a workflow that connects a text analysis model to measure sentiment in a tweet to COVID-19 data shared by JHU CSSE. With this aggregated data we will build out different graphs to represent the general attitude towards the Asian population during the international growth and spread of the COVID-19 virus. The period of time that we are looking at is between January 22, 2020 and April 26, 2020, but hope to continue this research beyond the scope of this project.

02 — Research & Definining the Solution

The COVID-19 Data
John Hopkins University University Center for Systems Science and Engineering has been the leader in accurate reporting of global data for the COVID-19 virus. Therefore, we chose this to be our primary data source.

Twitter API
We used the Twitter API in order to pull all tweets containing the selected keyword from the beginning to the end of the timeframe. The data rate limit only allowed us to pull 1000 tweets per day of our analysis timeframe.

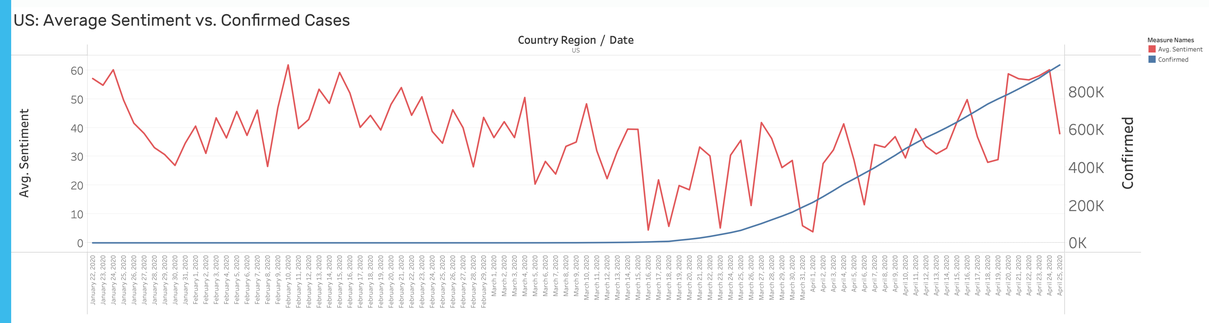
The software/programming languages we used to perform our analysis included:

Tableau
Used to visually format our quantitative data in a clear and appealing way for all users to easily understand. Multiple datasets are joined into one book to aggregate data.

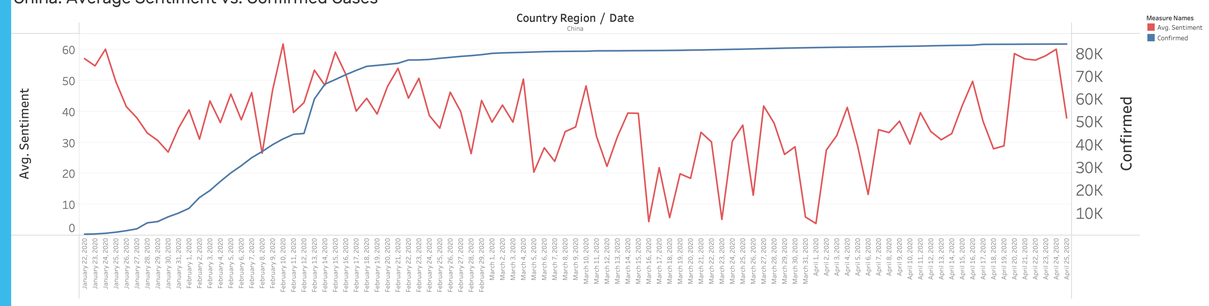
Python
Used to pull data from the JHU CSSE GitHub repository and tweets from the Twitter API. We also used Python for our final sentiment analysis model. Specific libraries include pandas, numpy, and TextBlob.

03 — Analysis & Results

US: Average Sentiment vs. Confirmed Cases



China: Average Sentiment vs. Confirmed Cases



Analysis
In both graphs there is a clear downward trend in the daily average sentiment towards the Asian population during this timeframe. In the US graph, the decrease in sentiment comes as a leading indicator for the increase in confirmed cases. On the other hand, the decrease in sentiment comes after the peak of confirmed cases in China. There are obvious political factors at work with the fast-changing climate, but the principle of negative Asian sentiment remains. While Twitter may not be the most reliable data source, many people use it as a platform to express their opinions and emotions on a public platform in response to virtually everything. Do to the origin of the COVID-19 virus it is unfortunate, but not surprising, to see many tweets negatively addressing this population of people.

04 — Reflection

Improvements:
In the future, we hope to obtain more data than just the random sampled daily data. The constraint on this is the time it takes to pull all of the data and with time this hurdle can be overcome. Another improvement would be the accuracy of the sentiment model as it is extremely difficult to train a perfect text analysis model.

Final Thoughts:
Our focus for this project was to provide insight into the rise of anti-Asian sentiment during the COVID-19 era. We are fully aware that there are many other populations of people who have to face this same sentiment every day and hope that in the future, this project can evolve into a voice for multiple populations to speak on the negative sentiment directed towards them.