**US Twitter Sentiment Towards the COVID-19 Vaccine VS Political Affiliation**

**Sophie Knight**

**Background:**

As the US continues to flail beneath the devastation of COVID-19, the country remains optimistic for an economic emergence. It has continued to amass enormous loans in anticipation of a rapid recovery. The timeline, however, remains ambiguous; assuming the upcoming Pfizer and Moderna vaccines would be distributed at a rate of about 6 million per week, it has been projected that cases will decline in late December and that re-opening will begin in February. However, this conclusion is largely based on the assumption (among others) that herd immunity would be reached, demanding that upwards of 160 million Americans receive the vaccine. Americans have historically not been so keen on needles, as only 40 percent of adults from the ages of 18-64 get regular shots recommended to them.1

A McKinsey poll revealed that this pathway might not be so narrow, noting that consumer sentiment is “more complex than likely/unlikely: more than 45 percent of adults want to wait and see.’2 Addressing this ambiguity is key to forecasting an American economic recovery. Although Twitter may be a narrow lens to view a broader and more complex issue, this project will attempt to analyze sentiment by area based on the social media platform.

**Data:**

Twitter.com (tweet/user data)

**Method:**

1. Create Twitter developer account
2. Stream tweets on python using tweepy API
3. Organize data into pandas dataframes
4. Convert pd dataframes into PostgreSQL
5. Query

**Limitations:**

Sentiment analysis is highly prone to error. To attempt to combat this, I used two different methods: textblob and Vader. Neither of them are perfect, since they weigh the positivity and negativity of words. Since the first has neutral while the second doesn’t, it was difficult to get a perfect comparison of the two results. However, it’s clear that there is enough of a discrepancy between them to render them faulty within this specific context. Also, just by reading the tweets, it’s easy to notice that many are labeled incorrectly as the sentiment analysis doesn’t account for sarcasm.

**Challenges:**

For the life of me, I could not figure out how to convert the dataframes into SQL. I tried many, many methods, until finally I realized I was not putting quotation marks in one specific spot.

I also struggled to find out how to pull the specific country from each tweet, as well as a Boolean verified.