Jed Meng

Wrangling Twitter Data

While working on this wrangling and analyzing project, my biggest obstacle was querying and looping through the tweets and gathering the data. This section of the project took most of my time and I think it really challenged me. It took me a while to realize I needed to have a try and except statement in my code. I would keep producing loops I thought were right but then would fail on me for some reason but once I realized there was just some missing tweets, I was able to fix my mistakes. After getting through the AP and looping through tweet id's, writing it to a JSON file was a just as difficult. Thankfully after reading the article provided and a bit more research and testing I was able to get everything into the JSON format and begin my analysis.

In my opinion, the wrangling section was a lot more fun as I got to look into the data and start forming my own ideas on how I wanted to data to act and look like. I found most of my issues just scrolling around in the dataframes and seeing what could be improved on. I also had to read everything on what the data actually meant as well since I felt like I did not really understand what I was looking for. There were a lot more things that I think I could add on to my cleaning steps such as changing all of the numbers that were id's to strings but I think with the analysis I was going for, the cleaning I performed was relevant and sufficient. Combining the stages of the dogs was probably the most difficult part of the wrangling. Figuring out a way to combine all of the columns and get rid of the None's was one part but then the issue of having multiple ages was the real challenge. I ended up just putting a comma in between and trimming the sides. My main goals were to get rid of retweets, fix capitalization issues, dropping columns I did not need, and create a column in each dataframe that I would be able to use to merge later, and finally merge and clean the rest of the data. I was able to accomplish this and use the resulting dataframe for my analysis successfully!