Data wrangling Process:

A guided data like what I did in this wrangling process is much easier than actual life, but we still trying to find out how to perform such a process. That's why the hardest part, in my opinion, was the Data Assessing, It's not easy to find out if your collected data was clean or not especially with the growing amounts of data, it really needs much experience and more practicing to learn how to find it out and also to make sure your data is now clean before working with it. It might be the main purpose of having big-data science and such other fields of science. Let's now have a small overview of what I found during this journey step-by-step.

1. Data Gathering (Collecting data):

This might seem to be the easiest, but it's not when you need access to outsource or an API, the Twitter team were helpful, but it's still free and for learning purposes. Not all other resources will be free or even easy, data resources vary and some of them might need more powerful tools like AI especially if we need to collect it from human speech or other voices. For me, the worst part was waiting till twitter accept creating my account, especially that they sent me an email asking for some clarifications and the mail was directed to the junk folder.

2. Assessing data (The hardest one):

This time I was lucky to have someone to tell me (Hey, make sure you have some unclean data!). I tried as hard as I can to find out what's wrong, I made the most effort to find out the unclean data. I was lucky to see some of them and that some names were a, an or the, in much more data and fewer mistakes it might be invisible and we might need a names dictionary to check names with.

3. Cleaning data:

Now, it's time to make sure you have good data. Try to find how to do them, and search for errors, if a code cleared the whole data instead of the unclean only, get back to your backup and fix the error after re-running the whole code again. It's great to have a backup of your data.

4. Make a clean copy:

It's good to save it to a file as a backup and to make your analysis to it. It's not part of Data wrangling Process in the course but it's the final part to save your work (The Ctrl+S)

THANKS FOR YOUR TIME