Final project plan:

Diagram

Description automatically generated

Question: What discussions have been happening around COVID on Twitter concerning vaccine hesitancy? Specifically:

1. What are the salient topics discussed around COVID and what each topic primarily concerns?
2. Relative engagement with those topics?
3. How positive/negative the response to the pandemic/vaccine has been?

# Nov 22th-27th :

## Data Collection:

1. Collect 1000 tweets within a 3-day window (TODO: each person collects 333 one day)
2. Filter tweets such that:
3. Language is English and time zone cover North America
4. Hashtags or words that mention: “COVID / COVID-19”, “vaccination”, “Pfizer”, “Moderna”, “Johnson & Johnson” (more brand names should be included)
5. Every tweet is unique, meaning that no two tweets should be identical

# Data Analytics:

1. Develop topics on 200 tweets: (should decide on topics together)

TODO: Each person in the group decides topics based on a sample of 66-67 tweets from their own data set, then we will discuss which topic to keep or add to limit the number of topics to 3 to 8.

# Nov 28th – Dec 2nd:

## Data Annotation:

TODO: Annotate the rest of the tweets with exactly one topic. Also code each tweet for positive/negative/neutral sentiment.

TODO: characterize each topic by 10 words with the highest tf-idf scores in each topic.

Dec 3rd – Dec 5th:

## Interpretation:

What are the most frequently used words in each topic and what dos it imply about people’s attitude towards pandemic / vaccination? (Probably need more information/data modifications to gain a better idea in answering this question.)

TODO: Feel free to add your ideas here!

Dec 6th – Dec 13th :

1. Improvements on better method
2. Report writing

Side Note:

* How should we present the data?
* Is our ultimate goal to answer the questions as well as possible with the results we would have?
* Need innovative design ideas that impact our results aside from the instructions.