Metrics:
- Overall Tweet Metadata [tweet_metadata.csv]
    - ID Needed (to get the contents)
    - Retweet Count
    - Favorite Count
    - Quote Count
    - Reply Count
    - Follower Count of the Poster
- [user_tweet_career.csv]

Overall focus:
- Whether or not a user is censored.
- There's a data set covering all users who have been completely censored, or a data set of all users who had at least one tweet censored.
- Datasets (Input)
    - Set A: Censored Tweets
    - Set B: Blacklisted Users (Censored Profiles)
    - Tweet Database
- Process
    - Make an analysis of tweets from both datasets.
        - Set A + Tweets found in database
        - Set B + Tweets found in that database
    - Will need to determine metrics in both sets, and their impacts.
        - Metrics for tweets
            - Interaction Data (Retweets/Favorites/Quotes/Replies/Follower Count)
        - Metrics for users
            - Consideration for popular users (high follower count/interaction/tweets) or users with minimal impact who were censored.
    - Will need to find similarities and trends between both sets.

User Data:
- Censored Tweet Count
- Followers
- Views(?)

Trying to establish lower/medium/upper limits of data to get better distributions.
Limitations:

Steps:

- We need more ways to analyze the data, with specific metrics.
- Interaction Metrics for both sets.
- Measure the moderation speed.
- Filter by country.
    - Initially divide data by country!!!
    - More of a distribution.
- Prediction system based on tweet text.
    - In each country, find the most common keywords that flag censorship.

Paper Takeaways - Ashwattha

- We perform the topical analysis by computing the most popular hashtags, mentions, URLs and then observing those entities. We measure the popularity by the number of unique accounts mentioning each entity. We use this metric instead of the tweet count in order to account for the same users using the same entities over and over.
- We additionally report the most frequent tweeting languages (measured by the number of tweets) and most reported locations (measured by the number of users) with respect to the censoring countries in Table 4.
- We found that censored users are mostly based in a country other than they are censored in, beyond the reach of the law enforcement of the censoring country.

Questions we need to answer -

- The dataset could be used to measure the effect of censorship on censored users' behavior. Do users forgo using their accounts after being censored, or does the censorship backfire?
- Furthermore, what is the effect of censorship on other users, e.g. does the public engage more with censored tweets?

● Can also look at 22mill additional tweets from users who had atleast 1 censored tweet to see how that affects future interactions after they have atleast 1 tweet censored
● Do users forgo using their accounts after being censored, or does the censorship backfire? Furthermore, what is the effect of censorship on other users, e.g. does the public engage more with censored tweets?
● Patterns in tweets that were censored / common words
● Tweets critical of government they are censored in
● Look at date tweet was censored in country, what policies came around before then