

Our project collects and analyzes data from Twitter. The project has two main components: the twitter data miner and the data analysis functions. The twitter data miner collects every single tweet containing specified key words as they are being tweeted in real time and stores the data and tweet ID in a CSV. The analysis functions allow us to compare trends in keywords.

Our repository contains a few files. Due to the nature of the data miners, they are stand alone functions (discussed with the professor). All miners print all tweet data in real time.

- streamMiner
  - Collects tweets containing a specified key word in real time. Parses the data, and then stores only the date and tweet ID (unique identifier for each tweet) in a CSV labeled for that word.
- activateMiners
  - Should activate all 3 data miners at the same time but only does the first because I did not code it for parallel processing. May slow down your computer.
- Main
  - Organizes all analysis functions to be called from the command line
- makeData
  - Makes a weeks worth of fake data for the analysis functions
- Pastminer
  - Collects previous tweets with a keyword. unreliable.
- checkDupes
- countTweets
- Covidproj
- Pandas\_graph

Sources:

<https://developer.twitter.com/en/docs/api-reference-index>