

# Identify drug dealers on Social Media

SFSU - CSC 664

Ahmad Afghan

Brooke Porter

Jessica Serrano

# Getting Our Data

- We were able to register a development account with Twitter for access to their API.
- After parsing for random tweets, we now have upwards of 2 million tweets available for use.
  - These tweets are very random, and must be filtered down.
- They are saved in numbered CSV files. Using Github Actions, we are able to automatically fetch tweets every half hour.
- Github limits based on file size. Each CSV file can hold around 700k tweets before a new file must be created.

# Current Progress

- We currently have our data, and continue to get more data as time goes on.
- We have parsing built, to allow the many CSV files to be combined into one database.
- The next thing for us to work on is the database parsing, for relevant tweets.

# Development Plan

- We parse the data and save it into a database.
- We will filter through the data we currently have using a stop list of terms that do not apply to drugs and drug dealing.
  - This should drastically reduce the number of tweets we need to parse.
- We will further filter using related keywords to reduce our list to relevant tweets.
- We will create an application to show the filtered data, resulting in a list of tweets and Twitter users who we believe to be drug users or sellers.