Scrum Report

What have we worked on?

- Created draft of final presentation/insight presentation
- Working through alternative data pulling techniques
    - Tweepy is limited by the free tiers 7 day limit (can only go back 7 days)
    - Investigating premium tier sandbox to potentially pull older tweets
    - Opportunities for academic license?
- Looked for other available datasets for sentiment analysis:
    - [Sentiment analysis to detect hate speech](#) - may be useful in detecting extreme negativity but most likely not applicable
    - [Airline sentiment](#) - includes positive, negative, and neutral but only lists the tweet ID, not the text
    - [Comments made on political figures in India](#) - also includes positive, negative, and neutral but is built using political language
- Looking into using pre-trained models for first pass at sentiment analysis
- Analysis using EWMA of trade volume vs price to add as feature in identifying stock "pumps"

What do we plan to work on next week?

- Continue investigation of alternate Twitter data sources
- Work with more recent subset to identify features

Potential Roadblocks:

- 7 day limit on free tier of Twitter API

Meeting Notes

- Maybe concatenate all tweets in a period of time about a stock and train off this big string
- Can you predict stock price using favorites count (or really any numerical feature)?
- X is seven day data for a large collection of stocks, Y is the change in price
    - Need to consider scale - some stocks may be more popular
- More about finding something in the data, not about more data - focus on data analysis not collection

Insight

- 3 minute presentation - concisely describe project and any insights we have so far
- Intro, what we've done, what's next (like a midterm for the project)