Garrett Van Beek

Ben Nguyen

Intro to Data Science

Professor Nick Mattei

**Final Tutorial Milestone 1:** “Can you Google the stock market?”

For our project we plan to collect data on search traffic from Google and Twitter and see how that data correlates with the performances of certain publicly traded stocks. For example, one question we would like to ask is “how does the search traffic about clean energy sources correlate with the performance of clean energy stocks?”. We would also like to see how the sentiment about certain topics correlates to their stock. For this, we plan to use a python library Twitter Sentiment Analysis or Tweepy. This tool gives us a sentiment rating that we can compare to our stock’s performance. We will most likely analyze our data as line graphs. This way we can see how stock performance, search traffic, and sentiment change over time. Consequently, we hope to find correlations and gain insight on the data.

We hope to tackle topics that are a bit more controversial or nuanced than clean energy. Some of the questions that we have thought of asking are: “How does search traffic on sex scandals within companies correlate to the stocks of the companies?” We can also observe the effects of influencers like Donald Trump or Kanye West have on the stock market. We can observe their twitter comments about certain topics and then observe how stocks relating to these topic’s perform. Another angle we can observe is how the general sentiment on topics will correlate to related stocks. For example, we can ask “How does the sentiment about Elon Musk correlate to the stock performance of TESLA.”

For our information on stock performance we plan on scraping or downloading data from Yahoo Finance, which contains a barrage of data on publicly traded stocks. Links to the datasets, tools, and websites we plan to use are provided below.

<https://cloud.google.com/public-datasets/>

<https://www.tweepy.org/>

<https://github.com/shaypal5/awesome-twitter-data>

<https://finance.yahoo.com/>