W200 - Project 2 Proposal
Wall Street Bets Posts Impact on Stock Price Movement

Team Members:
Vijay Ranganatha
Ricollis Jones
Ryan Goding

GitHub repository: https://github.com/UC-Berkeley-I-School/Project2_Ranganatha_Goding_Jones
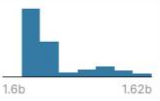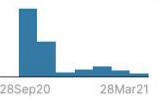
Primary Dataset:
Reddit posts from subreddit WallStreetBets, downloaded from https://www.reddit.com/r/wallstreetbets/ using praw (The Python Reddit API Wrapper). Data set can be seen at https://www.kaggle.com/gpreda/reddit-wallstreetsbets-posts,

Columns Present:

Column Name	Type	Description
title	String	Title of Post
score	int	Uptick and Downtick
id	String	ID
url	String	URL
comms_num	int	Comment Count
created	int	SystemTime
body	String	Body of Post
timestamp	DateTime	DateTime

Preview of the columns and values can be seen below.

About this file

Reddit posts

A title	# score	A id	∞ url	# comms_num	# created	A body	▭ timestamp
Title	Score of the post	ID	url	Comms num	Creation timestamp	Message body	Date/time timestamp
42733 unique values	[histogram] 0 — 348k	43724 unique values	43709 unique values	[histogram] 0 — 93.3k	[histogram] 1.6b — 1.62b	[null] 52% / Your daily trading di... 0% / Other (20822) 48%	[histogram] 28Sep20 — 28Mar21
It's not about the money, it's about sending a message. 🚀💎🙌	55	l6ulcx	https://v.redd.it/6j75regs72e61	6	1611862661.0		2021-01-28 21:37:41
Math Professor Scott Steiner says the numbers spell DISASTER for Gamestop shorts	110	l6uibd	https://v.redd.it/ah50lyny62e61	23	1611862330.0		2021-01-28 21:32:10

Variables:
We will be exploring reddit posts from the subreddit WallstreetBets to determine if sentiment or movement of a stock can be seen by the community's posts. Column values of date, stock ticker, sentiment, post score, and number of sub comments will be used.

Supplemental Datasets:
We will determine WallstreetBets top 10 focused stocks over our set timeframe and load historical price data from Yahoo Finance. A comparison to the S&P 500 will also be used to determine stock movement vs index movement. For each stock we supplement our primary dataset with columns of daily opening price values.

Report:
Our report will answer our overall question: "Does sentiment on WallStreetBets subreddit posts predict price movements?" We first plan to address this by tabulating the number of stock ticker mentions on the message board daily, communicating this visually, and overlaying this information on the stock's price movement. This will be done for each of the top 10 stocks that are the message board's primary focus. To further narrow the focus we may attempt to characterize certain slang words that the message board user's often use as either buy, sell, or hold sentiment.

- Compare the price movement of Top-5/10 stocks with S&P 500 before and after the Bets heated up
- Check the correlation of price movements based on the score/sentiment of comment
- How long did the comments able to hold the value high/low
- % Stocks which were discussed in the Reddit posts and how many actually got a high/low momentum

- Any Data insight to check if these can be used for Predicting Future Price Movements.
- Check if Negative Sentiment had higher impact or the Positive Sentiment