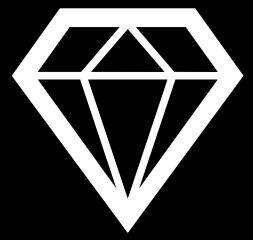


An Analysis of r/Wallstreetbets's effect on stock prices

Capstone Project - Ryan

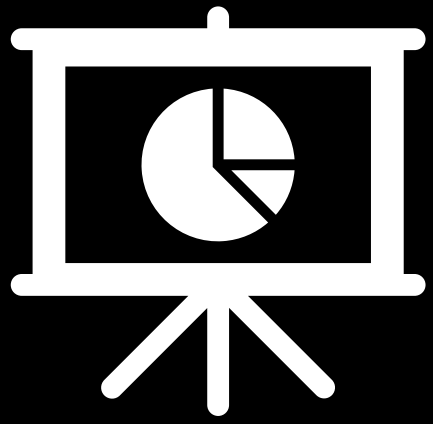
Business Problem Statement

- Johnny 💎🤝 is a MD of a foreign bank

- Seen WSB influence on gme
- Wants to use WSB's activity to give the bank a trading edge

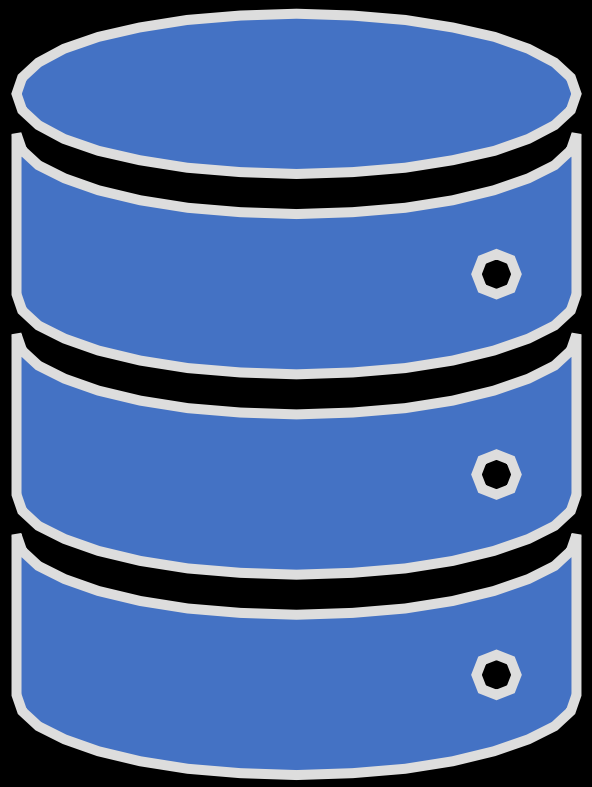
Data Problem Statements

Is there a relationship between the sentiment or number of comments on WSB and stock prices?



Can we use stock data and WSB comments data to predict the next day's price movement?

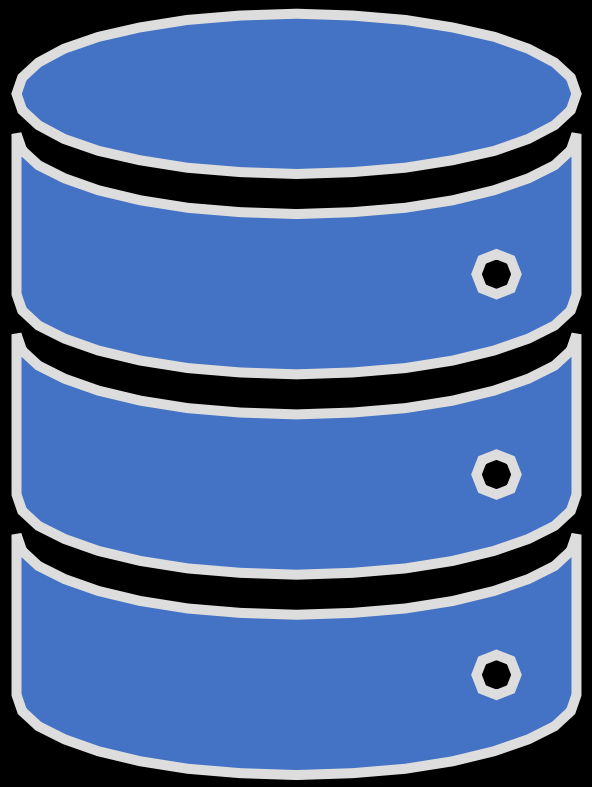
Data sources



Yahoo Finance: GME, TSLA, PLTR, NOK, BB, SP500, AMC

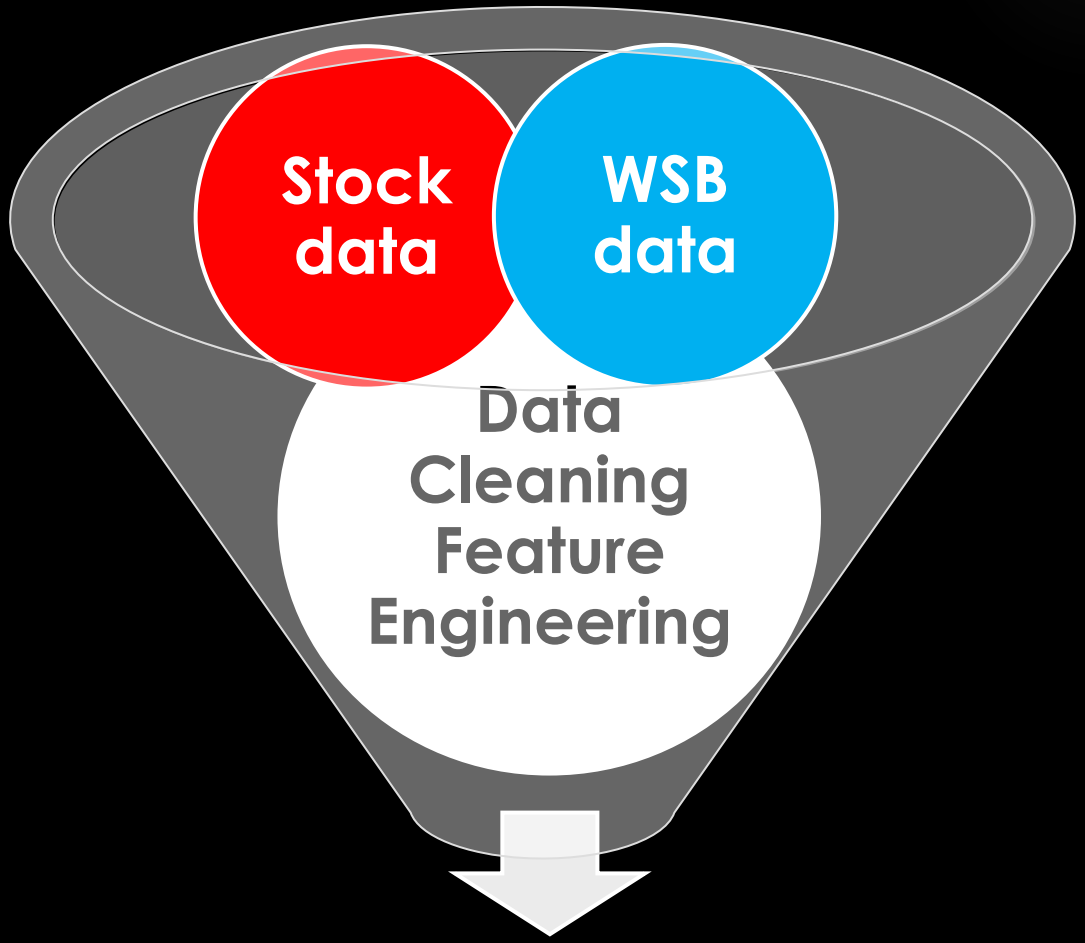
Kaggle: WSB comments (Feb-2018 to Feb 2021)

Data sources



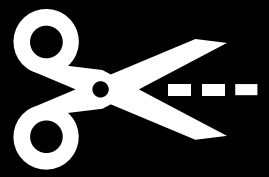
- Stock data columns:
  - **`Open, High, Low, Close, Volume`**
- WSB comments columns:
  - **`created_utc, text`**

Data preparation

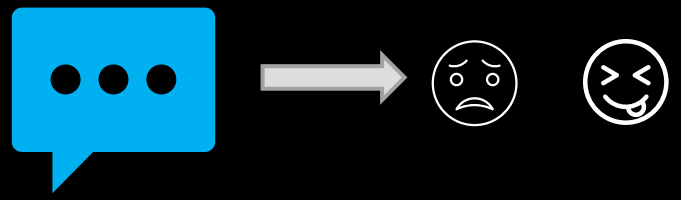


Final Dataframes

Data Cleaning / Feature Engineering
(WSB comments)



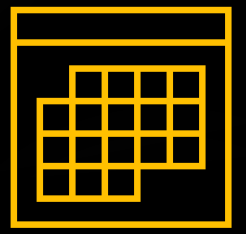
Remove NA values



Apply spacy's sentiment and subjectivity models.

Available Columns

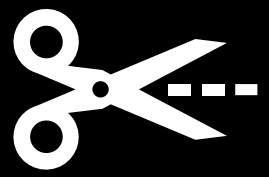
- **`created_utc`**
- **`text`**
- **`sentiment`**
- **`Subjectivity`**
- **`month`**
- **`day_of_week`**



- Convert unix timestamp **`(created_utc)`** to tz-aware datetime object
- Get **`day_of_week`** and month label from **`created_utc`**

Data Cleaning / Feature Engineering

(Stock data)

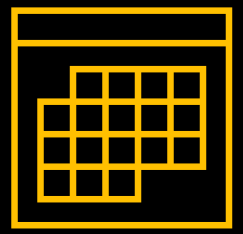


Remove all data prior to 16th Feb 2018

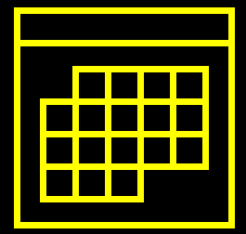
Compared **`Open`** and **`Close`** values to get **`pct_change`** column

Compared Open and Close values to get bool column **`up_today`** and **`up_tomorrow`**

Compared **`High`** and **`Low`** values to get **`pct_volatility`** column



Converted **`Date`** column to tz-aware datetime object

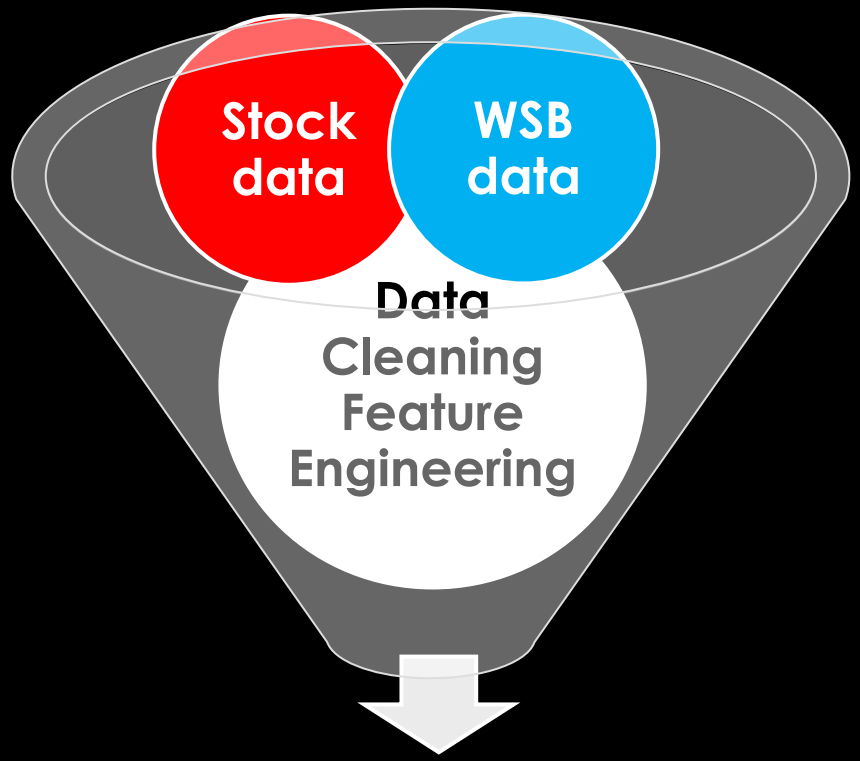


Get **`day_of_week`** and **`month`** label from **`Date`**

Available Columns

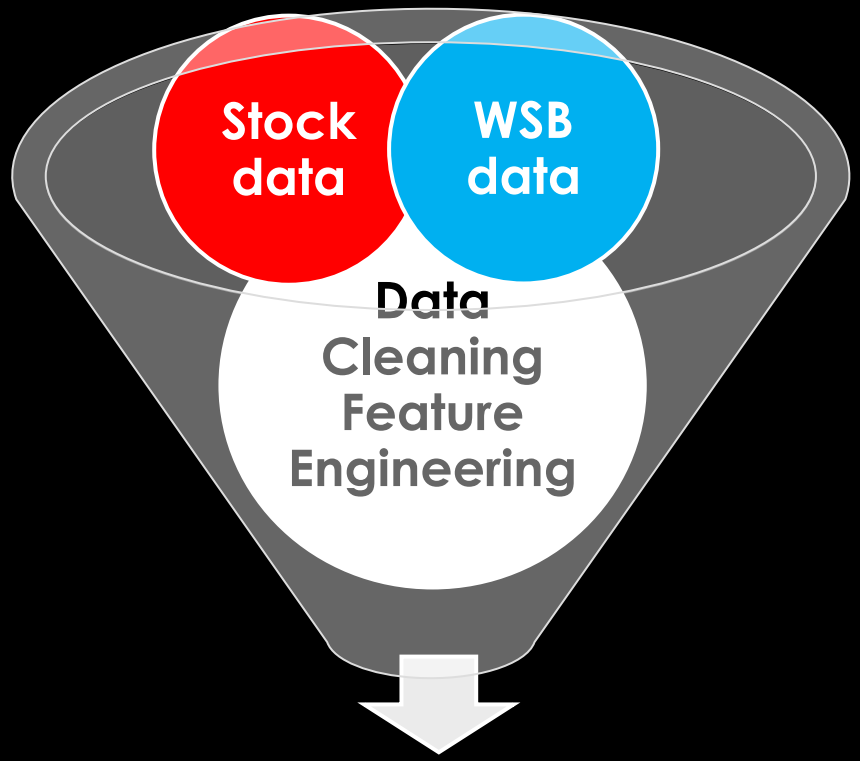
- **`Date`**
- **`Open`**
- **`High`**
- **`Close`**
- **`Low`**
- **`Adj Close`**
- **`Volume`**
- **`pct_change`**
- **`up_today`**
- **`up_tomorrow`**
- **`pct_volatility`**
- **`day_of_week`**
- **`month`**

Data preparation



- Group wsb data by date and obtain the following columns:
  - Mean subjectivity
  - Mean sentiment
  - Count of comments
- Left join wsb data on stock data to obtain 7 different dataframes
- Target variable is **`up_tomorrow`**

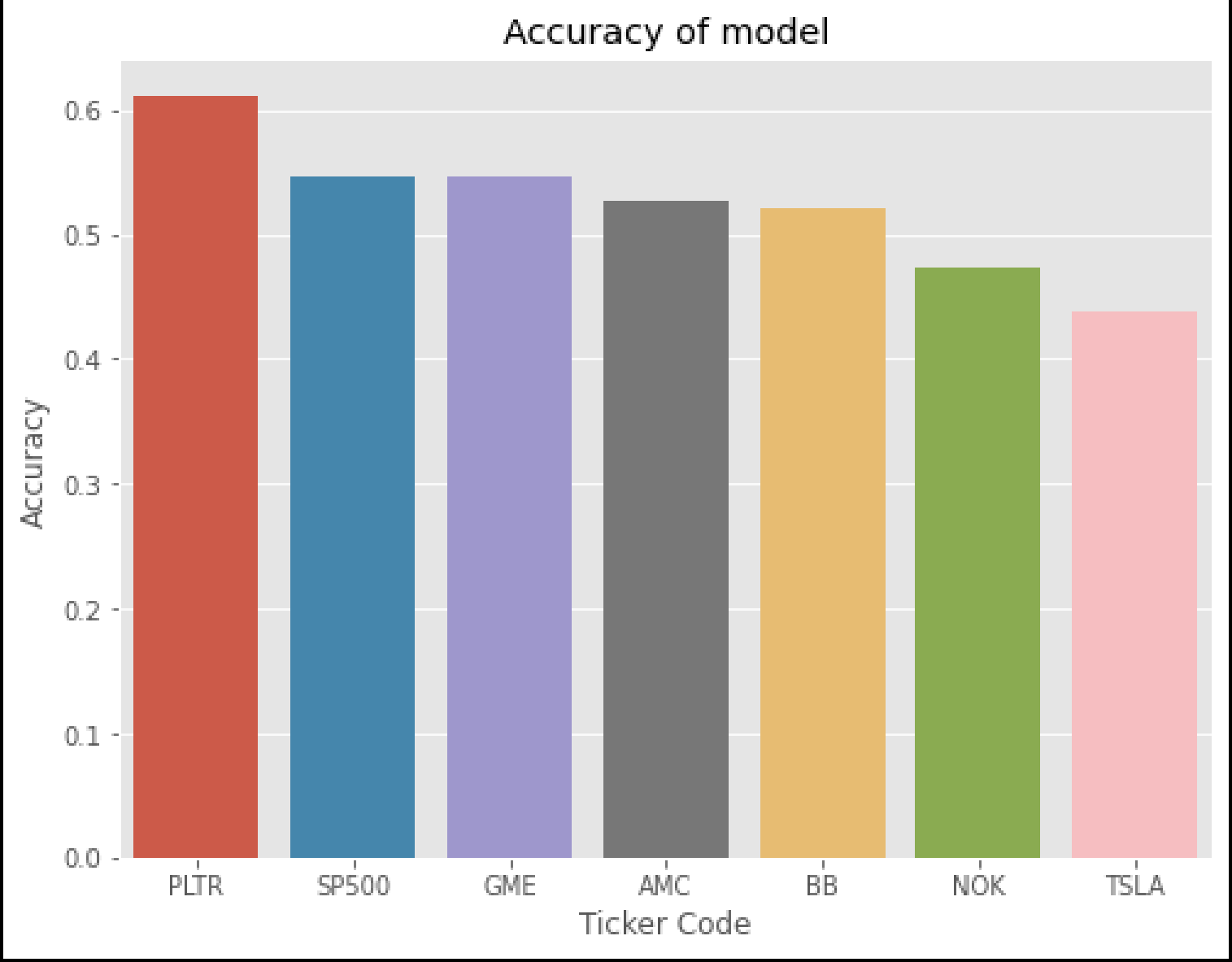
Final Dataframe Columns



7 Dataframes for each stock

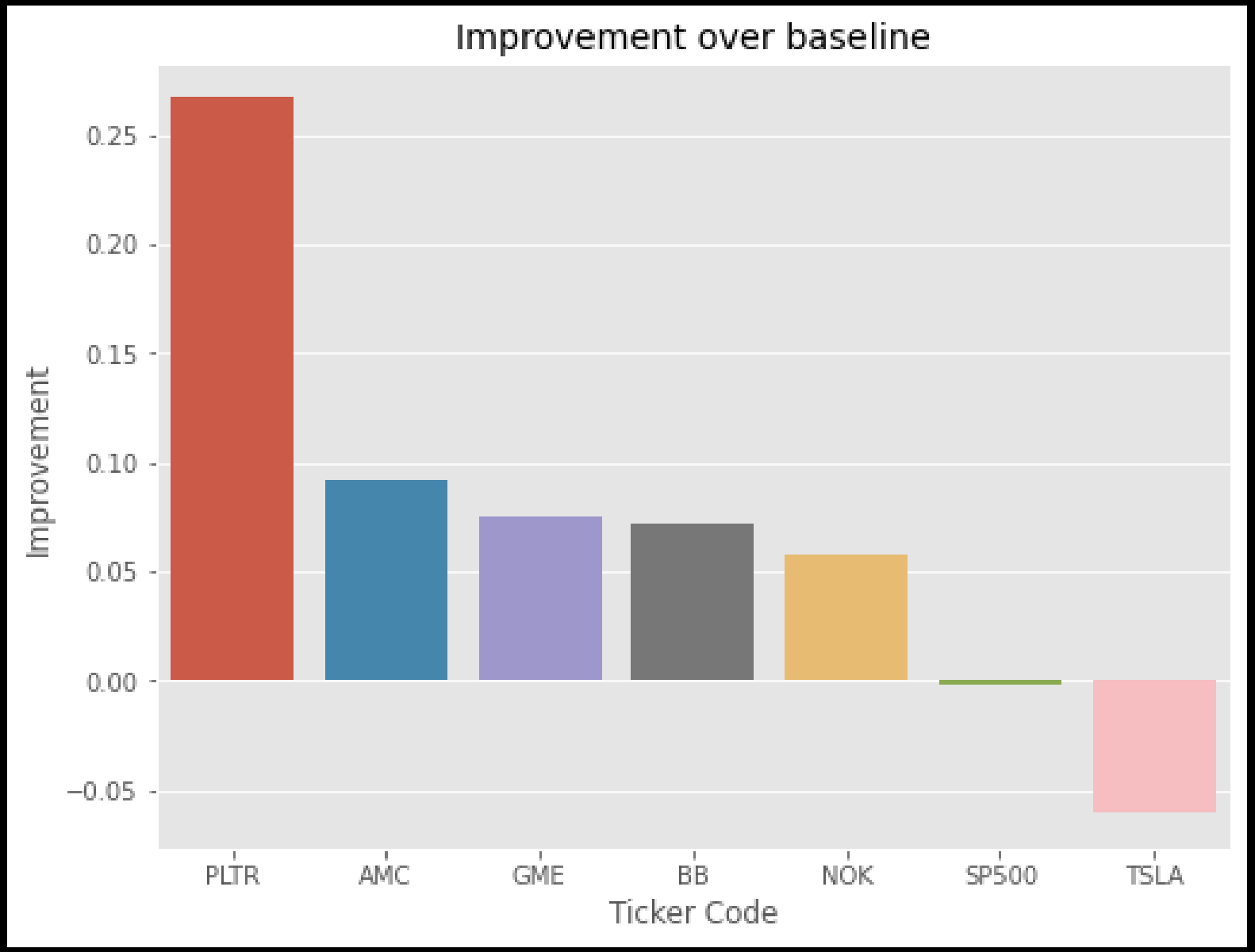
- Features:
  - Open : float64
  - High : float64
  - Low : float64
  - Close : float64
  - Volume : float64
  - up_today : bool
  - pct_change : float64
  - sentiment : float64
  - subjectivity : float64
  - comment_count : int64
  - day_of_week : category
  - month : category
- Prediction column:
  - up_tomorrow : bool

Results



- Model used:
- XGBoost Classifier:
- 80% : 20% **`train_test_split`**
- Best accuracy performer is PLTR, over 60% accuracy
- Model predicted the right movement 60% of the time

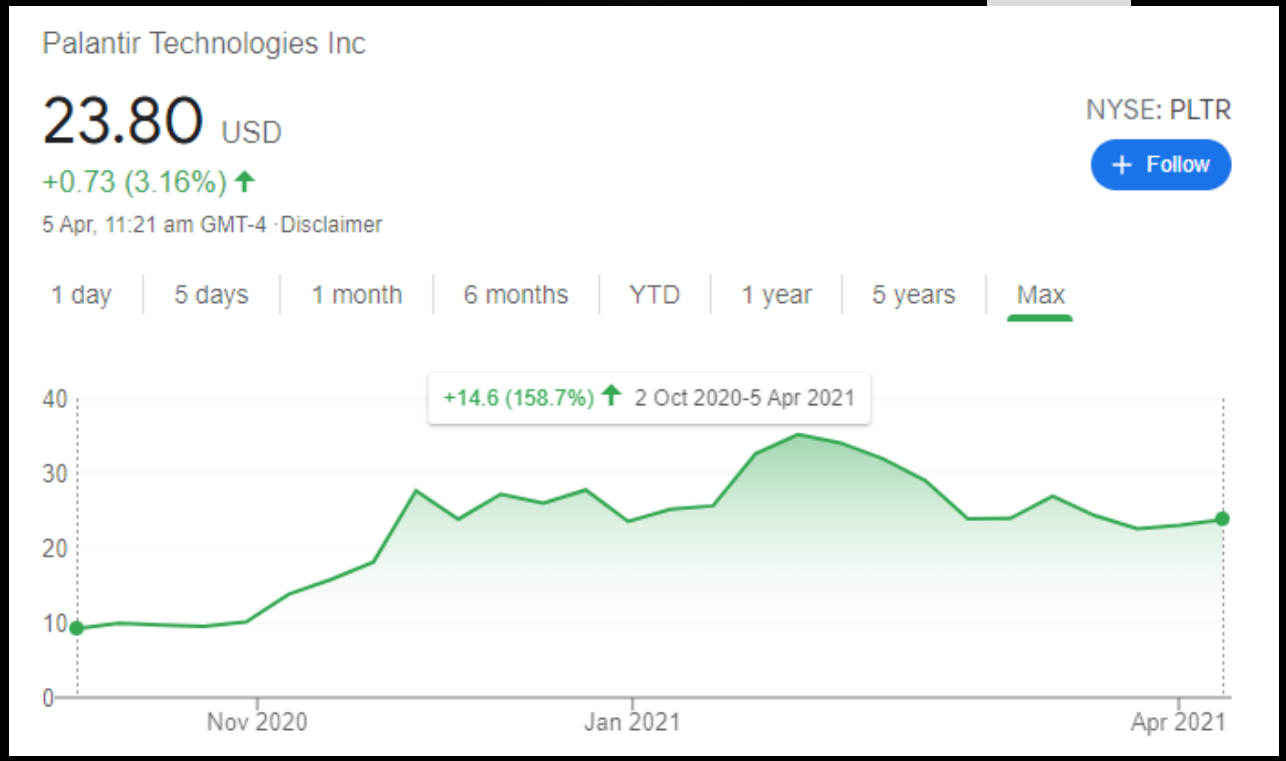
Results

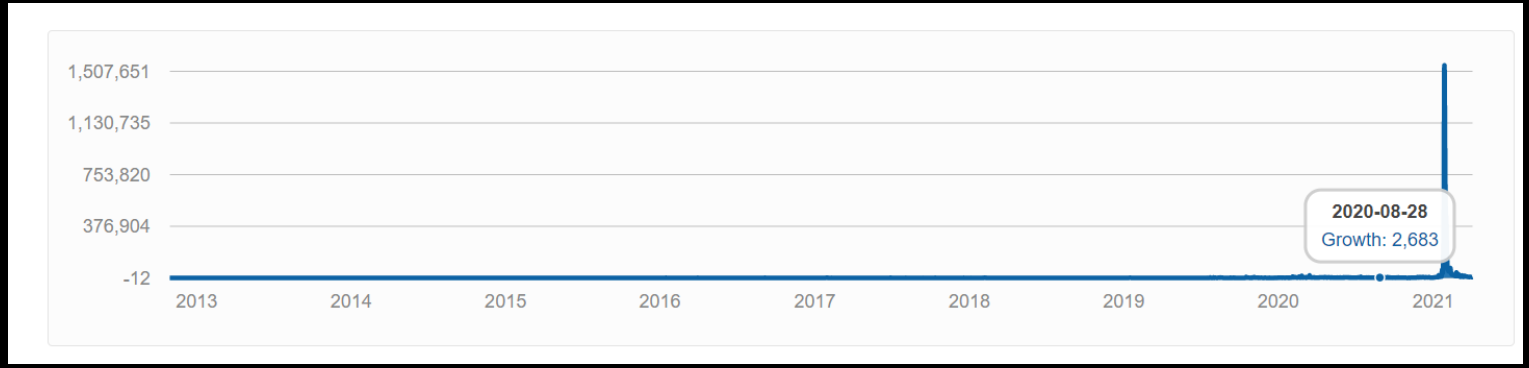


- Best improved model over baseline is PLTR
- Baseline: predict stock is up every day

Analysis of results

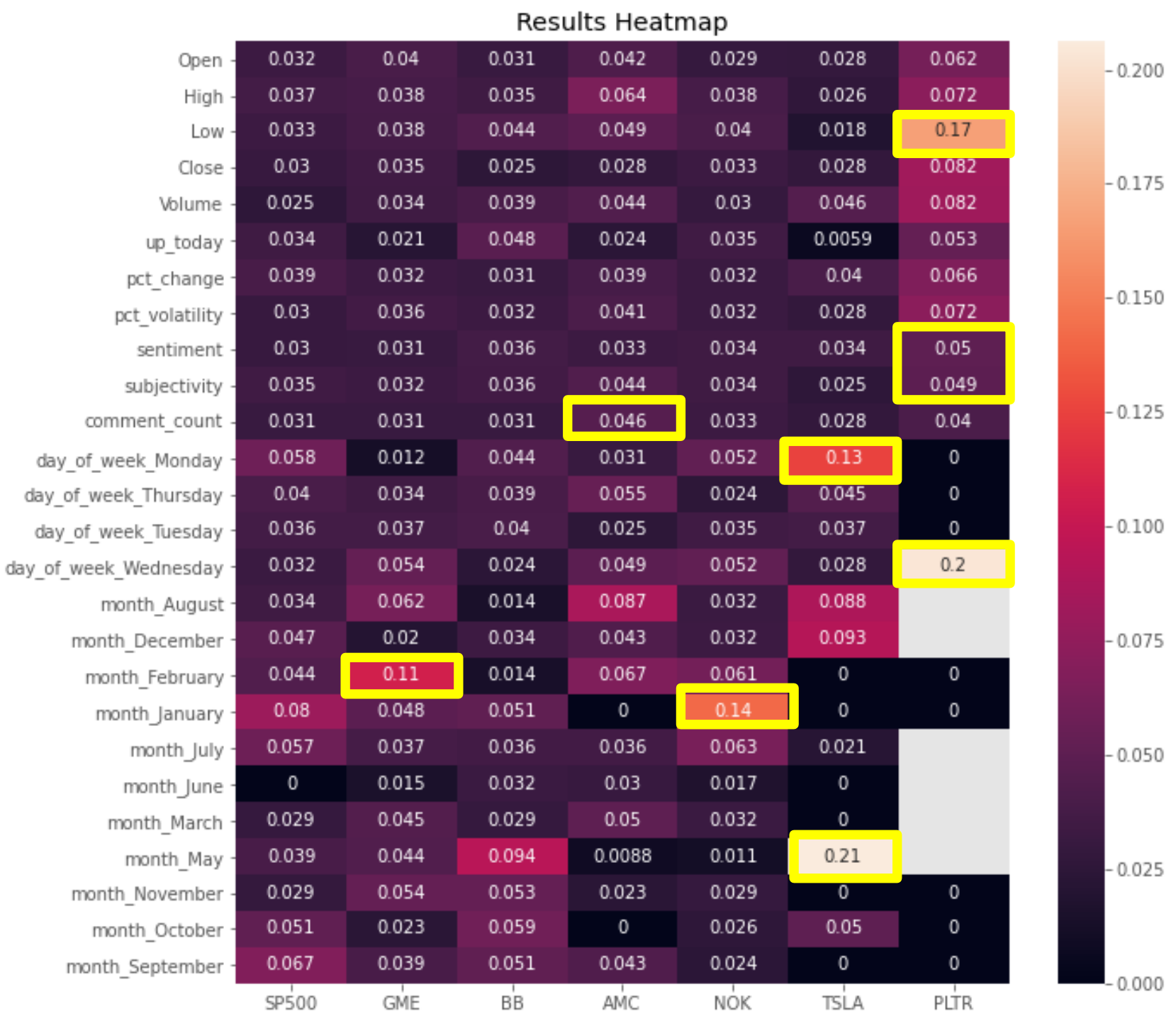
- PLTR IPO'd recently and stock data is only from 2nd Oct 2020 onwards
- For all other stocks, the analysis runs from Feb-2018 to Feb-2021
- Aug 2020 was the start of WSB's huge membership growth
- WSB's influence initially is low due to low membership count





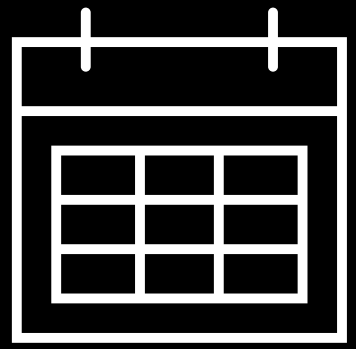
Feature Importances

- PLTR – affected by **`low`**, and whether or not it is **`wednesday`**
- TSLA – affected by whether or not it is **`monday`** and **`may`**
- Most affected by WSB **`sentiment`** & **`subjectivity`**– PLTR
- Most affected by WSB **`comment_count`** – AMC
- Most affected by the month of January – GME
- Most affected by the month of February



Future work

- Train a BERT Model on labelled data for sentiment analysis
- (Bidirectional Encoder Representations from Transformers)



- Narrow the timeframe down to Aug-2020 onwards as this is the period of WSB's explosive growth
  - Downsides: lack of data

Conclusion

- The results of PLTR show that WSB is starting to have influence on the stock market
  - The cult-like behavior in WSB is contagious
  - WSB encourages people to hold stocks through memes – something never done before
  - There is also the desire to get back at hedge funds for the 08 crisis
- The investing landscape has changed significantly since the Warren Buffet days
  - People were expecting a crash but no crash came
  - S&P broke 4000 during the middle of a pandemic
  - The economy and the stock market has decoupled