Stock Trading and Market Analysis Project Checkpoint

Angelo Vacca, Kasper Seglem, William Ruiz

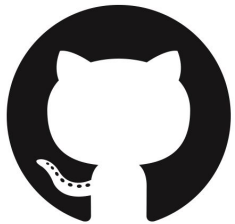
Introduction

We are going to present what we have gotten done so far and what we plan do in the future for the rest of the project.

- Dataset and Tools
- Previous Work
- Trello
- Correlated Stocks
- Proposed Work
- Evaluation
- Milestones

Dataset and Tools

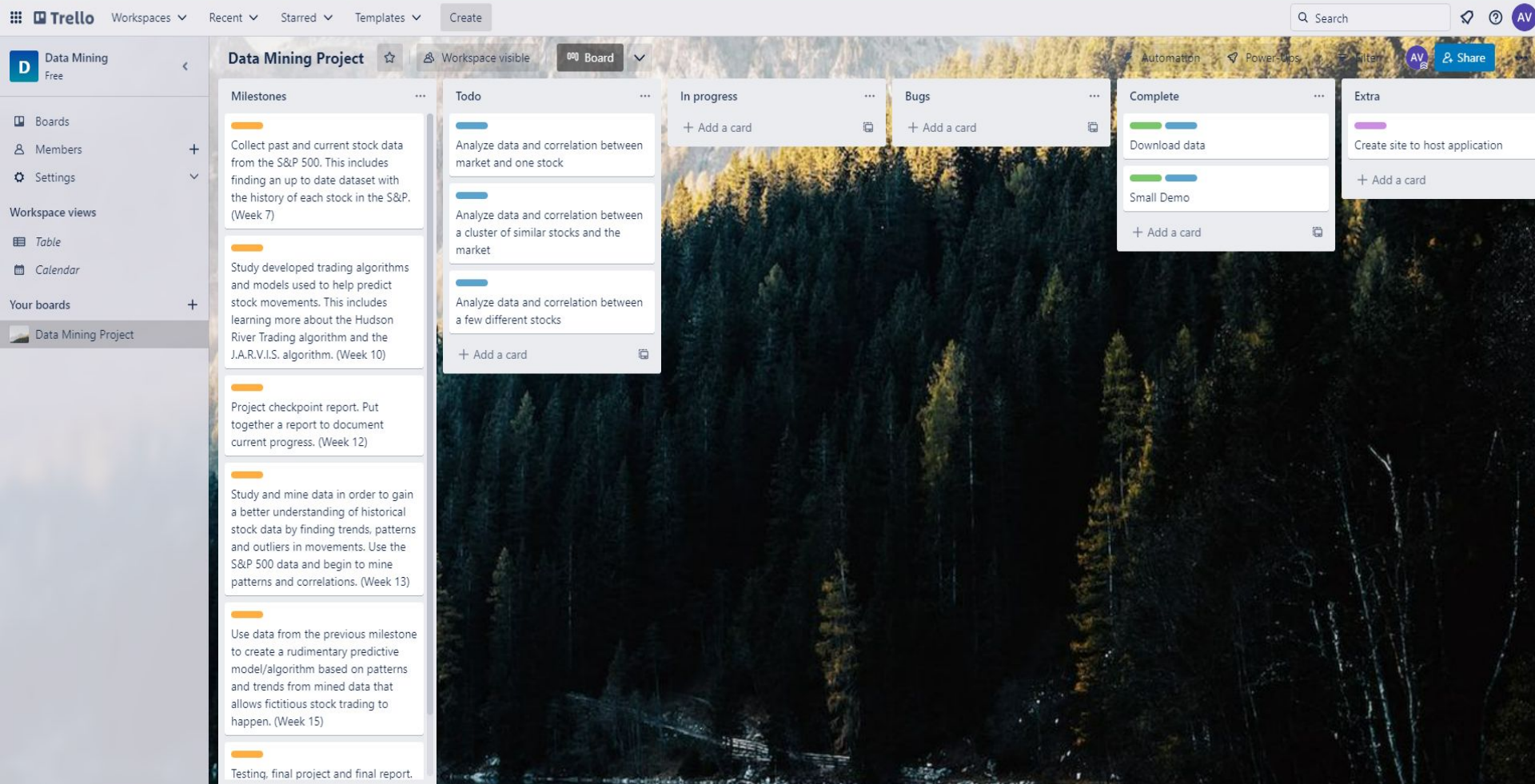
- S&P 500 companies
  - 2013 to 2018 (5 years)
  - date, open, high, low, close, volume, Name
  - 2013-02-08, 67.7142, 68.4014, 66.8928, 67.8542, 158168416, AAPL
- Python
  - matplotlib
  - numpy
  - pandas
  - seaborn
- GitHub



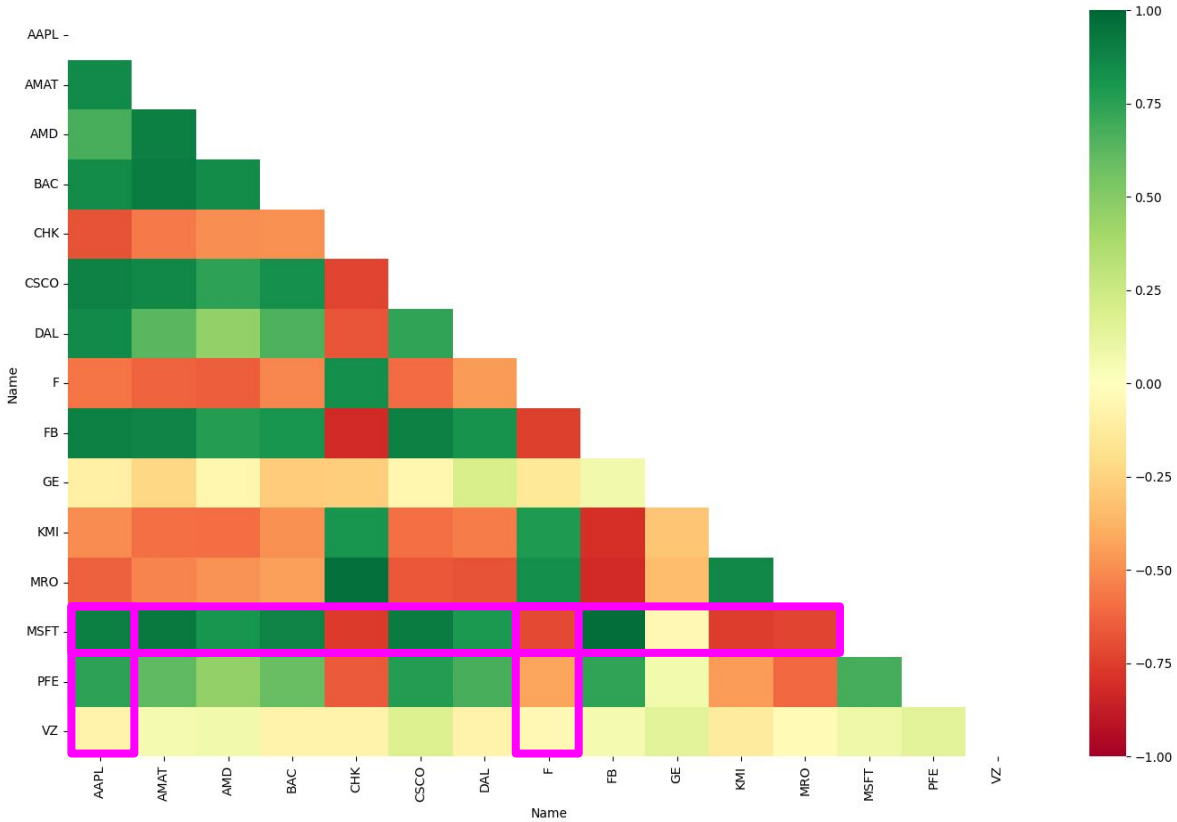
Previous Work

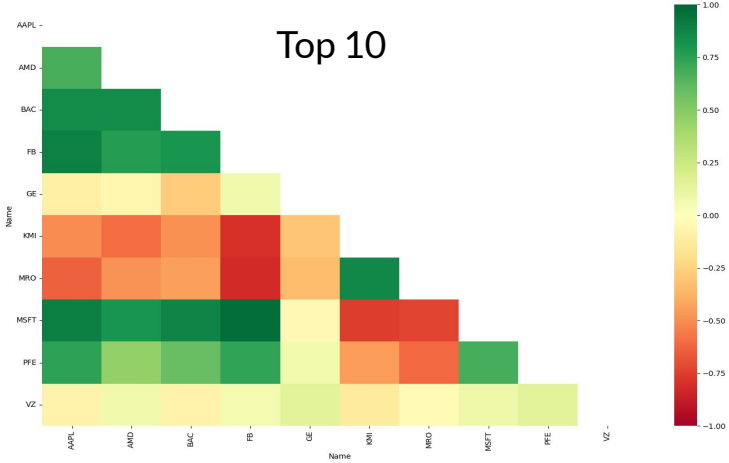
- Studied other trading algorithms
  - We spent much of the early time learning about algorithmic trading.
- Find and clean up dataset
  - For our dataset, we are using stocks on the S&P 500 from the years 2013 to 2018.
  - This dataset is provided by Kaggle, an open source dataset platform
- Find correlated stocks
  - Recently, we have gathered information about the stocks' correlation to each other.
- Set up Project Management
  - Using Trello boards we have a way to manage this project and divide work between the three of us
- Set up Project Sharing
  - Using github we have version control for our project

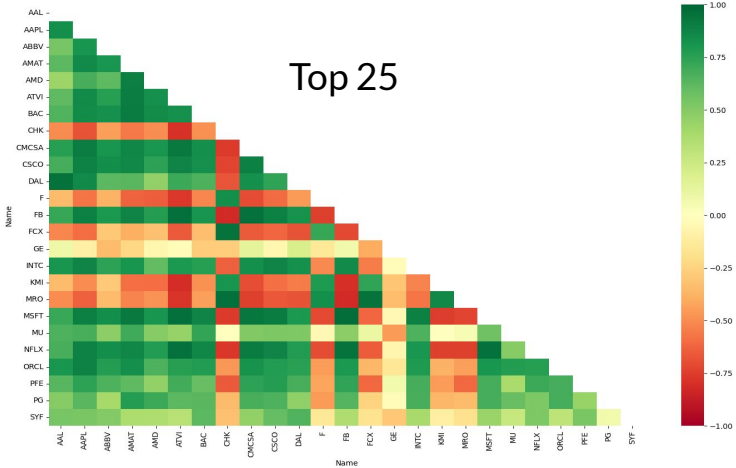
Trello Board

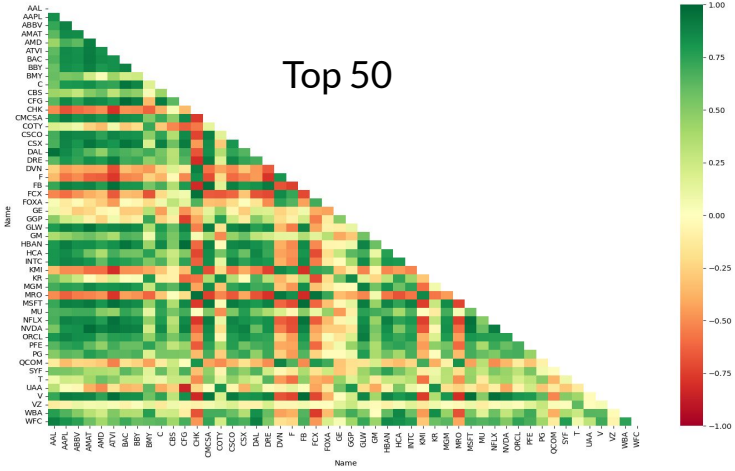


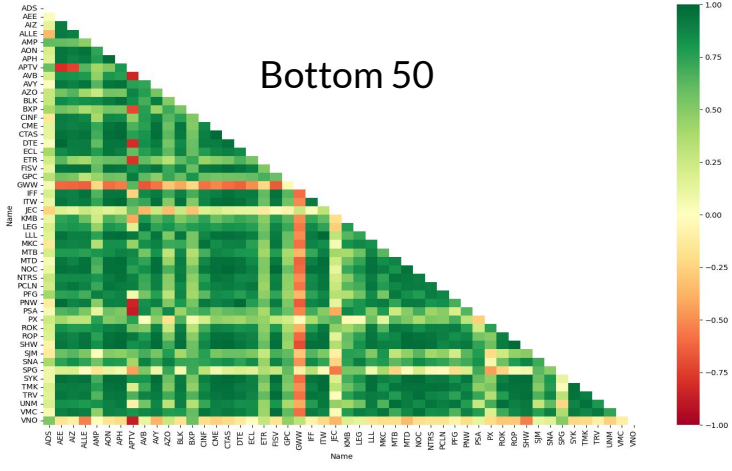
Correlated Stocks




Top 10
AAPL
AMD
BAC
FB
GE
KMI
MRO
MSFT
PFE
VZ
Name
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Top 25
AAL
AAPL
ABBV
AMAT
AMD
ATVI
BAC
CHK
CMCSA
CSCO
DAL
F
FB
FCX
GE
INTC
KMI
MRO
MSFT
MU
NFLX
ORCL
PFE
PG
SYF
Name
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Top 50
Name
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Bottom 50
Name
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Proposed/Future Work

- Use correlation data to build predictive models
  - Now that we have found what is correlated, we can start analyzing it and building predictive models based on the result
  - We would also like to cluster some correlated stocks with one another
- Grow our knowledge on the market
  - We hope to see if our analysis of data from 2013-2018 will grant an understanding of factors that affect the market the most, especially when it comes to crashes
- Rudimentary algorithmic trading
  - We are working on a trading algorithm that will be based on the data from this 2013-2018 dataset. We hope it will factor the most important discoveries we have made about the data and find some success with trading on the market.
  - We also hope to host this application on a simple website. Options including Heroku or a more complex website architecture
  - We hope to test this algorithm on later data so we are not only using the dataset we currently have

Evaluation

- Work can be evaluated by observing if our algorithm successfully predicts the actual change in the market.
- Metrics for evaluation will be the accuracy or error of our predictions, especially when compared to that of other prediction algorithms.
- We can compare our work with other prediction models and algorithms such as J.A.R.V.I.S and Hudson River Trading along with open source projects
- We can compare our work with the standard trading trends in the market
- We can compare our work with random guesswork

Milestones

- Study other trading algorithms and models (Week 7) COMPLETED
- Collect past and current stock data (Week 10) COMPLETED
- Find correlation between the top S&P 500 stocks COMPLETED
- Project Checkpoint Report (Week 12) (STARTED)
- Study and mine data for patterns, trends, outliers, etc. (Week 13) (STARTED)
- Create a predictive model/algorithm based off patterns and trends from mined data (Week 14)
- Create/use a site to host the application (Week 15)
- Final Project Report (Week 16)

We are on schedule!

Questions?