NASDAQ & INVESTMENT ANALYSIS

Hsiang-Yu Wang, Nikolai Saporoschetz, Yu-Hui Lin

Introduction

As students interested in finance we chose to focus on a NASDAQ dataset. A description of the dataset and review of our analysis and insights is outlined below.

Dataset Description

<u>Description</u>: The NASDAQ is a stock exchange headquartered in New York City. It is the second largest stock exchange in the world. The dataset we used includes pricing data from 2010-2021/19 for the 100 largest companies listed on the NASDAQ.

<u>Data Source</u>:

[NASDAQ-100 Stock Price Data (kaggle.com)]

<u>Features</u>:

- Date: Specific trading day
- Open: Price to start the trading day
- High; Highest price of the trading day
- Low: Lowest price of the trading day
- Close: Price at the end of the trading day
- Adj Close: Price at the end of the trading day accounting for dividends, stock splits, and new stock offerings.
- Volume: Number of stocks purchased in a trading day
- Name: Company stock ticker

	Open	High	Low	Close
count	271680.000000	271680.000000	271680.000000	271680.000000
mean	130.147060	131.678573	128.564517	130.173960
std	259.463324	262.249242	256.522832	259.455010
min	0.610000	0.660000	0.610000	0.650000
25%	32.549999	32.950001	32.150002	32.570000
50%	59.810001	60.504999	59.119999	59.849998
75%	117.139999	118.470001	115.820000	117.190002
max	3744.000000	3773.080078	3696.790039	3731.409912

	Adj Close	Volume
count	271680.000000	2.716800e+05
mean	126.929715	1.052670e+07
std	260.156874	3.924802e+07
min	0.612270	0.000000e+00
25%	28.001979	1.332175e+06
50%	55.599998	2.759400e+06
75%	114.705492	6.889500e+06
max	3731.409912	1.880998e+09

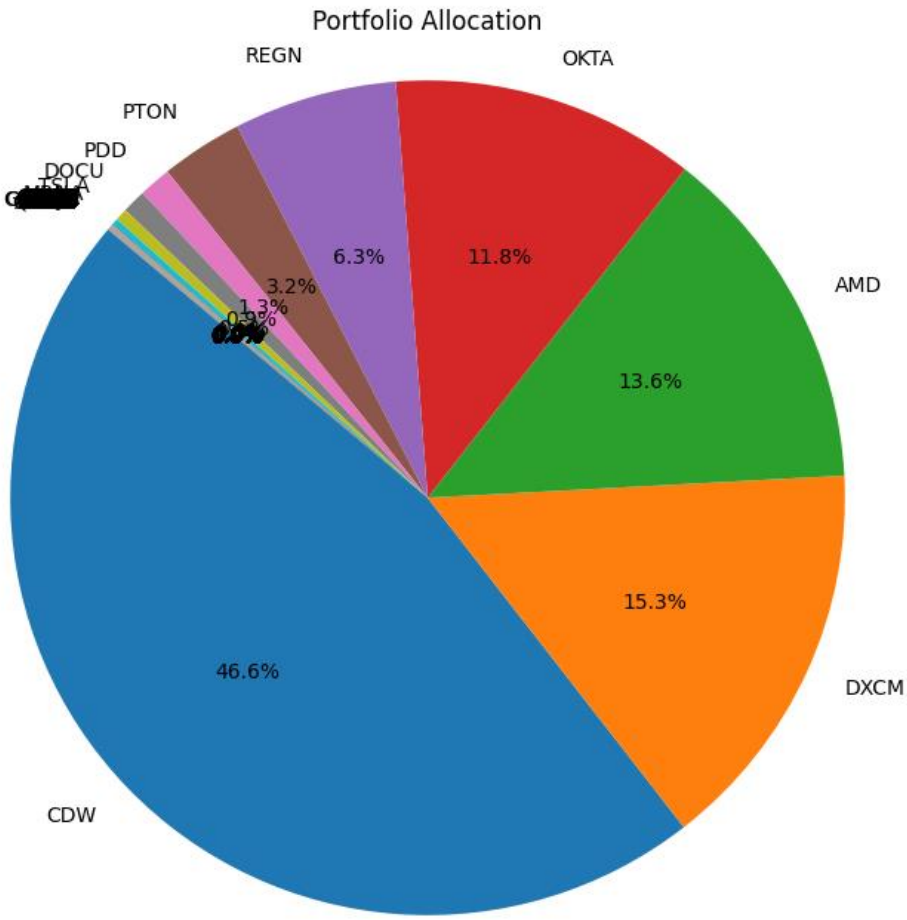
Analysis & Insights

(1) <u>Convex Programming for Portfolio Optimization</u>:

For investment professionals, stock selection is a frequent topic of discussion as these individuals seek to minimize risk while maximizing returns. Convex programming can be used as a data-driven approach to parsing data to derive insights on the best way to construct a portfolio.

In this analysis we developed a convex program to maximize the return-to-variance ratio for a portfolio. To do this, we created a covariance matrix of asset returns as well as the expected yearly returns for each stock. Our objective function was to maximize the return-to-variance ratio, as stated previously. We then adjusted the weights of each stock in the portfolio to minimize negative return-to-variance ratio while abiding by key constraints such as fully investing in the portfolio.

The results of our analysis can be seen below:



The portfolio weights for top 10 stocks:

1. CDW: 46.9%
2. DXCM: 15.4%
3. AMD: 13.6%
4. OKTA: 11.9%
5. REGN: 6.4%
6. PTON: 3.2%
7. PDD: 1.3%
8. DOCU: 0.9%
9. TSLA: 0.5%
10. MRNA: 0.2%

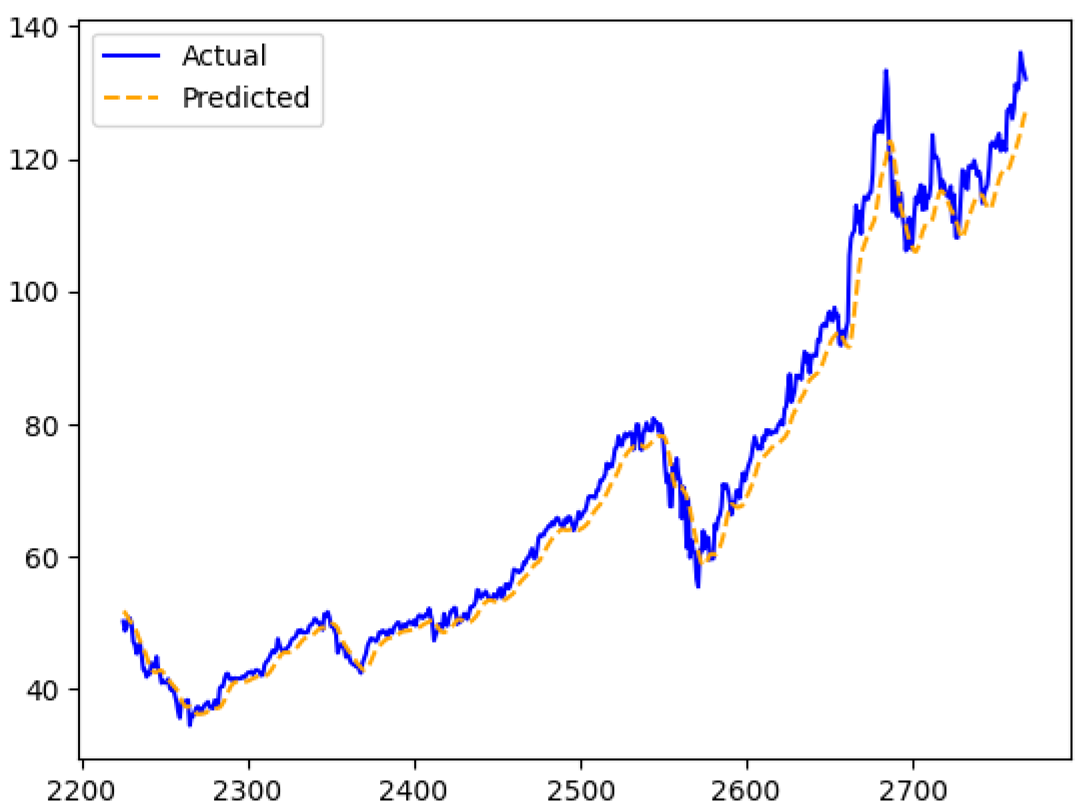
This portfolio allocation represents the optimal stock weightings to maximize the return-to-variance ratio. The optimal portfolio return is 52.62%, and the portfolio variance is 0.0484%. The key insight of this analysis is not only is this an optimal allocation, but convex programming can be used to help construct portfolios using data driven insights independent of investor bias.

(2) Supervised Machine Learning for Stock Price Prediction:

We built this stock price prediction model to enable individuals and investors to forecast future stock prices and form their investment strategies accordingly. By accurately predicting stock movements, people can make informed decisions about buying, selling, or holding stocks to maximize their returns.

As an example, we used an LSTM (Long Short-Term Memory) neural network model to predict the daily price of Apple (AAPL) stock. Based on our model setting, the model learned from the past 10 days of data to predict the next day's price recursively.

Finally, we plotted both the actual and predicted stock prices for the validation set to visualize the model's performance. The results are shown in the line chart below, with the actual prices in blue and the predicted prices in an orange dashed line. Our analysis proves that people can utilize LSTM as a time-series forecasting method to predict future stock prices.



(3) Dynamic programming for Investment Timing:

After obtaining the predicted stock prices from the previous LSTM model, we aim to develop an investment strategy that can maximize profit based on these predictions. To achieve this, we employ dynamic programming to determine the optimal buy and sell points, given a specific number of allowable transactions for a specific period.

First, we extract the necessary data, including the latest 20-day predicted prices and actual prices, from the previous predictions and validation data, respectively. This 20-day period represents approximately a month of trading days, which we use as an example for our investment strategy.

We defined a function called maxProfit that implements a dynamic programming approach to find the maximum profit possible by buying and selling a stock at most k times, given the stock prices for n days. It returns maximum profit and the optimal buy and sell points. We applied this function to 20-day predicted stock prices with a limitation of at most 3 transactions.

After obtaining the optimal buy and sell points based on the predicted prices, we validated the investment strategy using the actual prices from the validation data for the same month period. We calculated the actual profit by simulating the buy and sell transactions at the optimal points determined by the predicted prices. Finally, we checked whether the investment strategy based on the predicted prices would have been profitable (actual profit > 0) or not.

This approach combines the power of machine learning for stock price prediction with dynamic programming to devise an investment strategy. By validating the strategy against actual prices for the same period, we assess the effectiveness of the proposed approach.

An example output for the strategy employed on Apple's 20-day stock prices is shown below.

- Maximum predicted profit: 13.335449
- Buy points: [0, 8]
- Sell points: [7, 19]
- Actual profit: 3.632934600000013
- Does the strategy work: Yes

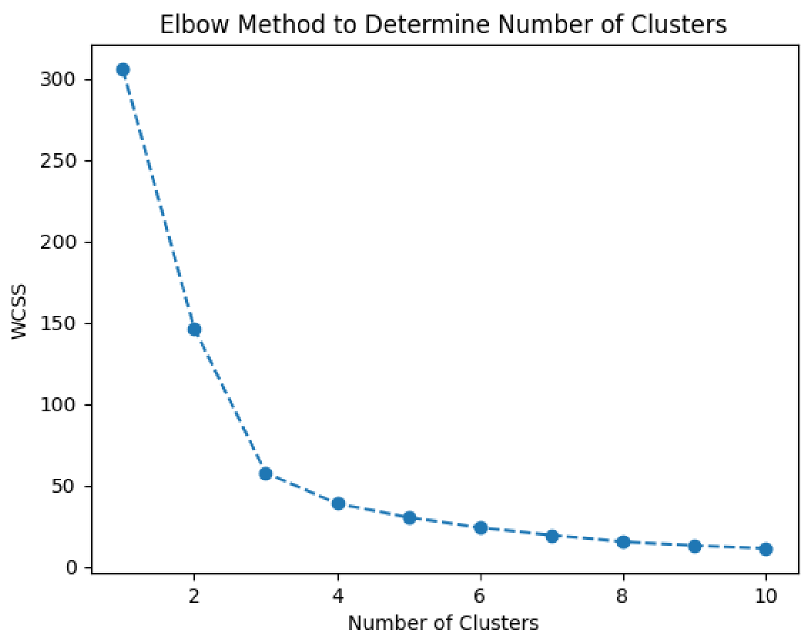
The results show that the best strategy is to make two transactions: buying on day 0 and selling on day 7 and then buying on day 8 and selling on day 19. This strategy can result in a 3.63 profit although the profit is not as high as the algorithm predicts.

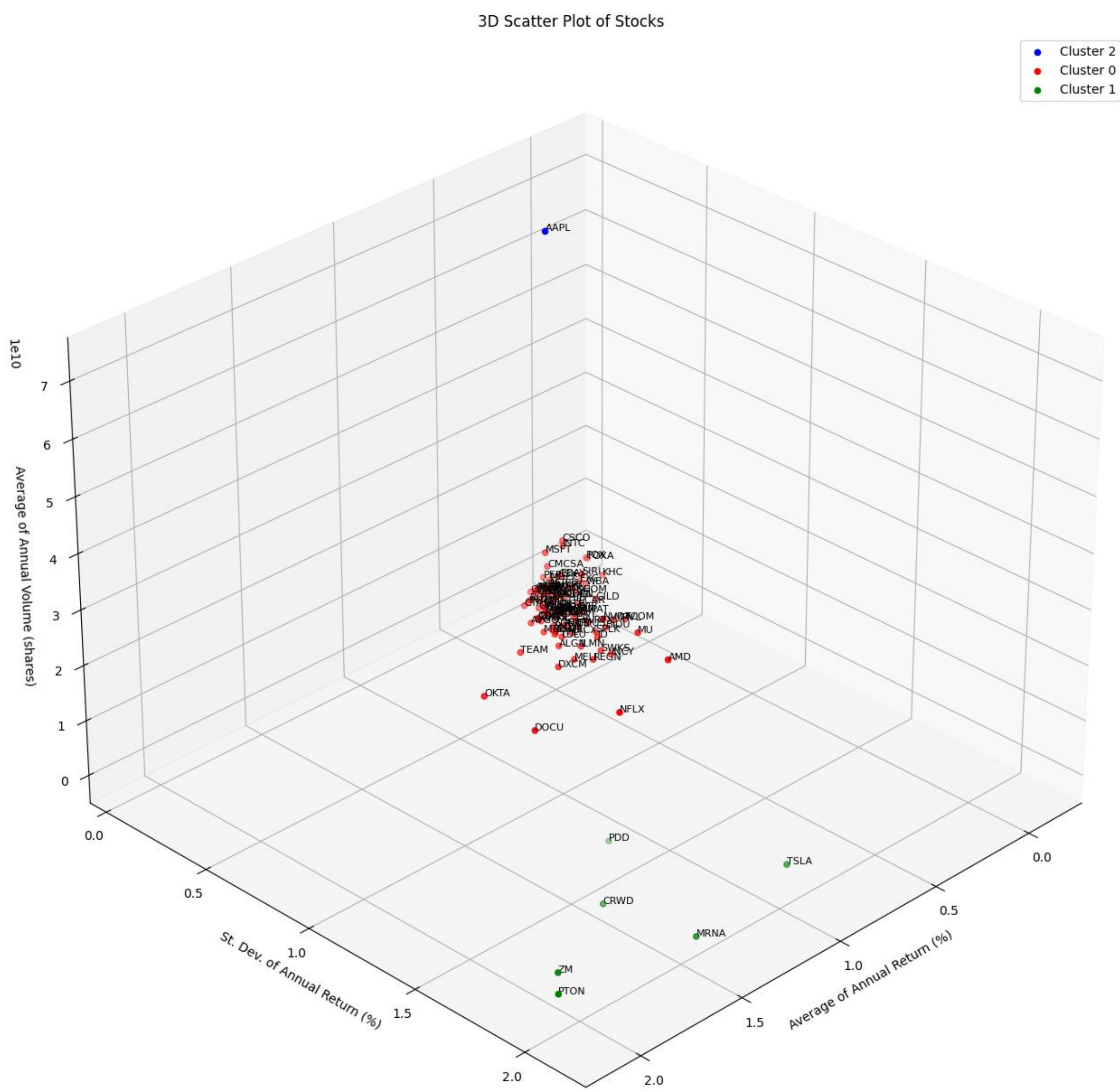
(4) Unsupervised Machine Learning for Grouping Stocks

We perform this analysis to identify groups of stocks that share similar patterns in terms of their average yearly return, standard deviation of yearly returns, and average yearly trading volume. Grouping stocks based on these characteristics can provide valuable insights for portfolio construction and risk management strategies.

To accomplish this task, we employ the K-means clustering algorithm. Additionally, we leverage the elbow method to determine the optimal number of clusters by analyzing the within-cluster sum of squares for different cluster counts. After applying the K-means clustering algorithm and the elbow method, our analysis reveals three distinct clusters of stocks.

1. PDD, CRWD, ZM, PTON, MRNA, and TSLA are in a cluster that features high return, high risk, and low volume.
2. AAPL is in a cluster alone. The cluster features extremely high trading volume and relatively lower return and risk.
3. The other stocks are in the other cluster that has relatively lower return, risk, and volume.





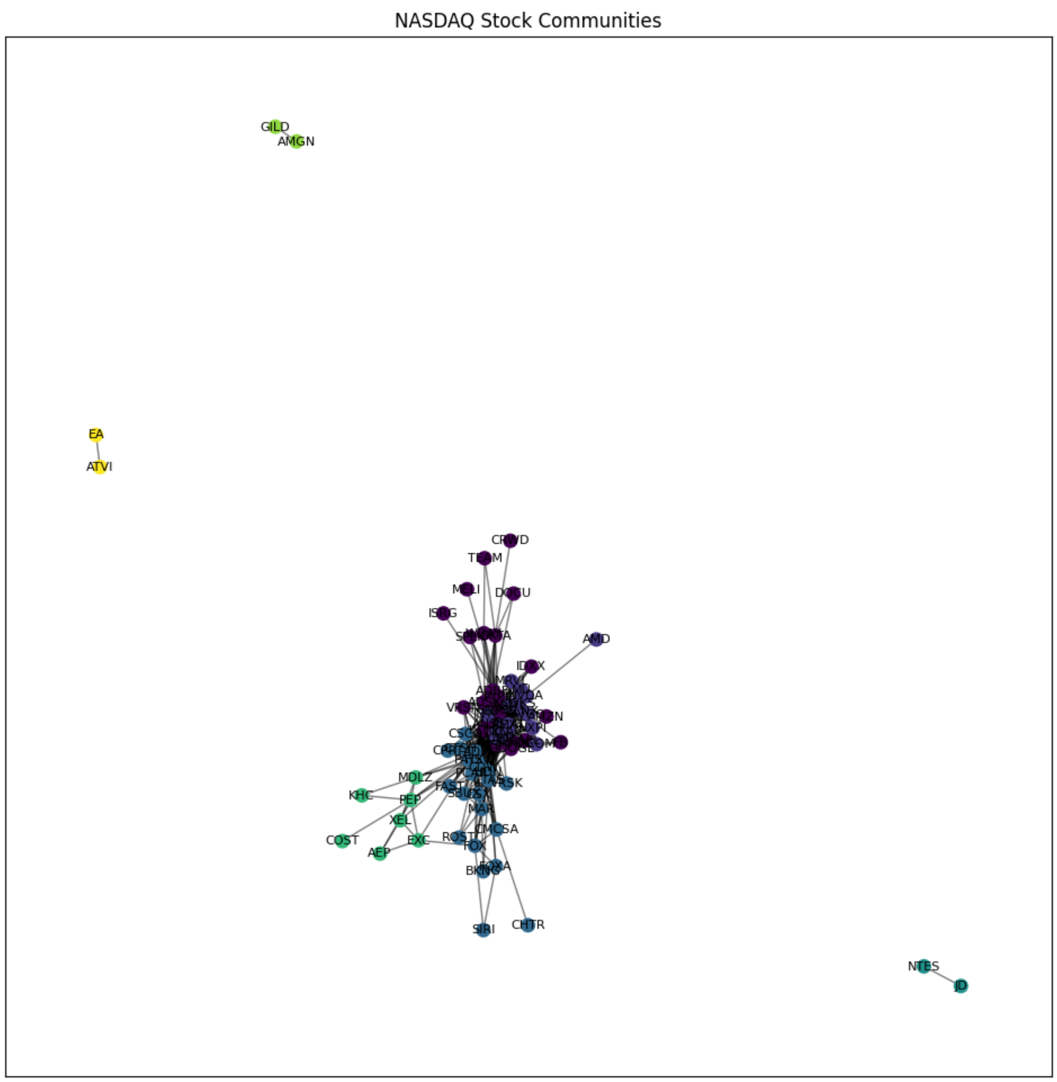
(5) Community Detection with Graph Utilization:

Community detection involves identifying groups of nodes with key connections in a graph. The graph provides the framework for representing networks of nodes and edges. With this method, we can uncover key relationships. This approach is well suited for our dataset containing stock prices due to its ability to unveil hidden market structures and to group stocks with similar behaviors into communities. These communities can aid in portfolio allocation and market research activities.

To obtain our insights we calculated the returns for each stock and constructed a correlation matrix. We then set a significant threshold for correlation at 0.5 to include significant relationships. Next, we created a graph representation of the stocks where nodes represent stocks and edges represent significant correlations between them. We then utilized the Louvain method which partitions a network into communities based on the density of their connections.

The results of this analysis can be seen below:

Community ID	Companies
0	JD, NTES
1	ADI, AMAT, KLAC, SWKS, TXN, MCHP, NVDA, ASML, AVGO, INTC, LRCX, MRVL, MU, NXPI, QCOM, XLNX, AMD
2	AAPL, ANSS, GOOG, GOOGL, INTU, MSFT, PYPL, SNPS, ADBE, ADSK, AMZN, CDNS, DOCU, OKTA, SPLK, VRSN, WDAY, IDXX, CRWD, FB, ISRG, MELI, TEAM
3	ADP, FISV, PAYX, CDW, CSCO, CSX, CTAS, CTSH, HON, MAR, PCAR, CMCSA, CPRT, FAST, SBUX, VRSK, BKNG, CHTR, FOX, FOXA, ROST, SIRI
4	MDLZ, PEP, AEP, EXC, XEL, COST, KHC
5	AMGN, GILD
6	ATVI, EA



- **Community 0**: JD - JD.com Inc., NTES - NetEase Inc.
- **Community 1**: ADI - Analog Devices Inc., AMAT - Applied Materials Inc., KLAC - KLA Corporation, SWKS - Skyworks Solutions Inc., TXN - Texas Instruments Incorporated, MCHP - Microchip Technology Incorporated, NVDA - NVIDIA Corporation, ASML - ASML Holding NV, AVGO - Broadcom Inc., INTC - Intel Corporation, LRCX - Lam Research Corporation, MRVL - Marvell Technology Group Ltd., MU - Micron Technology Inc., NXPI - NXP Semiconductors NV, QCOM - Qualcomm Inc., XLNX - Xilinx Inc., AMD - Advanced Micro Devices Inc.
- **Community 2**: AAPL - Apple Inc., ANSS - ANSYS Inc., GOOG - Alphabet Inc. (Class C), GOOGL - Alphabet Inc. (Class A), INTU - Intuit Inc., MSFT - Microsoft Corporation, PYPL - PayPal Holdings Inc., SNPS - Synopsys Inc., ADBE - Adobe Inc., ADSK - Autodesk Inc., AMZN - Amazon.com Inc., CDNS - Cadence Design Systems Inc., DOCU - DocuSign Inc., OKTA - Okta Inc., SPLK - Splunk Inc., VRSN - VeriSign Inc., WDAY - Workday Inc., IDXX - IDEXX Laboratories Inc., CRWD - CrowdStrike Holdings Inc., FB - Meta Platforms Inc., ISRG - Intuitive Surgical Inc., MELI - MercadoLibre Inc., TEAM - Atlassian Corporation Plc
- **Community 3**: ADP - Automatic Data Processing Inc., FISV - Fiserv Inc., PAYX - Paychex Inc., CDW - CDW Corp., CSCO - Cisco Systems Inc., CSX - CSX Corp., CTAS - Cintas Corporation, CTSH - Cognizant Technology Solutions Corporation, HON - Honeywell International Inc., MAR - Marriott International Inc., PCAR - PACCAR Inc., CMCSA - Comcast Corporation Class A, CPRT - Copart Inc., FAST - Fastenal Company, SBUX - Starbucks Corporation, VRSK - Verisk Analytics Inc., BKNG - Booking Holdings Inc., CHTR - Charter Communications Inc., FOX - Fox Corporation Class A, FOXA - Fox Corporation Class B, ROST - Ross Stores Inc., SIRI - Sirius XM Holdings Inc.
- **Community 4**: MDLZ - Mondelez International Inc., PEP - PepsiCo Inc., AEP - American Electric Power Company Inc., EXC - Exelon Corporation, XEL - Xcel Energy Inc., COST - Costco Wholesale Corporation, KHC - The Kraft Heinz Company
- **Community 5**: AMGN - Amgen Inc., GILD - Gilead Sciences Inc.
- **Community 6**: ATVI - Activision Blizzard Inc., EA - Electronic Arts Inc.

Classifying technology companies into different segments for investment research and monitoring can be challenging. Despite being challenging, finding distinct way to segment companies for portfolio monitoring is critical. Technology companies often have diverse product offerings that span multiple sectors and industries. And these companies also operate in dynamic and rapidly evolving markets characterized by continuous innovation. Our community detection approach segmented these companies effectively as seen by Community 2 versus Community 3. Community 2 primarily consists of companies that focus on software, e-commerce, and digital services. While Community 3 comprises companies primarily specializing in business solutions and consumer services that utilize software but are not necessarily frontier technology companies. Besides, we can also find that Community 1 contains mostly stocks related to semiconductors, and Community 4 encompasses companies from consumer goods, utilities, and retail sectors.

(6) Degree Centrality with Graph Utilization:

Building from our community detection analysis we utilized degree centrality. This metric quantifies the number of direct connections a node in our graph has. Higher values indicate a higher correlation with other stocks. This means that understanding the movements of these stocks is important to understanding movements in the prices of other stocks.

We conducted a centrality analysis on our previously developed graph to identify the top stocks based on their level of influence and connectivity within the network. The goal was to provide investment professionals with insight into important stocks to monitor due to their overall influence and connection with the market.

The results of this analysis can be seen below:

	Stock	Degree Centrality
0	TXN	0.5
1	PAYX	0.5
2	FISV	0.486486
3	ADI	0.472973
4	ADP	0.472973
5	HON	0.472973
6	PYPL	0.459459
7	SNPS	0.459459
8	MCHP	0.459459
9	INTU	0.418919

1. TXN (Texas Instruments Incorporated) - Degree Centrality: 0.5
2. PAYX (Paychex Inc.) - Degree Centrality: 0.5
3. FISV (Fiserv Inc.) - Degree Centrality: 0.486486
4. ADI (Analog Devices Inc.) - Degree Centrality: 0.472973
5. ADP (Automatic Data Processing Inc.) - Degree Centrality: 0.472973
6. HON (Honeywell International Inc.) - Degree Centrality: 0.472973
7. PYPL (PayPal Holdings Inc.) - Degree Centrality: 0.459459
8. SNPS (Synopsys Inc.) - Degree Centrality: 0.459459
9. MCHP (Microchip Technology Incorporated) - Degree Centrality: 0.459459
10. INTU (Intuit Inc.) - Degree Centrality: 0.418919

After our community analysis, we wanted to test for overall influence. Based on these results we have gathered insights into the companies that investment professionals should key in on while monitoring the NASDAQ. An investment professional who is invested in a NASDAQ oriented investment vehicle could monitor news surrounding these companies to develop their own sentiment around the future performance of NASDAQ. Companies like Texas Instruments and Paychex have a large degree of influence over performance, and it would be important to monitor them.

Conclusion

Our comprehensive analysis of the NASDAQ-100 dataset using various techniques has provided valuable insights for investment professionals and individuals interested in the stock market. Through convex programming, we optimized portfolio construction to maximize return-to-risk metric. Supervised machine learning enabled us to predict future stock prices, which we then leveraged with dynamic programming to devise profitable investment strategies. Unsupervised learning grouped stocks based on their characteristics, revealing distinct clusters for portfolio diversification. Community detection uncovered hidden market structures and identified key segments. Finally, degree centrality highlighted the most influential stocks to monitor for their impact on the broader market.

These data-driven approaches demonstrate the power of combining financial domain knowledge with advanced analytical techniques. The insights derived from our analysis can inform decision-making processes, risk management strategies, and investment research activities. As the financial landscape continues to evolve, leveraging such quantitative methods will become increasingly crucial for navigating market dynamics and making informed investment decisions.