Stocks vs. Cryptocurrencies

Connor Buxton, Ben Hines, Sargis Abrahamyan, Lucas Stefanic

Introduction

The stock market has long been a popular way to invest money. As such, it is of great interest and importance to investors, financial planners, and financial advisors. However, in recent years, another investment opportunity has seen a rise in popularity: cryptocurrencies. There are dozens of different cryptocurrencies active today, the most well-known of which is Bitcoin. With more and more people buying cryptocurrencies, investors should weigh them against the stock market and seriously consider the possible benefits.

We will compare the performance of four stocks with the largest market caps in each of the industries of manufacturing, finance, information, and retail to the performance of four of the most traded cryptocurrencies. The industries correspond to the US Census NAICS code for categorization, and as such, 'Finance' includes credit card and banking companies, 'Manufacturing' includes food, technology, and chemicals production, 'Information' includes telecommunications, and 'Retail' includes both online and physical storefronts. The stocks examined correspond to the companies Visa, JPMorgan Chase, Bank of America, Mastercard, Apple, Microsoft, MGP, Quaker Food, Comcast, Version, AT&T, T-Mobile US, Amazon, Walmart, Home Depot, and Costco. The cryptocurrencies examined are Bitcoin, Etherium, Dogecoin, and Litecoin.

We will look for any meaningful relationships between the performance of these stocks and the industry economic data, as well as determining the efficacy of predicting future performance of the portfolio using machine learning. Finally, we will make recommendations for a theoretical client about how they should invest their money between these stocks and cryptocurrencies based on our findings.

Individual Historical Stock Performance

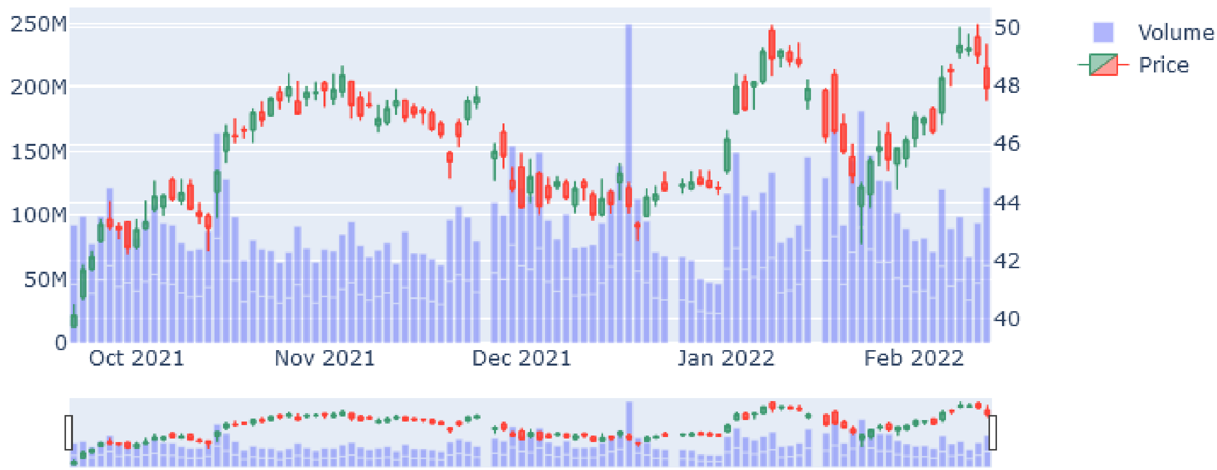
Looking at the history of these stocks gives us a sense of their performance across longer time frames and allows us to glean any patterns that might exist within their place in the markets.

Finance:

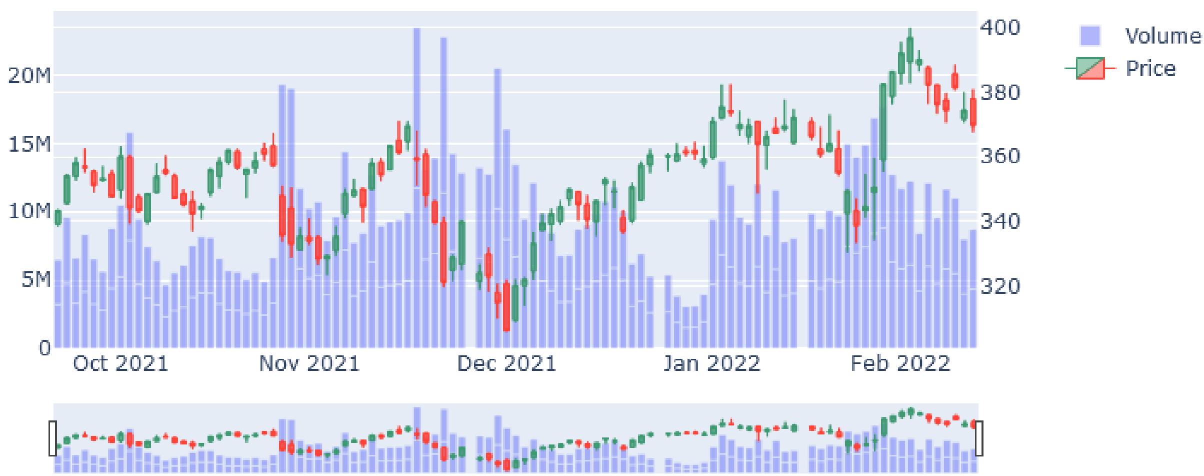
Data sourced from FinnHub API [1]

Here we see JPMorgan Chase displaying periodic gains following periodic losses, with the downturn at the end of January beginning to recover. Visa displays similar value recovery, where steep losses at the start of November was eventually recouped in early February 2022.

Bank of America (BAC)



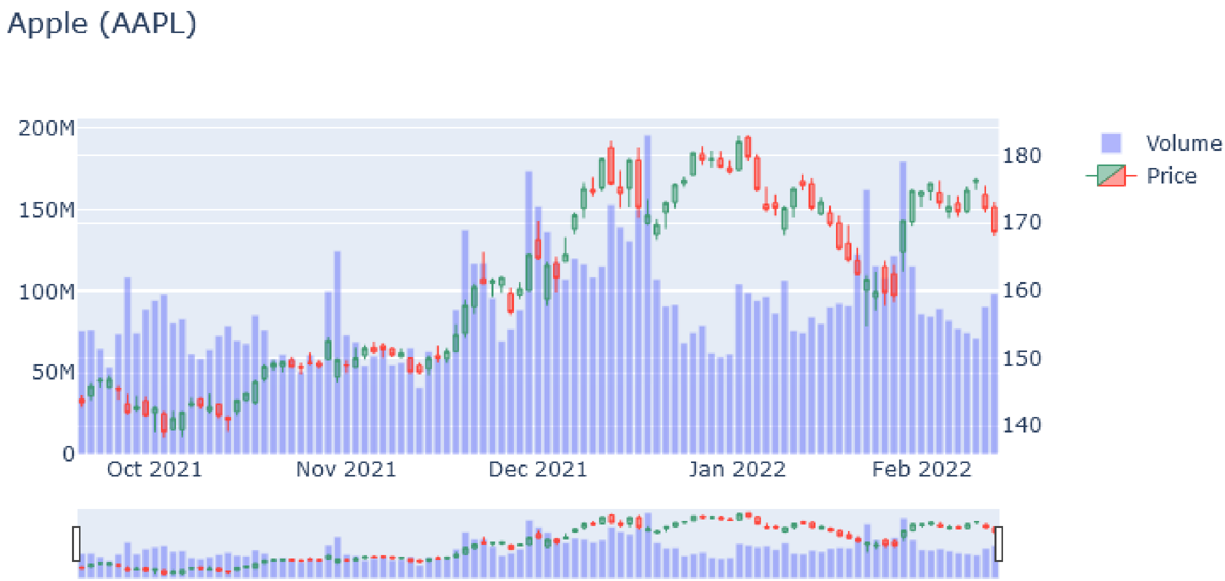
Mastercard (MA)



Data sourced from FinnHub API [1]

Both Mastercard and Bank of America shows a steady upward trend, with consistent recovery after periods of loss.

Manufacturing



Data sourced from FinnHub API [1]

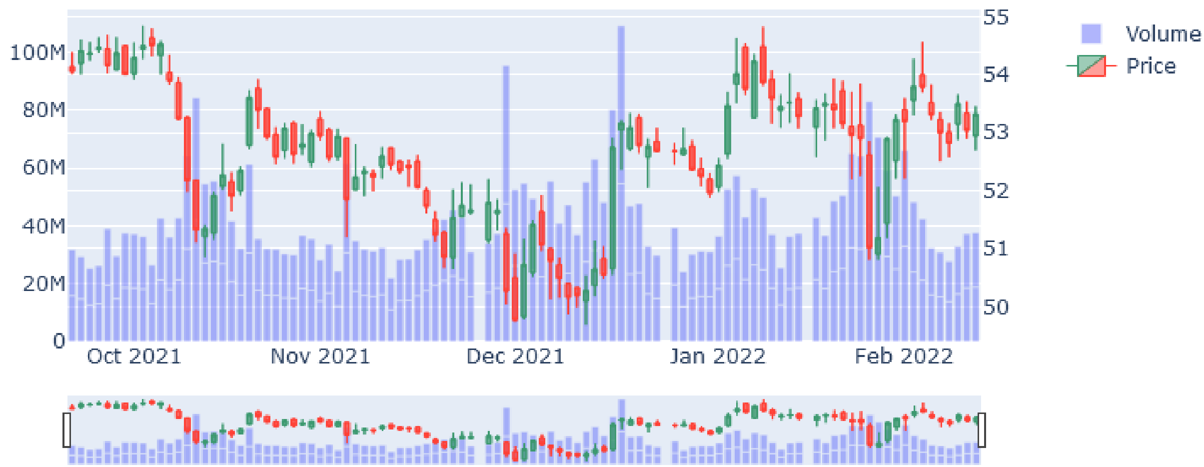
Apple shows strong appreciation in value from October, hitting a ceiling of $180 per share. Microsoft gained from October through late November but stagnated and began to decline in January of 2022.

Data sourced from FinnHub API [1]

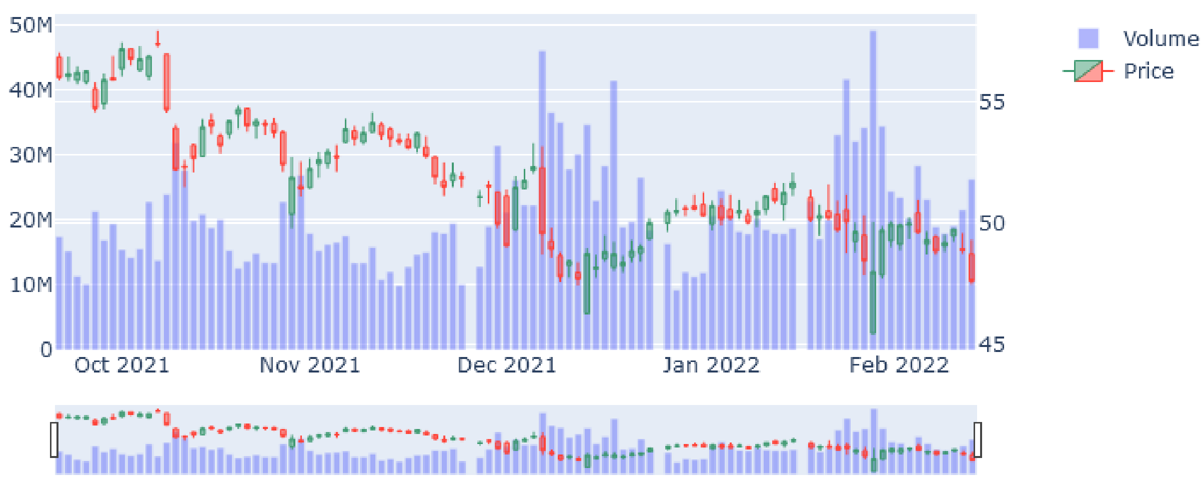
MGP rose strongly from November to December, but stagnated and fell in early February, however it seems poised for a recovery after hitting approximately $75 a share. Quaker However shows greater volatility, reaching a ceiling of approximately $178 a share before suffering a consistent decline since late November with no sign of recovery.

Information

Verizon (VZ)



Comcast (CMCSA)



Data sourced from FinnHub API [1]

Verizon keeps testing a ceiling of $53 a share from October and into early 2022, and having tested a floor of $50 and rebounding, the stock seems to have settled into an equilibrium. Comcast however, is showing steady downward pressure, depreciating share values from above $55 in October to under $50 in early 2022.

AT&T (T)

T-Mobile (TMUS)

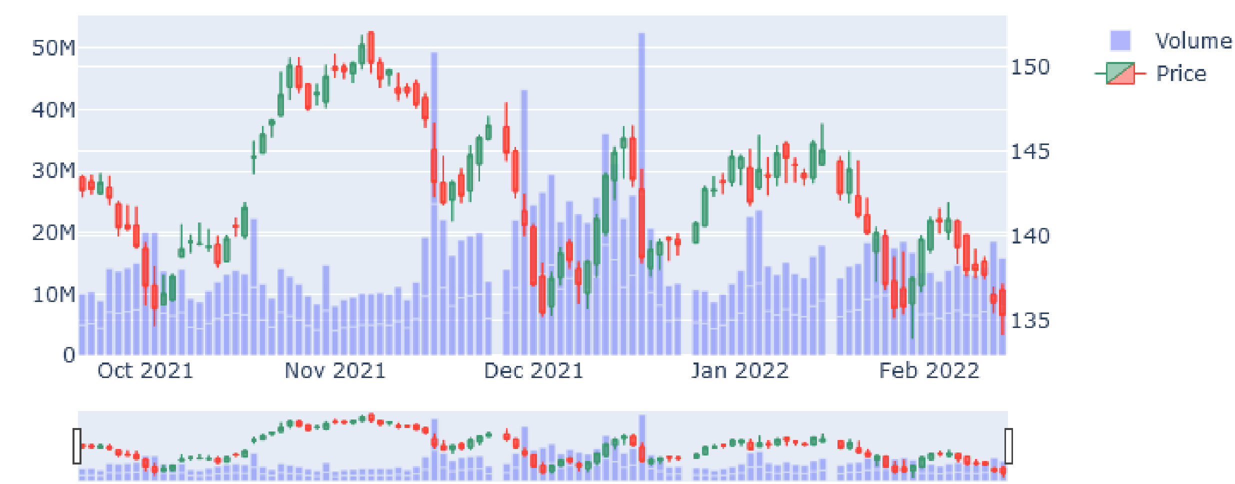
Data sourced from FinnHub API [1]

AT&T showed a steady decline from October to December, but after testing a floor of just over $22 a share, it appreciated quickly to its previous ceiling of $27 from October, establishing a reliable upper bound for the stock. A similar story for T-Mobile, as it displayed a steady loss of value from October to Early February where it rapidly rebounded after testing a floor of $100 to then test a ceiling of $130, and is now showing the beginnings of a period of value loss.

Retail

Amazon (AMZN)

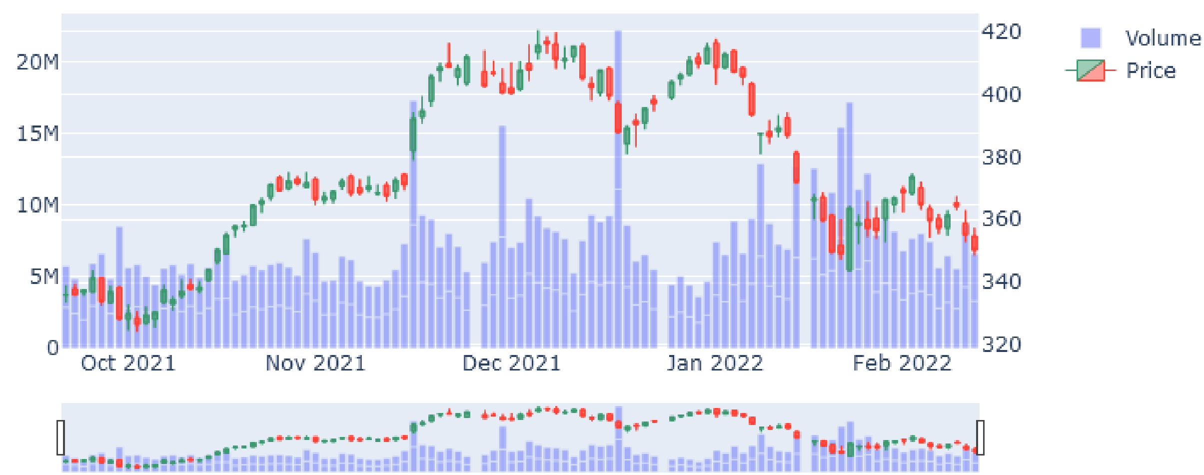
Walmart (WMT)



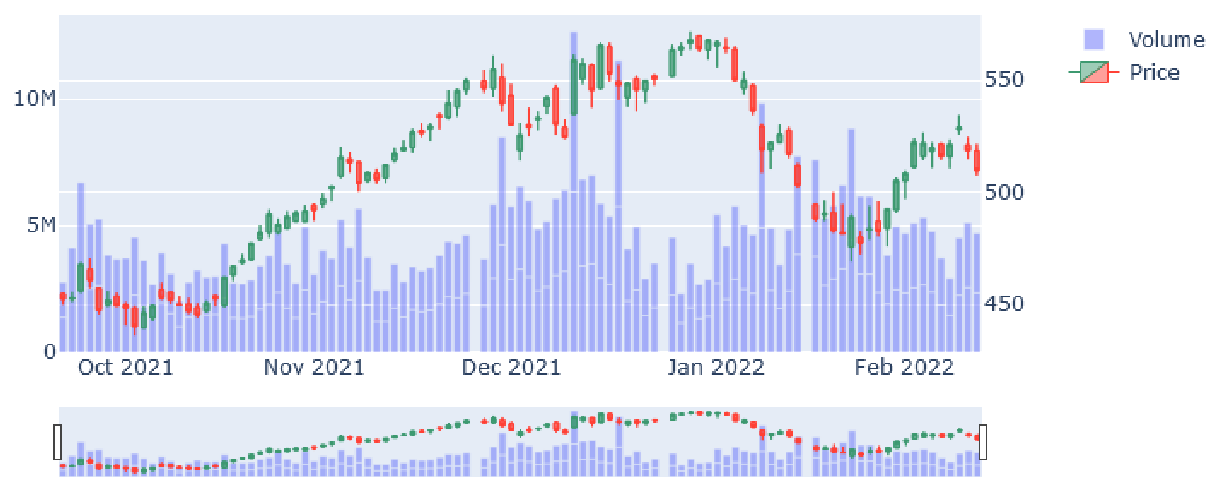
Data sourced from FinnHub API [1]

Amazon remained steady between $3400 and $3600 for a time before dropping precipitously in early 2022 and now shows signs of recovery. Walmart however shows much greater volatility, showing prices swinging from $135 to over $150, then back and forth between $148 and $136 before settling into a net downward trend.

Home Depot (HD)



Costco (COST)

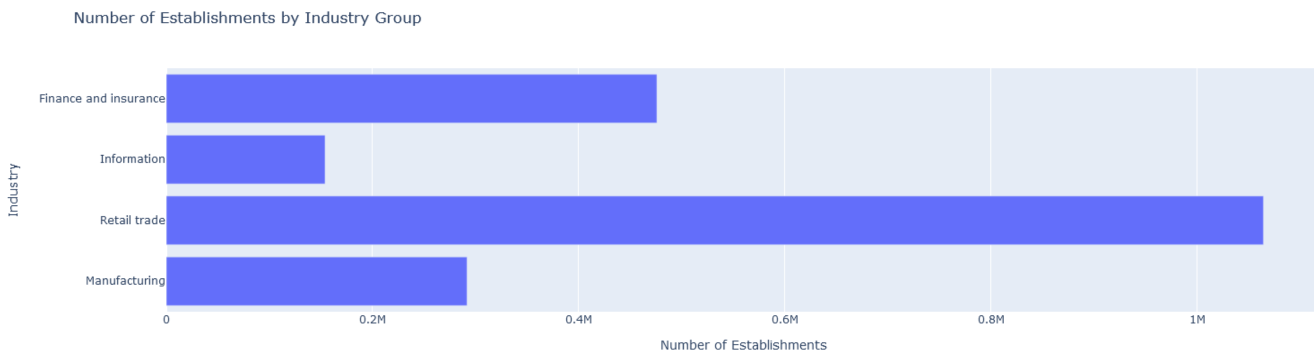


Data sourced from FinnHub API [1]

Home Depot saw significant appreciation from an October low hovering around $320 up to a ceiling of $420, and after testing it between late November and into January 2022, it began to steadily decline to a seasonal low of under $360. Costco saw similar movements, where between late October and into January 2022 a ceiling of over $550 was tested and eventually rejected, where the price then fell sharply, but then rebound to approximately $520.

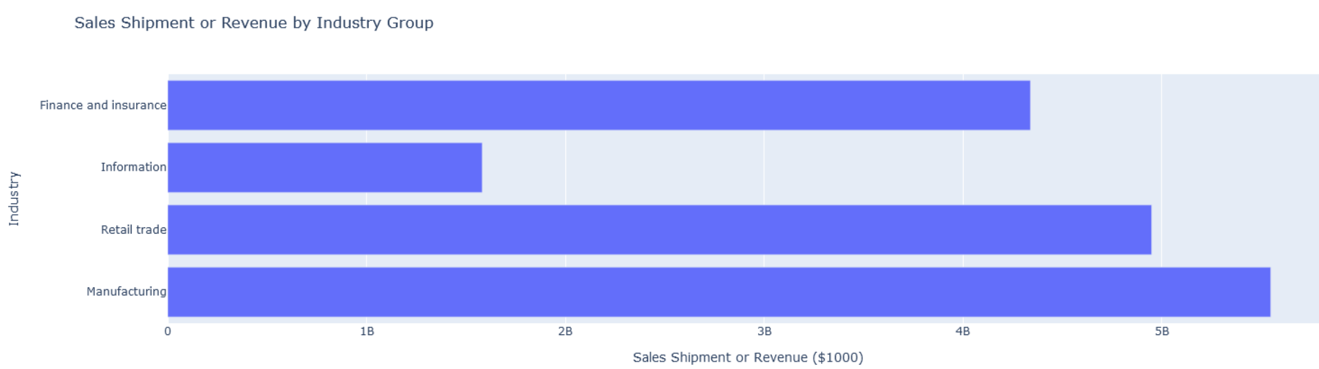
Industry Census Data

We can examine industry economic census data to gain some insight as to the factors that might affect economic health of the examined industries.



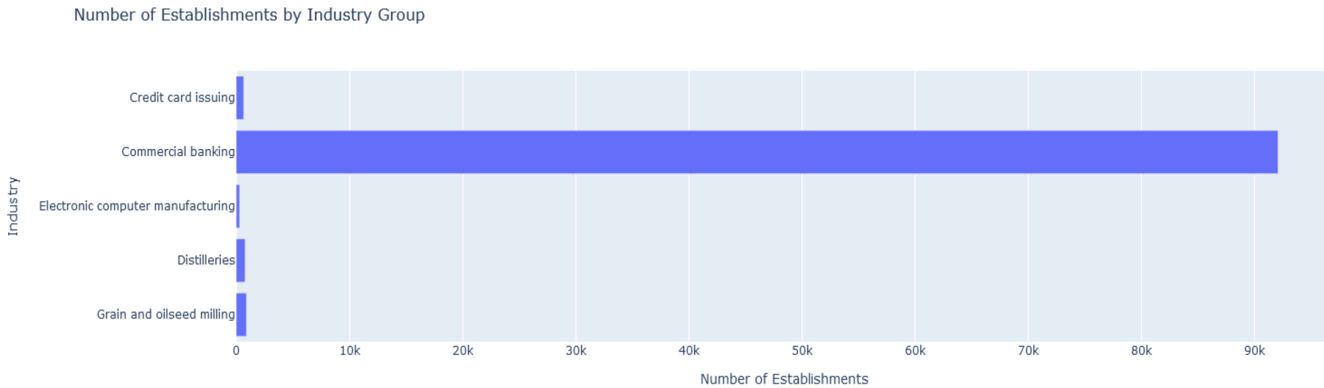
Data sourced from 2017 US Census Data. [3]

Here we see 'Retail Trade' has significantly more physical established properties across the US, with finance a distant second.



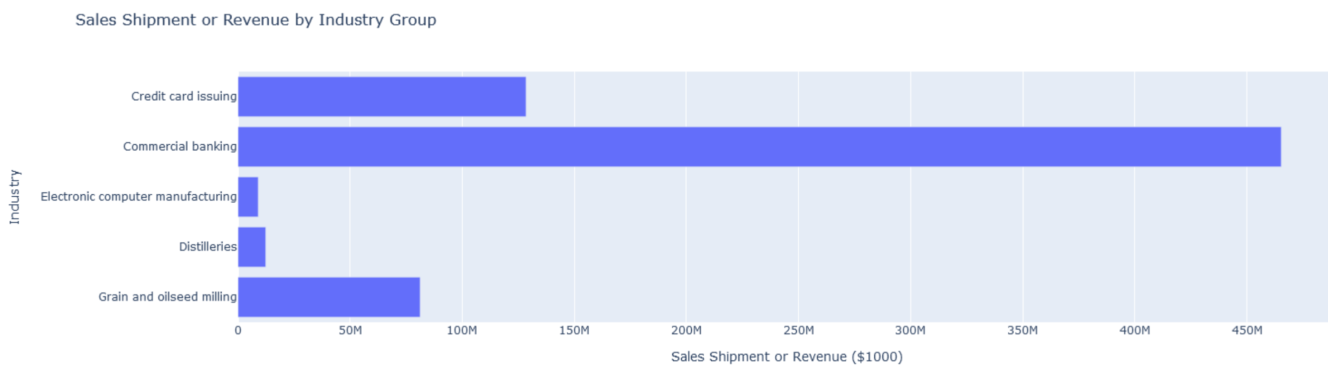
Data sourced from 2017 US Census Data. [3]

Looking at the figures for revenue, the 'Information' industry lags significantly behind the other three sectors, which for the industry as a whole is a concern.



Data sourced from 2017 US Census Data. [3]

Here we see just how the physical establishments are broken down in industry sub-categories and a clearer picture starts to emerge: Commercial Banking makes up the vast majority of established locations across the US.



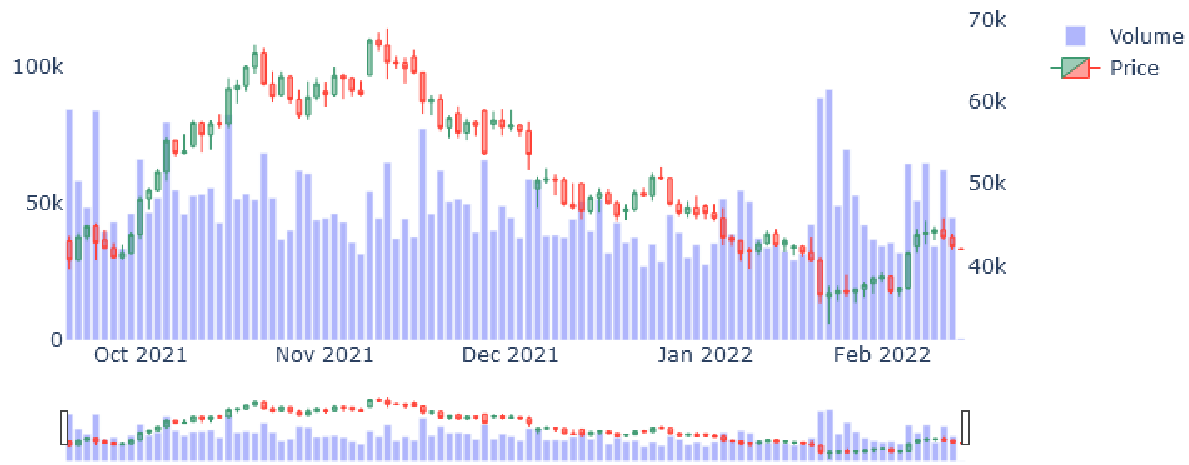
Data sourced from 2017 US Census Data. [3]

Breaking down revenue in a similar manner, we see a striking difference in Commercial Banking when compared to other sub-categories of industry. There exists a direct correlation between establishments and revenue within Commercial Banking. Interesting still, you have Grain and Oilseed Milling with relatively few establishments but disproportionately high revenue.

Historical cryptocurrency data

Looking at cryptocurrency performance across time allows an effective comparison to the stock portfolio.

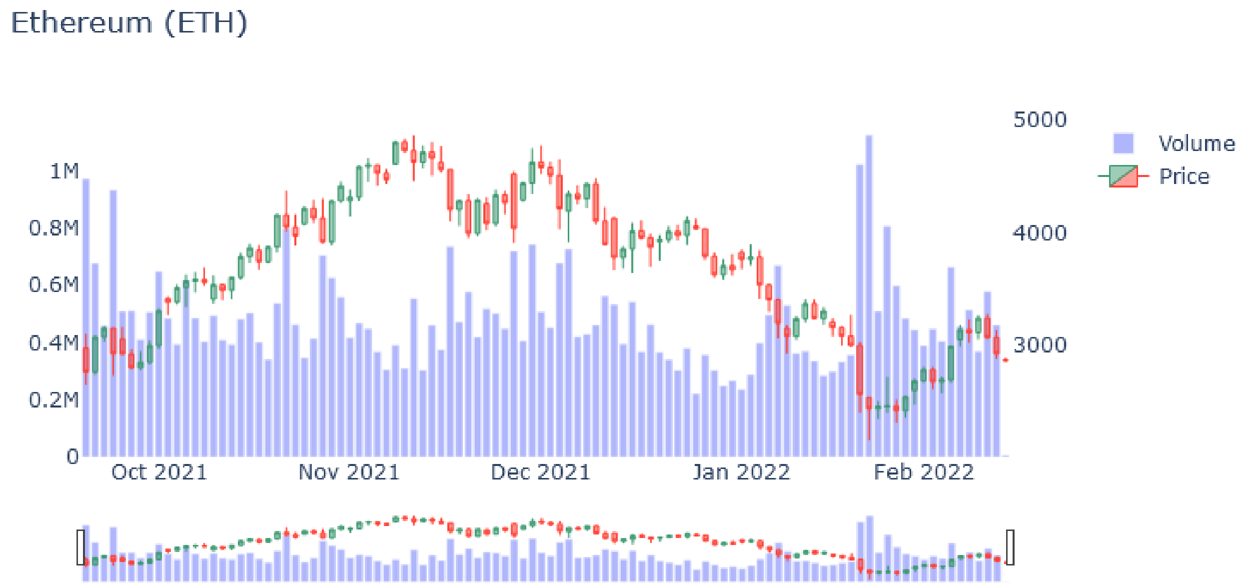
Bitcoin (BTC)

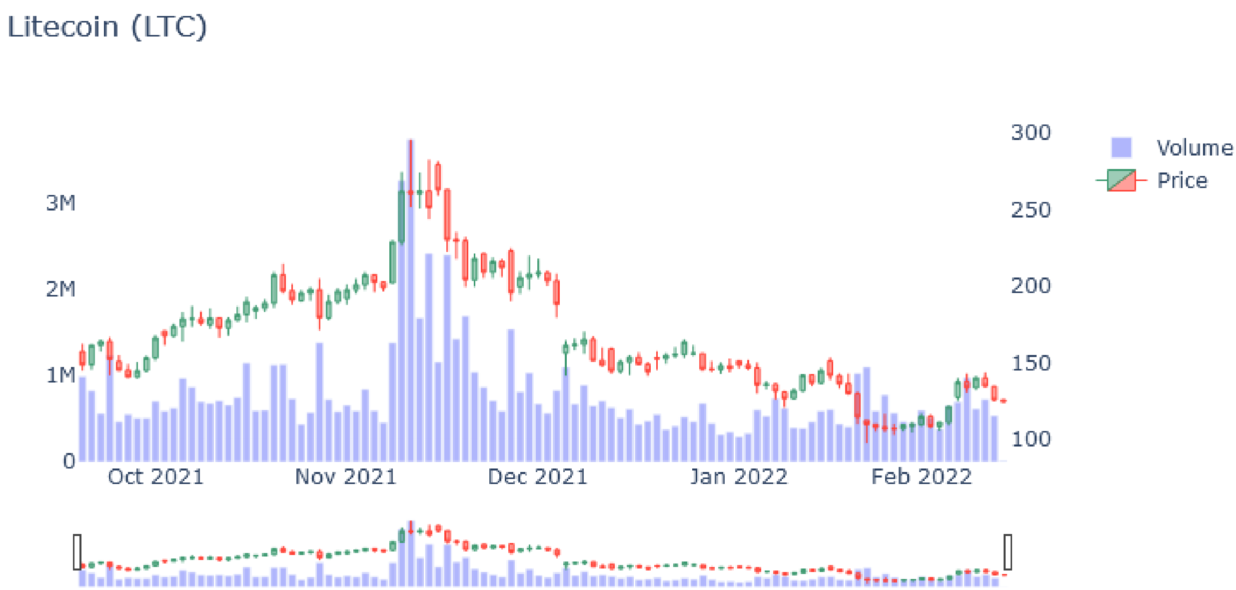


Dogecoin (DOGE)

Data sourced from AlphaVantage API [2]

Bitcoin shows deep and enduring losses in value after a high of just under $70,000 towards the end of November, and a minor recovery in February 2022 is looking to be offset by another recent devaluation. Dogecoin shows similarly poor trends, with its meager valuation in November hitting approximately $0.30 before dropping significantly to half that in February 2022.





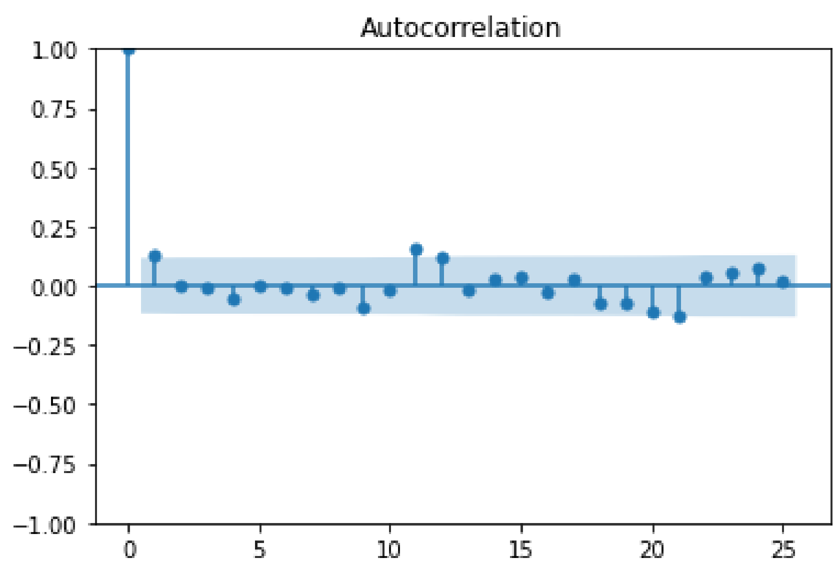
Data sourced from AlphaVantage API [2]

Ethereum consistently appreciated in value from October and into late November before peaking under $5000 and then steadily declined to a low below $2500 before recovering slightly to $3000. Litecoin shot up between October and November to a high of $300 before posting steep loses that pushed the value down to nearly $100.

Machine learning findings

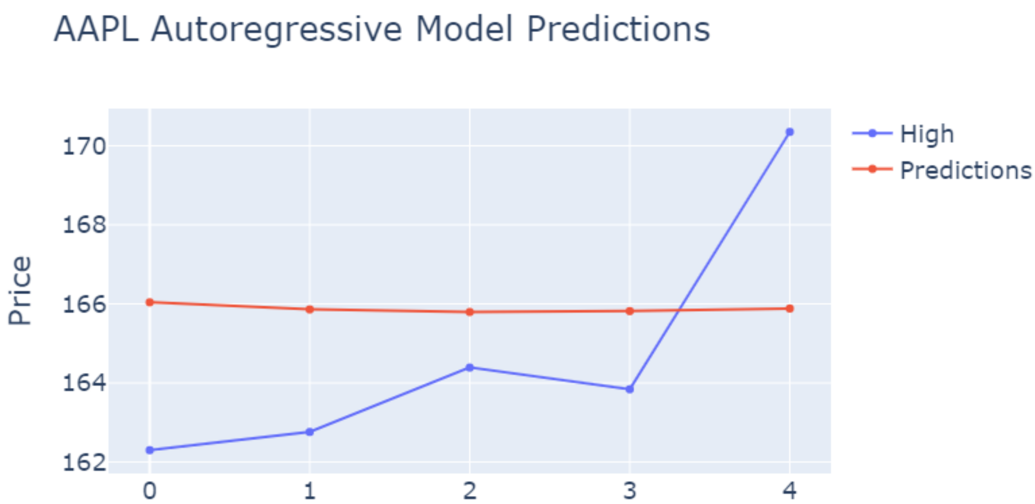
The prediction of stocks is arguably the most important aspect of the stock market, from the overall health of the market to the finer details of individual stock trends. It is no wonder that as technology has progressed and automation has become available for a variety of applications that the stock market would be of interest in utilizing artificial intelligence methods to predict the movement of stock prices at varying levels of granularity, from day-to-day movements to even finer levels of detail such as minute to minute. The rise of interest in cryptocurrency also raises the question of whether machine learning and other AI methods could be applied to the data in a similar context. Our interest is in the use of this tool to assist in decision making for consumer portfolios, in much the same way as an experienced professional might be able to. With sufficient accuracy, an algorithm which could predict stock movement effectively would not replace these professionals, but rather become another tool within their work which might provide them with greater efficiency and allow for more portfolio management by a single individual.

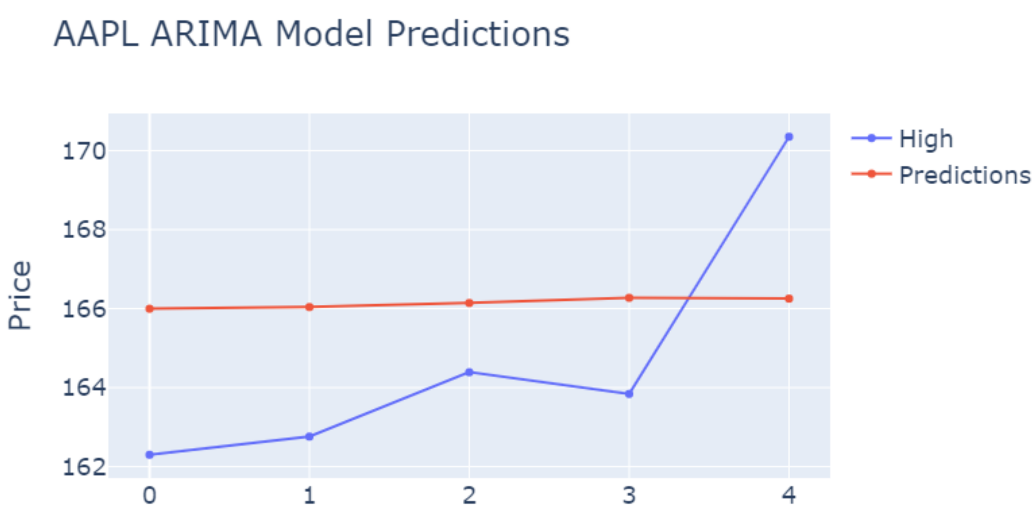
Within the scope of this project, we applied two time series models to specifically Apple Inc. stock as well as Bitcoin cryptocurrency. The models we chose to use are the Autoregressive model, which uses the patterns over time present in the data to create forecasts of future data, as well as the ARIMA model which incorporates and extends the AR model to further increase its performance. The autoregressive model is a type of regression which uses previous values in a time series to create a fit on the next point. The number of previous values used depends on the order of the model, often referred to as the lag parameter. Incorporating more previous observations can lead to a better model, but it can also cause it to perform worse if the previous points don't correlate with the current point in time. To help us find the right balance and pick the appropriate lag parameter, we can use the concept of autocorrelation. This approach looks at the linear relationship between points and previous points using a multitude of lag parameters and can be output into a chart to discern which values would be most effective. The chart below was created using the *statsmodels* library in Python, which contains many tools to allow for efficient time series analysis.

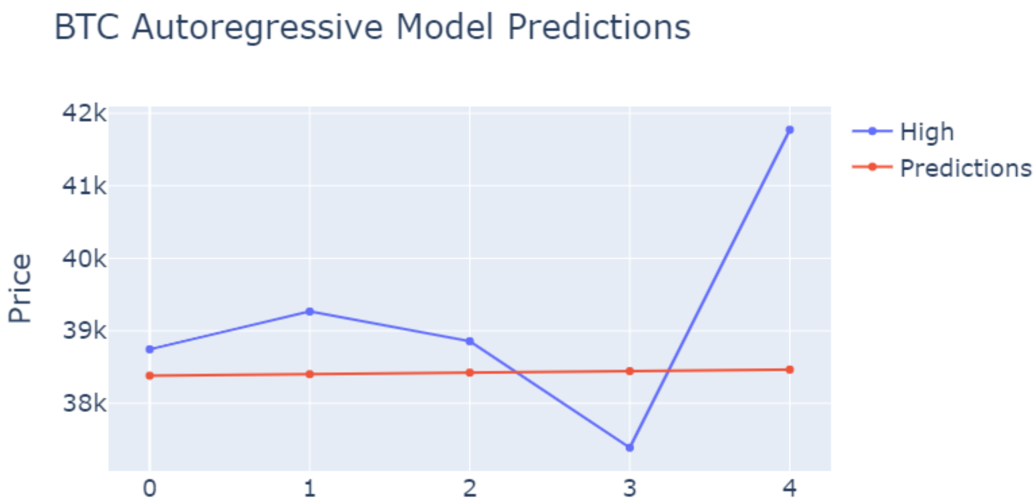


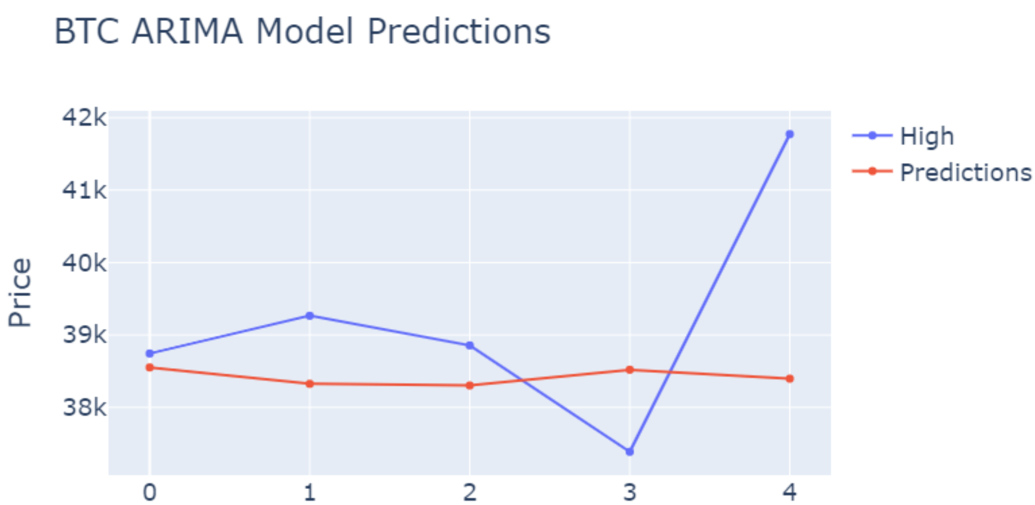
The ARIMA (Autoregressive Integrated Moving Average) model is an improved version of the AR model with incorporation of other elements, namely the concepts of integration and moving averages. This enables the model to take more information into account at the cost of a more complex model that needs greater computational resources. ARIMA is quite suited to the realities of financial time series data, in which data is often more complicated than a simple trend that an AR model could pick up on. The parameters involved are the lag parameter discussed earlier, the difference of the model to be used for stationarity and the size of the moving average window. These parameters lend themselves to great versatility and allow one to create the constituent models by simply setting certain parameters equal to zero. We utilized a python library called *pmdarima* for performing grid searches for the best triplet of parameters which minimized the value of AIC, which stands for Akaike Information Criteria and is a measure of the amount of information loss in a model.

We used both models in our efforts to see whether stock market data could be efficiently used for predictions of future data points, with our data coming from the historical cryptocurrency and stock market data tables we pulled from our SQL database. Taking one particular set from within each table, Apple and Bitcoin respectively, we set up our data for use in machine learning. Since these algorithms can be used with only the dependent variable, we eliminated all columns in our data except for the high price of the equity. The date present in the observation was also eliminated, since it had too many holes due to holidays and other occasions that it was not possible to set up frequency data for the dates. Thus, we simply utilized the index of the observations, which fulfilled the same purpose. Satisfied that the data was ready, we split it into training and testing sets for use in setting up the model and comparing predictions to actual values. We created the models using a variety of parameters, incorporating the recommendations from the ACF and PACF plots, as well as utilizing grid searches to evaluate the ARIMA models for the best parameters. The results of the automatic grid search showed that the AIC values for the majority of combinations were very close, which demonstrated that the combinations did not seem to have a significant impact on the model. These suspicions were confirmed when we plotted the predicted and actual observations on a chart, shown below for both Apple stock and Bitcoin:


AAPL Autoregressive Model Predictions
Price
170
168
166
164
162
0
1
2
3
4
High
Predictions


AAPL ARIMA Model Predictions
Price
170
168
166
164
162
0
1
2
3
4
High
Predictions





The results from these models and charts answer our question about the predictability of stock market and cryptocurrency data. The volatility of both presents a challenge to the way both the AR and ARIMA models function, which affects their predictions in a negative way. Both models rely on the previous data to be following a trend which can be detected, but the movement of the markets proves to be too erratic to develop any meaningful mathematical equation for these algorithms to produce effective results, which is why the output is a simple flat line.

Final investment recommendations

The sixteen stocks in our portfolio seem to represent the sounder investment plan over the four cryptocurrencies, as looking across the timeframe for the historical data in this report, ten of the stocks appreciated in market value, where all but one cryptocurrency depreciated. In addition to this, the chosen stocks show consistent value gains after periods of losses, thereby speaking to a measure of stability in their value, whereas the cryptocurrencies display a sustained downward trend starting around the fall/winter of 2021 with no obvious sign of reversal.

The census data offers limited but still noteworthy insights. The data shows that the financial industry accounts for a disproportionately high number of physical establishments within the dataset, with commercial banking responsible for the vast majority. A similar story plays out in the revenue data: commercial banking makes up a significant majority of the recorded revenue. While this doesn't offer any direct insight into our selection of stocks, it does imply a relationship between physical establishments and revenue for the commercial banking industry. As such, it can be said that if a significant discrepancy were to exist between a commercial bank's revenue and number of establishments, where the latter outstripped the former, it would make for a poor investment choice.

Lastly, while the machine learning model did not offer any real predictive depth, it does however predict the immediate date indexes with relative accuracy. If this predictive capacity can be repeated consistently, then the model would offer a sustained yet limited insight into a given stock's or cryptocurrency's value in the immediate trading days, which could be utilized for short-term investing.

Overall, between the portfolio of stocks or cryptocurrencies, we recommend the stocks. Their tendency towards appreciation, rebounding after losses in market value, and the potential for outside data to inform the economic picture within which the represented company exists allows for stocks to be a safer, more predictable investment.

Sources

1. F. (2022). *Finnhub - Free realtime APIs for stock, forex and cryptocurrency.* Finhub. https://finnhub.io/
2. *API Documentation | Alpha Vantage*. (2022). Alpha Vantage. https://www.alphavantage.co/documentation/
3. *US Census Data*. (2017). Census. https://data.census.gov/cedsci/table?q=ECNNAPCSIND2017.EC1700NAPCSINDPRD