Cryptocurrency Pattern Analysis

Adam Yanai - April '25

Introduction

The aim of this project is to analyze certain trends and patterns across the top eleven cryptocurrencies by market capitalization (BNB - Binance Coin, BTC - Bitcoin, XRP - Ripple, SOL - Solana, ADA - Cardano, USDT - Tether, DOGE - Dogecoin, ETH - Ethereum, TRX - TRON, USDC - USD Coin, and LDO - Lido). Moreover, the analysis will be used to predict the future movements of those cryptocurrencies through the utilization of machine learning models. Cryptocurrencies are integral components of the evolving financial ecosystem, and understanding their behavior and influence can prove crucial for investors, analysts, and institutions. The provided datasets are organized by date (one data point per day), price, volume, and market capitalization. Therefore, the central objective is to predict future values of cryptocurrency prices and market capitalization, as well as identify the underlying patterns and trends that influence their movements.

Although the data set includes price, volume, and market capitalization, this report will focus on the correlation between price and market capitalization and the effectiveness of these points in demonstrating the behaviour of cryptocurrencies. Furthermore, the study will explore predictions for a day, week, month, and year, and ultimately showcase that the month-based predictions are the most effective. The analysis primarily relies on linear regression and ridge regression models, but finds that the ridge regression models are far more effective, as they are better suited for time-series forecasting in financial markets.

The cryptocurrency data used in this project was provided by the Faculty of Engineering at the University of Sydney. The methodology for this project consisted of three major stages: (1) data transformation & cleaning, (2) data visualization & statistical pattern recognition, and (3) trend identification & prediction. Throughout the first stage, the data was cleaned, normalized, and transformed to be high quality and optimal for machine learning modeling. The second stage had the treated data visualized, which allowed for initial calculations and pattern identifications to be made. The third stage used the information gathered throughout the second stage to create models that make predictions on the price and market capitalization of the cryptocurrencies, using linear and ridge regression. Tools used include Python (Pandas, NumPy, Sklearn, Matplotlib) and Google Colab.

(I) Data Cleaning and Transformation

Initial Data Visualization

To gain a foundational understanding of the project, the dataset was plotted in its entirety. The data of all eleven cryptocurrencies was formatted in an identical CSV file structure, in which a single data point existed per day (taken at 00:00:00 UTC+0), with each cryptocurrency having varying timelines based on its creation and growth. Each data point presented the Price (cost of a unit), Volume (total value of all coins traded in a day), and Market Capitalization (total value of all coins in circulation). Plotting was done in three parts according to the three fields, and in four graphs for each part to facilitate viewing and understanding (the cryptocurrencies with the larger values were removed in sequential order). The three figures below (Figure 1, Figure 2, and Figure 3) are the plots for the three fields.

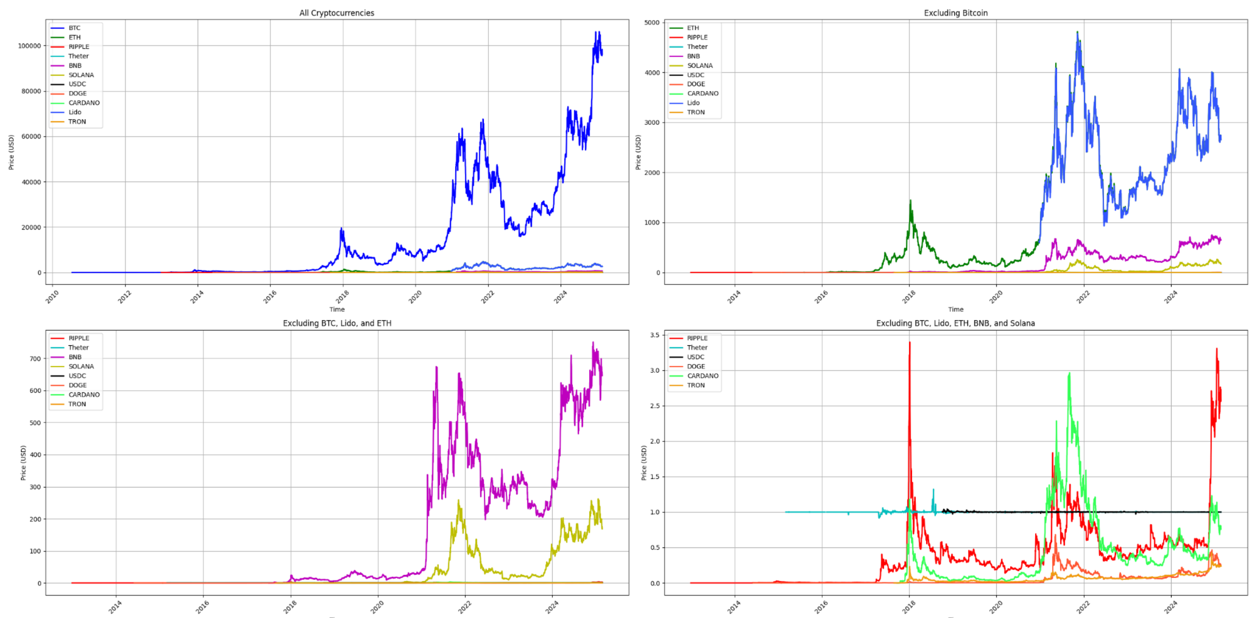


Figure 1: Initial Data Visualization (Price)

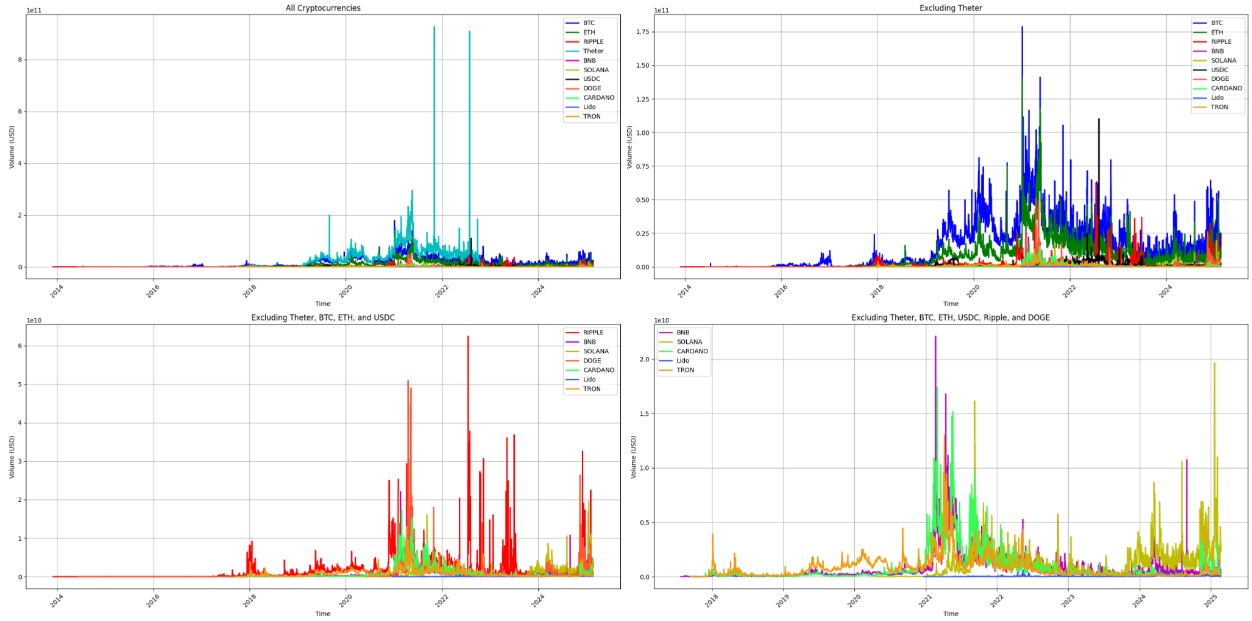


Figure 2: Initial Data Visualization (Volume)

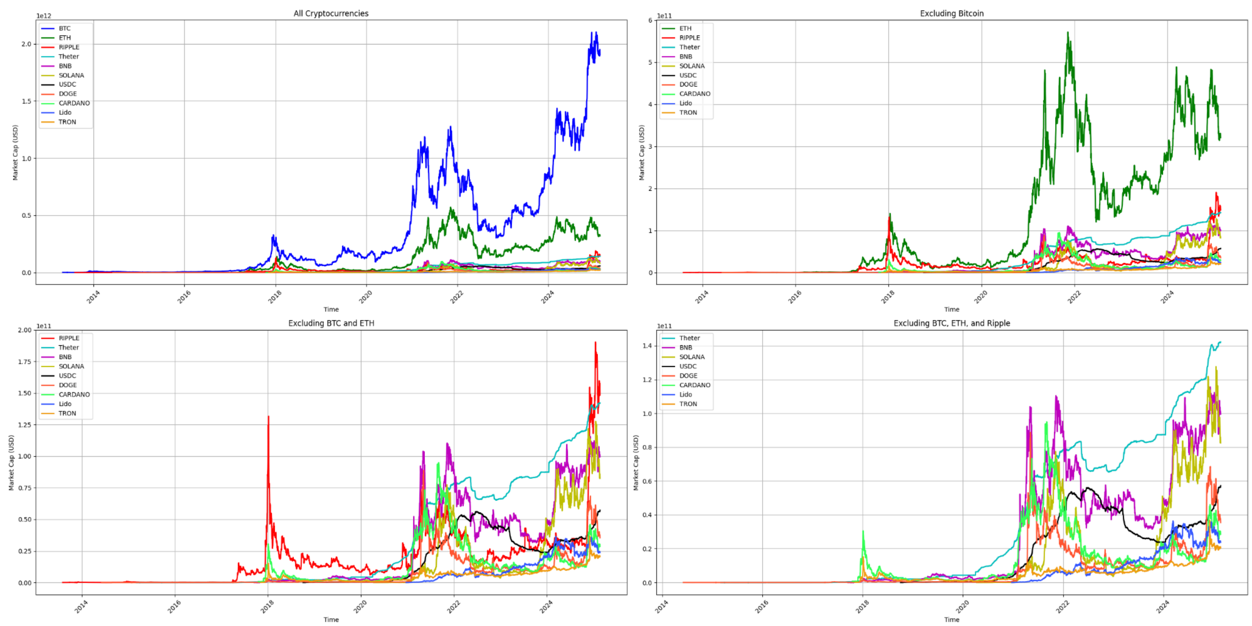


Figure 3: Initial Data Visualization (Market Capitalization)

Through the visualization of the data, I observed that the cryptocurrencies have different starting dates and that they have widely different value ranges, which would create difficulty in analyzing them together. Therefore, solutions must be found for the starting dates (cleaning and transformation) and the variance in the value ranges (normalization).

Data Cleaning & Transformation

To begin the data cleaning process, I first checked for missing values and duplicate rows in the dataset by reading through the files with pd.read_csv(file). I did so by searching for all null cells and cells that share exact values with other cells, and I organized the outputs by category as shown in Figure 4.

```
Missing values in BTC.csv: Date (0), Price (1), Volume (1232), Market_cap (1018)
Duplicate rows in BTC.csv: 0
Missing values in ETH.csv: Date (0), Price (1), Volume (1), Market_cap (2)
Duplicate rows in ETH.csv: 0
Missing values in RIPPLE.csv: Date (0), Price (0), Volume (333), Market_cap (220)
Duplicate rows in RIPPLE.csv: 0
Missing values in Theter.csv: Date (0), Price (23), Volume (23), Market_cap (23)
Duplicate rows in Theter.csv: 0
Missing values in BNB.csv: Date (0), Price (2), Volume (22), Market_cap (75)
Duplicate rows in BNB.csv: 0
Missing values in SOLANA.csv: Date (0), Price (0), Volume (0), Market_cap (0)
Duplicate rows in SOLANA.csv: 0
Missing values in USDC.csv: Date (0), Price (0), Volume (0), Market_cap (0)
Duplicate rows in USDC.csv: 0
Missing values in DOGE.csv: Date (0), Price (1), Volume (1), Market_cap (2)
Duplicate rows in DOGE.csv: 0
Missing values in CARDANO.csv: Date (0), Price (0), Volume (8), Market_cap (25)
Duplicate rows in CARDANO.csv: 0
Missing values in Lido.csv: Date (0), Price (0), Volume (0), Market_cap (0)
Duplicate rows in Lido.csv: 0
Missing values in TRON.csv: Date (0), Price (0), Volume (16), Market_cap (73)
Duplicate rows in TRON.csv: 0
```

Figure 4: Output of Missing & Duplicate Values Check

The output demonstrated that many cryptocurrencies had high levels of missing values, with Bitcoin and Ripple having thousands and hundreds of missing values, respectively. Fortunately, it also demonstrated that there were no duplicate rows, so I could focus entirely on the missing values. I wanted to ensure a high-quality dataset with minimal noise and consistency, so I decided to duplicate the dataset files and in the new files drop every row with missing values through df.dropna(inplace=True).

After cleaning the dataset, I wanted to restructure it for optimal use. First, I cleaned the date column by converting the datapoints into actual datetime objects through the function df['Date'] = pd.to_datetime(df['Date'], utc=True, errors='coerce') for better manipulation during analysis. I also resorted the date column in ascending order so that the data is shown from the earliest dates to the most recent ones, as well as made Date the index of the dataset with df.set_index('Date', inplace=True) so that the other data columns are all organized directly by the date. I also implemented another check to ensure that all the values in the Price, Volume, and Market Capitalization columns are real numbers.

Data Normalization

To tackle the problem of the wide variance in the value ranges, I decided to use a normalization, which standardizes the values of all cryptocurrency data sets to be between 0 and 1 (0 is the minimum of the cryptocurrency and 1 is the maximum) so that the focus is not on the values themselves, but rather on the patterns and trends. To do this, I used the equation shown in Figure 5, in which *X* is the original numeric value in a column, *Xmin* is the minimum value in that column, *Xmax* is the maximum value in that column, and *X'* is the new normalized value in the column.

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

Figure 5: Normalization Equation

In the code, this was implemented through importing and using MinMaxScaler, a tool that scales all values in a column between 0 and 1 based on the relationship between the values. The normalized values were attained through MinMaxScaler().fit_transform(df[numeric_cols]), and were input into new columns in the duplicate dataset files (which already had the cleaned versions of the Price, Volume, and Market Capitalization columns) as Price_normalized, Volume_normalized, and Market_cap_normalized (shown in Figure 6 below). Once the new cleaned, transformed, and normalized files were complete, they were saved under “Cleaned_[cryptocurrency].csv” in an entirely new data folder.

Cleaned_TRON

Date	Price	Volume	Market_cap	Price_normalized	Volume_normalized	Market_cap_normalized
2017-11-09 00:00:00+00:00	0.00238682	1224287.170176	156404161.61769100	0.0013732423927244600	9.27736850222454E-05	0.0010417584357038600
2017-11-10 00:00:00+00:00	0.00204444	990422.8195	133968506.2492820	0.0005660546258981280	7.48099797722263E-05	0.0004294137357534310
2017-11-11 00:00:00+00:00	0.00191476	707642.964341	125470649.38402900	0.0002603238308691020	5.30889523730697E-05	0.00019747856261372500
2017-11-12 00:00:00+00:00	0.00180434	814789.232403	118235245.9032630	0.0	6.13191250032363E-05	0.0
2017-11-13 00:00:00+00:00	0.00201768	894985.567811	132386427.48493500	0.0005029658221120640	6.74792059528841E-05	0.00038623347066449900
2017-11-14 00:00:00+00:00	0.00241536	1073924.144644	158479487.413139	0.0014405276864484600	8.12239252433381E-05	0.0010984010757625200
2017-11-15 00:00:00+00:00	0.0023203	1179822.352577	152242301.026122	0.0012164162631336900	8.935823125584E-05	0.0009281672241554670
2017-11-16 00:00:00+00:00	0.00224758	1324833.2011	147470996.18371	0.0010449731461186400	0.00010049687694250700	0.0007979422236318490
2017-11-17 00:00:00+00:00	0.00202291	1905625.229589	132729588.66321	0.0005152959582780250	0.0001451089641681970	0.0003955994965413840
2017-11-18 00:00:00+00:00	0.00202237	2184964.124249	132693856.773722	0.0005140228658249430	0.0001665656827917650	0.000394624252770346
2017-11-19 00:00:00+00:00	0.00198527	1832646.818395	130259776.10292	0.0004265566991409760	0.00013950331001648700	0.00032818998225128200
2017-11-20 00:00:00+00:00	0.00213324	1679383.57715	139968958.501911	0.0007754076070716130	0.00012773077742043500	0.0005931863144398040
2017-11-21 00:00:00+00:00	0.00212957	1420954.846383	139727776.589602	0.0007667552935478890	0.00010788022074424300	0.0005866036466593790
2017-11-22 00:00:00+00:00	0.00233308	1345920.56312	153081053.186036	0.0012465461178566300	0.0001021166499208710	0.0009510595993173040
2017-11-23 00:00:00+00:00	0.00213762	3417832.230171	140256267.319897	0.000785733801413278	0.00026126536319587600	0.0006010279410405780

Figure 6: Example of Cleaned Cryptocurrency File (TRX - TRON)

Confirmation of High Quality Data

Once the entire data cleaning, transforming, and normalizing process was over, I ran the cleaned cryptocurrency files through the missing values and duplicate rows checking program and confirmed that the data had been properly cleaned. The output demonstrated that no missing values were remaining in any of the cryptocurrency files, as shown in Figure 7 below.

```
Missing values in Cleaned_BTC.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_BTC.csv: 0
Missing values in Cleaned_ETH.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_ETH.csv: 0
Missing values in Cleaned_RIPPLE.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_RIPPLE.csv: 0
Missing values in Cleaned_Theter.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_Theter.csv: 0
Missing values in Cleaned_BNB.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_BNB.csv: 0
Missing values in Cleaned_SOLANA.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_SOLANA.csv: 0
Missing values in Cleaned_USDC.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_USDC.csv: 0
Missing values in Cleaned_DOGE.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_DOGE.csv: 0
Missing values in Cleaned_CARDANO.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_CARDANO.csv: 0
Missing values in Cleaned_Lido.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_Lido.csv: 0
Missing values in Cleaned_TRON.csv: Date (0), Price (0), Volume (0), Market_cap (0), Price_normalized (0), Volume_normalized (0), Market_cap_normalized (0)
Duplicate rows in Cleaned_TRON.csv: 0
```

Figure 7: Output of Missing & Duplicate Values Check (Using Cleaned Files)

(II) Data Visualization & Statistical Pattern Recognition

High Quality Data Visualization

With the data entirely cleaned and converted into a high-quality set, I could begin the analysis. First, I re-visualized the data to demonstrate the impact of the cleaning, transforming, and normalizing process and to better analyze the data before beginning to evaluate ways of detecting trends and patterns. I first re-plotted the cleaned Price, Volume, and Market Capitalization data using the same 'sequential removal by size' strategy that I had used for the initial data visualization, as demonstrated in Figure 8, Figure 9, and Figure 10 below.

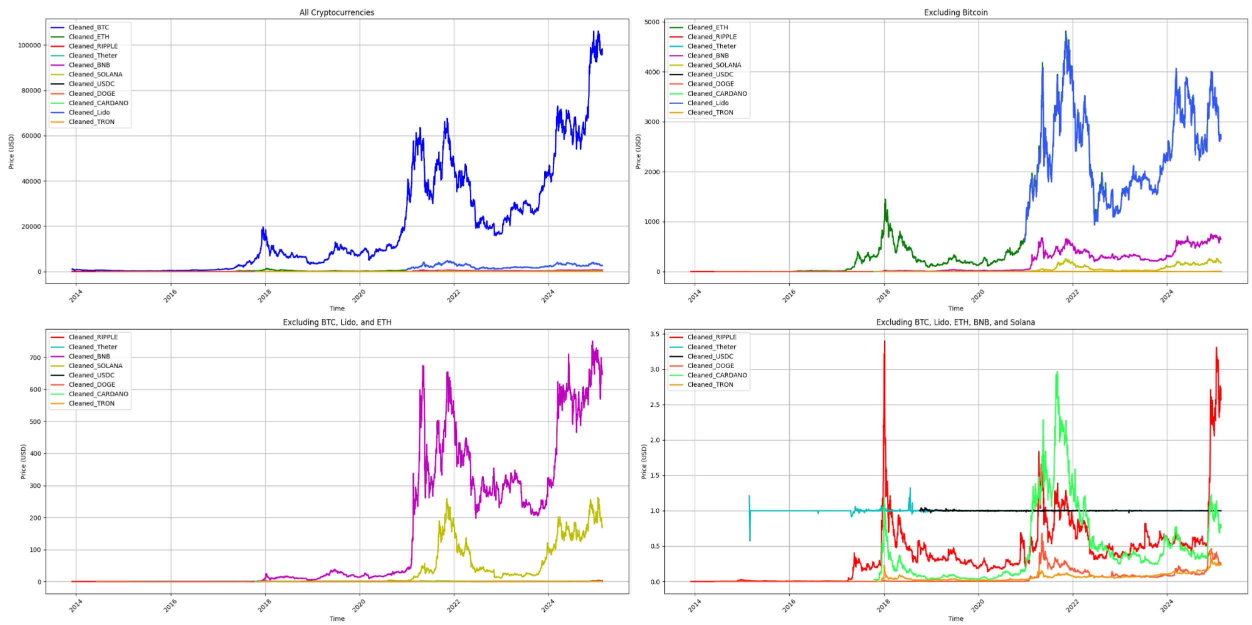


Figure 8: Cleaned Data Visualization (Price)

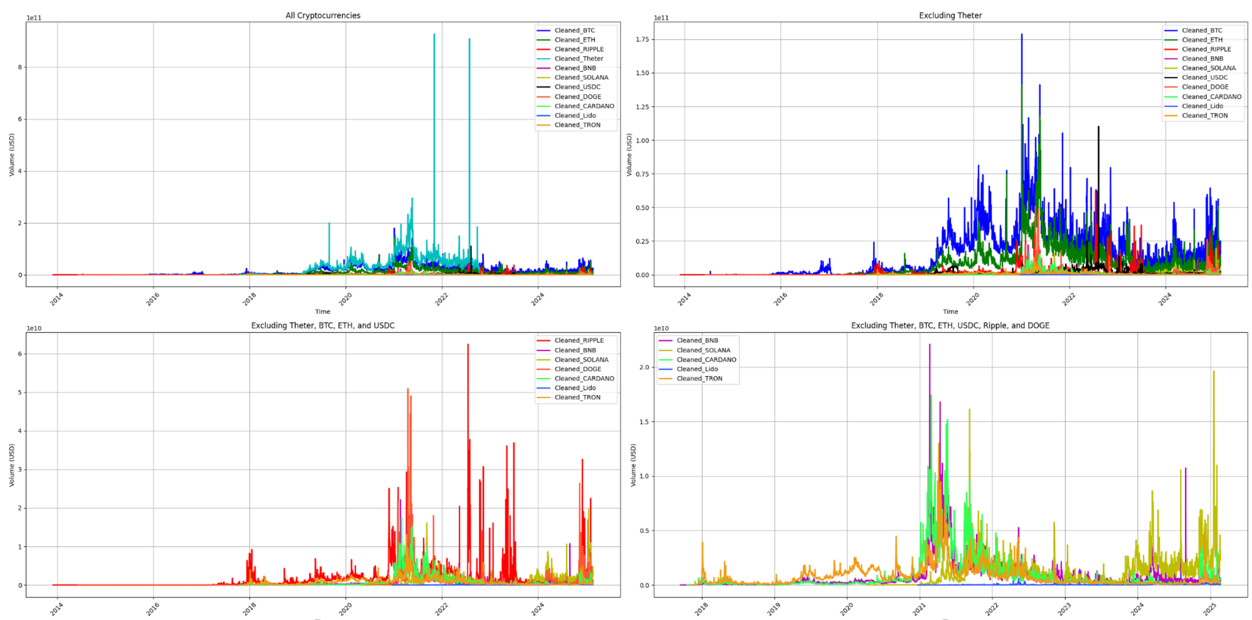


Figure 9: Cleaned Data Visualization (Volume)

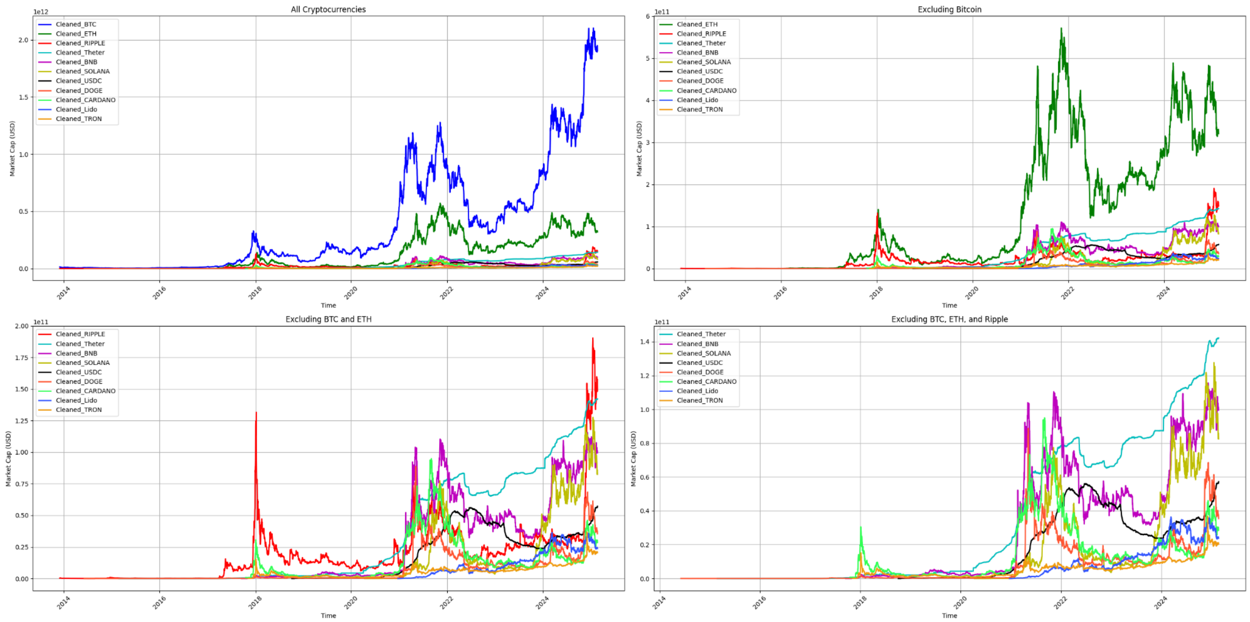


Figure 10: Cleaned Data Visualization (Market Capitalization)

Particularly for the Price plots, the data was significantly better formatted for machine learning modeling on the entire dataset. The dates across the cryptocurrencies were now more aligned and consistent, which facilitated making observations on the market of the eleven cryptocurrencies as a whole. The cleaned but not normalized plots are also effective in showing the similarities in some basic trends, even though they have incredibly different value ranges. For example, in the Market Capitalization plots, there is a clear pattern of peaks around 2022 for many of the currencies, followed by a dip into 2023 and a later increase to another peak around 2025. These very basic trends were much easier to identify with the cleaned data that limited some of the noise and inconsistency from earlier. However, it was still difficult to identify patterns due to the vast difference in value range that persisted. Fortunately, I hadn't yet plotted the normalized data, which is more focused on demonstrating trends and patterns do to its elimination of the role of the value range. The normalization plots are below in Figure 11, Figure 12, and Figure 13.

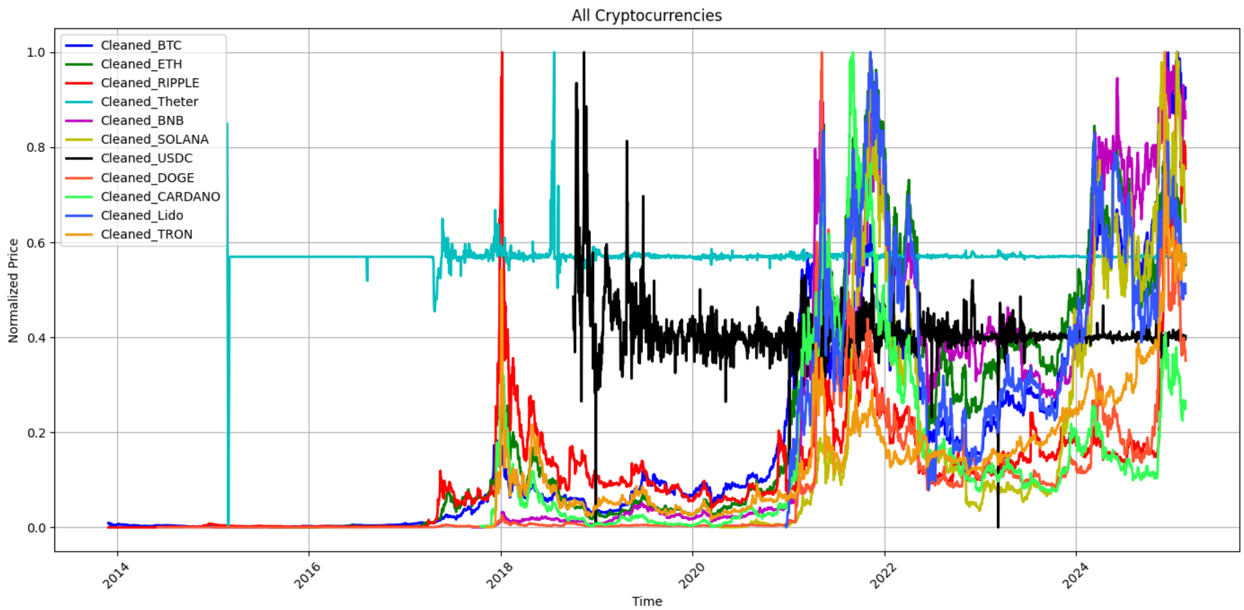


Figure 11: Normalized Data Visualization (Price)

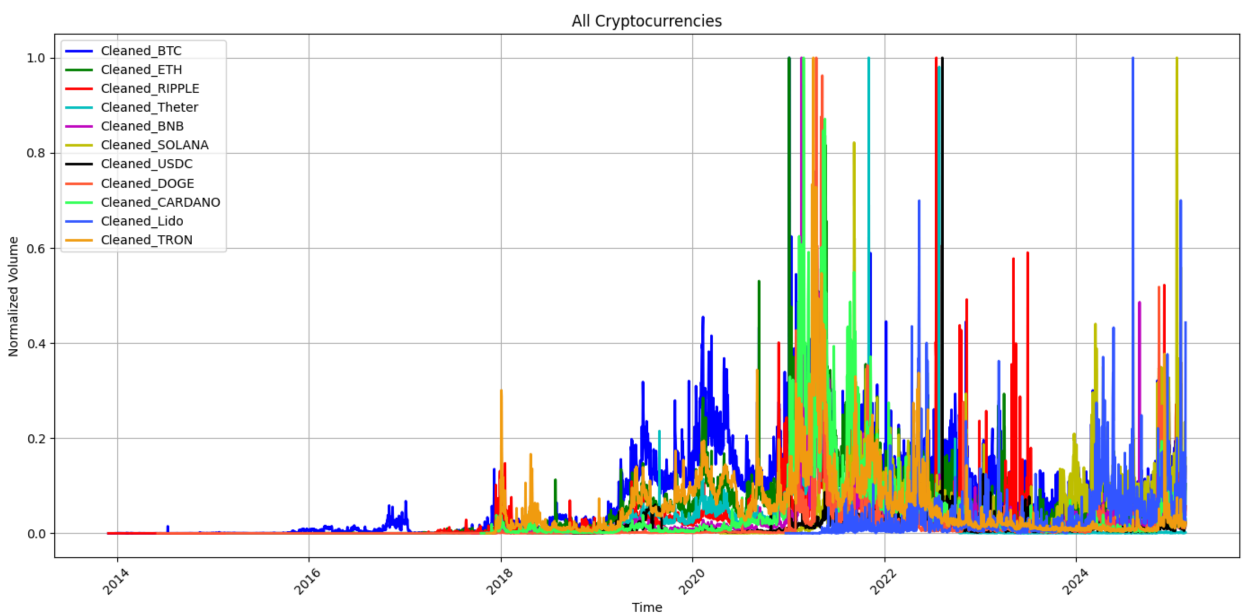


Figure 12: Normalized Data Visualization (Volume)

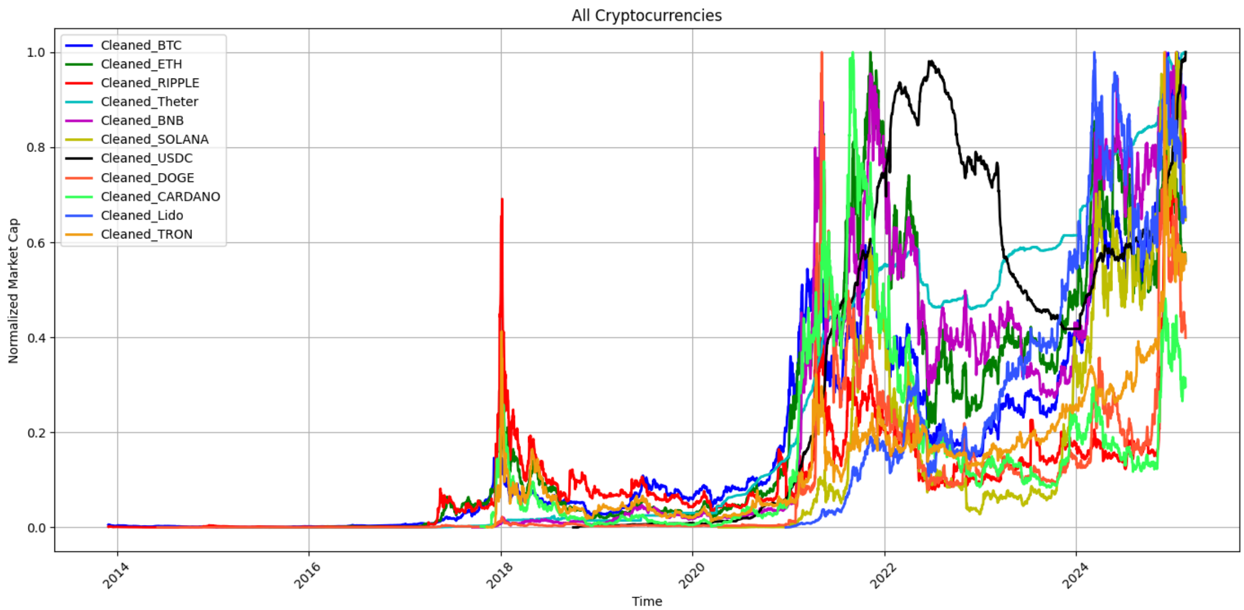


Figure 13: Normalized Data Visualization (Market Capitalization)

Since the problem of value range was now eliminated from plotting, I decided to remove the sequence-based plotting I did before and instead focus solely on all cryptocurrencies plotted together. However, another plotting problem popped up: the Cleaned_Thether and Cleaned_USDC datasets are based on Tether and USDC, which are cryptocurrency stablecoins that are pegged to the United States dollar, meaning that they are designed to remain stable and consistently be worth around one dollar ($1). Therefore, the activities of these currencies in the Price and Market Capitalization plots were completely different from the rest, and due to their distinction as stablecoins, I decided to recreate the Price and Market Capitalization plots with the stablecoins removed, as shown below in Figure 14 and Figure 15.

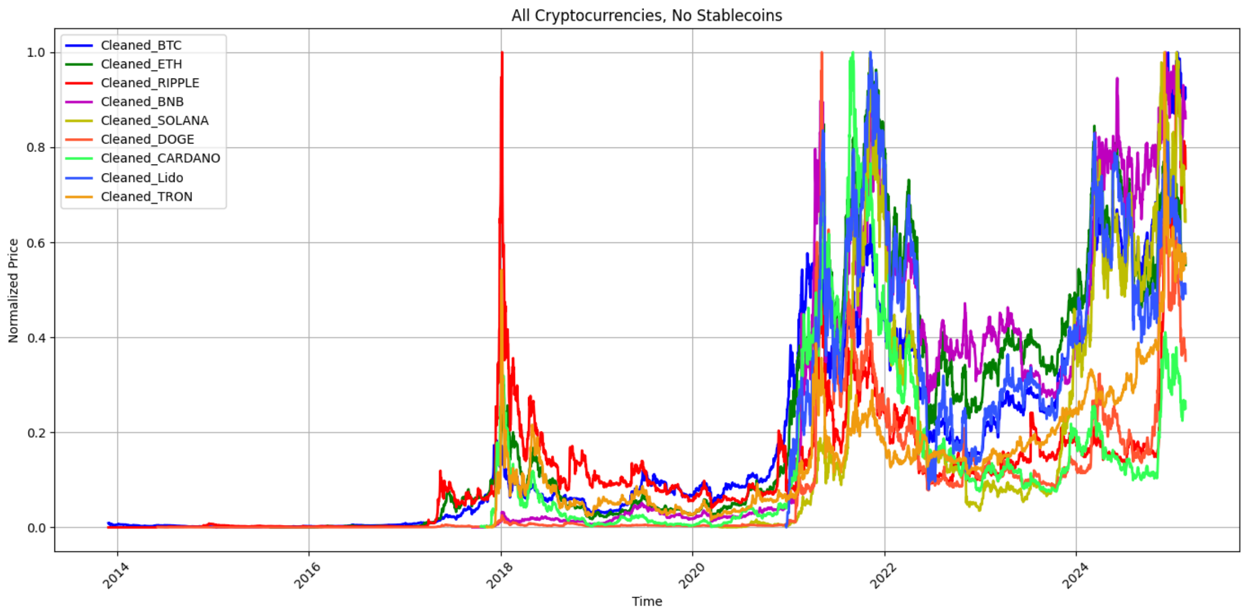


Figure 14: Normalized Data Visualization, No Stablecoins (Price)

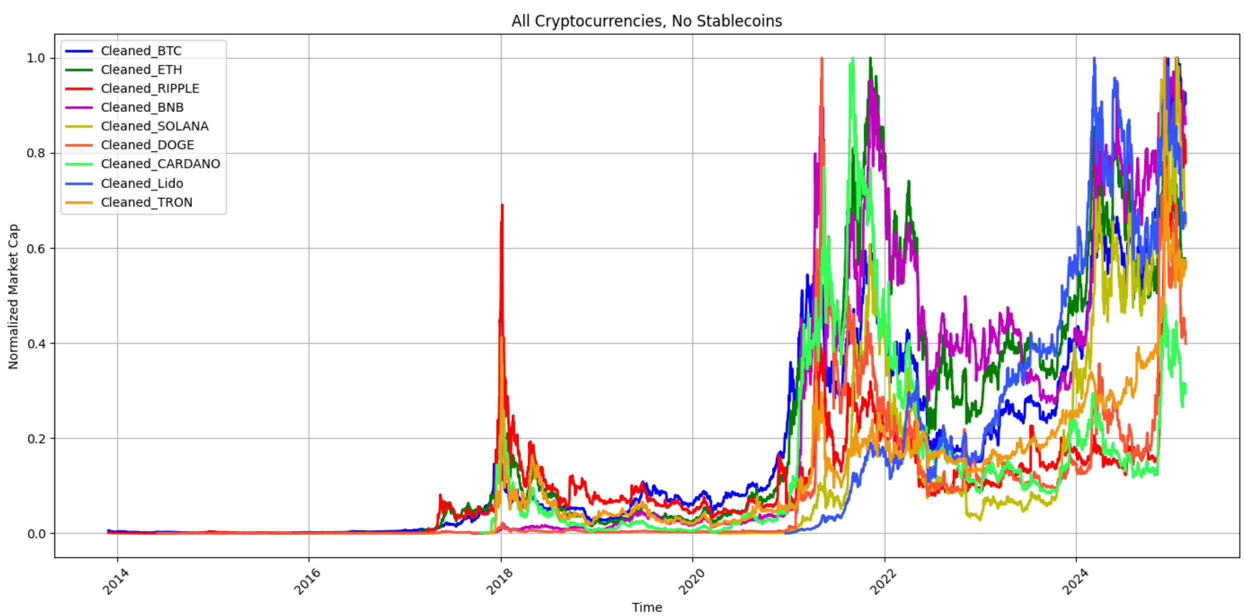


Figure 15: Normalized Data Visualization, No Stablecoins (Market Capitalization)

After removing the stablecoins, the patterns in the non-stablecoin cryptocurrencies are far easier to identify, with the pattern described earlier taking full effect. Many cryptocurrencies experienced a small peak in 2018, followed by a stabilization into 2020, a major peak in 2022, a re-stabilization in 2023 (with much higher values than in the 2020 stabilization), and another peak past 2024 and into 2025, for both the Price and Market Capitalization plots. Figure 14 and Figure 15 demonstrate the incredibly strong correlation between Price and Market Capitalization, and this correlation will be studied further in the following section.

Statistical Pattern Recognition

The visualization of the high-quality dataset that I created was impactful in gathering some initial patterns and trend identification in the data. Next, statistical measures such as mean, median, and standard deviation were calculated for each of the three categories (Price, Volume, Market Capitalization) for each of the eleven cryptocurrencies independently. Additionally, the correlation between each pair of the three categories was calculated for each of the cryptocurrencies. This was done through writing a function, calculate_statistics(file_paths), that loads the cleaned data (not normalized) from the files and calculates statistical measures (mean, median, standard deviation, and correlation) and outputs them in groups by the cryptocurrency, as shown in Figure 16.

```
Statistics for Cleaned_BTC:
  mean_price: 19259.849667069506
  median_price: 8710.022479695
  std_dev_price: 23551.451969969257
  mean_volume: 14724577587.563084
  median_volume: 8132239132.828295
  std_dev_volume: 17370915026.353333
  mean_market_cap: 367328876152.6327
  median_market_cap: 153318682581.47034
  std_dev_market_cap: 461620756914.3003
  correlation_price_volume: 0.49367769897227376
  correlation_price_market_cap: 0.9995555414767063
  correlation_volume_market_cap: 0.47951390639335634
```

Figure 16: Example Output of Base Statistical Calculations (Cleaned_BTC)

According to my earlier observations of patterns and trends being similar across cryptocurrencies for Price and Market Capitalization, I decided to study this relationship. I did so by creating a program that finds the top five most correlated cryptocurrency pairs and plots a heatmap of them all, for both Price and Market Capitalization. I found that the most correlated cryptocurrency pairs by Price were Cleaned_ETH and Cleaned_Lido (correlation = 0.9997), Cleaned_BNB and Cleaned_SOLANA (correlation = 0.8640), Cleaned_BTC and Cleaned_TRON (correlation = 0.8632), Cleaned_SOLANA and Cleaned_Lido (correlation = 0.8553), and Cleaned_ETH and Cleaned_SOLANA (correlation = 0.8543). By Market Capitalization, they were Cleaned_BTC and Cleaned_SOLANA (correlation = 0.9002), Cleaned_Theter and Cleaned_Lido (correlation = 0.8973), Cleaned_Theter and Cleaned_TRON (correlation = 0.8687), Cleaned_BNB and Cleaned_SOLANA (correlation = 0.8494), and Cleaned_ETH and Cleaned_BNB (correlation = 0.8482). Figure 17 and Figure 18 below demonstrate these findings.

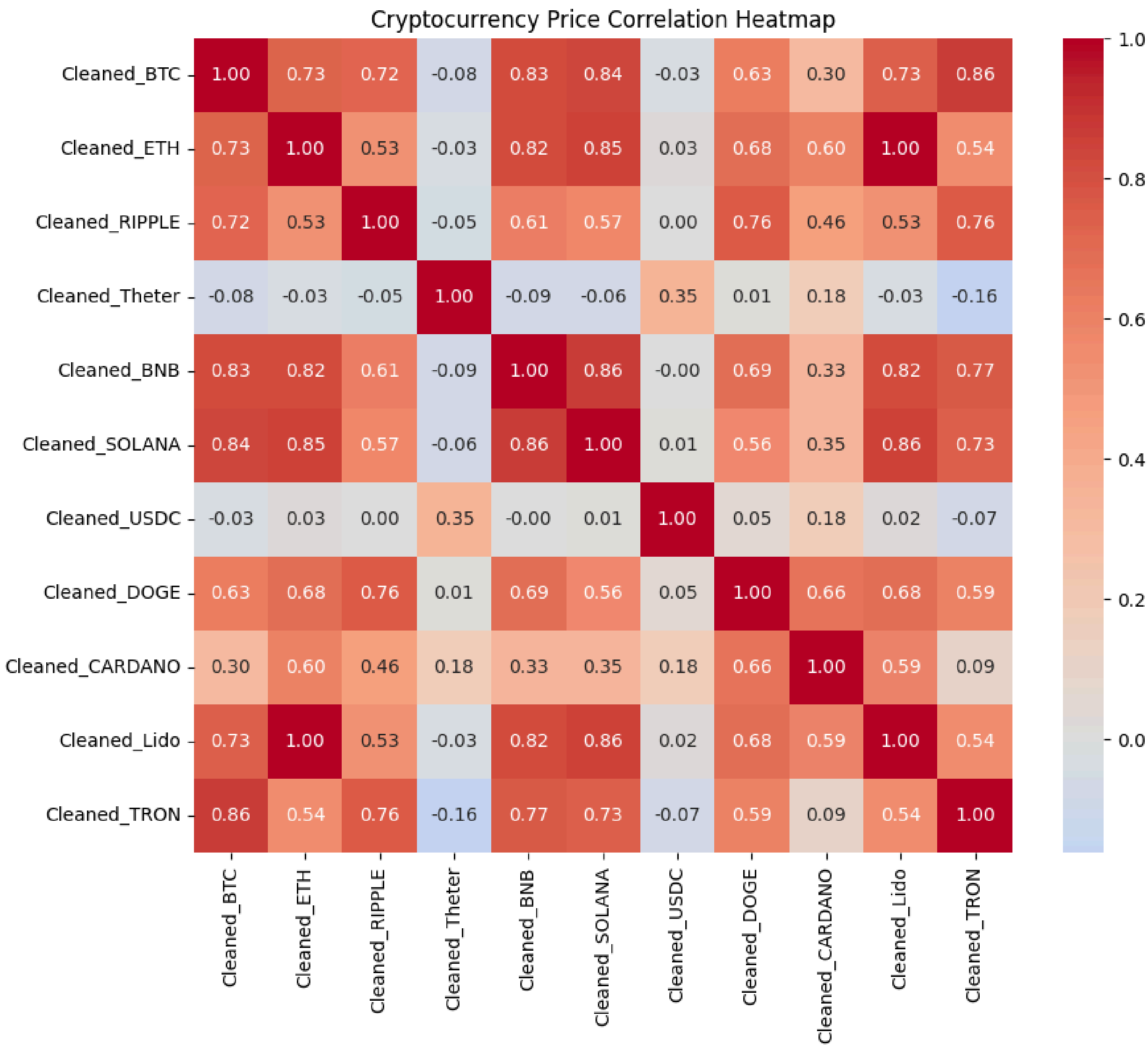


Figure 17: Heatmap of Cryptocurrency Correlations Based on Price

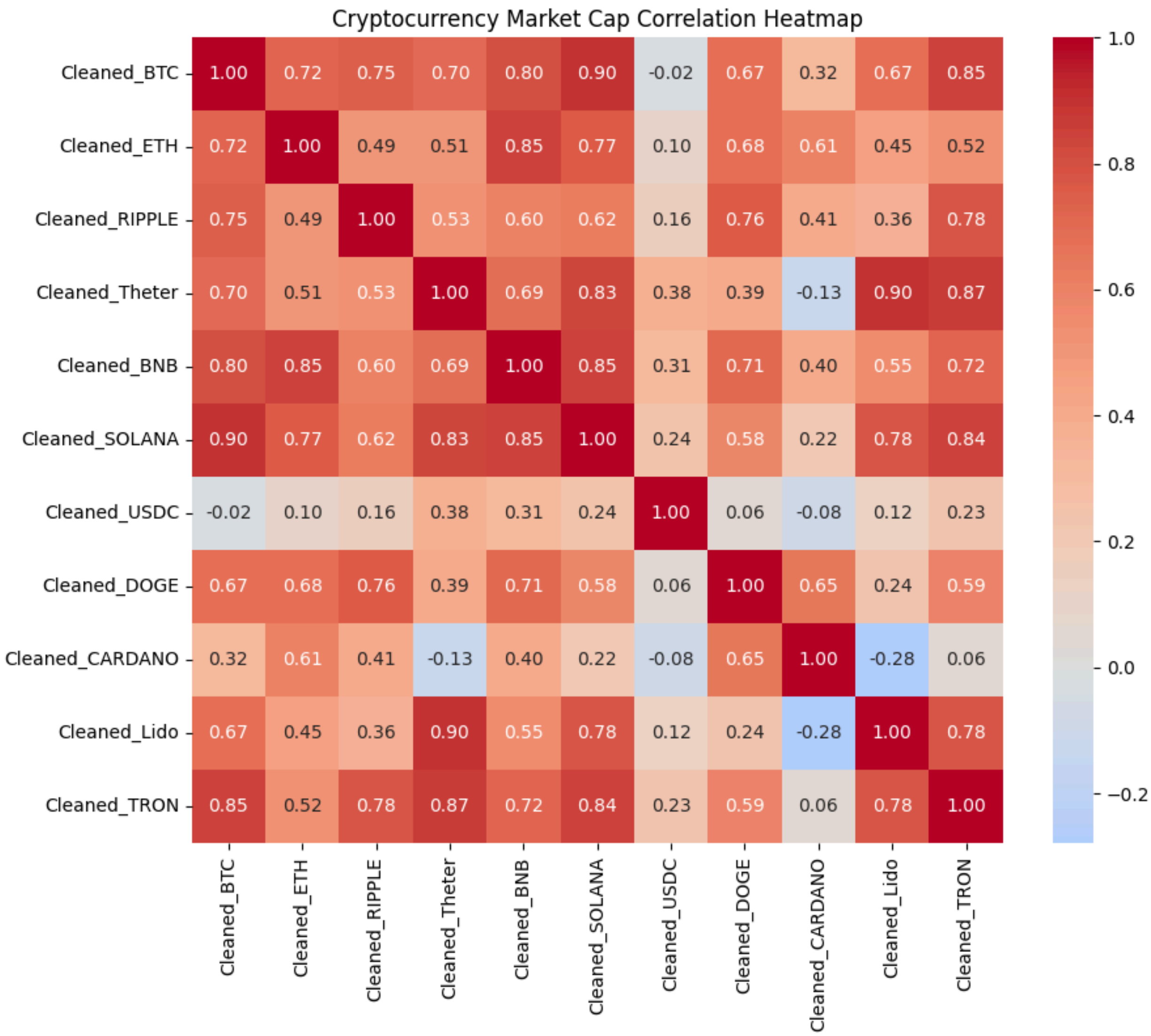


Figure 18: Heatmap of Cryptocurrency Correlations Based on Market Capitalization

These findings further demonstrate my initial pattern identifications and illustrate the connections and relationships between the behaviours of many of the cryptocurrencies. They also demonstrate – particularly in the Price heatmap – how the stablecoins are vastly different in behaviour than the other cryptocurrencies due to their consistency and stability-based design. Next, I wanted to revisit the correlation calculations from earlier between the varying categories (Price, Volume, Market Capitalization) within each cryptocurrency. However, I decided that since the correlations involving Price and Market Capitalization have been the most clear and strong, the rest of the project should focus entirely on those two categories and use them to identify patterns and make predictions. I created another program that again singled out the correlation between Price and Market Capitalization, and the results are shown below in Figure 19.

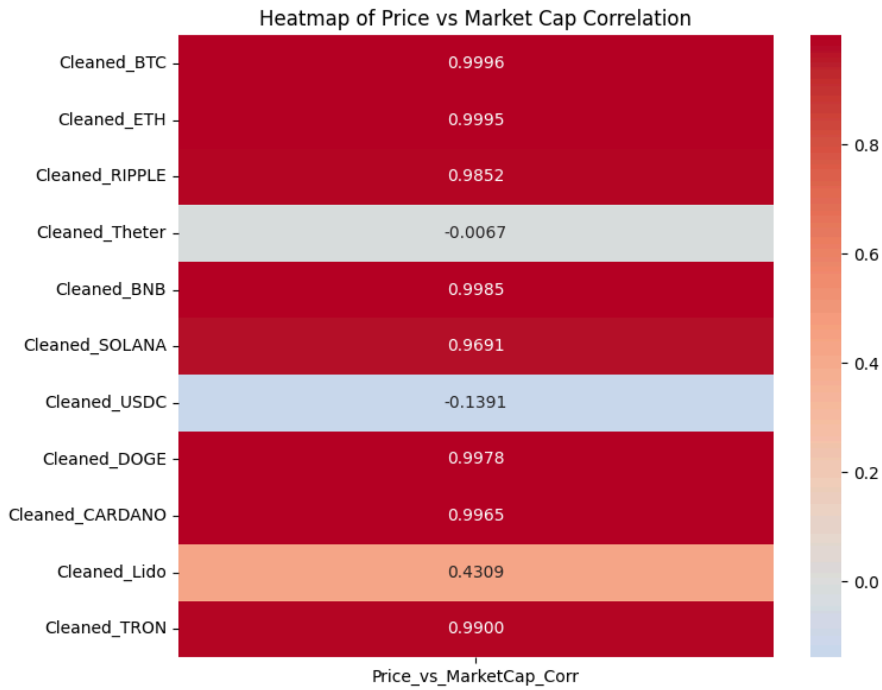


Figure 19: Heatmap of Price and Market Capitalization Correlations Per Cryptocurrency

As illustrated in the Price and Market Capitalization correlations heatmap, the Price and Market Capitalization categories are extremely closely connected and therefore a great pair of data sets to use for making predictions. This reinforced my notion to center on those two categories, and I continued by moving into identifying clear patterns and trends through programming, rather than just observation. In time-series data, trend identification is the finding of persistent movements and directions in the data, which are crucial for predicting future activity. The best trend identifications reveal the underlying market behavior by ignoring the noise, so I decided to develop trendlines and volatility lines for both Price and Market Capitalization for each cryptocurrency. The trendlines are polynomial regression curves that aim to fit the data, while the volatility lines are the rolling standard deviation that capture the fluctuations in the data over time and can highlight price stability/instability. However, after initially plotting and studying the Price and Market Capitalization data along with the trend and volatility lines, I realized that the trendlines were the most impactful in demonstrating patterns and that the degree of the polynomial I had set them to (three) did not follow the actual data as successfully as I hoped they would. Thus, I decided to replot the data with a stronger emphasis on the trendlines and a higher degree polynomial of ten. Below, Figure 20 displays the original, while Figure 21 displays the improved study.

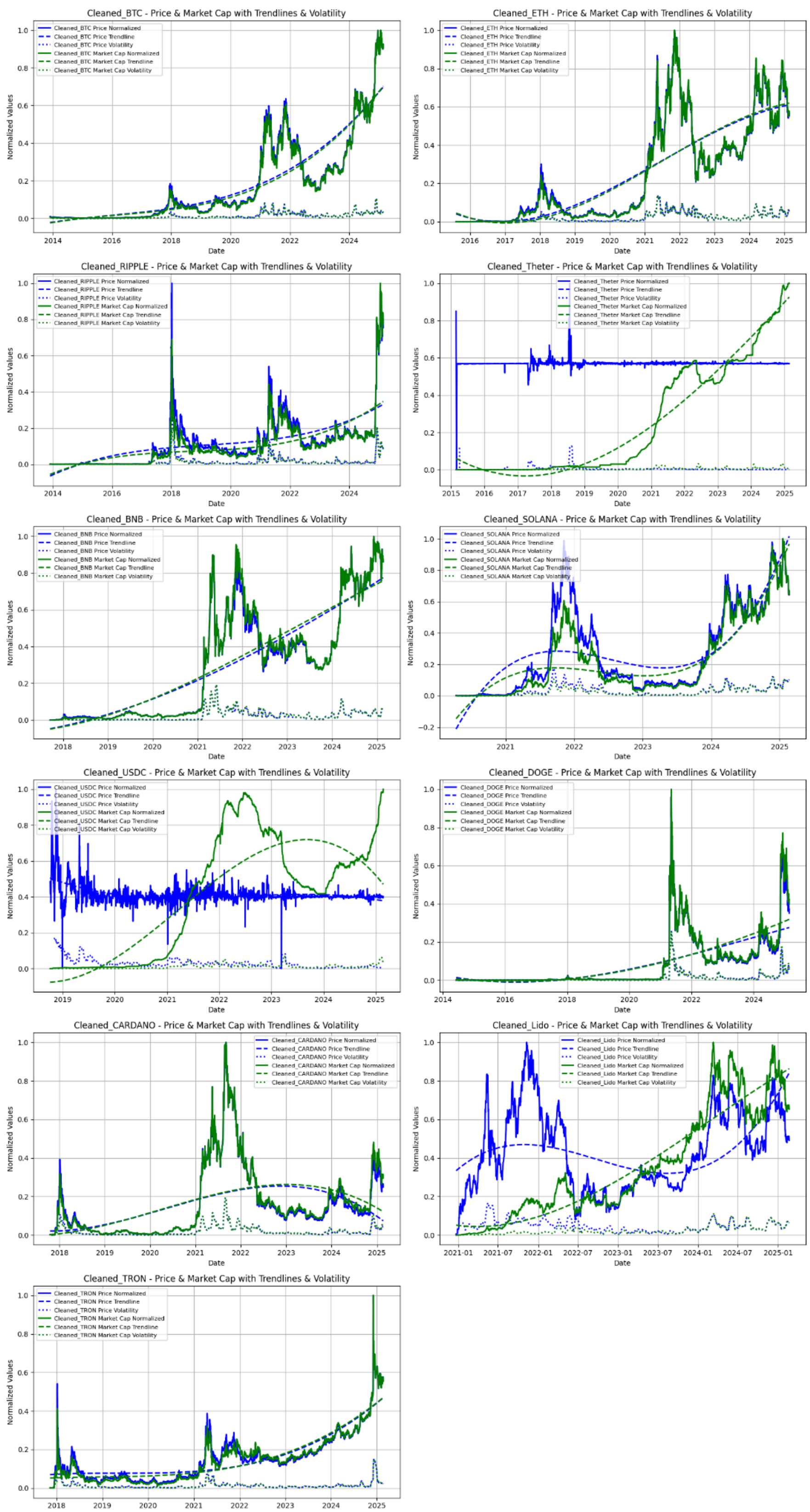


Figure 20: Price and Market Capitalization Plotting with Trendlines (Degree = 3) and Volatility Lines

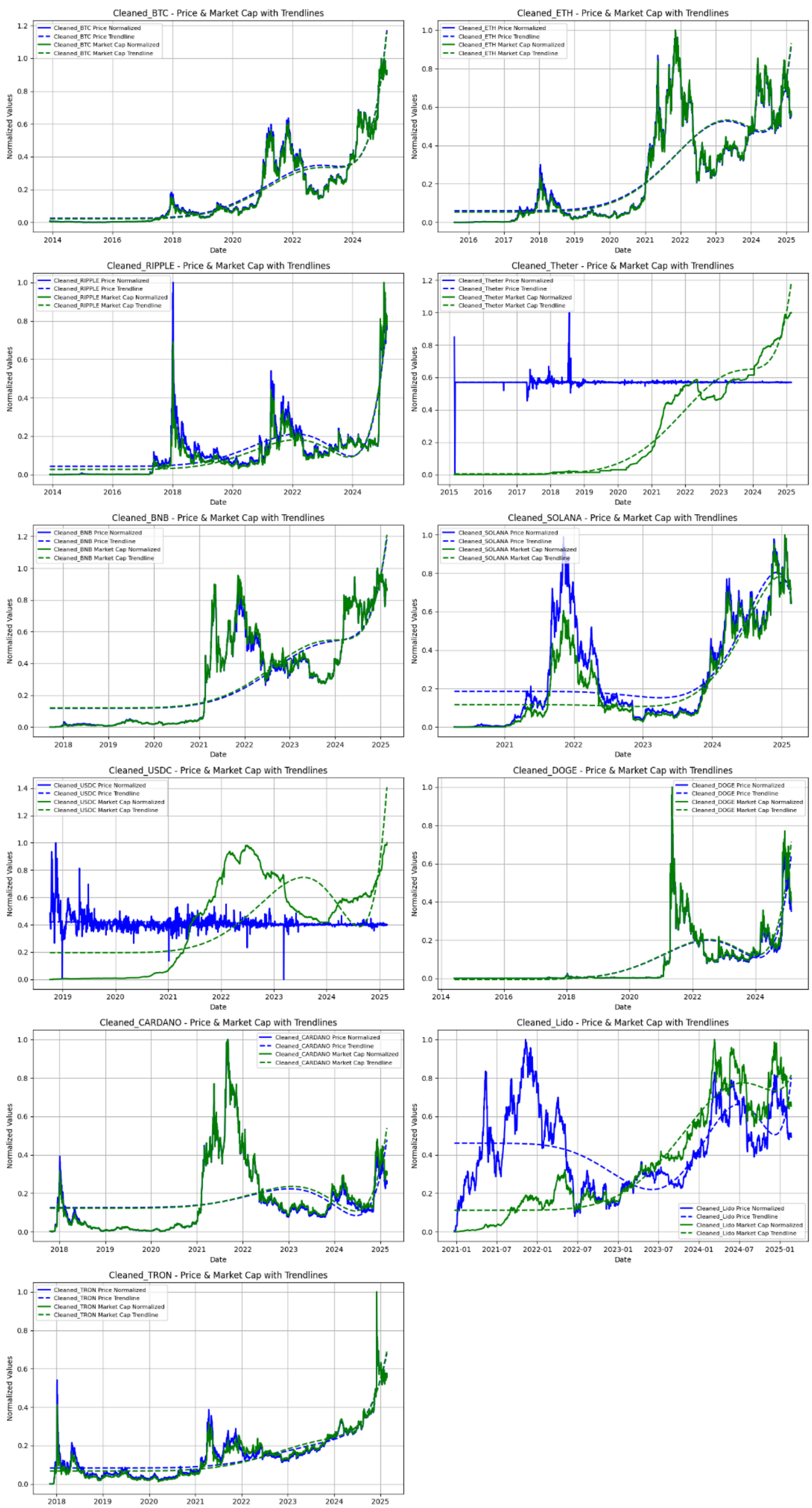


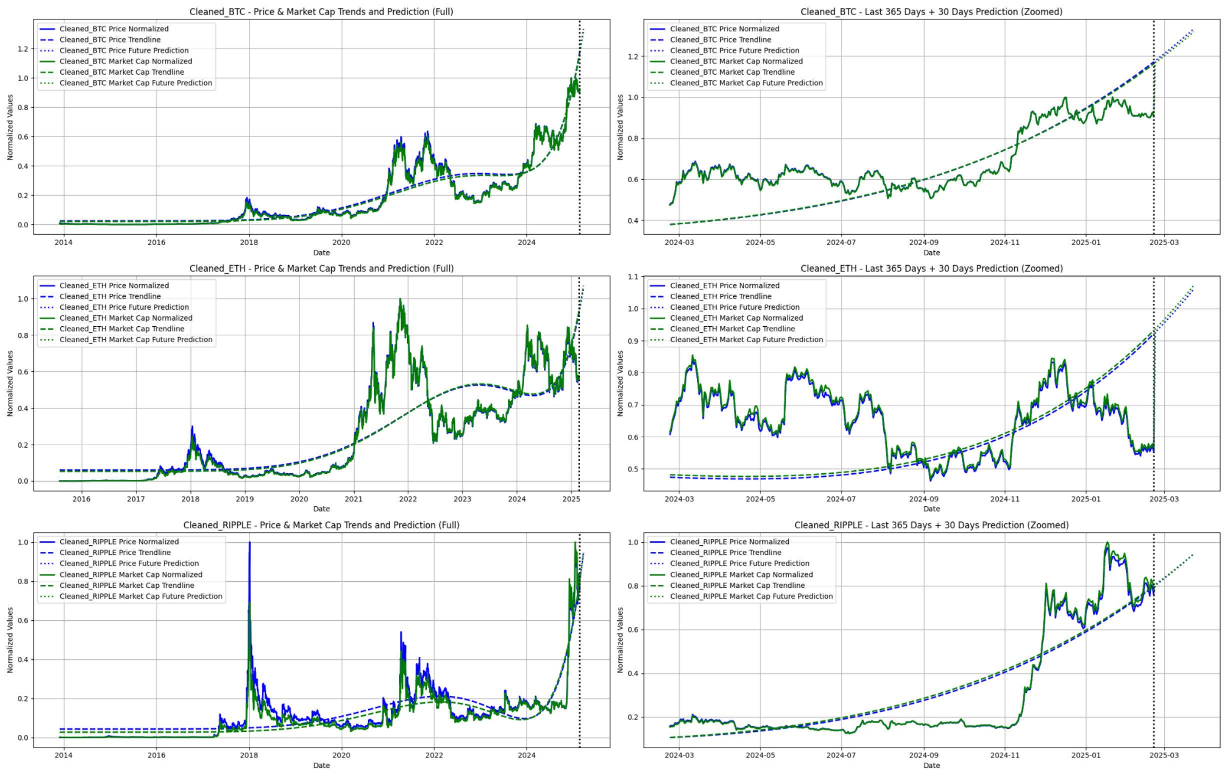
Figure 21: Price and Market Capitalization Plotting with Trendlines (Degree = 10)

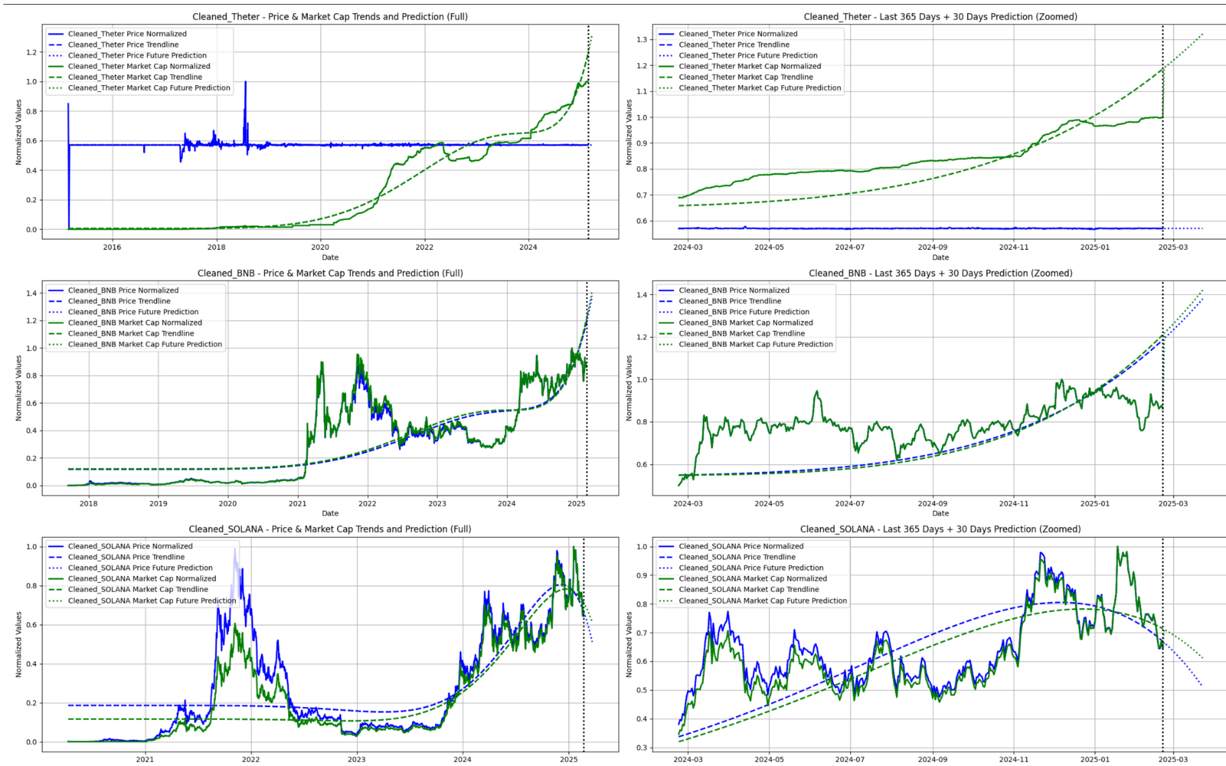
(III) Trend Identification & Prediction Modeling

Prediction Through Linear Regression

After properly implementing trendlines to track the underlying behaviour and patterns of the Price and Market Capitalization data, I moved on to making predictions and decided to implement the linear regression statistical method. Linear regression is used in modeling for relationships between dependent and independent variables, and assumes a linear relationship. I decided to begin with this method because it is simple to use in modeling basic trends, and even if it didn't work ideally for the non-linear trends in cryptocurrency, it would at least provide a strong baseline to work around. I implemented it through a monthly model (making predictions for the next 30 days based on historical linear trends), and created plots that showcased the predictions, as shown in Figure 22. I decided to make monthly predictions as they balance recency and stability. Although they don't track recent changes as well as day and week-based models, they are far less susceptible to the noise picked up by those more concentrated models. Additionally, although monthly models don't create as stable and non-noisy results as do yearly models, the monthly models pick up recent events far better and are more adaptable. These factors make monthly models the most reliable as they pick up recent events while disregarding most noise and hit the 'sweet spot' of modeling.

As I worked through creating models based on linear regression techniques, I quickly realized that the method would not be capable of creating meaningful predictions because, as illustrated in Figure 22, it is constrained and not accurate enough. The trendlines created through the linear regression model did not hug the historical data as closely as I wanted them to, and the resulting prediction lines seemed too inaccurate. This drove me to continue experimenting and try to find new solutions.


Cleaned_BTC - Price & Market Cap Trends and Prediction (Full)
Cleaned_BTC Price Normalized
Cleaned_BTC Price Trendline
Cleaned_BTC Price Future Prediction
Cleaned_BTC Market Cap Normalized
Cleaned_BTC Market Cap Trendline
Cleaned_BTC Market Cap Future Prediction
Normalized Values
Date
Cleaned_BTC - Last 365 Days + 30 Days Prediction (Zoomed)
Cleaned_ETH - Price & Market Cap Trends and Prediction (Full)
Cleaned_ETH Price Normalized
Cleaned_ETH Price Trendline
Cleaned_ETH Price Future Prediction
Cleaned_ETH Market Cap Normalized
Cleaned_ETH Market Cap Trendline
Cleaned_ETH Market Cap Future Prediction
Cleaned_ETH - Last 365 Days + 30 Days Prediction (Zoomed)
Cleaned_RIPPLE - Price & Market Cap Trends and Prediction (Full)
Cleaned_RIPPLE Price Normalized
Cleaned_RIPPLE Price Trendline
Cleaned_RIPPLE Price Future Prediction
Cleaned_RIPPLE Market Cap Normalized
Cleaned_RIPPLE Market Cap Trendline
Cleaned_RIPPLE Market Cap Future Prediction
Cleaned_RIPPLE - Last 365 Days + 30 Days Prediction (Zoomed)


Cleaned_Theter - Price & Market Cap Trends and Prediction (Full)
Cleaned_Theter Price Normalized
Cleaned_Theter Price Trendline
Cleaned_Theter Price Future Prediction
Cleaned_Theter Market Cap Normalized
Cleaned_Theter Market Cap Trendline
Cleaned_Theter Market Cap Future Prediction
Normalized Values
Date
Cleaned_Theter - Last 365 Days + 30 Days Prediction (Zoomed)
Cleaned_BNB - Price & Market Cap Trends and Prediction (Full)
Cleaned_BNB Price Normalized
Cleaned_BNB Price Trendline
Cleaned_BNB Price Future Prediction
Cleaned_BNB Market Cap Normalized
Cleaned_BNB Market Cap Trendline
Cleaned_BNB Market Cap Future Prediction
Cleaned_BNB - Last 365 Days + 30 Days Prediction (Zoomed)
Cleaned_SOLANA - Price & Market Cap Trends and Prediction (Full)
Cleaned_SOLANA Price Normalized
Cleaned_SOLANA Price Trendline
Cleaned_SOLANA Price Future Prediction
Cleaned_SOLANA Market Cap Normalized
Cleaned_SOLANA Market Cap Trendline
Cleaned_SOLANA Market Cap Future Prediction
Cleaned_SOLANA - Last 365 Days + 30 Days Prediction (Zoomed)

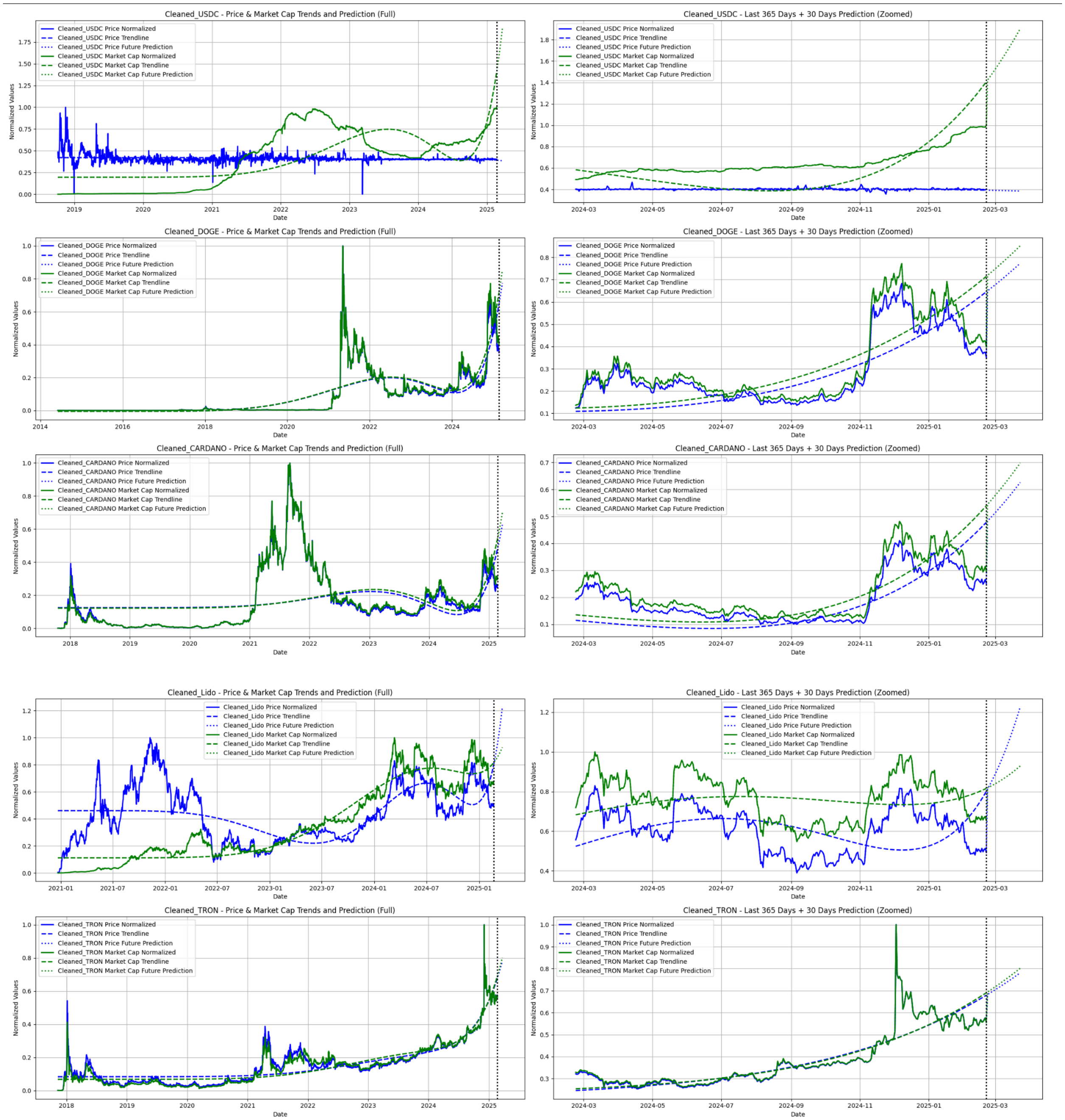
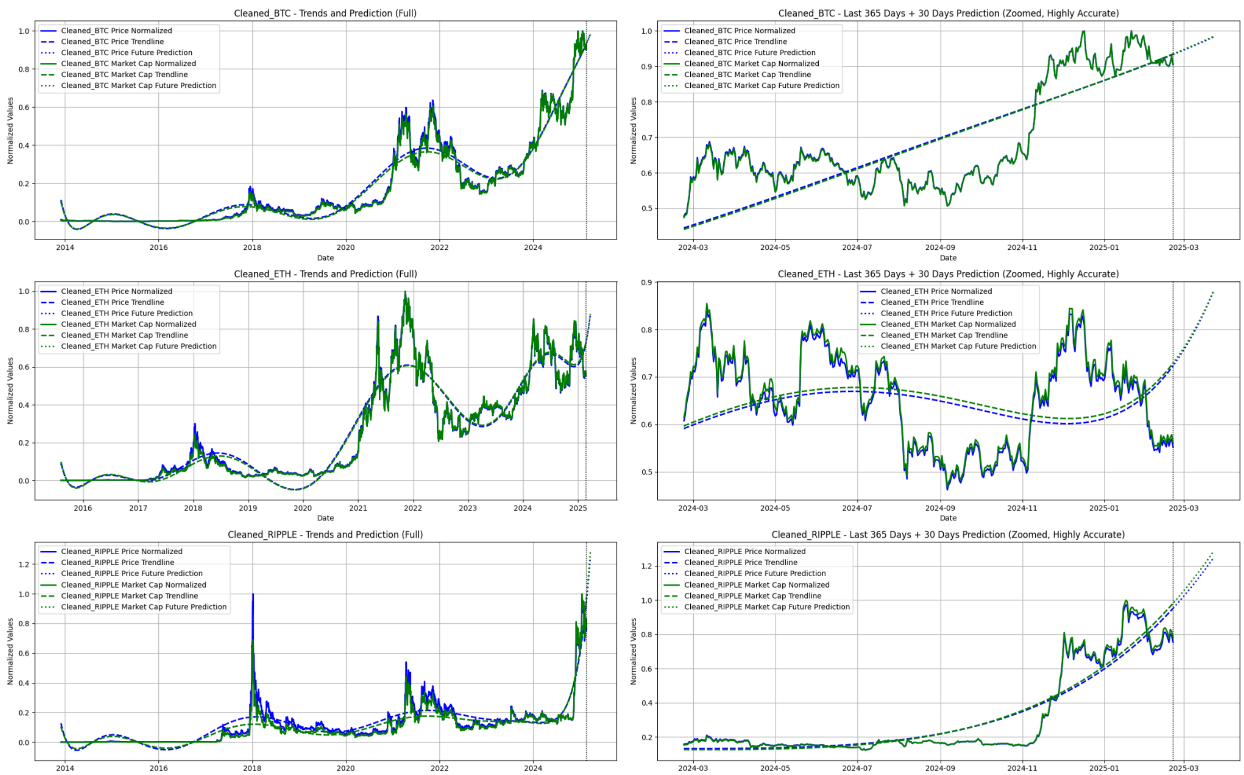
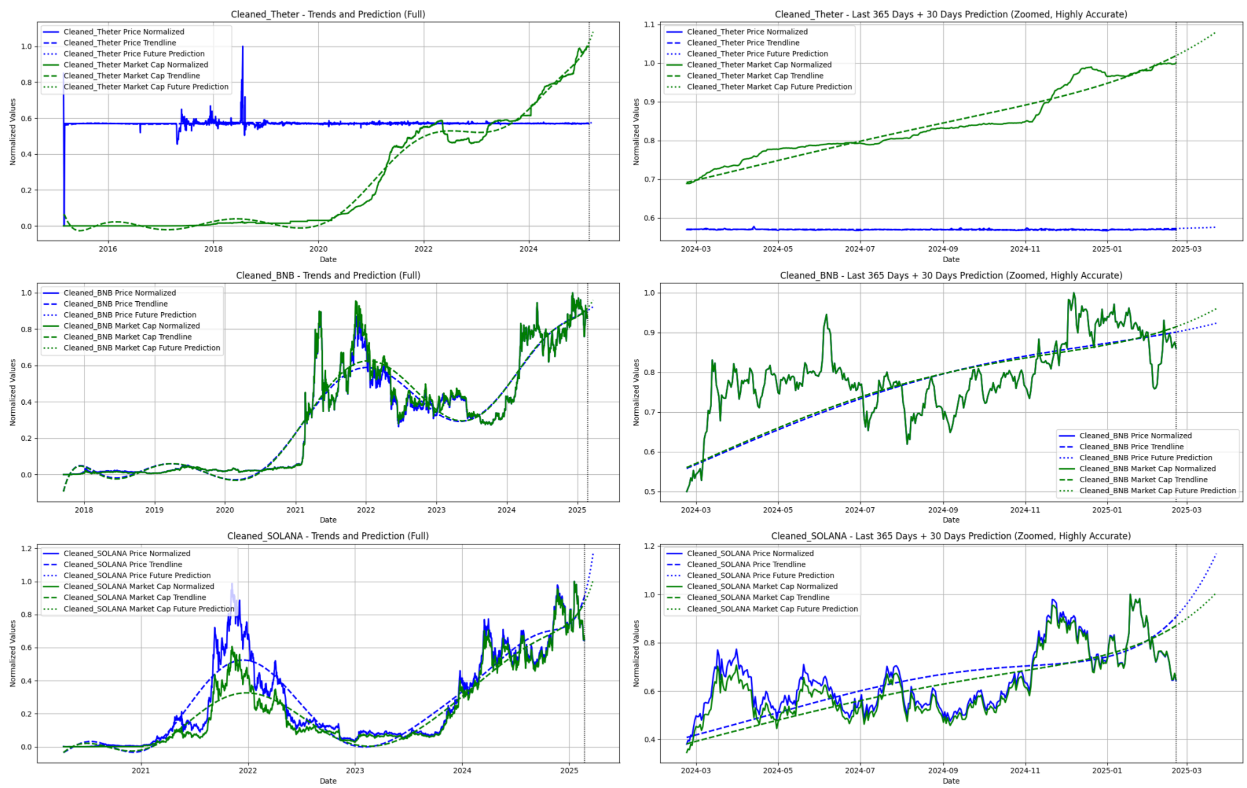


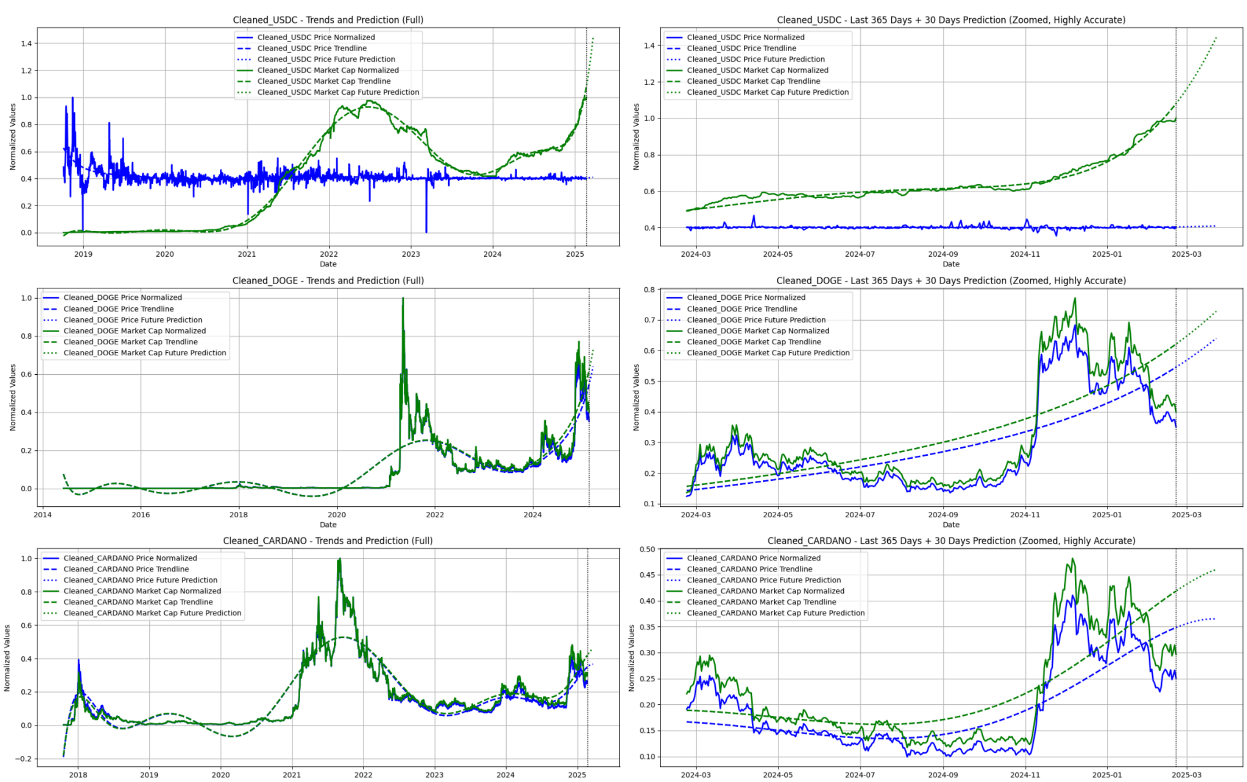
Figure 22: Linear Regression Predictions (Full View + Zoomed View)

Prediction Through Ridge Regression

While searching for other methods of modeling, I learned about ridge regression, a variant of linear regression that has built-in regularization that penalizes large coefficients and therefore reduces the risk of overfitting, so that even though I wanted to increase the fit, I wouldn't have to worry too much about overfitting. Ridge regression can improve prediction accuracy and is especially powerful in complex patterns, such as those in cryptocurrency data. The regularization parameter – alpha – also would allow me to intuitively tune the model to optimize it. Thus, I recreated the previous model with the implementation of ridge regression instead of linear regression, again using the historical data to make predictions 30 days ahead. As demonstrated below in Figure 23, the trendlines created using ridge regression fit the historical data far more accurately than the linear regression model, and the predictions thus appear far more reasonable and likely.




Cleaned_Theter - Trends and Prediction (Full)
Cleaned_Theter Price Normalized
Cleaned_Theter Price Trendline
Cleaned_Theter Price Future Prediction
Cleaned_Theter Market Cap Normalized
Cleaned_Theter Market Cap Trendline
Cleaned_Theter Market Cap Future Prediction
Normalized Values
Date
Cleaned_Theter - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)
Cleaned_BNB - Trends and Prediction (Full)
Cleaned_BNB Price Normalized
Cleaned_BNB Price Trendline
Cleaned_BNB Price Future Prediction
Cleaned_BNB Market Cap Normalized
Cleaned_BNB Market Cap Trendline
Cleaned_BNB Market Cap Future Prediction
Cleaned_BNB - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)
Cleaned_SOLANA - Trends and Prediction (Full)
Cleaned_SOLANA Price Normalized
Cleaned_SOLANA Price Trendline
Cleaned_SOLANA Price Future Prediction
Cleaned_SOLANA Market Cap Normalized
Cleaned_SOLANA Market Cap Trendline
Cleaned_SOLANA Market Cap Future Prediction
Cleaned_SOLANA - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)


Cleaned_USDC - Trends and Prediction (Full)
Cleaned_USDC Price Normalized
Cleaned_USDC Price Trendline
Cleaned_USDC Price Future Prediction
Cleaned_USDC Market Cap Normalized
Cleaned_USDC Market Cap Trendline
Cleaned_USDC Market Cap Future Prediction
Normalized Values
Date
Cleaned_USDC - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)
Cleaned_DOGE - Trends and Prediction (Full)
Cleaned_DOGE Price Normalized
Cleaned_DOGE Price Trendline
Cleaned_DOGE Price Future Prediction
Cleaned_DOGE Market Cap Normalized
Cleaned_DOGE Market Cap Trendline
Cleaned_DOGE Market Cap Future Prediction
Cleaned_DOGE - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)
Cleaned_CARDANO - Trends and Prediction (Full)
Cleaned_CARDANO Price Normalized
Cleaned_CARDANO Price Trendline
Cleaned_CARDANO Price Future Prediction
Cleaned_CARDANO Market Cap Normalized
Cleaned_CARDANO Market Cap Trendline
Cleaned_CARDANO Market Cap Future Prediction
Cleaned_CARDANO - Last 365 Days + 30 Days Prediction (Zoomed, Highly Accurate)

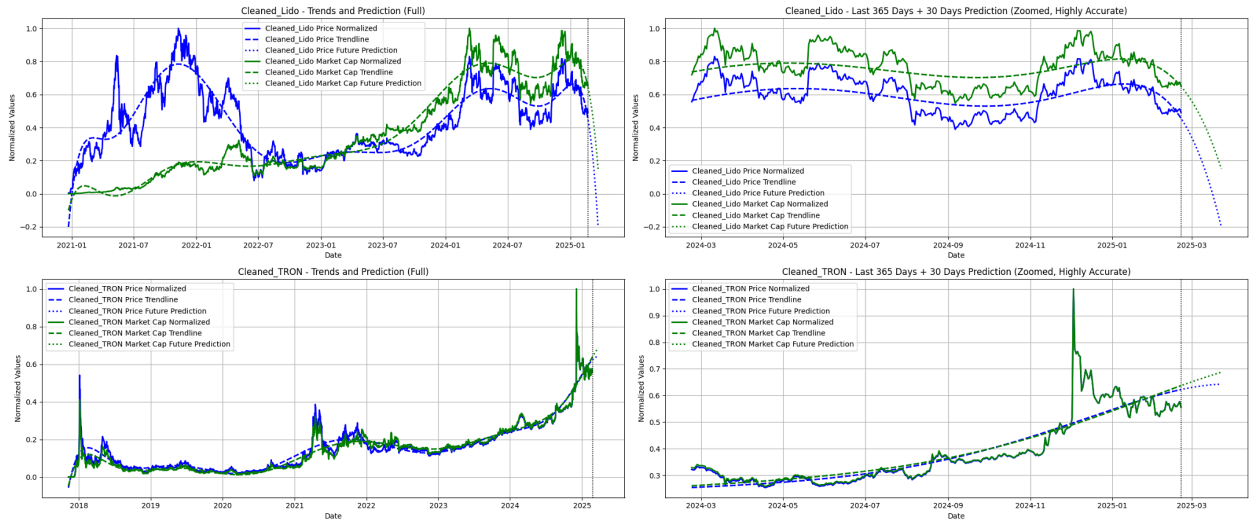


Figure 23: Ridge Regression Predictions (Full View + Zoomed View)

Some particular cryptocurrencies benefited heavily from the ridge regression approach to modeling. For example, BTC and ETH – modeled poorly by the linear regression model – were modeled accurately using the ridge regression model, featuring a highly accurate trendline. Unfortunately, the 30-day predictions could not be properly tested by evaluating the actual cryptocurrency values, as recent news of US tariffs has rocked the cryptocurrency market and rattled the risk appetite of typical traders and consumers. This has impacted all cryptocurrencies in an entirely unpredictable fashion, at least when the data used for modeling is derived only from historical data, without any data involving current events. In fact, the only accurate predictions were for Lido, which was predicted to go down in price (shown in Figure 24 below), and Theter and USDC, which were predicted to remain constant in price as they follow the US dollar. This demonstrates that proper testing of the ridge regression-based prediction model will only be possible later in the future, though its trendline tracking is promising.

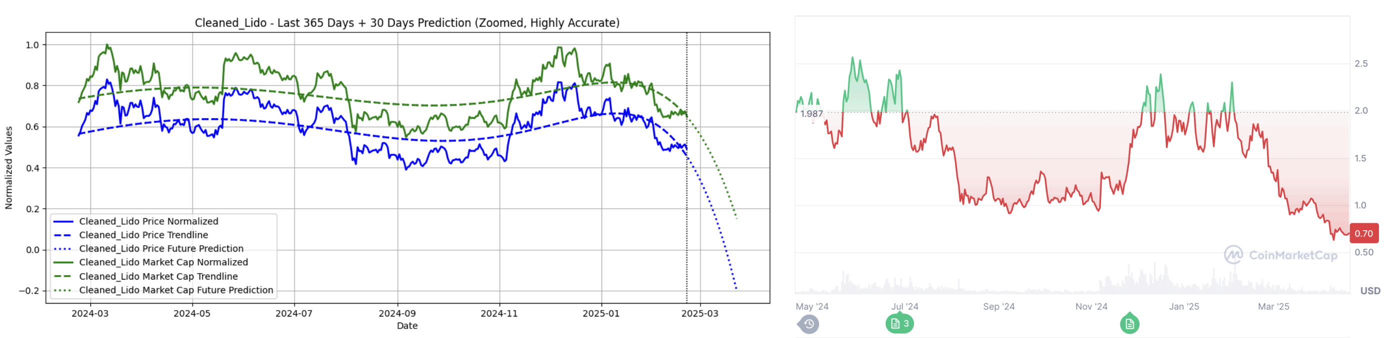


Figure 24: Ridge Regression Prediction for Lido & Actual Price on CoinMarketCap (Yearly View)

Conclusion and Final Interpretation of Results

The project was comprised of several critical parts. The initial data cleaning and transformation allowed me to optimize the model-building process and create impactful data analysis. Particularly, the normalization of the data allowed for more productive pattern and trend identification because it removed the impact of varying value ranges on the overall analysis of the data. The normalization was likely the most crucial aspect of the first section of the project.

Following the data cleaning, transformation, and normalization was the visualization and statistical pattern recognition, in which I used the high-quality data set I created to present the data in effective formatting that allowed for visual observations of trends and patterns that would later be the foundation of the more precise and program-based data analysis. The heatmaps were especially effective in demonstrating correlations across the data, both between two categories (Price and Market Capitalization) of the same cryptocurrency and between the Price and Market Capitalization of two separate cryptocurrencies. Furthermore, the analysis of the cryptocurrency data through the use of trendlines and volatility lines was an important aspect of preparing the data for prediction-making, as I learned about the relevance of polynomial degree application in tracking the historical behavior and movements of the cryptocurrency data and was able to apply this knowledge to creating better fitting polynomial functions of the data.

The final part of the project regarded the formal trend identification and prediction making. I began with modeling through linear regression, but learned that the method was not capable of creating accurate enough predictions and tracking due to its linear nature. Thus, I used the ridge regression method when developing models to ensure that the trendlines hugged the historical data closely and were able to generate more accurate predictions of future prices of the cryptocurrencies. The ridge regression method also allowed me further control over the model through the implementation of the alpha parameter, which controlled the regularization. Additionally, I based my models on making 30-day predictions to ensure they both blocked out noise and reacted to recent events effectively.

Cryptocurrencies are heavily influenced by external factors such as news, regulations, and current events. These factors make it incredibly difficult to predict cryptocurrency with only data on historical behavior, as often, there are many more factors at play. Volatility and market sentiment can always change quickly and unpredictably, and making very short or very long-term predictions is always uncertain. However, by leveraging historical data as well as unstructured data such as Twitter/X posts, news articles, and other forms of relevant and current information, powerful cryptocurrency prediction models can be built and implemented to effectively predict future movements.

References

Kessel, Andrew. “Bitcoin, Crypto Stocks Fall as Tariff Fears Grip Markets.” *Investopedia*, 7 Apr. 2025, https://www.investopedia.com/bitcoin-crypto-stocks-fall-as-tariff-fears-grip-markets-11710116.

"Lido DAO Price Today, LDO to USD Live Price, Marketcap and Chart." *CoinMarketCap*, https://coinmarketcap.com/currencies/lido-dao/. Accessed 19 Apr. 2025.

Appendices

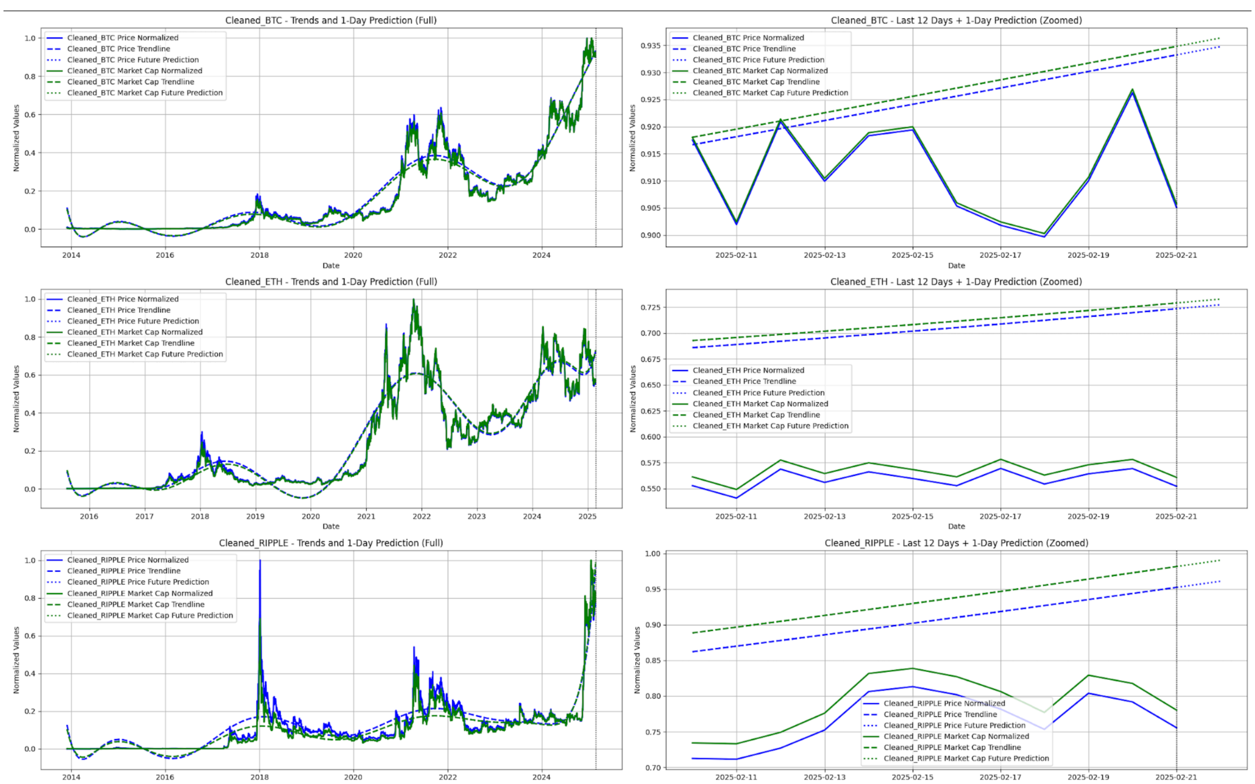
Code

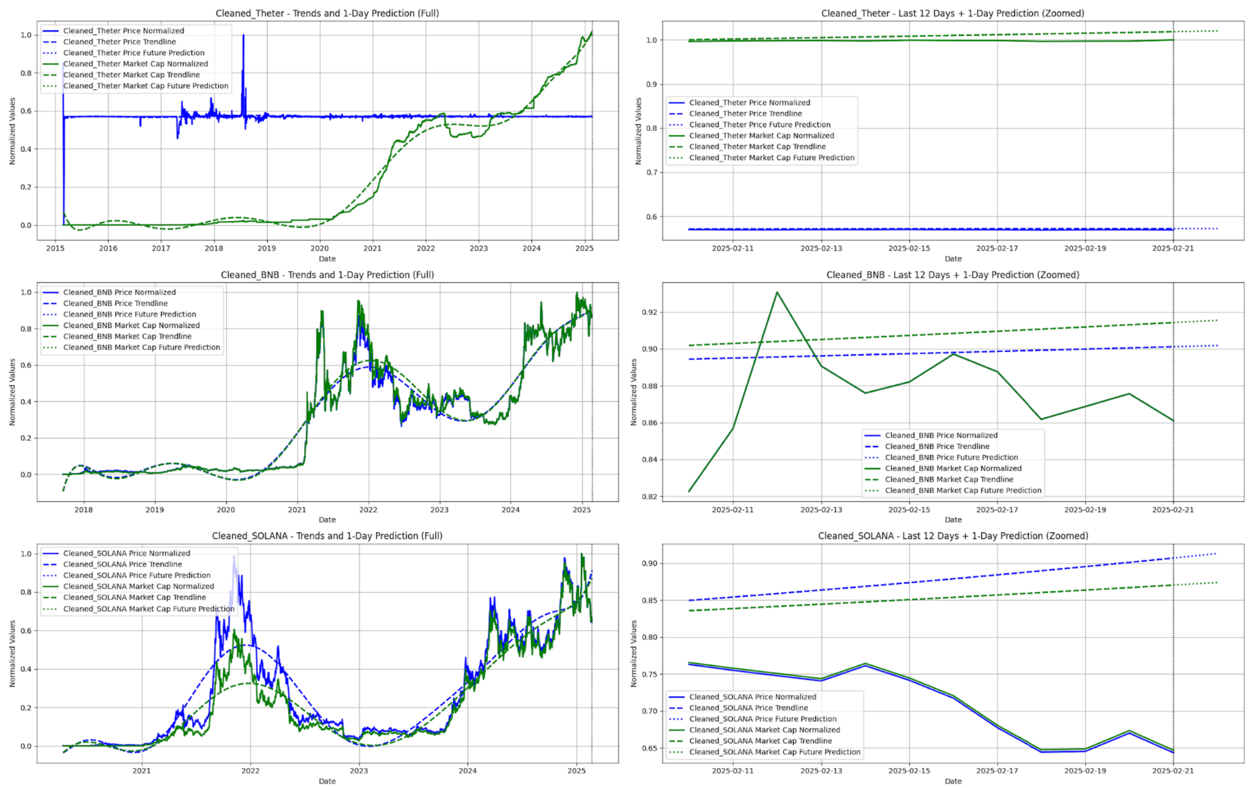
The Jupyter Notebook file can be accessed through my GitHub at:

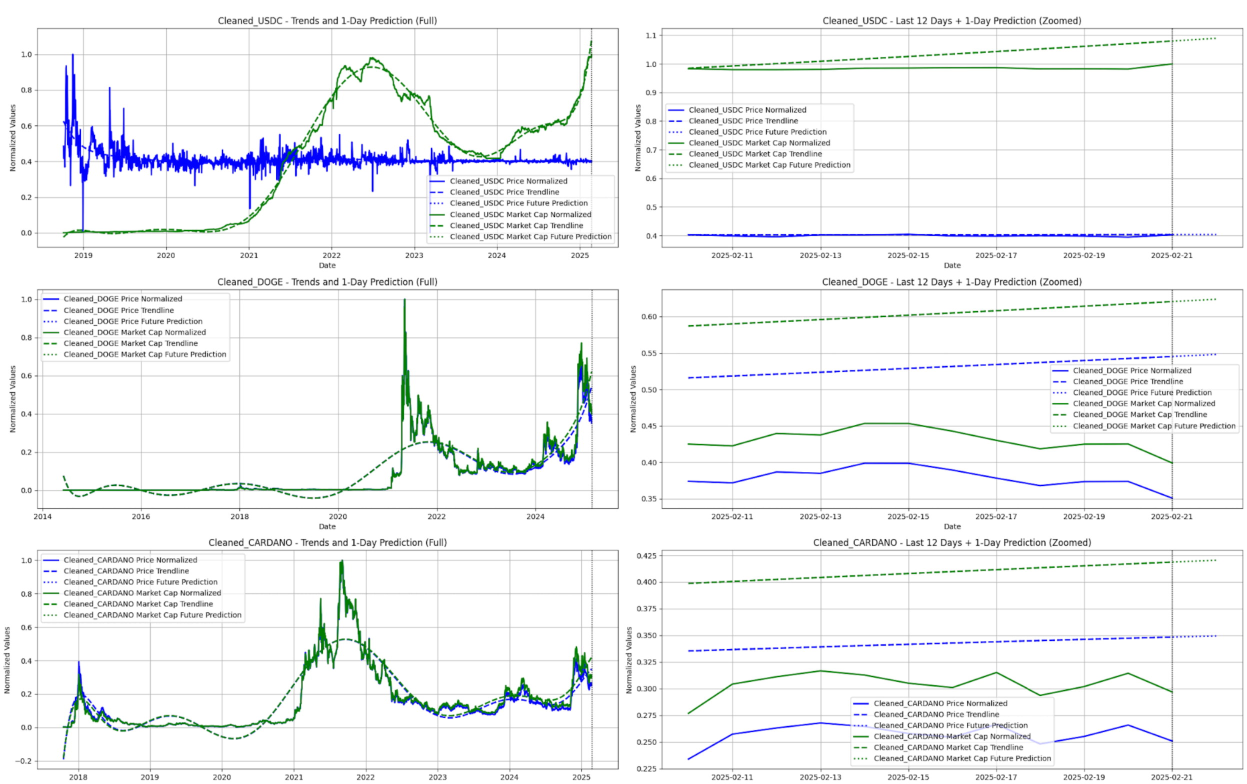
https://github.com/adamyanai/Cryptocurrency-Pattern-Analysis

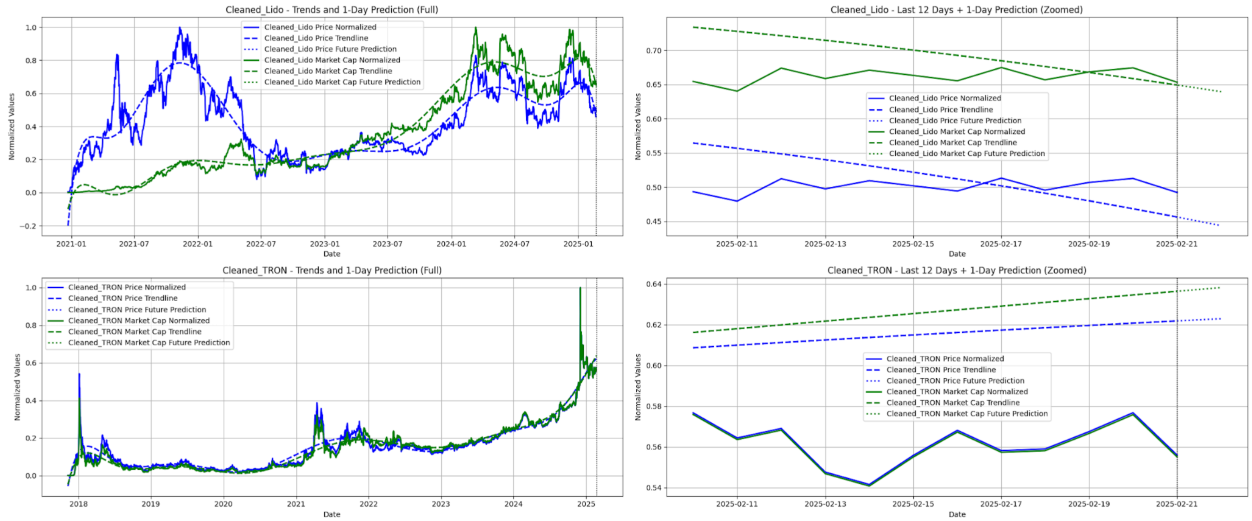
Additional Graphs and Visuals

Day-Based Ridge Regression Prediction Model

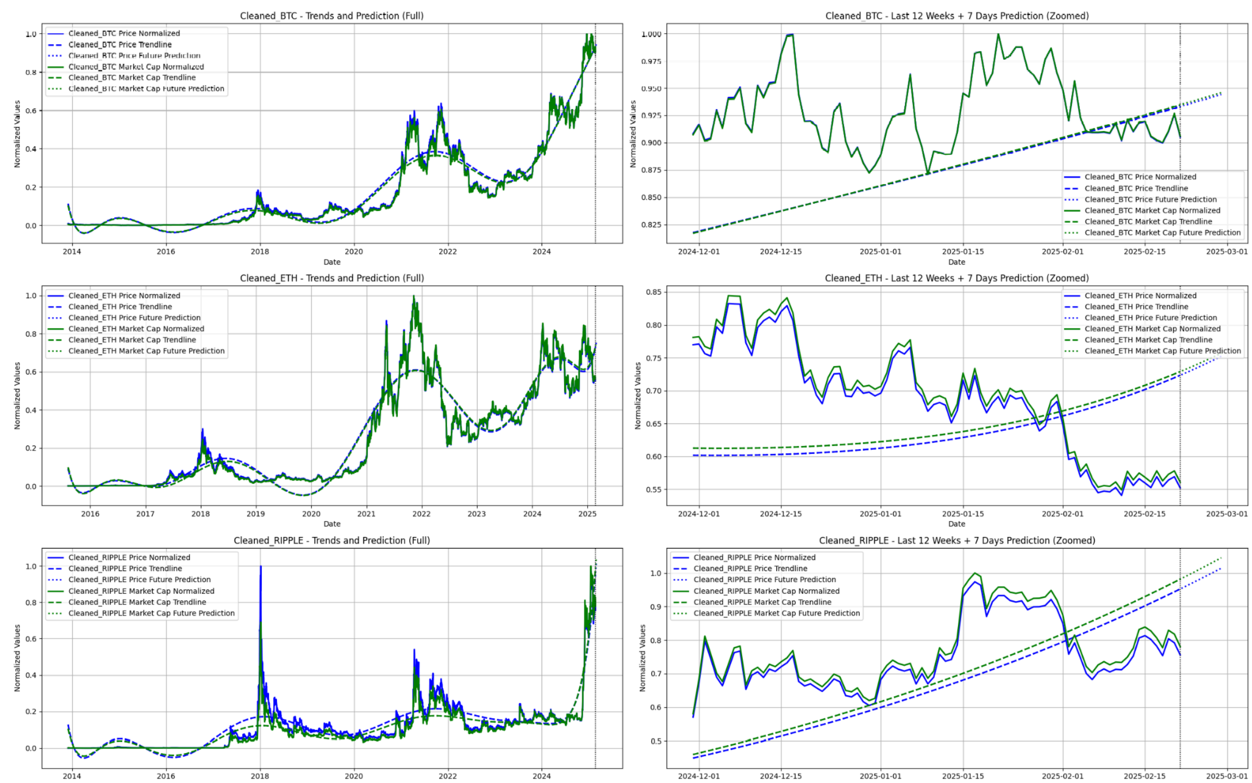


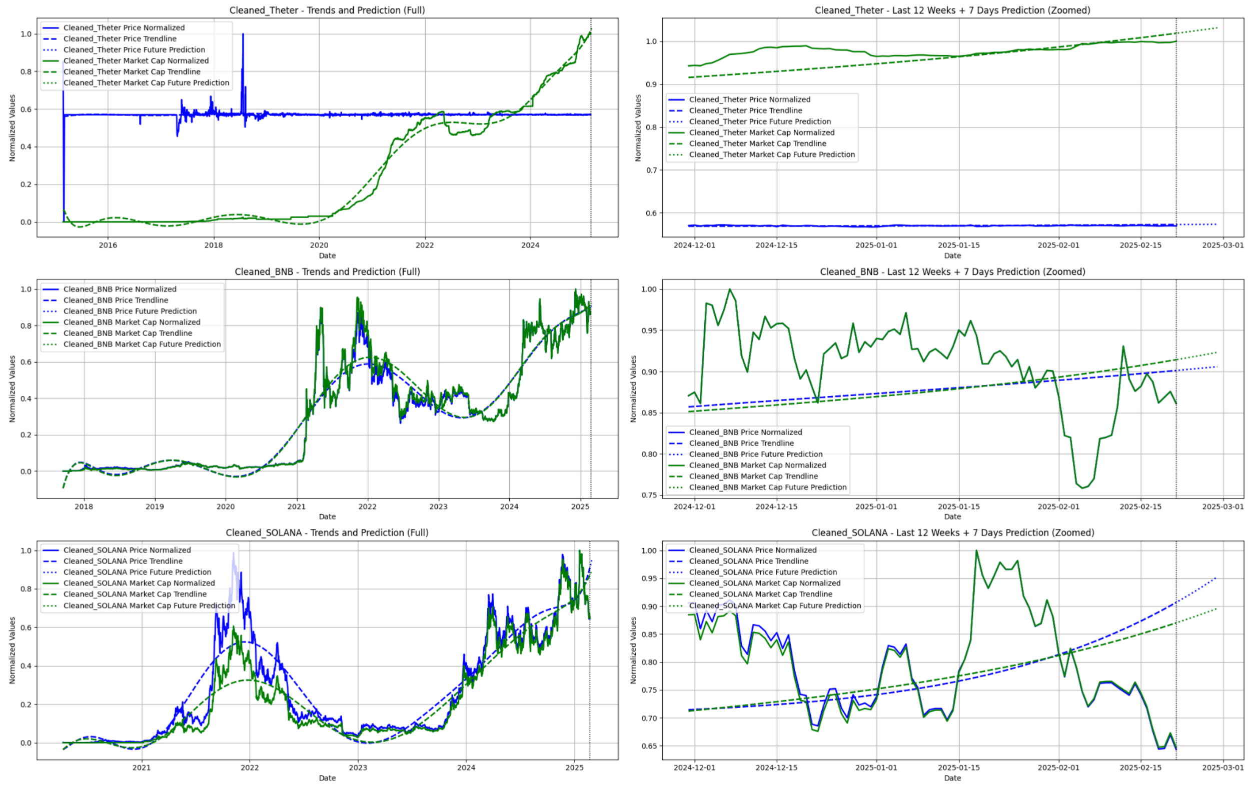

Cleaned_Theter - Trends and 1-Day Prediction (Full)
Cleaned_Theter Price Normalized
Cleaned_Theter Price Trendline
Cleaned_Theter Price Future Prediction
Cleaned_Theter Market Cap Normalized
Cleaned_Theter Market Cap Trendline
Cleaned_Theter Market Cap Future Prediction
Normalized Values
Date
Cleaned_Theter - Last 12 Days + 1-Day Prediction (Zoomed)
Cleaned_BNB - Trends and 1-Day Prediction (Full)
Cleaned_BNB Price Normalized
Cleaned_BNB Price Trendline
Cleaned_BNB Price Future Prediction
Cleaned_BNB Market Cap Normalized
Cleaned_BNB Market Cap Trendline
Cleaned_BNB Market Cap Future Prediction
Cleaned_BNB - Last 12 Days + 1-Day Prediction (Zoomed)
Cleaned_SOLANA - Trends and 1-Day Prediction (Full)
Cleaned_SOLANA Price Normalized
Cleaned_SOLANA Price Trendline
Cleaned_SOLANA Price Future Prediction
Cleaned_SOLANA Market Cap Normalized
Cleaned_SOLANA Market Cap Trendline
Cleaned_SOLANA Market Cap Future Prediction
Cleaned_SOLANA - Last 12 Days + 1-Day Prediction (Zoomed)

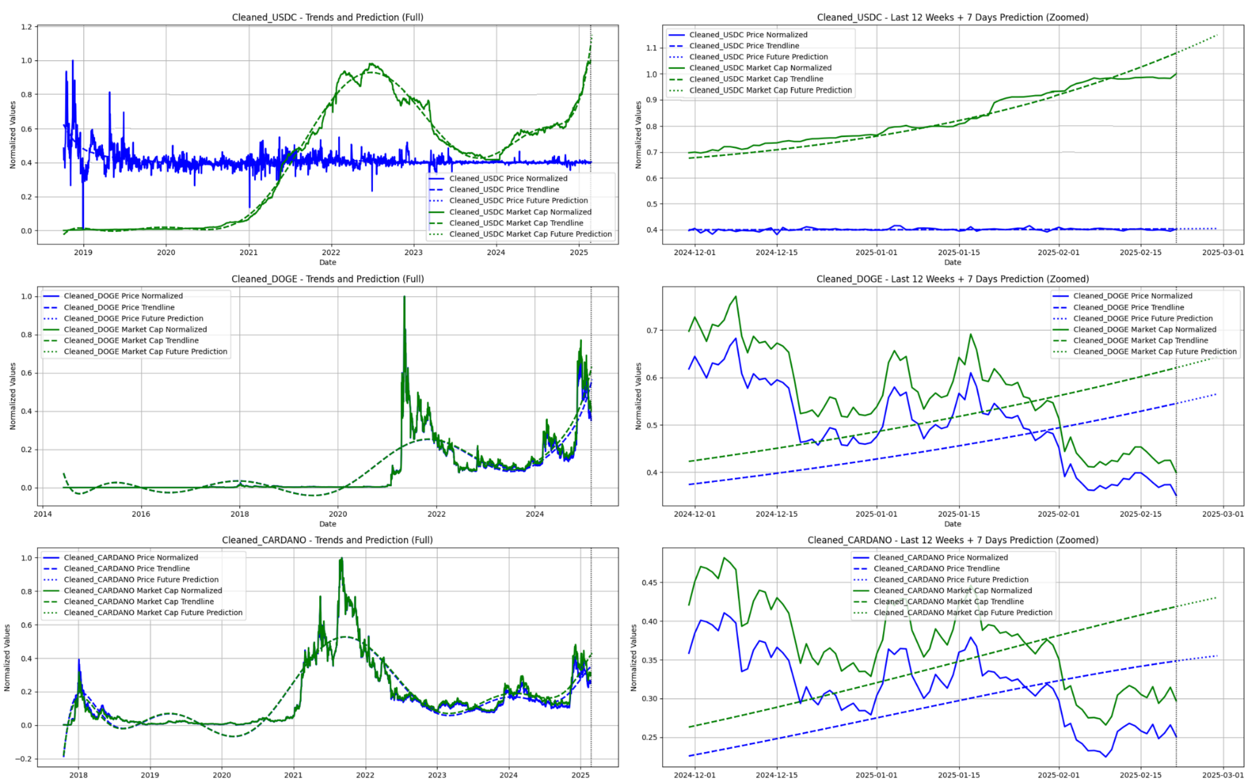

Cleaned_USDC - Trends and 1-Day Prediction (Full)
Cleaned_USDC Price Normalized
Cleaned_USDC Price Trendline
Cleaned_USDC Price Future Prediction
Cleaned_USDC Market Cap Normalized
Cleaned_USDC Market Cap Trendline
Cleaned_USDC Market Cap Future Prediction
Normalized Values
Date
Cleaned_USDC - Last 12 Days + 1-Day Prediction (Zoomed)
Cleaned_DOGE - Trends and 1-Day Prediction (Full)
Cleaned_DOGE Price Normalized
Cleaned_DOGE Price Trendline
Cleaned_DOGE Price Future Prediction
Cleaned_DOGE Market Cap Normalized
Cleaned_DOGE Market Cap Trendline
Cleaned_DOGE Market Cap Future Prediction
Cleaned_DOGE - Last 12 Days + 1-Day Prediction (Zoomed)
Cleaned_CARDANO - Trends and 1-Day Prediction (Full)
Cleaned_CARDANO Price Normalized
Cleaned_CARDANO Price Trendline
Cleaned_CARDANO Price Future Prediction
Cleaned_CARDANO Market Cap Normalized
Cleaned_CARDANO Market Cap Trendline
Cleaned_CARDANO Market Cap Future Prediction
Cleaned_CARDANO - Last 12 Days + 1-Day Prediction (Zoomed)

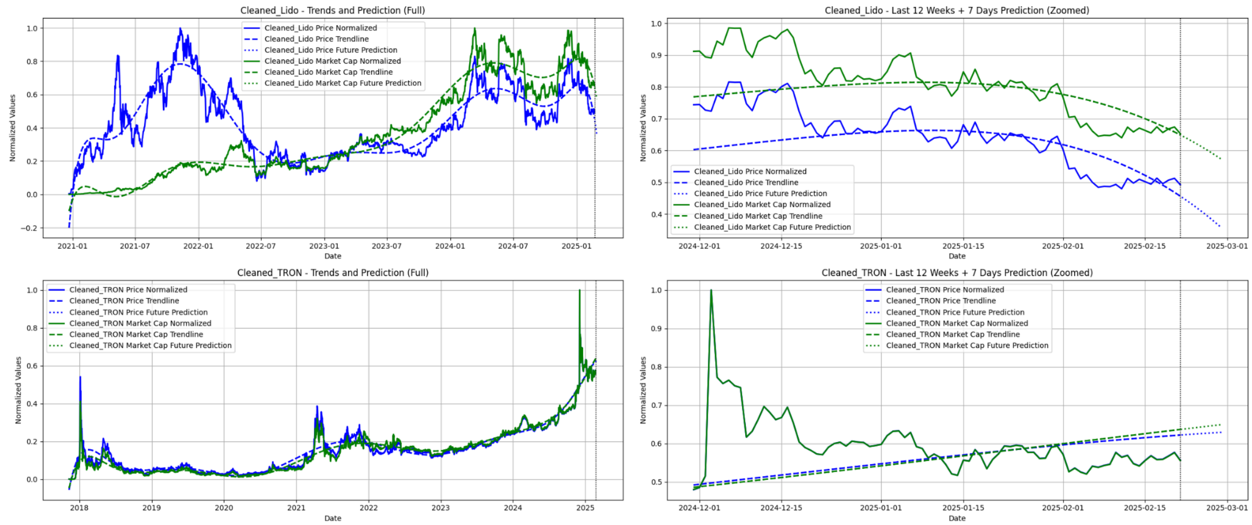


Week-Based Ridge Regression Prediction Model

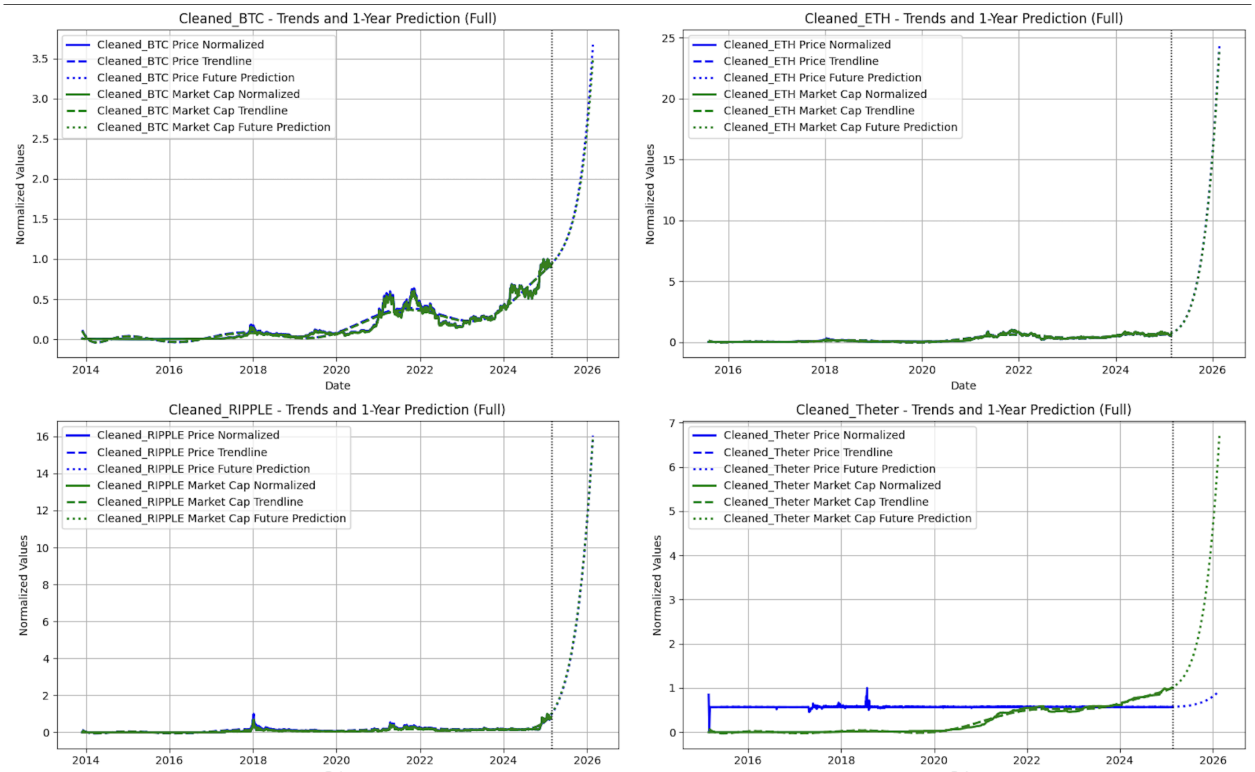


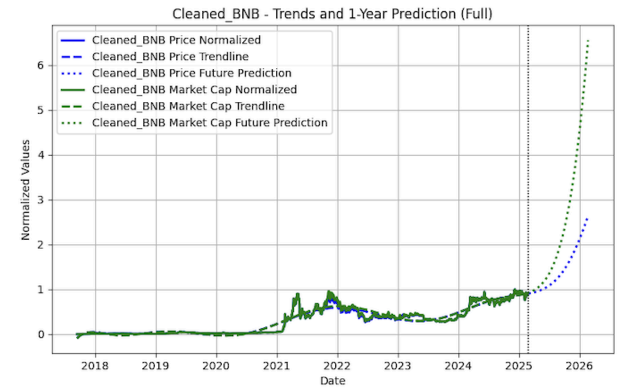

Cleaned_Theter - Trends and Prediction (Full)
Cleaned_Theter Price Normalized
Cleaned_Theter Price Trendline
Cleaned_Theter Price Future Prediction
Cleaned_Theter Market Cap Normalized
Cleaned_Theter Market Cap Trendline
Cleaned_Theter Market Cap Future Prediction
Normalized Values
Date
Cleaned_Theter - Last 12 Weeks + 7 Days Prediction (Zoomed)
Cleaned_BNB - Trends and Prediction (Full)
Cleaned_BNB Price Normalized
Cleaned_BNB Price Trendline
Cleaned_BNB Price Future Prediction
Cleaned_BNB Market Cap Normalized
Cleaned_BNB Market Cap Trendline
Cleaned_BNB Market Cap Future Prediction
Cleaned_BNB - Last 12 Weeks + 7 Days Prediction (Zoomed)
Cleaned_SOLANA - Trends and Prediction (Full)
Cleaned_SOLANA Price Normalized
Cleaned_SOLANA Price Trendline
Cleaned_SOLANA Price Future Prediction
Cleaned_SOLANA Market Cap Normalized
Cleaned_SOLANA Market Cap Trendline
Cleaned_SOLANA Market Cap Future Prediction
Cleaned_SOLANA - Last 12 Weeks + 7 Days Prediction (Zoomed)

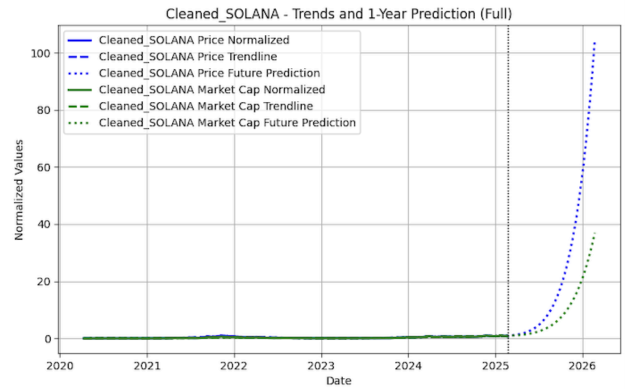

Cleaned_USDC - Trends and Prediction (Full)
Cleaned_USDC Price Normalized
Cleaned_USDC Price Trendline
Cleaned_USDC Price Future Prediction
Cleaned_USDC Market Cap Normalized
Cleaned_USDC Market Cap Trendline
Cleaned_USDC Market Cap Future Prediction
Normalized Values
Date
Cleaned_USDC - Last 12 Weeks + 7 Days Prediction (Zoomed)
Cleaned_DOGE - Trends and Prediction (Full)
Cleaned_DOGE Price Normalized
Cleaned_DOGE Price Trendline
Cleaned_DOGE Price Future Prediction
Cleaned_DOGE Market Cap Normalized
Cleaned_DOGE Market Cap Trendline
Cleaned_DOGE Market Cap Future Prediction
Cleaned_DOGE - Last 12 Weeks + 7 Days Prediction (Zoomed)
Cleaned_CARDANO - Trends and Prediction (Full)
Cleaned_CARDANO Price Normalized
Cleaned_CARDANO Price Trendline
Cleaned_CARDANO Price Future Prediction
Cleaned_CARDANO Market Cap Normalized
Cleaned_CARDANO Market Cap Trendline
Cleaned_CARDANO Market Cap Future Prediction
Cleaned_CARDANO - Last 12 Weeks + 7 Days Prediction (Zoomed)

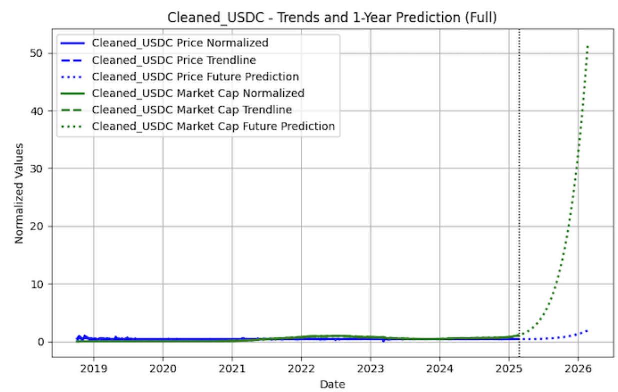


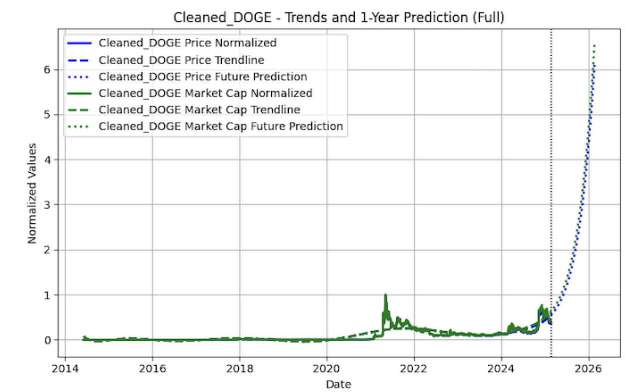
Year-Based Ridge Regression Prediction Model

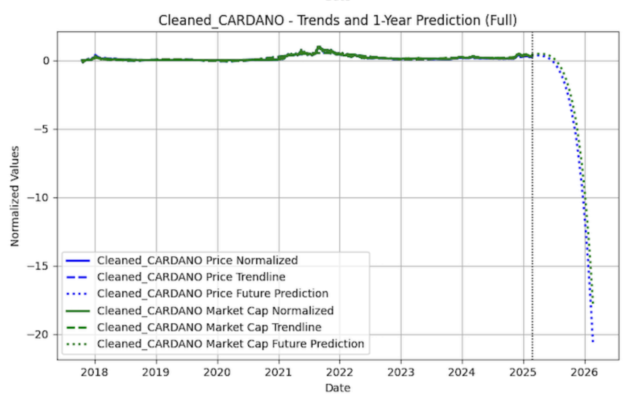



Cleaned_BNB - Trends and 1-Year Prediction (Full)
Cleaned_BNB Price Normalized
Cleaned_BNB Price Trendline
Cleaned_BNB Price Future Prediction
Cleaned_BNB Market Cap Normalized
Cleaned_BNB Market Cap Trendline
Cleaned_BNB Market Cap Future Prediction
Normalized Values
0
1
2
3
4
5
6
2018
2019
2020
2021
2022
2023
2024
2025
2026
Date

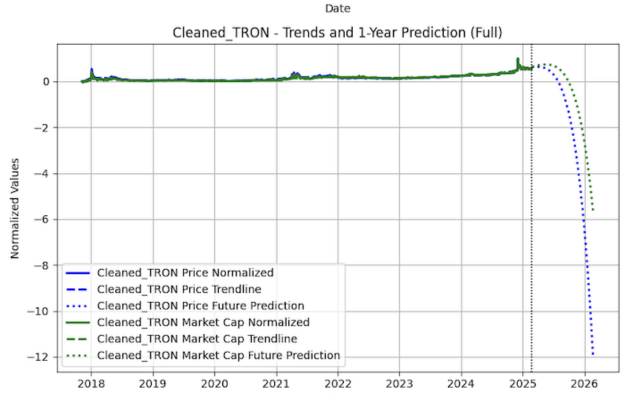

Cleaned_SOLANA - Trends and 1-Year Prediction (Full)
Cleaned_SOLANA Price Normalized
Cleaned_SOLANA Price Trendline
Cleaned_SOLANA Price Future Prediction
Cleaned_SOLANA Market Cap Normalized
Cleaned_SOLANA Market Cap Trendline
Cleaned_SOLANA Market Cap Future Prediction
Normalized Values
0
20
40
60
80
100
2020
2021
2022
2023
2024
2025
2026
Date


Cleaned_USDC - Trends and 1-Year Prediction (Full)
Cleaned_USDC Price Normalized
Cleaned_USDC Price Trendline
Cleaned_USDC Price Future Prediction
Cleaned_USDC Market Cap Normalized
Cleaned_USDC Market Cap Trendline
Cleaned_USDC Market Cap Future Prediction
Normalized Values
0
10
20
30
40
50
2019
2020
2021
2022
2023
2024
2025
2026
Date


Cleaned_DOGE - Trends and 1-Year Prediction (Full)
Cleaned_DOGE Price Normalized
Cleaned_DOGE Price Trendline
Cleaned_DOGE Price Future Prediction
Cleaned_DOGE Market Cap Normalized
Cleaned_DOGE Market Cap Trendline
Cleaned_DOGE Market Cap Future Prediction
Normalized Values
0
1
2
3
4
5
6
2014
2016
2018
2020
2022
2024
2026
Date


Cleaned_CARDANO - Trends and 1-Year Prediction (Full)
Cleaned_CARDANO Price Normalized
Cleaned_CARDANO Price Trendline
Cleaned_CARDANO Price Future Prediction
Cleaned_CARDANO Market Cap Normalized
Cleaned_CARDANO Market Cap Trendline
Cleaned_CARDANO Market Cap Future Prediction
Normalized Values
0
−5
−10
−15
−20
2018
2019
2020
2021
2022
2023
2024
2025
2026
Date


Cleaned_Lido - Trends and 1-Year Prediction (Full)
Cleaned_Lido Price Normalized
Cleaned_Lido Price Trendline
Cleaned_Lido Price Future Prediction
Cleaned_Lido Market Cap Normalized
Cleaned_Lido Market Cap Trendline
Cleaned_Lido Market Cap Future Prediction
Normalized Values
0
−100
−200
−300
−400
−500
2021
2022
2023
2024
2025
2026
Date


Cleaned_TRON - Trends and 1-Year Prediction (Full)
Cleaned_TRON Price Normalized
Cleaned_TRON Price Trendline
Cleaned_TRON Price Future Prediction
Cleaned_TRON Market Cap Normalized
Cleaned_TRON Market Cap Trendline
Cleaned_TRON Market Cap Future Prediction
Normalized Values
0
−2
−4
−6
−8
−10
−12
2018
2019
2020
2021
2022
2023
2024
2025
2026
Date