

The Future of Crypto: Google Trends Decomposition Analysis & Forecasting Models

A Submission to Ocean Data Challenge hosted through Desights AI

Dominikus Brian 钟鸿盛 | domi@dreambrook.tech

www.dreambrook.tech

Introduction

The cryptocurrency is not just a new form of value store and exchange, it is a revolution of its own. Beginning with its use to provide peer-to-peer payment network (or digital money) like Bitcoin, today's cryptocurrency, or crypto for short, have evolved way beyond its humble start. Underlying the crypto world, there lies amazing technology called Blockchain. In simple term, blockchain is a decentralized and shared digital ledger that records transactions transparently and immutably across nodes in the network. Today's Crypto community has slowly turned into industry of its own introducing a whole spectrum of enigmatic pattern, trends, and economic framework. In this report we will explore the trend, correlations, and dynamics related to 20 selected Crypto projects to derived insights and build models that predict the future of crypto.

Key Findings:

- (1) Our exploratory data analysis (EDA) underlines the span and general pattern of the Google Trend and Price related data. The data being analyzed span from the earliest entry on 2014-09-17 up to the latest on 2024-04-07.
- (2) Time series decomposition was performed to extract trend, seasonal cycle, and residuals that made up the Google Interest Trend data.
- (3) Analysis on the time-series decomposition help us distinguish cluster (a) with projects on the rise such as Solana, SingularityNet, Fetch.ai, and Ocean Protocol; and cluster (b) containing old project such as Dogecoin, Litecoin Filecoin, Tezos that are facing stagnant/downfall trend.
- (4) Based on the Google Trends's Correlation across projects we characterize Highly correlated projects cluster with correlation of about >0.8 , and up to 0.92 with Bitcoin-Ethereum-Chainlink-Litecoin-Monero as the prominent group members.
- (5) By introducing additional Google Trend data to understand Crypto Narrative, we worked toward building interpretable Event/Entity driving the market sentiment to explain our decomposed Time-series data.
- (6) Based on Lag Characteristics in Correlation of Google Trend and Price/Trade Volume we highlight the tendency for the correlation to accumulate at longer lag time.
- (7) Using NeuralProphet Framework we build forecasting models for Google Trend and Token Price for all 20 projects investigated here. We deployed these models to predict Trend and Price for all 20 projects for the following 52 weeks (up until April 2025). The developed models performed extraordinarily well with the R^2 value for most fall between the range of 0.75-0.88, while the highest goes up to 0.919.
- (8) We highlight the correlation between Bitcoin, Ethereum, Ocean, with the rest of other projects. Ocean and Bitcoin, also Ethereum and Solana are the most correlated, both with correlation value of 0.89. The Kucoin's KCS token is the least correlated with both Ocean and Bitcoin (0.31), while with Ethereum, Filecoin have the least correlation (0.41).

Exploratory Data Analysis

We begin our exploration by first going over the raw data for these 20 projects. Namely, we have a series of Google Interest Trend data covering the worldwide scope, and the prices data for each token. To understand the form and shape of the data we first plot all the data without tampering with its form and shape and are visualized in Fig. 1 for Token Price time-series data and Fig. 2 for Price value histogram, and Fig 3 for the Google Trend time-series. By carefully inspecting these data we can notice not all the data are of the same length and duration. Therefore, we first survey the data based on year, month, and days. The earliest data entry are for Bitcoin-USD pair that dated on 2014-09-17 and the shortest data entry is for ROSE-USD pair, with earliest entry dated on 2020-11-19. Having identified the timeline involved we tabulate all the data in year-to-year basis, the result of which are shown in Fig.4. We now then introduce a time cutoff on 2021-01-01, in which all 20 projects data overlapped. By doing so we can then identify two blocks of dataset, namely the “Aligned Dataset” and “Extra Dataset”. In the Aligned data set, we then have 3 full years (365 days/year) + 100 days of data up until the screenshot of the data that was performed on April 9th 2024. In the investigation and analysis that follows we focused our analysis on the Aligned Dataset, and only include or use data from the “Extra Dataset” for completeness shown in appendices when appropriate.

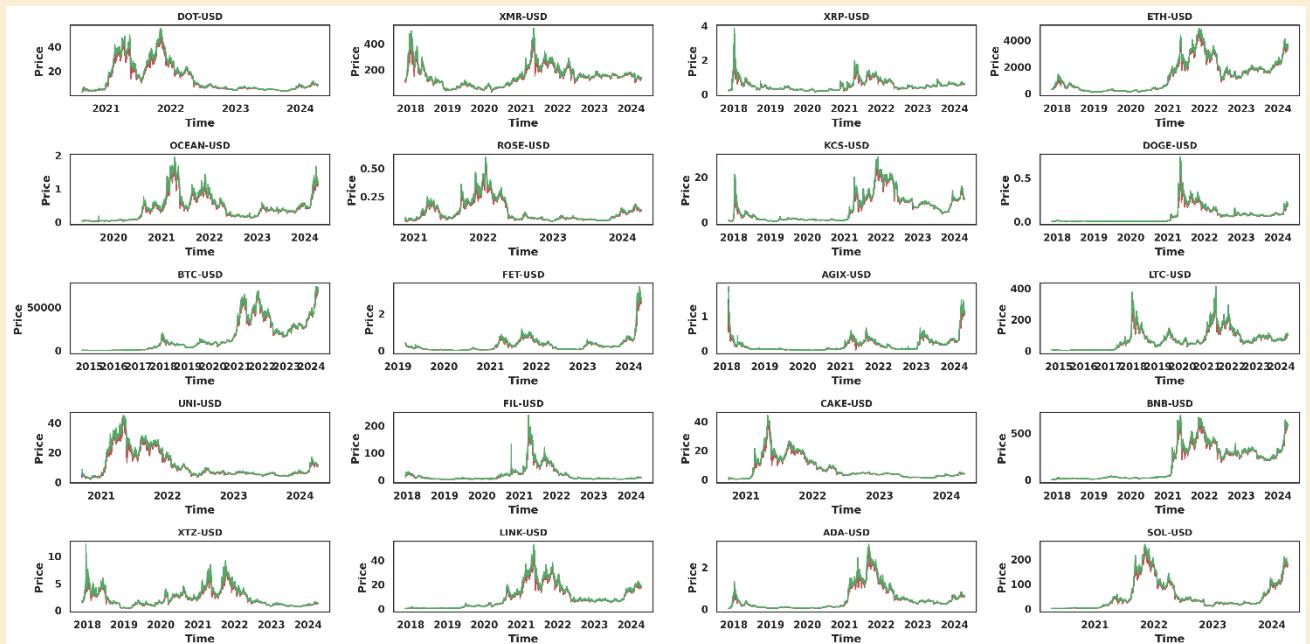


Figure 1 Overview of the Raw Token Price data

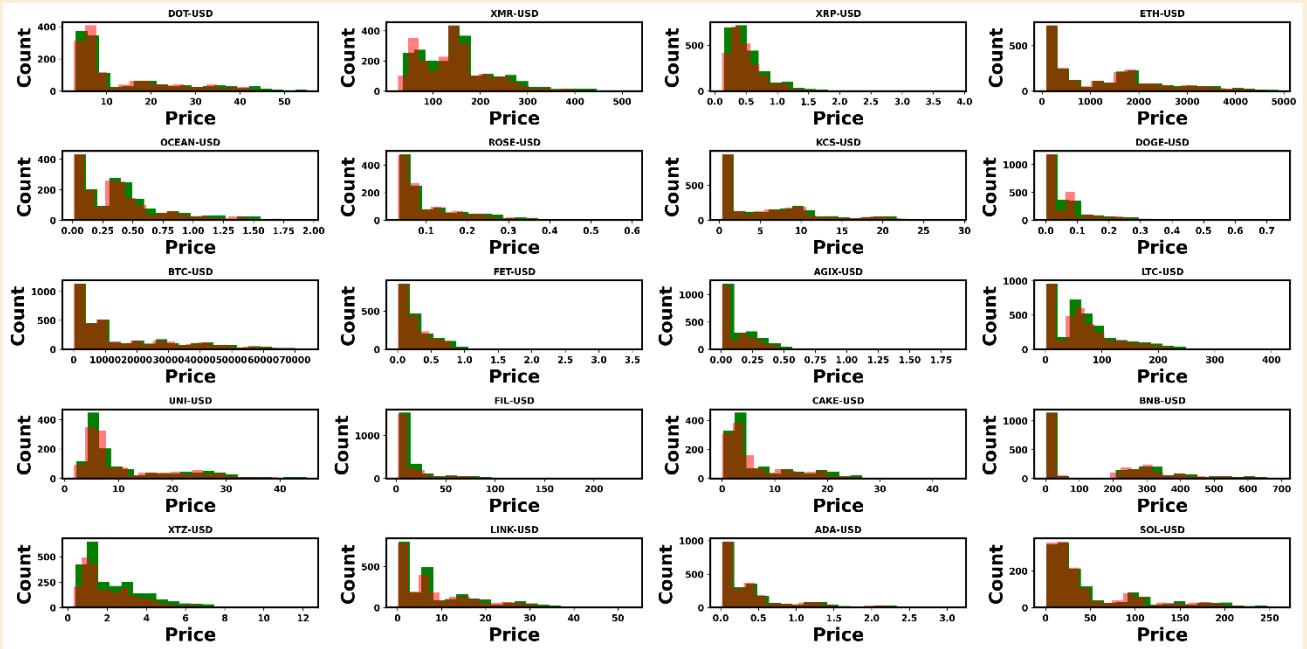


Figure 2 Histogram distribution of the overall price data for daily low (red) and high (green).

Having performed the demarcation of data based on the price data, we now address the form and nature of the data. In the price data (sample are shown in Fig.4) there are several variables of interest, the high, low, close, and open price for the token as well as the trade volume for the token. As shown in Fig1 and also partly in Fig. 5. we can see that for most rows of data entry the differences between this different daily price marker are minimal. For completeness, the full raw-data for trade volume are also given in **Appendix A1**.

In the above we have described that the aligned dataset is defined based on the price data timeline, in order to complement the timing properly, we also produced aligned component of the Google Trend data based on the same timeframe. Here we note that Google Trend data is reported on weekly basis and therefore is spaced by 7 days instead of daily. In order to ensure integrity of the size and form of the data, for cases in which operations were to be performed on both price and trend data, preprocessing of the data to yield the same dataset size. The aligned dataset that contained both Google Trend and price is called "Combined Dataset" and this naming convention is to be used throughout.

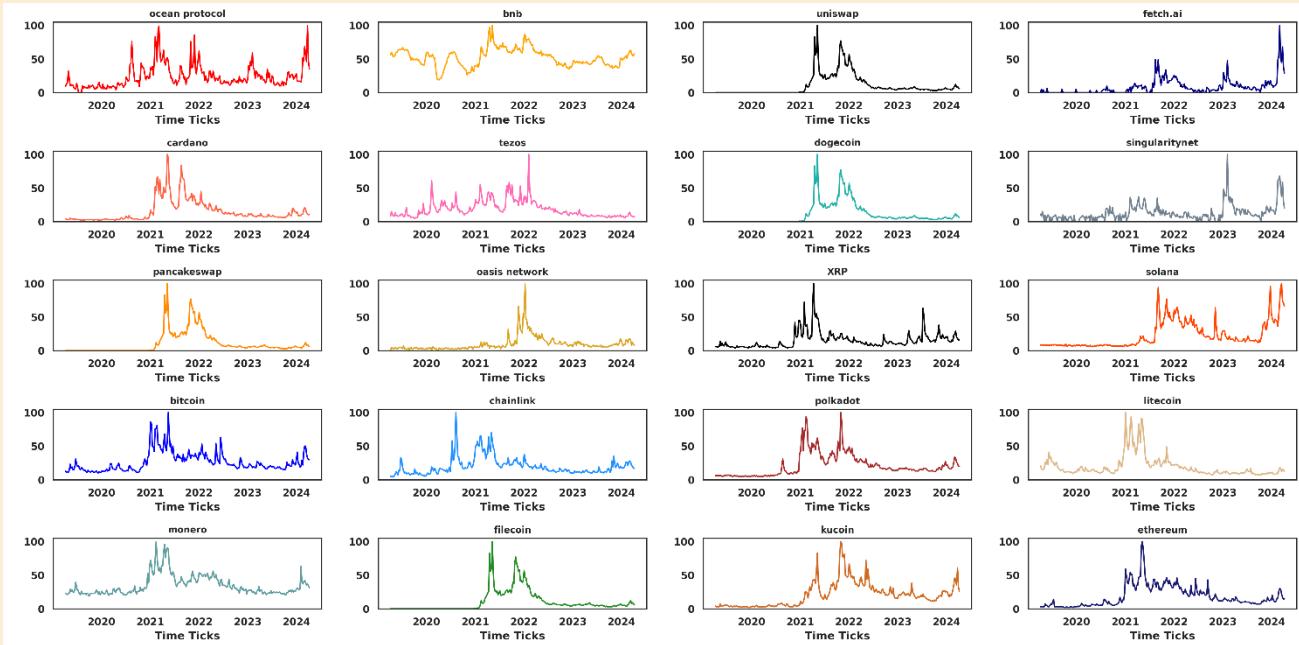


Figure 3 Raw data for the Google Interest Trend

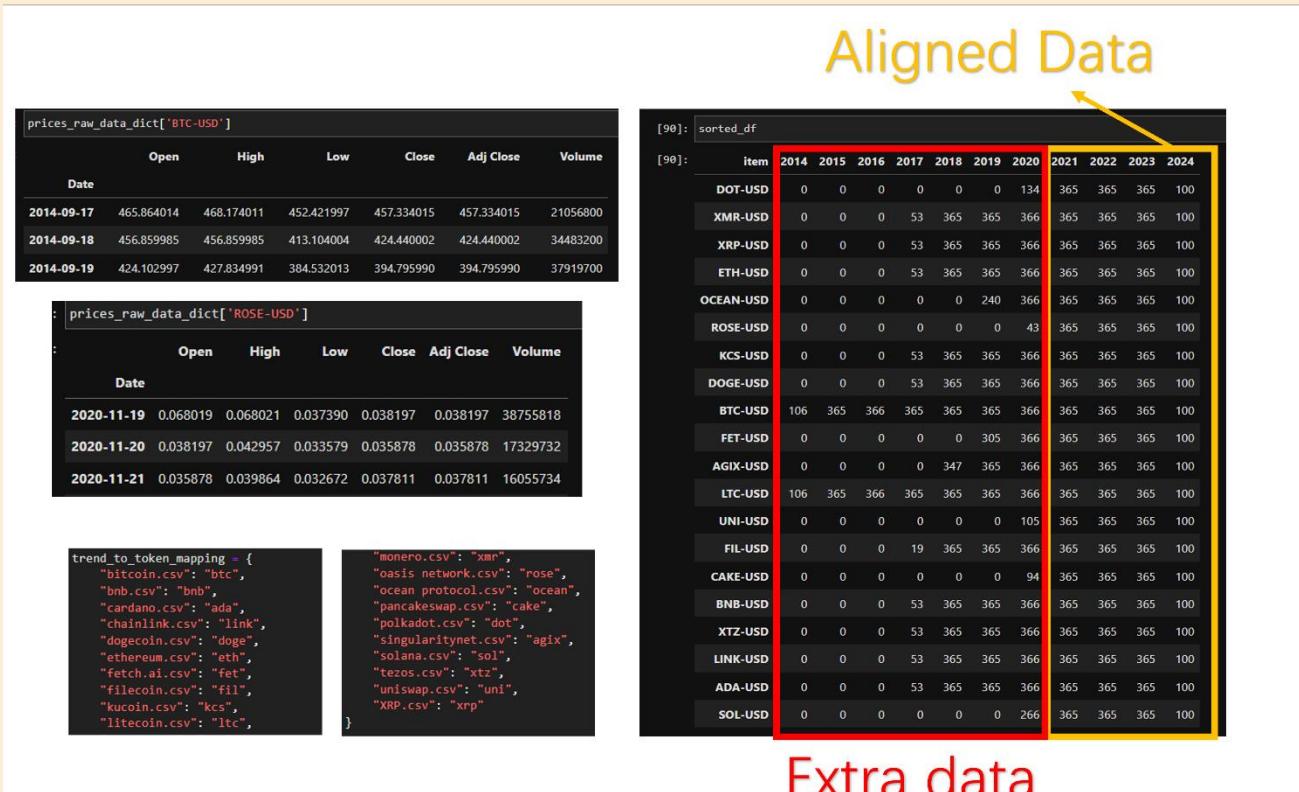


Figure 4 Data Information, mapping of token naming, and demarcation of raw dataset used in this study.

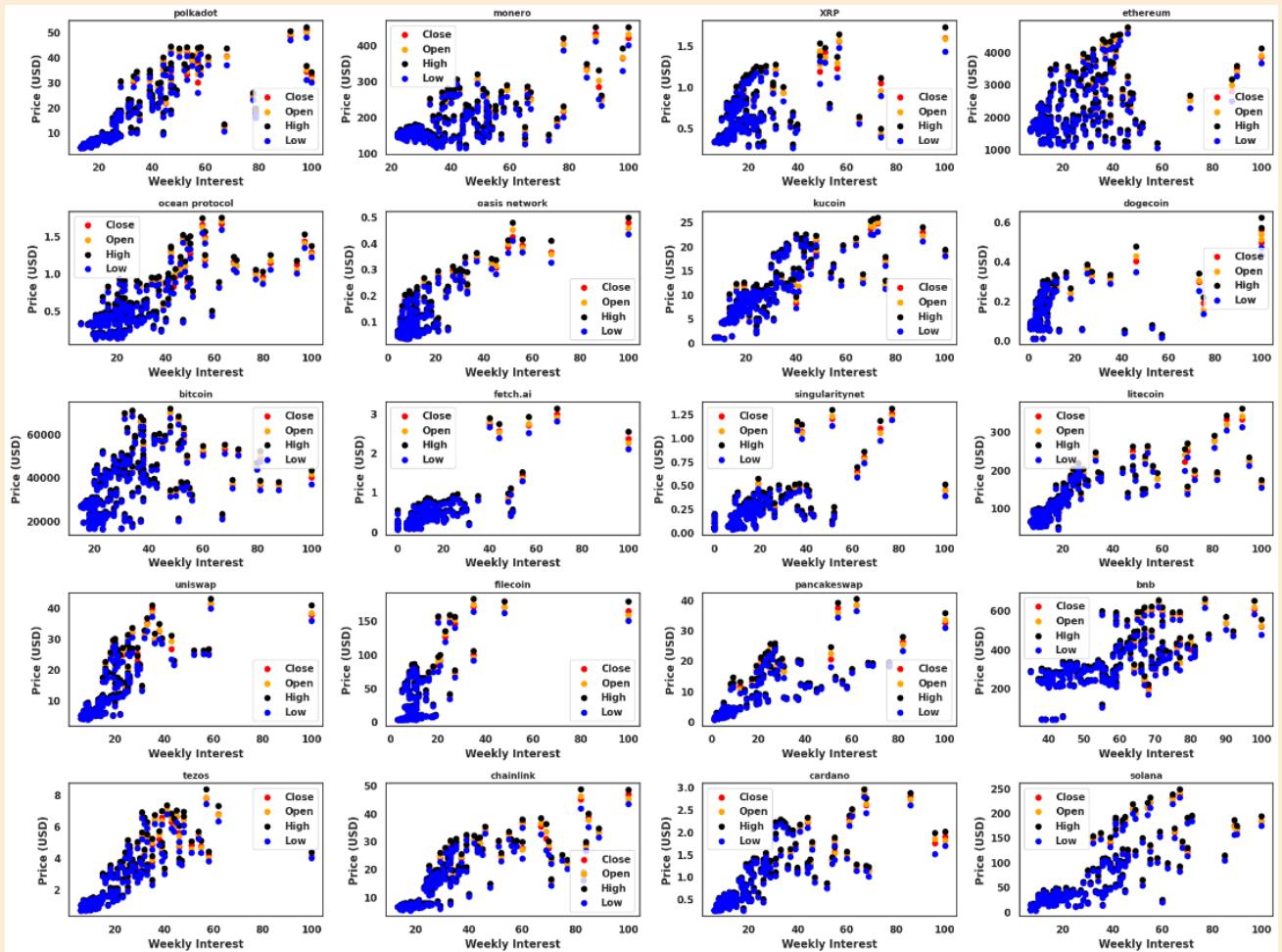


Figure 5. Variability across the Open Close, Low High Weekly average plotted against the Google Trend Interest value

Data Analysis of Past and Current Google Trend Cycle by Means of Time-Series Decomposition

Before moving forward onto the deep-dive for understanding intricate relationship contained in our Google Trend data, let us briefly revisit definitions and meaning of these Google Trend data. The Google Trend data or Google Interest Trend is a “*Numbers represent search interest relative to the highest point on the chart for the given region and time*”, that being said, full value of 100 is to represent peak popularity for the given search term and 50 is less popular, if value 0 is given it represents that there was not enough data for the given term at particular time point. In this study the Google Trend data used are of Worldwide scale and is not dissected further based on region. It is important to note that by nature the reporting of Google Trend data is inherently contain a delay of 3 days or so. So, the reported data for a particular day, are to certain degree based on searches and popularity of the term in about 2-3 days prior. Though in general this doesn’t

affect the investigation and analysis that follow, such clarification is necessary. In fact, later we will note that in order to further improve on the developed model, exploration to refine this inherent delay may be useful.

The Google Trend data is a weekly based data of 172 weeks length, starting from the first week of 2021, and ends at the week of 7th April 2024. We first proceed to present part of the decomposed Google interest trend data. Here the analyses were performed on all 20 crypto projects and shown in 5 part, with 4 project on each. In Fig.6. and Fig.7, the data shown belong to part 1 and part 4, respectively. Result from the remaining part 2, 3, and 5 were discussed in this section, while the full figures were included in **Appendix A2**.

Google Interest Trend Time-Series Decomposition [1/5]

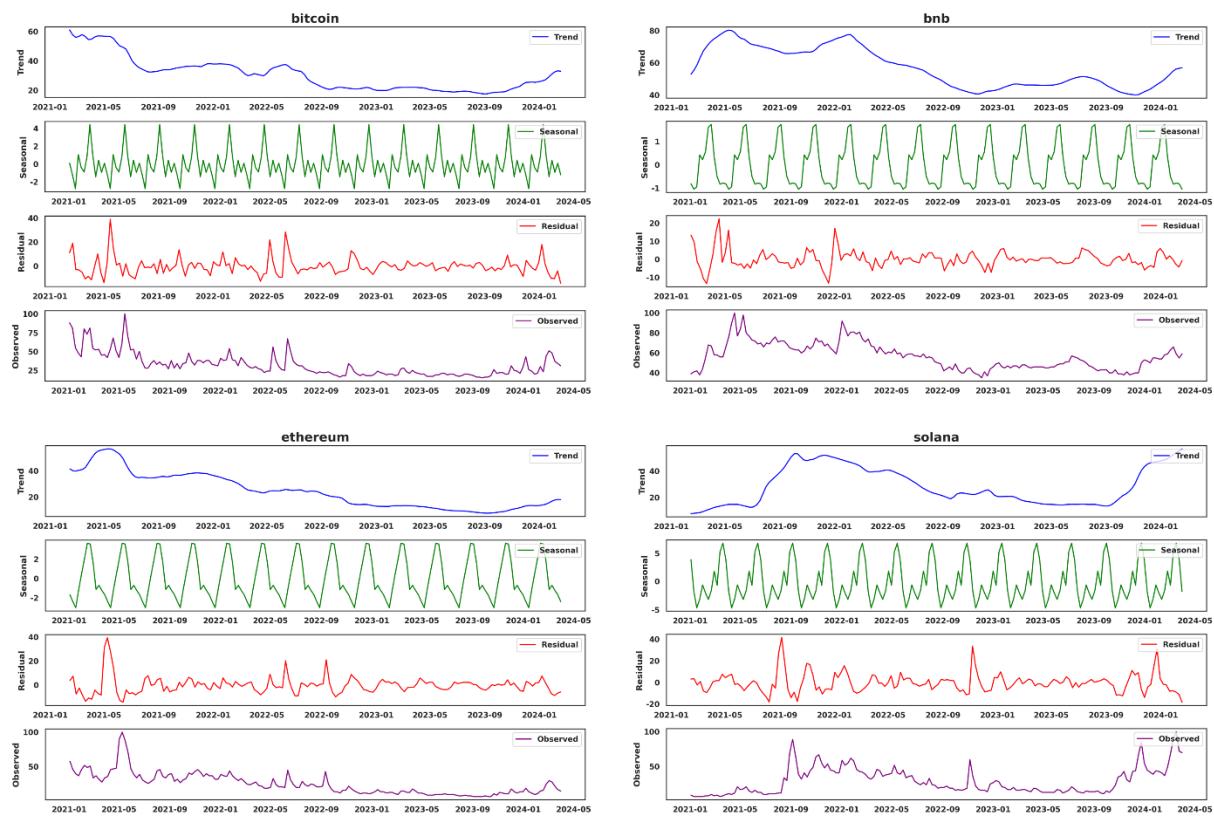


Figure 6. Decomposition of Google Interest Trend data into Trend, Seasonal, Residual and Observed (original data) for Bitcoin, BNB, Ethereum, and Solana.

Google Interest Trend Time-Series Decomposition [4/5]

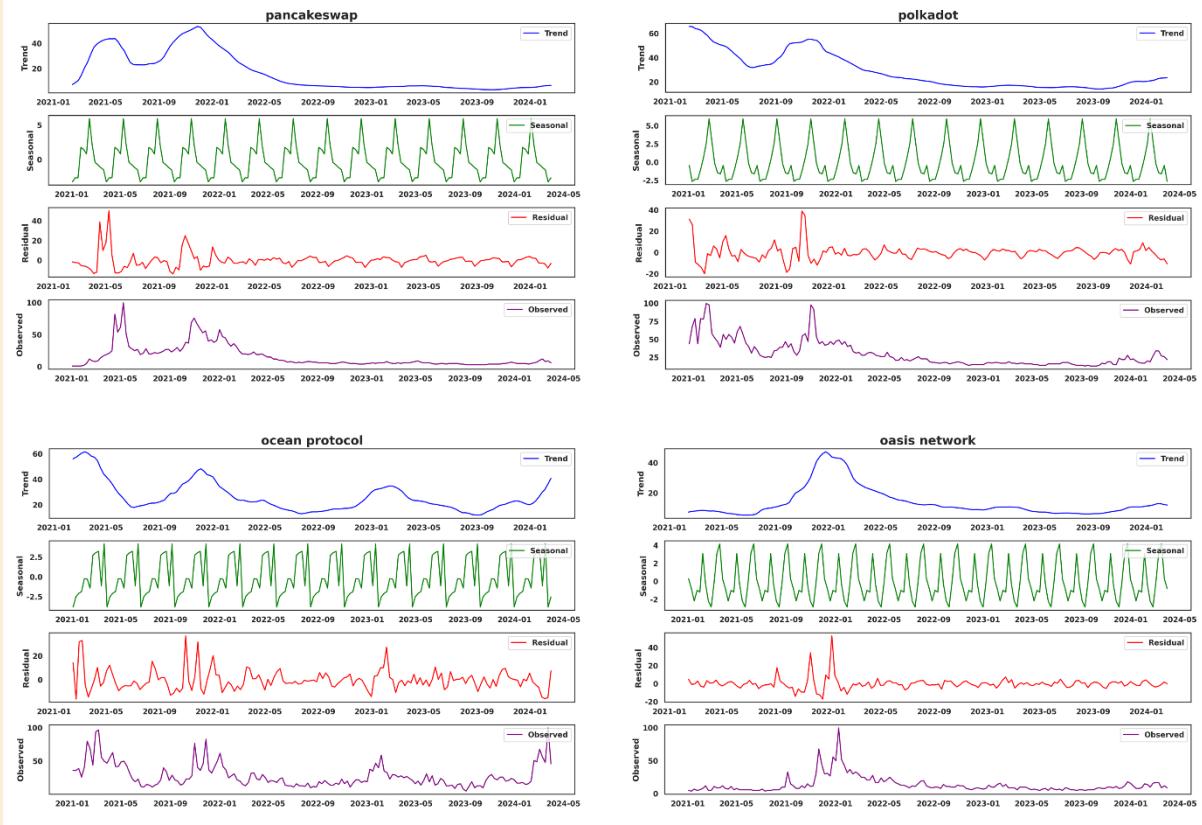


Figure7. Decomposition of Google Interest Trend data into Trend, Seasonal, Residual and Observed (original data) for PancakeSwap, Polkadot, Ocean Protocol, and Oasis Network.

To understand each part of the components, here we provide general understanding of each. The trend data ironed out the short-term fluctuations of the data providing core trend of how the data evolve over time. Increasing or decreasing spike in this trend data signal provide direction of movement in trend, while a flat trend exhibit stability of the data.

Seasonal data shown in this work were decomposed using period value of 12 in order to better capture the monthly trend of the data. The seasonal data demonstrate the cyclic nature that are observed in the time-series, highlighting the periodical cycle inherent in the data. The residual curve represents the remaining part of the data after the general trend and seasonality component were extracted. Generally, these residual values should be random spikes by nature, much like white noise. That being said, the appearance of significant spikes on the residual would then indicate anomaly or movement that are different than usual.

Based on characteristics of the decomposed curves, we observed the following similarity between projects. According to trend, we distinguish three type of trend cluster: (a) up-and-rise, (b) stagnant-downward, (c) neutral-movement. Projects that belong to the first cluster (a) are Solana, SingularityNet, Fetch.ai, Ocean Protocol, the second cluster (b) contain old project such as Dogecoin, Litecoin Filecoin, Tezos, while the rest simple belong to cluster (c). As to why the older generation (OG) project seems to gain less interest can be attributed to the diminishing hype of things from the past, so it is harder for this OG project to generate a

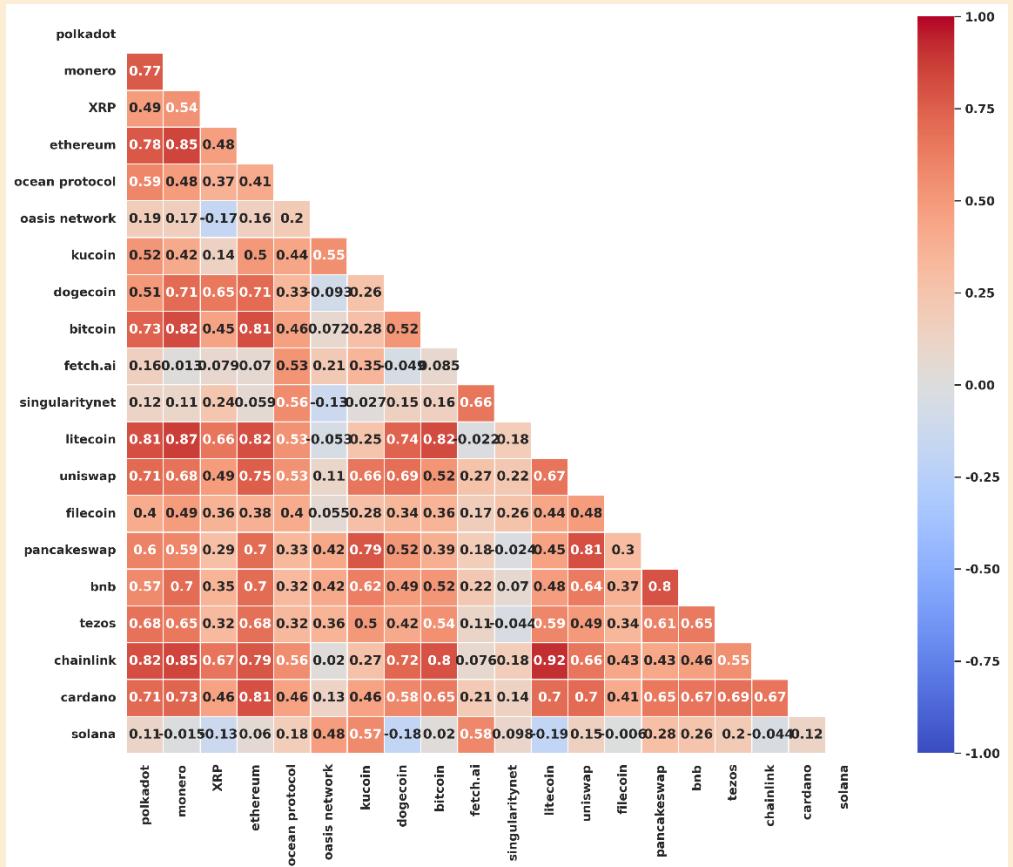
significant or meaningful amount of new peaking flux in interest. In terms of the seasonal cycle patterns, we identify three categorizations: (1) multi-peak, (2) double-peak, (3) single peak. The multi-peak contains projects such as Kucoin, SingularityNet, Dogecoin, Oasis Network, and Fetch.ai. The double-peak contains projects such as XRP, Tezos, Bitcoin, Ocean Protocol, Litecoin, and Chainlink. Whilst the rest of other projects belong to the single peak category. The fluctuations in the residuals are generally negligible and random for most of the time, but at certain event-driven instances the residuals were observably strong. On later section we will wring in additional data in attempt to identify some of the key driving force behind the outlier residuals. But first let us also analyze the temporal autocorrelation (correlation of data point with previous data points) for these Google Trend Time series data. In **Appendix A3.** we present the autocorrelation of the Google Trend data for up to 24 weeks (i.e. correlation memory of about half-year). Interestingly we found that projects that have shorter memory correlation, diminishing close to zero after 10 weeks or so, such kucoin, SingularityNet, Fetch.ai and few others are also the projects belonging to the double-peak cluster of seasonal cycle. One possible explanation to this is that only when the memory is short that the break in cycle (thus, the double peak) can happen. For the rest of other projects, the memory become weaker after 15 weeks or so but constant correlation was observed at longer period, signifying a build of momentum in the trend.

Noting of these special features discussed above, we later will use the information in choosing and developing appropriate ML framework that accommodate the trend, seasonal cycle, and residue for forecasting both the Google Trend and Token Price. Before that, let us also explore the correlation between Google Trends of different projects, between tokens of different projects, and also between token and Google trends of a given projects.

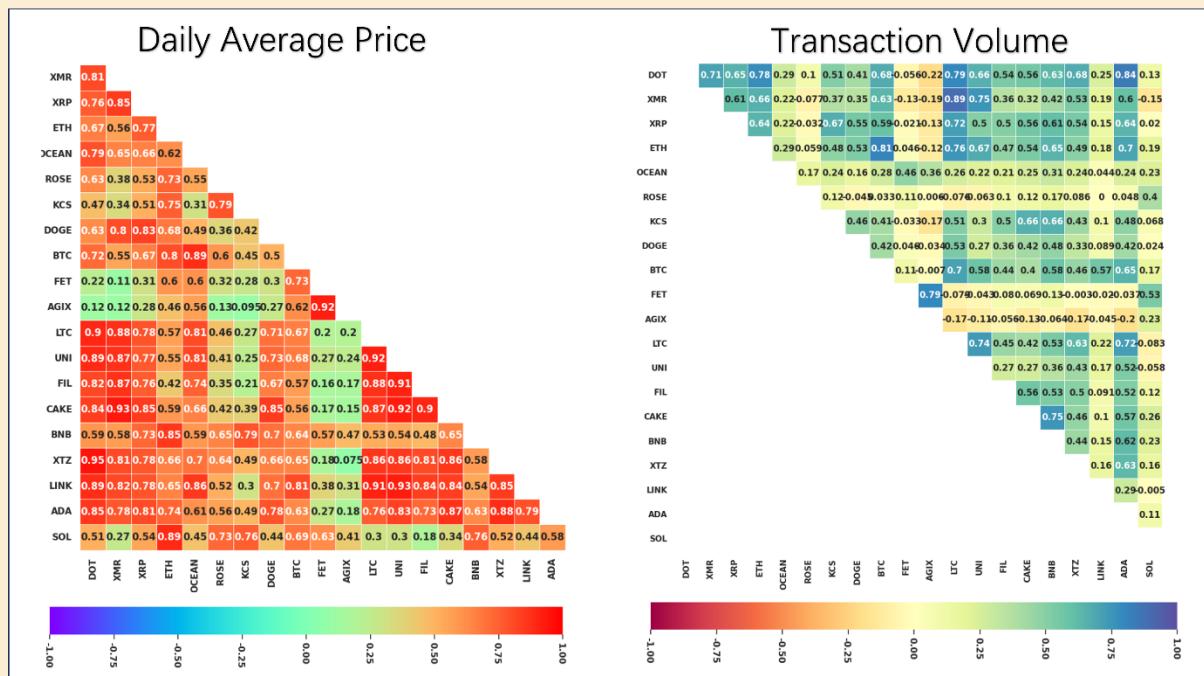
Correlation of Token Price and Google Trends Across Projects

In this section we analyze and present the Pearson correlations between: (1) Google Trends of different Crypto Projects in Fig. 8, (2) Token Price and Volume for different Crypto Projects in Fig 9. And Fig 10, and (3) Correlation between Crypto Project's Google Trend and the associated token price and trade volume.

In Fig. 9 we can observe that there are two camp of correlations that are pretty apparent. First are the Highly correlated projects with correlation of about >0.8 or so. Within this cluster we have Bitcoin-Ethereum-Chainlink-Litecoin-Monero as the prominent members. The correlation between Litecoin and Chainlink being the strongest with value of 0.92. On the other spectrum, we have projects that almost do not correlate with other project's Google Trend, namely, SingularityNet, Fetch.ai, Oasis Network, and Solana being the most typical member of the cluster.

**Figure 8.** Pearson Correlation Heatmap between Google Trends of Different Crypto Projects

In Fig. 9. We can observe similar pattern and clustering for the correlations between different projects' daily average price and transaction volume.

**Figure 9.** Pearson Correlation Heatmap between Daily Average Price (left) and also Transaction Volume (right) of Different Crypto Projects

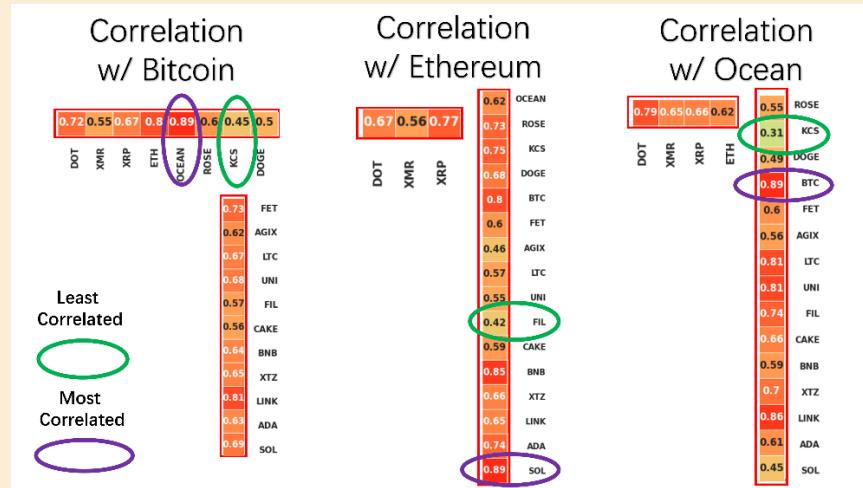


Figure 10. Focused Pearson Correlation Heatmap between Google Trends of Different Crypto Projects and how they relate to Bitcoin, Ethereum, and Ocean Token.

To highlight the correlation between tokens a bit further, we also specifically highlight the correlations between Bitcoin, Ethereum, Ocean, with the rest of other projects. Ocean and Bitcoin also Ethereum and Solana are the most correlated with correlation value of 0.89. The Kucoin's KCS token is the least correlated with both Ocean and Bitcoin, while with Ethereum, Filecoin have the least correlation.

Having analyzed the static correlation between Google Trends or Price of different projects, we now continue with the cross-correlation between the Google Trend and Price/Volume of the project Crypto Token. In Fig. 11, we exhibit the above mentioned correlation denoted as A: Trade Volume x Closing Price , B : Google Trend x Closing Price, C: Google Trend x Trade Volume. Based on this results we can highlight that for Dogecoin and Litecoin the Google Trend increase also correlate closely with increase of Trade Volume. While for project like Kucoin and Uniswap (Token Swap Exchanges), the relationship between trade volume and Google trend are weakest. When it come to price and Google Trend, Bitcoin and Ethereum are the lowest, this could be interpreted as price hike on these tokens doesn't preceded with increase in interest whatsoever.

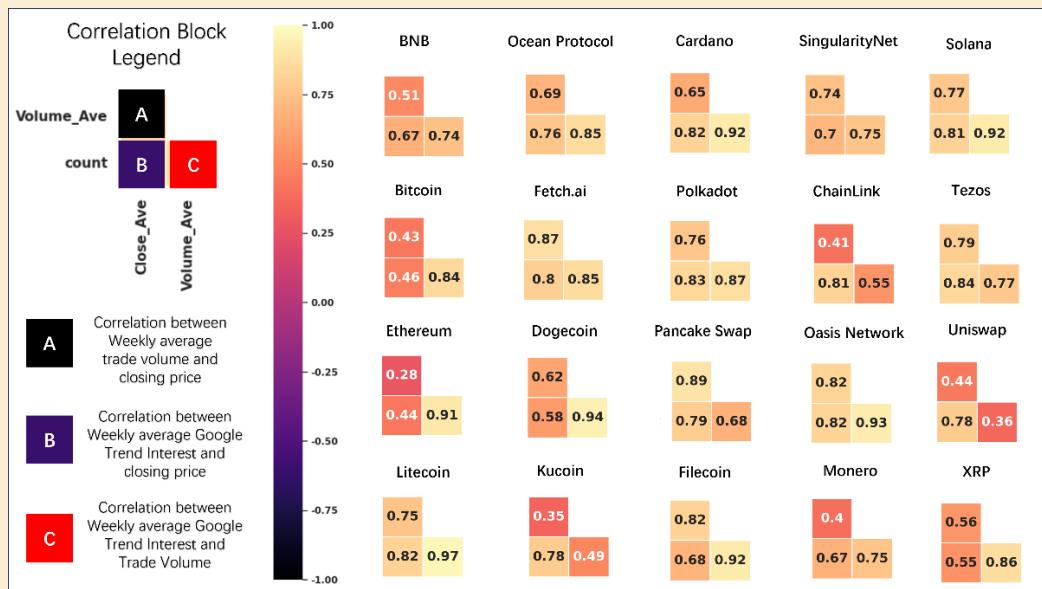


Figure 11. Cross-Correlation between Google Trend and Token Weekly average (Closed) Price or Volume.

Correlation Time Lags Analysis

The correlation in Fig. 11 is based on a single weekly average. We then also look toward the influence of time lag in the direction of how the correlation varies. In Fig.12, we picture a portion of these result in which instead of a single week average we analyzed data based on 2-, 4-, and 8-weeks average. Thereby giving a bi-weekly, monthly, and bi-monthly average correlation. The Directional Trend, for most of the project investigated here is that by incorporating longer time lag, correlation are enhanced. This is especially true with correlation that include Google Trend. One factor that may contribute to this phenomenon is the accumulated awareness of particular project. For full data on all other case please see **Appendix A4**.

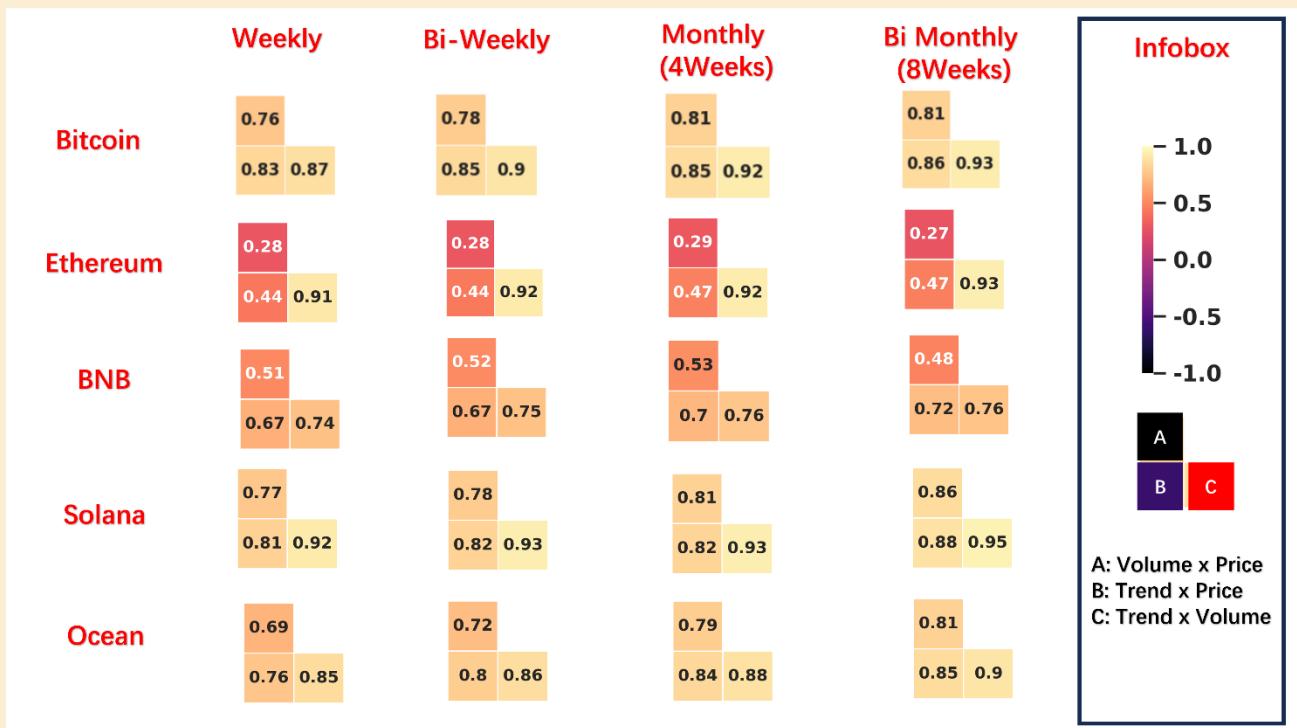


Figure 12 Time-Lag Dependence of the Correlation between Google Trend and Price or Volume

Event/Entity Driven Trend : Additional Data for More Interpretability and Understanding.

In the crypto market, the role of narrative and event can't never be undermined. Here in the attempt of incorporating more data for providing insights, we selected 8 keywords that been suggested through Google Search as highly related to Crypto Market. Namely, 'halving', 'ETF', 'tokenomics', 'ftx', 'microstrategy', 'blackrock', and 'coinbase'. We then follow the same protocol to extract the Google Trend data based on these 8 keywords for the same duration of time with the Aligned Dataset timeframe and shown in Fig.13

below. The aim for this exercise is to improve Interpretability of the data we have worked with. For instance, we can attribute the recent spike in ETF narrative and Halving along with recent rise in MicroStrategy purchase in bitcoin as drivers in the spike of decomposed residual data in the Google Trend we observed. Other events such as the IPO of Coinbase, or the bust of FTX, also appear as a extreme spike that influence several of the Crypto Projects involved.

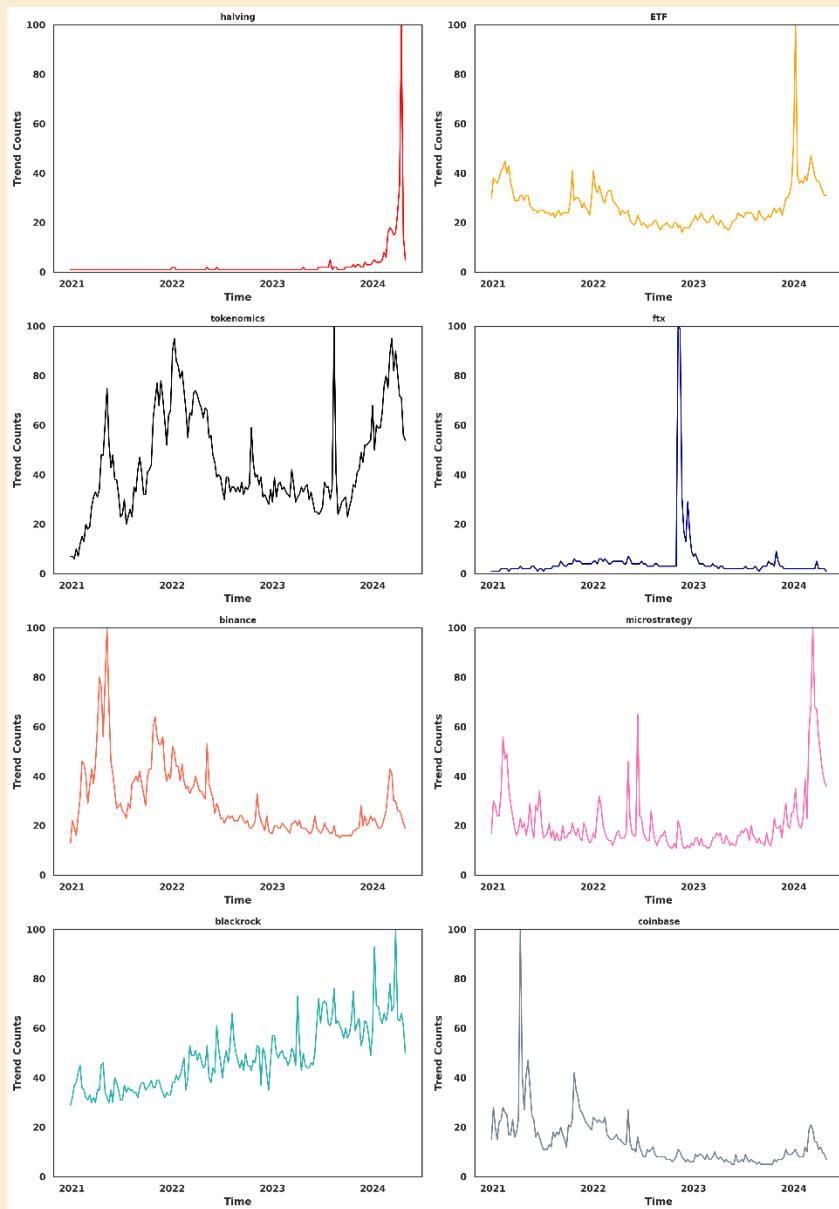


Figure 13. Event Driven Keyword to analyze different market narrative across crypto world.

Machine Learning Model

In this section we proceed with the construction of the ML model for Google Trend and Price Forecasting. Based on our analysis regarding the correlations and also the time-series decomposition we decided to build our ML pipeline for forecasting using the Neural Prophet framework. We first perform preprocessing of input

data preprocessing for all 20 projects studied here. Prior to deploying a full search we performed test development by exploring different possible combinations of hyperparameters, as well as other methods for forecasting. We have run several test using LSTM, XGBoost, Seasonal-ARIMA, and decided to stick working with NeuralProphet. After several rounds of pre-development we settle with hyperparameter space given by, `{learning_rate: [0.01, 0.1, 0.5, 1], batch_size : [8,16], optimizer :[AdamW, SGD], loss_func=[SmoothL1,MAE,MSE]}`. We then parallelly submitted and run all the computation jobs to then produce best models for both Google Trend Forecasting as well as Price Forecasting. To analyze the performance of these models we employ tracking of Learning Curve based on the Mean Average Error (MAE), and Mean Squared Error (MSE) shown in **Fig. 14** and **Fig. 15** for 8 out of 20 projects. The rest of the learning Curve data are shown in **Appendix A5**.

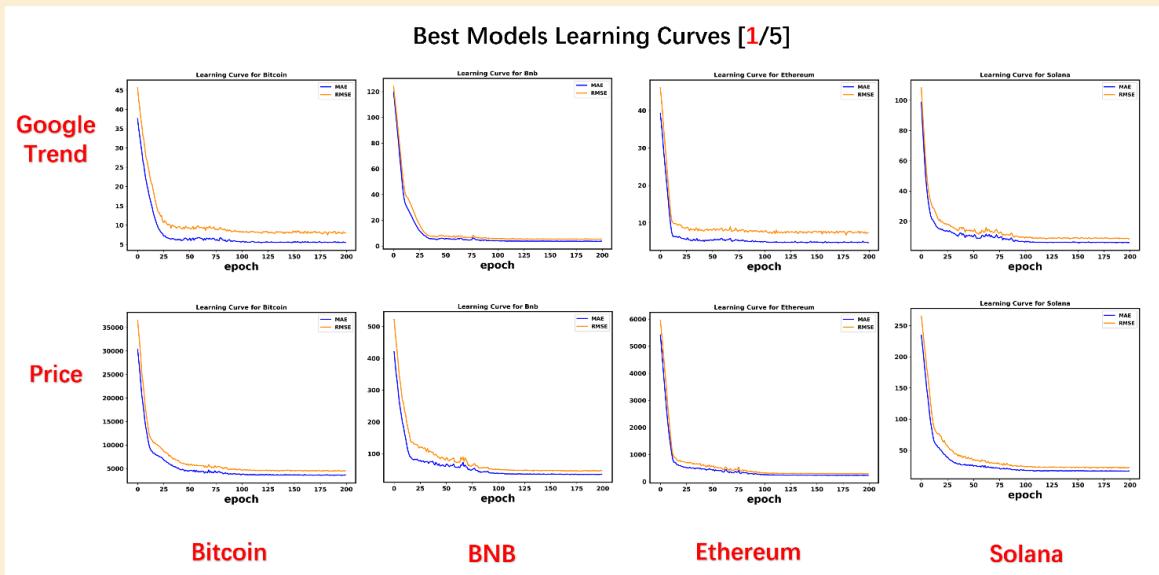


Figure 14. Learning Curve of Best models for Crypto Project Google Trend (upper panels) and token price (bottom panels) forecasting. Part [1/5]

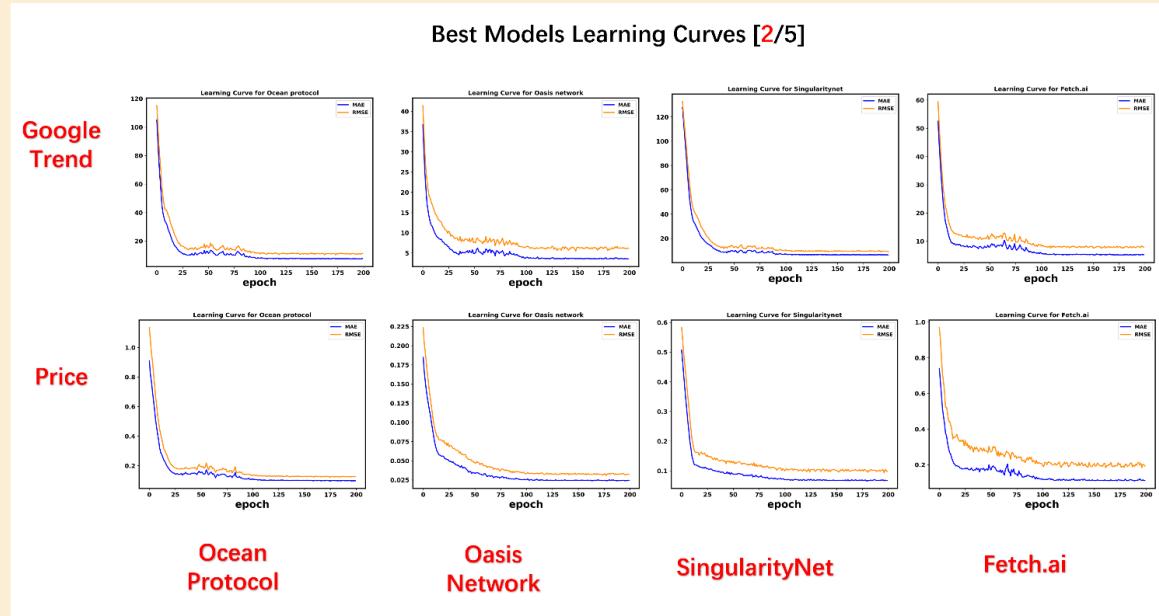


Figure 15. Learning Curve of Best models for Crypto Project Google Trend (upper panels) and token price (bottom panels) forecasting. Part [2/5]

The resulting Best Models are plotted in **Fig. 16**, **Fig. 17**, and **Fig. 18**, for part of the projects, and the rest are shown in **Appendix A6**. Using the best model, we then deployed a forecasting task for the upcoming 52 weeks ahead of the last data entry. The performance of these models was reported based on the MAE, RMSE, and R^2 value. The R^2 value for most fall between the range of 0.75-0.88, while the highest goes up to 0.919.

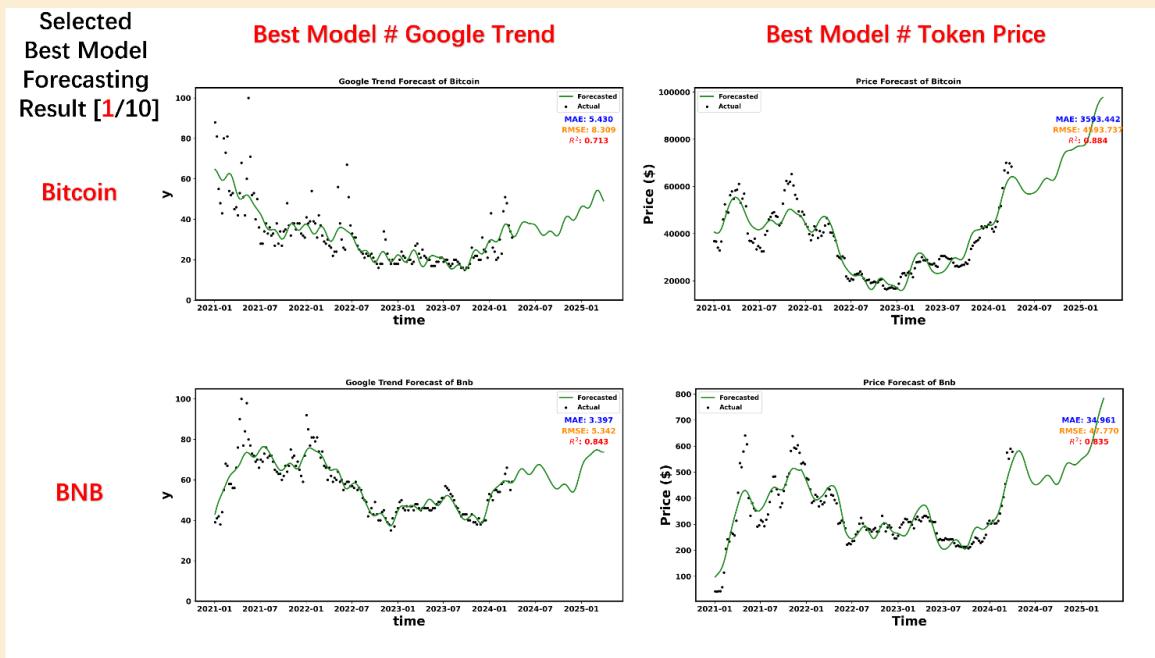


Figure 16 Selected best Model Forecasting Result for Google Trend and Token price of Bitcoin and BNB

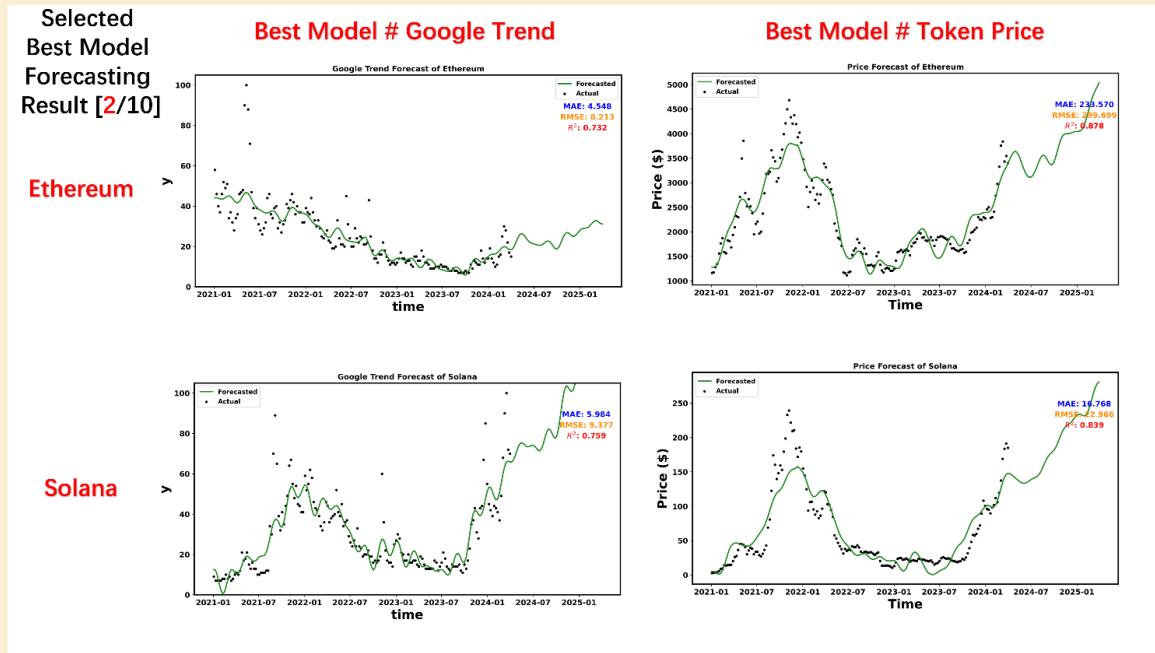


Figure 17 Selected best Model Forecasting Result for Google Trend and Token price of Ethereum and Solana

These models can be considered as starting point since the aim is to build a general enough framework. Optimization for each particular project is not performed exhaustively in this current work, but will be planned for works that follows.

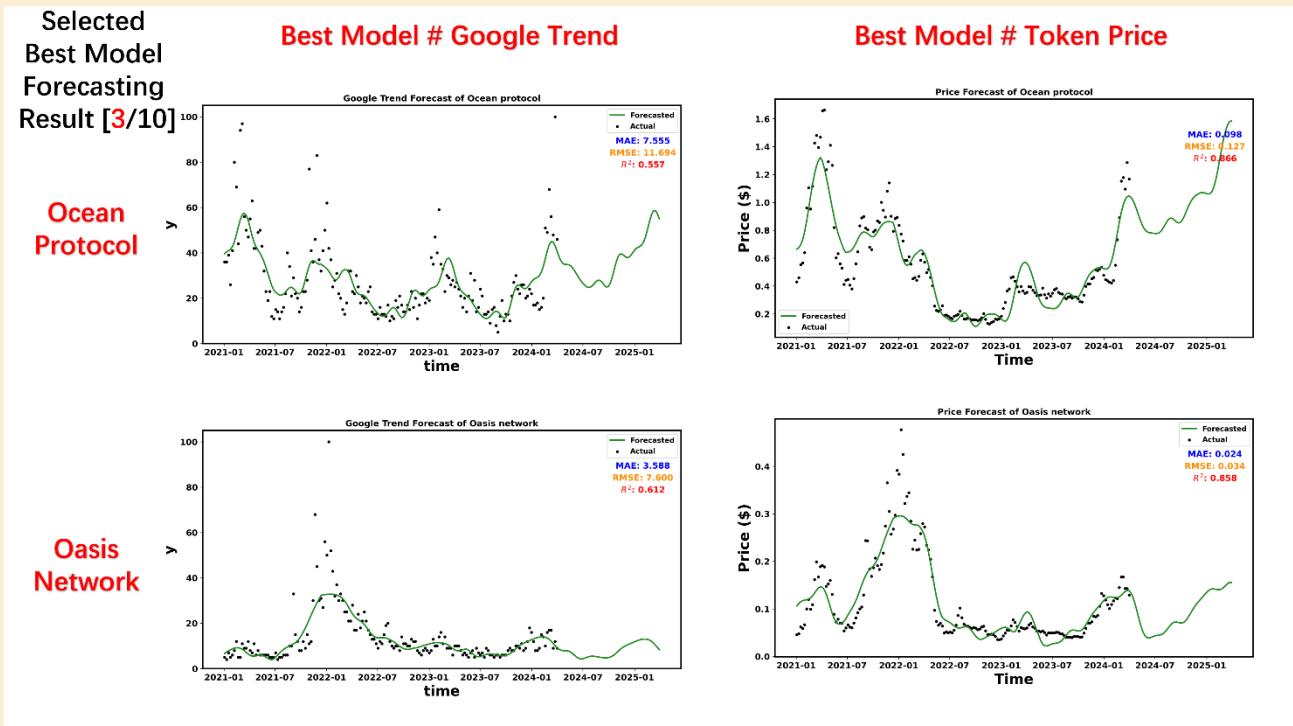


Figure 18 Selected best Model Forecasting Result for Google Trend and Token price of Ocean Protocol and Oasis Network

Methodology

NeuralProphet Framework

The NeuralProphet model is a specialized package designed for time-series forecasting tool built on PyTorch ecosystem. The framework heavily borrows from Facebook's Prophet model. Initially we attempted to work directly with the Facebook's original prophet model but unfortunately the fact that the original codes and other infrastructure were mainly developed in R, the python wrapper is not robust enough and we decide to work with a Python native NeuralProphet instead.

NeuralProphet include several key aspects that make it attractive for our application: 1. Model Components, It integrates various components such as trends, seasonality, various custom regressor, and treatment of special events in the data pattern. 2. Data Preprocessing, in order to construct a meaningful and powerful forecast model, data quality is of utmost importance. The NeuralProphet framework have many of this processing capability built-in. The model offers options for data normalization (min-max or z-score) and handles missing data through imputation if auto-regression is used. This preprocessing is crucial for aligning and preparing data for accurate forecasting. 3. Extendability, one other main reason we consider the use of NeuralProphet is the modularity fo perform both single and multi-step forecasts, including single or multiple time series when necessary. In this work we demonstrated one of the basic applications. In the future more advanced extension of the existing model can be expected.

The Pearson Correlation Coefficient measures the linear relationship between two datasets. Ranging from -1 to 1, a value of 1 implies a perfect positive linear relationship, -1 implies a perfect negative linear

relationship, and 0 implies no linear relationship. The Pearson Correlation Coefficient between two variables X and Y is defined as:

$$r_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

In the above \bar{X} and \bar{Y} are the means of the X and Y variables, respectively.

The AutoCorrelation Function (ACF) allows us to measures the correlation of a signal with itself at different lags or time delay. It is a useful tool in time series analysis to identify repeating patterns or seasonal effects in our weather dataset. The ACF at lag or delay time (in our context in the unit of hours) k is defined as the correlation between time series observations that are k time periods (hours) apart. Mathematically, the ACF at a given time lag is given as:

$$ACF(k) = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}$$

Where y_t is the value at time t and \bar{y} is the mean of the series, N is our total number of observations.

Conclusion

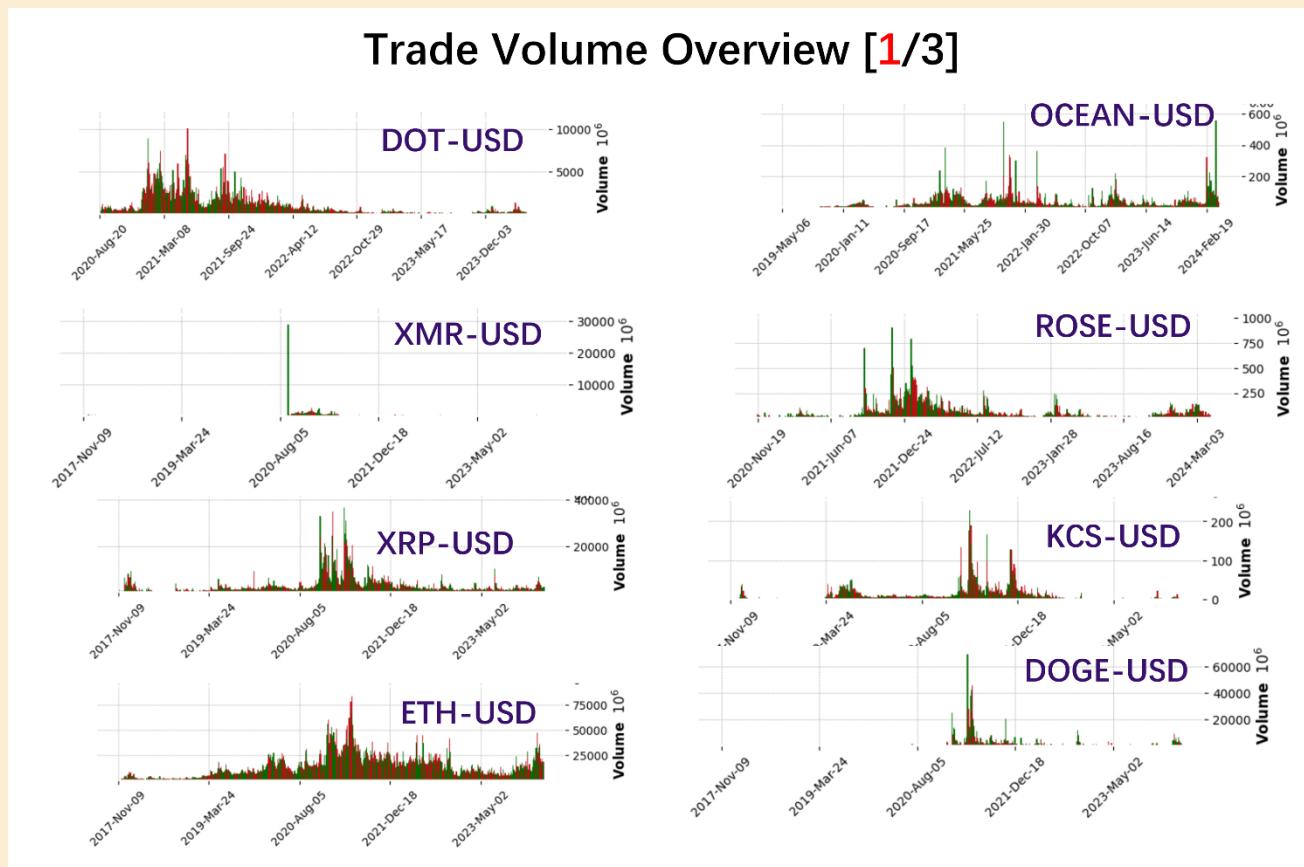
This investigative study presents a thorough data analysis and exploration of correlations, time-lag characteristics, and time-series decomposition concerning Google Trends and token prices for 20 selected crypto/blockchain projects. By decomposing the time-series data, we have identified several clusters of crypto projects that is moving up in popularity such as Fetch.ai, SingularityNet, Solana, Ocean and some others that are stuck or in downfall trend, such as Dogecoin and Litecoin. Our analysis also includes a detailed exploration of various factors that contribute to understanding the data better, such as the incorporation of event-driven trends that explain outlier spikes in the residual data from our decomposed time-series.

In addition to our in-depth analysis, we build strong mini-library of forecasting models for predicting the Google Trend as well as price for the upcoming year with R^2 score that goes as high as 0.88 for most cases. Moreover, in order to demonstrate the utility of our exploratory data analysis tools and pipeline in full we also include all the results and analysis output produced in this work.

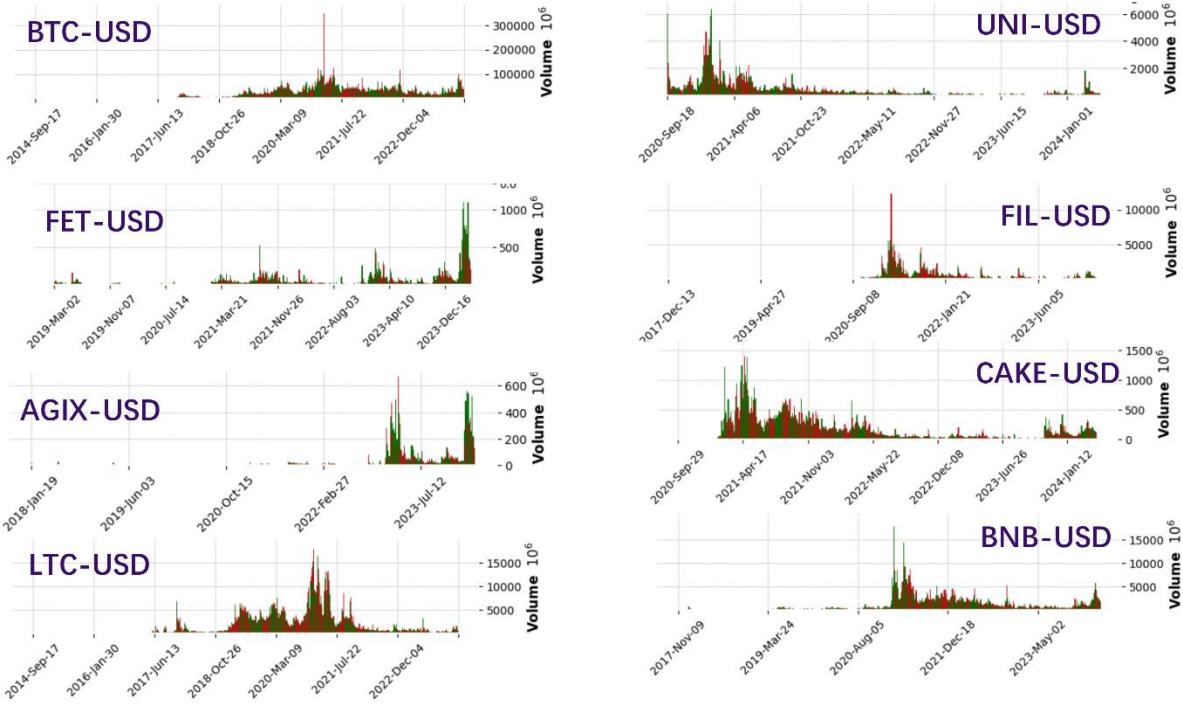
Looking ahead, we plan to expand our developed forecasting models and the presented data into a "CryptoForecast MiniApp." This application, based on the Streamlit package, will be hosted on a decentralized cloud (Akash) and connected to the Ocean marketplace and Predictoor, enhancing accessibility and utility for users interested in real-time data for Google Trends and Crypto Token Price forecasts.

Appendices

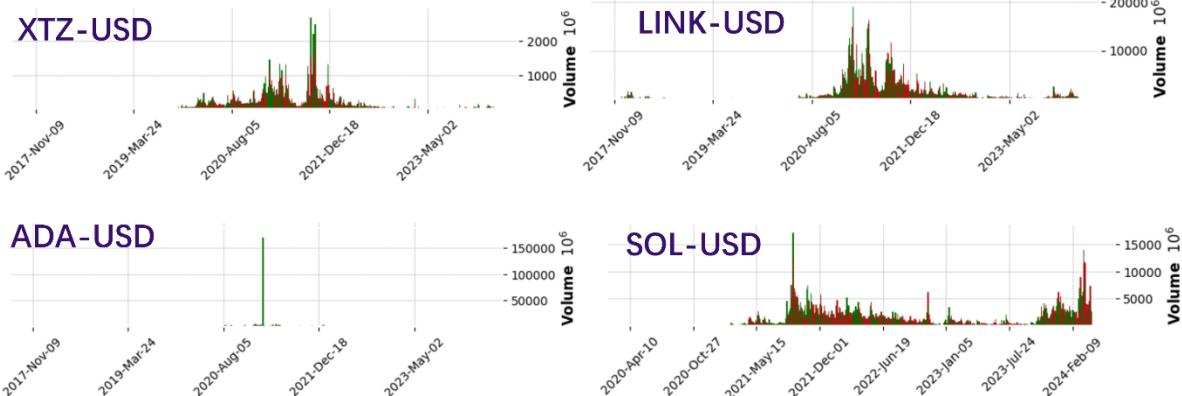
A1. Trade Volume Raw Data



Trade Volume Overview [2/3]

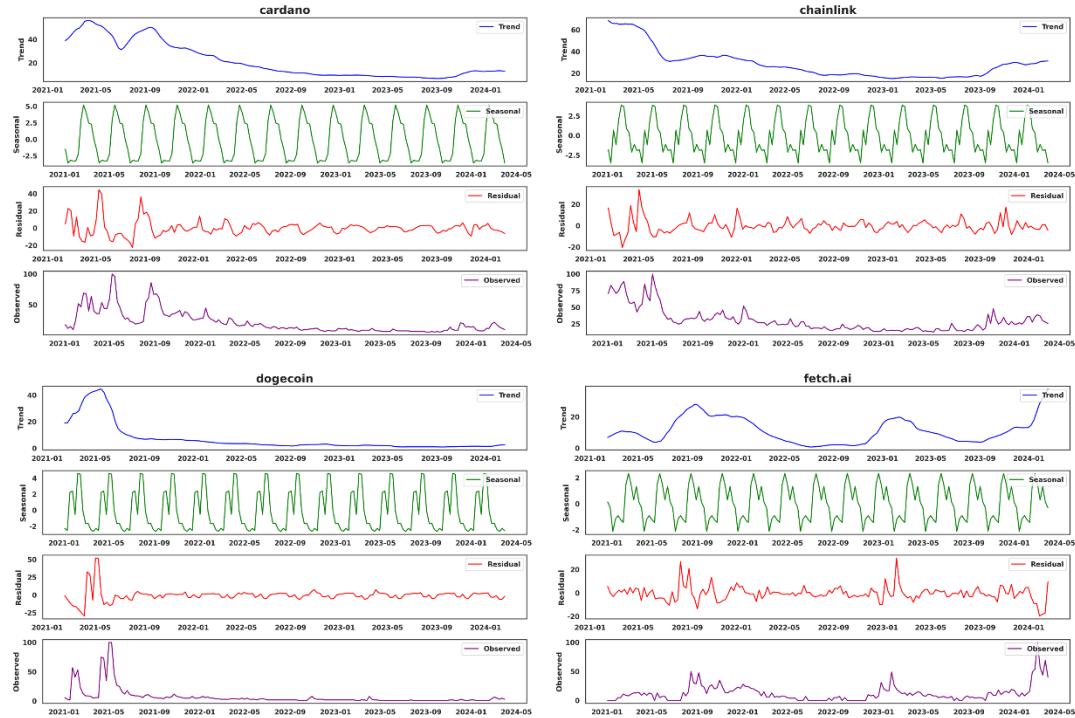


Trade Volume Overview [3/3]

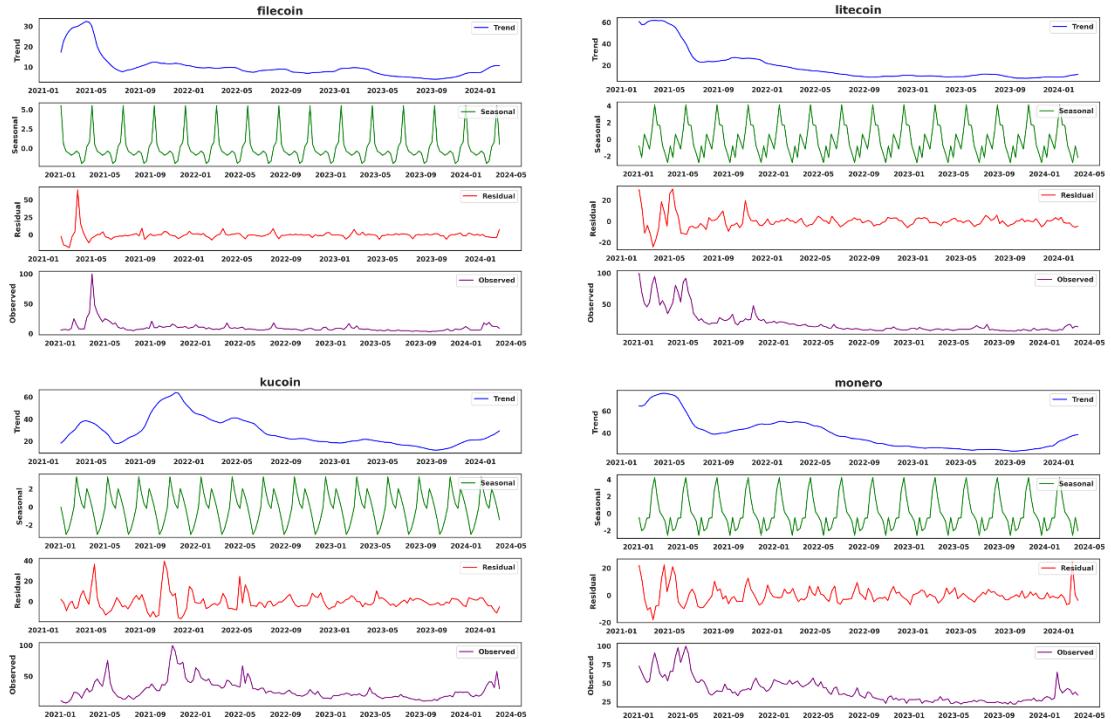


A2. Google Interest Trend Time-Series Decomposition

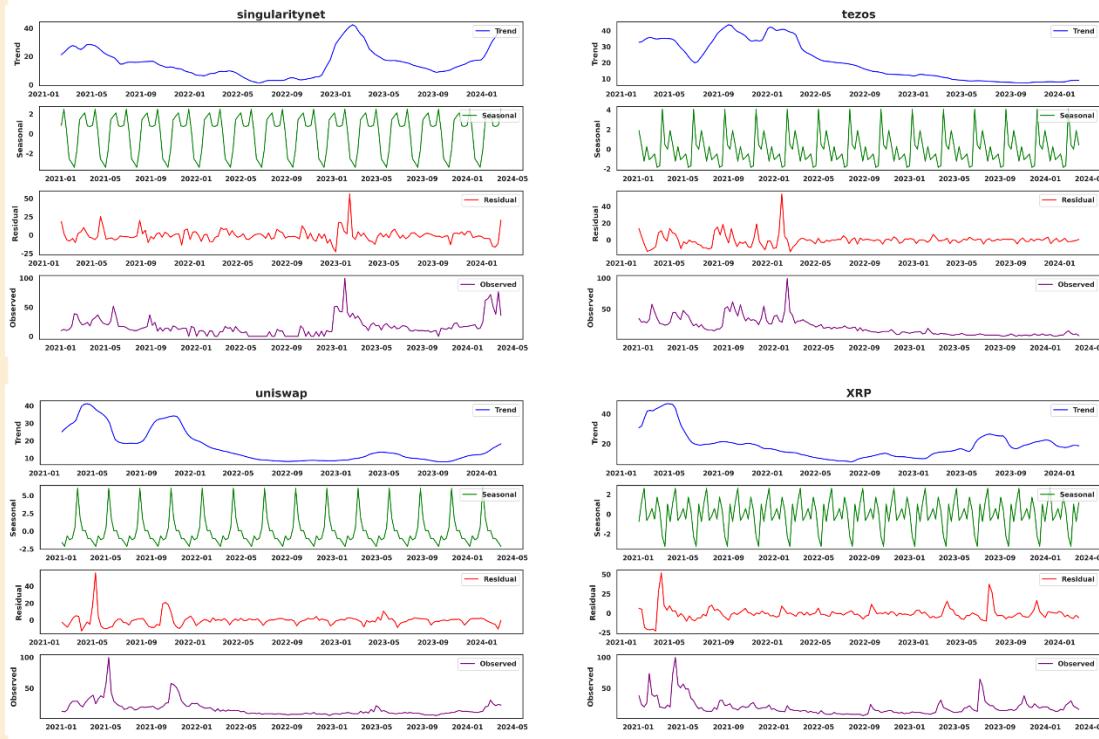
Google Interest Trend Time-Series Decomposition [2/5]



Google Interest Trend Time-Series Decomposition [3/5]

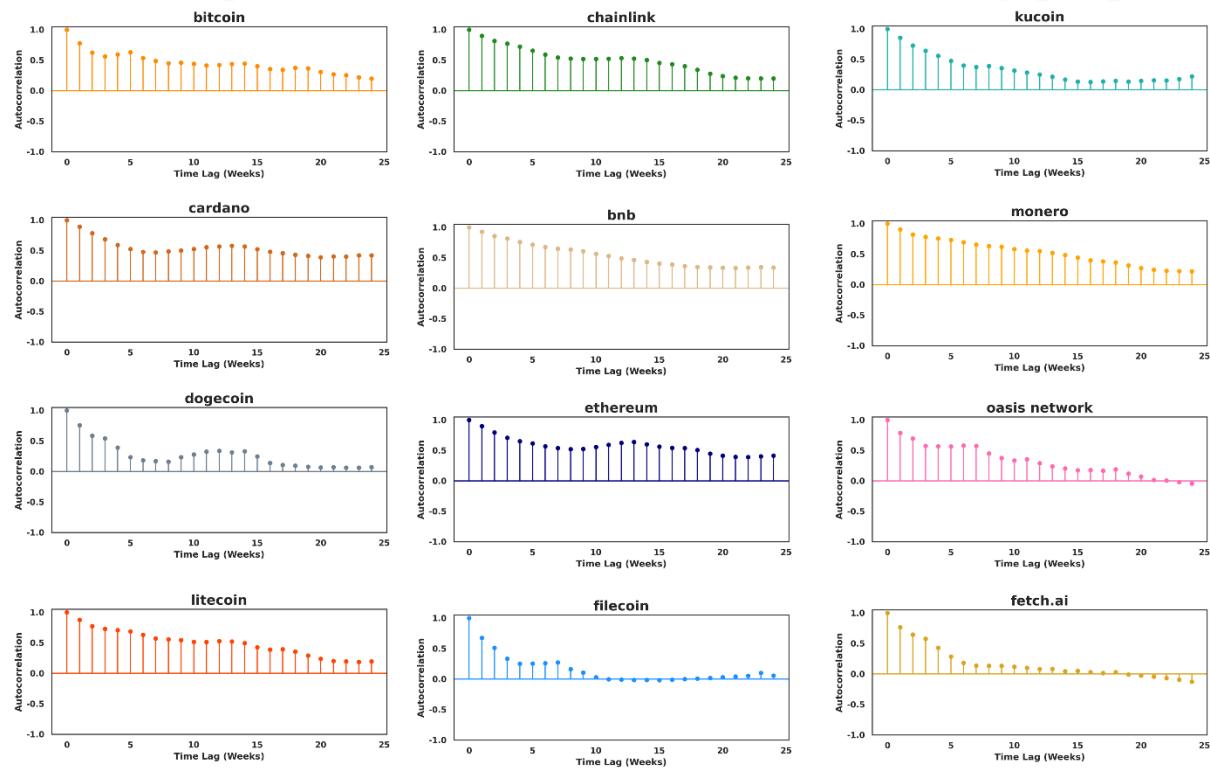


Google Interest Trend Time-Series Decomposition [5/5]

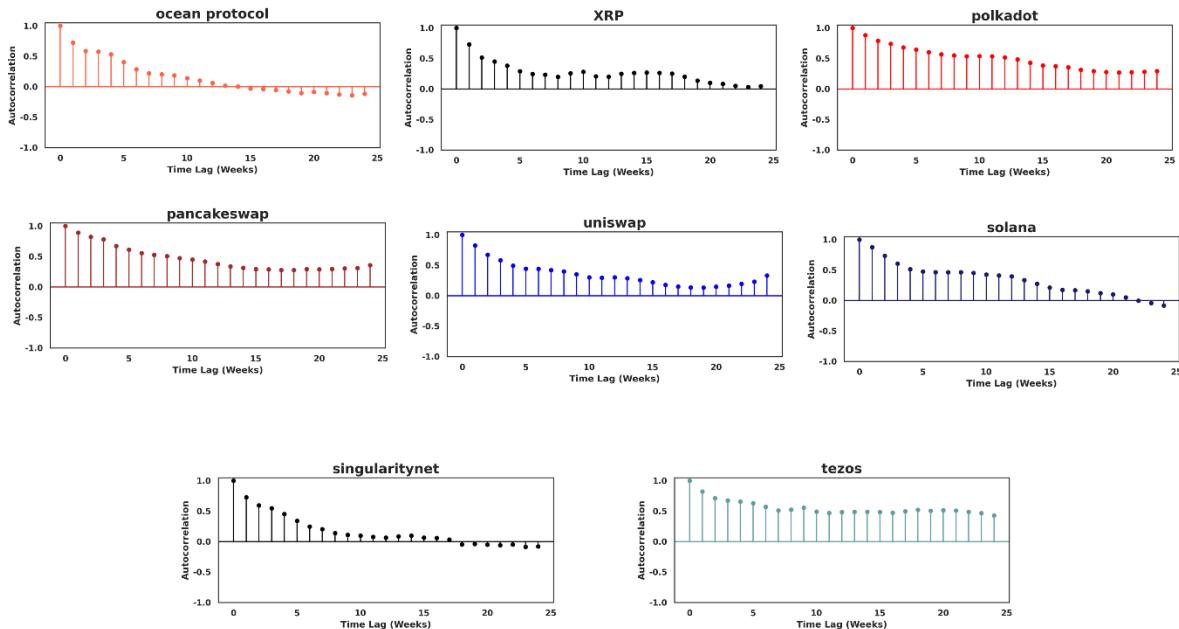


A3. Autocorrelation

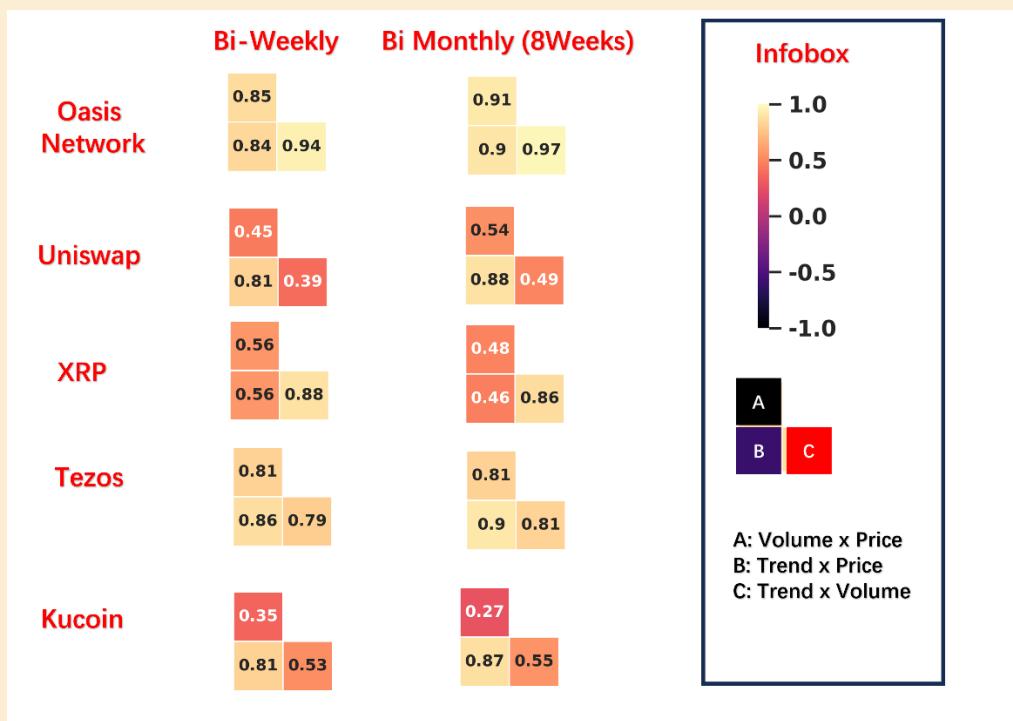
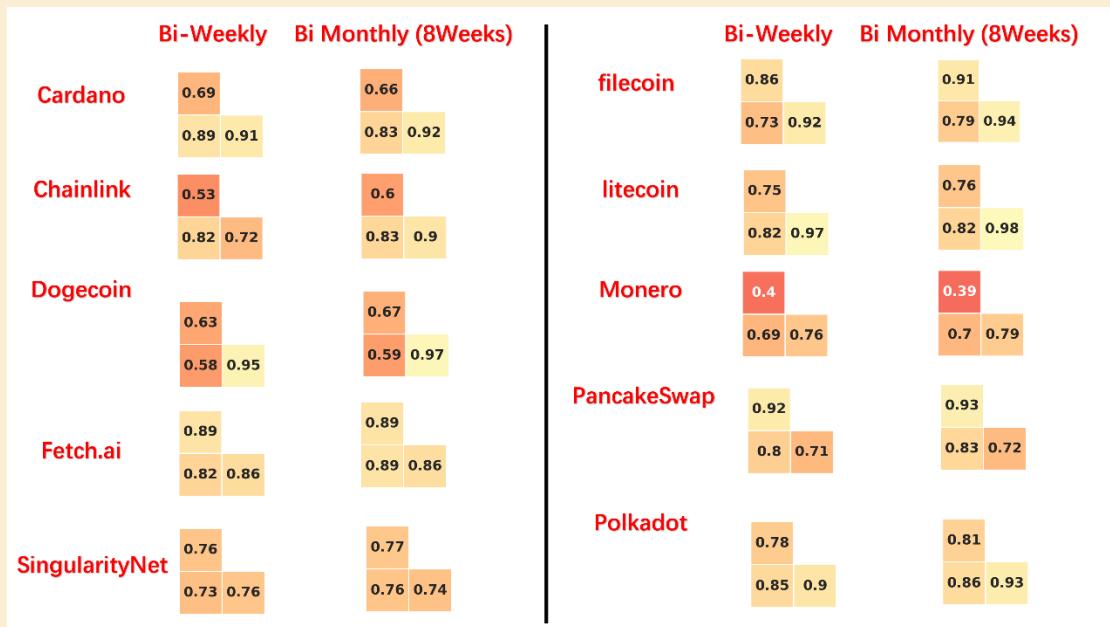
Google Interest Trend Autocorrelation Memory [1/2]



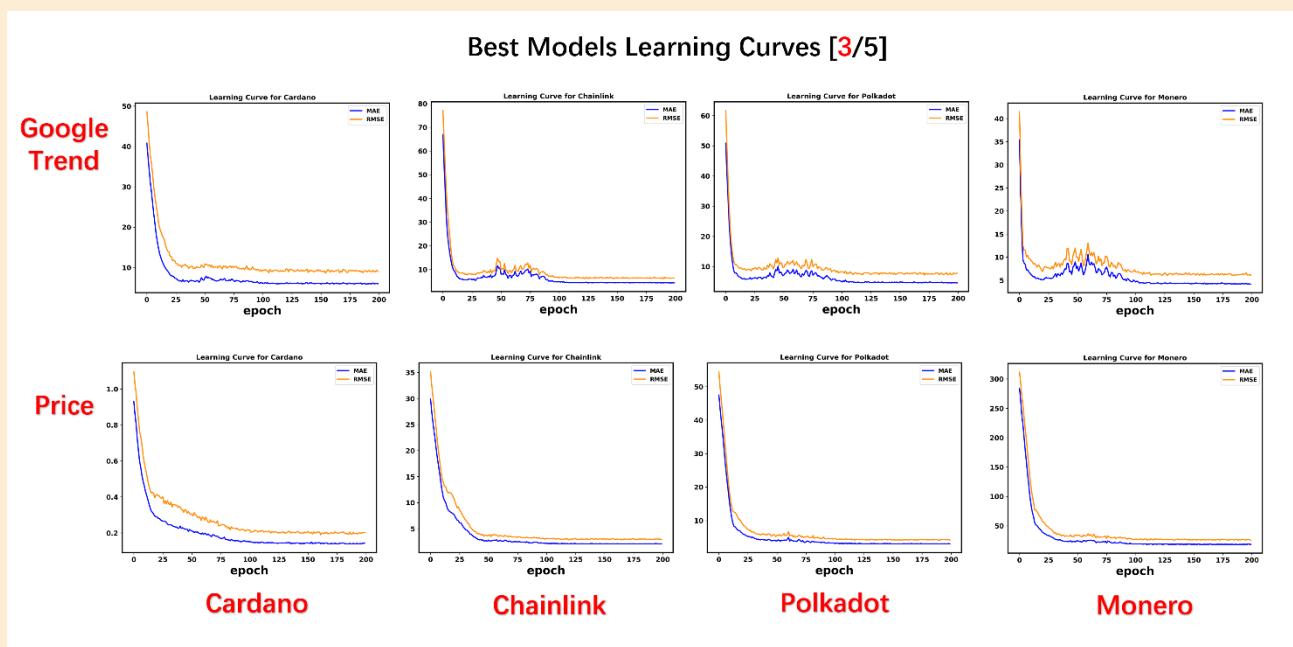
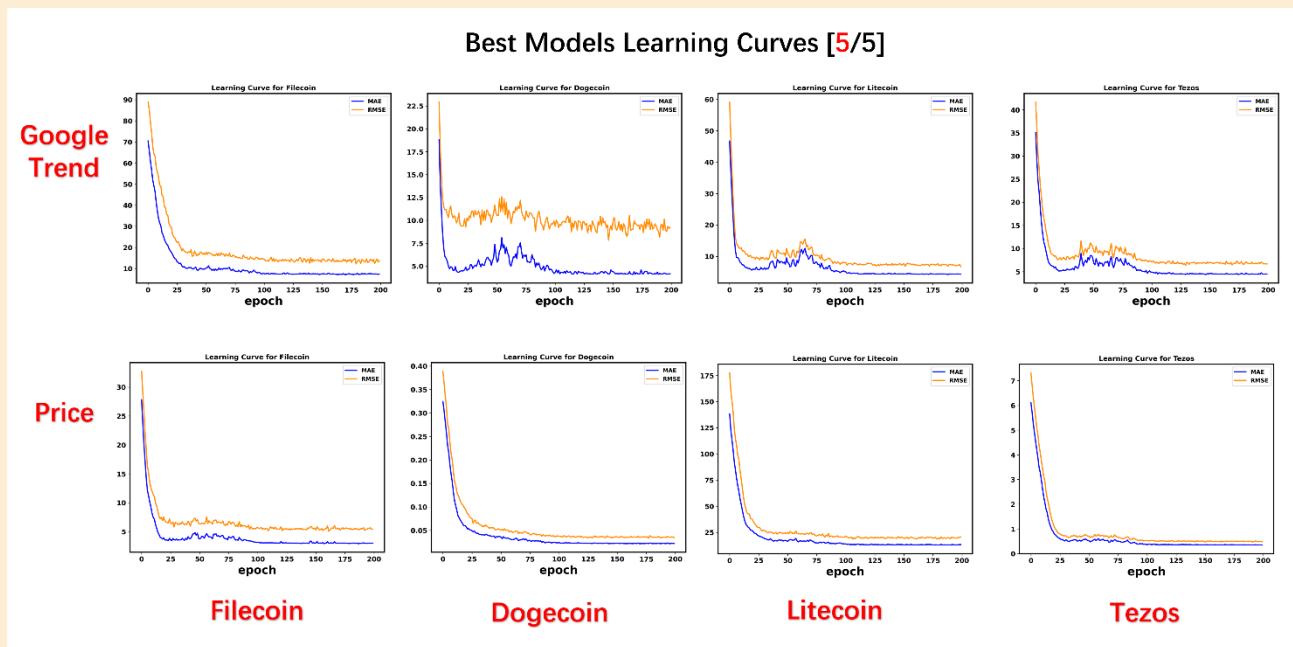
Google Interest Trend Autocorrelation Memory [2/2]

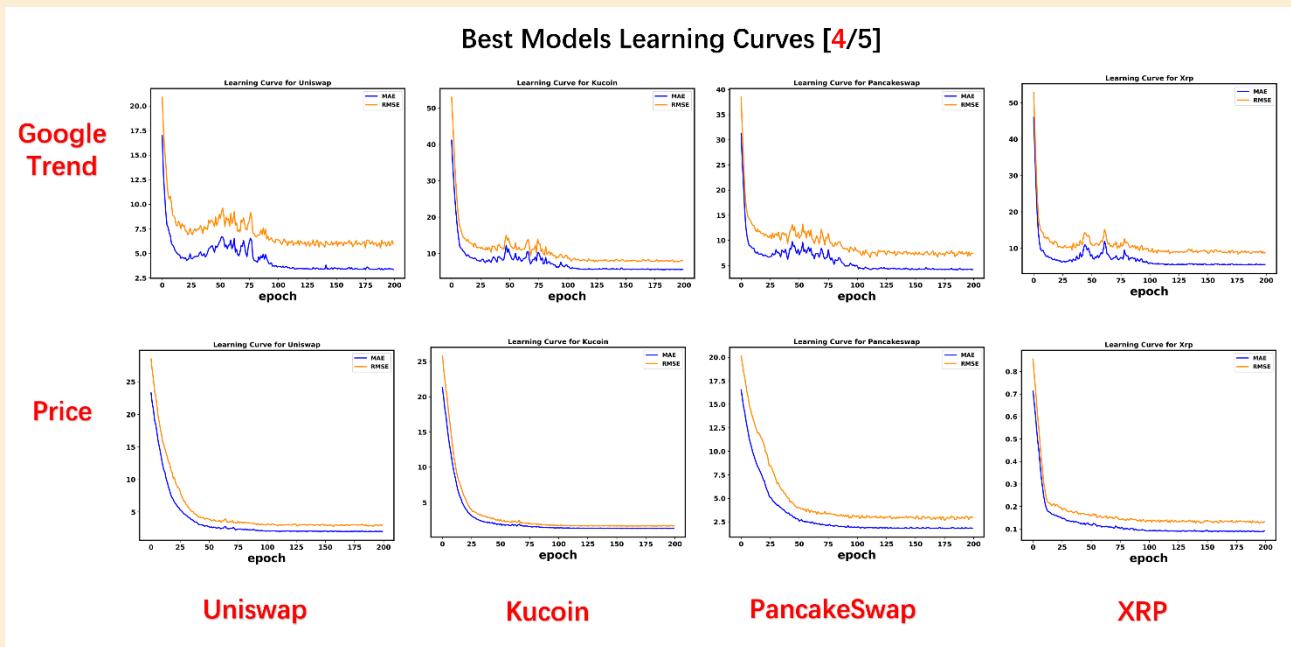


A4. Time-Lag

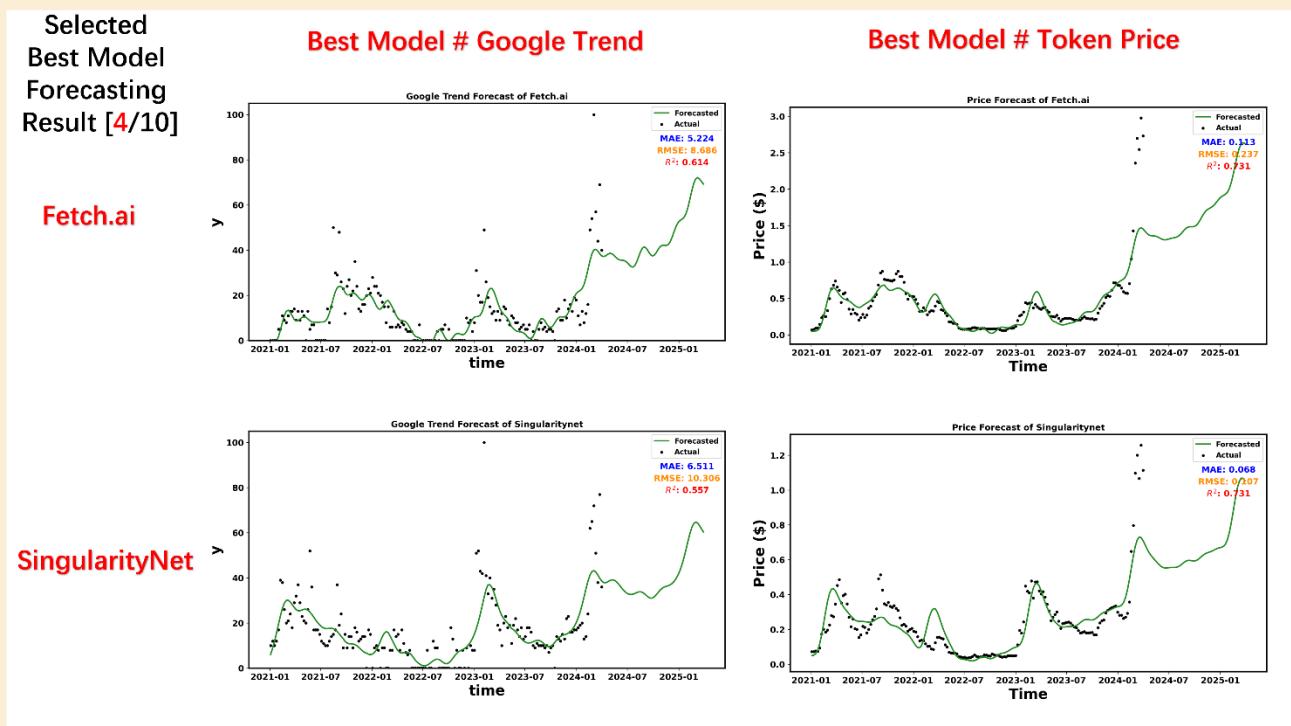


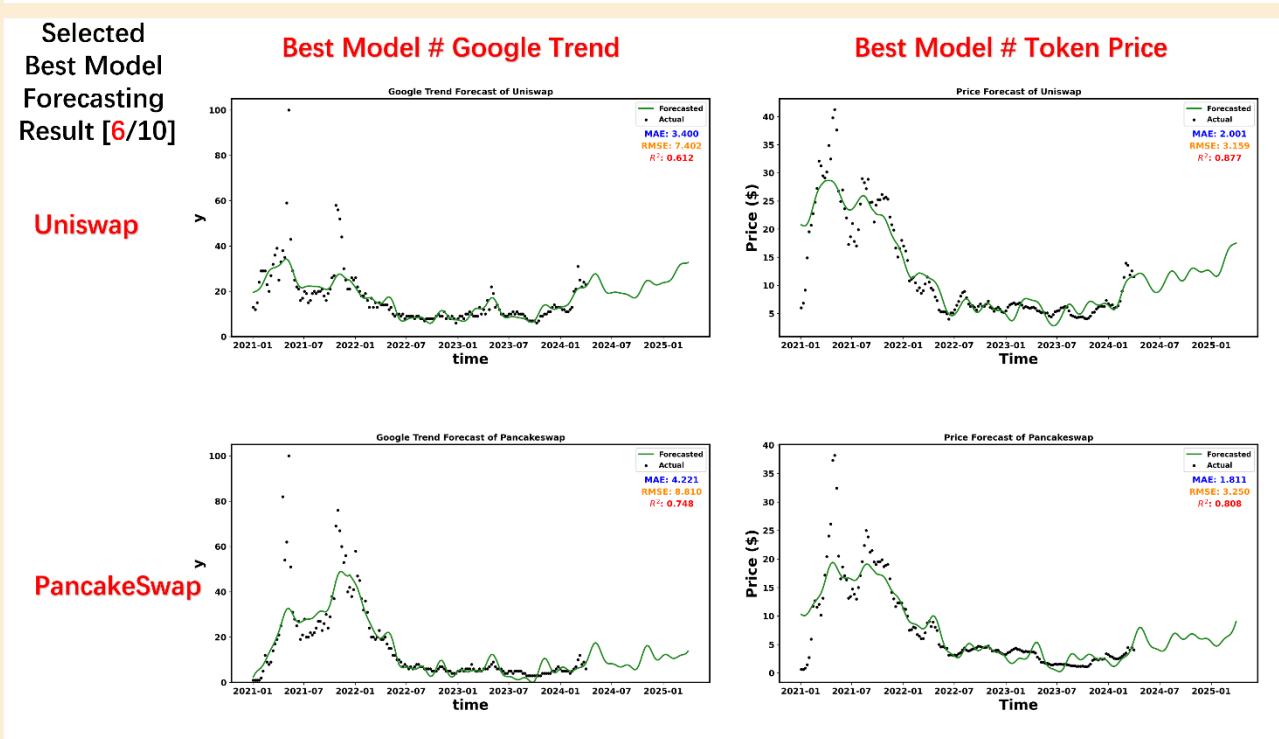
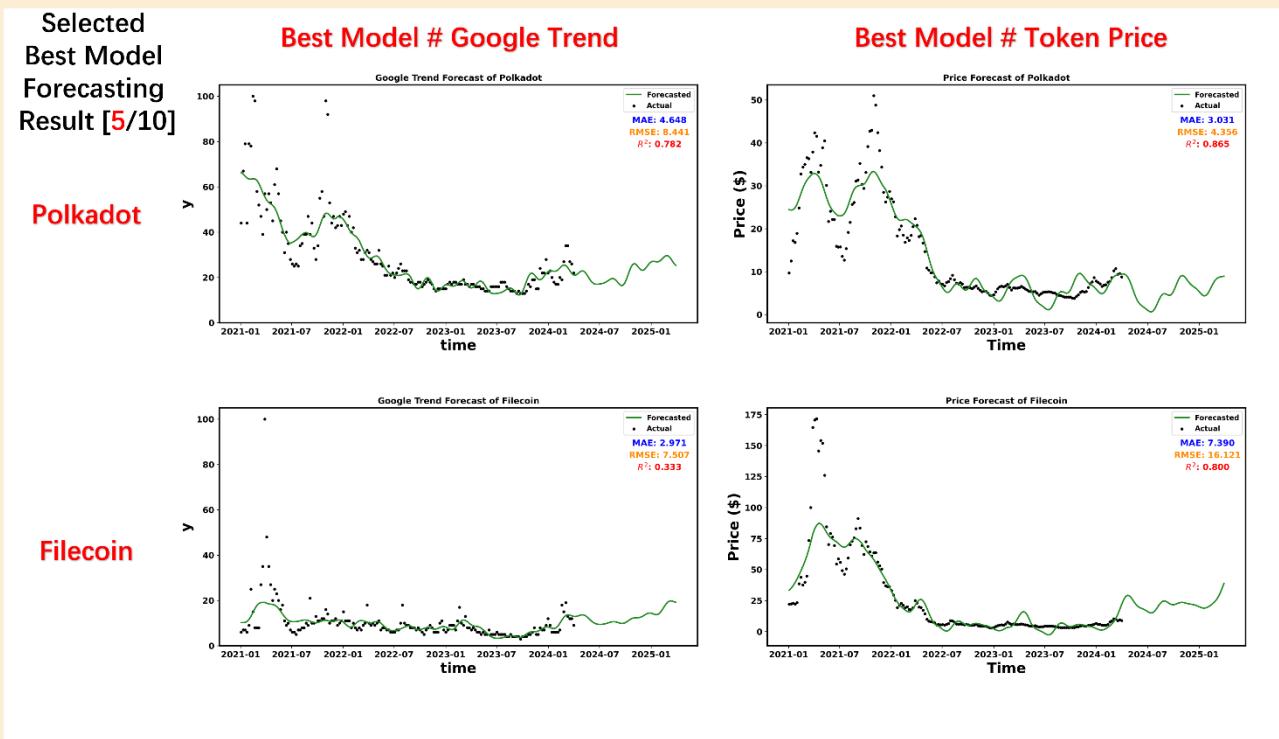
A5. Learning Curve





A6. Forecasting Model for Google Trend and Price

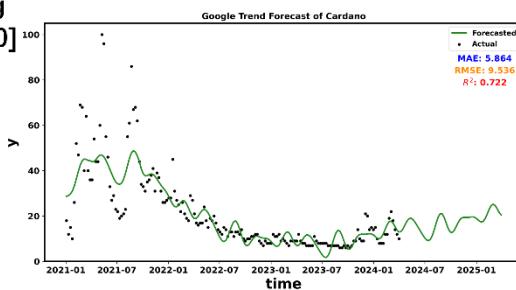




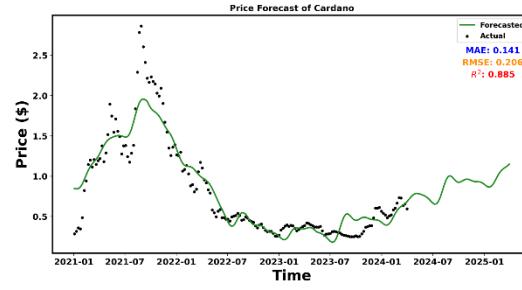
**Selected
Best Model
Forecasting
Result [7/10]**

Cardano

Best Model # Google Trend

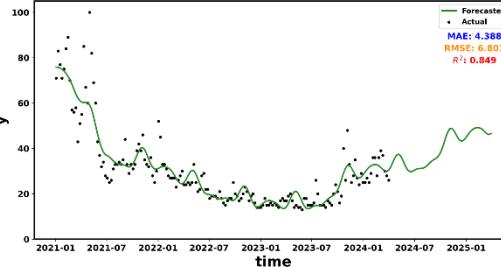


Best Model # Token Price

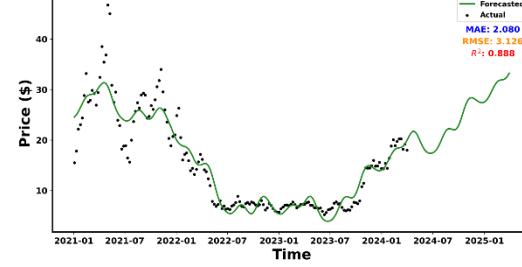


Chainlink

Google Trend Forecast of Chainlink



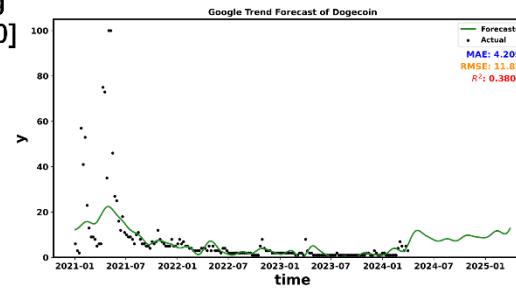
Price Forecast of Chainlink



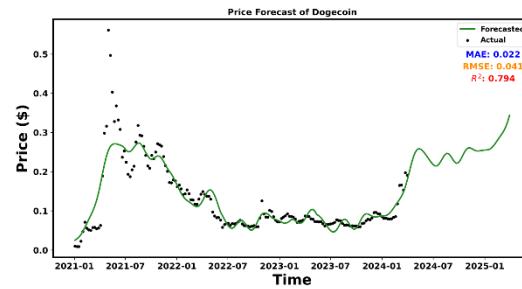
**Selected
Best Model
Forecasting
Result [8/10]**

Dogecoin

Best Model # Google Trend

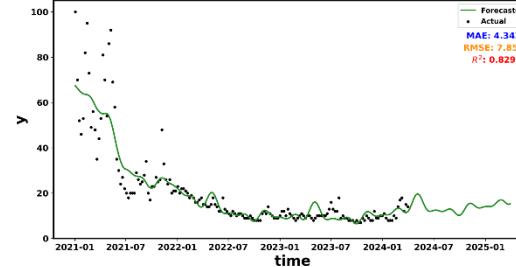


Best Model # Token Price

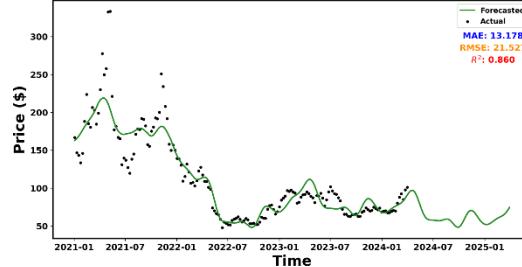


Litecoin

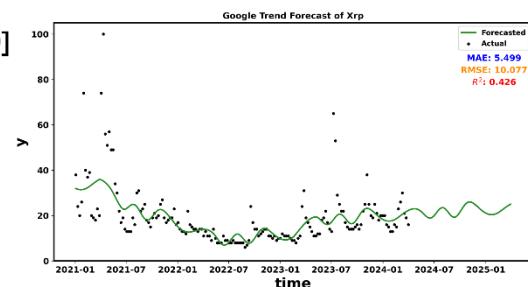
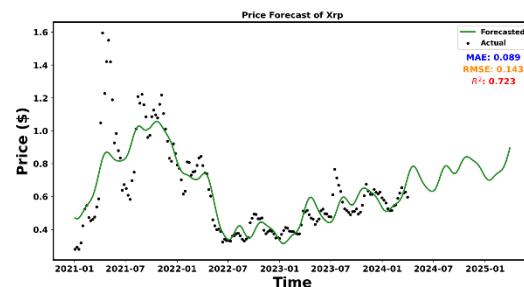
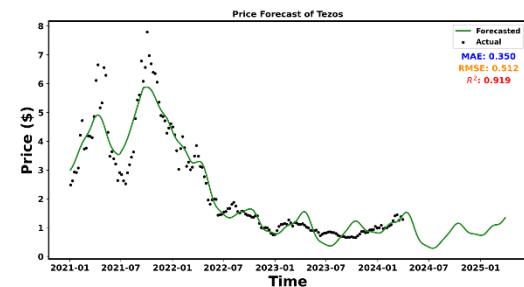
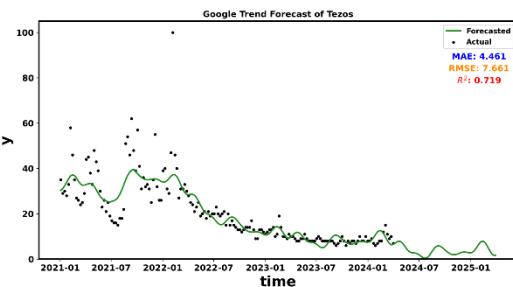
Google Trend Forecast of Litecoin



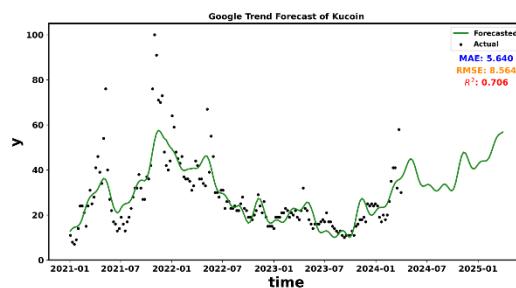
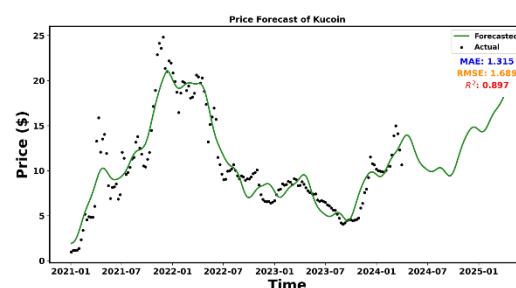
Price Forecast of Litecoin



**Selected
Best Model
Forecasting
Result [9/10]**

XRP**Best Model # Google Trend****Best Model # Token Price****Tezos**

**Selected
Best Model
Forecasting
Result [10/10]**

Kucoin**Best Model # Google Trend****Best Model # Token Price****Monero**