Final Report:

The Effects of Google Search Trends on Cryptocurrencies

Problem Statement

Since Cryptocurrencies first came onto the market in 2009, they have garnered an immense amount of praise and criticism. Supporters of cryptocurrency have heralded them as being a decentralized currency that has protection from inflation and is self-governed and managed. Critics have expressed concern over its environmental impact, its uses in illegal activities and its high volatility.

Whichever your stance on cryptocurrency may be, it is certainly a topic that many people have expressed their opinion on online. In fact, recently Elon Musk has made headlines by announcing that Bitcoin, Ethereum, and Dogecoin was going to be an acceptable mode of payment for their future project of landing on Mars. Having all this attention and people talking about these coins shot the prices up overnight to record highs.

I think that this is an aspect of cryptocurrencies that is worth exploring. Do people's interest in cryptocurrencies affect the price of the coin? And can we use this data to make predictions about the price of cryptocurrencies?

Datasets

- CCi30 Index – The CCi30 is a rules-based index designed to objectively measure the overall growth, daily and long-term movement of the blockchain sector. We were looking to mimic this index
- Top 30 cryptocurrencies from Yahoo Finance – To mimic the CCi30 index we pulled data from the Yahoo Finance API. To measure the price of stock, we pulled daily 'Closing' price of each stock.
    - Stocks included in our data: ['BTC-USD', 'ETH-USD', 'USDT-USD', 'BNB-USD', 'ADA-USD', 'DOGE-USD', 'XRP-USD', 'USDC-USD', 'DOT1-USD', 'HEX-USD', 'UNI3-USD', 'BCH-USD', 'LTC-USD', 'SOL1-USD', 'LINK-USD', 'MATIC-USD', 'THETA-USD', 'XLM-USD', 'VET-USD', 'ETC-USD', 'TRX-USD', 'ICP1-USD', 'FIL-USD', 'XMR-USD', 'EOS-USD', 'AMP1-USD', 'AAVE-USD', 'SHIB-USD', 'ALGO-USD', 'CRO-USD']
- Google Search Trends Dataset – Looked at popularity of search terms from top crypto currencies- "Cryptocurrency", "Bitcoin", "Ethereum", 'XRP", "Dogecoin".

Data Wrangling

Originally, I was pulling daily stock data from Yahoo Finance API from 1/1/2015 to 8/1/2021 which gave us 2401 rows and 30 columns. However, many cryptocurrencies did not exist in 2015. There were some columns of stock closing prices that had over 1500 or more missing values. Removing the rows and columns with many missing values left us with 1025 rows and 19 columns. For any other missing values found in the middle of the dataset, we used forward fill to take the last closing price and fill missing values with those values.

Our Google Trends Dataset measured the popularity of different google search terms measured as weekly data. To match my stock price data and trends data, I resampled Yahoo Stock Data from daily measurements to weekly measurements

EDA

The first step in our project was to mimic the CCi30 index. The CCi30 index takes in data from the top 30 crypto currencies and returns a value that represents the overall health of the cryptocurrency sector. It is formatted like a typical historical stock data table with “Date”, “Open”, “Close”, ”High”, ”Low”, and “Volume” as columns. Here we only use the closing price of the stock as a metric to evaluating the valuation of the stock.

The closing price is typically used in stock market valuations as the most accurate measure of stock performance. Seen in Figure 1.

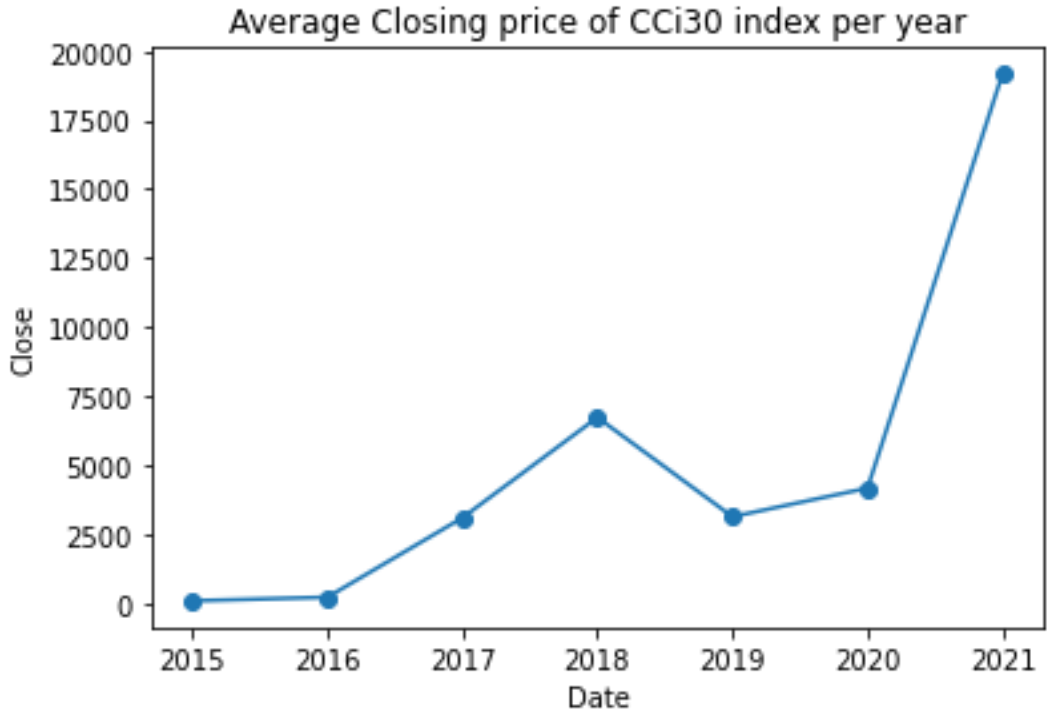


Figure 1: Closing Price CCi30

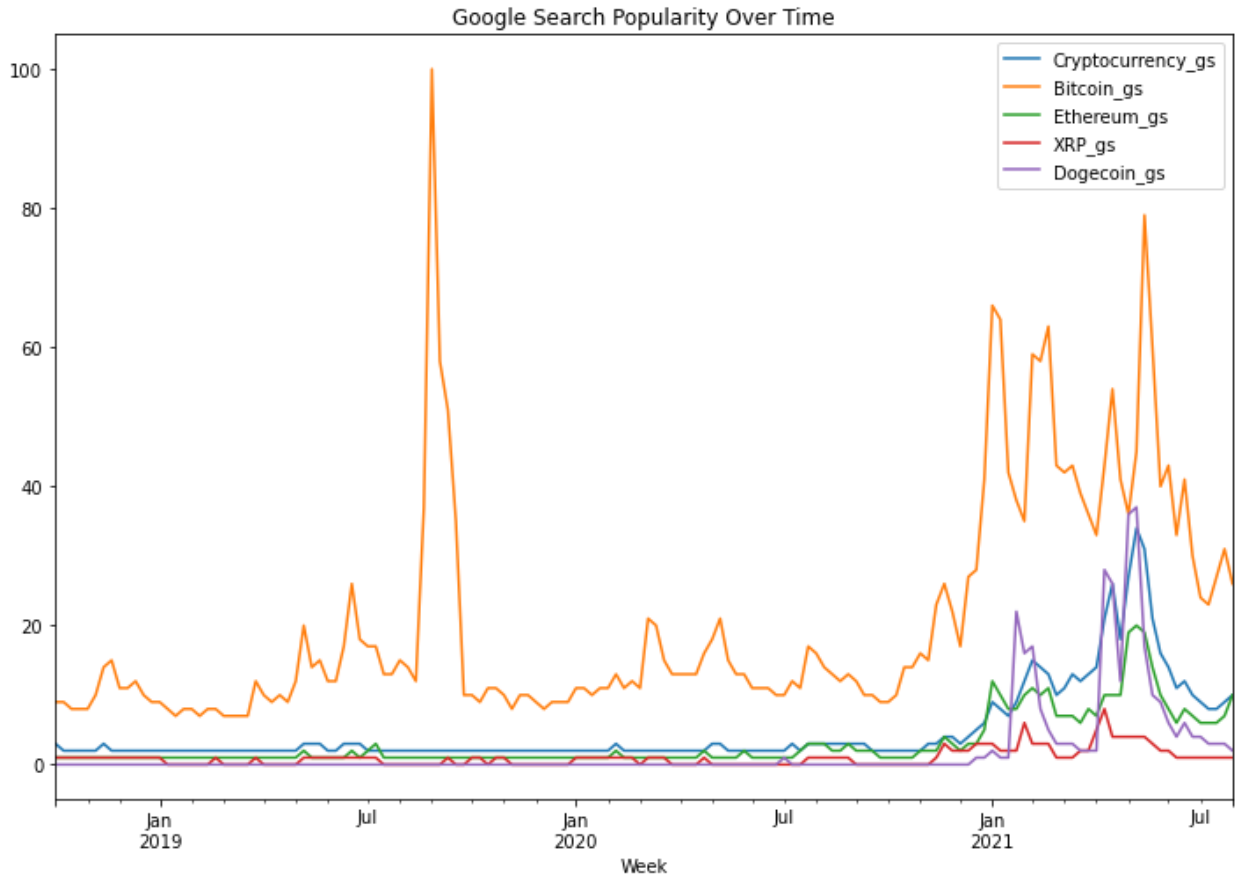
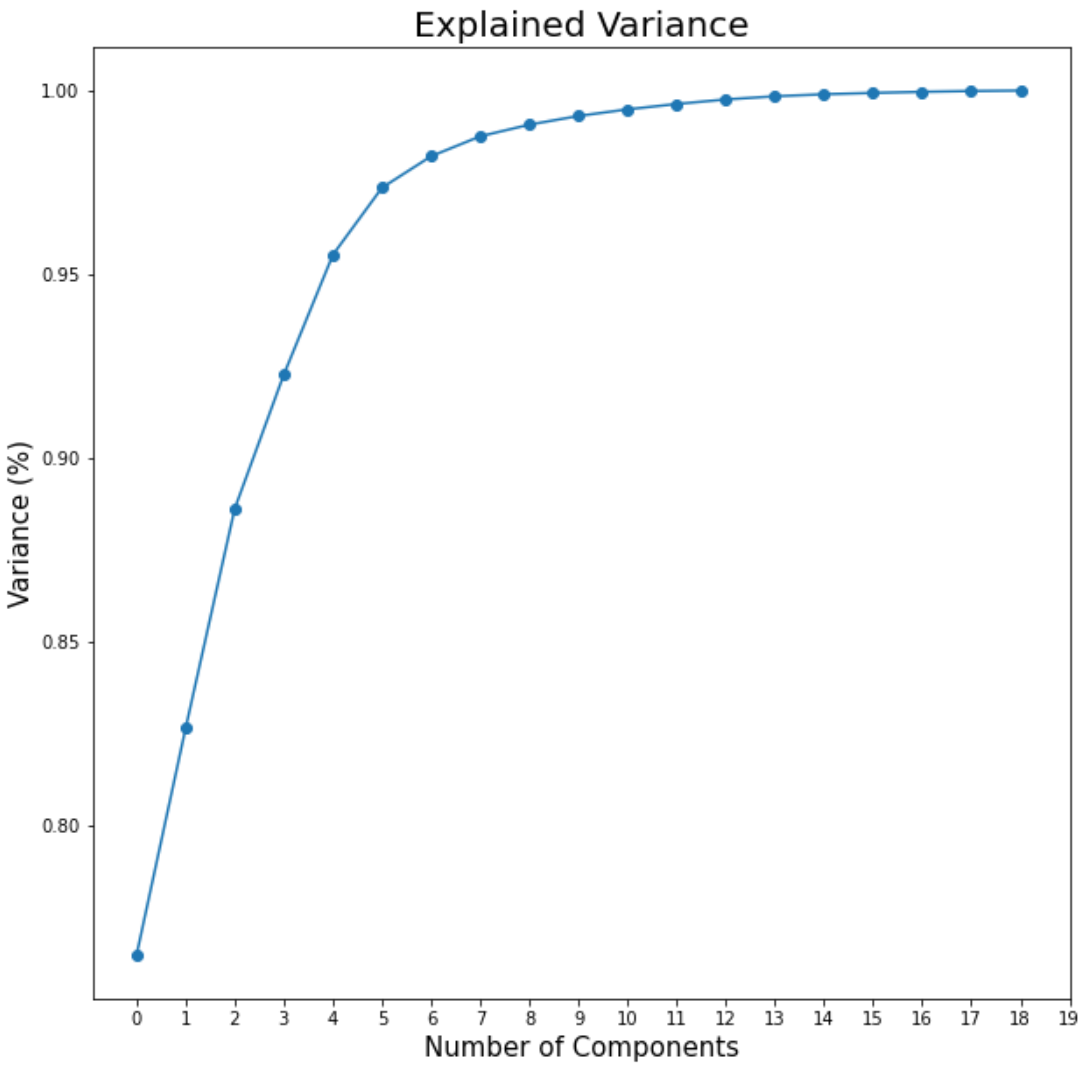


Figure 2: Google Search Trends

Google search trends all follow a similar structure. It is no surprise that specific terms such as bitcoin has the most searches as it is the largest and most popular cryptocurrency on the market.

PCA Analysis

I ran a PCA analysis on Yahoo Stock Prices to reduce the number of features from 30 down to 4 and decorrelate the features. Our PCA analysis showed an optimal number of 4 components that explains 92% variance of the distribution.



Further Research

Recommendations

Final Project report as PDF. The report should clearly explain the problem, your approach, and your findings. Include ideas for further research, as well as up to 3 concrete recommendations on how your client can use your findings. Please give the document an app