Using Principal Component Analysis on Crypto Correlations to Build a Diversified Portfolio

María Guinda
WorldQuant University
mariaguinda@cherrypeaks.com

under the guidance of

Ritabrata Bhattacharyya
WorldQuant University
ritabrata.bhattacharyya@wqu.org

Abstract

A simple look at cryptoassets' historical can lead us think that in recent years most have followed Bitcoin's wake. If so, it would be very difficult to build an exposure to this market without being highly exposed to Bitcoin, and on the other hand a portfolio with many cryptos poses a great operational risk due to the lack of institutional custody. To this aim, this paper presents an updated correlation analysis of 31 crypto assets, among them and with some equity and gold indices. Furthermore, we conduct a PCA to identify the group of cryptos that present different correlation patterns and may help us build a diversified portfolio.

The correlation update shows that these cryptoassets, which account for aprox. 80% of the market, have been positively correlated since 2017 and Ether has been the asset with the highest results. These correlations increase during bear markets, especially in the current bear period started in April 2021. When analyzing Bitcoin against equity markets, we confirmed that correlation is very volatile and swings from positive to negative continuously, which makes it very difficult to use Bitcoin as an equity hedge. As a closing, we have observed that the only times that Bitcoin presented negative correlation with equity indexes coincides with times when gold also showed negative correlation, which could reveal the use of the digital asset as a store of value.

Finally, the PCA show a great number of assets from different category, size and design around a single, highly concentrated cluster. This confirms the great speculation that exists in the market, which moves all the assets en masse. When using the PCA to build a diversified portfolio we achieved better results in terms of return, risk-adjusted return and with a lower correlation to Bitcoin.

Keywords: *crypto market, Bitcoin, correlation, principal component analysis, portfolio construction, portfolio optimization*

JEL Classification: *G11, G12, G14, C13*

1. Introduction

There are over 10,000 crypto assets[1] in the market, and it seems most of them move as one. A quick reading of the historical data can suggest that cryptocurrencies move in unison with Bitcoin, the oldest and largest cryptocurrency. Therefore, many investors looking for an exposure to the crypto market and the technology behind it, blockchain, might only consider investing in Bitcoin. But is this really true? Do all crypto respond equally to external stimuli, such as relevant news or movements in other markets? And what happens if an investor wants to have exposure to the crypto market without being so exposed to Bitcoin? Is it possible to build a diversified crypto portfolio or a hedge against Bitcoin? Are there times or trends in those correlations?

A highly correlated market would show a high degree of speculative investment, which makes everything move at the same time instead of moving independently, driven by each projects' particular circumstances. This would be a clear sign that the market is still very immature and that investors are not moved by the fundamentals behind each project (compliance with milestones, ecosystem improvement, protocol updates, transaction throughput improvement...) but by short-term situations that make the market highly correlated and volatile. A further study of the crypto movements and correlation using Principal Component Analysis (PCA) might tell us whether this coins cluster under fundamental factors (such as privacy coins, platform coins, exchange coins, tokens…) or trading factors (trading volume, price level, liquidity…). This can allow us to identify currencies that have clearly different behaviors and thus build a portfolio with greater diversification factors.

2. Theoretical Framework

The publication of Bitcoin's white paper by Nakamoto (2008)[25] gave rise to blockchain technology. However, Blockchain itself is not a new technology but the result of the intelligent combination of multiple disciplines that already existed; distributed networks, cryptography, game theory and economics. This rare combination has captured the intellectual interest of many scientists, nurturing the sector with great literature ever since.

In general terms, the paper by Dwyer (2015)[4] provides a good understanding around the economic aspect of crypto assets, additionally we supported on Corbet et al. (2019)[3] publication, which with a very well-organized format, provides an exhaustive study on the scientific works that has been published on crypto as a new asset class since Nakamoto's paper.

2.1. Literature review on factors influencing Bitcoin's and other crypto's prices

Bitcoin, Ethereum and the rest of crypto assets developed to form a new investment proposition and hence, many scientists have researched the reason and possible use of this new form of money from an economic point of view. Qi et al. (2020)[13] made an exhaustive analysis of the literature on influencing factors of Bitcoin's price and volatility.

These analyzes reach multiple conclusions, sometimes contrary. Finding influence on both external factors, such as macroeconomic, gold, federal fund rate, Google trends, etc., and internal factors, such as number of transactions, miners hash power, and the difficulty of Proof-of-Work consensus algorithms.

Several studies have concluded that Bitcoin's price movements are not correlated with macroeconomic factors, such as inflation and GDP (Kristoufek, 2013)[8], but to market's online sentiment, as Google trends. This idea has been reinforced by others such as Ciaian

[1] The term crypto asset refers to the coins or tokens, that power a blockchain decentralized network, and can be bought and sold setting market prices. Currently there are thousands of different cryptos and only some are cryptocurrencies, that is, projects that want to become a virtual currency. This is the reason why the correct term when referring to all crypto should be crypto assets instead of cryptocurrencies. In this report, however, we will use the terms currencies, crypto, crypto asset and tokens interchangeably.

et al. (2014)[2], who concluded that the number of visits to certain forums and Wikipedia had a substantial influence on Bitcoin's price, and Polasik et al. (2015)[12], who added the amount and sentiment of news as an influencing factor in Bitcoin's yield. This was later reinforced by Mai at al. (2018)[11], who pointed to social media and investors' sentiment as the main predictors for forecasting Bitcoin and other crypto assets' movement. (Burggraf, Huynh, Rudolf, & Wang, 2020[1]).

However, we find opposite positions. Like that of van Wijk (2013)[16], which argues that there is an important correlation between the price of Bitcoin and macroeconomic indicators. Dyhrberg (2015)[5] too, using a GARCH model, concludes that Bitcoin's price moves aligned with that of the dollar and gold, and found a relatively strong impact by the federal funds rate. The most recent study by Wang, Chen, & Zhao, (2020)[17] also found a significant correlation of Bitcoin with traditional stock market indices, such as Dow Jones and S&P 500. However, Thampanya, Nasir and Huynh (2020)[15] disagreeing with previous studies, concluded that neither cryptocurrencies nor gold served as a possible hedging instrument for the stock market, since their correlations with the market were positive in most cases. All this, therefore, means that adding cryptocurrencies or gold to a portfolio of equities would not improve portfolio's risk-adjusted performance. Despite having conflicting opinions in this regard, the general conclusion is that crypto assets present a considerable higher volatility versus traditional financial assets.

On the other hand, from a purely internal approach and using a least squares regression model, Hayes (2016)[7], suggested that the driving factors for Bitcoin's price were mainly internal; miners' competition level, unit output, and the proof-of-work difficulty.

2.2. Literature review on crypto market correlation

If we analyze the price of the main cryptos in recent years, we see a clear interdependence between them, which in unison follow Bitcoin's path. Price correlation analysis has been a recurring topic of analysis during recent years.

For the present matter we have focused on reviewing literature that centers in the correlation of Bitcoin and other main crypto, between them and with other financial assets. Lahajnar and Rožanec (2020)[10] explored the degree of correlation of the ten cryptocurrencies with the highest capitalization, specifically analyzing the differences during bull and bear markets. According to their findings, these coins are mostly positively correlated over time, decreasing during bull periods, and increasing significantly during bear period. Therefore, a high correlation during the bear market prevents having a diversified portfolio of crypto assets with effective hedge, which must be taken into account when investing funds in this market. Giudici and Polinesi (2019)[6], analyzing how price information flew among different crypto exchanges and between crypto markets and equity markets, concluded that Bitcoin prices are highly correlated between different crypto exchanges, with the largest exchanges acting as price-setters. Moreover, based on their findings, Bitcoin's prices were not influenced by classic market prices; however, their volatilities were, with a negative and lagged effect.

In this line of reasoning, we find another study by Kumar and Ajaz (2019)[9], that through wavelet-based method studied different patterns of cryptocurrency pair movements. From their results, they conclude that correlations followed an aperiodic cyclical pattern, with Bitcoin movements being the main impulse for the rest of the cryptocurrencies. Therefore, they suggested that constructing a portfolio based on multiple crypto assets may be risky at that point of time as the market was mainly driven by Bitcoin.

A high interdependence between the different crypto projects reveals the speculative use that is being given to this new technology. These are mainly moved by sentimental factors, instead of by the advances in blockchain technology on the different projects. We see this hypothesis in the study by Qi et al. (2020) [13] that analyzes the correlation between Bitcoin and the Blockchain equity index, composed of 50 of the most representative blockchain concept stocks. According to their results, there was no significant influence between Bitcoin and the Blockchain Index.

2.3. Other crypto correlation tools and resources

Finally, we draw on some of the reports and price analysis' tools offered by industry analysts and exchanges. Binance produces an analysis on market correlations every year. In their latest Binance Research (2020a)[19] the results showed that largest cryptos presented record-high correlations during 1Q 2020. Bitcoin and Ethereum, the first and second crypto, had a 0.93 among them and 0.81 average correlation with the other 15 largest cryptoassets. Furthermore, for 2019, according to Binance Research (2020b) [20], Ether (ETH) was the highest correlated asset. With an average correlation coefficient of 0.69 throughout 2019, it is consistently among the most correlated assets. During that year, the correlation changes suggested that cryptocurrencies could be more strongly correlated during bear markets and less correlated with sideways movements or bull markets.

More dynamic and visual sources can be accessed through Coinmetrics, Cryptodatadownload and Cryptowat. Coinmetrics (2021)[21] allows us the access to a historical linear representation of the correlation between different cryptocurrencies for both Spearman and Pearson correlation coefficients. This site serves as a starting point to present our initial hypothesis. On the other side, Cryptowat.ch (2021)[24] and Cryptodatadownload (2021)[23] provides correlation heatmaps that can be adjusted for different time periods and correlation type, also providing a good initial visual thought.

2.4. This study purposes

After this review, it is clear that the interest in decoding Bitcoin's behavior and the rest of the crypto market arises as a common factor. A better understanding on risk factors can be of great help to innovative investors who want to add this type of asset to their portfolios. However, either because it is a very young market with little history, or because of the immaturity of the projects that support the currencies, but it is common to find contradictory conclusions.

This study wants to contribute to the existing field and add more clarity. After the last rise in the market at the end of 2020 and beginning of 2021, with the consequent steep decline, this new information may change the conclusions regarding the behavior of the different currencies between them and especially with respect to Bitcoin. Additionally, since the crypto markets have suffered a considerable higher increase in volatility compared to other markets, this study wants to update the conclusions on the interdependence (or not) of this new technology with the main equity indexes and gold. To see if it can really be a tool to diversify a global portfolio. Finally, through the use of PCA we identify possible groups of cryptocurrencies that might be useful in order to build a more stable crypto portfolio or a portfolio that doesn't have as much exposure to Bitcoin.

3. Methodology

As in most quantitative analyzes, success is highly dependent on the information quality. In this study, our first great mission was to choose and obtain historical data for the crypto market that represented a good price reference. We then processed the information to be able to use it with R and carry out a correlation study of the different assets. Next, using results from the correlation analysis, through a PCA we define our assets using only two variables. Finally, we were able to build two investment portfolios based on the PCA groups to test their efficiency against a fully diversified crypto portfolio.

3.1. Collection and treatment of data

When looking for reference prices, the cryptocurrency market, due to its great fragmentation, presents additional difficulties. There are hundreds of different exchanges where the thousands of crypto assets actively trade on a 24-hour base. Prices can vary significantly between different exchanges, especially for the less liquid currencies. That is why, for reference purposes, it is important to choose a good information provider.

We have collected our crypto historical data from *coinmarketcap.com*, an industry information provider that summarizes price data since 2013. In order to pull the data, we have manually accessed the data online, as its API is currently only available by paid subscription.

As presented in *Table 1*, we have selected a mix of cryptos, including the 20 currencies with the highest capitalization (as of 24th June 2021), 5 additional coins (NEO, Tezos, IOTA, Maker and Zcash) selected as being clear fundamental examples in the blockchain environment and 6 additional ones with considerable smaller capitalization but being representative in the Services and Media category tokens (Storj, Golem, Enjin Coin, Steem, BAT and FunFair). This selection excludes stablecoins (coins which are design to maintain a constant price level against USD, EUR or other crypto), coins with less than 1 year of trading history and coins with current price lower than 0.1 (as price history present minimum variations of 0.01, which for these currencies, in percentage terms, distorts the real profitability). If readers are interested in the description of each of the projects, we recommend consulting *Messari* database[26], which provides an exhaustive description of each project, its monetary base, and other technicalities.

Although, there are over 10,500 different cryptoassets listed in *Coinmarketcap* (as of June 2021)[22], our 31 selected assets accumulate 82% of the total market capitalization, which if we exclude stablecoins would increase to 90% of the market. Therefore, we can affirm that with only 31 coins we can represent the crypto market with a high degree of precision.

Coinmarketcap website provides historical tables with prices, volumes and market cap for all cryptocurrencies traded in crypto exchanges. These tables include USD daily values for OHLC prices (open, high, low and close), where prices are a volume weighted average of market pair prices for each cryptoasset.

Crypto assets trade 24-hours, 365 days per year, which means that, contrary to traditional markets, there is no opening and closing trading times. Therefore, when talking about closing price for a day we must define a constant time, and this value will be the opening time for the next day. *Coinmarketcap* sets as the closing price the latest data in range UTC time.

Additionally, each token has been categorized based on five main types: Currency, Infrastructure, Financial, Services and Media & Entertainment. We have used m*essari.io* classification for this purpose. **Currency** coins refer to tokens that are primarily used as money, payments and/or store-of-value, these are Bitcoin, Tether, Dogecoin, and XRP, for example. **Infrastructure** are the native tokens that fuel blockchain projects which have been designed to be the protocol foundation for multiple use-cases. These include Ethereum, Cardano, Polkadot, and Solana, among much more. **Financial** coins are those that support decentralized networks dedicated to cryptoasset financial services, such as crypto exchanges like Binance Coin. **Services** coins refers to tokens which have a specific application implemented in another blockchain-network, usually infrastructure networks such as Ethereum. Our study includes coins like Storj, Enjin Coin and Golem, an Ethereum token that enables a marketplace for computing power. Finally, **Media & Entertainment** tokens are those from decentralized social networks and content creators. Theta Token, included in our study is the native asset of the Theta Network, a protocol which aims to improve the quality of streaming video content.

Table 1. Selected Crypto Assets for Correlation Analysis

#	Name	Ticker	Included	Excluded	Price *(USD)*	Market Cap. *(M UDS)*	Market Share *(%)*	Category
1	Bitcoin	BTC	YES		33,907	635,487	46.6%	Currency
2	Ethereum	ETH	YES		1,961	228,314	16.8%	Infrastructure
3	Tether	USDT	NO	Stablecoin	1.00	62,651	4.6%	Currency
4	Binance Coin	BNB	YES		303.88	46,625	3.4%	Financial
5	Cardano	ADA	YES		1.35	43,087	3.2%	Infrastructure
6	Dogecoin	DOGE	YES		0.24	31,007	2.3%	Currency
7	XRP	XRP	YES		0.65	29,988	2.2%	Currency
8	USD Coin	USDC	NO	Stablecoin	1.00	25,463	1.9%	Currency
9	Polkadot	DOT	YES		16.09	15,354	1.1%	Infrastructure
10	Uniswap	UNI	NO	New Coin	17.41	10,013	0.7%	Financial
11	Binance USD	BUSD	NO	Stablecoin	1.00	9,579	0.7%	Currency
12	Bitcoin Cash	BCH	YES		486.31	9,129	0.7%	Currency
13	Litecoin	LTC	YES		133.00	8,878	0.7%	Currency
14	Solana	SOL	YES		30.18	8,229	0.6%	Infrastructure
15	Chainlink	LINK	YES		18.81	8,136	0.6%	Financial
16	Polygon	MATIC	YES		1.19	7,498	0.6%	Infrastructure
17	THETA	THETA	YES		6.92	6,924	0.5%	Media
18	Wrapped Bitcoin	WBTC	NO	Stablecoin	33,972	6,490	0.5%	Currency
19	Stellar	XLM	YES		0.26	6,124	0.4%	Currency
20	Dai	DAI	NO	Stablecoin	1.00	5,123	0.4%	Currency
21	VeChain	VET	NO	Small Price	0.08	5,062	0.4%	Infrastructure
22	Ethereum Classic	ETC	YES		41.01	4,770	0.4%	Infrastructure
23	Internet Computer	ICP	NO	New Coin	34.81	4,695	0.3%	Infrastructure
24	TRON	TRX	NO	Small Price	0.07	4,683	0.3%	Infrastructure
25	Filecoin	FIL	YES		56.87	4,635	0.3%	Infrastructure
26	Monero	XMR	YES		221.22	3,968	0.3%	Currency
27	EOS	EOS	YES		3.82	3,646	0.3%	Infrastructure
28	Klaytn	KLAY	YES		1.10	2,717	0.2%	Infrastructure
29	SHIBA INU	SHIB	NO	New Coin	68×10^{-7}	2,697	0.2%	Media
30	Algorand	ALGO	YES		0.86	2,664	0.2%	Infrastructure
39	Neo	NEO	YES		34.09	2,405	0.2%	Infrastructure
41	Tezos	XTZ	YES		2.76	2,348	0.2%	Infrastructure
42	IOTA	MIOTA	YES		0.83	2,306	0.2%	Infrastructure
44	Maker	MKR	YES		2,203	2,184	0.2%	Financial
57	Zcash	ZEC	YES		113.25	1,366	0.1%	Currency
70	Enjin Coin	ENJ	YES		1.05	876	0.1%	Services
74	Basic Attention	BAT	YES		0.55	830	0.1%	Media
111	Golem	GLM	YES		0.22	220	0.0%	Services
134	Storj	STORJ	YES		0.69	198	0.0%	Services
163	FunToken	FUN	YES		0.02	176	0.0%	Media
171	Steem	STEEM	YES		0.23	98	0.0%	Media

Source: Market data from coinmarketcap.com. Categories from messari.io. Data as of 24th June 2021. Full table included in Appendix.

The equities and gold data has been collected from *yahoo! Finance* data provider using their webpage. As shown in *Table 2*, we have selected 5 equity ETFs indices and one gold ETF. For the equity indices, we have selected representative ETF, as in order to compare real performance we must use investable assets. Therefore, we have selected ETFs with high liquidity and standard costs that follows the different indexes with sufficient efficiency.

Table 2. Selected ETFs for Equity and Gold Indexes

Index	Category	ETF Name	Ticker	Code
MSCI WORLD INDEX	Global Index	iShares MSCI World ETF	URTH	World
S&P ASIA 50 INDEX	Asia Index	iShares Asia 50 ETF	AIA	Asia
EURO STOXX 50 INDEX	Europe Index	iShares Core EURO STOXX 50 UCITS ETF	EUEA.AS	Europe
DOW JONES	USA Index	SPDR Dow Jones Industrial Average ETF	DIA	DowJones
NASDAQ 100 INDEX	Technology Index	Invesco NASDAQ 100 ETF	QQQ	Nasdaq
GOLD	Gold	SPDR Gold Shares	GLD	Gold

All market data has been treated using Excel (workbook with original and treated data can be consulted in the *GitHub* repository). We have made three groups of study data. The first group (*Crypto Group*) includes only crypto data; daily price since 2017. The second group (*Equity Group*) includes all crypto and equity indexes data. The last group (*Bitcoin Group*) includes only Bitcoin and the equities data. We have worked with daily values for both crypto and equities, where price returns have been computed using log return, as shown in *equation 1*:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln\,(P_{t-1}) \tag{1}$$

Where P_t represent the asset price for that period and P_{t-1} the price level from previous periods. When computing daily returns, these prices are closing prices from subsequent days.

3.2. Period selection

For the correlation analysis we have divided our time span into six periods defined by the dominance of the crypto bull and bear market. *Figure 1* illustrates Bitcoin's price evolution and the resulting bull and bear periods. According to our criteria, to be defined as bull (or bear) market, based on Bitcoin's price history, we consider a market change when:

- Bitcoin's price changed more than ±20.0%.
- For the next 90 days, Bitcoin's price doesn't return to the top or bottom before this change.

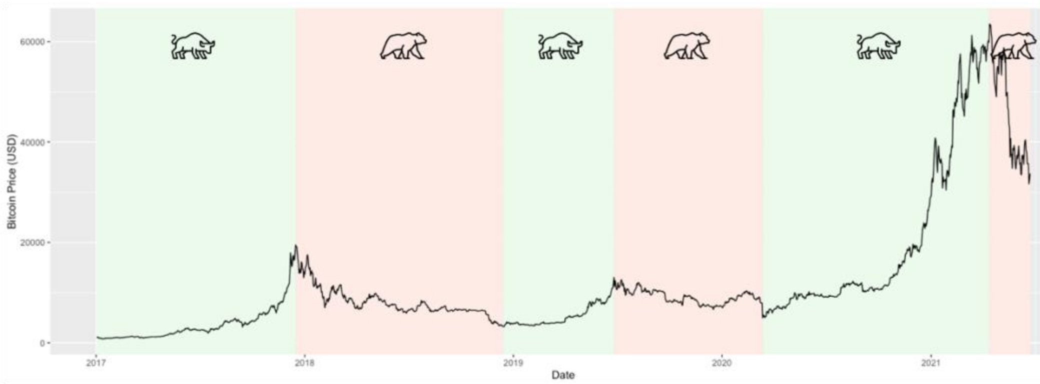

Figure 1. Bitcoin's Price Evolution and Bull and Bear Market Identification

We additionally included eight study periods with relevant macro-economic impact to evaluate the potential play of Bitcoin against other markets and Gold.

Table 3. Selected Study Periods

PERIOD	START	END	Lasted *(days)*
TOTAL	2017-01-03	2021-06-23	1,632
BULL Crypto 1	2017-01-03	2017-12-16	347
BEAR Crypto 1	2017-12-17	2018-12-14	362
BULL Crypto 2	2018-12-15	2019-06-26	193
BEAR Crypto 2	2019-06-27	2020-03-12	259
BULL Crypto 3	2020-03-13	2021-04-12	395
BEAR Crypto 3	2021-04-13	2021-06-23	71
Brexit Vote	2017-06-01	2017-07-01	30
COVID 1Q20	2020-01-01	2020-04-10	100
Trump First Impeachment Trial	2020-01-16	2020-02-15	30
UK Leaves EU	2020-01-31	2020-03-01	30
Bitcoin Halving	2020-05-18	2020-06-17	30
Biden Victory	2020-11-03	2020-12-03	30
Capitol Attack	2021-01-06	2021-01-26	20
Suez Canal Obstructed	2021-03-23	2021-04-22	30

3.3. Correlation analysis

A correlation analysis is used to evaluate the grade of relationship between two or more variables. When applying to assets performance, our principal goal is to find whether the price movement of an asset is influenced by other asset's price movement; do they tend to move together? Do they tend to move in opposite directions, rising one when the other drops value? How strong is this relation?

For our analysis we have used Pearson correlation coefficient, which is a parametric test (depends on the data's distribution) that measures the linear dependence between two numerical variables. The coefficient is calculated as the result of the following formula:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \tag{2}$$

Where $\bar{x}$ and $\bar{y}$ correspond to the variables' mean values, and x_i and y_i are the variables' sample.

The result is a coefficient that range from -1 to +1. Where -1 denotes strong negative correlation, which means that when asset *x* increases in value, asset *y* decreases. 0 indicates that there does not exist an association between the two variables. And a +1 coefficient denotes a strong positive correlation, this means that asset *x* and *y* tend to move in the same direction, rising together and falling together.

The limits to define the Pearson correlation results were initially established by Cohen (Cohen, 1988)[18] and improved by Rosenthal (Rosenthal, 1996)[14]. From their definitions we can interpret a less than 0.1 correlation coefficient (in absolute value) as a **weak** relation, 0.1 to 0.3 as **moderate**, 0.3 to 0.5 as **strong**, and over 0.7 as a **very strong** relation.

In order to visually represent our results, we have used correlation heat matrix, which using a color code (in our study red for positive correlation, blue for negative correlation, white for no correlation) shows in a single matrix the degree of correlation between multiple assets. Allowing to identify and compare between matrices in a more visual way.

We have also analyzed average coins' correlations for each period, which for example would tell us the average correlation of Bitcoin with the rest of cryptoassets during a certain period. This is computed as:

$$\overline{r_i} = \frac{\sum_{j=1}^{n-1} r_{ij}}{n} \tag{3}$$

Where r_{ij} correspond to the Pearson correlation coefficient (for a certain period) of asset i with the rest of assets (represented as j), and n is the number of assets in the study.

Lastly, we have analyzed the historical correlation of various pairs of assets. Rolling correlation is simply calculating a correlation coefficient as a rolling window calculation. Therefore, if we choose a 50-day correlation window we will obtain for each day the correlation coefficient of two assets for the previous 50 days. Accordingly, by applying a rolling correlation of 50 days to the entire period, we can visualize how the relationship between two assets has changed throughout the period. This technique is widely used in time series forecast as it can detect shifts in trends, detect forecasting errors and signal special events.

3.4. Principal Component Analysis (PCA)

Once we have calculated the correlation matrix, we obtain a data table with *n* different assets defined by *n* variables. Same number of components and variables. Where each variable is the correlation coefficient of that asset with the rest of asset in our study. Principal Component Analysis (PCA) is a powerful technique that allow us to reduce the number of variables and visualize our assets (components) using the two best dimensions that better represent our data.

There are two typical methods to perform PCA using *R*; i) spectral decomposition, which is based on variables' correlation, and ii) singular values decomposition (SVD), which is constructed from the observation's covariances. We have used the R function *prcomp()* which uses SVD and presents marginally better numerical accuracy.

As our variables are assets' return correlation, therefore, this PCA study allow us to identify assets that have similar behaviors in terms of correlation with the rest of cryptoassets, detecting if there is a common pattern by category or other circumstance. In this way, we can identify different groups of assets with the same correlation patterns, and this might help us build a better diversified portfolio, since we can avoid having too many assets from a single group and thus too much exposure to one type of correlation.

Using this PCA analysis we have plotted our assets using the first two principal components and identify them in different colors as their natural category shown in *Table 1* (Currency, Infrastructure, Services, Financial and Media). Afterwards, we have checked whether the groups are kept separate or presented an overlap, which would indicate that the correlation patterns between groups are mixed. Once the groups are displayed, we can identify the asset that better represent the group, these are the one that appear closer to the group's center.

Finally, we have used a K-means clustering technique, a data grouping method which aims to distribute a set of observations into *k* defined number of groups, in which each observation goes to the group whose mean value is closest. This technique, frequently used to perform unsupervised learning tasks, allow us to analytically identify the groups of crypto that present same correlation patterns. Finally, we have chosen the assets that better represent each cluster.

3.5. Portfolio construction

In order to test our analysis, we have built three portfolios. Our first portfolio includes the 31 selected cryptoassets. This, theoretically, represent the most diversify crypto portfolio. Additionally, using the results from the PCA we have built two portfolios. One includes five assets, each being the one that best represents the different crypto categories as per our PCA analysis. And the third portfolio includes 3 assets, being the result of the three clusters under the K-means analysis.

In traditional markets, when building a diversified portfolio, it is very common to use indexed ETFs, which allow us to obtain exposure to a certain market or sector. However, in the crypto market, these investible indexes do not exist, so diversifying involves buying many assets. Which, compared to traditional markets, implies an additional operational risk. While for traditional financial assets we can use a custodian to keep our investments, in crypto there are still no accredited custodians to secure our money. Crypto exchanges are companies that facilitate crypto investments and keep our assets (acting as custodians); however, these companies have no guarantees or insurances[2] to respond in the event of theft or hacking of their systems. The alternative, which is to save the investments ourselves, supposes a great operational risk, that increases with the number of assets, since each crypto involves different security measures that we must know when working with them. That is why building a diversified portfolio with few cryptocurrencies can be a great advantage against the market.

Each portfolio has been evaluated during the same period ("Test Period"), using the correlation results from the previous year ("Train Period"). In our study the training period, or calibration period, goes from Aug-2019 to Aug-2020, and the Portfolio Testing Period goes from Aug-2020 to July-2021.

The investment exposure to each asset is the same at the beginning of the investment, so that in portfolio 1 each asset begins with an exposure of 1/30[3] of the total portfolio, in the second portfolio each asset begins with a weight of 20% and in the third portfolio a 33% exposure per asset. These exposures are not rebalanced during the testing period.

Finally, we compare all portfolio performance, drawdown, standard deviation, Sharpe Ratio and correlation to Bitcoin and the different equity indexes to see whether the PCA diversification method results in a more efficient investment.

3.5. Code repository

All these computations have been performed using *R* programming language and *RStudio*. The relevant libraries and sources can be accessed through the *GitHub* repository at *https://github.com/cherrypeaks/Crypto_Correlation*.

3.5. Interactive online tool

Additionally, we have adapted the code and made an interactive online tool available to everyone, without the need for programming knowledge. This tool includes all the functions that we have used to prepare the study. The application can be accessed through the following link. *https://cherrited.shinyapps.io/Capstone_MariaGuinda/*

The application is organized in three tabs. The first two tabs have dynamic tools for correlation analysis. With this application we intend to complete the services offered by *Coinmetrics* and *Cryptowat* by adding more currencies, functionalities and adaptability. The last tab includes the PCA analysis and a summary of the portfolio construction. It allows us to apply the analysis to different periods of time and modify the number of clusters for a more optimal use.

This application is informative and not intended to serve as an investment advice.

[2] There are currently a couple of exchanges that already have anti-theft insurance. However, these insurances only cover hot wallets and account for approximately 10% of the funds that the exchange hold.
[3] Although the correlation study includes 31 assets, in this last analysis one of the assets was excluded because it did not present enough historical information in the calibration period. The model does these checks automatically.

4. Results and Discussion

Throughout this section we present our study findings, starting with a description of the current and historical crypto market correlations and ending with the construction of a diversified crypto portfolio resulted from a PCA analysis.

4.1. Correlation heatmaps for each period

Our first analysis covers the 25 first selected crypto (excluding the 6 least liquid assets) during the full study period, from 2017 until June 2021. As shown in *Figure 2*, red dominates our map, showing a clear positive correlation of all assets during the full period. Coefficients range from 0.09 to 0.70. During this period ETH showed the highest average correlation with the rest of assets, with a 0.56 coefficient. During this time there were four assets that presented a clear nonrelation with the rest of crypto (or very small), which translates into white columns in our matrix. These were MATIC, FIL, KLAY and MKR.

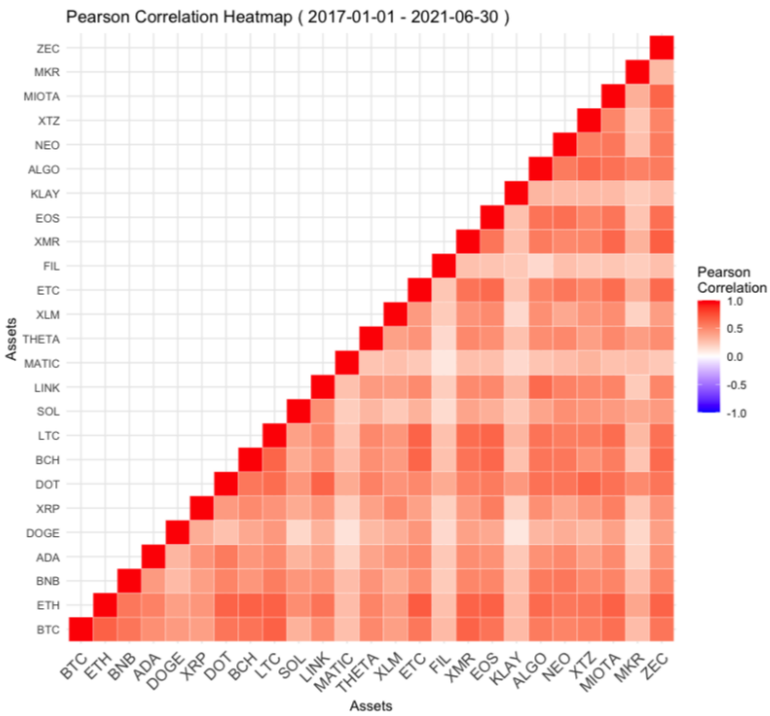

Figure 2. Correlation Heatmap for the Full Period

Our next analysis was aimed to identifying correlation changes between the bull and bear markets. In *Figure 3* we present the correlation matrices for the six bull and bear periods since 2017 (as defined in *Section 3.2.*). The results confirm that the correlation in the crypto market increases during bear markets, being much more pronounced in the last market change that took place in April 2021. In *Figure 4* we have summarized in a single table the average correlations of each token with the rest during the different periods. From this table we see a clear increase in the red color during bear periods.

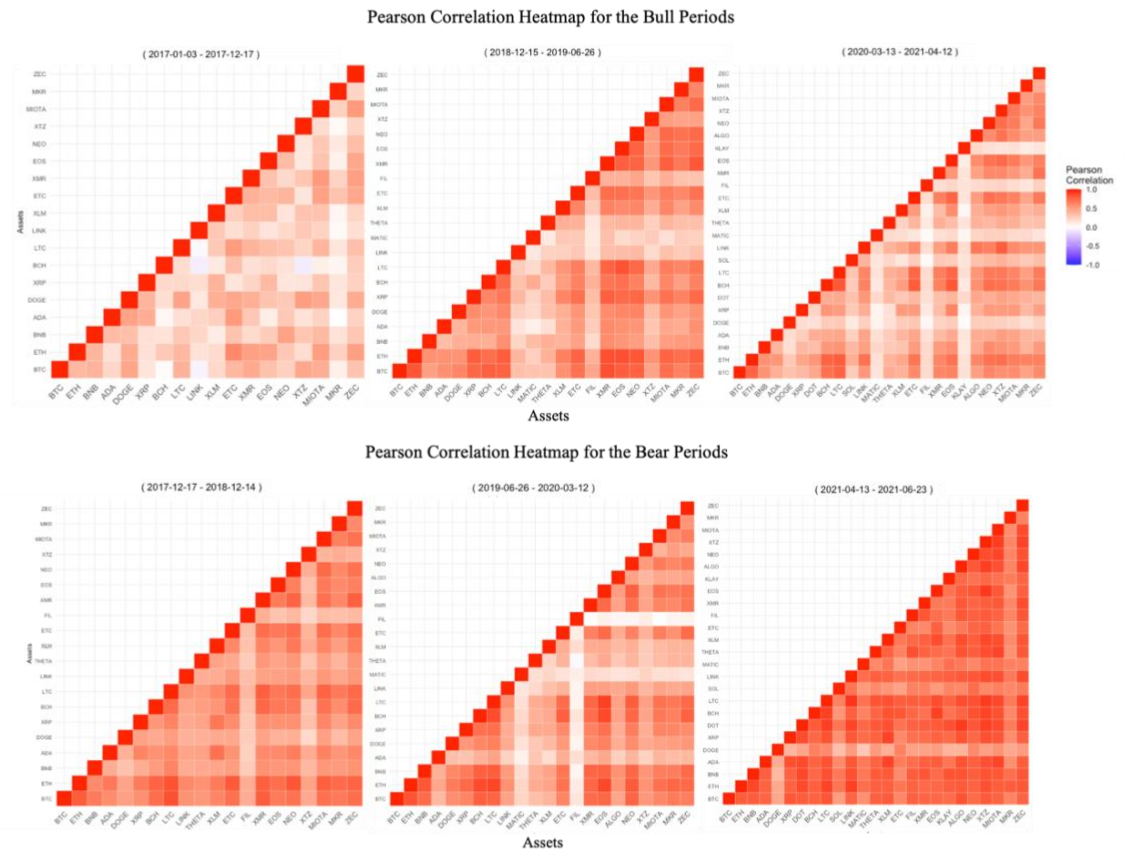

Figure 3. Correlation Heatmap for the Bull and Bear Periods

The numerical results showed that during the first bull period (Jan17 to Dic17) average correlations ranged from 0.15 to 0.38, the second bull period (Dic18 to Jun19) coefficients ranged from 0.23 to 0.65 and during the last bull period (Mar20 to Apr21) the average coins' correlations remained in the range of 0.13 to 0.55. And total average, computed as the average correlation of all pairs during the period, were 0.30, 0.50 and 0.4 for the first, second and third period respectively. During these periods Ethereum showed the highest average correlation with the rest of assets, slightly above of Bitcoin.

For the bear periods instead, results are clearly higher, with a total average correlation that went from 0.55 in the first bear market, to 0.47 during the second and 0.73 during the current bear market. Throughout the first bear period (Dic17 to Dic18) average coins' correlations ranged from 0.32 to 0.68, the second bear period (Jun19 to Mar20) results varied from 0.17 to 0.64 and during the current bear market (Apr21 to present) the average correlations have gone from 0.47 to 0.79. Highest coefficient came from Ethereum during the first and second periods and Litecoin in the current period, closely followed by IOTA and ZEC.

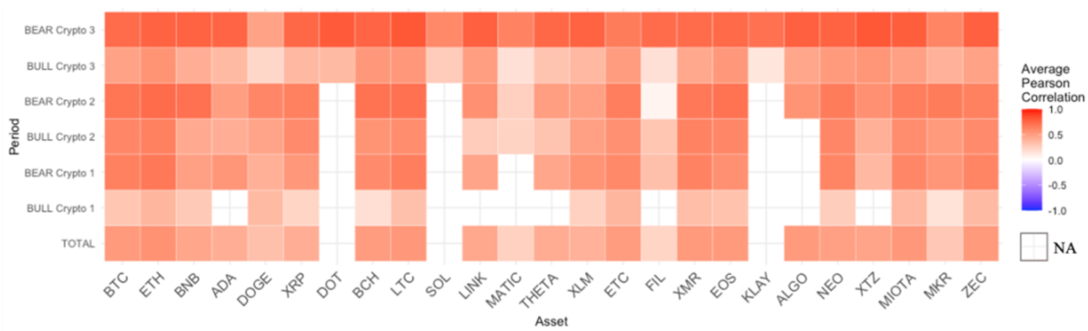

Figure 4. Average Correlation for Bull and Bear Periods

We can therefore confirm that the correlation between cryptocurrencies during bear markets has been on average higher than the correlation during bull markets.

Our next analysis focused on the relation of Bitcoin, the cryptocurrency with the highest capitalization, versus some traditional markets during certain time periods. *Figure 5* summarizes the correlation heatmap for these periods.

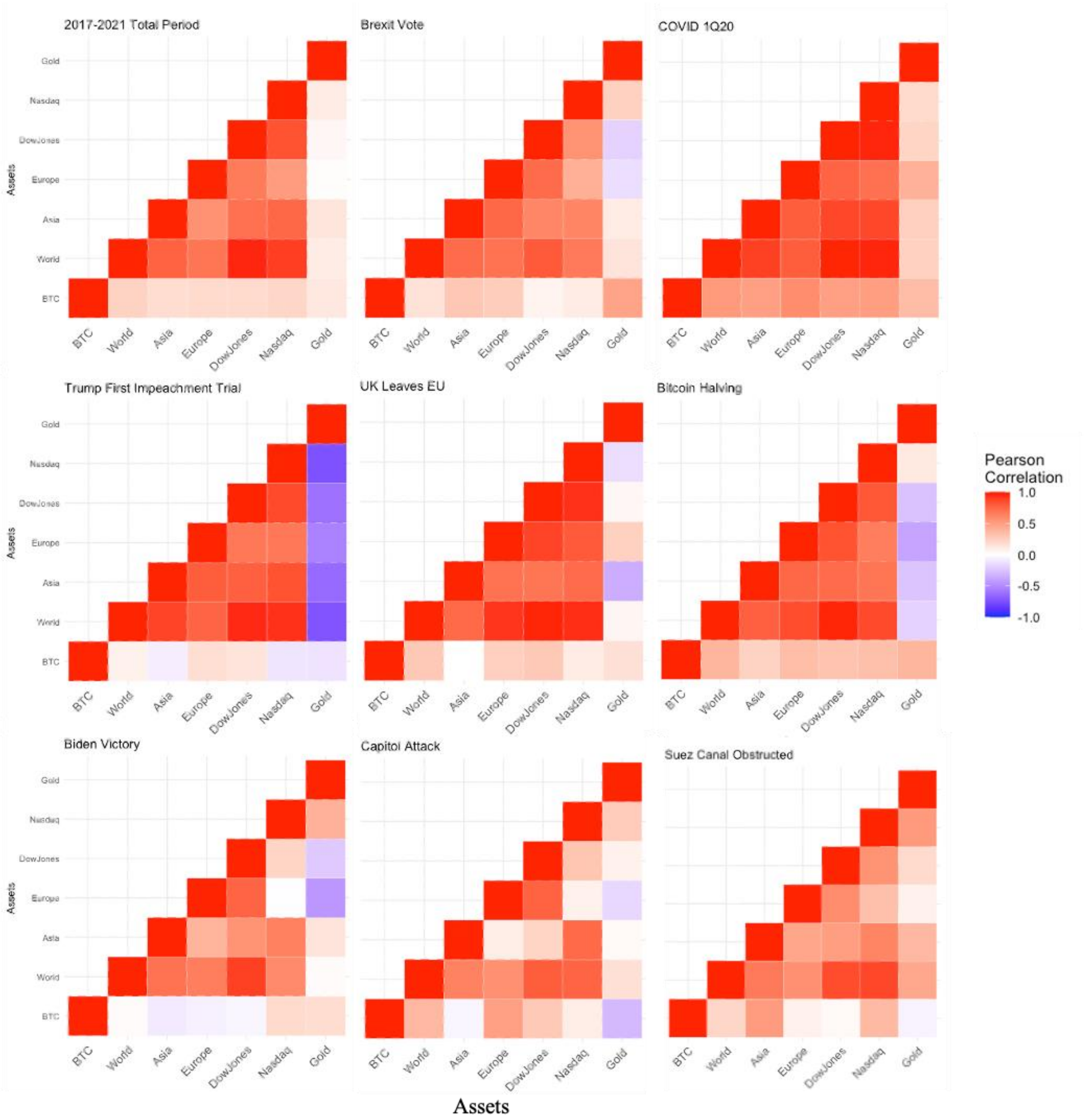

Figure 5. Pearson Correlation Heatmap for Relevant Macro Periods

For the full study period (2017 to June 2021) Bitcoin revealed a weak positive correlation of around 0.2 with the equity indexes in our study, and 0.12 with gold. During these particular macroeconomic events Bitcoin reveals a weak positive correlation with every index, except for the COVID19 period (1Q2020) where all assets had a moderate and strong correlation across the markets.

On the other hand, the relation with gold has remained very volatile, fluctuating from negative 0.3 during Capitol Attack to positive 0.47 during Brexit vote. Interestingly, the only times that Bitcoin has shown a negative correlation with any equity index coincide with times when gold also showed a negative correlation. This fact could reveal the use of Bitcoin as a store of value against traditional markets crises.

4.2. Moving correlation

Our last correlation analysis consists of a rolling correlation among various pairs. This visual analysis let us identify trends in correlation during different crypto market sentiments. In *Figures 6, 7, 8, 9 & 10* we have plotted the 50-days Pearson correlation coefficient over the different bull and bear crypto markets.

Figure 6 represents the relation of Bitcoin and Ethereum since 2017. Although the correlation for the full period has been 0.68, correlation since mid-2018 has been 0.82, which denotes a very strong positive correlation. It is worth highlighting that these two currencies account for 70% of the crypto market capitalization, and historically they have accounted for an average of 80% of the entire market.

The two correlation drops in mid and late 2017 in this BTC-ETH pair was due to a rally in Ethereum versus Bitcoin, which acquired greater attention and popularity as the project progressed. Since 2018, however, the two currencies have followed a very similar path. These trends can be visualized in *Figure 7*, which represents the value of a $100 investment in both Bitcoin and Ethereum as of 1st January 2017. Apart from these falls, the rest of the period shows a high correlation with slight occasional rises and falls, both during bull and bear market, which denotes not dependency on market sentiment.

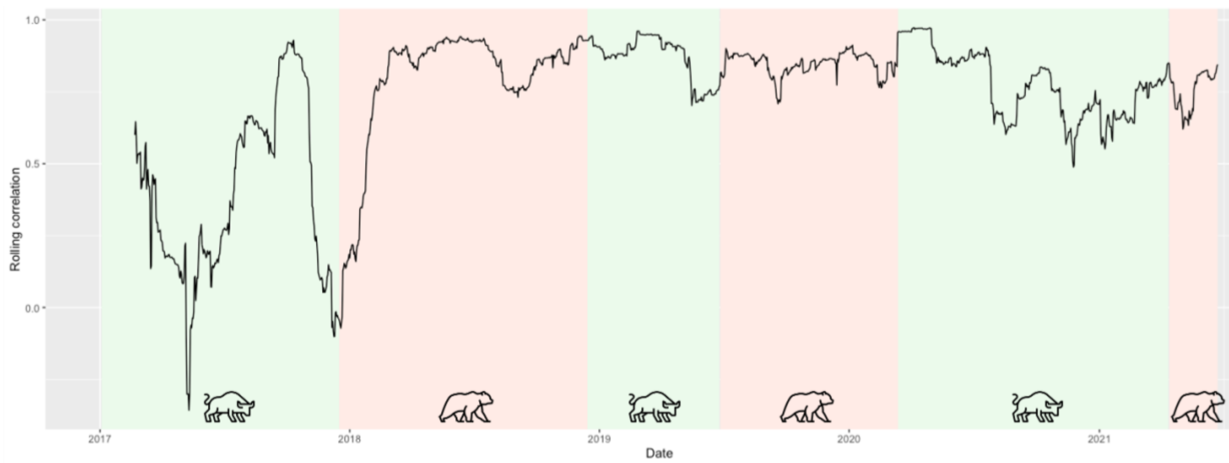

Figure 6. Rolling 50 Days Correlation for BTC and ETH

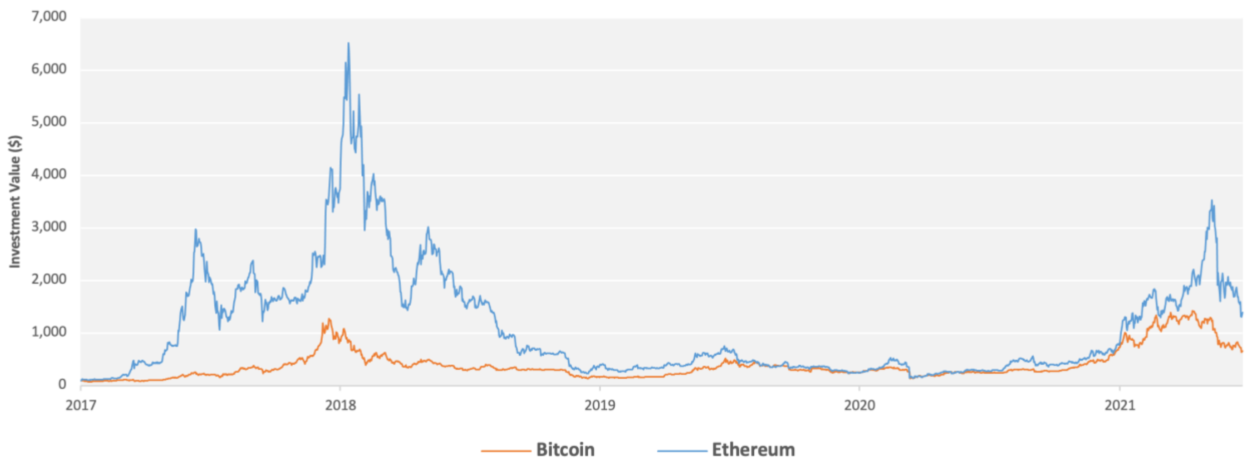

Figure 7. Investment value of $100 in Bitcoin and Ethereum at 1st Jan 2017

Figure 8 shows the relationship of Bitcoin and a gold ETF since 2017. We can see how the correlation between these two assets has remained very volatile during the six bull and bear periods. The relationship has swinged from negatively moderate to positively moderate and strong. From this pair we would highlight a higher mean correlation during bull markets, and an increase in correlation that occurs when the market changes from bear to bull.

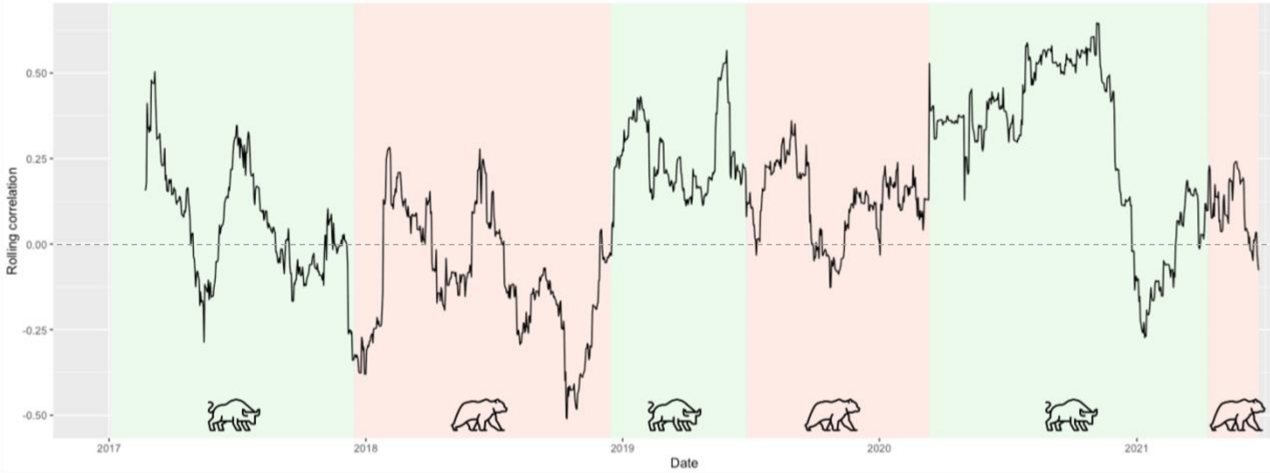

Figure 8. Rolling 50 Days Correlation for BTC and Gold ETF

Finally, we analyze the historical relationship of Bitcoin with some of the most important stock indices, where the conclusions have been the same for all the analyzed markets. In *Figures 9, 10 and 11* we have represented this relationship for a Dow Jones, Nasdaq and an Asian ETF index. Although the Pearson correlation of Bitcoin with these markets since 2017 was 0.19, 0.21 and 0.18 respectively, the 50-day correlation has been continuously varying from negative to positive. Consequently, we cannot conclude a relationship between Bitcoin and capital markets, and contrary to the conclusions of previous studies, the results show erratic and changing relationships that denote the little interconnection between these markets.

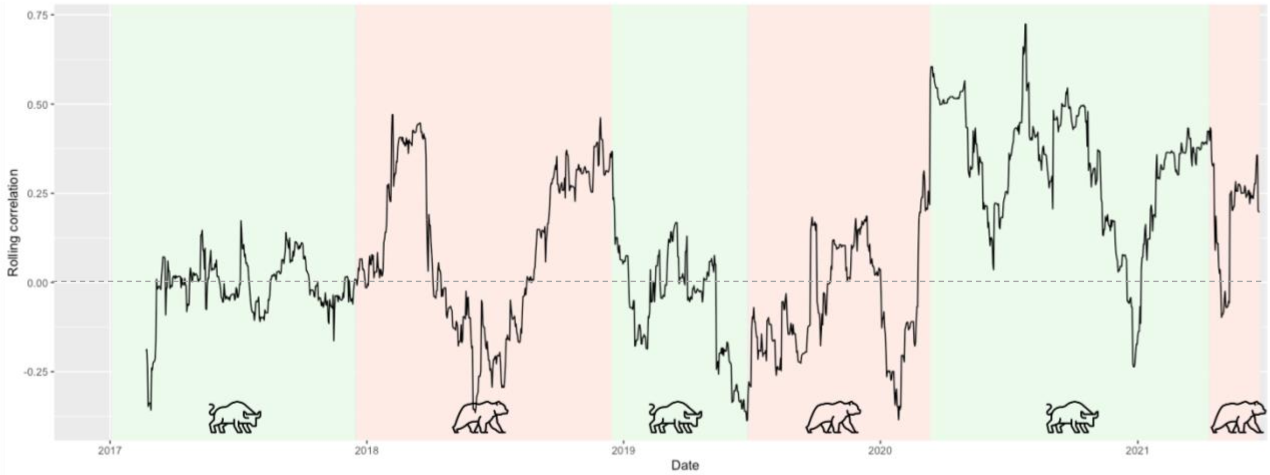

Figure 9. Rolling 50 Days Correlation for BTC and Dow Jones Index ETF

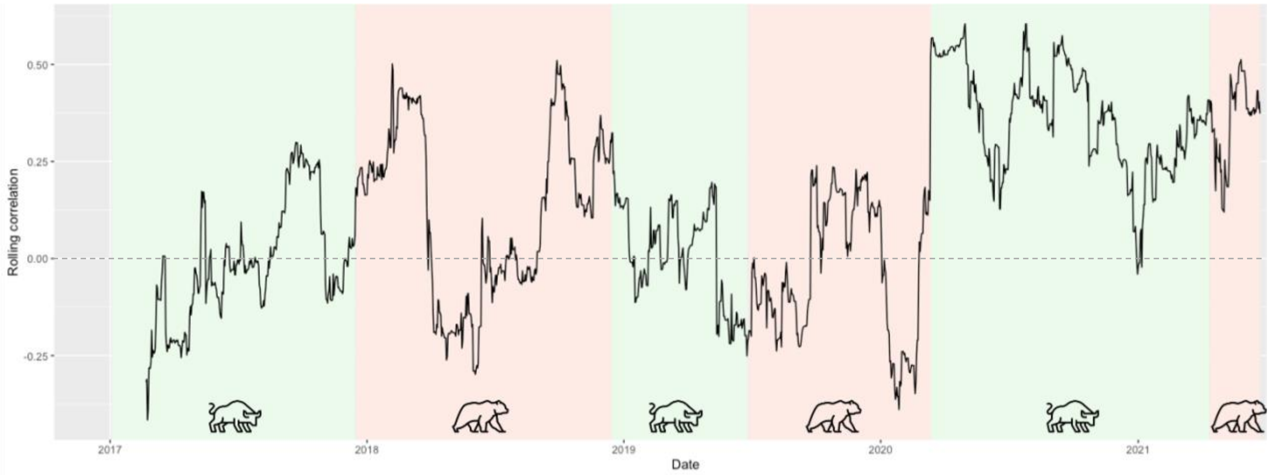

Figure 10. Rolling 50 Days Correlation for BTC and Nasdaq Index ETF

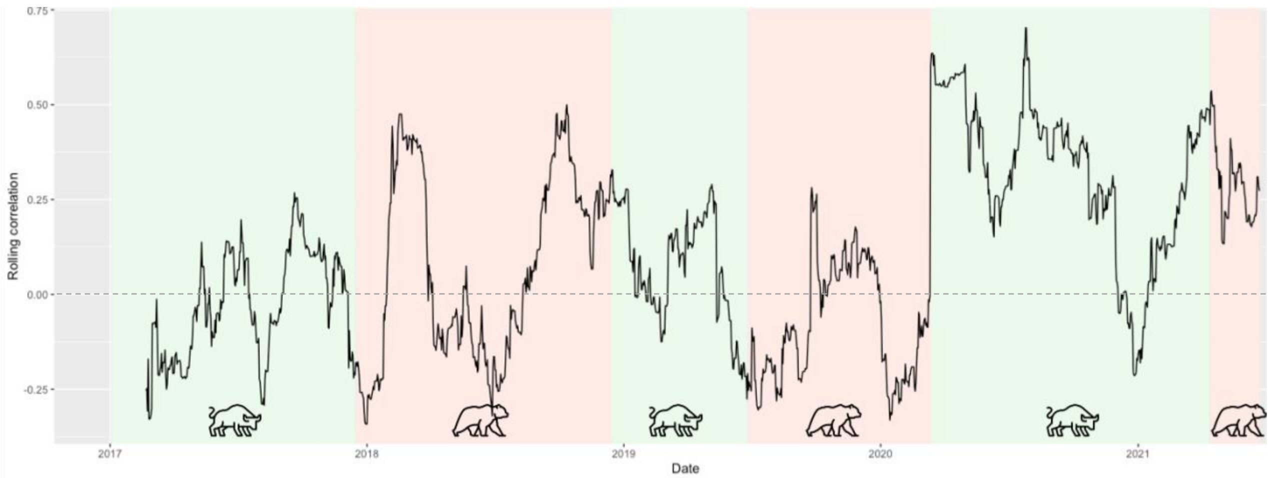

Figure 11. Rolling 50 Days Correlation for BTC and Asian Index ETF

4.3. Principal Component Analysis

Once we have the information on correlations, we can represent each crypto asset based on its correlation with the rest of the coins in the study. Before building the correlation matrix we checked if the calibrating period included enough data points for each asset. For our analysis it resulted in only one discard, Polkadot. Therefore, our study included 30 assets. Taking into account that we have included 30 crypto, that means a total of 30 variables that describe each token (29 correlation pairs plus a value of 1 for their own correlation). However, visualizing 30 variables and drawing conclusions from their possible relationships is very difficult. That is why, with the help of the Principal Component Analysis (PCA), we have reduced the number of variables while maintaining plenty information and help us draw conclusions about the correlations between assets.

Our first step was applying PCA to the set of information. The set was composed of the correlation results of each asset with the rest of the assets during the training period (Oct-2019 to Oct-2020). We then represented the eigenvalues to check what was the percentage of variation explained by each of the new calculated variables. As we can see in *Figure 12*, the first dimension represents 71.7% of the variation and the second 5.2%. The rest of the dimensions represent decreasingly lower values. This means that with only two new variables (*Dim1* and *Dim2*) we can represent 76.9% of the information contained in the 30 correlation variables. This gives us clues about the behavior of assets, which seems that despite showing complex relationships between them, they will follow a unique marked pattern.

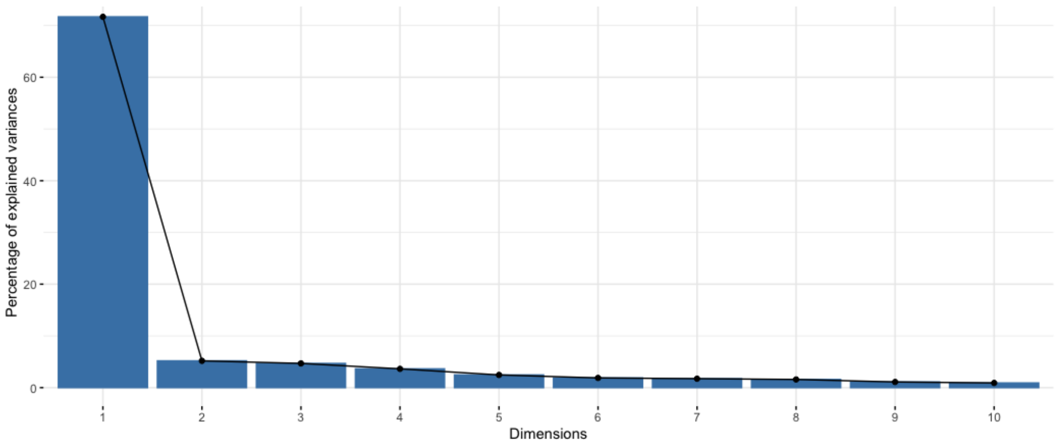

Figure 12. PCA: Percentage of Variances Explained by Each Component

Before representing our assets based on our new two variables (*Dim1* and *Dim2*), we will visualize how the different 30 correlations build these new dimensions. *Figure 13* shows the degree of contribution to *Dim1* and *Dim2* using a scaled of 1. Each variable is represented as a vector (eigenvector) that starts from the center. The longer the vector, the more that variable contributes to explain our new variables. Therefore, the correlation coefficient with MATIC (blue vector) does not represent much the behavior of our assets. While XMR (red vector) has a high degree of representation.

In addition, this graph, also called the influence graph, allows us to see the relationship and intensity that exists between the variables and the new dimensions. The closer an eigenvector lays over *Dim1* and *Dim2* axes, the greater its influence on these new variables is. Therefore, when analyzing *Figure 13*, we see a great negative influence of most of the variables with *Dim1*. This grouping of variables denotes their redundancy, and in our case these alignments indicate that the correlations between cryptoassets are related and follow a single pattern. This large grouping in a single direction also tells us that when representing the assets using only variables *Dim1* and *Dim2* as information, most of the assets will lay grouped very close to the negative *Dim1* axis.

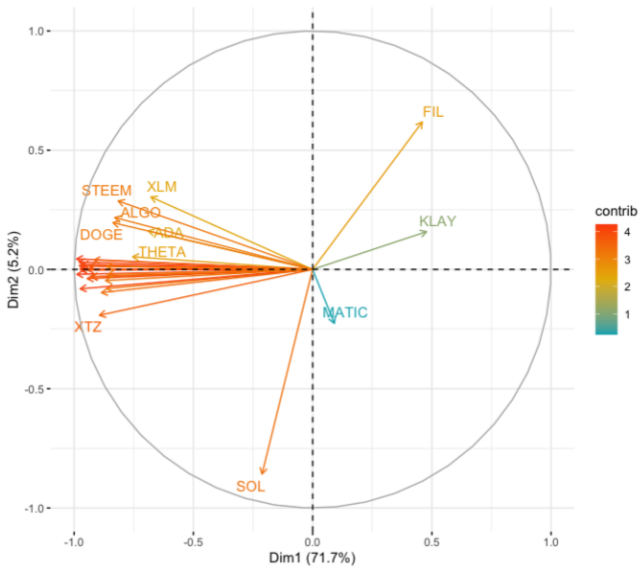

Figure 13. PCA: Variables' Contribution to Principal Components 1 and 2

This contribution per dimension can be represented in a single graph, in which the projection of each vector on the *Dim1* and *Dim2* axes is translated into a single value. As we see in *Figures 14* and *15*, our conclusions from the influence graph coincide with the results of these graphs. Dimension 1 has great influence by many of the variables, but in dimension 2 we clearly see the influence of SOL, which builds over 50% of *Dim2*, and FIL, contributing with 25%. We could also identify these facts represented in the influence graph from *Figure 13*.

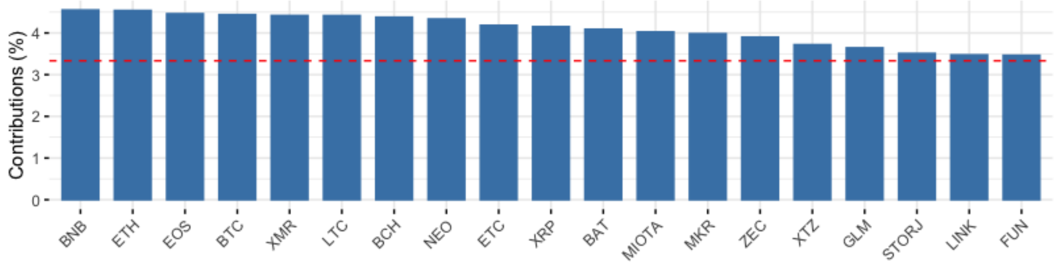

Figure 14. Contribution of Each Variable to Dimension 1 (Percentage terms)

Figure 15. Contribution of Each Variable to Dimension 1 (Percentage terms)

Our following graph, *Figure 16*, plots in a single representation the variable contribution and the assets represented by *Dim1* and *Dim2*. As we had expected, most assets have a high negative component of *Dim1*, since most of the variables that helped building *Dim1* laid over that axis. This, therefore, tells us that the relationships that a crypto *A* price has with another crypto *B* is very similar to the relationship it has with *C* and *D*. And therefore, their returns are closely related.

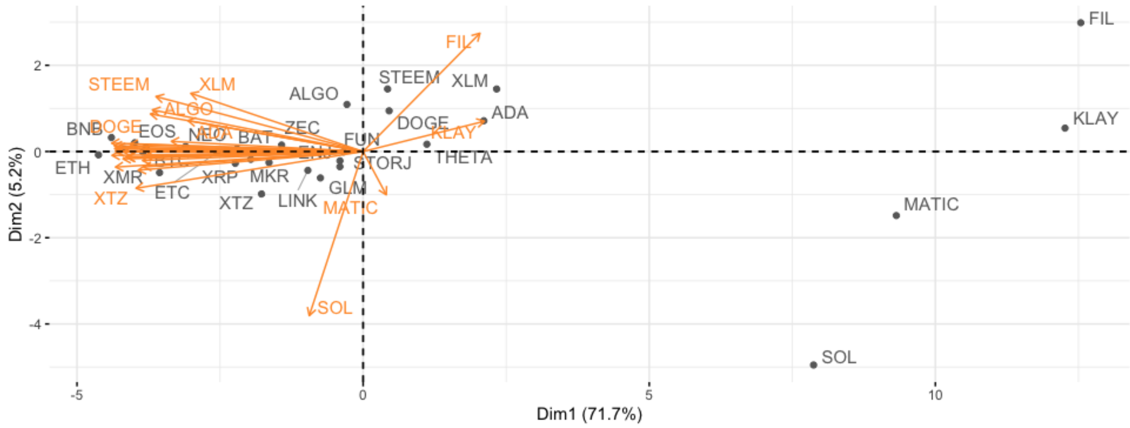

Figure 16. PCA: Variables and Elements as of PC1 and PC2

Focusing only on the points of the graph in *Figure 16*, which represent the different crypto in our study, if two points appear together on the graph means that their price return in relation to other cryptos is very similar. That is, they are affected in the same way by how the rest of the market moves. And therefore, if we want to build a diversified portfolio, we should not invest in points that are very close together, since the investment would behave the same way. Hence, in order to build a portfolio with a greater diversifying effect, it would be necessary to invest in points that are far apart, since these assets will respond differently to the same market changes.

When we pose this study, our initial hypothesis was that when representing the different assets under the PCA analysis the assets should be grouped by type. This hypothesis stems from the belief that different news and facts would affect crypto groups differently. As in traditional markets, a rise in oil has different effects on different industries and stocks. For this reason, in the following graph, *Figure 17*, we represent our assets grouped by their different design categories (Currencies, Financial, Infrastructure, Media and Services). Our surprise was that, far from appearing in differentiated groups, the sets present a great overlap. This means that market movements affect assets of different categories in the same way, and therefore diversifying based on categories should not result in a good hedging portfolio.

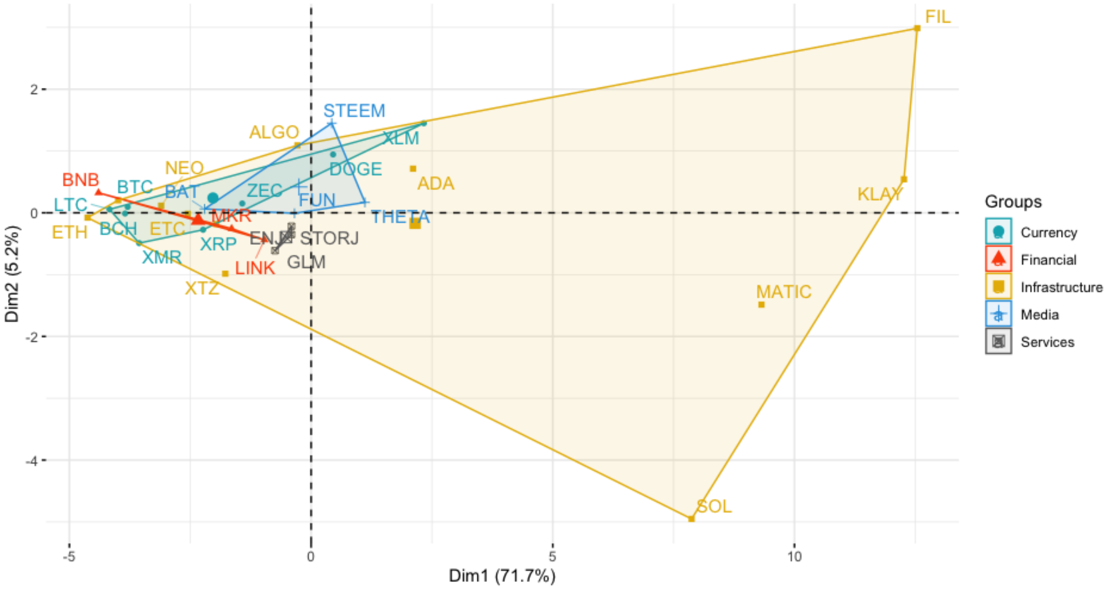

Figure 17. PCA: Elements Association based on Categories (PC1 and PC2)

Despite the great overlap, we decided to build a portfolio (we will call this **Portfolio 2**) that includes one asset from each category. The way to choose that asset is by selecting the element of each group that is closest to the center of the group. The center of the group is calculated as the mean of all the items in that category. The result, which shown in *Table 4*, includes XRP, MKR, ADA, FUN, and ENJ.

Table 4. Portfolio 2 Constituents by Category Grouping

Group	Principal Crypto	Distance to Group's Mean
Currency	XRP	0.55
Financial	MKR	0.70
Infrastructure	ADA	0.89
Media	FUN	0.44
Services	ENJ	0.65

To finish this PCA analysis, we are going to group the items using the K-means clustering method, which will place each asset into the group whose mean is closest to. By choosing three clusters we obtain a very differentiated first group (blue) that includes FIL, KLAY, MATIC and SOL, four recently created infrastructure tokens which together account for less than 2% of market dominance. The second group (red) includes 11 assets of different categories but which, except for ADA and DOGE, are assets with a market share of less than 0.6%, and therefore assets with little market weight. The sum of this group market dominance is 7.3%. Our latest cluster (yellow) is also the most numerous and largest in terms of capitalization. It includes the two heavyweights BTC and ETH, which together account for 63% of the market, but the rest of the group's assets account for almost 10% of the market. Furthermore, when analyzing *Figure 18*, we must emphasize that this last group, despite containing the largest number of assets, occupies the smallest surface on our *Dim1-Dim2* graph. Which suggests that cluster 3 includes assets whose price behavior when analyzed as a reaction to the rest of the market (correlations) is very similar.

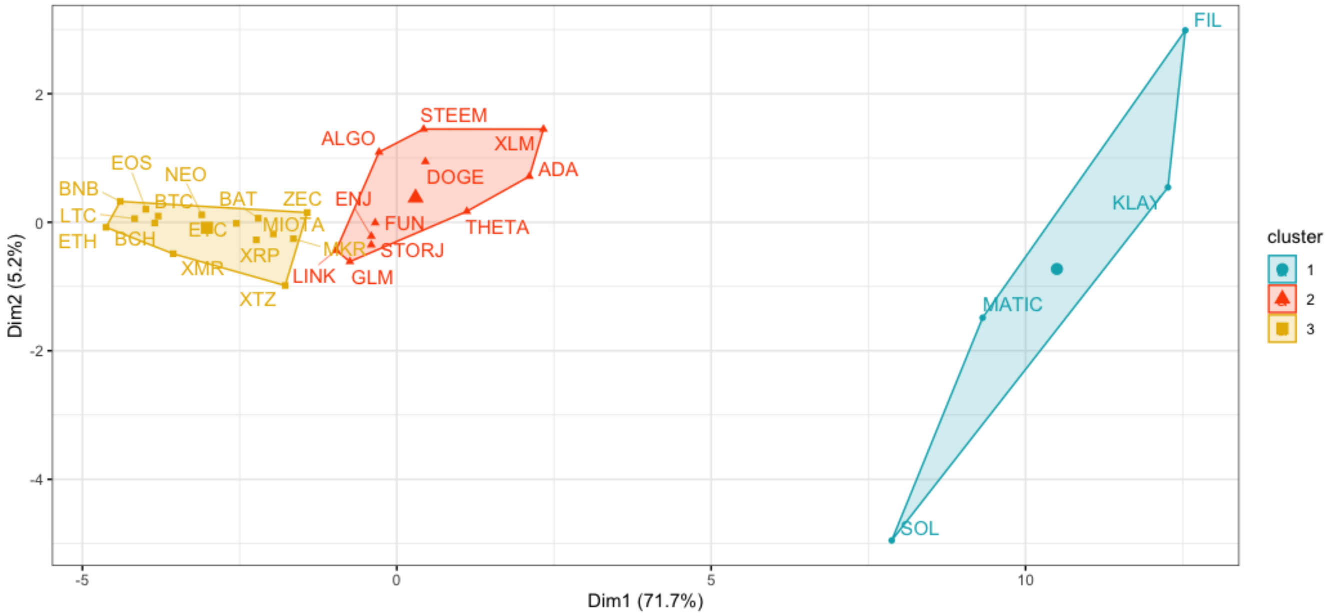

Figure 18. PCA: Elements Association based on K-Means Clusters

So, let's imagine a random market stimulus, we will call it *Z*. This might be news on regulation, a tweet from a crypto character or whatever. If we know that the assets from *Cluster 3* are going to respond to factor *Z* with a drop in value, if we want to build a diversified portfolio, we will avoid investing in too many assets in *Cluster 3*. Too much exposure to a group will suppose a greater fall. However, if we know that *Cluster 1* responds positively to *Z*, then by investing in a balanced way in the different clusters there will be higher probability that the decline in *Cluster 3* is cushioned by *Cluster 1*.

Table 5. Portfolio 3 Constituents by Cluster Grouping

Cluster	Principal Asset	Distance to Cluster's Mean
1	MATIC	1.04
2	DOGE	0.58
3	NEO	0.22

With this last idea in mind, we built our latest portfolio, **Portfolio 3**. This is made up of only one asset from each cluster, choosing the one closest to the center of its cluster. The resulting portfolio is made up of MATIC, DOGE and NEO which together account for only 3% of the market capitalization.

4.4. Portfolio optimization analysis

In the last section of our study, we are going to check how the resulted PCA portfolios behave compared to the fully diversified portfolio. If our PCA calibration period (Train

Period) was from Aug 2019 to Aug 2020, we will now evaluate the portfolios from Aug 2020 to July 2021.

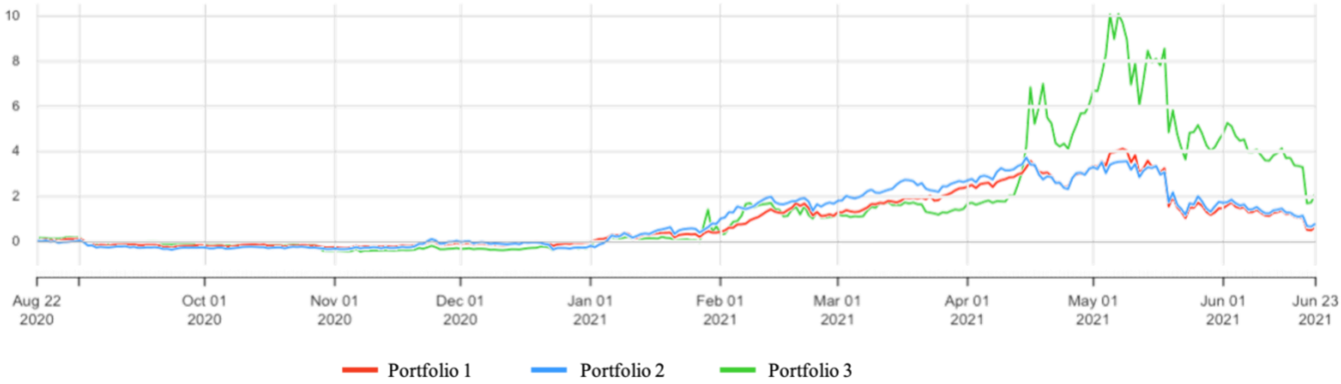

Figure 19. Cumulative Return for Portfolio 1, 2 and 3

Table 6 shows a summary of the results. Our second portfolio, constructed with one asset from each category show a slightly improvement in terms of return and Sharpe Ratio, however nothing remarkable. On the other hand, we see that the portfolio that invests only in one asset per cluster (Portfolio 3) obtains not only the best profitability but also its risk-adjusted profitability improves substantially compared to the rest of the portfolios. Although presenting a higher annualized standard deviation, Portfolio 3 obtains a Sharpe Ratio of 0.95 versus 0.56 on the original portfolio. This validates our initial hypothesis.

Table 6. Portfolio Summary Results *(Aug 20 – July 21)*

	Portfolio 1: All Crypto	**Portfolio 2:** Categories	**Portfolio 3:** Clusters
Number of assets	30	5	3
Assets	All study assets	XRP, MKR, ADA, FUN, ENJ	MATIC, DOGE, NEO
Market share (Market capitalization)	82.2%	5.6%	3.0%
Annualized Return	51.6%	62.5%	156.5%
Worst Drawdown	70.2%	64.7%	75.6%
Cumulative Return	65.7%	80.3%	213.8%
Annualized Standard Deviation	0.92	1.01	1.64
Sharpe Ratio	**0.56**	**0.62**	**0.95**

Finally, we analyze the correlation of the three portfolios with Bitcoin and the different market indices. The results are shown in *Table 7*. We see how the correlation of the third portfolio with respect to Bitcoin falls from a very strong positive correlation of 0.78 (first portfolio) to 0.46. Which is still strong, but the drop is considerable. Furthermore, this clusters portfolio has the lowest correlation of the three portfolios across all equity indices (World, Asia, Europe, Dow Jones and Nasdaq). And in the case of gold, its correlation remains very low, at 0.06.

Table 7. Portfolio Correlation with Market Indexes *(Aug 20 – July 21)*

	Portfolio 1: All Crypto	**Portfolio 2:** Categories	**Portfolio 3:** Clusters
BTC	0.78	0.66	0.46
World	0.28	0.26	0.12
Asia	0.20	0.18	0.07
Europe	0.17	0.14	0.09
Dow Jones	0.21	0.20	0.11
Nasdaq	0.25	0.22	0.09
Gold	0.07	0.04	0.06

5. Conclusion

When an investor seeks to build a diversified equity portfolio, they try to divide their portfolio between different types of stocks; geography, industry, size, value and growth, etc. However, when we analyze the cryptocurrency market it seems that, despite having different types of currencies, they all move in unison. This can make it very difficult to build a diversified portfolio, since in response to bad news the portfolio will move *en bloc*.

This study aimed to update the existing information regarding crypto market correlations, especially after the last rally and fall of 2021. In addition, through a PCA analysis we have represented the cryptoassets in a single graph according to their market behavior. This allow us to check whether it is possible to identify assets with different behaviors, or if on the contrary, there is overlap between classes and all move in unison.

5.1. Correlation analysis

When analyzing the correlations of our group of cryptocurrencies (which together account for 82% of the market capitalization) we have been able to verify that the average correlation is a moderate positive correlation of 0.45. By focusing our analysis during bull and bear markets, we observed that the correlation increases considerably during bear periods, especially in the current bear period that started in April 2021, where the average correlation between assets has been of 0.73. These conclusions coincide with the recent study by Lahajnar and Rožanec (2020)[10], who explored the degree of correlation of the ten cryptocurrencies with the highest capitalization and already observed this correlation increase during bear periods. Therefore, a high correlation during the bear market makes very difficult building a diversified portfolio of crypto assets with an effective hedge to Bitcoin.

One of the key findings of our correlation study comes from the hand of Ethereum, which compared to other studies such as that of Kumar and Ajaz (2019) [9] that positioned Bitcoin as the market price setter, our study shows Ethereum as the asset with the highest correlations with the rest of the market, changing the perspective we had on Bitcoin. This fact was already pointed out by Binance in its latest market correlation study (2020b). Binance also highlighted the increase in correlation during bear markets. Despite this discovery, it is worth noting that when carrying out a 50-day rolling correlation analysis of BTC and ETH we have verified that the correlation of these two assets, which now accumulates 70% of the market capitalization, has been 0.82 since mid 2018, that is a very strong correlation.

A correlation study can lead to wrong conclusions if we analyze it in a very long and static period of time, as results will only reflect the average correlation during the full period ignoring big swings. And yet, by doing an analysis of the correlations in shorter periods and seeing how it evolves over time, as we did in the rolling correlation study, we can check whether the correlations vary a lot over time and identify periods in which they are likely to change.

When analyzing the relationship of Bitcoin with the main equity markets, we have confirmed that the correlation between these assets is not only very volatile, but it also goes from positive to negative indistinctly. This, by definition, makes it impossible to use Bitcoin as a hedge against capital market. And it emphasizes the great instability and volatility of this young market.

In relation to previous studies, we do not agree with those that affirm that there is a positive correlation of Bitcoin with traditional stock market indices, such as the study by Wang, Chen, & Zhao, (2020)[17]. And we reaffirm the findings of Thampanya, Nasir and Huynh (2020)[15] who concluded that neither cryptocurrencies, nor gold, serve as a possible hedging instrument for the stock market, since their correlations with these markets were positive in most cases.

An interesting result observed from our study is that the only times that Bitcoin has shown a negative correlation with any equity index coincides with times when gold also showed a negative correlation. This fact could reveal the use of Bitcoin as a store of value against traditional markets crises.

5.2. PCA

When conducting our PCA, as our variables are assets' return correlation, the study allow us to identify assets that have similar behaviors in terms of correlation with the rest of cryptoassets, seeing if there is a common pattern by category or other circumstance. In this way, we managed to identify different groups of assets with the same correlation patterns, and could help us building a better diversified portfolio, since we can avoid having too many assets from a single group and thus too much exposure to one type of correlation.

While in traditional markets, assets behave differently in response to the same stimuli. There are stocks that fall due to a rise in the price of wheat, others that are greatly affected by negative employment expectations, and so on. All this translates into assets that have disparate correlations between them, and that are usually grouped by sectors, industries, sizes or countries. The crypto market, theoretically, consists of different types of assets, from distinct sectors and sizes, however what we have been able to verify is that they seem to respond *en masse* to market stimuli.

The influence graph resulting from the PCA showed variables that were grouped very close to the *Dim1* axis, evidencing the great redundancy of many of these variables. Which in terms of our study supports the idea of a great correlation between groups. Unsurprisingly, when representing assets in terms of Dim 1 and Dim 2, the vast majority ended up close to the *Dim1* axis in a big single group. This again indicates that the vast majority of assets will follow a unique path in response to market changes.

As we could see in *Figure 17*, when we group each asset according to its natural category (Currency, Financial, Infrastructure, Media and Services), far from appearing in differentiated groups, the assets presented a great overlap. This means that market movements affect assets of different categories in the same way, and therefore diversifying based on categories should not result in good performance.

Finally, we have grouped our elements using k-mean clustering technique, which ignores categorical classification and groups the assets only based on *Dim1* and *Dim2* factors. The assets with the most similar values will be grouped together under three different clusters. These groups contain the assets that show the most similar market behavior. The first cluster contained only four assets which accounted for less than 2% of market share. All four correspond to recently created infrastructure projects. The second and third group were closer together. The second group included 11 assets and accounted for 7.3% of the market. The final group was the biggest, the one that contained BTC and ETH, and accounted for 63% of the market. This again, reinforced our hypothesis of a greatly correlated market.

The resulted assets' clustering in these two large groups, without a common factor of category or internal design, confirms our hypothesis that currently the crypto market is moved by speculative sentiments and not by fundamentals behind each individual project.

5.3. Portfolio optimization

In traditional markets, when building a diversified portfolio, it is very common to use indexed ETFs. However, in the crypto market, these investible indexes do not exist, so diversifying involves buying many assets, which implies an additional operational risk compared to traditional markets.

With this idea in mind, and after analyzing the results of our PCA, we decided to build two portfolios that we compared with a base portfolio. The base portfolio (Portfolio 1) consisted of a balanced investment in the 31 selected study assets. Portfolio 2 consisted of one asset from each category, choosing the ones that were closest to the groups center in the PCA analysis. And finally, Portfolio 3 was formed with only one asset from each of the clusters from the K-means analysis.

Results not only demonstrated that Portfolio 3, composed of assets with a theoretically different market behavior, obtained better results in terms of returns and risk-adjusted returns. But also, its correlation with Bitcoin and the rest of the equity markets decreased considerably, making Portfolio 3 a better hedging tool.

5.4. Closing and potential study improvements

Finally, it should be noted that the results of this analysis want to highlight the great correlation that exists today between most crypto assets. Where different categories and sizes move under the same patterns. This makes an investment in multiple similar assets very fruitful during bull markets, but disastrous during bear markets. This report wanted to present the PCA analysis as a potential tool to identify cryptos with different behaviors, which allows to build portfolios that do not move *en bloc*.

Future work requires acquiring more data and make it automatically pulled from the original source. That would require an investment in a paid subscription to *Coinmarketcap*. From a technical point of view, the portfolio optimization tool could be upgraded incorporating a second analysis on historical Sharpe ratios, so that the selected assets for Portfolio 3 were the result of a least square optimization of PCA clusters and best Sharpe Ratios. This may result in a diversified portfolio (with less exposure to Bitcoin) and with better risk-return metrics.

Disclaimer

Appendix

A.1. Code repository

All the used code can be accessed through the GitHub repository:

https://github.com/cherrypeaks/Crypto_Correlation

An interactive tool has also been created to test the different tools used for the development of this study. This can be accessed at:

https://cherrited.shinyapps.io/Capstone_MariaGuinda/

A.2. Crypto market data

As of 24th June 2021, the market information for the currencies with the highest capitalization as reported by *coinmarketcap.com* was:

#	Name	Ticker	Included	Excluded	Price *(USD)*	Market Cap. *(M UDS)*	Market Share *(%)*	Cumulative MS *(%)*
1	Bitcoin	BTC	YES		33,907	635,487	46.6%	46.6%
2	Ethereum	ETH	YES		1,961	228,314	16.8%	63.4%
3	Tether	USDT	NO	Stable Coin	1.00	62,651	4.6%	68.0%
4	Binance Coin	BNB	YES		303.88	46,625	3.4%	71.4%
5	Cardano	ADA	YES		1.35	43,087	3.2%	74.6%
6	Dogecoin	DOGE	YES		0.24	31,007	2.3%	76.8%
7	XRP	XRP	YES		0.65	29,988	2.2%	79.1%
8	USD Coin	USDC	NO	Stable Coin	1.00	25,463	1.9%	80.9%
9	Polkadot	DOT	YES		16.09	15,354	1.1%	82.0%
10	Uniswap	UNI	NO	New Coin (<1Y)	17.41	10,013	0.7%	82.8%
11	Binance USD	BUSD	NO	Stable Coin	1.00	9,579	0.7%	83.5%
12	Bitcoin Cash	BCH	YES		486.31	9,129	0.7%	84.2%
13	Litecoin	LTC	YES		133.00	8,878	0.7%	84.8%
14	Solana	SOL	YES		30.18	8,229	0.6%	85.4%
15	Chainlink	LINK	YES		18.81	8,136	0.6%	86.0%
16	Polygon	MATIC	YES		1.19	7,498	0.6%	86.6%
17	THETA	THETA	YES		6.92	6,924	0.5%	87.1%
18	Wrapped Bitcoin	WBTC	NO	Stable Coin	33,972	6,490	0.5%	87.5%
19	Stellar	XLM	YES		0.26	6,124	0.4%	88.0%
20	Dai	DAI	NO	Stable Coin	1.00	5,123	0.4%	88.4%
21	VeChain	VET	NO	Small Price	0.08	5,062	0.4%	88.7%
22	Ethereum Classic	ETC	YES		41.01	4,770	0.4%	89.1%
23	Internet Computer	ICP	NO	New Coin (<1Y)	34.81	4,695	0.3%	89.4%
24	TRON	TRX	NO	Small Price	0.07	4,683	0.3%	89.8%
25	Filecoin	FIL	YES		56.87	4,635	0.3%	90.1%
26	Monero	XMR	YES		221.22	3,968	0.3%	90.4%
27	EOS	EOS	YES		3.82	3,646	0.3%	90.7%
28	Klaytn	KLAY	YES		1.10	2,717	0.2%	90.9%

29	SHIBA INU	SHIB	*NO*	New Coin (<1Y)	0.0000068	2,697	0.2%	91.1%
30	Algorand	ALGO	**YES**		0.86	2,664	0.2%	91.3%
31	Aave	AAVE	*NO*	Outside top 20	204.52	2,618.10	0.2%	91.5%
32	Amp	AMP	*NO*	Outside top 20	0.06	2,602.51	0.2%	91.7%
33	PancakeSwap	CAKE	*NO*	Outside top 20	13.51	2,510.55	0.2%	91.8%
34	Bitcoin BEP2	PTCB	*NO*	Outside top 20	34,022	2,487.01	0.2%	92.0%
35	FTX Token	FTT	*NO*	Outside top 20	26.12	2,463.89	0.2%	92.2%
36	Crypto.com Coin	CRO	*NO*	Outside top 20	0.10	2,460.60	0.2%	92.4%
38	Bitcoin SV	BSV	*NO*	Outside top 20	129.79	2,436.10	0.2%	92.6%
37	Theta Fuel	TFUEL	*NO*	Outside top 20	0.46	2,435.69	0.2%	92.7%
39	Neo	NEO	**YES**		34.09	2,405	0.2%	92.9%
40	UNUS SED LEO	LEO	*NO*	Outside top 20	2.48	2,368.84	0.2%	93.1%
41	Tezos	XTZ	**YES**		2.76	2,348	0.2%	93.3%
42	IOTA	MIOTA	**YES**		0.83	2,306	0.2%	93.4%
43	Terra	LUNA	*NO*	Outside top 20	5.28	2,203.07	0.2%	93.6%
44	Maker	MKR	**YES**		2,203	2,184	0.2%	93.8%
45	Cosmos	ATOM	*NO*	Outside top 20	10.02	2,111.87	0.2%	93.9%
46	Avalanche	AVAX	*NO*	Outside top 20	11.39	1,964.68	0.1%	94.0%
47	TerraUSD	UST	*NO*	Outside top 20	1.00	1,908.74	0.1%	94.2%
48	Huobi Token	HT	*NO*	Outside top 20	10.74	1,880.61	0.1%	94.3%
49	Kusama	KSM	*NO*	Outside top 20	208.44	1,765.54	0.1%	94.5%
50	The Graph	GRT	*NO*	Outside top 20	0.61	1,754.13	0.1%	94.6%
51	Hedera Hashgraph	HBAR	*NO*	Outside top 20	0.19	1,670.63	0.1%	94.7%
52	BitTorrent	BTT	*NO*	Outside top 20	0.00	1,647.01	0.1%	94.8%
54	Chiliz	CHZ	*NO*	Outside top 20	0.24	1,421.38	0.1%	94.9%
53	THORChain	RUNE	*NO*	Outside top 20	6.06	1,416.76	0.1%	95.0%
55	TrueUSD	TUSD	*NO*	Outside top 20	1.00	1,405.70	0.1%	95.1%
56	Decred	DCR	*NO*	Outside top 20	105.48	1,376.83	0.1%	95.2%
57	Zcash	ZEC	**YES**		113.25	1,366	0.1%	95.3%

References

11.1. Journal article

[1] Burggraf, T., Huynh, T. L. D., Rudolf, M., & Wang, M. **(2020)**. "Do FEARS drive Bitcoin?" *Review of Behavioral Finance.*

[2] Ciaian, P., Rajcaniova, M. and Kancs, D.A. **(2014)**. "The economics of bitcoin price formation." *Applied Economics*, Vol. 48 No. 19, pp. 1799-1815.

[3] Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. **(2019)**. "Cryptocurrencies as a financial asset: a systematic analysis." *International Review of Financial Analysis*, 62, 182-199.

[4] Dwyer, G. P. **(2015)**. "The economics of Bitcoin and similar private digital currencies." *Journal of Financial Stability*, 17(apr), 81–91.

[5] Dyhrberg, A.H. **(2015)**. "Bitcoin, gold and the dollar – a GARCH volatility analysis." *Finance Research Letters*, Vol. 16, pp. 85-92.

[6] Giudici, P., & Polinesi, G. **(2019)**. "Crypto price discovery through correlation networks." *Annals of Operations Research.*

[7] Hayes, A.S. **(2016)**. "Cryptocurrency value formation: an empirical study leading to a cost of production model for valuing bitcoin." *Telematics and Informatics*, Vol. 34 No. 7, pp. 1308-1321.

[8] Kristoufek, L. **(2013)**. "Bitcoin meets google trends and Wikipedia: quantifying the relationship between phenomena of the internet era." *Scientific Reports*, Vol. 3 No. 1, p. 3415.

[9] Kumar, A. & Ajaz, T. **(2019)**. "Co-movement in crypto-currency markets: evidences from wavelet analysis." *Financial Innovation*; Heidelberg Vol. 5, Iss. 1, (Jul 2019): 1-17.

[10] Lahajnar, S. & Rožanec, A. **(2020)**. "The correlation strength of the most important cryptocurrencies in the bull and bear market." *Investment Management & Financial Innovations*; Sumy, 17, N3, 67-81.

[11] Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, H. L. R. **(2018).** "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis." *Journal of Management Information Systems*, 35(1), 19-52.

[12] Polasik, M., Piotrowska, A.I., Wisniewski, T.P., Kotkowski, R. and Lightfoot, G. **(2015)**. "Price fluctuations and the use of bitcoin: an empirical inquiry." *International Journal of Electronic Commerce*, Vol. 20 No. 1, pp. 9-49.

[13] Qi, T., Wang, T., Zhu, J., & Bai, R. **(2020)**. "The correlation and volatility between bitcoin and the blockchain index." *International Journal of Crowd Science*, Bingley, 4, N.º 2.

[14] Rosenthal, J. A. **(1996)**. "Qualitative descriptors of strength of association and effect size." *Journal of Social Service Research*, 21(4), 37-59.

[15] Thampanya, N., Nasir, M.A., & Huynh, T.L.D. **(2020)**. "Asymmetric correlation and hedging effectiveness of gold & cryptocurrencies: From pre-industrial to the 4th industrial revolution." *Technol Forecast Soc Change*, 159.

[16] van Wijk, D. **(2013)**. "What Can Be Expected fromthe Bitcoin." *Erasmus Universiteit*, Rotterdam.

[17] Wang, X., Chen, X., & Zhao, P. **(2020)**. "The Relationship between Bitcoin and Stock Market." *International Journal of Operations Research and Information Systems*, 11(2), 22-35.

11.3. Books

[18] Cohen, J. **(1988)**. "Statistical Power Analysis for the Behavioral Sciences" (2nd ed.). *Hillsdale: Lawrence Erlbaum.*

11.4. Online sources

[19] Binance Research **(2020a)**. "How Has the Recent Market Turmoil Impacted Cryptocurrencies?" Retrieved from https://research.binance.com/en/analysis/correlations-q1-2020 05-06-2021

[20] Binance Research **(2020b)**. "2019 - Annual Crypto-Correlations Review." Retrieved from https://research.binance.com/en/analysis/annual-crypto-correlations-2019 05-06-2021

[21] Coinmetrics. **(2021)**. "Correlations." Retrieved from https://charts.coinmetrics.io/correlations/ 05-06-2021

[22] Coinmarketcap.com **(2021)**. Market information. 24-06-2021

[23] Cryptodatadownload. **(2021)**. “Correlations.” Retrieved from https://www.cryptodatadownload.com/analytics/correlation-heatmap/ 05-06-2021

[24] Cryptowat. **(2021)**. “Correlations.” Retrieved from https://cryptowat.ch/es-es/correlations

[25] Nakamoto, S. **(2008)**. “Bitcoin: A peer-to-peer electronic cash system.” Retrieved from https://bitcoin.org/bitcoin.pdf 05-06-2021

[26] Messari.io. **(2021).** Crypto descriptive information.