March 14th, 2022

Analysis of Cryptocurrencies

DATA/BIOSTAT 557

CRYPTOCURRENCIES GROUP

Mathew Chau
Nayantara Mohan
Rohit Lokwani
Sravan Hande
Tharun Kumar Reddy Karasani

1. Abstract

Cryptocurrencies are playing a pivotal role in digitalizing money. With the exponential rise in the previous decade, the cryptocurrency market is a three trillion-dollar market attracting many new investors every day [1]. In our work, we aim to help new cryptocurrency investors devise an investment strategy through ex-post-facto research. We use 2-sample one-tailed T-tests to compare group means and aim to test the statistical difference. We test if tokens give better returns than coins and if minable coins give better returns than non-mineable ones. We also test the market sentiment if Mondays give better results than investments on other weekdays. The final method aims to detect a time-lagged cross-correlation between traded volume and daily average returns. At a significance level of 0.05, we get statistically significant results with a p-value of 4.385e-06($p<0.0001$) and 0.03825 for the first two tests respectively. For the third test, we do not have enough evidence to reject the market sentiment. For the correlation test, we found a -0.15 correlation for log return and traded volume with a one-day lag. From the investor's point of view, our work suggests that investing in tokens over coins gives better returns. If the investor still wants to invest in coins, mineable is better than non-mineable. The investor can invest on any of the days. We would recommend an investor trade a particular cryptocurrency if it had a low return the previous day.

2. Introduction

The cryptocurrency market is a three trillion-dollar industry [1] attracting many new investors daily. Over the past decade, the overall market growth has been exponential, and its assets gained wide acceptance as digital currencies. Its decentralized nature offers freedom to the user, keeping their transactions confidential and unforgeable [2]. Cryptocurrencies are also playing a pivotal role in digitalizing money. Like the equities stock market, we see a variety of products and categories in the cryptocurrency market [3]. Figure 2.1 captures a simplistic structure of the types of cryptocurrencies available for public trading.

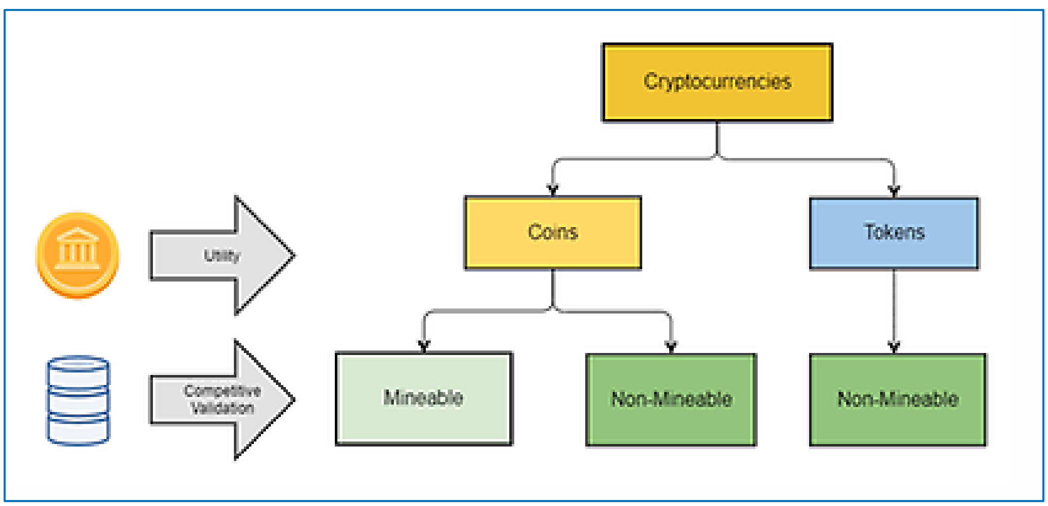

Figure 2.1: Types of Cryptocurrencies available for public trading

Apart from the benefits offered above, the cryptocurrency market has some caveats. Its inherent market traits – high volatility, unregulated structure – are the primary concerns that drive away many traditional and new investors to appreciate its potential and adapt. We followed a structured approach of breaking down the bigger question into multiple smaller hypotheses through our analysis.

2.1 Goal and Motivation

Our analysis aims to help new cryptocurrency investors devise an investment strategy through ex-post-facto research. As it is a growing and volatile market, the motivation for this project is to understand the cryptocurrency market dynamics and how knowing it better can help us formulate and suggest more informed investment decisions to new investors.

Our motivation is curiosity-driven too. We read about many market sentiments [4] that sprout from empirical studies. We were curious to see if these sentiments have any statistical significance. Hence, we performed statistical tests on a couple of market sentiments - returns on Mondays are significantly higher than on other days of the week, the impact of traded volume on price returns.

2.2 Problem Scope

The analysis was performed on the historical data of cryptocurrency returns and volume from November 2019 to November 2020. The trade date range was limited to 1 year in 2020 as it captures the properties of the matured market, which began during the start of the decade. It also contains recent trends of cryptocurrencies like Dogecoin, which gathered public attention over the last few years. The date range also captures a good picture of volatility and market reaction to the COVID pandemic.

After the background research, we finalized the below questions which we analyzed in detail

1. Do coins and tokens have the same average yearly returns?
2. Do mineable and non-mineable coins have the same average yearly returns?
3. Are Mondays more favorable for trading cryptocurrencies?
4. Is there a cross-correlation (or lagged correlation) between traded volume and returns in the cryptocurrencies?

3. Dataset

The dataset used for this analysis is “Historical data on the trading of cryptocurrencies” [8]. The dataset contains daily trading data for 4137 crypto coins and tokens. It has a total of 5 years of data from Dec 2015 to Nov 2020 with 2.4M records. The data is collected from multiple cryptocurrency trading platforms. As stated in the above section, the current analysis is limited to the most recent 1-year period i.e., November 2019 to November 2020. Each record in the dataset has the attributes - Trade Date, Volume, Price USD, Market cap, Capitalization Change 1day, USD Price change 1day, Crypto Name, Crypto Type, Ticker, Max Supply, Site URL, GitHub URL, Mineable, Platform Name, and Industry Name. Along with these, we derive a few aggregate fields for better statistical analysis which are further discussed in each hypothesis test.

Table 3.1 describes the attributes we utilize for the analysis,

Attribute	Description
Trade Date	Date of crypto transactions in the format MM/DD/YYYY
Volume	Volume traded for specific crypto on a specific day
Market cap	The market capitalization of specific crypto on a specific day
USD Price change 1day	Change in price of specific crypto on a specific day
Crypto Name	Name of the cryptocurrency
Crypto Type	Type of the crypto – Coin/Token
Mineable	Boolean column indicating if crypto is minable or not

Table 3.1: Dataset attributes used for analysis

We aim to analyze the dataset under an exploratory study design and discover the underlying patterns. We hypothesize that these patterns can be used to validate the market sentiment and identify key findings that can help an investor form a strategic investment plan.

4. Exploratory Data Analysis

This section includes the key findings from data exploration.

To start with, we wanted to analyze the market share of each of the crypto. The 95% Market capitalization is attributed to only 20 cryptocurrencies. We saw that Bitcoin has the highest market share covering 69%, followed by Ethereum 11%, XRP 4%, and Tether 3%.

Out of 4100 cryptocurrencies, 3000 gave a positive yearly return. All the top 20 cryptos have a positive yearly return. Among the top 20 cryptocurrencies based on market cap, we have 7 Tokens and 13 Coins out of which 8 are Mineable and 12 are non-Mineable. The average yearly return of tokens is greater than that of coins. Mineable cryptocurrencies have greater average yearly returns than non-mineable.

Figure 4.1 compares the average yearly returns for coins vs tokens and mineable vs non-mineable coins.

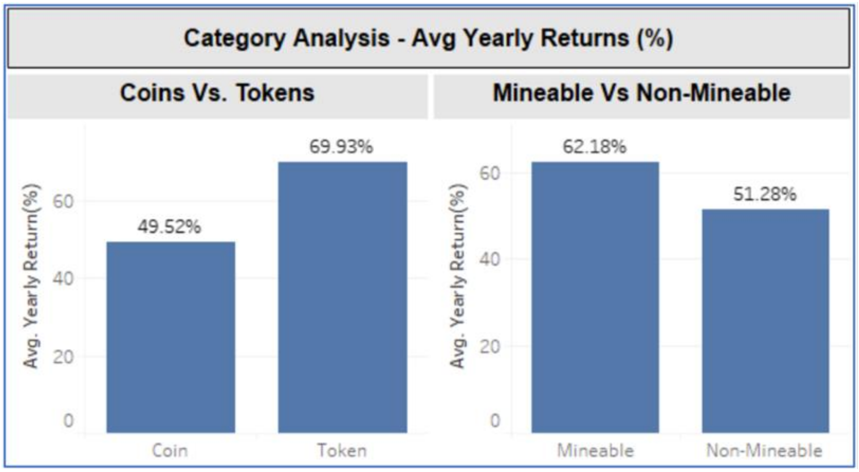

Figure 4.1: Average yearly returns: (Left) Coins vs Tokens, (Right) Mineable vs non-mineable coins

Another aspect of the analysis is to check if there is any specific day of a week resulting in higher profit/loss in comparison with other days. We notice Monday and Wednesday represent a higher average daily return, while Sunday shows the highest dip (Table 4.2). Out of all the currencies, Ripple, stellar has the highest average daily return on Monday.

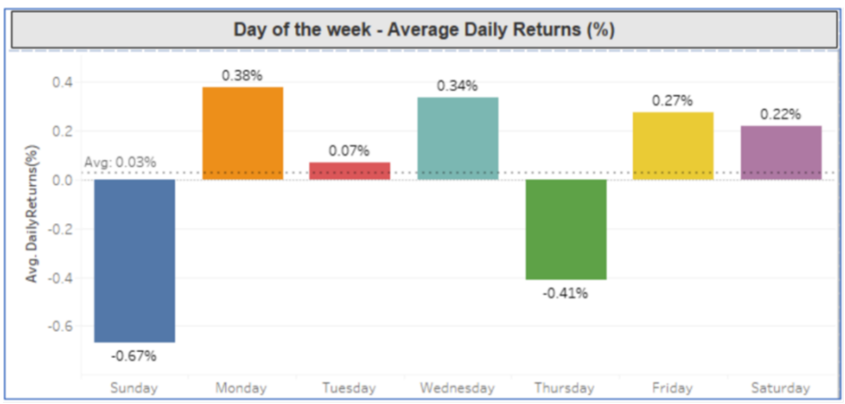

Figure 4.2: Average daily return (Day of the week)

We wanted to further split the data into trades resulting in profit and loss and compare the average gain/loss on each day of the week. The below plot shows the comparison between the number of trades with positive returns and negative returns and the respective returns. Profitable trades with higher daily returns are observed on Monday.

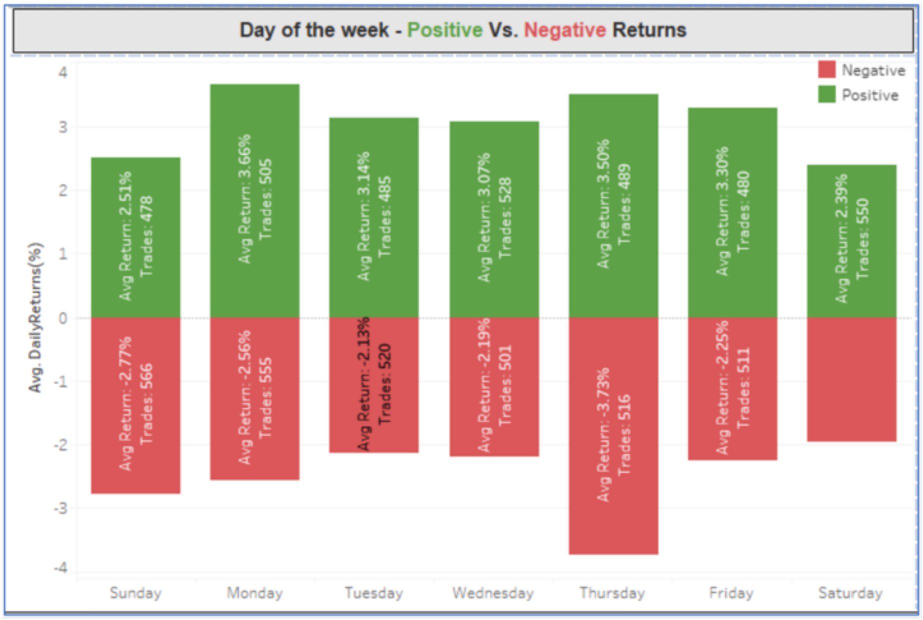

Figure 4.3: Positive vs Negative returns (Day of the week)

We wanted to explore if there is any correlation between the attributes like traded volume, trades, market cap, crypto price, and crypto returns. We see a positive linear relationship between absolute daily return and traded volume (Figure 4.4).

We further check for any correlation within cryptocurrencies to explore the scenarios where the profit of crypto results in the loss of another crypto. We observed a higher correlation between the daily returns of a few groups of cryptocurrencies. Bitcoin, Binance Coin, and XRP daily returns have higher correlations.

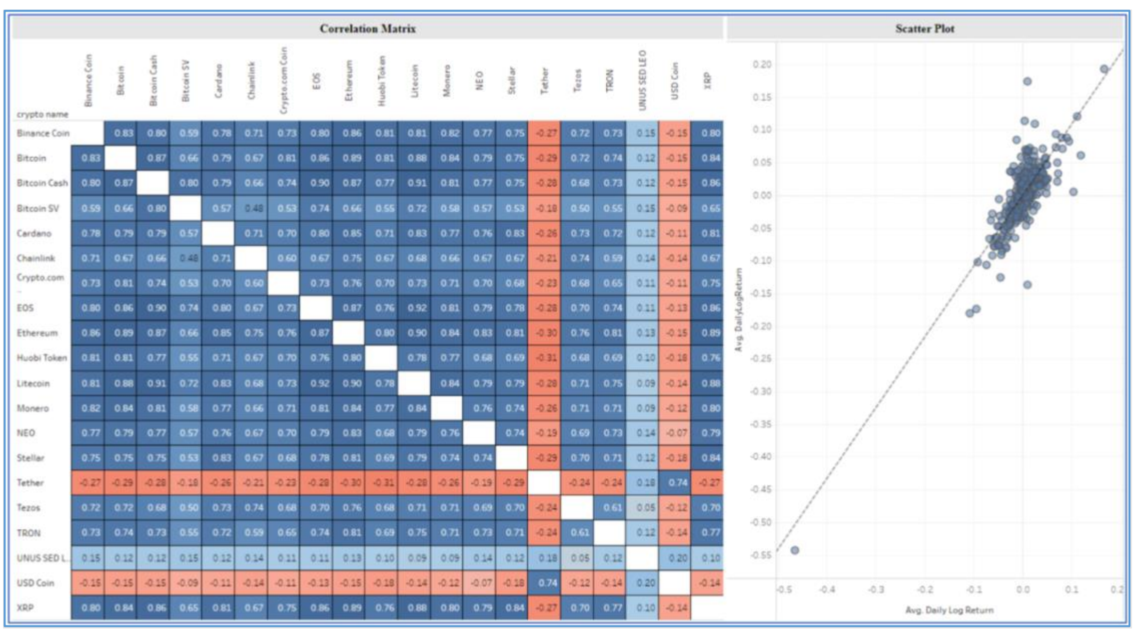

Figure 4.4: (Left) Correlation Matrix showing the relationship between different cryptocurrencies, (Right) Scatterplot of traded volume and average log returns

5. Statistical Methods

This section speaks about the statistical methods that we have used to answer the questions stated in section 2.2. To state them, we use one-tailed equal variance 2-sample T-tests to answer the first three questions and then test the lagged cross-correlation for the last question. The software used for all of these tests was RStudio. The programming languages used are R and Python.

5.1 Question 1

The first question is as follows:

Do coins and tokens have the same average yearly returns?

We address this question by comparing the log yearly returns of the coins and tokens in the dataset. We are comparing the means of two sample groups with equal variances. The observed mean of tokens was higher than coins. Hence, the statistical test used was a left-tailed equal variance 2-sample T-test. We conduct this test for all the cryptocurrencies in the dataset. We also conduct the test for the top 20 cryptocurrencies as they cover 95% of the market capitalization. We do not emphasize the results for the top 20 for two reasons: the small sample size would greatly reduce the power of the test and the normality assumption might not hold.

5.1.1 Data Cleaning/Aggregation

We calculate the log return i.e., the log of the year-end price divided by the start of the year price using the price in USD column (Example: log (Price of Bitcoin on 01-Nov-20/price of bitcoin on 01-Nov-19)). So, the final data to be analyzed is aggregated by the name of the cryptocurrency and the crypto type (coin/token) with log yearly return values.

5.1.2 Dataset details and Descriptive Statistics

Following are the variables in the dataset used for analysis

- **Variables:** Name of the cryptocurrency, Log Yearly Return, Crypto type
- **Derived fields**: Log Yearly Return – log (Year-end price / Year start price)

For analysis 1, table 5.1 shows the size, mean, and variance of both the sample groups.

Group Name	Sample Size	Sample Mean	Sample Variance
Group 1 (Coins)	685	-0.182	1.48
Group 2 (Tokens)	354	0.181	1.68

Table 5.1: Sample size and descriptive statistics of sample groups for Test 1

The violin chart (Figure 5.1) shows the distribution of yearly log return values for both groups. We see that both resemble a normal distribution plot.

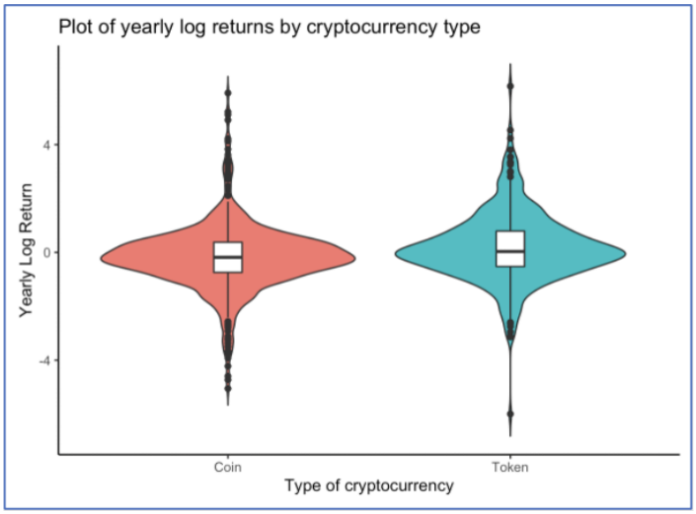

Figure 5.1: Violin plot for depicting the distribution of yearly log returns for coins and tokens

When performing the test for the top 20 cryptocurrencies, the mean of 13 coins and 7 tokens is 0.334 and 0.375 respectively. The variance of these groups is 0.151 and 0.324 respectively.

5.1.3 Statistical method

The following are the details of the hypothesis test we performed:

Test: Left-tailed equal variance 2-sample t-test

Null hypothesis: The mean of yearly average returns is the same for coins and tokens. μ (coins) = μ (tokens)

Alternative hypothesis: The mean of yearly average returns for coins is less than tokens. μ (coins) < μ(tokens)

Metric/Topic	Details
Test distribution	T-distribution with 1037 degrees of freedom
Significance level	0.05
Confidence Interval	95%

Table 5.2: Statistical test details for Test 1

5.1.4 Test Assumptions

- **Independence**: Since all the data points are aggregated annually for each cryptocurrency individually, the independence assumption holds true.

- **Equal variance:** The values of variances come out to be 1.48 and 1.68 which are not very different, hence we assume equal variances.

- **Large sample size or normality**: We have a large sample size here. We also plotted histograms (Figure 5.2) to check if the data is normally distributed. The assumption holds true strongly.

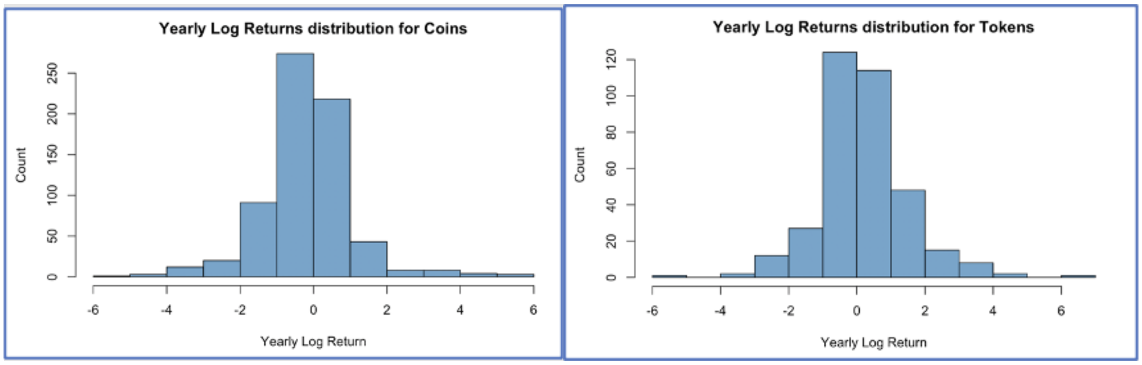

Figure 5.2: Distribution of yearly log returns for coins and tokens in the dataset

5.2 Question 2

Do mineable and non-mineable coins have the same yearly returns?

We address this question by comparing the log yearly returns of the mineable and non-mineable coins in the dataset. We are comparing the means of two samples with equal variances. The observed mean of mineable was higher than non-mineable coins. Hence, the statistical test used was the right-tailed equal variance 2-sample T-test. We conduct this test for all the cryptocurrencies in the dataset. We also conduct the test for the top 20 cryptocurrencies as they cover 95% of the market capitalization. Again, we do not emphasize the results for the top 20 for two reasons: the small sample size would greatly reduce the power of the test and the normality assumption might not hold true.

5.2.1 Data Cleaning and Aggregation

The log yearly return for a cryptocurrency is obtained the same way as mentioned in 5.1.1 (using the log of year-end price/year start price) and is aggregated over cryptocurrency name, mineable/non-mineable with log yearly return values. We analyzed the two groups for comparison: all coins that are mineable and all that are not mineable.

5.2.2 Dataset Details and Descriptive statistics

- **Variables:** Crypto name, Log year return, Mineable/non-mineable
- **Derived fields:** Log Yearly Return – log (Year-end price / Year start price)

For analysis 1, table 5.3 shows the size, mean, and variance of both the sample groups.

Group Name	Sample Size	Sample Mean	Sample Variance
Group 1 (Mineable coins)	434	-0.1191	1.451
Group 2 (Non-mineable coins)	251	-0.2902	1.528

Table 5.3: Sample size and descriptive statistics of sample groups for Test 2

The violin chart (figure 5.3) shows the distribution of yearly log return values for both groups. We see that both resemble a normal distribution plot.

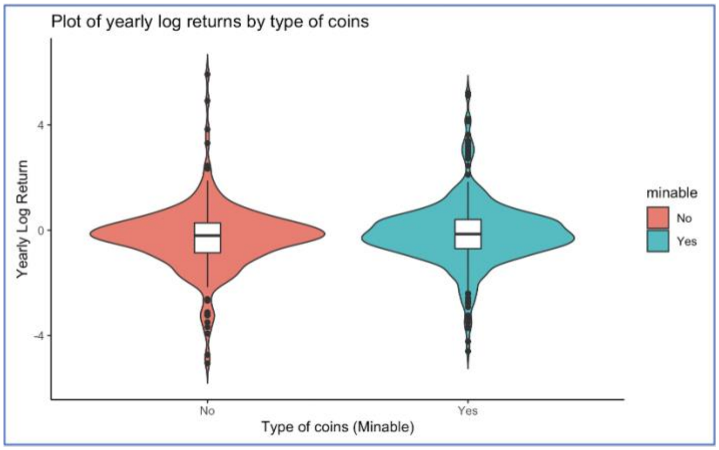

Figure 5.3: Violin plot for depicting the distribution of yearly log returns for mineable and non-mineable coins

When performing the test for the top 20 cryptocurrencies, the mean of 8 mineable coins and 4 non-mineable coins is 0.439 and 0.194 respectively. The variance of these groups is 0.104 and 0.205 respectively.

5.2.3 Statistical method

Test: Right-tailed **e**qual variance 2-sample t-test

Null Hypothesis: the mean for the log return for the mineable and non-mineable coins are equal, μ (mineable) = μ (non-mineable)

Alternative Hypothesis: the mean for the log return for the mineable and non-mineable coins are not equal, μ (mineable) > μ (non-mineable)

Metric/Topic	Details
Test distribution	t-distribution with 683 degrees of freedom
Significance level	0.05
Confidence Interval	95%

Table 5.4: Statistical test details for Test 2

5.2.4 Test Assumptions

- **Independence:** All the data points are aggregated annually for each coin, thus we can assume independence in the sample.

- Equal Variance: The variance of mineable is 1.451 and that of non-mineable is 1.528. They do not differ much, so we assume equal variance.

- Large Sample Size/Normality: We have a large sample size**.** We plotted histograms to check if the data is normally distributed. The assumption holds true strongly

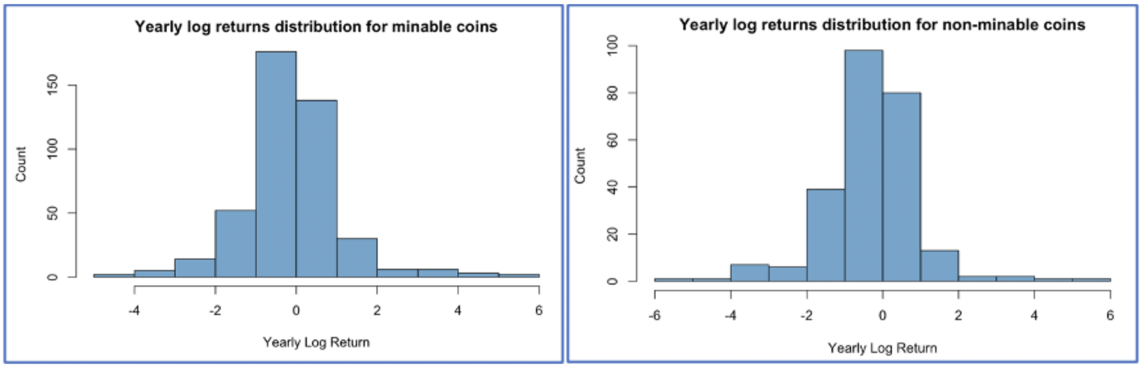

Figure 5.4: Distribution of yearly log returns for mineable and non-mineable coins in the dataset

5.3 Question 3

Are Mondays more favorable for trading cryptocurrencies?

We address this question by comparing the daily log returns for Mondays and other days in the week. We are comparing the means of two samples with equal variances. The observed mean of daily log returns for Mondays was higher than other days. Hence, the statistical test used was the right-tailed equal variance 2-sample T-test. We conduct this test for all the cryptocurrencies in the dataset.

5.3.1 Data Cleaning/Aggregation

The daily returns for every cryptocurrency are given in the dataset. An average of the daily returns for all cryptocurrencies in a particular trade date is averaged and the log is taken to avoid skewness in data. The data set is for 364 days (52 weeks).

5.3.2 Dataset details and Descriptive Statistics

- **Variables:** Name of the cryptocurrency, Trade Date, Volume, Price(USD)
- **Derived fields**: Log Daily Return – log (Today's price / Yesterday's price)
  Day of the week – derived from Trade Date
  Average Daily Log Returns – Average of log daily returns grouped by day of week

Group Name	Sample Size	Sample Mean	Sample Variance
Group 1 (Avg daily log returns for Monday)	52	-0.00017	0.00074

Group 2 (Avg daily log returns for all days but Monday)	312	-0.00020	0.00125

Table 5.5: Sample size and descriptive statistics of sample groups for Test 3

The violin chart (figure 5.5) shows the distribution of yearly log return values for both groups. We see that both resemble a normal distribution plot.

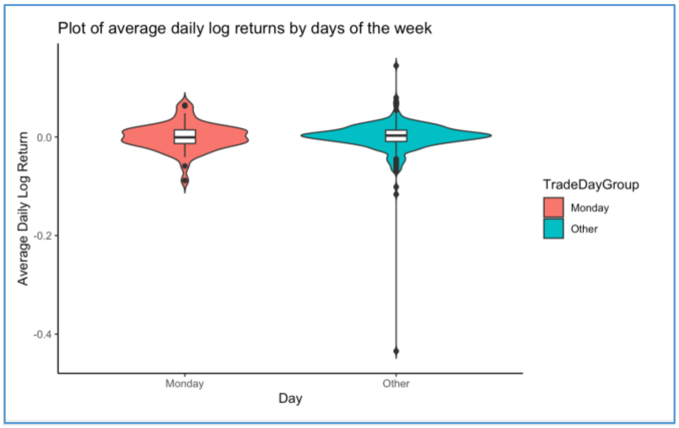

Figure 5.5: Violin plot for depicting the distribution of yearly log returns for Monday and other days

5.3.3 Statistical method

The following are the details of the hypothesis test we performed:

Test: One-tailed equal variance 2-sample t-test

Null hypothesis: The mean of log daily average returns on Monday is the same as for other days in the week. μ (Monday) = μ (any other day)

Alternative hypothesis: The mean of log daily average returns on Monday is greater than other days in the week. μ (Monday) > μ(any other day)

Metric/Topic	Details
Test distribution	t-distribution with 362 degrees of freedom
Significance level	0.05
Confidence Interval	95%

Table 5.6: Statistical test details for Test 3

5.3.4 Test Assumptions

- **Independence**: Since all the data points are collected on different trade dates and individually for each cryptocurrency, the independence assumption does hold true.

- Equal variance: The values of variances come out to be 0.00074 and 0.0012 which are not very different, hence we assume equal variances.

- Large sample size or normality: QQ plots were plotted to check if the data is normally distributed. The assumption holds fairly true.

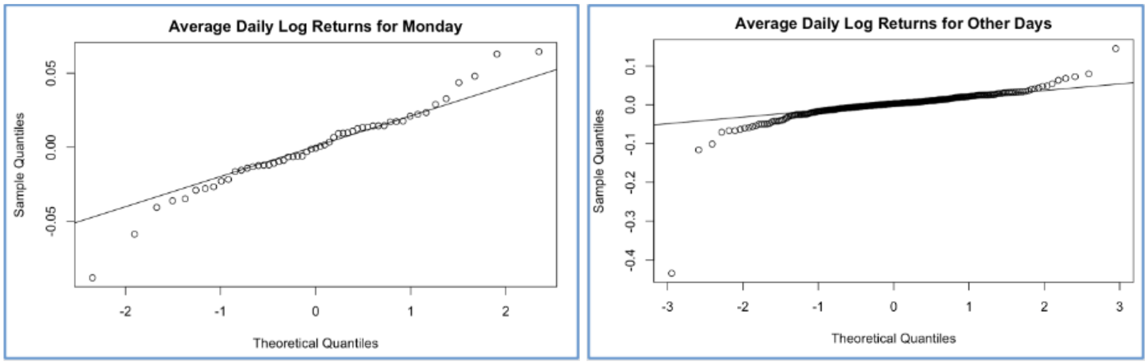

Figure 5.6: Distribution of yearly log returns for Monday and other days using QQ plot

5.4 Question 4

Is there a cross-correlation (time-lagged cross-correlation) between daily traded volume and returns in the cryptocurrencies?

5.4.1 Data Cleaning/Aggregation

The log yearly return for a cryptocurrency is obtained the same way as mentioned in 5.1.1 and 5.1.2 using the log of year-end price/year start price.

5.4.2 Dataset details and Descriptive Statistics

- **Variables:** Trade Date, Traded Volume, Log Daily Return.

- **Derived fields:** Log Daily Return – log (Today's price / Yesterday's price)

- **Sample size:** 20 cryptocurrencies and each coin/token has ~365 data points (Will change based on different lag values as the series hang out)

Variable name	Data size	Sample Mean	Sample Variance
Log daily return	365	7.6e-4	1.5e-3
Traded volume	365	6.1e9	4.5e18

Table 5.7: Sample size and descriptive statistics of sample groups for Test 4

5.4.3 Statistical method

This test seeks for the correlation between the log daily return and the volume of the cryptocurrencies. We calculated the time-lagged cross-correlation with up to 20 days lag – that is, bringing the return values forward and backward in time for up to 20 days and compared the correlation of that with the present traded volume. Positive lag means the number of days the return data has been brought forward in time. For instance, +2 lag indicates we are computing the coefficient with present return and volume two days ago. Negative lag means the other way round.

We have implemented this test for both the average of the top 20 cryptocurrencies as well as each cryptocurrency.

5.4.4 Test Assumptions

- **Level of measurement**: Both traded volume and USD Price change are continuous variables
- **Related pairs**: Traded volume and USD Price change in each record indicate the values for specific crypto on a day
- **Absence of outliers**: Data points that are above 3.29 standard deviations from the mean are filtered out.

6. Results and Interpretation

This section details the results of each test, primarily focusing on their statistical significance, stating the confidence intervals, p-values, power, and their final interpretation. This section also states how these tests could be used to make investment decisions.

6.1 Test 1: Comparing log yearly returns of coins and tokens:

The results for the full data give a p-value of 4.834e-06 ($p<0.0001$) which is highly statistically significant. Further, Cohen's effect size value ($d = .292$) suggested a small to moderate practical significance.

The following table highlights the results from the tests:

Data	95% CI	Test statistic	P-value	Power	Result
Full data	-∞, -0.22827	-4.4465	4.834e-06	0.998	Rejects null hypothesis for equal means
Top 20	-∞, 0.340	-0.1721	0.4326	0.072	Does not reject the null hypothesis for equal means

Table 6.1: Results of analysis for test 1 comparing means of yearly log returns for coins and tokens

At a significance level of 0.05, we reject the null hypothesis of equal means for yearly log return of coins and tokens for all cryptocurrencies. From the investment point of view, for all the cryptocurrencies, we have evidence for tokens having higher yearly log returns as compared to coins

Note: For the top 20 cryptocurrencies, the results contradict but might not be valid, since the normality assumption does not hold true. Even if the test results are considered valid, the power of the test is very low.

6.2 Test 2: Comparing log yearly returns of mineable and non-mineable coins

The results for the full data give a p-value of 0.03825 which is statistically significant at a significance level of 0.05. Further, Cohen's effect size value (d = .1407) suggested a small practical significance.

The following table highlights the results from the tests:

Data	95% CI	Test statistic	P-value	Power	Result
Full data	0.0122, ∞	1.774	0.03825	0.551	Rejects null hypothesis for equal means
Top 20	-1.2286, ∞	-0.1721	0.129	0.265	Does not reject the null hypothesis for equal means

Table 6.2: Results of analysis for test 2 comparing means of yearly log returns for mineable and non-mineable coins

At a significance level of 0.05, we **reject the null hypothesis** of equal means for the yearly log return of all minable and non-minable coins. From the investment point of view, for all the cryptocurrencies, **we have evidence for minable coins have higher yearly log returns as compared to non-minable coins.**

Note: For the top 20 cryptocurrencies, the results contradict but might not be valid, since the normality assumption does not hold true. Also, the power of the test is quite low.

6.3 Test 3: Comparing log average daily returns on Mondays and Other Days

The results for the full data give a p-value > 0.05 which is not statistically significant at a significance level of 0.05. The following table highlights the results from the tests:

Data	95% CI	Test statistic	P-value	Power	Result
Full data	-0.0007, ∞	0.86265	0.1942	0.216	Does **not** reject the null hypothesis for equal means
Top 20	-0.0083, ∞	0.00474	0.4981	0.050	Does **not** reject the null hypothesis for equal means

Table 6.3: Results of analysis for test 3 comparing means of yearly log returns for Mondays and other days of the week

At a significance level of 0.05, we **cannot reject the null hypothesis** of equal means for avg daily log return of all cryptocurrencies on Mondays and Other days.

From an investment point of view, **we can disregard the market sentiment bias towards Monday**. Trading cryptocurrencies have no statistical significance for favoring Mondays.

6.4 Test 4: Checking cross-correlation (time-lagged cross-correlation) between daily traded volume and returns in the cryptocurrencies

For the correlation of overall log daily return and trading volume, we have produced a Correlogram as shown below.

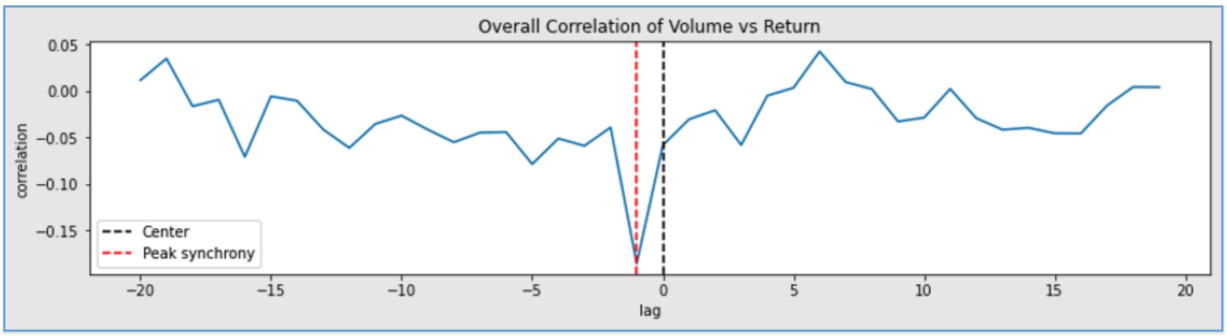

Figure 6.1: Overall Correlation of volume vs return

We have found **that the overall correlation has its peak of -0.15 at –1-day lag, with a p-value of 0.00039 < 0.5, indicating an association between the volume of a specific day and the return of the past day.** That means if the crypto has a high return for yesterday, there tends to be a lower volume for today; if the crypto has a lower return for the previous day, people tend to trade at a higher volume on the next day. This indicates that the **reaction time for the market is around 1 day as of the previous return.**

For the individual coins, we have calculated the correlation of all the top 20 cryptocurrencies and summarized them in the table below. For better visualization, we have presented only –10 to 10 days lag in the following table.

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
Binance Coin	-0.042	-0.078	-0.066	-0.065	-0.02	-0.083	-0.041	-0.028	0.042	-0.154	0.106	-0.004	0.011	-0.026	0.057	0.053	0.044	-0.018	-0.06	-0.022	0.01
Bitcoin	-0.062	-0.067	-0.072	-0.061	-0.059	-0.091	-0.092	-0.054	-0.033	-0.145	0.027	0	0.019	-0.011	0.026	0.031	0.077	0.049	0.005	-0.047	0.014
Bitcoin Cash	-0.035	-0.013	-0.062	-0.107	-0.105	-0.097	-0.041	-0.027	-0.025	-0.076	0.028	-0.031	0.028	0.015	-0.04	-0.116	-0.035	0.001	-0.09	-0.316	-0.11
Bitcoin SV	-0.027	-0.009	-0.046	0.02	0.043	0.003	0.068	0.025	0.062	0.019	0.199	-0.022	0.044	0.007	0.038	0.024	0.063	-0.008	-0.006	-0.034	-0.002
Cardano	0.001	-0.035	-0.052	-0.003	0	-0.016	-0.003	0.072	0.082	0.037	0.062	0	-0.027	-0.042	-0.016	-0.008	-0.001	0.01	-0.026	-0.043	-0.016
Chainlink	-0.017	-0.045	-0.133	-0.031	-0.014	-0.049	-0.045	-0.045	-0.033	-0.044	0.096	0.028	-0.051	-0.093	-0.03	-0.053	-0.001	-0.033	-0.093	-0.084	-0.045
Crypto.com Coin	-0.03	-0.062	-0.067	-0.135	-0.077	-0.046	-0.091	-0.094	-0.106	-0.049	-0.049	-0.052	-0.005	-0.055	-0.034	-0.039	-0.069	-0.097	-0.059	-0.04	-0.125
EOS	-0.059	-0.038	-0.063	-0.03	-0.048	-0.107	-0.081	-0.015	-0.064	-0.26	-0.054	-0.019	0.008	-0.055	-0.016	0.027	0.069	0.002	0.013	-0.048	-0.015
Ethereum	-0.028	-0.037	-0.076	-0.026	-0.017	-0.066	-0.054	-0.039	-0.033	-0.152	0.007	-0.018	-0.021	-0.057	-0.007	-0.001	0.045	0.001	-0.003	-0.034	0.009
Huobi Token	-0.043	-0.022	-0.054	-0.049	-0.007	-0.017	-0.109	-0.018	0.021	-0.076	0.109	0.063	0.077	-0.002	0.059	0.035	0.053	-0.022	-0.05	-0.124	-0.043
Litecoin	-0.02	-0.049	-0.073	-0.03	-0.042	-0.071	-0.044	-0.004	-0.009	-0.143	0.046	-0.002	0.009	-0.042	-0.004	-0.003	0.035	-0.003	-0.017	-0.058	-0.004
Monero	-0.079	-0.114	0.004	-0.155	0.063	-0.019	0.033	0.019	0.05	-0.019	0.058	0.015	0.007	-0.025	0.043	-0.036	0.074	-0.026	-0.022	0.024	-0.034
NEO	-0.012	-0.025	-0.074	-0.053	-0.068	-0.07	-0.027	-0.002	-0.003	-0.042	0.034	-0.028	-0.03	-0.055	-0.078	-0.064	-0.043	-0.062	-0.068	-0.12	-0.099
Stellar	0.031	0.018	0.036	0.047	0.024	0.015	0.02	0.071	0.063	0	0.113	0.062	0.068	0.011	0.044	0.083	0.079	0.054	0.035	0.026	0.021
TRON	-0.051	-0.054	-0.047	-0.048	-0.058	-0.065	-0.017	-0.024	-0.006	-0.126	-0.013	-0.062	-0.048	-0.075	-0.018	-0.038	-0.013	-0.045	-0.053	-0.076	-0.072
Tether	-0.021	0.008	0.03	-0.051	-0.005	0.02	0.002	-0.067	0.006	0.098	-0.05	-0.02	0.005	-0.009	-0.073	-0.012	0.072	-0.015	-0.018	0.011	0.04
Tezos	0.023	0.007	-0.042	-0.003	-0.012	-0.054	-0.059	0.01	-0.029	-0.167	0.047	-0.03	-0.059	-0.15	-0.054	0.003	0.008	-0.035	-0.057	-0.091	-0.067
UNUS SED LEO	-0.029	0.038	0.05	0.022	0.11	0.055	-0.008	-0.013	-0.015	0	0.006	0.026	0.021	0.024	0.041	0.013	0.02	0.039	0.013	-0.025	-0.004
USD Coin	-0.058	-0.007	-0.017	-0.027	-0.038	-0.027	-0.038	-0.01	-0.033	0	0.063	-0.038	-0.019	-0.027	-0.037	-0.061	-0.006	-0.055	-0.022	0.024	-0.004
XRP	0.003	0.005	0.006	-0.014	-0.006	-0.068	-0.037	-0.03	0.02	-0.141	-0.004	-0.006	-0.003	-0.034	-0.007	0.017	0.053	-0.003	0.002	-0.033	0.028

Figure 6.2: Time-lagged correlation of top 20 cryptocurrencies for return vs traded volume

Notice that most correlations are in the range of [-0.1, 0.1], indicating that there is very weak or no correlation between traded volume and log return for most day lags. However, do notice that there are a few values with significant correlations. For Bitcoin Cash with 9 days lag, we have a correlation of –0.316; and for EOS with –1-day lag, we have a correlation of –0.26. The correlograms for those two cryptos are shown below.

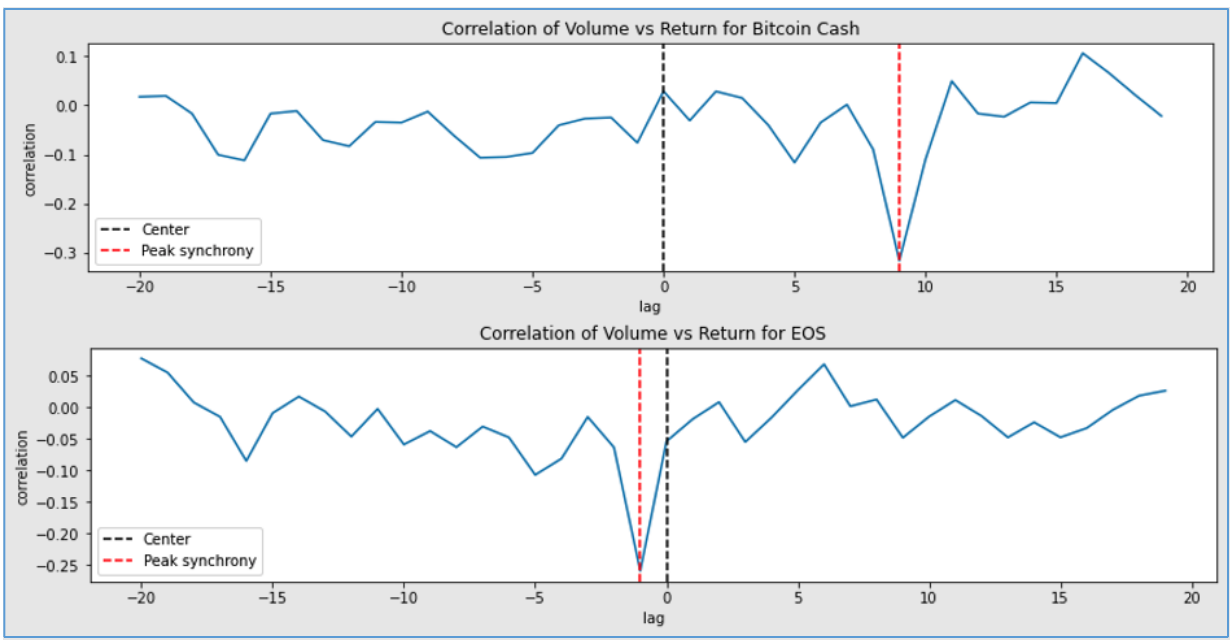

Figure 6.3: Time-lagged correlation of Bitcoin Cash and EOS for return vs traded volume

This indicates that the volume for Bitcoin Cash for today has a negative correlation with the return nine days later. Some simple strategies could be formulated such as "if the trade volume for Bitcoin Cash is low nine days ago, we could buy in more today as we expect a higher return."

7. Discussion

We started our work by understanding cryptocurrencies, analyzing data, and doing market research about the sentiments. Our goal was to help an investor devise an informed strategy using statistical evidence. Our test results in the above section and the conclusions drawn from them are summarized as follows:

• From Test 1, we have evidence for suggesting an investor invest in tokens over coins as they give higher yearly log returns.

• If the investor still wants to invest in coins because of their market coverage or whatever reasons, we have evidence of suggesting the investor invest in minable coins over non-minable coins as they show higher yearly log returns in test 2.

• For the Monday market sentiment, we can conclude that trading cryptocurrencies have no statistical significance for favoring Mondays.

• From the correlation tests, we found that overall next-day volume has a slight negative correlation with the previous-day return. An additional finding includes that it might be a good idea to invest in Bitcoin Cash and have special attention to the volume nine days ago.

A couple of challenges faced throughout the process included dealing with real-world datasets which required transformations and manipulation. Figuring out the right tests and respective strategies to avoid misinterpretation of statistical results was a challenge as well. We also learned the appropriate way of applying statistical tests to this time series data keeping in mind that the assumptions are fulfilled.

The limitations of this analysis include that we have considered the period in which COVID hit around the world due to the available data. This period could be seen as an anomalous period for the cryptocurrency market. Any extrapolation for the upcoming years could be deemed inappropriate. We conducted the test on 1037 cryptocurrencies available in the dataset but the number of cryptocurrencies being traded in the market was 4138 at the same point in time. This could also be a sampling bias and could lead to missing out on the most recent cryptocurrencies.

9. Future scope

We further plan to deepen and extend our statistical analysis as follows:

9.1 Volatility Analysis using ARCH, GARCH, TGARCH

We want to explore another analysis area which is volatility. As we deal with cryptocurrency data, volatility analysis plays a vital role in understanding the assets' stability and risk aspects. ARCH and its extension models like GARCH [5] and TGARCH will help us analyze and forecast the volatility. Through this analysis, an investor will be informed of how new information (news) affects the price of a purchased asset and decide when to exit his position.

9.2 Asset Price Analysis using ARIMA, SARIMA, LSTM

Like any other asset in the stock market, cryptocurrencies also have time series components that explain few data patterns. These patterns can be understood by leveraging autoregression techniques [6] that model the cryptocurrency's price. Using ARIMA, SARIMA, and LSTM models, modeling and forecasting the price helps investors plan effective market entry and exit strategies based on the historical trend and seasonality.

9.3 Portfolio Analysis

Most of the cryptocurrency platforms and brokers facilitate the traders with custom portfolios that track the prices of a few cryptocurrencies. However, deciding which portfolio category is better for investment is hard to make. Hence, we plan to extend the current analysis to a list of portfolios offered across various top cryptocurrency platforms, both in terms of risk and return profiles. The same analysis can be further extended to compare cryptocurrency with S&P500 portfolios.

10. Acknowledgments

We learned a lot about the different statistical techniques and methods to test accurate and relevant hypotheses and interpret results throughout this quarter. We want to acknowledge Prof. Brian Leroux for his teachings, consistent guidance and frequent project sync-up calls throughout the quarter. The constructive feedback from our Teaching Assistants Anushna and Marc on assignments has also helped us rethink some of our test strategy decisions that further guided our iterations for this project.

11. References

[1] *The World's Cryptocurrency Is Now Worth More Than $3 Trillion* (2021, November 8). Time https://time.com/6115300/cryptocurrency-value-3-trillion/

[2] *Key Characteristics of Cryptocurrency and Why Do They Matter to You (2020, August 24). Medium* https://medium.com/the-capital/key-characteristics-of-cryptocurrency-and-why-do-they-matter-to-you-5f33e483a40f

[3] *Understanding the different types of Cryptocurrency* (2021, September 17). SoFi Learn https://www.sofi.com/learn/content/understanding-the-different-types-of-cryptocurrency/

[4] *When is the best time to buy Bitcoin*? Unbanked https://unbanked.com/best-time-buy-bitcoin/

[5] *A systematic review of forecasting the cryptocurrency volatility using GARCH model (2018, April 26) Medium* https://medium.com/coinmonks/a-systematic-review-of-forecasting-the-cryptocurrency-volatility-using-garch-model-d1855c9779c

[6] *Price Analysis and Forecasting for Bitcoin Using Auto-Regressive Integrated Moving Average Model* (2021, December) Research Gate https://www.researchgate.net/publication/356886219_Price_Analysis_and_Forecasting_for_Bitcoin_Using_Auto_Regressive_Integrated_Moving_Average_Model

[7] *Four ways to quantify synchrony between time series data* https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9

[8] *Historical data on trading of cryptocurrencies.* George Zakharov. Kaggle Datasets. https://www.kaggle.com/georgezakharov/historical-data-on-the-trading-of-cryptocurrencies

12. Appendices

This appendix was not used and was intentionally left blank.