Analysi$ of Cryptocurrencie$

DATA/BIOST 557 – Applied Statistics and Experimental Design

Team:
Matthew Chau
Nayantara Mohan
Rohit Lokwani
Sravan Hande
Tharun Kumar Reddy Karasani

Introduction

Cryptocurrencies have received massive recognition in the past decade.

Benefits of owning cryptocurrencies

- Easy Transactions
- Short Settlement Times and Low Fees
- Outsized Returns
- Exponential Industry Growth
- Portfolio Diversification
- Inflation Hedge

Apprehensions for investing in cryptocurrencies

- High volatility
- Unregulated structure

Goal

To help new cryptocurrency investors devise an investment strategy

Motivation

The motivation for this project is to understand the cryptocurrency market dynamics and help investors make more informed investment decisions.

We would also like to statistically test the validity of existing **market sentiments** like **returns on Mondays are significantly higher than on other days of the week.**

Dataset

Dataset Name - Coins and Tokens – Historical data of crypto trading.

Source – Kaggle (Link)

Description - Price and volume data of 4100 cryptocurrencies for the period from Jan-2016 to Nov-2020. Additional fields like cryptocurrency type and mineable information are made available.

Total features – 17

Total observations – 2,382,643

Total unique cryptocurrencies - 4,138

trade_date	volume	price_usd	price_btc	market_cap	capitalization_change_1_day	USD_price_change_1_day	BTC_price_change_1_day	crypto_name	crypto_type	ticker	max_supply	site_url	github_url	minable	platform_name	industry_name
1/1/2016	36278900	434.33	1	6529299589	0	0	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)
1/2/2016	30096600	433.44	1	6517390487	-0.001823948	-0.002049133	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)
1/3/2016	39633800	430.01	1	6467429942	-0.007665728	-0.007913437	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)
1/4/2016	38477500	433.09	1	6515713340	0.007465624	0.007162624	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)
1/5/2016	34522600	431.96	1	6500393256	-0.002351252	-0.002609157	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)
1/6/2016	34042500	429.11	1	6458942098	-0.006376715	-0.006597833	0	Bitcoin	0	BTC	21000000	https://bitcoin.org/	https://github.com/bitcoin/	1	XRP	Proof of Work (PoW)

Data snapshot

Forms of Cryptocurrencies

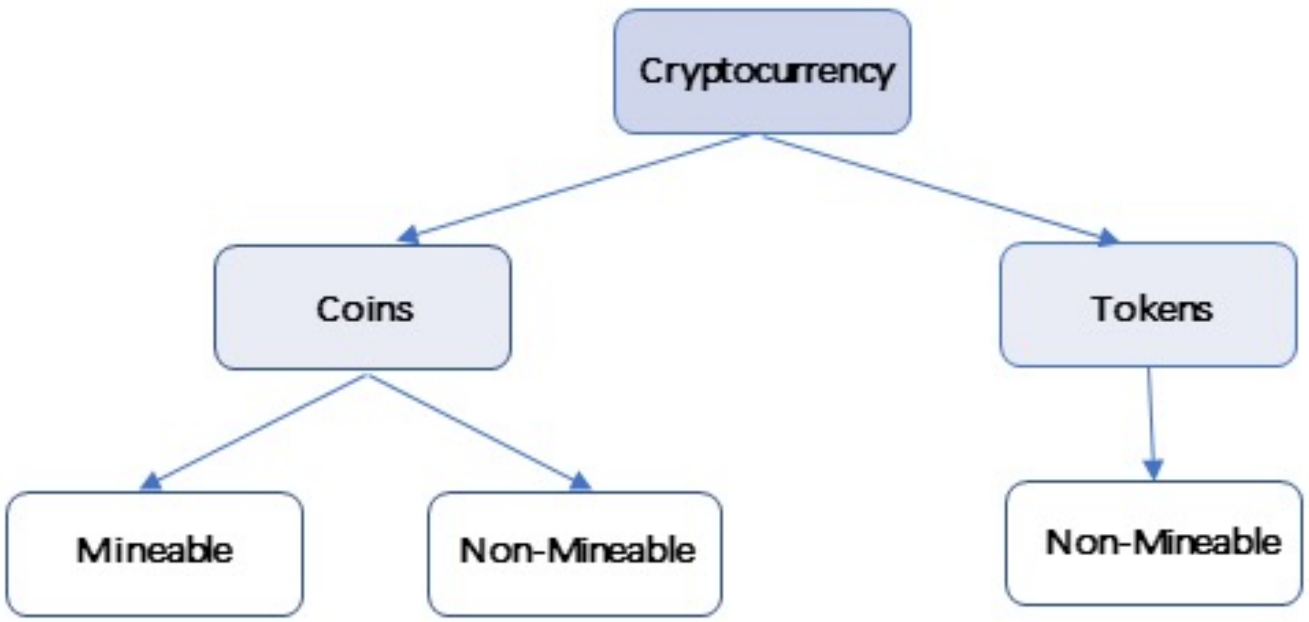
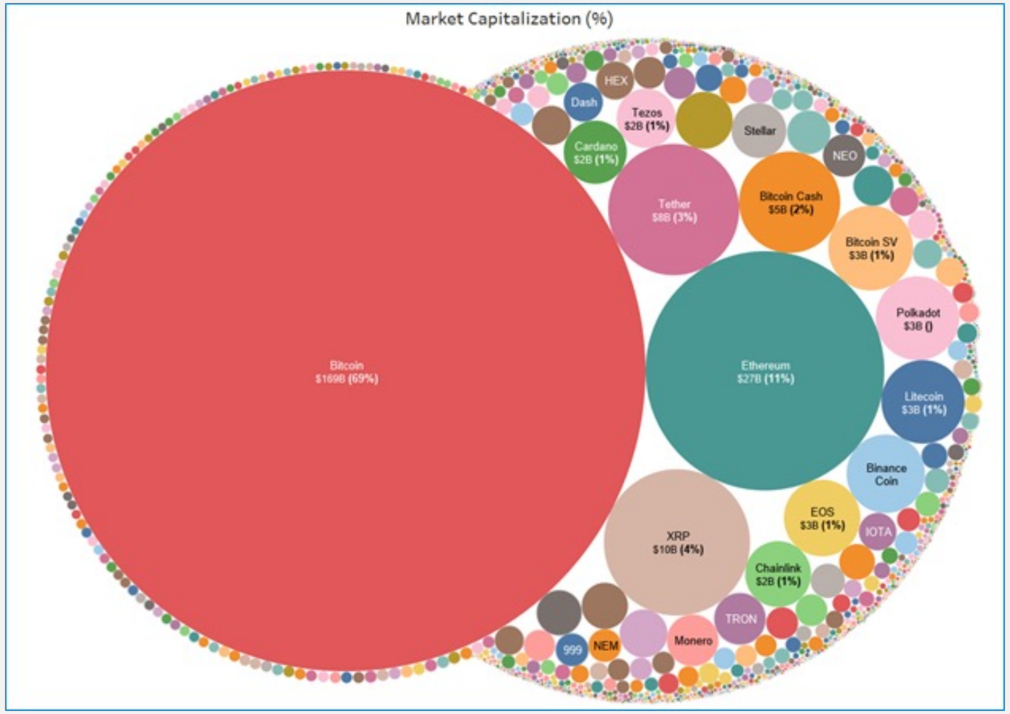
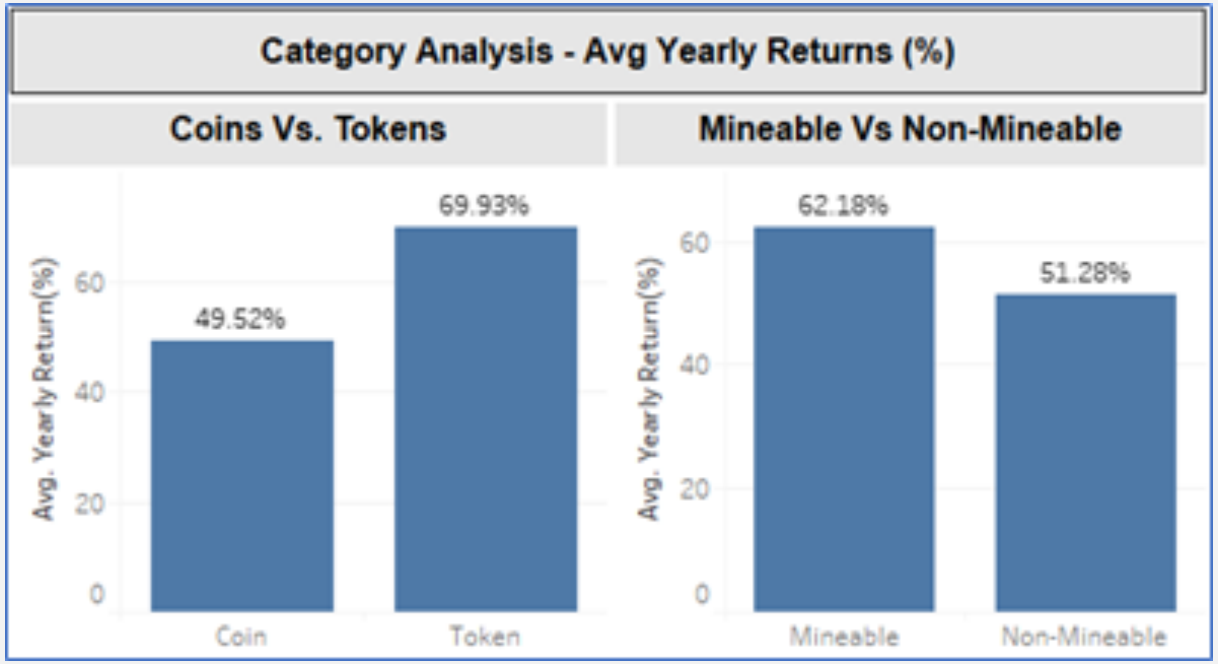


Figure: Forms of Cryptocurrencies available for trading

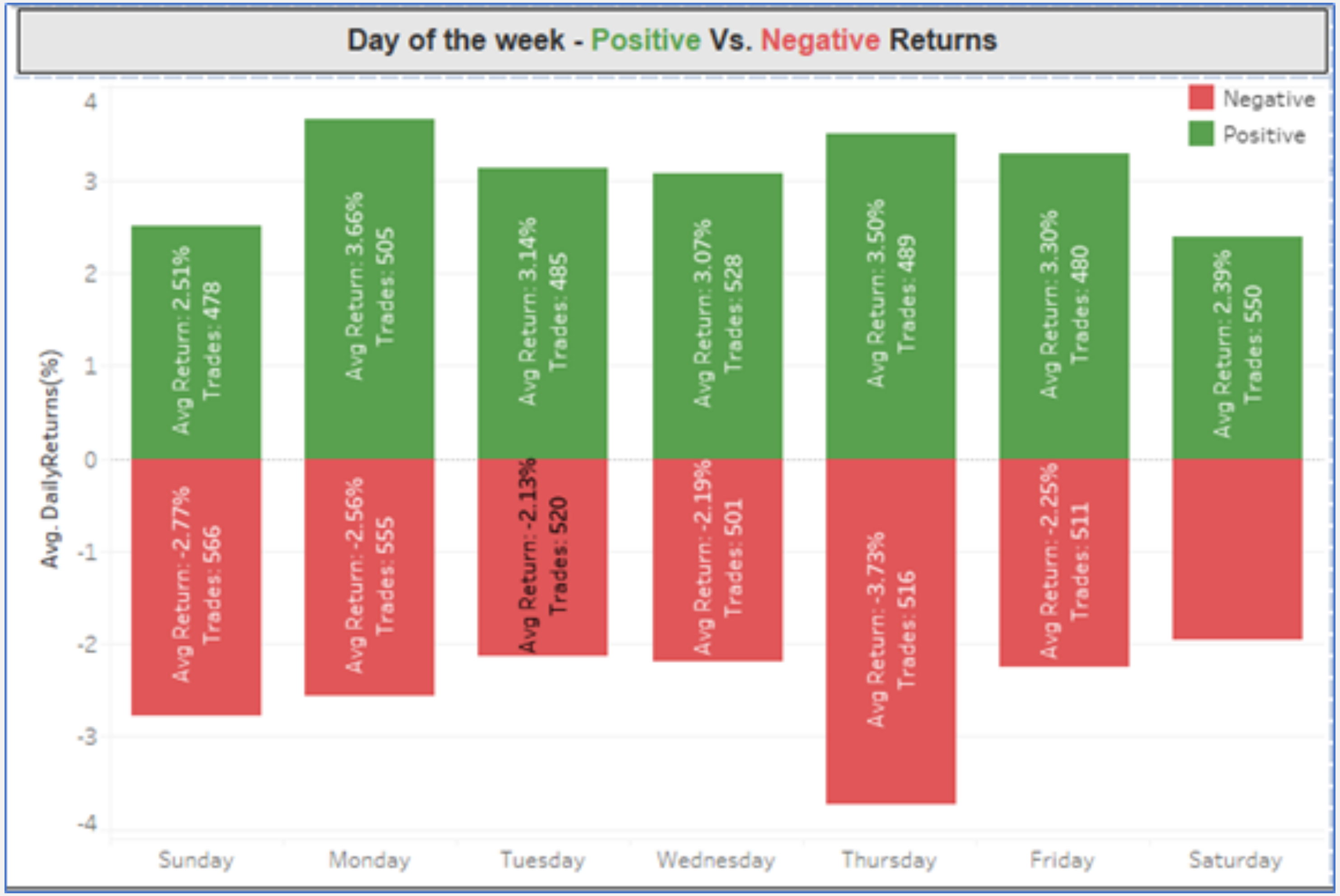
Top Cryptocurrencies

EDA Visualizations I





EDA Visualizations II



The average returns of trades with positive outcome are higher on Monday

Problem Scope

To help new cryptocurrency investors devise an investment strategy based on investment return analysis for the period November 2019-2020

Questions

- Do coins and tokens have the same average yearly returns?

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

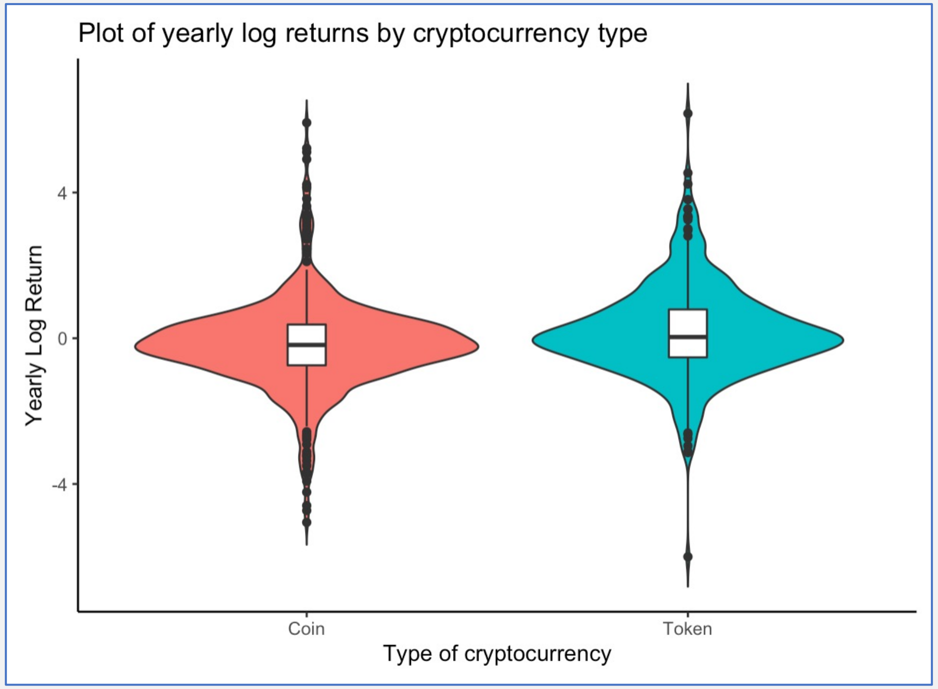
- Are Mondays more favorable for trading cryptocurrencies?

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

Question 1

Do coins and tokens have the same average yearly returns?

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



Dataset

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

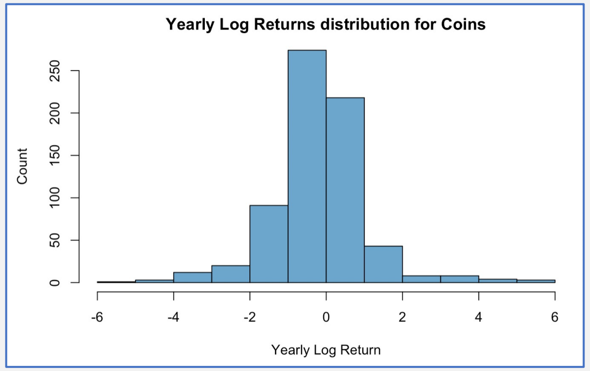
Question 1 – Hypothesis Testing

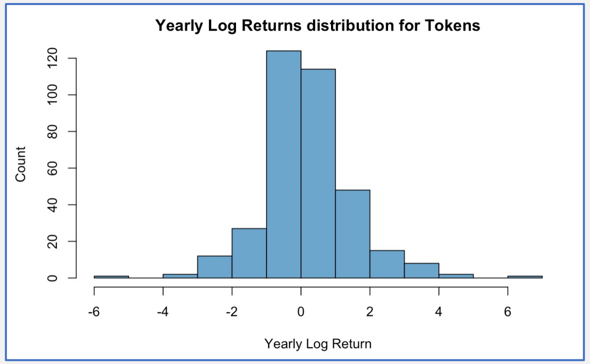
Test: Left-tailed equal variance 2-sample t-test

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

Alternative hypothesis: The mean of yearly average returns for coins is greater 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%





Test Assumptions

- **Independence :** Strongly holds true

- **Large sample size or normality:** Strongly holds true

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

Question 1 – Results & Interpretation

Results

Data	Test statistic	P-value	Result
Full data	-4.4465	4.834e-06	Rejects null hypothesis for equal means
Top 20	-0.1721	0.4326	Does not reject null hypothesis for equal means

Interpretation

At 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 top 20 cryptocurrencies, the results contradict but might not be valid, since the normality assumption does not hold true

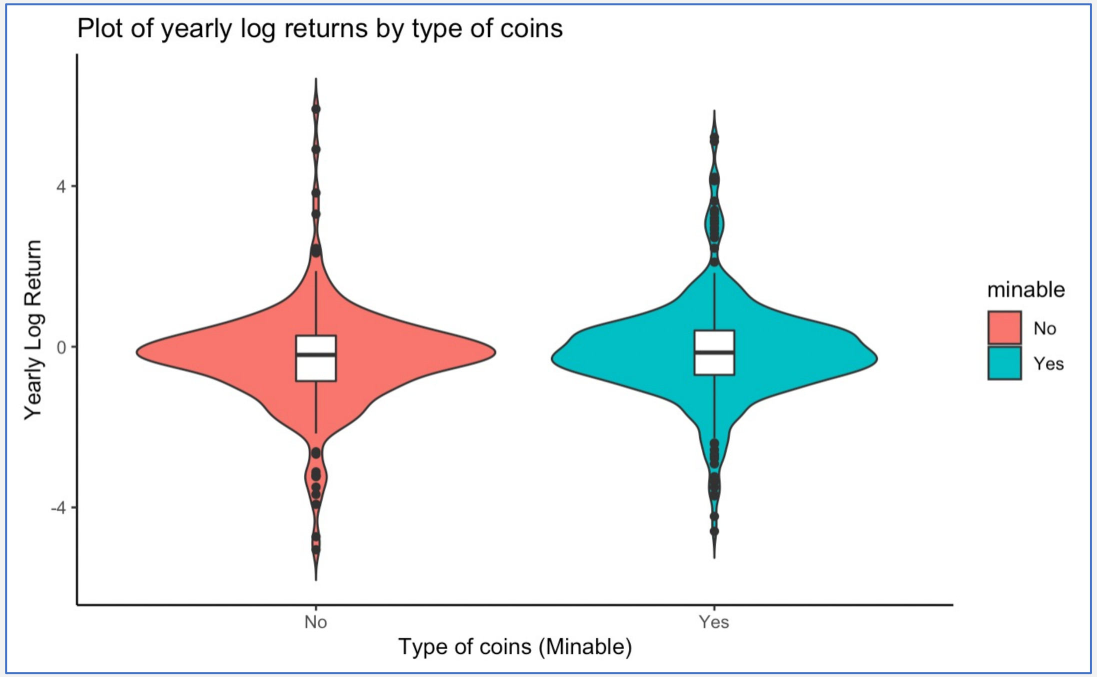
Question 2

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

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

Dataset

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



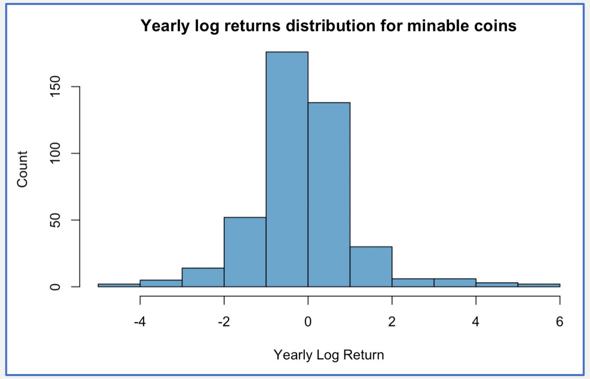
Question 2 – Hypothesis Testing

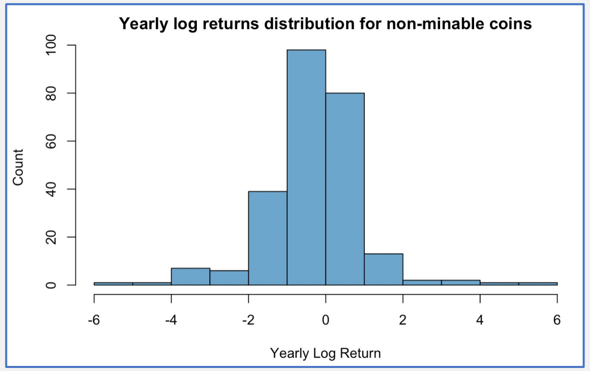
Test: Right-tailed equal variance 2-sample t-test

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

Alternative hypothesis: The mean of yearly average returns for coins is greater than tokens. μ (mineable) > μ(non-mineable)

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





Test Assumptions

- **Independence :** Strongly Holds true

- **Large sample size or normality:** Strongly holds true

- **Equal variance:** The values of variances come out to be 1.451 and 1.528 which are not very different, hence we assume equal variances.

Question 2 – Results & Interpretation

Results

Data	Test statistic	P-value	Result
Full data	1.774	0.03825	Rejects null hypothesis for equal means
Top 20	-0.1721	0.129	Does not reject null hypothesis for equal means

Interpretation

At significance level of 0.05, we **reject the null hypothesis** of equal means for 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 having higher yearly log returns as compared to non-minable coins**

Note: For top 20 cryptocurrencies, the results contradict but might not be valid, since the normality assumption does not hold true

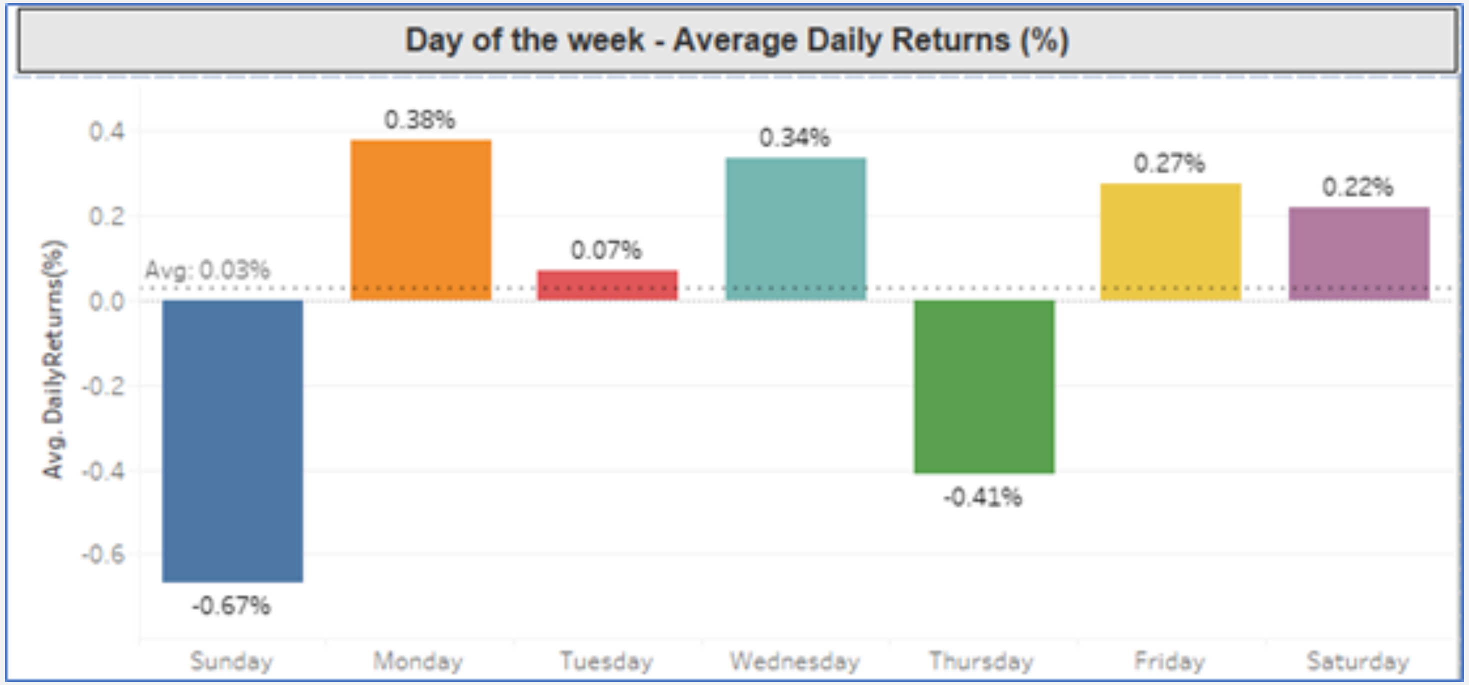
Question 3 – Background (Market Sentiment)

THEORY #1

Since the market isn't as active over the weekend. Getting in before the market starts back up means you have a better chance of landing a good price.

THEORY #2

As people start to buy Bitcoin on Monday, the price and demand increase. Once the week finishes, the demand drops off. The cycle then continues each week.



Question 3

Are Mondays more favorable for trading cryptocurrencies?

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

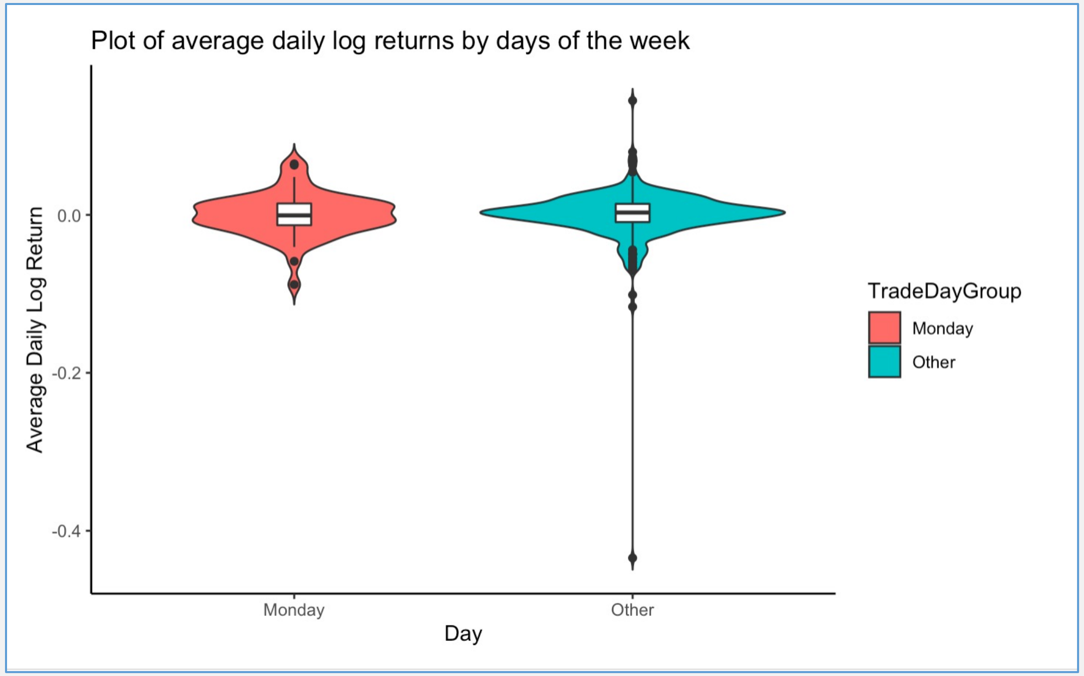
Dataset

- **Variables:** Trade Date, Volume, Price, Cryptocurrency name
- **Derived fields**:

• Log_Return = log (Today's price / Yesterday's price)

• day_of_the_week = derived from using the trade_date (YYYY-MM-DD format)

• Avg_DailyLogReturns = average(Log_Return grouped by day_of_the_week



Question 3 – Hypothesis Testing

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

Null hypothesis: μ (monday) = μ (any other day)

Alternative hypothesis: The mean of yearly average returns for coins is greater than tokens.
μ (monday) > μ(any other day)

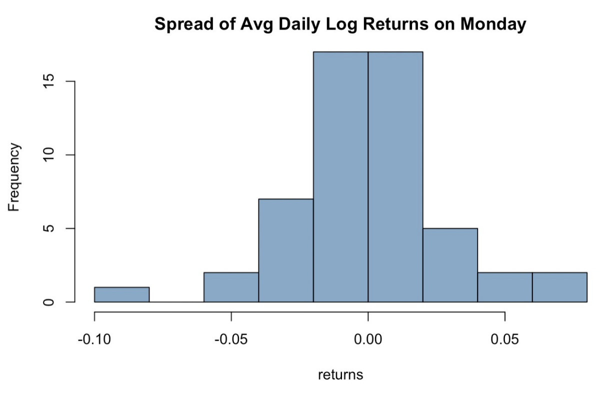
Test Assumptions

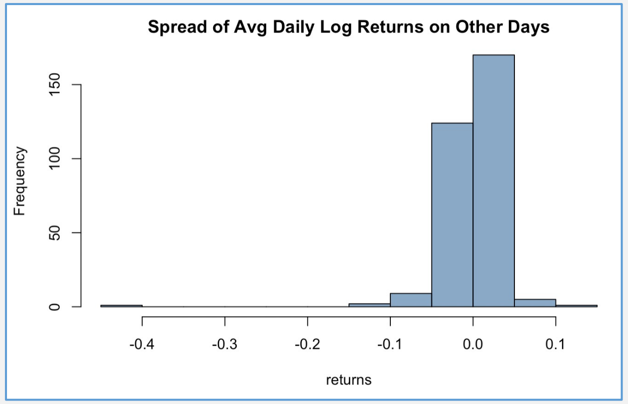
- **Independence :** Strongly holds true

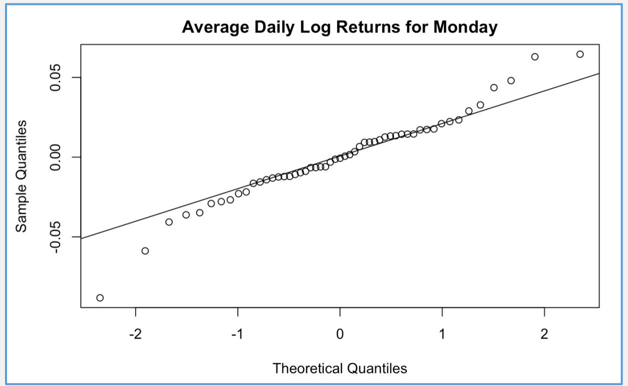
- **Large sample size or normality:** Holds true (large sample size)

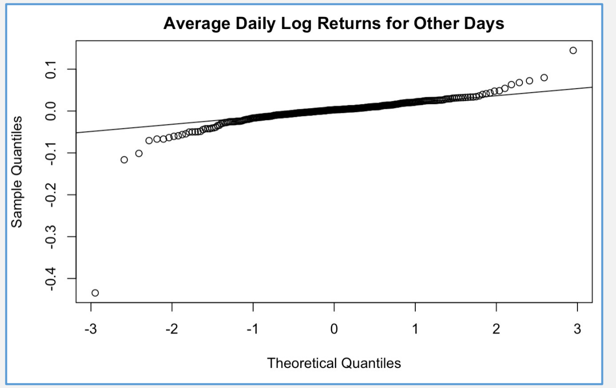
- **Equal variance:** The values of variances come out to be 0.00074 and 0.00125 which are not very different, hence we assume equal variances

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









Question 3 – Results & Interpretation

Results

Data	Test statistic	P-value	Result
Full data	0.86265	0.1942	Does **not** reject null hypothesis for equal means
Top 20	0.00474	0.4981	Does **not** reject null hypothesis for equal means

Interpretation

At 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.

Question 4

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

The motive is to find if an investor can rely on the market sentiment for his investment plan. Which means, we are interested to find if the currencies' returns are correlated with the volume traded on a specific day. We also aim to analyze cross correlation between returns and a lag in the traded volume by a few days.

Dataset

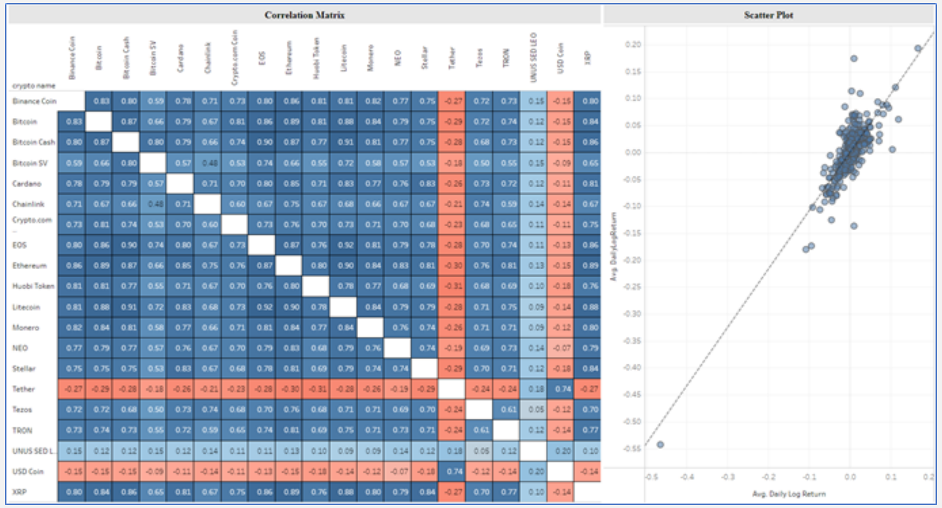
- **Variables:** Trade Date, Traded Volume, Log Daily Return.
- **Derived fields**: Log Daily Return – log (Today's price / Yesterday's price)
- **Sample size:** 20 crypto currencies and each coin/token has ~365 data points (Will change based on different lag values as the series hang out)

Question 4 – Statistical Analysis

Method: Cross correlation between Traded Volume and daily return of the cryptocurrencies

We are considering lag values up to 20 for each of the currencies to identify the correlation at different lag values.

Positive lag means the number of days the return data has been brought forward in time i.e, +2 lag indicates we are computing the coefficient with present return and volume two days ago. Negative lag means the otherwise.

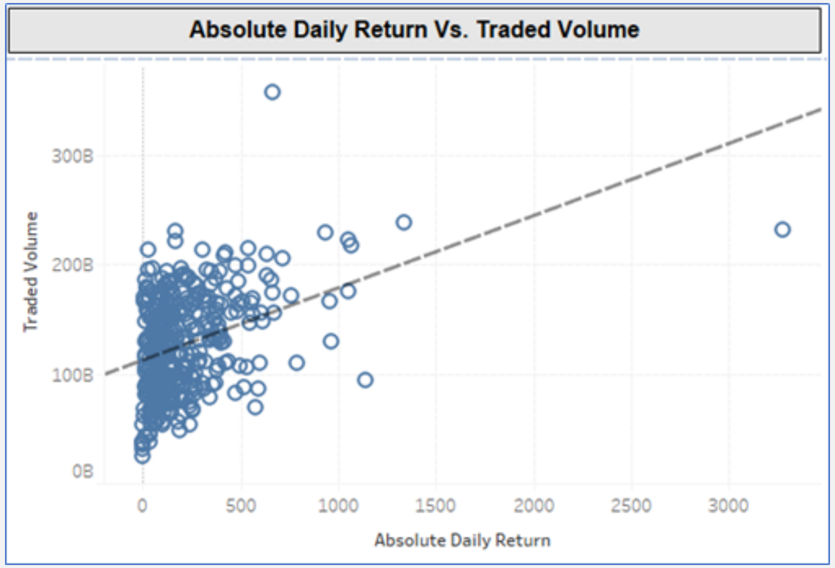


Test Assumptions

- **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 a specific crypto on a day

- **Absence of outliers**: Data points that are above 3.29 standard deviations from the mean are filtered out.



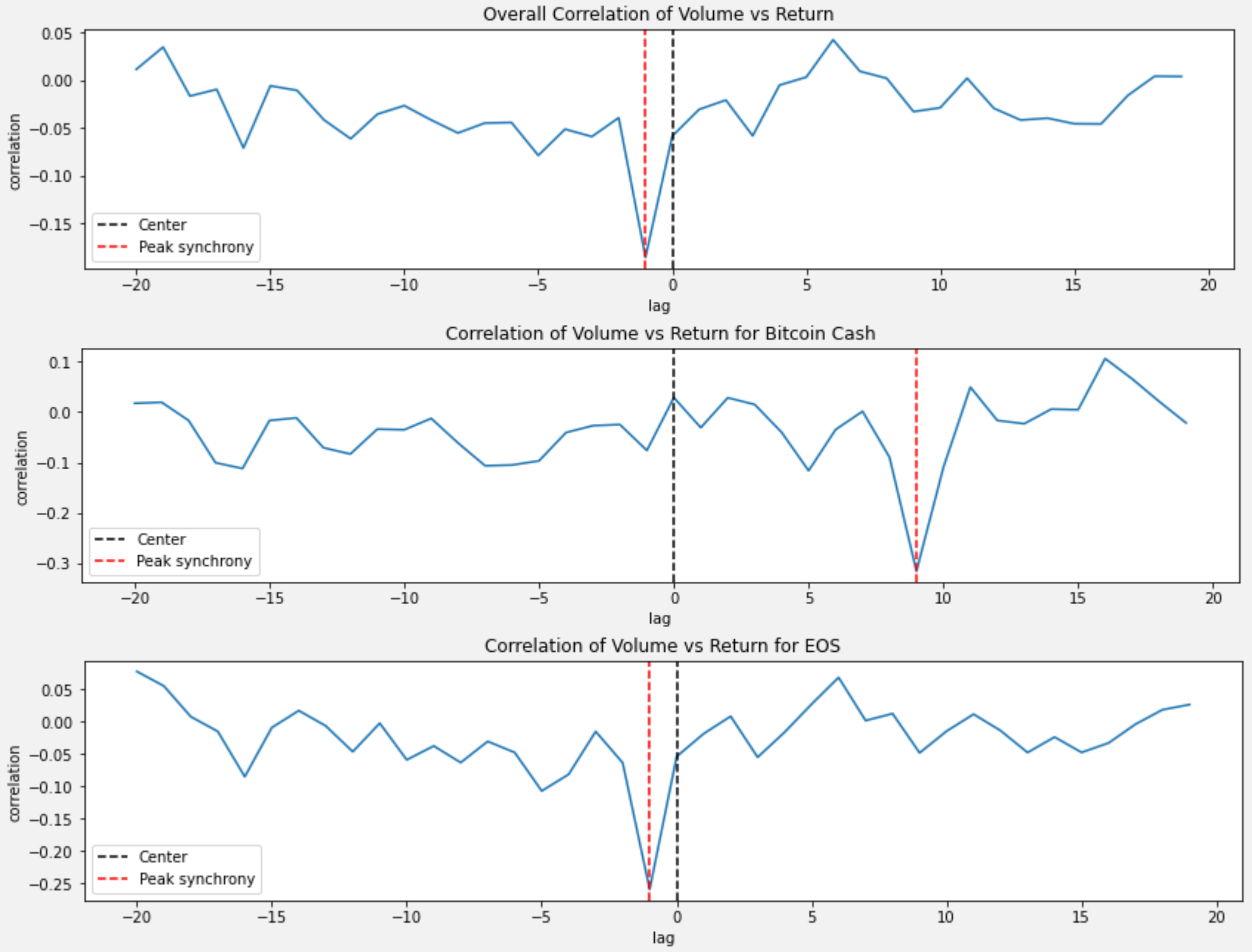
Question 4 – Results & Interpretation

Correlation coeff. for returns with lag/lead in traded volume.

	-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

Question 4 – Results & Interpretation

Results



Interpretation

We have produced a Correlogram (correlation coefficient vs lag) for the overall volume VS returns. We have found that the overall correlation has its peak of -0.15 at –1 day lag, indicating an association between 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 found low correlations for most data except for two:

- Bitcoin Cash -0.316 with lag 9 days
- EOS -0.26 with lag -1 days

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."

Overall Conclusion

- From Test 1, we have evidence for suggesting an investor to 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 to **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. Also, it might be a good idea to invest on Bitcoin Cash, and have special attention to the volume nine days ago.

Future Work

- ARCH, GARCH, TARCH
- Portfolio analysis
- Specific top cryptocurrency comparison and suggest changes
- Predict future prices for specific top cryptocurrencies

Challenges and Learnings

- Application of statistical tests on real world messy data sets
- Ability to identify appropriate tests for a given problem
- Aggregating the time-series data such that Independence assumptions hold true

Questions?

Thank You!

Appendix

````
```{r}
#Avg Daily Returns b/w Monday and Other Days
monday<- (df[df$TradeDayGroup == 'Monday',])$Avg_DailyLogReturn
other<- (df[df$TradeDayGroup == 'Other',])$Avg_DailyLogReturn
t_test = t.test(monday, other, var.equal = T, alternative = "greater")
t_test
```
````

```
	Two Sample t-test

data:  monday and other
t = 0.0047427, df = 363, p-value = 0.4981
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.008384144          Inf
sample estimates:
    mean of x     mean of y 
-0.0001787078 -0.0002028900 
```

```{r}
#Average log yearly returns of tokens is greater than coins for the year 2019-2020
coins_sample <- df[df$crypto_type=='Coin',]
tokens_sample <- df[df$crypto_type=='Token',]

print (paste(mean(coins_sample$YearlyLogReturn), mean(tokens_sample$YearlyLogReturn)))
print(paste(var(coins_sample$YearlyLogReturn), var(tokens_sample$YearlyLogReturn)))

t.test(coins_sample$YearlyLogReturn, tokens_sample$YearlyLogReturn, var.equal=TRUE, alternative='less' )
```

```
[1] "-0.181873766423358 0.180619576271186"
[1] "1.48440766922848 1.68033293687546"

        Two Sample t-test

data:  coins_sample$YearlyLogReturn and tokens_sample$YearlyLogReturn
t = -4.4465, df = 1037, p-value = 4.834e-06
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf -0.2282798
sample estimates:
 mean of x  mean of y
-0.1818738  0.1806196
```

```{r}
#Average log yearly returns of non-minable is greater than minable for the year 2019-2020

coins_sample <- df[df$crypto_type=='Coin',]

minable <- coins_sample[coins_sample$minable=='Yes',]
non_minable <- coins_sample[coins_sample$minable=='No',]

print (paste(mean(minable$YearlyLogReturn), mean(non_minable$YearlyLogReturn)))
print(paste(var(minable$YearlyLogReturn), var(non_minable$YearlyLogReturn)))

t.test(minable$YearlyLogReturn, non_minable$YearlyLogReturn, var.equal=TRUE, alternative='greater' )
```

```
[1] "-0.119167004608295 -0.290299003984064"
[1] "1.45133776239008 1.5289931307522"

        Two Sample t-test

data:  minable$YearlyLogReturn and non_minable$YearlyLogReturn
t = 1.7741, df = 683, p-value = 0.03825
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.01224944        Inf
sample estimates:
mean of x mean of y
-0.119167 -0.290299
```