7COM1079-0901-2024 - Team Research and Development Project

Final report title: Bitcoin

Group ID: A96

Dataset number: DS100

Prepared by: Udyanraje Bhosale 23023589
Muhammad Ahmed 23010566
Muhammad Hasnain 23031497
Muhammad Awais 22080974
Muhammad Faizan Tariq 23013839

University of Hertfordshire
Hatfield, 2024

Table of Contents

1. Introduction

1.1. Problem statement and research motivation **(100 words)**

The leading cryptocurrency, Bitcoin, has drawn much attention from investors and researchers with its volatile nature. How open and close prices of Bitcoin relate to each other is very important in predicting market trends for better investment decisions. Previous literature suggests that price changes in cryptocurrencies are highly correlated over short time horizons because of their speculative nature and sentiment in the market (Kristoufek, 2015). This study seeks to ascertain whether there is a significant correlation between the opening and closing prices of Bitcoin. The "bitcoin cz.csv" dataset shall be used to test the null hypothesis of no correlation against the alternative hypothesis that a positive correlation exists.

1.2. The data set **(75 words)**

This is a historical Bitcoin trading dataset that includes seven columns: Date, Open, High, Low, Close, Adj Close, and Volume. The column named "Date" represents the trading date. Opening, High, Low, and Close refer to the opening, highest, lowest, and closing prices for each day, respectively. The "Adj Close" represents the closing price adjusted for stock splits and dividends. Finally, "Volume" means the number of Bitcoin transactions on each trading day. This dataset encompasses several days and thus allows the analysis of price trends and correlations. The unit of measurement for Bitcoin is satoshi, the smallest denomination of Bitcoin, equivalent to 0.00000001 BTC.

1.3. Research question **(50 words).**

“Is there a significant correlation between the opening (Open) and closing (Close) prices of Bitcoin?”

To answer this question, we treated the Close price as the dependent variable and the Open price as the independent variable. Since the dependent variable did not follow a normal distribution, we utilized the Spearman correlation method to measure the strength and direction of the monotonic relationship between the two variables.

1.4. Null hypothesis and alternative hypothesis (H0/H1) **(100 words)**

This study tries to establish if there is any significant correlation between the Closing price and Opening price of Bitcoin.

H_0 (Null Hypothesis): The Closing price of Bitcoin, which is Close, does not relate to its Opening price, which is Open. That is to say, in this hypothesis, both variables are independent of each other.

Alternative Hypothesis (H_1): There is a relationship between the Closing price and the Opening price of Bitcoin; therefore, for every change in one variable, there is a regular change in another variable either positively or negatively.

2. Background research

2.1. Research papers (at least 3 relevant to your topic / DS) **(200 words)**

Here's a draft for the Research Papers Review:

Understanding the correlation between Bitcoin's open and close prices is essential for predicting market movements and improving trading strategies. Multiple studies have investigated Bitcoin price dynamics, providing valuable insights:

Kristoufek (2015) explored the main drivers of Bitcoin prices using wavelet coherence analysis, revealing that speculative behavior and market sentiment significantly influence cryptocurrency prices. This highlights the importance of studying price correlations.

Phillip, Chan, and Peiris (2018) examined the behaviors of cryptocurrency markets and found that many times, price movements are highly interdependent; thus, the relationship between open and close prices is of great relevance.

Bouri, Molnár, Azzi, Roubaud, and Hagfors (2017) investigated the volatility of Bitcoin and its role as a diversifier in investment portfolios, emphasizing how important it is to understand price trends for informed decision-making
.
While these studies provide insight into the market behavior of Bitcoin, none of them directly address the correlation between open and close prices, thus leaving a gap in research that our study will aim to fill. Our findings will add to the general understanding of cryptocurrency price trends and may guide future work on predicting Bitcoin prices and optimizing trading strategies.

2.2. Why RQ is of interest (research gap and future directions according to the literature) **(100 word**s)

This question is interesting because the correlation of open and close prices gives insight into market behaviour and predictability of the prices, important for investors and traders. Available literature on Bitcoin's volatility and its market drivers and interdependencies by Kristoufek, 2015 and Phillip et al., 2018 are scant, while no focused study concerning open-close price relations has been undertaken. This research closes that gap through correlation analysis and hence provides a different perspective on short-term price tendencies.

Further studies may extend the results by considering the correlations in other cryptocurrencies or, more interestingly, incorporating external factors such as market sentiment and macroeconomic events to enhance their predictive models.

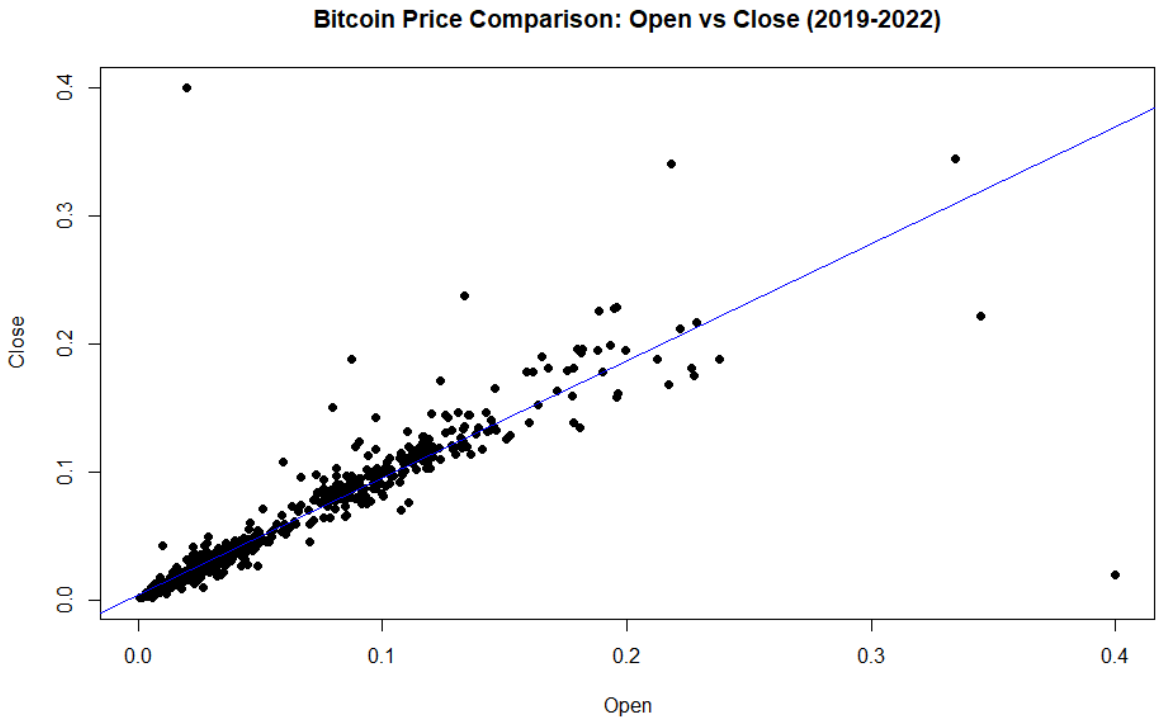
3. Visualisation

3.1. Appropriate plot for the RQ (**50 words)**

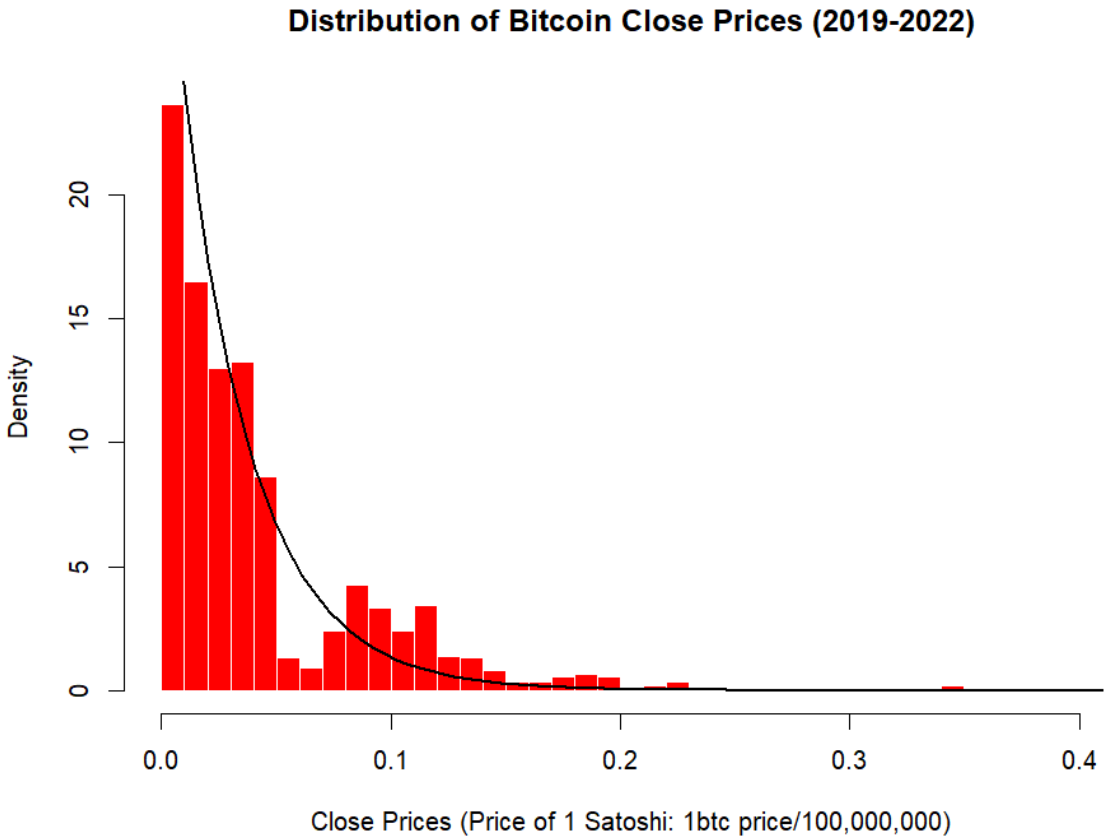
Explanation of plot choice:

Scatter Plot: The scatter plot is selected to show the relationship between the independent variable, Open, and dependent variable, Close. It will allow us to see the strength and direction of their relationship.

Histogram: The histogram is used to analyze the distribution of the independent variable, Open. Fewer bins are used in order to capture the overall patterns without overcomplicating the visualization.



Caption for the Scatter Plot
Title: Scatter Plot: Correlation Between Open and Close Prices
X-axis: Open Price in Satoshi
Y-axis: Close Price in Satoshi



Caption for the Histogram
Title: Histogram: Distribution of Open Prices
X-axis: Open Price in Satoshi
Y-axis: Frequency

3.2. Useful information for the data understanding **(50 words)**

The plot indicates that the independent variable, Open, is not normally distributed; hence, it is skewed. Also, from the scatterplot, the correlation between the Open and Close of Bitcoin reflects a good relationship between the opening and closing prices of Bitcoin on each trading day.

4. Analysis

4.1. Statistical test used to test the hypotheses and output (**75 words)**

Since the dependent variable was not normally distributed, we opted for Spearman rank correlation. Spearman's rank correlation is a non-parametric approach and denotes how strong a relationship exists between different variables in terms of monotonic dependence. This fits our research questions because our questions assess the relationship that exists between Opening and Closing Prices.

RESULTS:
Spearman rank correlation coefficient: 0.9816525
- Test statistic value: 3841393
- p-value: 0

The zero p-value implies the highest possible evidence against the null hypothesis; therefore, rejection and an assurance of the statistical existence of a significant relationship between Open and Close prices.

4.2. The null hypothesis is rejected /not rejected based on the p-value (**100 words)**

Reject null with p-value of 0: A p-value of 0 presents very strong evidence against the null hypothesis; thus, from here, one should decide there is no significant relationship between open and close prices. Since that is below the generally accepted level of 0.05, one would strongly reject the null hypothesis. This therefore means that the relationship between Open and Close Prices of Bitcoin exists at a statistical significance level. The strength and direction of the relationship are supported by the Spearman rank correlation coefficient of **0.9816525.**

5. Evaluation – group's experience at 7COM1079

5.1. What went well **(75 words)**

Our group was quite collaborative, and the roles were well divided throughout. We had effectively divided the work based on individual strengths, which ensured smooth progress on the project.
Through regular communication, challenges were quickly resolved, and everyone contributed significantly to the research and analysis. The team was easy to work with, maintaining a positive and productive environment to ensure the successful completion of project tasks such as data analysis and writing of reports.

5.2. Points for improvement **(75 words)**

The main area of improvement concerning our experience within the group can be identified regarding time management: There were phases, especially

for the cleaning and the analysis part of the project, which may be smoother by better planning. Also, during the first days, communicating better among the group regarding the roles that each should handle would avoid some confusion. In the future, we could have more rigid timelines and even intermediate check-ins to make sure we are on pace throughout the project.

5.3. Group's time management (**50 words)**

Time management could have been done better for the group, as some tasks took longer than anticipated, especially at the cleaning and analysis of data stages. Miscommunication and lack of clear deadlines were the causes of delays. Setting clearer timelines and periodic check-ins in the future would ensure more efficient progress.

5.4. Project's overall judgement (**50 words)**

Overall, it was a very successful project with solid analysis and much meaningful insight from the Bitcoin dataset. The group collaborated really well, and the research question was well addressed. On a future project, some areas for improvement would be in time management and task coordination. The outcome was of value and met expectations.

5.5. Note any changes to the group since submission of Assignment 1. Add new or amended GitHub Ids for new members **(75 words, write only if applies to your group arrangements)**

5.6. Comment on the GitHub log output **(50 words)**
Commit Message: `Update teamResearch.R`
Explanation: This update reflects a refinement in the project codebase, potentially improving the functionality and accuracy of research processes. This change enhances team collaboration and data quality.

Commit Message: `Create teamResearch.R`

Explanation: This commit marks the creation of a critical file (`teamResearch.R`), which forms the foundation of subsequent developments. It demonstrates the project's initiation of key analytical functionality.

Commit Message: `Delete teamResearch.R`

Explanation: This removal likely indicates a pivot or restructuring in the project's scope, emphasizing adaptability and a commitment to maintaining code relevance and efficiency.

6. Conclusions

6.1. Results explained (**75 words)**

Our analysis returned a very strong statistically significant correlation in Bitcoin's Open and Close prices, as evidenced by a Spearman rank correlation coefficient of 0.9816525. That means the pattern in the closing price usually follows up with its opening price being high or low. Thus, the p-value of 0 supports the rejection of the null hypothesis and ascertains the fact that the relation between these two variables is not just by random chance, but rather a meaningful connection.

6.2. Interpretation of the results (**75 words)**

Indeed, according to these results, it follows from our research hypothesis that Bitcoin Open and Close are strongly related to one another, and for any of them being changed; another can surely be guessed and would act consequently in their market or strategize it accordingly. The result indicates that improved market forecasting and decision-making in Bitcoin trading are possible, as the understanding of the early price trend can lead to an understanding of daily price behavior and future market movements.

6.3. Reasons and/or implications for future work, limitations of your study (**50 words)**

Other possible avenues of future research might comprise investigating other parameters-such as trading volume or market sentiment-that drive Bitcoin's price behavior. Regarding this paper, one should bear in mind the limitation that this work cannot generalize for sudden market fluctuation or events that previous data cannot catch. An extension with more variables could give a better overview.

7. Reference list ***(not included in the work count)***

Field, A. (2013) ***Discovering Statistics Using IBM SPSS Statistics***. 4th ed. London: SAGE Publications.

Bouri, E., Jain, A., Lee, K., & Saka, A. (2017) 'On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?', *Finance Research Letters*, 20, pp. 27-34. Available at: https://doi.org/10.1016/j.frl.2016.10.010 (Accessed: 6 January 2025).

Nakamoto, S. (2008) 'Bitcoin: A Peer-to-Peer Electronic Cash System', *Bitcoin.org*. Available at: https://bitcoin.org/bitcoin.pdf (Accessed: 6 January 2025).

8. Appendices

A. R code used for analysis and visualisation ***(not included in the word count)***
Analysis.R code with the appropriate statistics to test the hypotheses.

```
## Setting the working directory to where the dataset is!
getwd()
setwd("G:/team-research/ds&code")
## Importing the Dataset!
dataset <- read.csv("./Bitcoin CZ.csv")
head(dataset,2)
## Display the first 2 rows of the dataset!
head(dataset, 2)
## Running some descriptive Stats on our dependent and independent
        variables!
print("The data covers a time range or interval of:")
min_timestamp <- min(dataset$Date)
max_timestamp <- max(dataset$Date)
cat("Minimum timestamp:", min_timestamp, "\n")
cat("Maximum timestamp:", max_timestamp, "\n")
print('This shows we are dealing with 3 years of Historical Data!!')
## Total number of Records!!
total_records <- nrow(dataset)
# Printing the total number of records
cat("Total number of records in the dataset:", total_records, "\n")
print('Our dependent Variable is Close(price) for any particular day')
print('Our Independent Variable is Open(price) for any particular day')
## Lets deal with null Values first!
## Lets check a Volume value of zero that indicates periods of no trading
        activity,
## which might be useless still we will keep these records as we are only
        concerned
## with the open and close!!
zero_count<- sum(dataset$Volume==0)
cat("Number of Days with no trading activity are: ",zero_count)
# Count rows where 'Open' has "null"
null_open_count <- sum(dataset$Open == "null")
# Count rows where 'Close' has "null"
null_close_count <- sum(dataset$Close == "null")
# Print the results
cat("Rows with 'null' in 'Open':", null_open_count, "\n")
cat("Rows with 'null' in 'Close':", null_close_count, "\n")
# As significant amount of rows have these null values we
# use Replace Missing Values with Surrounding Non-Null Average.
install.packages("zoo")
library(zoo)
# Replace "null" with NA
dataset[dataset == "null"] <- NA
# Replace NA values with surrounding average in 'Open' column
dataset$Open <- na.approx(dataset$Open, na.rm = FALSE)
# Replace NA values with surrounding average in 'Close' column
dataset$Close <- na.approx(dataset$Close, na.rm = FALSE)
# Checking the null count again
null_open_count_again <- sum(dataset$Open == "null")
null_open_count_again
```

```
null_close_count_again <- sum(dataset$Close == "null")
null_close_count_again
# Descriptive stats about our dependent and independent variables:
summary(dataset$Open)
summary(dataset$Close)
# Histogram for the dependent variable to see data spread!
hist(dataset$Close,
    main = "Distribution of Bitcoin Close Prices (2019-2022)",
    xlab = "Close Prices (Price of 1 Satoshi: 1btc price/100,000,000)",
    col = "red",
    border = "white",
    probability = TRUE,
    breaks = 50)  # Adjust this number for finer bins
# Fit an exponential decay curve to the histogram
# Define the exponential function (decay)
decay_function <- function(x, a, b) {
  a * exp(-b * x)
}
# Fit the model (using nonlinear least squares fitting)
fit <- nls(density ~ decay_function(x, a, b),
          data = data.frame(x = hist(dataset$Close, plot = FALSE)$mids,
                            density = hist(dataset$Close, plot = FALSE)$density),
          start = list(a = 1, b = 0.001))
# Extract the fitted parameters
params <- coef(fit)
# Add the exponential decay line to the histogram
curve(params["a"] * exp(-params["b"] * x),
      col = "black",
      add = TRUE,
      lwd = 2)
# Time for the scatter plot to get some insights about the relationship between
# our dependent and independent variables!
x <- dataset$Open
y <- dataset$Close
plot(x,
    y,
    main = "Bitcoin Price Comparison: Open vs Close (2019-2022)",
    xlab = "Open",
    ylab = "Close",
    pch = 19,
    frame = T)
model <- lm(x ~ y, data = dataset)
abline(model, col = "blue")
# Performing Spearman Correlation:
result <- cor.test(x,y, method="spearman")
cat("Spearman Correlation Coefficient:", result$estimate, "\n")
cat("P-value:", result$p.value, "\n")
cat("Test Statistic is: " , result$statistic)
```

B. GitHub log output.

	A	B	C	D	E	F
1	91d7368	udyanraje	25-11-2024	Udyanraje Commit		
2	cbe91c0	Muhamma	25-11-2024	Faizan commit		
3	f3f1a2a	ahmedher	21-11-2024	Update teamResearch.R		
4	7719d5d	ma23att	21-10-2024	Create teamResearch.R		
5	07f77b7	udyanraje	07-11-2024	Delete teamResearch.r		
6	f7b30ef	ma23att	25-11-2024	Create teamResearch.r		
7						