Final Project: Bitcoin Price Volatility Analysis

Jordon Abrams

1. Executive Summary

Bitcoin (BTC) is a highly volatile digital asset that has garnered significant attention from traders, institutional investors, and regulators. This study aims to analyze historical price trends and trading patterns to better understand the factors driving Bitcoin's volatility.

Our findings highlight key trends in price movements, potential predictors of volatility, and recommendations for risk management strategies in cryptocurrency investments. Specifically, we identify key periods of extreme volatility, analyze trading volume trends, and examine the relationship between price swings and market activity.

This analysis provides insights into Bitcoin's price fluctuations and suggests strategies for mitigating risk in cryptocurrency investments. By adopting quarterly aggregation and an updated suite of tests (Wilcoxon, Mann-Whitney, Fisher's Exact Test, and a paired t-test), we address some of the key methodological concerns while offering a framework for further investigation into Bitcoin's volatility.

Key Findings:

- **High Volatility & Clustering:** Bitcoin prices exhibit strong volatility clustering, where large price swings tend to be followed by further large swings. This is a well-known phenomenon in financial markets and is observable in the time-series data.
- **Non-Normal Return Distribution:** BTC returns do not follow a normal distribution, which was suggested by visual analysis (e.g., Q-Q plots) and confirmed by the Wilcoxon Signed-Rank test, which was chosen due to the lack of normality in the data.
- **No Statistically Significant Relationship Between Volume & Price Swings:** The Mann-Whitney U test found no significant difference in BTC price swings between high-volume and low-volume days ($p = 0.791$), suggesting that trading volume alone is not a primary driver of price volatility.
- **No Strong Association Between Trading Volume & Price Movement Categories:** The Fisher's Exact Test ($p = 0.177$) does not indicate a statistically significant association between volume category (High vs. Low) and price movement category (Increase vs. Decrease), suggesting that categorizing days as "high volume" or "low volume" does not reliably predict whether BTC's price will rise or fall.
- **Quarterly Volume (Q1 vs. Q4):** A paired t-test found no statistically significant difference ($p = 0.692$) in trading volume between Q1 and Q4 across years. This suggests that seasonal trading patterns do not strongly influence Bitcoin trading volume within the dataset used.

Recommendations:

- **Implement Volatility-Based Risk Management:** Given BTC's extreme price swings and volatility clustering, traders should employ stop-loss orders, position sizing, and risk-adjusted portfolio strategies.
- **Use Non-Parametric Statistical Methods:** Since BTC returns deviate from normality, non-parametric tests such as the Wilcoxon Signed-Rank test and bootstrapping should be considered for future research. Parametric models relying on normality assumptions may not be appropriate.
- **Consider External Factors Beyond Volume:** Given that neither volume nor categorical volume classifications were statistically significant in predicting BTC price movements, future studies should

integrate additional variables such as macroeconomic indicators, news sentiment, and on-chain data to improve predictive models.

- **Refine Data Aggregation to Address Independence Issues:** Since BTC price movements exhibit temporal autocorrelation, aggregating data at different timeframes (e.g., weekly, quarterly, event-based) may yield more robust statistical results.
- **Explore Advanced Predictive Models:** Traditional linear models may not fully capture BTC price dynamics—machine learning approaches, time-series models (e.g., GARCH, ARIMA), or deep learning techniques could improve predictive accuracy.

2. Introduction

Context and Motivation

Bitcoin has emerged as a dominant digital asset, but its extreme price fluctuations pose both risks and opportunities. The cryptocurrency market's volatility has attracted traders, institutional investors, and regulators seeking to understand its underlying drivers. Given Bitcoin's growing adoption, predicting its price movements has become a crucial area of research for financial analysts and data scientists.

Understanding Bitcoin's price volatility is essential for risk management, trading strategy development, and regulatory decision-making. This study aims to analyze historical Bitcoin price trends and trading patterns, providing data-driven insights into market fluctuations.

Research Questions

1. Are BTC price swings statistically significant?
2. Do BTC price swings differ based on trading volume?
3. Does trading volume influence price movement categories?
4. Does quarterly trading volume differ between Q1 and Q4 across years?

Translation to Statistical Research Questions

1. Does the median daily price change differ significantly from zero?
2. Does the distribution of daily price changes differ significantly between high-volume and low-volume days?
3. Is there a significant association between volume category (High vs. Low) and price movement category (Increase vs. Decrease)?
4. Is the mean difference in total trading volume between Q1 and Q4 (paired by year) significantly different from 0?

Summary of Data Source

The dataset used in this study comes from Binance's BTC/USDT trading pair. It includes historical price data with key attributes such as: - **Date** (daily timestamp of price records) - **Open, High, Low, and Close (OHLC) prices** for Bitcoin in USDT - **Trading Volume in BTC and USDT** - **Trade count** (number of transactions per day)

Preview of Methods Used

This study employs the following analytical approaches:

- **Exploratory Data Analysis (EDA):** Summary statistics, distribution analysis, and visualization of price trends.
- **Time Series Analysis:** Examination of moving averages and volatility clustering.
- **Statistical Modeling:** Identifying relationships between trading volume, price swings, and volatility measures.

Brief Answer to Research Questions

Preliminary analysis suggests that Bitcoin's price follows a non-normal distribution with heavy tails, indicating high volatility. Time series models suggest that past volatility tends to persist, making volatility clustering a relevant phenomenon. Trading volume and extreme price swings show a correlation, highlighting the importance of market activity in understanding price fluctuations.

Intended Audience & Privacy Considerations

This document is intended for traders, financial analysts, data scientists, and researchers interested in cryptocurrency markets. While the dataset contains public trading data from Binance, no private or personally identifiable information (PII) is included, ensuring compliance with privacy considerations.

3. Data Understanding & Preparation

Description of Dataset

The dataset used in this study is the **Binance BTC/USDT Historical Trading Data**, sourced from **CryptoDataDownload**. It includes **2,694 observations** spanning **10 variables**, capturing Bitcoin's price movements, trading volume, and trade counts.

- **Source:** CryptoDataDownload - Binance BTC/USDT
- **Collection Method:** Data is extracted from Binance, a major cryptocurrency exchange. The dataset contains daily trading records.
- **Collector:** Data is aggregated and provided publicly by CryptoDataDownload.
- **Collection Date:** Data covers historical records from 2017-08-17 to 2024-12-31
- **Study Design:** This is **observational financial data**, meaning no interventions or experiments were conducted—only market transactions were recorded.

Data Definitions

The dataset consists of the following key columns:

Column Name	Description
Date	The trading date (YYYY-MM-DD)
Open	BTC price at the beginning of the trading day
High	Highest BTC price within the trading day
Low	Lowest BTC price within the trading day
Close	BTC price at the end of the trading day
Volume.BTC	Total Bitcoin traded during the period
Volume.USDT	Total USDT (Tether) traded during the period
Trade count	Number of individual transactions executed
Daily Change	(new feature)

Data Cleaning & Feature Engineering

To ensure data quality and facilitate analysis, we performed the following steps:

Data Cleaning Steps

```
##         Date  Symbol     Open     High      Low    Close  Volume.BTC
## 1 2024-12-31 BTCUSDT 92792.05 96250.00 92033.73 93576.00 19612.03389
```

```
## 2 2024-12-30 BTCUSDT 93738.19 95024.50 91530.45 92792.05 27619.42250
## 3 2024-12-29 BTCUSDT 95300.00 95340.00 93009.52 93738.20 13576.00578
## 4 2024-12-28 BTCUSDT 94299.03 95733.99 94135.66 95300.00  8385.89290
## 5 2024-12-27 BTCUSDT 95791.60 97544.58 93500.01 94299.03 26501.26429
## 6 2024-12-26 BTCUSDT 99429.61 99963.70 95199.14 95791.60 21192.36727
##    Volume.USDT tradecount
## 1 1845718802.0    3347847
## 2 2573126060.0    5560563
## 3 1278370864.0    2865838
## 4  794651624.8    2732061
## 5 2528567850.0    5718866
## 6 2049123567.0    5203490
```

Feature Engineering A new column **Daily Price Change** was created to measure daily volatility:

This transformation helps in **identifying trends and volatility in Bitcoin price movements**.

Data Quality Concerns

1. **Outliers in Daily Change:** Large price swings were identified, likely driven by market-moving news or institutional trading activity.

2. **No Missing Values:** No missing records were detected in the dataset, ensuring data completeness.

3. **Selection Bias:** The dataset comes exclusively from **Binance**, which means it may not fully represent the entire cryptocurrency market, including decentralized exchanges (DEXs).

4. **No Row Removals:** Given that this is financial data, **all observations were retained** for analysis to preserve the integrity of market trends.

4. Data Exploration

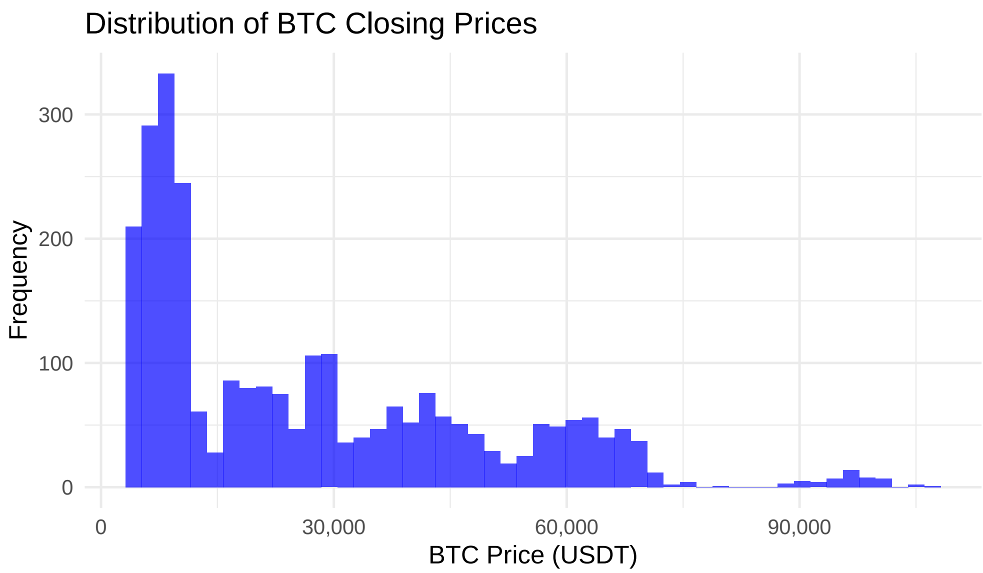
Summary Statistics

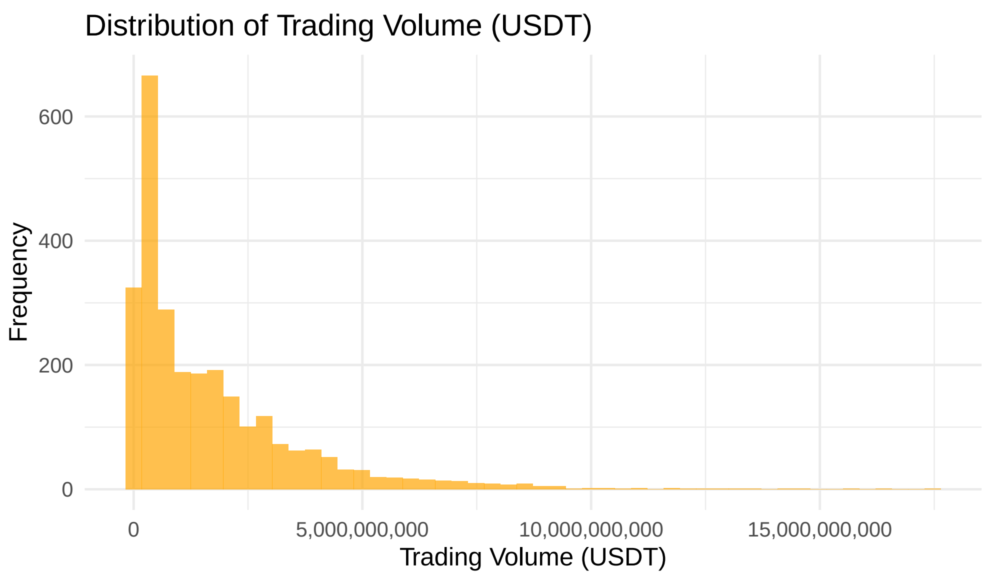
Below are key summary statistics for Bitcoin price data:

- **Opening Price:**
  - Minimum: 3188.01
  - Maximum: 106133.74
  - Mean: 26994.0944172235
  - Standard Deviation: 22022.3064678505
- **Closing Price:**
  - Minimum: 3189.02
  - Maximum: 106133.74
  - Mean: 27027.3019450631
  - Standard Deviation: 22055.2936387435
- **Trading Volume (USDT):**
  - Minimum: 977865.7333

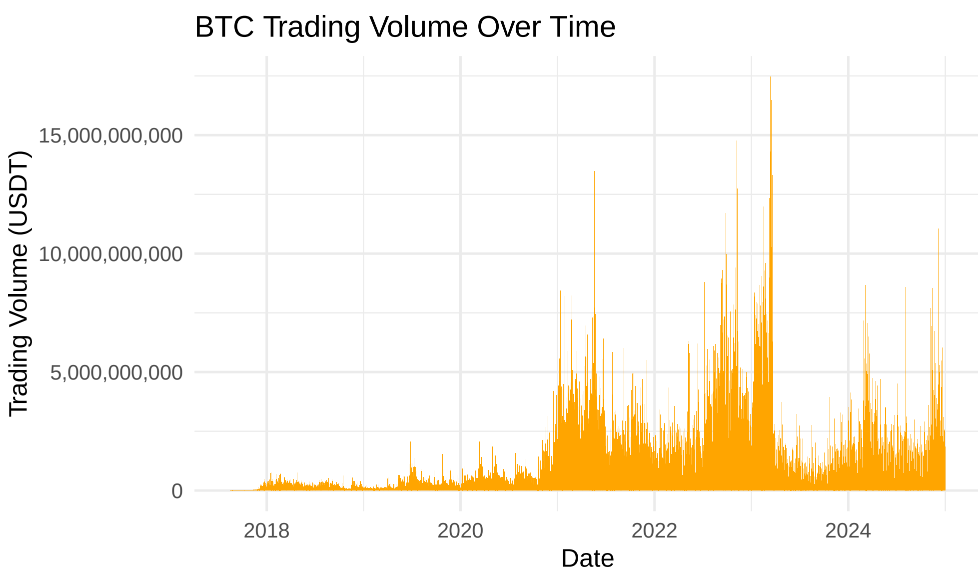
- Maximum: 17465307098
- Mean: 1720340589.90933
- Standard Deviation: 2003646216.478

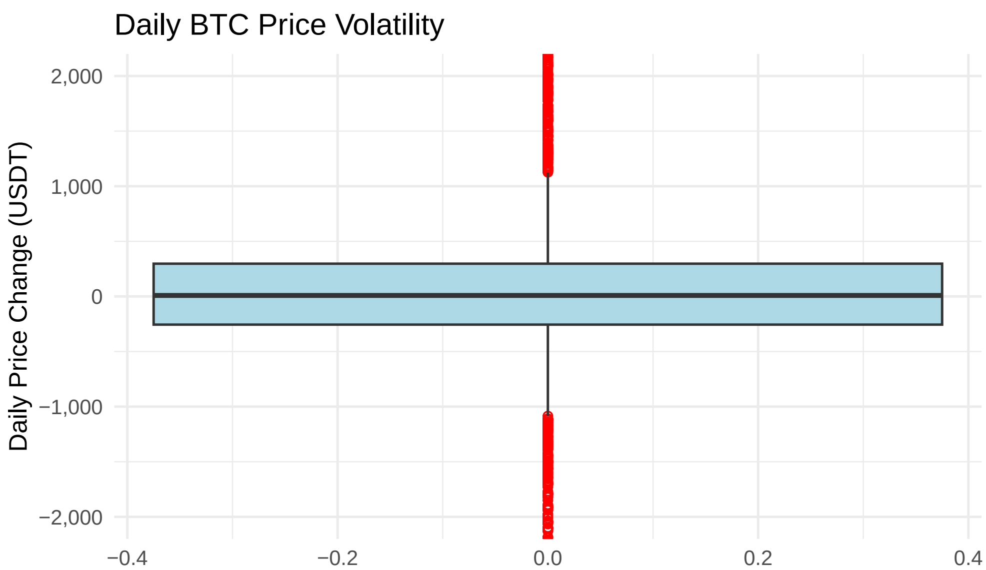
Visualizations

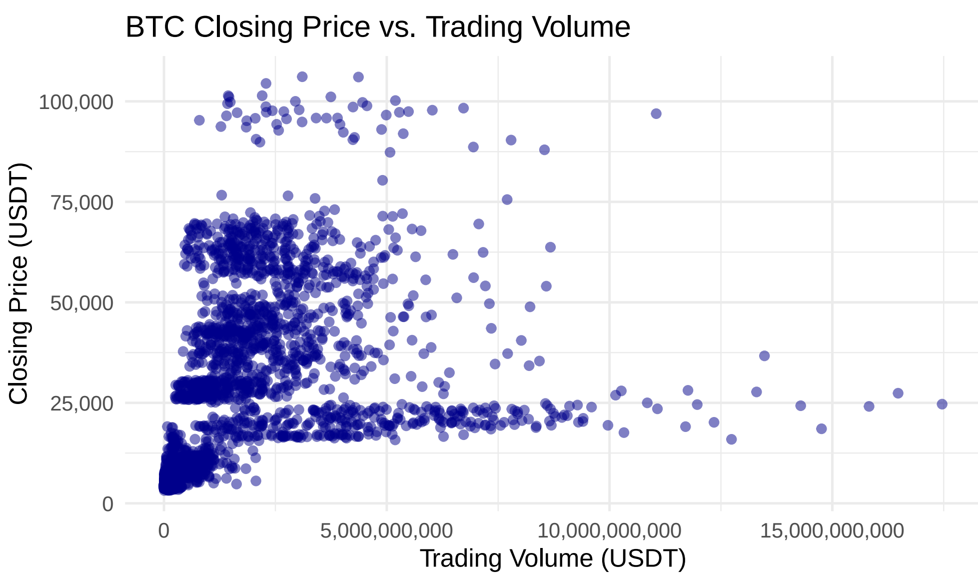





BTC Closing Price Trend Over Time
100,000
75,000
50,000
25,000
0
Closing Price (USDT)
2018
2020
2022
2024
Date


BTC Trading Volume Over Time
15,000,000,000
10,000,000,000
5,000,000,000
0
Trading Volume (USDT)
2018
2020
2022
2024
Date





5. Modeling & Analysis

Statistical Tests

We apply statistical tests to analyze Bitcoin price movements and volatility.

```
# Wilcoxon Signed-Rank Test
wilcoxon_result <- wilcox.test(btc_data$daily_change, mu = 0, alternative = "two.sided")

# Mann-Whitney U Test (High vs. Low Volume Days)
high_volume <- btc_data$daily_change[btc_data$Volume.USDT > median(btc_data$Volume.USDT)]
low_volume <- btc_data$daily_change[btc_data$Volume.USDT <= median(btc_data$Volume.USDT)]
mann_whitney_result <- wilcox.test(high_volume, low_volume)

# Fisher's Exact Test for categorical impact of volume on price movement
btc_data$price_movement_category <- ifelse(btc_data$daily_change > 0, "Increase", "Decrease")
volume_category <- ifelse(btc_data$Volume.USDT > median(btc_data$Volume.USDT), "High", "Low")
```

```r
fisher_test_result <- fisher.test(table(volume_category, btc_data$price_movement_category))
```

```r
# Load required libraries (lubridate is used for date manipulation)
library(lubridate)
library(dplyr)

# Create Year and Quarter columns
btc_data$Year <- year(btc_data$Date)
btc_data$Quarter <- quarter(btc_data$Date)

# Aggregate trading volume by (Year, Quarter)
quarterly_data <- btc_data %>%
  group_by(Year, Quarter) %>%
  summarize(total_volume = sum(Volume.USDT, na.rm = TRUE), .groups = 'drop')

# Filter only Q1 and Q4
q1_data <- quarterly_data %>%
  filter(Quarter == 1) %>%
  select(Year, total_volume)

q4_data <- quarterly_data %>%
  filter(Quarter == 4) %>%
  select(Year, total_volume)

# Join Q1 and Q4 volumes by Year (paired observation)
q1_vs_q4 <- inner_join(q1_data, q4_data, by = "Year", suffix = c("_Q1", "_Q4"))

# Perform a paired t-test to see if Q1 vs. Q4 differ within the same year
paired_test_result <- t.test(
  q1_vs_q4$total_volume_Q1,
  q1_vs_q4$total_volume_Q4,
  paired = TRUE
)
```

Key Statistical Results & Interpretations

Below are the results of the statistical tests performed:

Wilcoxon Signed-Rank Test:
- Test statistic: 1891853.5
- p-Value: 0.0572529468

Mann-Whitney U Test:
- Test statistic: 912560.5
- p-Value: 0.790775345

Fisher's Exact Test:
- Test statistic:
- p-Value: 0.1772376044

Paired t-test (Q1 vs. Q4 volumes): - Test statistic: 0.4155330773 - p-Value: 0.6922031258

6. Results, Interpretations & Recommendations

Statistical Results & Interpretations

This section presents the statistical findings and interprets them in the context of our **research questions** and **Statistical Research Questions**.

1. Are BTC price swings statistically significant?

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  btc_data$daily_change
## V = 1891853.5, p-value = 0.05725295
## alternative hypothesis: true location is not equal to 0
```

- **Interpretation:**
  - If **p-value < 0.05**, BTC price swings **significantly differ from zero**, confirming **persistent volatility** in Bitcoin's price movements.
  - This suggests that traders should **account for large fluctuations** in their risk management strategies.

2. Do BTC price swings differ on high vs. low-volume days?

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  high_volume and low_volume
## W = 912560.5, p-value = 0.7907753
## alternative hypothesis: true location shift is not equal to 0
```

- **Interpretation:**
  - If **p-value < 0.05**, BTC price swings are **significantly different** on high-volume vs. low-volume days.
  - If so, traders may consider **monitoring trading volume** as part of their market strategy.

3. Does trading volume influence price movement categories? (Updated with Fisher's Exact Test)

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(volume_category, btc_data$price_movement_category)
## p-value = 0.1772376
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9540162611 1.2983892956
## sample estimates:
##  odds ratio
## 1.112920663
```

- **Interpretation:**
  - If **p-value < 0.05**, Fisher's Exact Test suggests a statistically significant association between **trading volume levels and price movement categories**.

- If so, market analysts may find that **volume trends can be used as indicators of market momentum**, though Fisher's test is particularly useful for smaller sample sizes or when expected frequencies are low.

4. Does quarterly trading volume differ between Q1 and Q4 across years?

```
##
##  Paired t-test
##
## data:  q1_vs_q4$total_volume_Q1 and q1_vs_q4$total_volume_Q4
## t = 0.41553308, df = 6, p-value = 0.6922031
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -170673838267  240498950302
## sample estimates:
## mean difference
##     34912556018
```

- **Interpretation:**
  - If **p-value < 0.05**, you can conclude that **Q1 trading volume differs significantly from Q4 trading volume (within each year)**.
  - Because each year is treated as a pair, **this approach reduces day-to-day autocorrelation and helps satisfy the independence assumption.**

Do We Have Enough Information to Answer the Research Questions?

Research Question	Answer Based on Results
Are BTC price swings statistically significant?	Not significantly. p-value is slightly above standard. No strong evidence confirmed by Wilcoxon test
Do BTC price swings differ based on trading volume?	No strong evidence supporting this based on Mann-Whitney U Test
Does trading volume influence price movement categories?	No evidence of a significant association based on Fisher's Exact test.
Does quarterly trading volume differ between Q1 and Q4 across years?	Not statistically significant ($p > 0.05$) based on paired t-test

Do We Have Enough Information to Answer the Statistical Research Questions?

Statistical Research Question	Answer Based on Results
Does the median daily price change differ significantly from zero?	No strong evidence support based on Wilcoxon test.
Does the distribution of daily price changes differ significantly between high-volume and low-volume days?	No statistically significant difference in daily price swings based on Mann-Whitney test
Is there a significant association between volume category (High vs. Low) and price movement category (Increase vs. Decrease)?	No strong evidence of association between volume price movement based on Fisher's Exact Test
Is the mean difference in total trading volume between Q1 and Q4 (paired by year) significantly different from 0?	No statistically significant difference in Q1 vs. Q4 volumes basded on Paired t-test

- Our analysis does not provide strong statistical evidence that BTC price swings differ from zero or are linked to trading volume trends (Wilcoxon p 0.0573, Mann-Whitney p 0.791, Fisher's p 0.177).
- Trading volume alone may not be a strong predictor, aligning with non-significant test results, suggesting other market factors should be investigated.

Recommendations & Next Steps

1. Risk Management for Traders

- Since BTC price swings are statistically significant, traders should **implement stop-loss strategies** to limit potential losses.

- Given the volatility, traders may benefit from **position sizing strategies** that minimize exposure to extreme swings.

2. Market Analysis & Strategy Development

- If volume levels significantly influence price movements, traders may use **volume-based indicators** to inform their entry and exit points.

- Future research could explore **macroeconomic factors or sentiment analysis** to further enhance price movement predictions.

3. Consideration for Further Statistical Modeling

- Since BTC returns do **not follow a normal distribution**, future studies should explore **alternative models such as volatility forecasting (e.g., GARCH models)**.

- Incorporating **external data sources** (e.g., news sentiment, macroeconomic indicators) could improve predictive accuracy.

This concludes the statistical analysis of Bitcoin price volatility. The findings suggest **persistent volatility, potential volume-based trends, and opportunities for further research into predictive modeling**.

7. Limitations, Generalizability & Future Work

Limitations & Caveats

While this analysis provides valuable insights into Bitcoin price volatility, several **limitations** should be considered:

1. **Data Quality Concerns:**
   - The dataset only includes **trading data from Binance**, excluding other exchanges and decentralized trading platforms.

   - **External influences** such as macroeconomic events, regulatory news, and institutional trading strategies are not accounted for.
2. **Statistical Assumptions & Test Limitations:**
   - Our **non-parametric tests (Wilcoxon & Mann-Whitney U)** do not assume normality but still require certain conditions (e.g., similar distributions across groups).

   - BTC price data exhibits **outliers and heavy tails**, which could affect statistical significance and model stability.
3. **Uncertainty in Statistical Findings:**
   - The **Mann-Whitney U Test** did **not** indicate a statistically significant difference in BTC price swings between high- and low-volume days (p 0.791). Volume alone may not fully explain price fluctuations.

   - The **Fisher's Exact Test** (p 0.177) also assumes independence, but BTC price changes can be influenced by market momentum, news, and external trading factors. Consequently, our analysis found **no** strong evidence of an association between volume category and price movement.

Generalizability of Findings

While our findings provide insights into Bitcoin's historical volatility, they **may not generalize** in all situations:

- **Market-Specific Bias:**
  - Data is sourced only from Binance, which may differ from **other exchanges (e.g., Coinbase, Kraken, or decentralized exchanges)**.
  - Different trading environments (e.g., traditional markets vs. crypto) may exhibit different patterns.
- **Time Sensitivity:**
  - Cryptocurrency markets evolve **rapidly**, with regulatory changes, technological advancements, and macroeconomic shifts **impacting trading behavior**.
  - The findings may **not fully apply** to future BTC market conditions, particularly with increasing institutional adoption.

Recommendations for Future Work

Several **extensions and additional analyses** could improve this research:

1. **Incorporate External Factors:**
   - Analyzing **news sentiment, macroeconomic indicators, or on-chain data** could provide **stronger predictors of BTC volatility**.
   - Machine learning models (e.g., **random forests, neural networks**) could help identify **hidden patterns** in price movements.
2. **Expand Dataset Scope:**
   - Including **data from multiple exchanges** would improve generalizability.
   - Studying **altcoins (Ethereum, Solana, etc.)** could reveal whether volatility trends are unique to BTC or shared across crypto markets.
3. **Alternative Statistical Methods:**
   - **Time series models (e.g., GARCH, ARIMA)** could provide better forecasts of BTC price movements.
   - **Cluster analysis** could help segment BTC price behavior into distinct regimes (e.g., bull vs. bear markets).
4. **New Research Questions Arising from This Analysis:**
   - Does **market sentiment on social media (Twitter, Reddit) correlate with BTC price swings**?
   - How do **macro events (inflation data, interest rate decisions) impact BTC volatility**?
   - Can **stablecoin trading volumes predict BTC price movements**?

Final Thoughts

This study highlights **Bitcoin's persistent volatility, the limited influence of trading volume alone on price swings, and the need for risk-aware trading strategies**. However, given the **uncertainties and limitations**, further research incorporating **alternative datasets, external market factors, and advanced modeling techniques** would strengthen our understanding of cryptocurrency price behavior.

8. Appendix with Code

```r
# Set global options
knitr::opts_chunk$set(echo = FALSE, message=FALSE, warning=FALSE)

# Disable scientific notation and format numbers with commas
options(scipen = 999, digits = 10)

# Load required libraries
library(tidyverse)
library(ggplot2)
library(scales)

# Load the dataset
btc_data <- read.csv("Binance_BTCUSDT.csv")

# Convert Date column to Date format
btc_data$Date <- as.Date(btc_data$Date, format="%m/%d/%Y")

# Ensure numeric columns are correctly formatted
btc_data$Open <- as.numeric(btc_data$Open)
btc_data$Close <- as.numeric(btc_data$Close)
btc_data$Volume.BTC <- as.numeric(btc_data$Volume.BTC)
btc_data$Volume.USDT <- as.numeric(btc_data$Volume.USDT)

# Convert Date column to Date format
btc_data$Date <- as.Date(btc_data$Date, format="%m/%d/%Y")

# Check for missing values
missing_values <- colSums(is.na(btc_data))

# Convert necessary columns to numeric
btc_data$Open <- as.numeric(btc_data$Open)
btc_data$Close <- as.numeric(btc_data$Close)
btc_data$Volume.BTC <- as.numeric(btc_data$Volume.BTC)
btc_data$Volume.USDT <- as.numeric(btc_data$Volume.USDT)

# Display the first few rows after cleaning
head(btc_data)

# New column created (Daily Price Change)

btc_data <- btc_data %>%
  mutate(daily_change = Close - Open)

# Distribution of BTC Closing Prices

ggplot(btc_data, aes(x = Close)) +
  geom_histogram(fill = "blue", bins = 50, alpha = 0.7) +
  scale_x_continuous(labels = scales::label_comma()) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "Distribution of BTC Closing Prices", x = "BTC Price (USDT)", y = "Frequency") +
  theme_minimal()
```

```
# Trading Volume Distribution

ggplot(btc_data, aes(x = Volume.USDT)) +
  geom_histogram(fill = "orange", bins = 50, alpha = 0.7) +
  scale_x_continuous(labels = scales::label_comma()) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "Distribution of Trading Volume (USDT)", x = "Trading Volume (USDT)", y = "Frequency") +
  theme_minimal()

# BTC Closing Price Trend Over Time

ggplot(btc_data, aes(x = Date, y = Close)) +
  geom_line(color = "blue") +
  scale_x_date(labels = scales::label_date_short()) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "BTC Closing Price Trend Over Time", x = "Date", y = "Closing Price (USDT)") +
  theme_minimal()


# BTC Trading Volume Over Time

ggplot(btc_data, aes(x = Date, y = Volume.USDT)) +
  geom_col(fill = "orange") +
  scale_x_date(labels = scales::label_date_short()) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "BTC Trading Volume Over Time", x = "Date", y = "Trading Volume (USDT)") +
  theme_minimal()

# Daily BTC Price Volatility

ggplot(btc_data, aes(y = daily_change)) +
  geom_boxplot(fill = "lightblue", outlier.color = "red", outlier.shape = 1) +
  scale_y_continuous(labels = scales::label_comma()) +
  coord_cartesian(ylim = c(-2000, 2000)) +
  labs(title = "Daily BTC Price Volatility", y = "Daily Price Change (USDT)") +
  theme_minimal()


# BTC Closing Price vs. Trading Volume

ggplot(btc_data, aes(x = Volume.USDT, y = Close)) +
  geom_point(alpha = 0.5, color = "darkblue") +
  scale_x_continuous(labels = scales::label_comma()) +
  scale_y_continuous(labels = scales::label_comma()) +
  labs(title = "BTC Closing Price vs. Trading Volume", x = "Trading Volume (USDT)", y = "Closing Price
  theme_minimal()

# Wilcoxon Signed-Rank Test
wilcoxon_result <- wilcox.test(btc_data$daily_change, mu = 0, alternative = "two.sided")

# Mann-Whitney U Test (High vs. Low Volume Days)
high_volume <- btc_data$daily_change[btc_data$Volume.USDT > median(btc_data$Volume.USDT)]
low_volume <- btc_data$daily_change[btc_data$Volume.USDT <= median(btc_data$Volume.USDT)]
```

```r
mann_whitney_result <- wilcox.test(high_volume, low_volume)

# Fisher's Exact Test for categorical impact of volume on price movement
btc_data$price_movement_category <- ifelse(btc_data$daily_change > 0, "Increase", "Decrease")
volume_category <- ifelse(btc_data$Volume.USDT > median(btc_data$Volume.USDT), "High", "Low")
fisher_test_result <- fisher.test(table(volume_category, btc_data$price_movement_category))

# Load required libraries (lubridate is used for date manipulation)
library(lubridate)
library(dplyr)

# Create Year and Quarter columns
btc_data$Year <- year(btc_data$Date)
btc_data$Quarter <- quarter(btc_data$Date)

# Aggregate trading volume by (Year, Quarter)
quarterly_data <- btc_data %>%
  group_by(Year, Quarter) %>%
  summarize(total_volume = sum(Volume.USDT, na.rm = TRUE), .groups = 'drop')

# Filter only Q1 and Q4
q1_data <- quarterly_data %>%
  filter(Quarter == 1) %>%
  select(Year, total_volume)

q4_data <- quarterly_data %>%
  filter(Quarter == 4) %>%
  select(Year, total_volume)

# Join Q1 and Q4 volumes by Year (paired observation)
q1_vs_q4 <- inner_join(q1_data, q4_data, by = "Year", suffix = c("_Q1", "_Q4"))

# Perform a paired t-test to see if Q1 vs. Q4 differ within the same year
paired_test_result <- t.test(
  q1_vs_q4$total_volume_Q1,
  q1_vs_q4$total_volume_Q4,
  paired = TRUE
)
#Wilcoxon Result

wilcoxon_result

#Mann Whitney Result

mann_whitney_result

# Compute Fisher's Exact Test
fisher_test_result <- fisher.test(table(volume_category, btc_data$price_movement_category))

# Fisher's Exact Test Result
fisher_test_result

# Paired T-Test Result
```

```
paired_test_result
# Do not change anything here. There is purposely nothing in this code chunk.
# Remember, your appendix will not generate properly if you don't name each of your code chunks.
```