

Cryptocurrency Market Analysis: Advanced SQL Techniques

Introduction

This project conducts a comprehensive analysis of Bitcoin (BTC) and Ethereum (ETH) price movements using advanced SQL queries on high-frequency trading data. The minute-by-minute dataset spans from 2020 to 2025, capturing significant market events including the 2021 bull market and subsequent bear market. Through SQL analysis, this project explores volatility patterns, price correlations, trading volume anomalies, and market behavior across different timeframes.

Dataset Overview

The analysis uses two primary datasets:

- **BTCUSDT Minute Kline Data:** Contains minute-by-minute OHLCV (Open, High, Low, Close, Volume) data for Bitcoin against USDT
- **ETHUSDT Minute Kline Data:** Contains minute-by-minute OHLCV data for Ethereum against USDT

Both datasets include approximately 1.5 million rows each, providing a rich foundation for time-series analysis of cryptocurrency markets.

Data Source Information

The BTCUSDT and ETHUSDT minute kline data were obtained from Binance's historical data API. To access this data, one can use the following resources:

1. CCXT Library:

- GitHub: <https://github.com/ccxt/ccxt>
- Documentation: <https://docs.ccxt.com/>
- Example for fetching historical OHLCV data

2. Binance API Documentation:

- REST API: <https://binance-docs.github.io/apidocs/spot/en/#kline-candlestick-data>
- Futures API: <https://binance-docs.github.io/apidocs/futures/en/#kline-candlestick-data>

Key Analytical Queries

The project includes 10 specialized queries that explore different aspects of cryptocurrency markets:

1. **Volatility Analysis:** Comparing daily price ranges between BTC and ETH
2. **Correlation Studies:** Examining price relationships during specific market conditions
3. **Arbitrage Opportunities:** Identifying potential trading opportunities through ratio analysis
4. **Volume Analysis:** Detecting abnormal trading activity and volume spikes
5. **Market Crash Patterns:** Analyzing price behavior during significant downturns
6. **Intraday Patterns:** Comparing price movements at different times of the day
7. **Trading Hour Analysis:** Identifying the most active and volatile trading hours
8. **Weekend vs. Weekday Trading:** Contrasting market behavior between weekdays and weekends
9. **Monthly Performance:** Tracking month-over-month price changes and trading volumes
10. **Extreme Price Movements:** Detecting and analyzing minutes with unusual price volatility

AI-Generated SQL Mistakes

During the development of these queries, I identified several mistakes that AI systems commonly make when generating SQL code:

Mistake 1: Incorrect Date Function Usage

```
-- Original problematic code
JOIN
    "BTCUSDT.csv" yesterday_btc ON CAST(yesterday_btc.datetime AS DATE) =
```

Explanation: The AI incorrectly used the `datetime()` function which isn't supported in many SQL dialects, resulting in the error "Scalar Function with name datetime does not exist!" This demonstrates how AI systems often assume SQLite syntax will work universally across different database systems. The correct approach would be to use database-specific date arithmetic or a more widely supported function.

Mistake 2: Incorrect Window Function Usage

```
-- Problematic window function usage
SELECT
    SUBSTR(btc.datetime, 1, 7) as month,
    AVG(btc.close) as avg_btc_price,
```

```
(AVG(btc.close) - LAG(AVG(btc.close), 1) OVER (ORDER BY SUBSTR(btc.datetime, 1, 7)))  
LAG(AVG(btc.close), 1) OVER (ORDER BY SUBSTR(btc.datetime, 1, 7)))
```

Explanation: This query produced the error "column 'datetime' must appear in the GROUP BY clause or must be part of an aggregate function." The AI incorrectly assumed that window functions could be applied directly to aggregate functions in this way. Most SQL implementations don't allow window functions to be nested within aggregate functions or vice versa. The correct approach is to use a CTE or subquery to first calculate the aggregates and then apply window functions.

Mistake 3: Inefficient Correlated Subquery

```
-- Correlated subquery with performance issues  
(eth.close / btc.close) - (  
    SELECT AVG(e.close / b.close)  
    FROM "ETHUSDT.csv" e  
    JOIN "BTCUSDT.csv" b ON e.datetime = b.datetime  
    WHERE CAST(b.datetime AS DATE) = CAST(btc.datetime AS DATE)  
) as ratio_deviation
```

Explanation: While this query may execute correctly, it represents a significant performance mistake. The correlated subquery runs for every row in the outer query, making it extremely inefficient for large datasets like minute-by-minute cryptocurrency data. The AI failed to recognize that this could be rewritten more efficiently using window functions or a pre-aggregation step. This type of mistake could cause query timeouts or excessive resource consumption in production environments.

Background Research: Cryptocurrency Market Analysis

Market Volatility and Price Movements

The cryptocurrency market is known for its high volatility compared to traditional financial markets. Our volatility analysis query (Query 1) comparing BTC and ETH daily price ranges aligns with research by Dyhrberg (2016)^[1], who found that Bitcoin exhibits volatility patterns that place it somewhere between gold and the US dollar. More recent research by Liu & Tsyvinski (2021)^[2] shows that cryptocurrency returns and volatility cannot be explained by exposure to traditional asset classes, supporting our approach of analyzing crypto assets independently.

Technical Analysis and Trading Patterns

Our queries examining intraday patterns (Query 6) and extreme price movements (Query 10) are supported by research from Detzel et al. (2021)^[^3], who found that technical analysis can generate significant profits in cryptocurrency markets. Their study showed that simple moving average strategies and other technical indicators have predictive power for Bitcoin returns. Similarly, Corbet et al. (2019)^[^4] demonstrated that technical trading rules can be effective in cryptocurrency markets, with particular success during periods of high volatility.

Weekend vs. Weekday Trading

Query 8, which compares weekend and weekday trading patterns, is particularly relevant given the 24/7 nature of cryptocurrency markets. Hudson & Urquhart (2021)^[^5] found that Bitcoin returns on weekends differ from those on weekdays, with weekend returns showing different statistical properties. This supports our analysis of trading volume and price range differences between weekdays and weekends.

Arbitrage Opportunities

Our analysis of potential arbitrage opportunities (Query 3) is grounded in research by Makarov & Schoar (2020)^[^6], who documented significant price deviations across cryptocurrency exchanges. They found that these deviations could persist for several hours, creating potential arbitrage opportunities. However, Hautsch et al. (2019)^[^7] noted that settlement latency in blockchain-based markets creates limits to arbitrage, which explains why some price discrepancies may persist.

Monthly Performance and Seasonality

Query 9's examination of monthly performance aligns with findings from Grobys et al. (2020)^[^8], who identified seasonal patterns in cryptocurrency returns. Their research suggests that certain months historically show better performance than others, which could be captured by our monthly performance comparison query.

Correlation Between Cryptocurrencies

Query 2's analysis of correlation between BTC and ETH during specific market conditions is supported by multiple studies showing that cryptocurrency correlations tend to increase during market stress periods. This "correlation clustering" phenomenon has implications for portfolio diversification within the cryptocurrency asset class.

References

- [^1]: Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar–A GARCH volatility analysis. *Finance Research Letters*, 16, 85–92.
- [^2]: Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6), 2689–2727.
- [^3]: Detzel, A. L., Liu, H., Strauss, J., Zhou, G., & Zhu, Y. (2021). Bitcoin: Learning, predictability, and profitability via technical analysis. *Journal of Empirical Finance*, 63, 52–70.
- [^4]: Corbet, S., Eraslan, V., Lucey, B., & Sensoy, A. (2019). The effectiveness of technical trading rules in cryptocurrency markets. *Finance Research Letters*, 31, 32–37.
- [^5]: Hudson, R., & Urquhart, A. (2021). Technical trading and cryptocurrencies. *Annals of Operations Research*, 297(1), 191–220.
- [^6]: Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics*, 135(2), 293–319.
- [^7]: Hautsch, N., Scheuch, C., & Voigt, S. (2019). Limits to arbitrage in markets with stochastic settlement latency. *Journal of Economic Dynamics and Control*, 99, 1–28.
- [^8]: Grobys, K., Ahmed, S., & Sapkota, N. (2020). Technical trading rules in the cryptocurrency market. *Finance Research Letters*, 32, 101396.