Data Analytics Report and Executive Summary

Project Title

Correlation and Inter-market Analysis of Bitcoin, Ethereum, and Cardano on Binance

---

Executive Summary

This report investigates the interrelationships among three major cryptocurrencies—Bitcoin (BTC/USDT), Ethereum (ETH/USDT), and Cardano (ADA/USDT)—using 1-minute OHLCV data extracted from Binance over a two-year period. The analysis explores whether statistically significant correlations exist that can be leveraged to optimize trading strategies. Data was collected using a custom Binance API client, and the extraction process was divided into two notebooks: one for data extraction/cleaning and another for visualization/analysis. Key steps included reindexing the data to a full 1-minute interval, identifying and handling missing rows, computing percentage changes and movement classifications, and performing statistical analyses such as correlation and rolling volatility. Our findings show strong positive correlations—BTC and ETH are highly correlated (≈0.84), while ADA has a moderately strong relationship with BTC (≈0.80) and ETH (≈0.68). These results suggest that BTC may serve as a leading indicator for the other assets, which has significant implications for risk management and trading strategy optimization.

See Appendix A for screenshots of the data extraction and cleaning process, and Appendix B for analysis and visualization outputs.

---

1. Research Question Summary

Research Question:
What is the nature of the relationship between Bitcoin, Ethereum, and Cardano on Binance, and can these correlations be leveraged to optimize trading strategies?

Justification & Context:

- The volatile nature of the cryptocurrency market demands a deeper understanding of how major assets interact.
- By analyzing 1-minute data for BTC/USDT, ETH/USDT, and ADA/USDT, we aim to determine if price movements in one asset can serve as predictive signals for the others.
- Such insights are valuable for portfolio diversification, risk management, and the development of algorithmic trading strategies.

Hypotheses:

- **Null Hypothesis (H_0):** There is no statistically significant correlation between the price movements of Bitcoin, Ethereum, and Cardano on Binance.
- **Alternate Hypothesis (H_1):** There is a statistically significant correlation among these assets, with evidence of predictive relationships.

Refer to Appendix A, Figure 1 for the initial raw data extraction output.

---

2. Data Collection

Data Description:

- **Source:** Binance API (accessed via a custom Python client using the ccxt library).
- **Assets:** BTC/USDT, ETH/USDT, ADA/USDT.
- **Timeframe:** Two years of historical data, ending at yesterday's midnight.
- **Frequency:** 1-minute intervals capturing OHLCV (Open, High, Low, Close, Volume) data.

Methodology:

- Data was pulled in a round-robin manner across the three symbols, with batches saved incrementally into separate CSV files.
- This approach ensures that even if API rate limits or occasional missing intervals occur, partial data is saved for later review.

Advantages & Disadvantages:

- **Advantages:**
  - High-frequency, granular data.
  - Continuous saving allows for early review and troubleshooting.
- **Disadvantages:**
  - API rate limitations and occasional missing intervals.
  - Gaps in data that require cleaning and careful handling.

See Appendix A, Figure 2 for screenshots of the data extraction and file-saving process.

---

3. Data Extraction and Preparation Process

Extraction Process:

- Custom Python scripts using the Binance client extracted the data.
- Each CSV (e.g., `binance_klines_BTCUSDT_2years.csv`) was then loaded into a Jupyter Notebook for reindexing and cleaning.

Data Cleaning:

- The `timestamp` column was converted to a datetime object and set as the DataFrame index.

- A full date range (1-minute intervals) was generated based on the minimum and maximum timestamps.
- The DataFrame was reindexed to this complete range, revealing missing rows.
- For example, ADA/USDT had 1,048,160 expected rows with 408,660 missing.
- The decision was made to drop rows that were completely empty (i.e., all columns are NaN) to ensure that subsequent analyses are based solely on recorded trading data.

See Appendix A, Figures 3 and 4 for screenshots of the reindexing process, missing row check, and confirmation of the cleaned data saved as new CSV files.

---

4. Data Analysis Process

Additional Data Processing:

- New columns were computed:
    - **`pct_change`:** The percentage change in closing price from one minute to the next.
    - **`movement`:** A categorical variable indicating whether the change was "up," "down," or "flat."

Statistical Analysis & Visualizations:

- **Time Series Analysis:**
    - Line plots were generated for raw closing prices and percentage changes across all symbols.
- **Correlation Analysis:**
    - Data for each symbol was merged on the timestamp to compute a correlation matrix for the percentage changes.
    - The resulting matrix showed strong positive correlations (BTC vs. ETH ≈ 0.84; BTC vs. ADA ≈ 0.80; ETH vs. ADA ≈ 0.68).
- **Rolling Analysis:**
    - Rolling standard deviation (volatility) of percentage changes was calculated using a 60-minute window.
    - Rolling correlations were also computed to evaluate dynamic relationships over time.
- **Technical Analysis Visualizations:**
    - Multiple charts were produced using mplfinance, including:
        - Candlestick charts (with volume, moving averages).
        - Overlays with Bollinger Bands.
        - Renko charts for alternative price movement visualization.
        - Charts with RSI subplots.
        - A line chart for ADA with an overlay of percentage change.

See Appendix B, Figures 1–6 for the key outputs: time series plots, correlation heatmap, rolling volatility plots, and various mplfinance charts with technical indicators.

Justification:

- The methods normalize differences in price scale and capture dynamic relationships.

- Rolling metrics and technical overlays provide actionable insights for trading decisions.
- Limitations include missing data periods and assumptions in the statistical techniques.

---

5. Data Summary and Implications

Key Findings:

- Strong positive correlations between BTC and ETH indicate that BTC may serve as a reliable leading indicator.
- ADA shows a slightly lower, yet significant, correlation with BTC and ETH.
- Normalized percentage changes enable a cross-comparison despite differences in absolute price levels.
- Volatility analysis highlights periods of increased market instability.

Implications for Trading:

- Traders might monitor BTC's price movements to adjust positions in ETH and ADA.
- The technical indicators (e.g., Bollinger Bands, RSI) provide additional signals that could improve trading strategies.
- Recommendations include dynamically adjusting risk exposure based on identified volatility trends.

Limitations & Future Directions:

- The analysis is based on historical data with gaps that were dropped.
- Future research could involve:
    1. Expanding the asset universe to include additional cryptocurrencies.
    2. Applying machine learning techniques (e.g., LSTM networks) to improve predictive accuracy.

Refer to Appendix B, Figure 7 for a summary visualization of rolling volatility and Appendix B, Figure 8 for an integrated technical chart with RSI.

---

6. Sources and References

- **Data Source:** Binance API (accessed via ccxt library).
- **Tools & Libraries:** Python, pandas, matplotlib, seaborn, mplfinance.
- **References:**
    - Binance API Documentation.
    - mplfinance documentation and tutorials.
    - Relevant academic literature on financial time series analysis.

Appendices

Appendix A: Data Extraction & Cleaning

- **Figure A1:** *Raw Data Extraction Output*

```
Processing ADA/USDT from file binance_klines_ADAUSDT_2years.csv
                  timestamp    open    high     low   close  volume
0       2023-03-05 22:33:00  0.3369  0.3369  0.3369  0.3369     0.0
1       2023-03-05 22:34:00  0.3369  0.3369  0.3369  0.3369     0.0
2       2023-03-05 22:35:00  0.3369  0.3369  0.3369  0.3369     0.0
3       2023-03-05 22:36:00  0.3369  0.3369  0.3369  0.3369     0.0
4       2023-03-05 22:37:00  0.3369  0.3369  0.3369  0.3369     0.0
...                     ...     ...     ...     ...     ...     ...
639495  2025-03-02 19:48:00  1.0397  1.0459  1.0397  1.0399   162.9
639496  2025-03-02 19:49:00  1.0466  1.0467  1.0412  1.0455  2372.2
639497  2025-03-02 19:50:00  1.0443  1.0443  1.0392  1.0415  3951.0
639498  2025-03-02 19:51:00  1.0400  1.0413  1.0400  1.0413     3.3
639499  2025-03-02 19:52:00  1.0426  1.0426  1.0426  1.0426     9.5

[639500 rows x 6 columns]
```

```
Processing ETH/USDT from file binance_klines_ETHUSDT_2years.csv
                  timestamp     open     high      low    close   volume
0       2023-03-05 22:33:00  1563.87  1564.15  1561.51  1562.12  48.4015
1       2023-03-05 22:34:00  1562.14  1562.82  1562.14  1562.82   5.9740
2       2023-03-05 22:35:00  1563.32  1564.55  1563.32  1564.38   1.8214
3       2023-03-05 22:36:00  1565.50  1565.61  1564.06  1565.61  36.4439
4       2023-03-05 22:37:00  1565.61  1565.61  1565.61  1565.61   0.0000
...                     ...      ...      ...      ...      ...      ...
639495  2025-03-02 19:48:00  2496.55  2496.55  2496.55  2496.55   0.0000
639496  2025-03-02 19:49:00  2499.33  2499.33  2499.29  2499.29   3.3149
639497  2025-03-02 19:50:00  2499.08  2499.08  2499.08  2499.08   0.0220
639498  2025-03-02 19:51:00  2499.08  2499.08  2499.08  2499.08   0.0000
639499  2025-03-02 19:52:00  2499.08  2499.08  2499.08  2499.08   0.0000

[639500 rows x 6 columns]
```

```
Processing BTC/USDT from file binance_klines_BTCUSDT_2years.csv
                  timestamp      open      high       low     close   volume
0       2023-03-05 22:33:00  22404.50  22411.02  22397.89  22397.89  1.36425
1       2023-03-05 22:34:00  22392.22  22409.44  22392.22  22407.91  0.18043
2       2023-03-05 22:35:00  22410.55  22428.09  22410.55  22428.09  0.29686
3       2023-03-05 22:36:00  22434.58  22434.58  22417.31  22429.73  0.83456
4       2023-03-05 22:37:00  22429.64  22431.52  22423.88  22427.08  1.19932
...                     ...       ...       ...       ...       ...      ...
639495  2025-03-02 19:48:00  93421.24  93491.57  93421.24  93491.57  0.00431
639496  2025-03-02 19:49:00  93491.57  93491.57  93449.26  93449.26  0.00153
639497  2025-03-02 19:50:00  93449.26  93449.26  93449.26  93449.26  0.00000
639498  2025-03-02 19:51:00  93449.26  93449.26  93449.26  93449.26  0.00000
639499  2025-03-02 19:52:00  93449.26  93449.26  93449.26  93449.26  0.00000

[639500 rows x 6 columns]
```

- **Figure A2:** *Reindexing & Missing Row Check*

  BTC:

```
[639500 rows x 6 columns]
Total expected rows: 1048160
Missing rows: 408660
/tmp/ipykernel_118530/1184356804.py:15: FutureWarning: 'T' is deprecated and will be removed in a future version, please use 'min' instead.
  full_range = pd.date_range(start=df.index.min(), end=df.index.max(), freq='T')
Sample missing timestamps: [Timestamp('2023-03-06 06:53:00'), Timestamp('2023-03-06 06:54:00'), Timestamp('2023-03-06 06:55:00'), Timestamp
('2023-03-06 06:56:00'), Timestamp('2023-03-06 06:57:00')]
```

  ETH:

```
[639500 rows x 6 columns]
Total expected rows: 1048160
Missing rows: 408660
/tmp/ipykernel_118530/1184356804.py:15: FutureWarning: 'T' is deprecated and will be removed in a future version, please use 'min' instead.
  full_range = pd.date_range(start=df.index.min(), end=df.index.max(), freq='T')
Sample missing timestamps: [Timestamp('2023-03-06 06:53:00'), Timestamp('2023-03-06 06:54:00'), Timestamp('2023-03-06 06:55:00'), Timestamp
('2023-03-06 06:56:00'), Timestamp('2023-03-06 06:57:00')]
```

  ADA:

```
[639500 rows x 6 columns]
Total expected rows: 1048160
Missing rows: 408660
/tmp/ipykernel_118530/1184356804.py:15: FutureWarning: 'T' is deprecated and will be removed in a future version, please use 'min' instead.
  full_range = pd.date_range(start=df.index.min(), end=df.index.max(), freq='T')
Sample missing timestamps: [Timestamp('2023-03-06 06:53:00'), Timestamp('2023-03-06 06:54:00'), Timestamp('2023-03-06 06:55:00'), Timestamp
('2023-03-06 06:56:00'), Timestamp('2023-03-06 06:57:00')]
```

-

- **Figure A3:** *Cleaned Data Confirmation*

```
Data cleaning and review complete.
```

```python
# For example, display the first few rows of the cleaned BTC/USDT DataFrame:
print("BTC/USDT Cleaned Data Sample:")
display(cleaned_dfs['BTC/USDT'].head())
```

```
BTC/USDT Cleaned Data Sample:
```

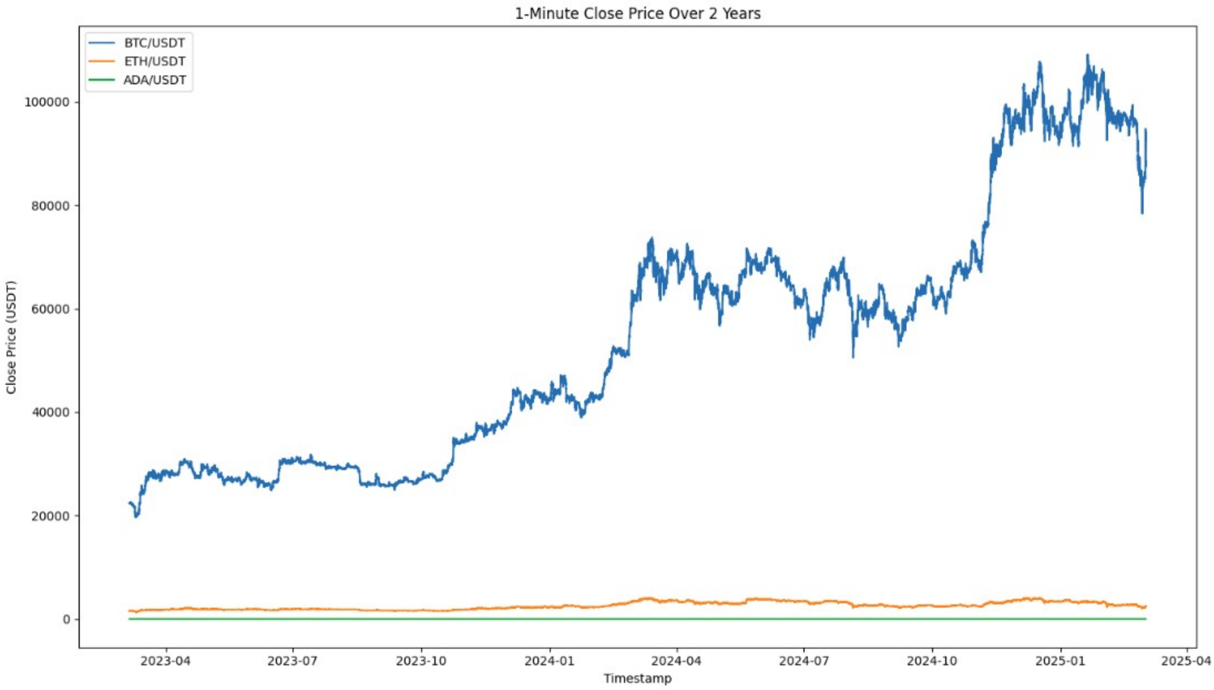
	open	high	low	close	volume
2023-03-05 22:33:00	22404.50	22411.02	22397.89	22397.89	1.36425
2023-03-05 22:34:00	22392.22	22409.44	22392.22	22407.91	0.18043
2023-03-05 22:35:00	22410.55	22428.09	22410.55	22428.09	0.29686
2023-03-05 22:36:00	22434.58	22434.58	22417.31	22429.73	0.83456
2023-03-05 22:37:00	22429.64	22431.52	22423.88	22427.08	1.19932

```python
import os

for symbol, df in cleaned_dfs.items():
    # Drop rows where all values are NaN
    df_dropped = df.dropna(how='all')
    print(f"{symbol}: After dropping empty rows, remaining rows = {df_dropped.shape[0]}")

    # Define a new filename for the saved cleaned data
    cleaned_filename = f"binance_klines_{symbol.replace('/', '')}_2years_cleaned_dropped.csv"

    # Save the DataFrame to CSV
    df_dropped.to_csv(cleaned_filename, index_label='timestamp')
    print(f"Cleaned data saved to {cleaned_filename}\n")
```

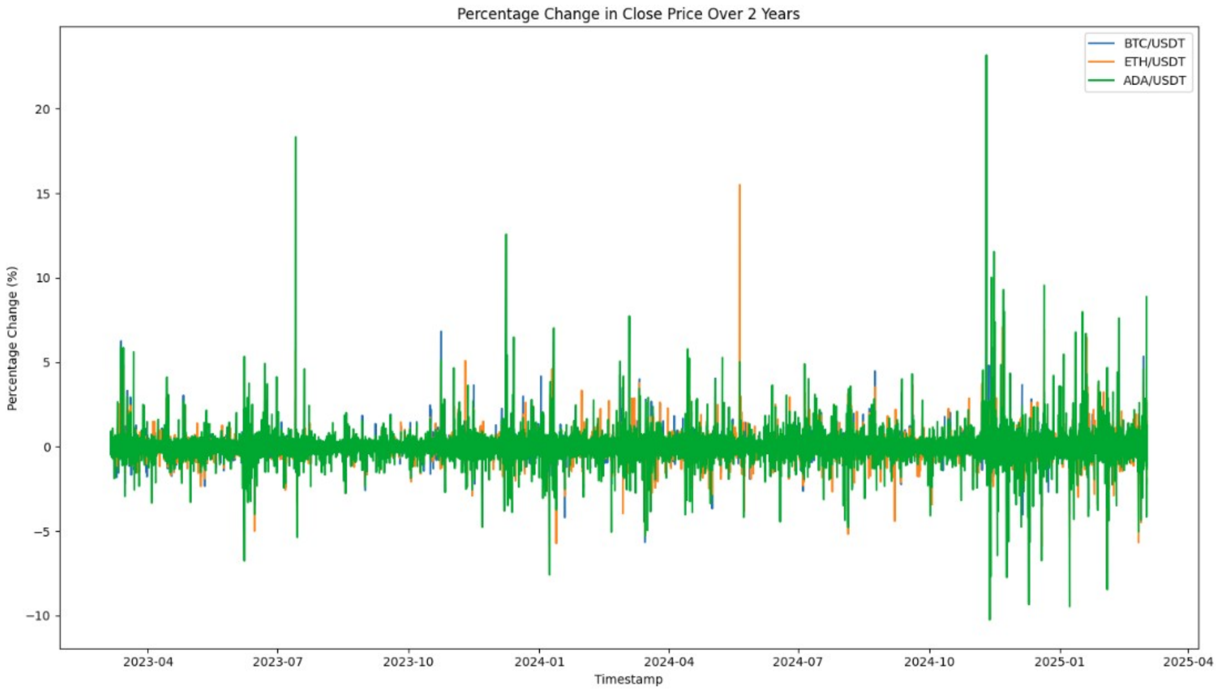
```
BTC/USDT: After dropping empty rows, remaining rows = 639500
Cleaned data saved to binance_klines_BTCUSDT_2years_cleaned_dropped.csv

ETH/USDT: After dropping empty rows, remaining rows = 639500
Cleaned data saved to binance_klines_ETHUSDT_2years_cleaned_dropped.csv

ADA/USDT: After dropping empty rows, remaining rows = 639500
Cleaned data saved to binance_klines_ADAUSDT_2years_cleaned_dropped.csv
```

Appendix B: Data Analysis & Visualization

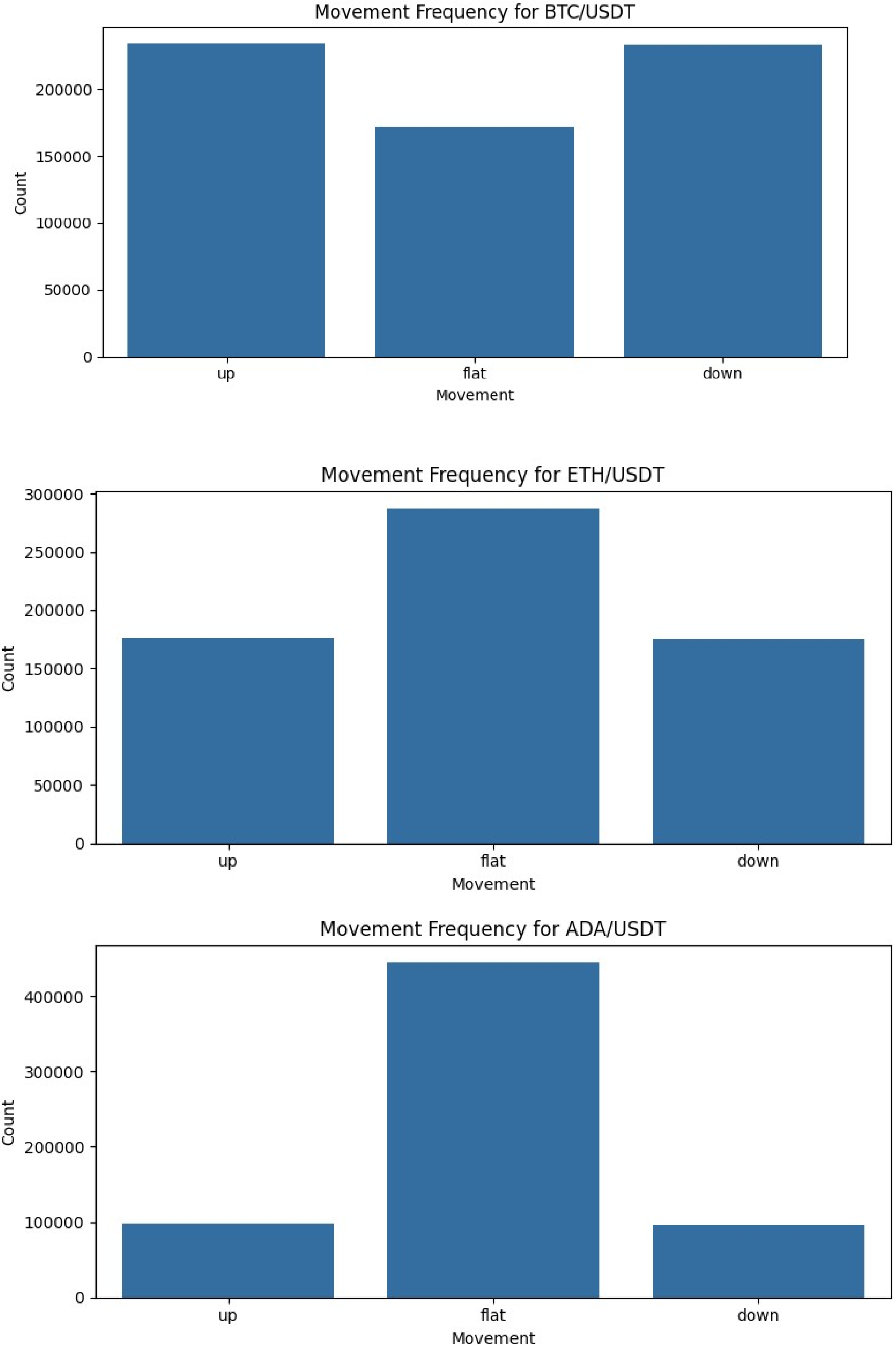
- **Figure B1:** *Time Series Plot of Closing Prices*



- **Figure B2:** *Time Series Plot of Percentage Changes*



- **Figure B3:** *Movement Frequency Distribution*

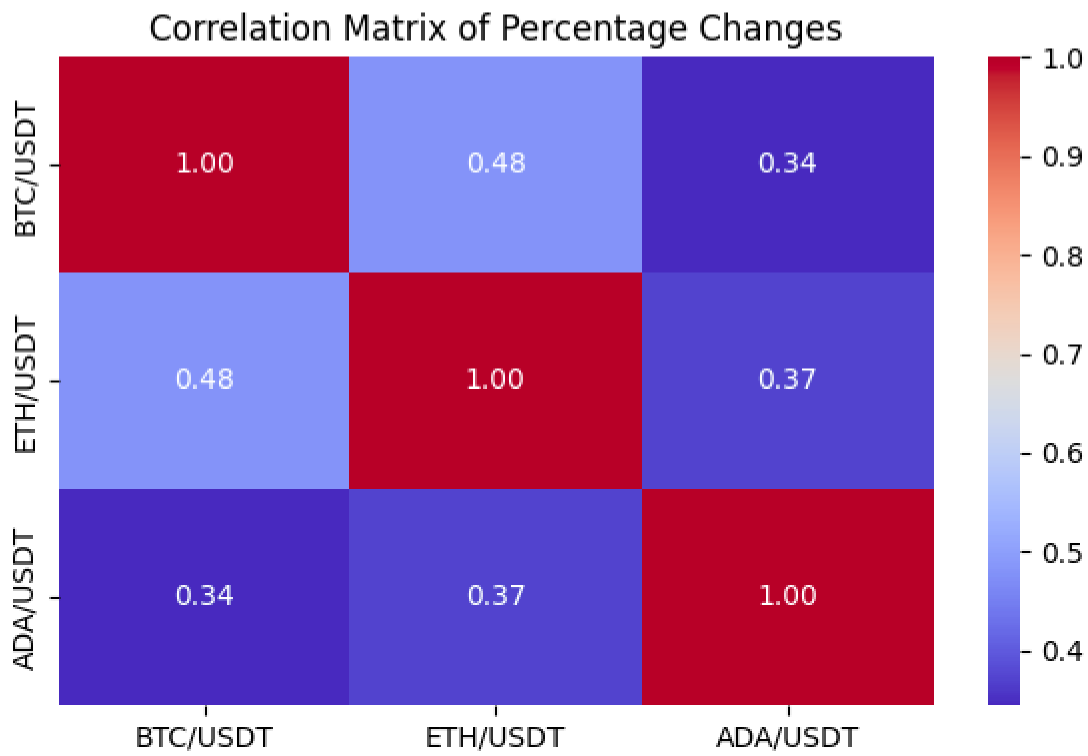


- **Figure B4:** *Correlation Matrix & Heatmap*

```
Merged percentage change data shape: (639500, 4)
```

	timestamp	BTC/USDT	ETH/USDT	ADA/USDT
0	2023-03-05 22:33:00	NaN	NaN	NaN
1	2023-03-05 22:34:00	0.044736	0.044811	0.0
2	2023-03-05 22:35:00	0.090057	0.099820	0.0
3	2023-03-05 22:36:00	0.007312	0.078625	0.0
4	2023-03-05 22:37:00	-0.011815	0.000000	0.0

```python
# Set 'timestamp' as the index for correlation computation
merged_pct_df.set_index('timestamp', inplace=True)
corr_pct = merged_pct_df.corr()
plt.figure(figsize=(6, 4))
sns.heatmap(corr_pct, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of Percentage Changes')
plt.tight_layout()
plt.show()
```



• **Figure B5:** *Rolling Volatility Plot*

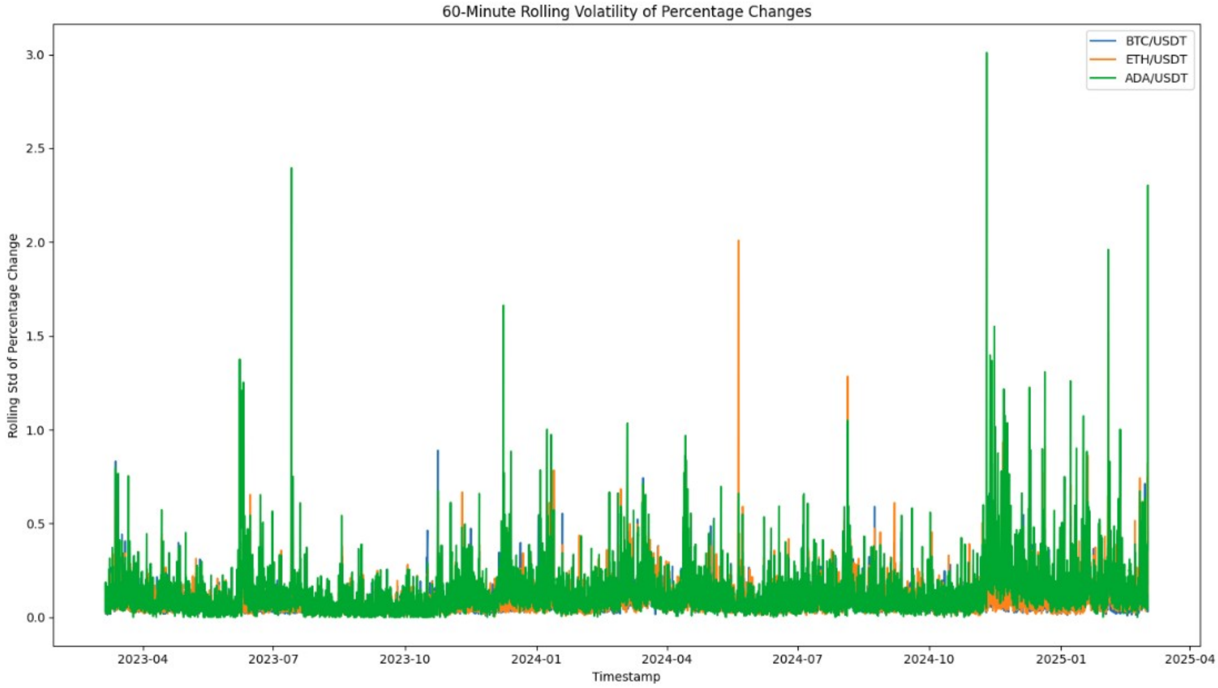
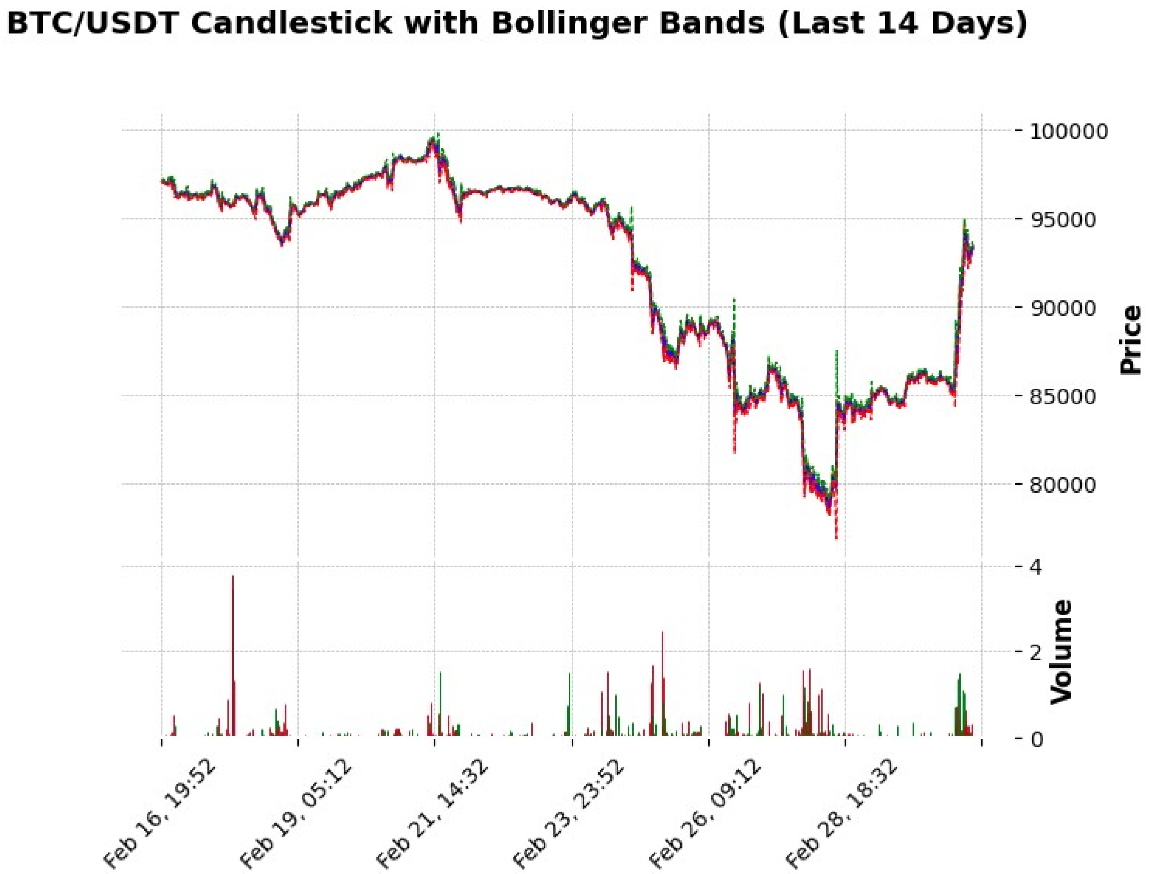
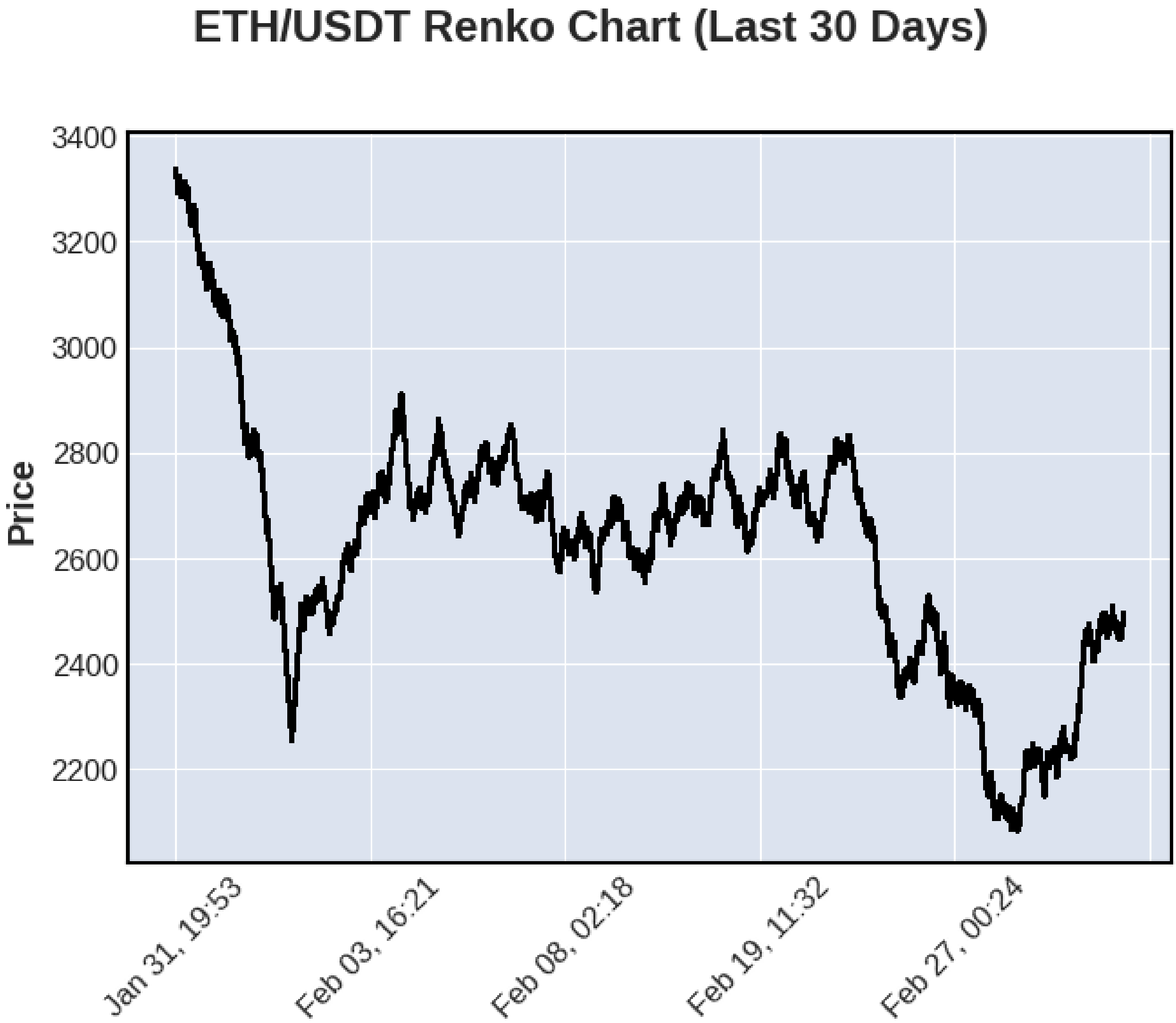


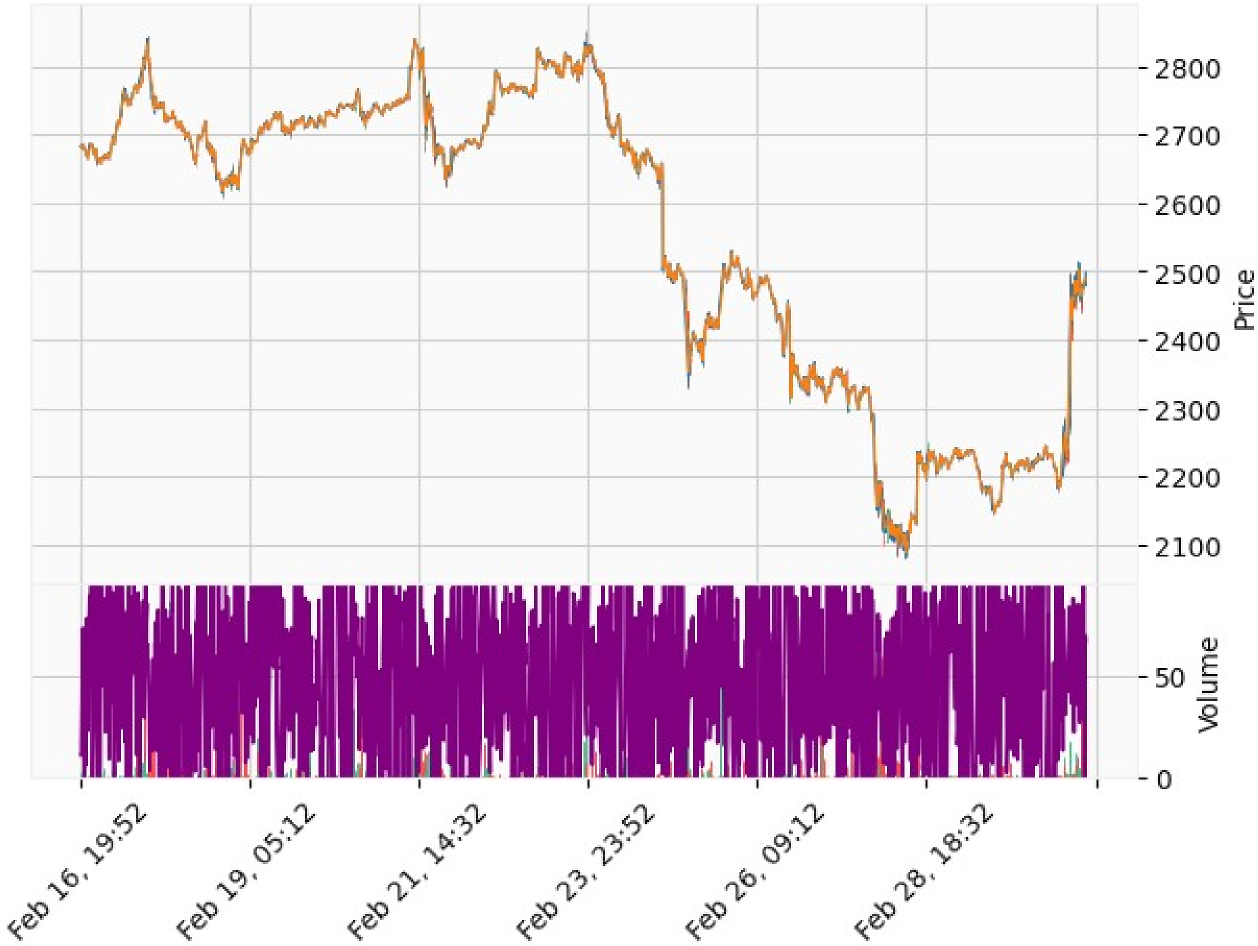
Figure B6: *Candlestick Chart with Bollinger Bands*

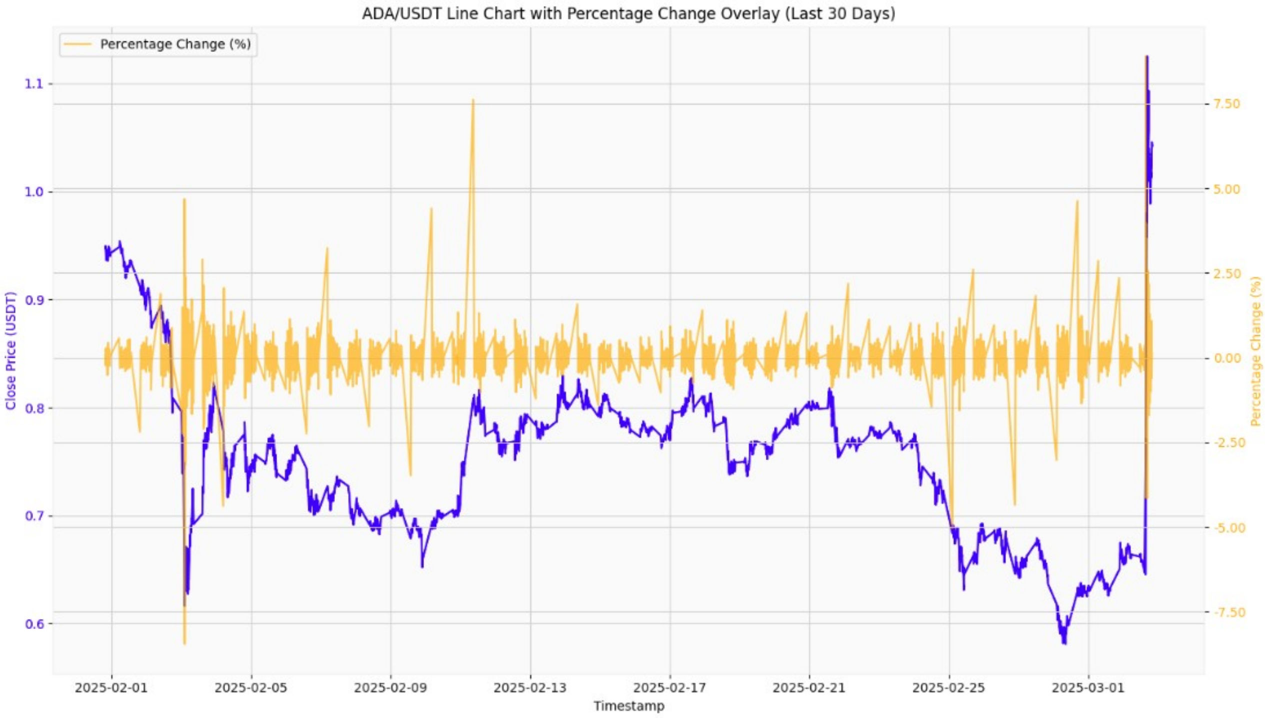


- **Figure B7:** *Alternative Technical Charts (Renko, RSI, ADA Overlay)*



ETH/USDT Candlestick Chart with RSI (Last 14 Days)





ADA/USDT Line Chart with Percentage Change Overlay (Last 30 Days)
Percentage Change (%)
Close Price (USDT)
Percentage Change (%)
Timestamp
1.1
1.0
0.9
0.8
0.7
0.6
7.50
5.00
2.50
0.00
-2.50
-5.00
-7.50
2025-02-01
2025-02-05
2025-02-09
2025-02-13
2025-02-17
2025-02-21
2025-02-25
2025-03-01