Investment for Cryptocurrency ETFs and Price Prediction:

In the context of the approval of the Bitcoin ETF Fund in January 2024, both funds and Bitcoin have seen a 130% increase in value. Recently, in May 2024, the Ethereum ETF has also been approved. The cryptocurrency landscape has never been more dynamic, with traditional funds and capital flows rushing to fill the market. Understanding this, my personal project aims to help individuals assess the health of these assets, measure and visualize the results, and thus assist you in making the most suitable choices for yourselves.

Given the landmark approval of the Bitcoin ETF Fund in January 2024, we've witnessed an astonishing 100% surge in the value of both Bitcoin and related investment funds. This milestone was followed by another significant event in May 2024, with the approval of the Ethereum ETF. The cryptocurrency market has never been more vibrant and bustling, as traditional funds and capital are eagerly pouring into the space, seeking opportunities and growth.

Recognizing the unprecedented activity and the growing interest in the crypto market, I have embarked on a personal project designed to aid individuals in navigating this complex landscape. The project's core objective is to provide a comprehensive assessment of the health of major cryptocurrency assets, specifically focusing on Bitcoin and Ethereum, which are currently the most influential and widely discussed.

Through meticulous data collection and analysis, this project aims to measure key indicators that reflect the performance and stability of these assets. By utilizing advanced data visualization techniques, we will present the findings in a clear and accessible manner, allowing users to gain insightful and actionable information. This approach will empower individuals to make informed decisions based on the latest market trends and data-driven insights.

In essence, my project will serve as a valuable tool for anyone looking to understand the current state of the cryptocurrency market. Whether you are a seasoned investor or a newcomer to the crypto world, having access to reliable and up-to-date information is crucial. By assessing the health of Bitcoin and Ethereum, measuring their performance, and visualizing the results, we aim to provide a robust foundation for making the most suitable investment choices tailored to your personal needs and goals.

Stay tuned for detailed analyses, visual reports, and ongoing updates as we delve into the exciting and ever-evolving world of cryptocurrencies. Together, we can navigate this vibrant market and uncover the opportunities it holds.

1, Data Sources and Collection:

For our project, we source our data from Investing.com. Investing.com is a comprehensive financial platform that provides a wide range of market data, including real-time quotes, charts, financial news, technical analysis, and tools for various financial instruments across global markets. This platform is well-regarded for its accuracy and breadth of data, making it an ideal source for our analysis.

Data Details:

We collect price data for the following assets:

- **ARKB, BRRR, BITB, FBTC, and HODL**:
  - Date Range: 12/01/2024 to 23/05/2024
  - These are newly launched funds that have shown significant growth and interest in the market.
- **GBTC**:
  - Date Range: 28/05/2015 to 23/05/2024
  - Grayscale Bitcoin Trust (GBTC) is one of the oldest and most established Bitcoin investment funds. Its long history provides valuable insights into Bitcoin's performance over time.
- **IBIT**:
  - Date Range: 16/01/2024 to 24/05/2024
  - This fund is another recent addition to the market, reflecting the growing institutional interest in Bitcoin.
- **ETH**:
  - Date Range: 31/12/2017 to 24/05/2024
  - Ethereum data captures the performance of the second-largest cryptocurrency by market capitalization, providing crucial insights post-approval of the Ethereum ETF.

- **BTC**:
  - Date Range: 24/05/2015 to 24/05/2024
  - Bitcoin's price data over an extensive period helps analyze long-term trends and market behavior.

Data Type:

The data is available in CSV and XLSX formats, which are suitable for in-depth analysis and visualization.

2, Data Cleaning and Basic Analysis:

Data Cleaning:

To ensure the integrity and usability of our dataset, we undertake a comprehensive data cleaning process. This involves several steps, leveraging SQL for its efficiency and precision in handling large datasets. Here's a detailed outline of our data cleaning approach:

1. **Setting Data Types**:
   - **Volume**: Converted to BIGINT to handle large numerical values representing the trading volume.
   - **Date**: Formatted as DATE to facilitate chronological analysis and accurate time series operations.
   - **Other Numerical Columns**: Converted to FLOAT or DOUBLE for precise calculation of prices, returns, and other financial metrics.
2. **Handling Missing Values**:
   - Identify and address any missing or null values to ensure the dataset is complete.
   - Depending on the context, we may fill missing values using methods such as forward fill, backward fill, or interpolation.
3. **Data Consistency**:
   - Ensure that all columns follow a consistent format, particularly dates and numerical values.
   - Remove any duplicate rows that may skew analysis results.
4. **Validation**:
   - Validate the data to confirm that there are no anomalies or outliers that could impact the analysis.
   - Perform checks to ensure the data matches expected ranges and patterns.

Basic Analysis:

Once the data is cleaned and validated, we move on to performing basic analyses to gain initial insights and prepare for more detailed studies.

1. **Performance Analysis by Time Periods**:
   - **Monthly Analysis**: Calculate and compare the performance of each asset group on a month-to-month basis. This helps identify short-term trends and anomalies.
   - **Quarterly Analysis**: Aggregates data to assess the performance over each quarter. This smooths out short-term volatility and provides a clearer picture of medium-term trends.

- **Yearly Analysis**: Evaluates long-term performance by comparing annual returns, which helps in understanding broader market cycles and trends.
2. **Calculation of Key Indicators**:
   - **Relative Strength Index (RSI)**: Measures the speed and change of price movements to identify overbought or oversold conditions.
   - **Moving Average Convergence Divergence (MACD)**: Utilizes the difference between two moving averages to identify potential buy or sell signals.
   - **Signal Line**: Derived from the MACD, it helps confirm buy or sell signals and identify market trends.
3. **Grouping and Conditional Analysis**:
   - Use SQL queries with WHERE clauses to group data by specific indicators (e.g., RSI, MACD) to identify patterns and validate data integrity.
   - This helps in spotting periods of potential trend reversals and validating the robustness of our dataset.

Pre-Analysis Validity Checks:

Before delving into deeper analysis using Python, we conduct several validity checks to ensure the reliability of our data and preliminary indicators:

1. **Data Range Checks**:
   - Verify that the date ranges for each asset match the expected periods.
   - Confirm that the numerical values are within realistic bounds and there are no outliers due to data entry errors.
2. **Indicator Validation**:
   - Check the calculated RSI, MACD, and Signal values for logical consistency.
   - Ensure that the indicators align with known market conditions during the corresponding periods.

3, Data analysis:

a. Modeling Price Charts and Indicators

To effectively analyze and visualize the data, we use Python to model price charts and key indicators. The code for this analysis is contained in a Jupyter Notebook titled Code Test Variable.ipynb. This notebook includes:

- **Price**: Historical prices of each asset.
- **RSI (Relative Strength Index)**: A momentum oscillator that measures the speed and change of price movements.

Biểu đồ tương tác với RSI cho BTC
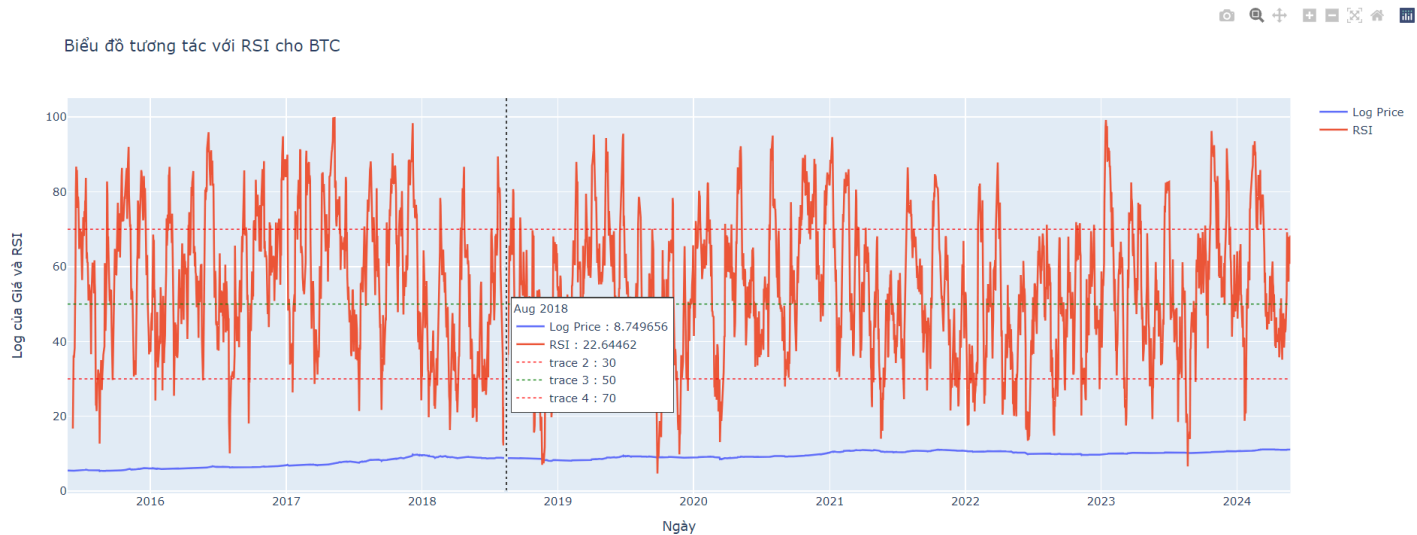
- **MACD (Moving Average Convergence Divergence)**: A trend-following momentum indicator that shows the relationship between two moving averages of an asset's price.

Biểu đồ MACD cho IBIT
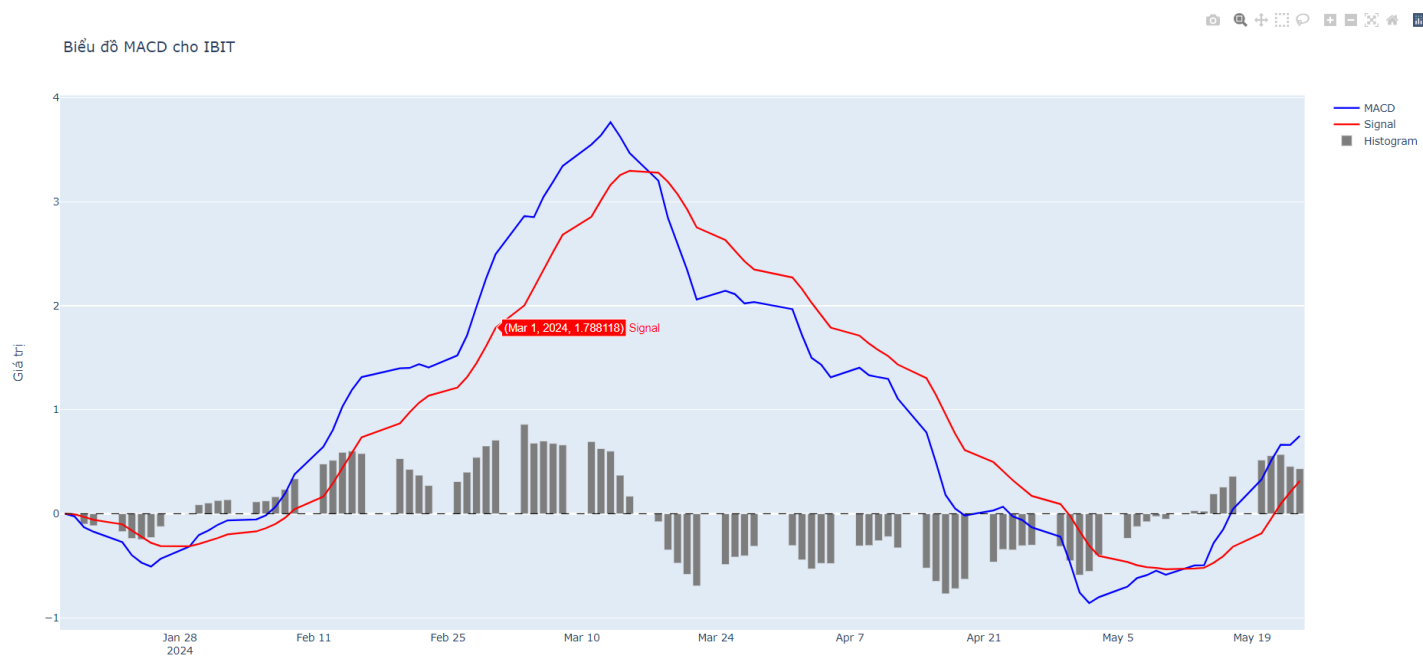
- **Stock RSI**: An oscillator that measures the level of the RSI relative to its range over a set period.

Stochastic Indicator for ETH
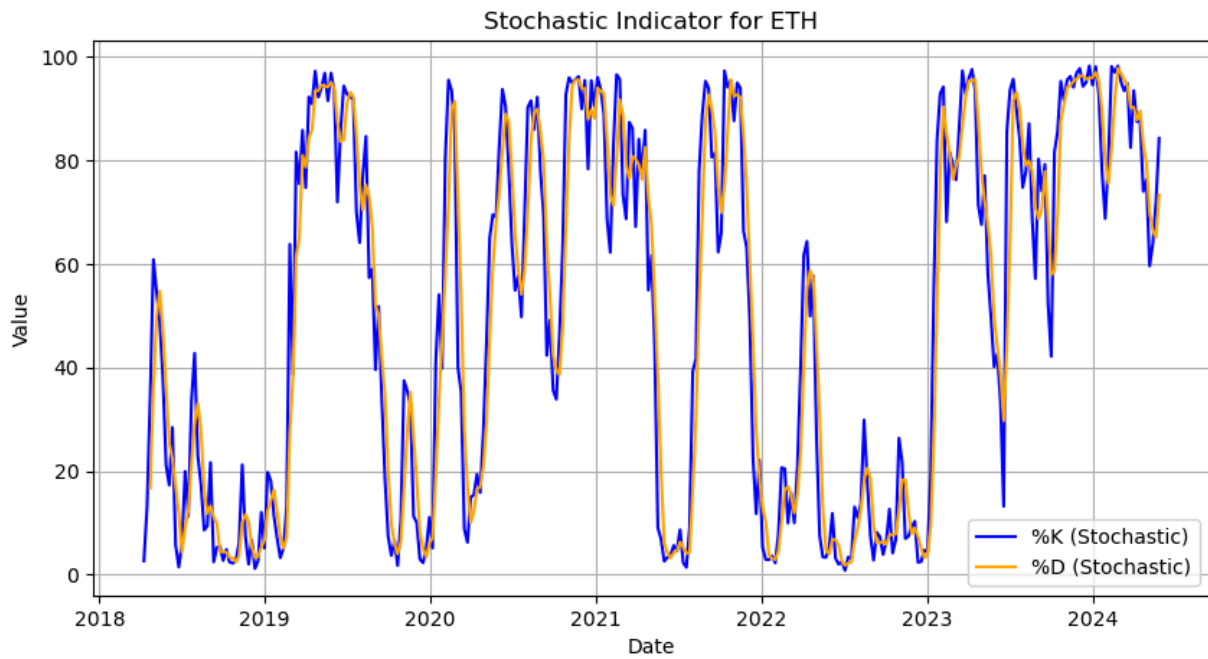
Note: To visualize the price or volume of all assets in a single chart, we use a logarithmic scale. This is essential due to the significant discrepancies in prices and volumes among different assets.

Biểu đồ tổng hợp log của giá cho tất cả các sheet

Most of the visualizations are stored in folders named "PRICE," "MACD," "MACD and Volume," and "StockRSI."

b. Exploratory Data Analysis (EDA) and Regression:

For a comprehensive exploratory data analysis, we focus on the year 2024, excluding any NaN values from our dataset. The following steps outline our EDA and regression analysis process:

1. **Trimming the Timeline**:
   - Reduce the dataset to only include data from the year 2024.
   - Remove NaN values to ensure the accuracy of our analysis.
2. **Visualization Techniques**:
   - **Boxplot**:
     - Boxplots are used to represent the distribution of data points, helping us understand the spread, symmetry, range, and presence of outliers in the dataset.
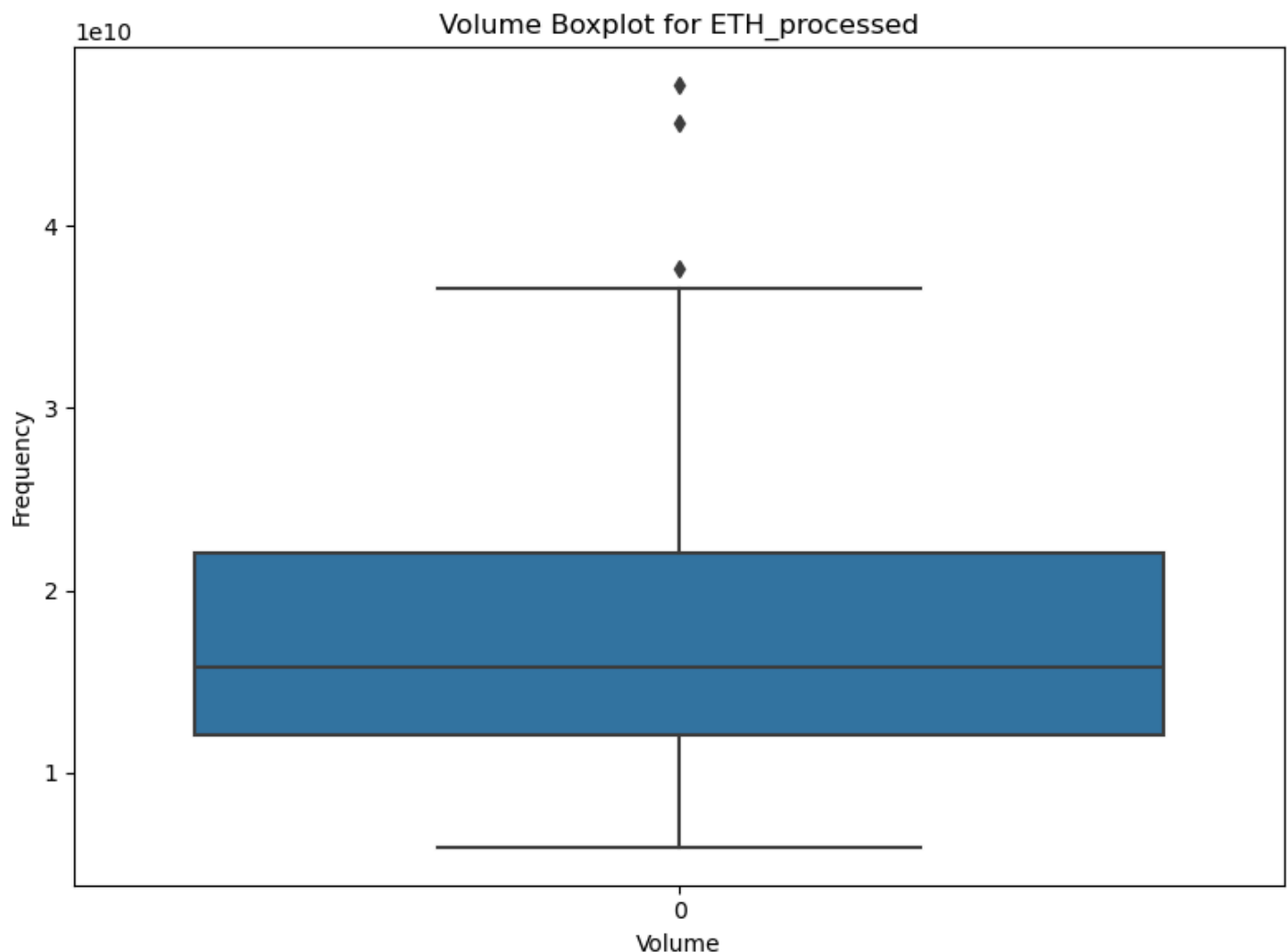
   - **Correlation Heatmap**:
     - Correlation heatmaps help identify the relationships between different variables. For instance, we might observe that Stock RSI and RSI have a strong correlation with price, while MA (Moving Average) might not be as strongly related to price but more so with Stock RSI and volume.
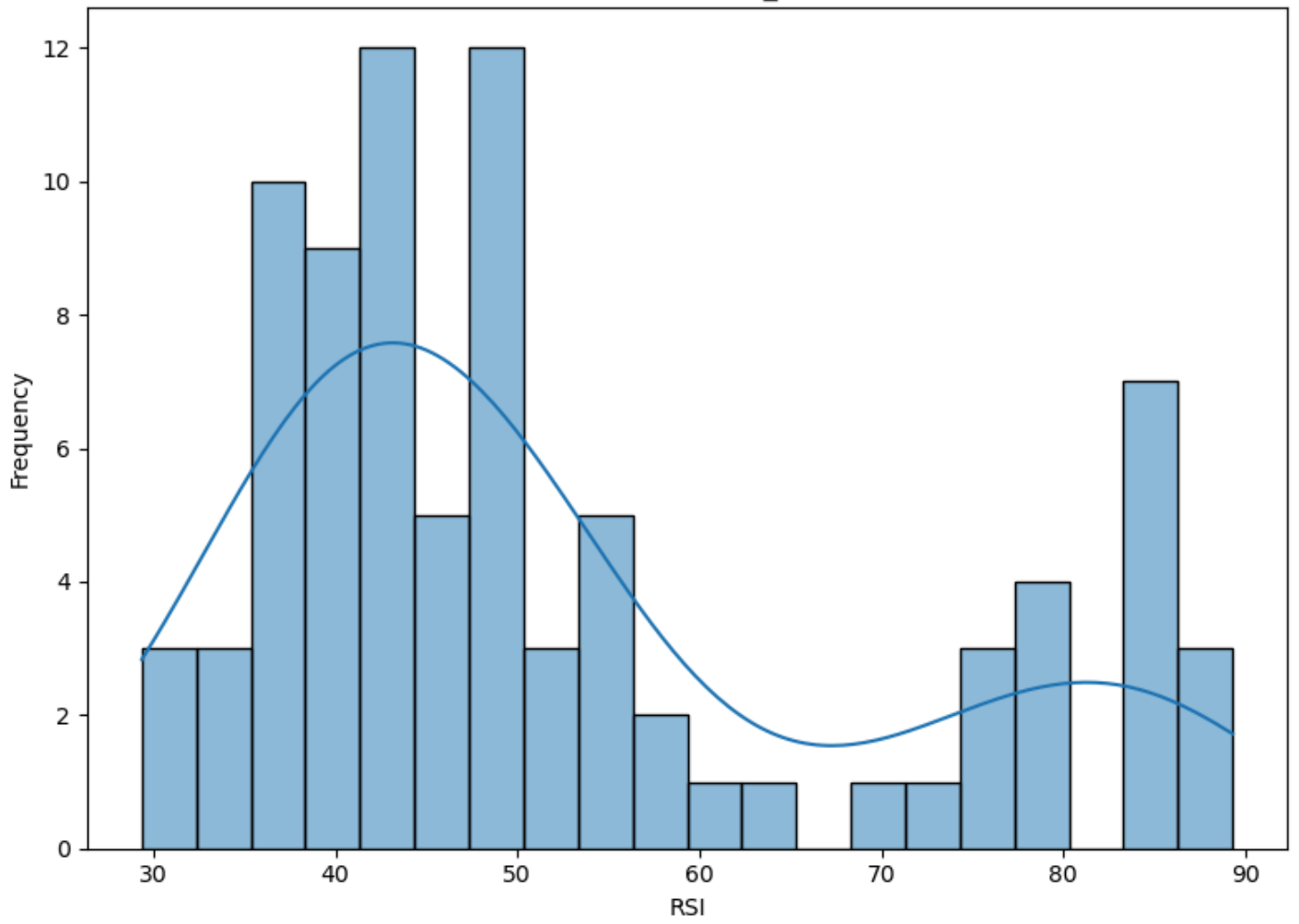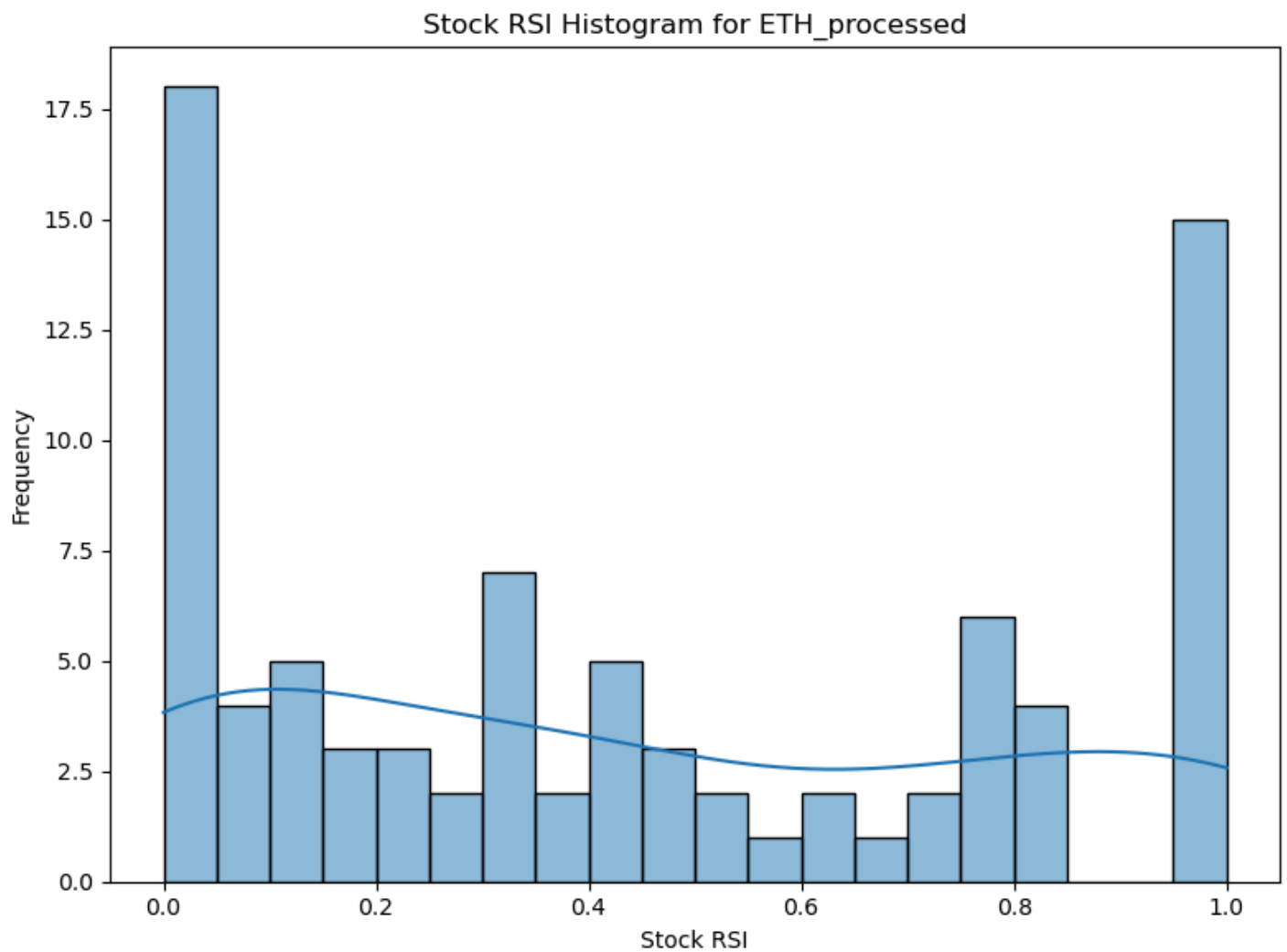
Correlation Heatmap for BTC_processed
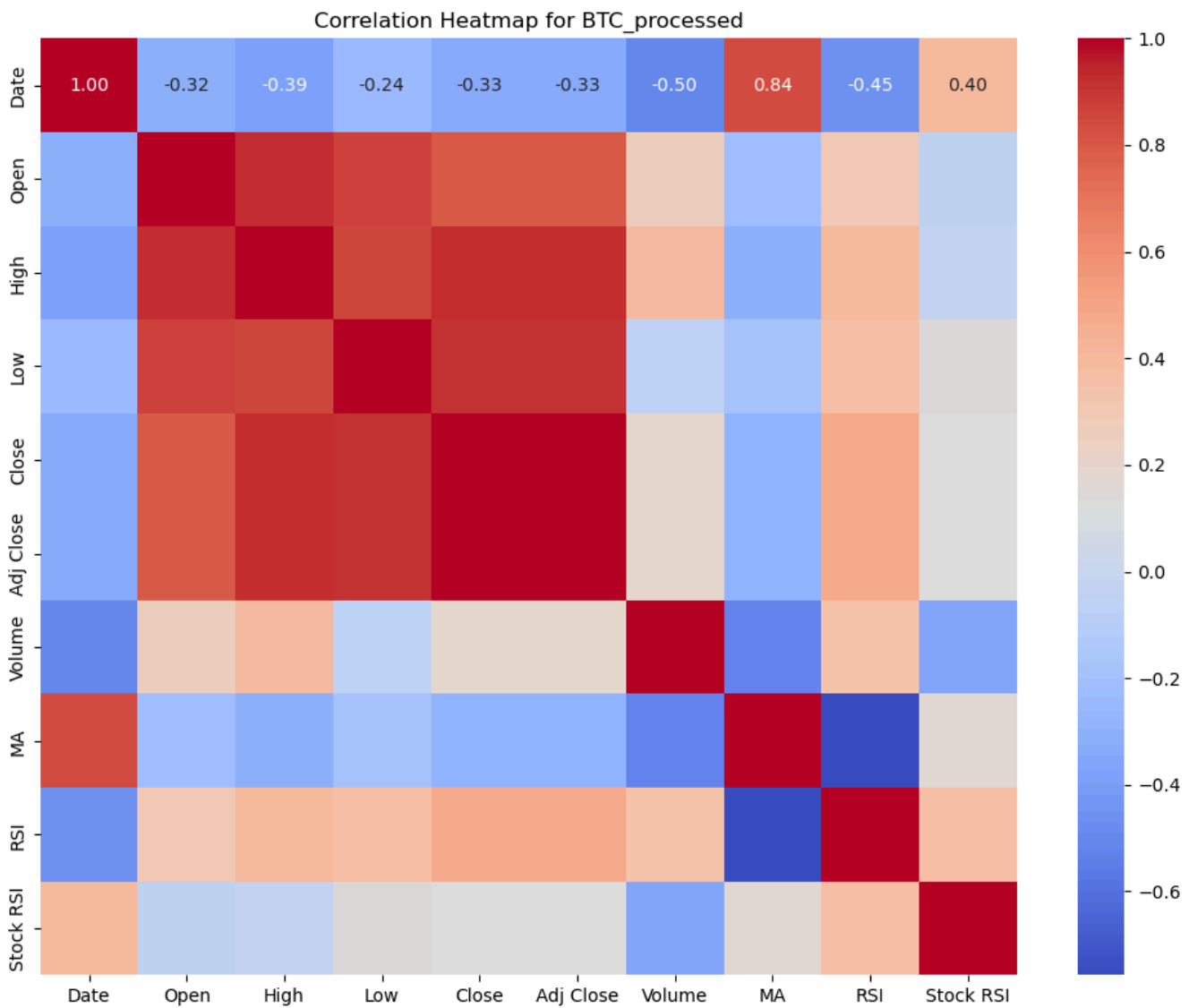
- **Histogram**:
  - Histograms of Stock RSI and RSI show the frequency of overbought and oversold conditions throughout 2024 up to May 24, 2024.
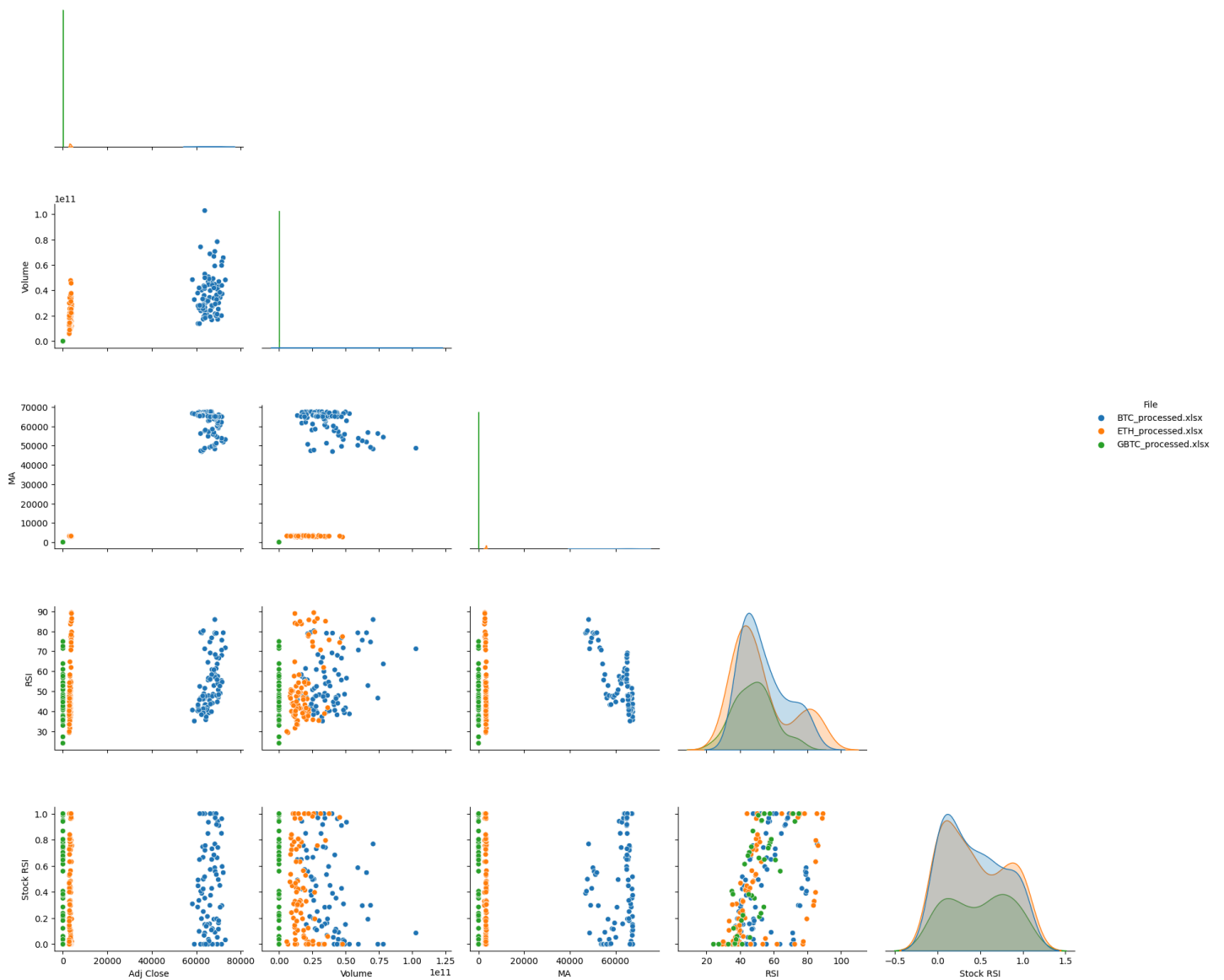
RSI Histogram for ETH_processed

Stock RSI Histogram for ETH_processed

- ○ **Pairplot**:
  - ▪ Pairplots provide a multi-dimensional view of the data, illustrating the relationships between various variables in different formats.

Super Pairplot Matrix

○ **Linear Regression**:
  ■ Linear regression is used to evaluate the Adjusted P-value, which helps understand the relationship between the Moving Average (MA) and future analysis.

Numeric Date (Days)

Regression Plot for Adj Close in HODL
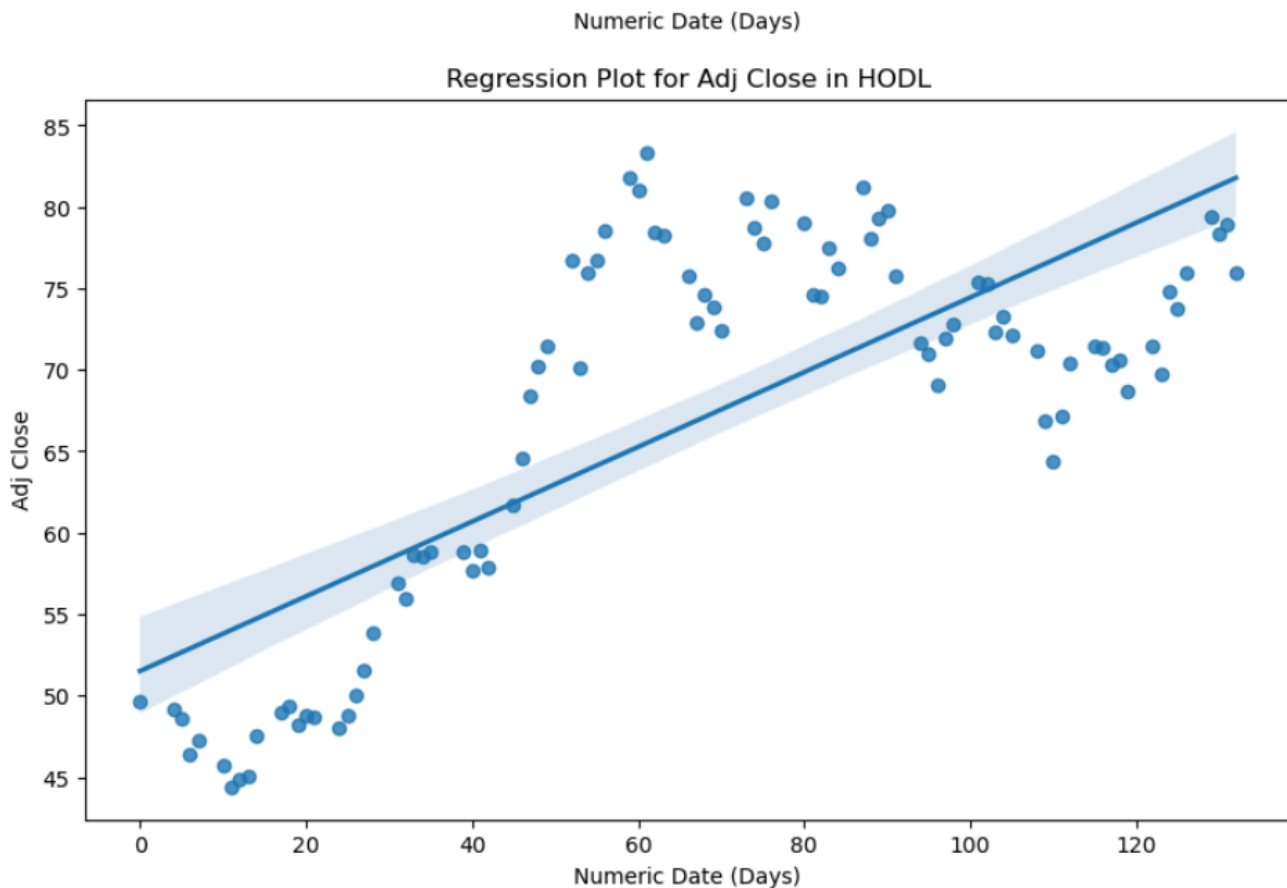
Most of the results are saved in the "EDA" folder, which includes subfolders like "New Table" (showing the common time range from 12/1/2024 to 24/5/2024), "Boxplot," "Relation Heatmap," and "Pairplot Matrix."
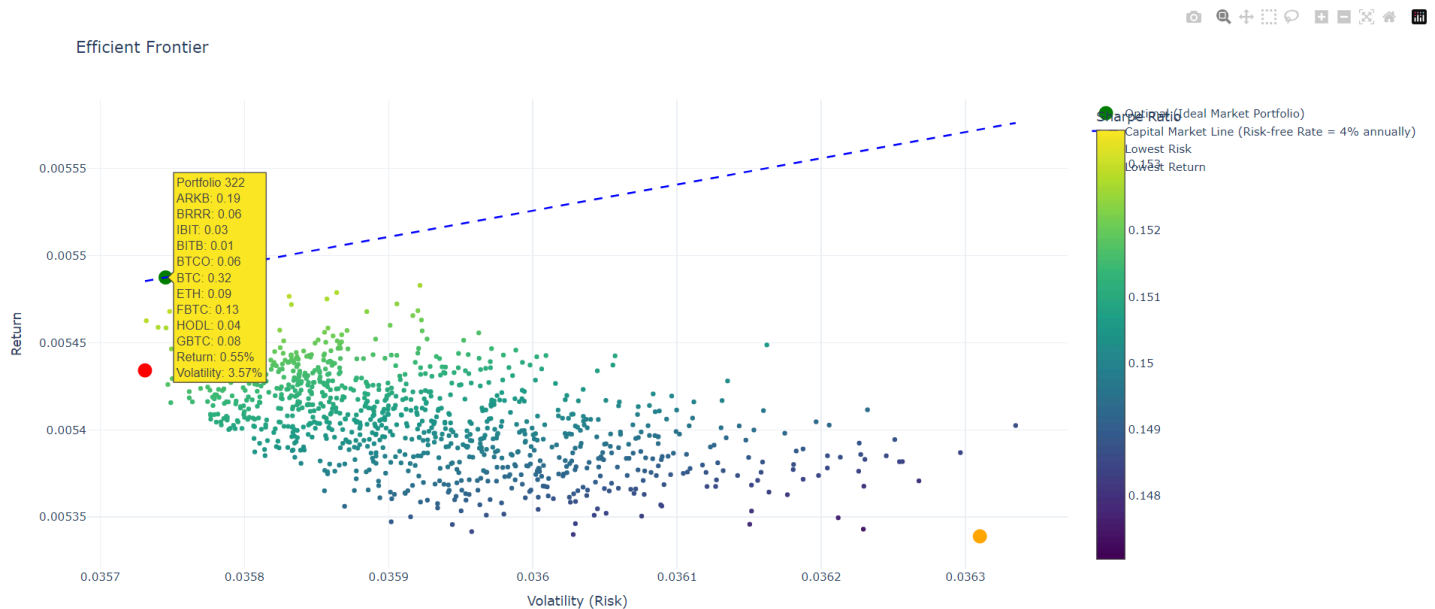
c. Efficient Frontier:

The Efficient Frontier theory, introduced by Nobel Laureate Harry Markowitz in 1952, is a key concept in modern portfolio theory (MPT). It rates portfolios based on their returns (y-axis) versus risk (x-axis).

Theory Overview:

- The Compound Annual Growth Rate (CAGR) is commonly used as the return component.
- Standard deviation (annualized) represents the risk metric.

The efficient frontier graphically represents portfolios that maximize returns for a given level of risk. Ideally, investors seek portfolios with high returns and low standard deviation. The key to achieving this is diversification, as less synchronized securities (lower covariance) reduce the overall portfolio risk.

Optimal Portfolio Determination:

- ○ Randomly assign weights to the assets.
- ○ Identify the optimal portfolio, as well as portfolios with the highest and lowest risk based on the investor's risk appetite.

Results are stored in the "Efficient Frontier" folder. For example, Portfolio #322 has the highest asset allocation in BTC at 32% and the lowest in BITB at 1%. It has an average daily return of 0.55%, which translates to an approximate annual return of 200.75%, with a volatility of 3.57%. This relatively stable outcome will be used to test investment strategies through machine learning and AI models.

d. Machine Learning:

While attempting to apply machine learning for model training, we encountered several challenges and errors, indicating that the models are not yet fully developed. These issues will be addressed and completed in future iterations of the project.

e. Conclusion:

Due to the challenges faced in using machine learning for model training, we revert to traditional methods combined with indicators like RSI, Stock RSI, and MA. Using these indicators, we define buy and sell signals:

def loc_tin_hieu_mua(df):

    return df[(df['RSI'] < 40) & (df['Stock RSI'] < 0.35) & (df['Adj Close'] <= 1.15 * df['MA'])]

ETH Adj Close với các tín hiệu Mua/Bán

FBTC Adj Close với các tín hiệu Mua/Bán
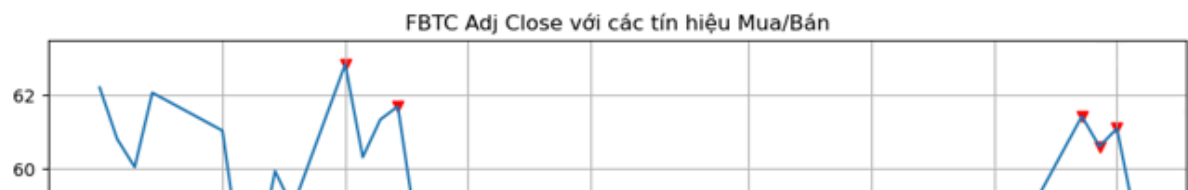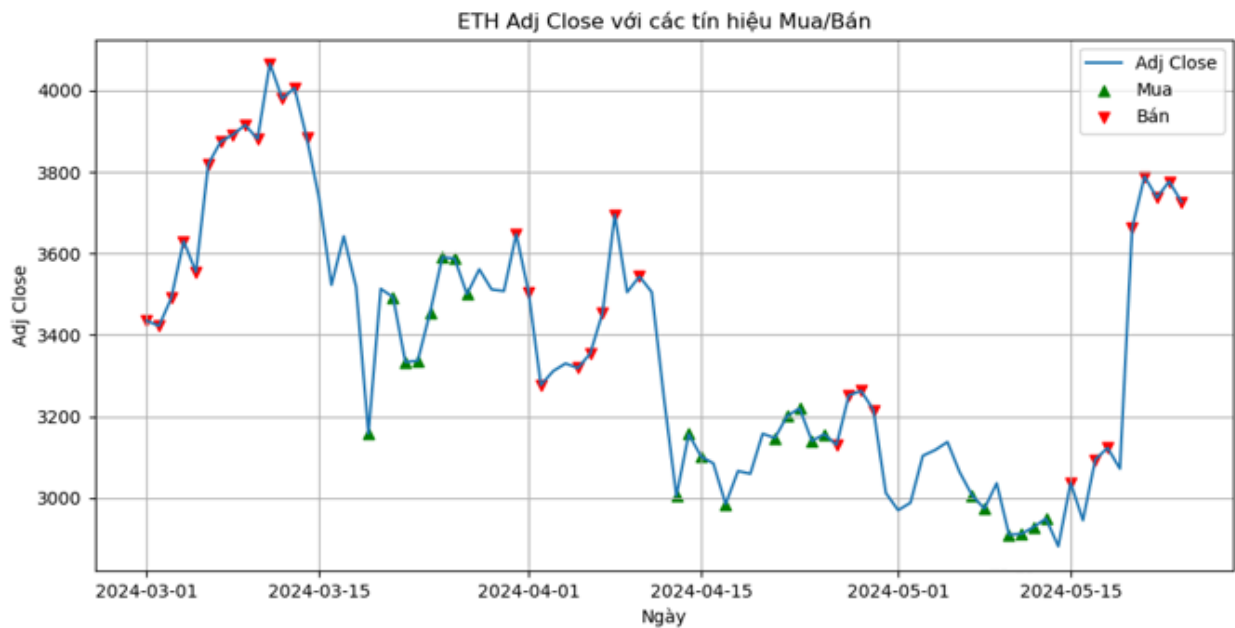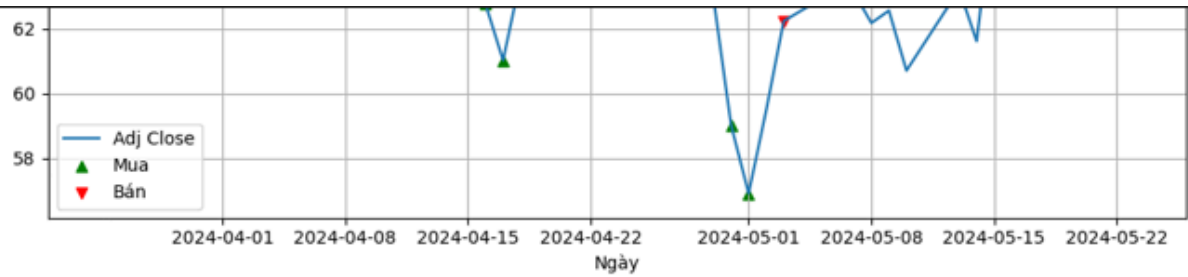

```
def loc_tin_hieu_ban(df):

    return df[(df['RSI'] > 70) | (df['Stock RSI'] > 0.75)]
```

The results are stored in the "Conclusion" folder. The returns for each asset are as follows:

- **ARKB**: 26.24%
- **BRRR**: 26.22%
- **BTC**: 23.61%
- **BTCO**: 26.16%
- **ETH**: 39.75%
- **FBTC**: 26.37%
- **GBTC**: 26.30%
- **HODL**: 26.12%
- **IBIT**: 26.43%

These results demonstrate the effectiveness of traditional methods combined with indicators for achieving significant returns in the cryptocurrency market.

4, Visualization:

We will use Tableau to create two interactive dashboards that provide comprehensive insights into the performance of ETFs and cryptocurrencies.

Dashboards:

1. Price and Volume Visualization:

- **Logarithmic Price Charts**: Display historical prices of all assets on a logarithmic scale for easy comparison.
- **Volume Charts**: Show trading volumes to understand market activity.
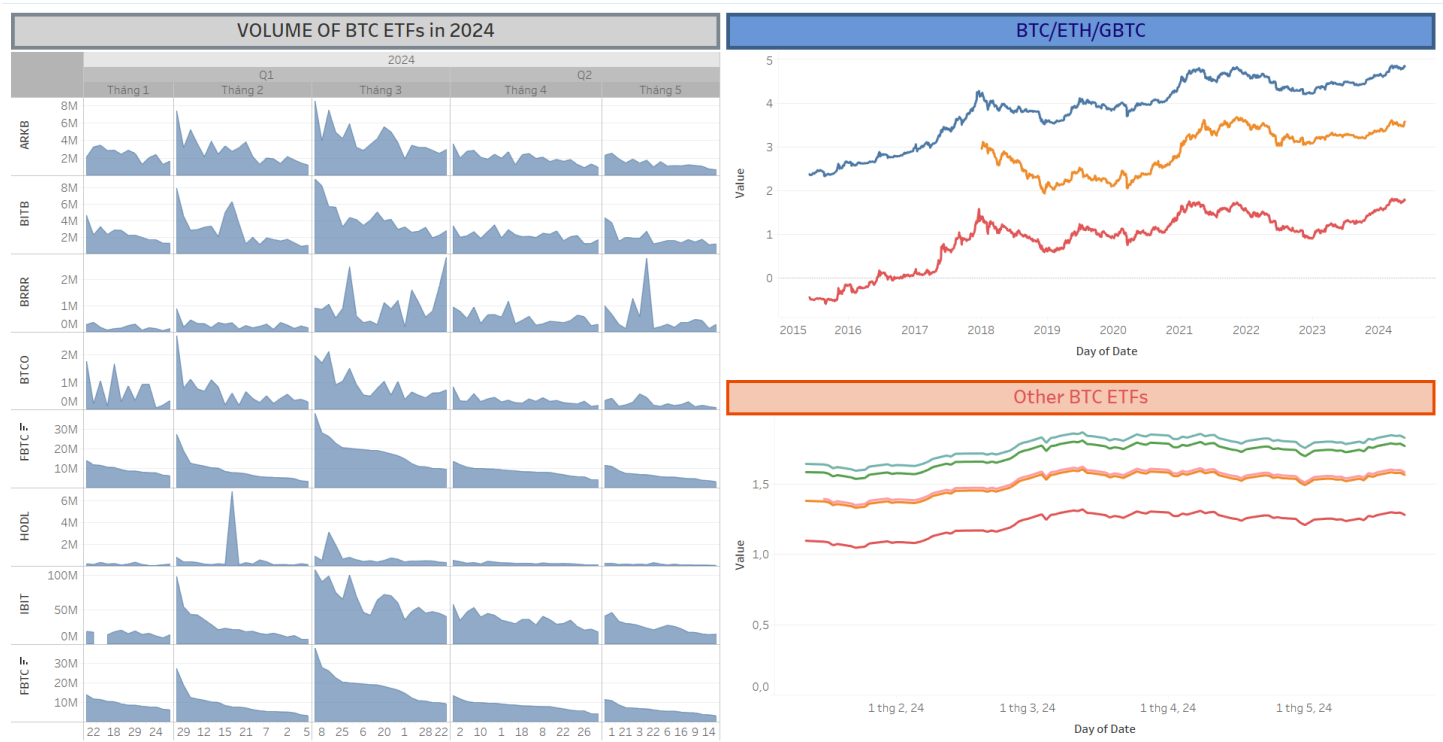
2. Indicators and Performance of ETFs & Crypto:

- **Summary Table**: Includes start date, end date, asset name, lowest price, highest price, and average volume.
- **Histograms**: Display frequency distribution of RSI and Stock RSI values.
- **Line Charts**: Visualize RSI, MACD, Stock RSI, MA, and close price trends.
- **Key Features**:
  - Performance Metrics
  - Interactive Filters
  - Annotations

Crypto/ETFs Dashboard

Start Date 26/03/2024 **End Date** 24/05/2024 **Cryto/ETFs Name** BRRR
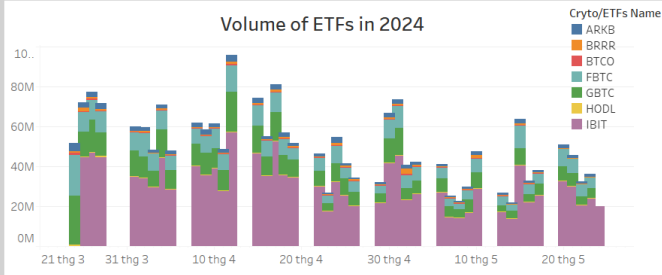
Last day: Thứ Sáu, 24 tháng 5, 2024
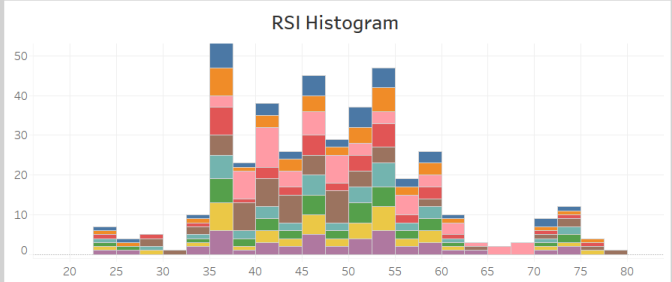
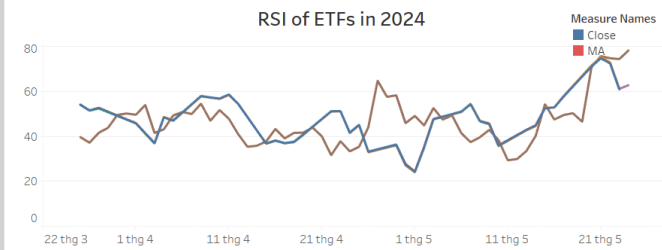Lowest Price in the Period: $16.04

Highest Price in the Period: $72,715.36

Volume of ETFs in 2024

Cryto/ETFs Name: ARKB, BRRR, BTCO, FBTC, GBTC, HODL, IBIT

RSI Histogram

RSI of ETFs in 2024 — Measure Names: Close, MA

Moving Average - Close Price
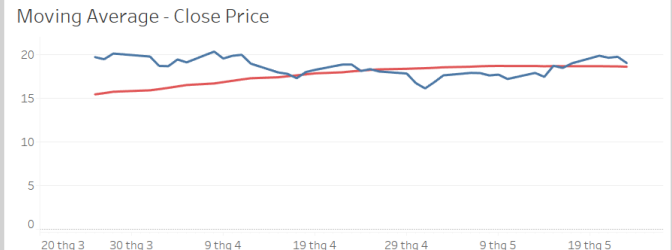
Outcomes:

- **Enhanced Understanding**: Clear insights into price and volume trends and performance indicators.
- **Actionable Insights**: Interactive dashboards enable deep data exploration.
- **Comprehensive Analysis**: Summary tables, histograms, and line charts provide a holistic market view.

These visualizations will make complex data accessible and help users make informed decisions in the ETF and cryptocurrency markets.