**Project Report:**

**Identifying Fairy Stocks in the US Market**

**Introduction**

The goal of this project is to identify abnormal stock behavior, referred to as "妖股," in the US market using historical stock price data obtained from Yahoo Finance API. This report outlines the steps taken to achieve this objective, including data collection, analysis, and database management. The project utilizes Python and SQLite to build a database of stock prices and analyzes abnormal behavior using three key metrics: Unusual Price Movements, High Volatility, and High Trading Volume.

**Project Implementation**

**Data Collection and Database Creation**

Data Source: The project uses the Yahoo Finance API (yfinance) to obtain historical stock price data for 40 well-known US stocks. These stocks are specified in the my\_list variable.

Database Management: The data is stored in an SQLite database named 'us\_stock\_data.db.' Each stock is represented as a separate table in the database. The database schema includes fields for Date, Open, High, Low, Close, Volume, and Log Return, which is calculated by myself.

Data Cleaning and Maintenance: To ensure data accuracy and relevance, the script performs the following actions:

Calculates log returns and adds them as a new column.

Retains only the latest data points (maximum 774 rows) to maintain the desired historical data.

Deletes older data to ensure data in the last 60 days.

**Abnormal Stock Behavior Detection**

The script identifies abnormal stock behavior based on the following three metrics:

1. Unusual Price Movements:

I used Log\_Return to analyze the Unusual Price Movements because as we all know, the price is a random walk. So it cannot be analyzed using any distributions. Then, I want to test if Log\_Return is normally distributed. By using the JB test and our finance, in general, the Log\_Return of stocks is not normally distributed. But in fact, it is normally distributed with a heavy tail. Therefore, I want to use the normal distribution but make the threshold for abnormal returns bigger than a normal distribution.

Calculates the Z-score of the Log\_Return (logarithmic return) for each stock.

Defines a threshold for abnormal returns (Z-score: 15.0).

Identifies dates with unusual return rates exceeding the threshold.

Stores these dates as potential instances of unusual price movements.

1. High Volatility:

I believe that if a stock has a 10% price change rate within 30 minutes, we can say that it is High Volatility(Extreme price fluctuations in a short period of time).

Computes the maximum price change rate (High - Low) relative to Low for each stock.

Defines a threshold for high volatility (e.g., maximum price change rate above 0.1).

Identifies dates with high volatility based on the threshold.

Stores these dates as potential instances of high volatility.

1. High Trading Volume:

Just as the same as Log\_Return, it is like normally distributed with heavy tail. Therefore, I want to use normal distribution but make the threshold for abnormal returns bigger than normal distribution.

Computes the Z-score of the trading volume for each stock.

Defines a threshold for high trading volume (e.g., Z-score above 15.0).

Identifies dates with unusually high trading volume based on the threshold.

Stores these dates as potential instances of high trading volume.

**Final Result and Information Store**

An additional SQLite database named 'fairy\_stocks.db' is created to store information about the identified 妖股. This database has the following structure:

fairy\_stocks\_list table:

fairy\_stock\_symbol (TEXT, PRIMARY KEY): Symbol of the potential 妖股.

reasons\_to\_be\_fairy\_stocks (TEXT): Reasons for considering the stock as a potential 妖股.

tip (TEXT): A tip is provided based on the number of abnormal data points identified. If there are three or fewer abnormal data points, a tip suggests checking for obvious reasons or news that could justify the change. This is because some obvious reasons or news could make the stock have Unusual Price Movements or High Volatility or High Trading Volume. But it is reasonable, so we cannot say it is a妖股. For example, in our result, AMZN has an Unusual return rate on 2023-08-04 09:30:00 because it published a nice financial report. So we will give a tip to make people check it, if the abnormal data is less than or equal to 3 data.

**Results**

|  |  |  |
| --- | --- | --- |
| Fairy\_stock\_symbol | Reasons to be fairy stocks | tip |
| AMZN | Unusual return rate on 2023-08-04 09:30:00 | The abnormal data is 1 (less than or equal to 3 data), please check if any obvious reason or news that would justify such a change. If yes, it may not be a fairy stock. |
| VFS | High volatility on 2023-08-15 09:30:00, 2023-08-15 10:00:00, 2023-08-15 10:30:00, 2023-08-15 11:00:00, 2023-08-15 11:30:00, 2023-08-15 12:00:00, 2023-08-15 14:00:00, 2023-08-15 14:30:00, 2023-08-15 15:00:00, 2023-08-15 15:30:00, 2023-08-16 09:30:00, 2023-08-16 10:00:00, 2023-08-16 15:30:00, 2023-08-17 09:30:00, 2023-08-17 14:30:00, 2023-08-18 09:30:00, 2023-08-18 10:00:00, 2023-08-18 11:30:00, 2023-08-18 12:00:00, 2023-08-18 12:30:00, 2023-08-21 09:30:00, 2023-08-21 10:00:00, 2023-08-22 09:30:00, 2023-08-22 10:00:00, 2023-08-22 10:30:00, 2023-08-22 11:00:00, 2023-08-22 11:30:00, 2023-08-22 12:00:00, 2023-08-22 12:30:00, 2023-08-22 14:30:00, 2023-08-22 15:00:00, 2023-08-22 15:30:00, 2023-08-23 09:30:00, 2023-08-23 10:00:00, 2023-08-23 10:30:00, 2023-08-24 09:30:00, 2023-08-24 10:00:00, 2023-08-24 10:30:00, 2023-08-24 11:00:00, 2023-08-24 11:30:00, 2023-08-24 13:30:00, 2023-08-25 09:30:00, 2023-08-25 10:00:00, 2023-08-25 11:30:00, 2023-08-25 12:00:00, 2023-08-25 12:30:00, 2023-08-25 13:00:00, 2023-08-28 09:30:00, 2023-08-28 10:00:00, 2023-08-28 11:00:00, 2023-08-29 09:30:00, 2023-08-29 10:30:00, 2023-08-29 11:00:00, 2023-08-29 11:30:00, 2023-08-30 09:30:00, 2023-08-30 10:30:00, 2023-08-31 09:30:00, 2023-09-01 09:30:00, 2023-09-01 13:30:00 |  |
| COIN | High volatility on 2023-06-30 09:30:00, 2023-08-29 10:00:00 | The abnormal data is 2 (less than or equal to 3 data), please check if any obvious reason or news that would justify such a change. If yes, it may not be a fairy stock. |
| TUP | High volatility on 2023-07-11 10:30:00, 2023-07-21 10:00:00, 2023-07-21 14:00:00, 2023-07-21 14:30:00, 2023-07-24 09:30:00, 2023-07-24 10:00:00, 2023-07-24 11:00:00, 2023-07-24 13:00:00, 2023-07-24 15:30:00, 2023-07-25 09:30:00, 2023-07-25 10:00:00, 2023-07-25 11:00:00, 2023-07-25 12:00:00, 2023-07-26 09:30:00, 2023-07-26 12:00:00, 2023-07-26 12:30:00, 2023-07-26 13:00:00, 2023-07-26 14:00:00, 2023-07-26 14:30:00, 2023-07-27 09:30:00, 2023-07-27 11:00:00, 2023-07-27 11:30:00, 2023-07-27 12:00:00, 2023-07-27 12:30:00, 2023-07-27 13:30:00, 2023-07-27 15:30:00, 2023-07-28 09:30:00, 2023-07-28 10:00:00, 2023-07-28 10:30:00, 2023-07-28 14:30:00, 2023-07-31 09:30:00, 2023-07-31 10:00:00, 2023-07-31 10:30:00, 2023-07-31 11:00:00, 2023-07-31 12:00:00, 2023-07-31 13:30:00, 2023-08-01 09:30:00, 2023-08-01 10:00:00, 2023-08-01 10:30:00, 2023-08-01 11:00:00, 2023-08-02 09:30:00, 2023-08-02 11:00:00, 2023-08-03 09:30:00, 2023-08-03 10:30:00, 2023-08-03 11:00:00, 2023-08-04 09:30:00, 2023-08-04 10:00:00, 2023-08-04 13:00:00, 2023-08-04 13:30:00, 2023-08-07 09:30:00, 2023-08-10 09:30:00, 2023-08-11 09:30:00, 2023-08-14 09:30:00, 2023-08-17 09:30:00, 2023-08-18 09:30:00, 2023-08-18 10:30:00, 2023-08-18 11:00:00, 2023-08-21 09:30:00 |  |
| GME | High trading volume on 2023-06-08 09:30:00 | The abnormal data is 1 (less than or equal to 3 data), please check if any obvious reason or news that would justify such a change. If yes, it may not be a fairy stock. |
| CVNA | High volatility on 2023-06-08 09:30:00, 2023-06-08 10:00:00, 2023-06-08 11:30:00, 2023-06-08 12:00:00, 2023-06-09 09:30:00, 2023-06-09 10:30:00, 2023-06-12 09:30:00, 2023-06-16 09:30:00, 2023-07-10 10:00:00, 2023-07-14 09:30:00, 2023-07-19 09:30:00, 2023-07-19 10:00:00, 2023-07-20 09:30:00, 2023-07-21 09:30:00 |  |

**Conclusion**

This project successfully accomplishes the task of identifying abnormal stock behavior (妖股) in the US market. It utilizes historical stock price data from Yahoo Finance, implements three key metrics for abnormality detection, and maintains a database of identified potential 妖股. The use of SQLite databases ensures data integrity and allows for easy retrieval of information about abnormal stocks.

Future enhancements could include the automatic check through news fetch or something else to determine if any obvious reason or news on the date would justify such a change, and thereby it is not a 妖股(Now, we need to search for info by ourselves). In addition, we can adjust the tuning parameters to make them more precise.