

1. Introduction

The goal of task is to analyze high-frequency market data from the `aggtrade` and `depth20` datasets. The data is aggregated at a per-second granularity and then subjected to feature engineering to extract meaningful insights into market behavior. It employs unsupervised learning, specifically K-Means clustering, to detect market regimes based on these engineered features. After applying the K-Means algorithm, the optimal number of clusters is determined, and the market is segmented into distinct regimes, such as trending, volatile, and liquid conditions.

2. Data Processing and Aggregation

a. Aggregation of `aggtrade` and `depth20` Data

We start by aggregating `aggtrade` data at the per-second level:

- **Price:** The mean of the prices within each second is calculated.
- **Quantity:** The total quantity of trades for each second is summed.
- **NumTrades:** The number of trades within the second is summed.
- **Other Features:** The `IsMarketMaker` flag and `M` field are carried forward based on logical rules (e.g., majority vote or presence of a market maker).

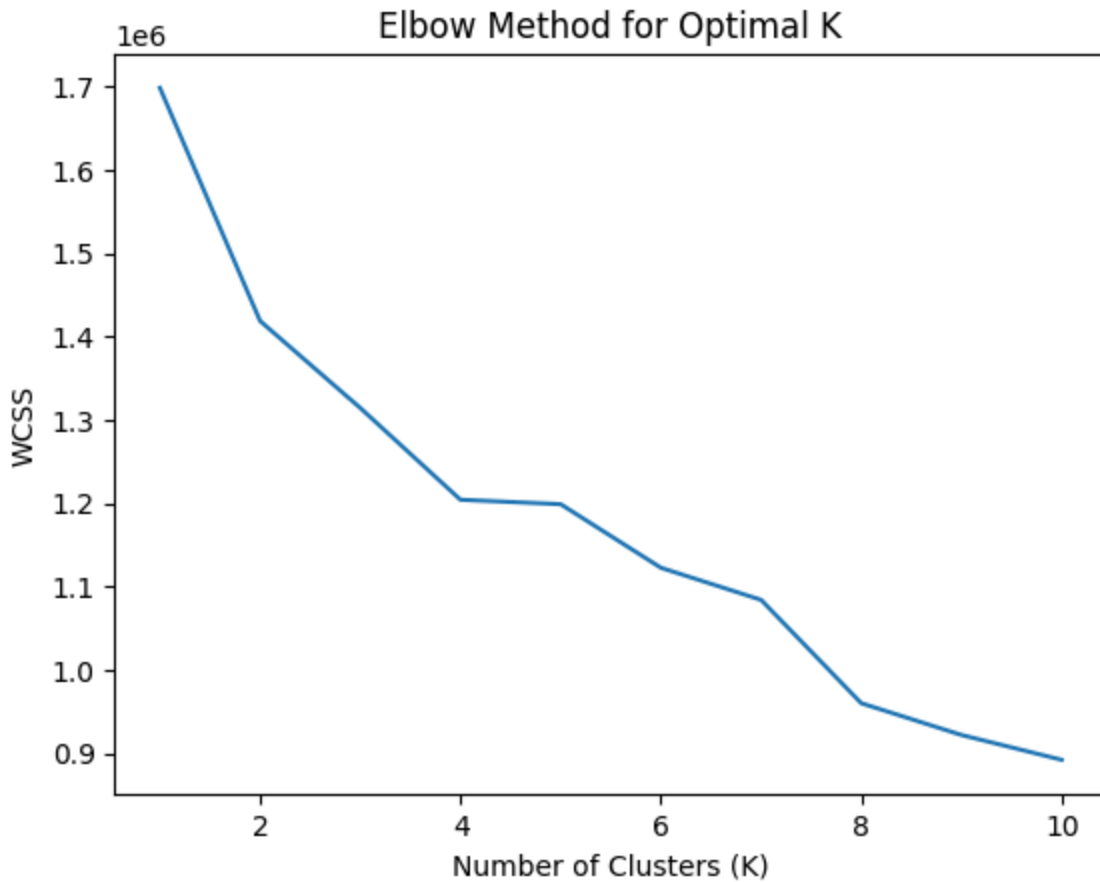
Similarly, `depth20` data is aggregated in the same manner, calculating key features related to order book depth.

After aggregation, both datasets (`aggtrade` and `depth20`) are merged based on their timestamps (`TimeRounded`), which ensures that each second's market data contains both trade and order book information.

3. Clustering and Regime Detection

a. K-Means Clustering

K-Means clustering is applied to the aggregated and normalized features to segment the market into different regimes. The elbow method is used to determine the optimal number of clusters, which is found to be **4**. The K-Means algorithm is then applied with `k=4` to classify the market data into four distinct regimes.



The resulting clusters represent different market conditions, which are labeled as distinct regimes, such as:

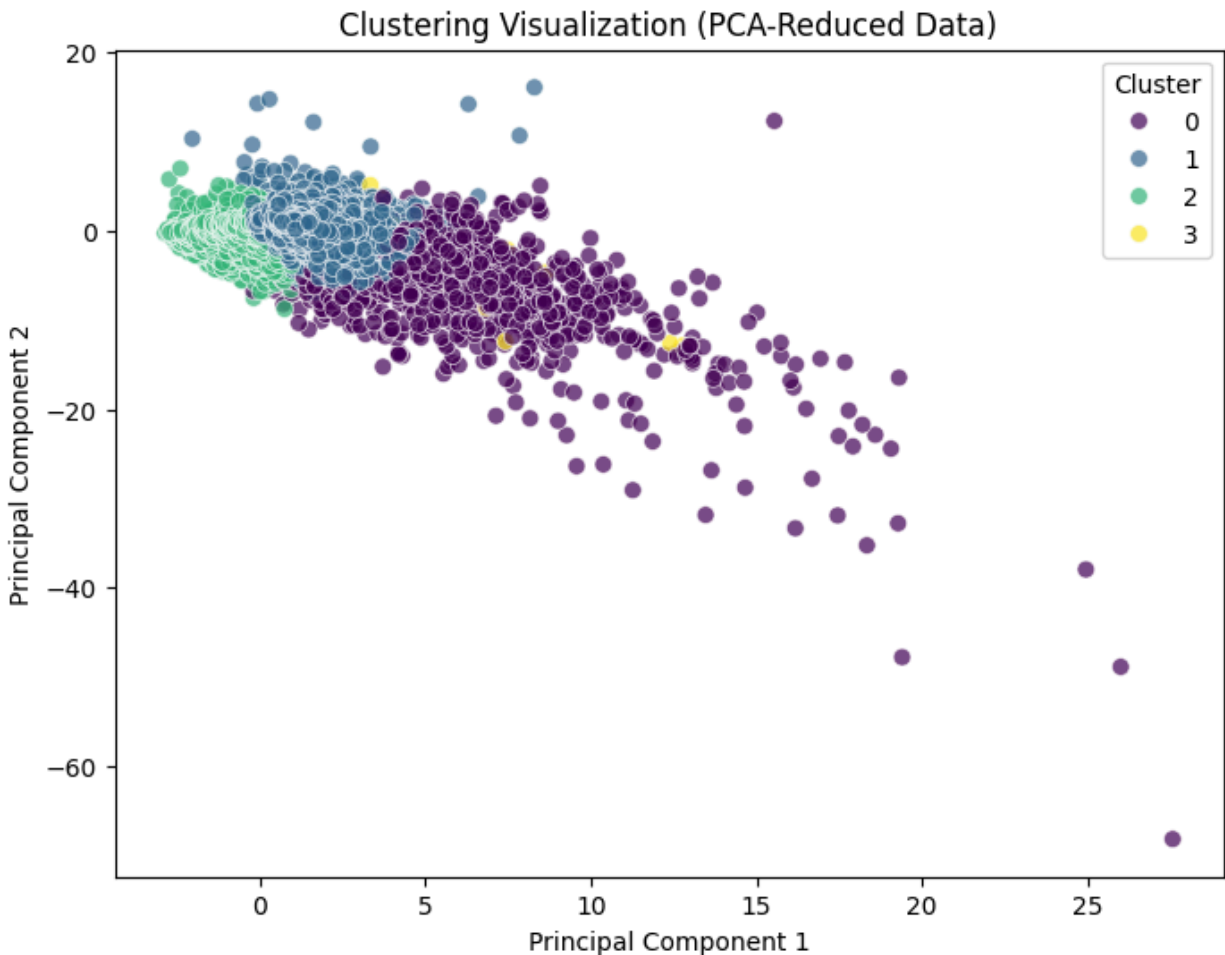
- **Trending:** Markets with consistent upward or downward price movements.
- **Volatile:** Markets with high price fluctuations and low predictability.
- **Stable:** Markets with relatively consistent prices and low fluctuations.
- **Liquid:** Markets with high trading volumes and active order books.

b. Clustering Metrics

To assess the quality of the clustering, several metrics such as the **Davies-Bouldin Index** and **inertia** is used. it shows the fit is moderately good

Davies-Bouldin Index: 1.39235637356841.

4. Clustering Results



The clustering process revealed the following market regimes, each characterized by different behaviors:

- **Regime 0: Trending & Liquidity & Stable**
 - This regime represents a market in a clear trend (upward or downward) while being both stable and liquid. Such conditions are ideal for trend-following strategies.
- **Regime 1: Stable & Liquid**
 - Markets in this regime show stability and liquidity, but there are no strong directional movements. These markets are ideal for range-bound trading strategies.

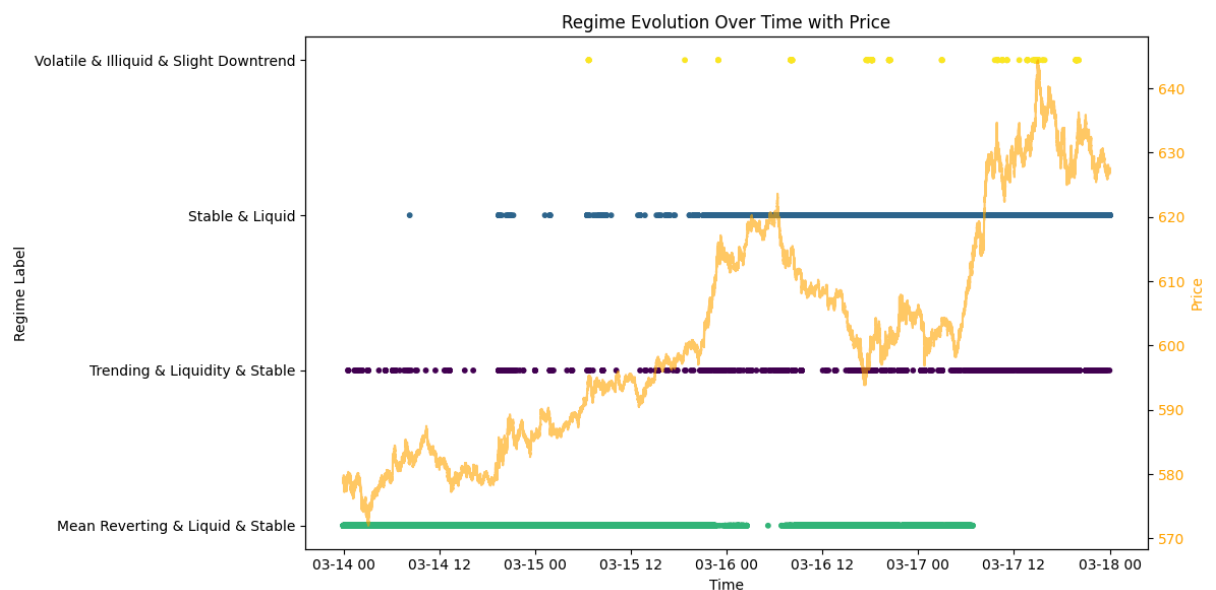
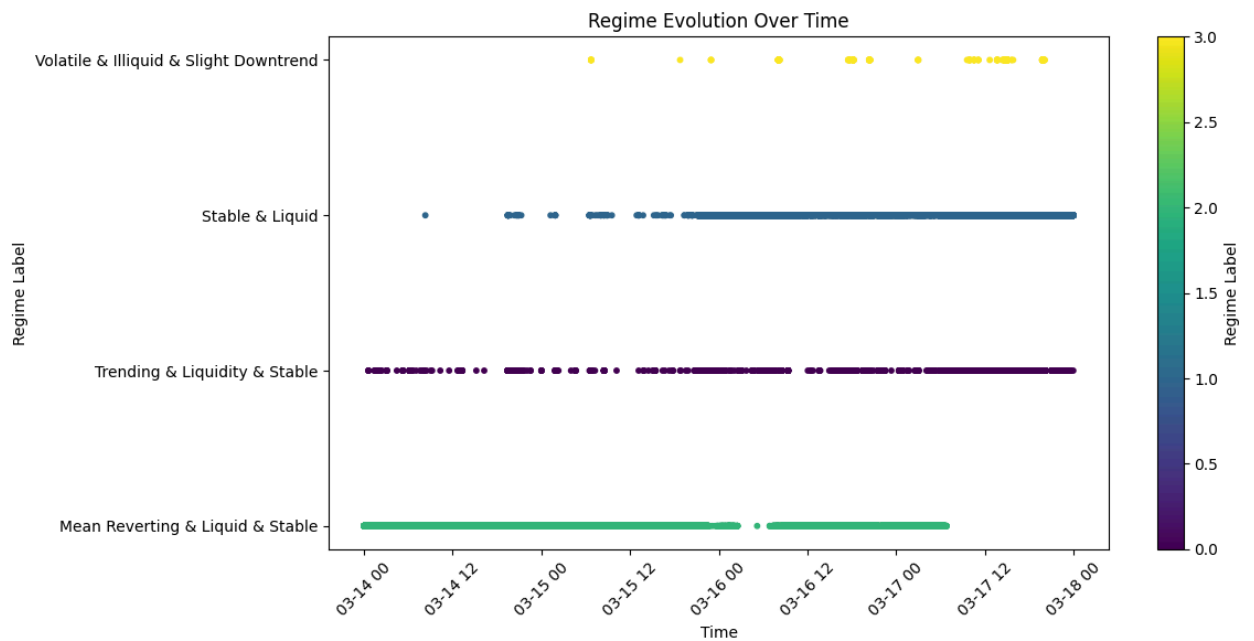
- **Regime 2: Mean Reverting & Liquid & Stable**

- This regime shows a stable market with mean-reverting price behavior. Traders can capitalize on price oscillations around a central value in these conditions.

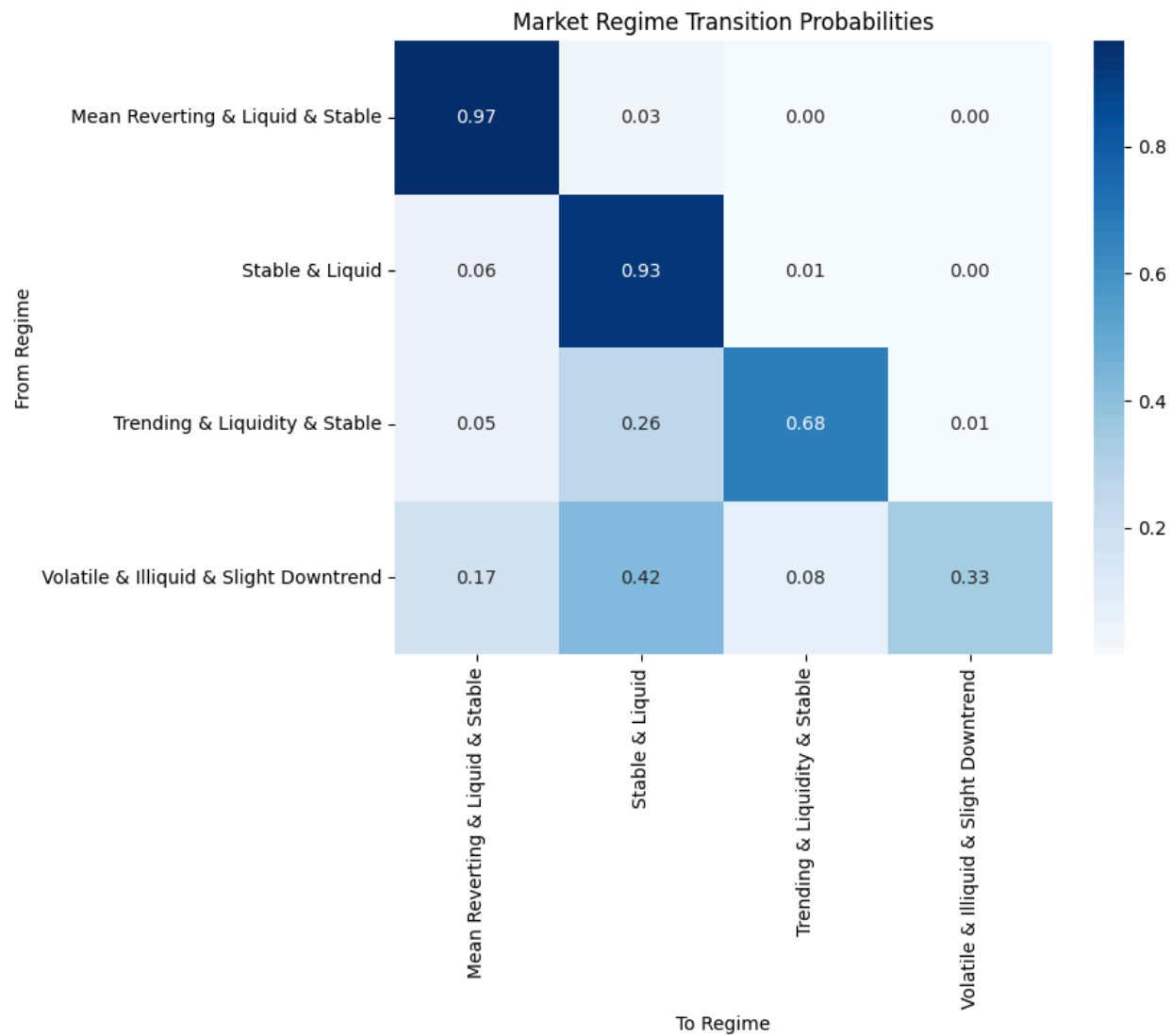
- **Regime 3: Volatile & Illiquid & Slight Downtrend**

- Markets in this regime are volatile, illiquid, and show a slight downward trend. This regime may present opportunities for contrarian strategies but also poses higher risks

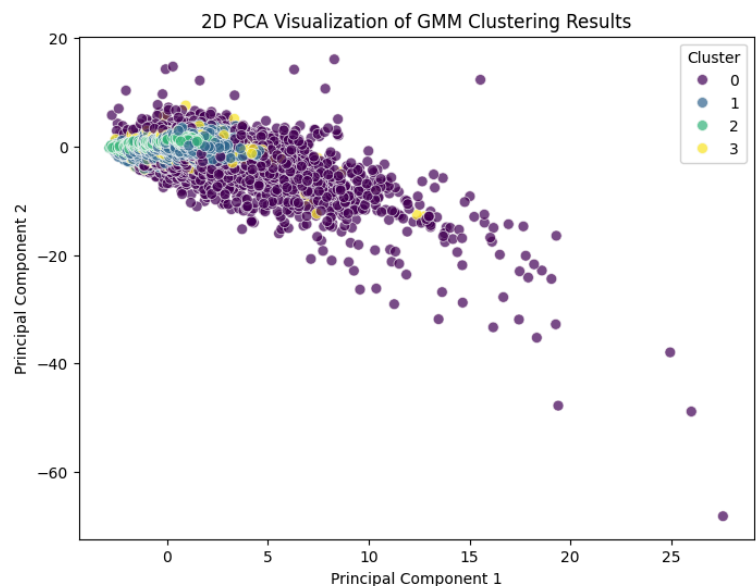
5. Visualizations



Probability of going from one regime to other



6. GMM



The following four regimes were identified by the GMM clustering, each representing a unique market condition:

- **Regime 0: Trending & Liquidity & Volatile**

- This regime represents markets that are in a clear trend (either upward or downward), while also being liquid. The market exhibits volatility during the trending phase, which could be ideal for momentum or trend-following strategies.

- **Regime 1: Stable & Liquid**

- Markets in this regime are both stable and liquid but show little to no directional movement. This type of market is conducive to range-bound trading strategies, where traders capitalize on price oscillations within a defined range.

- **Regime 2: Mean Reverting & Illiquid & Stable**

- This regime shows a market that is stable, but prices are mean-reverting. The market exhibits periods where the price moves back towards a central value. However, it is also illiquid, which can create higher spreads and lower trading activity. Traders can exploit these oscillations for short-term profits.

- **Regime 3: Volatile & Illiquid & Downtrend**

- A highly volatile market that is also illiquid and shows a slight downward trend. This regime is associated with increased risks and presents opportunities for contrarian trading strategies that bet on price reversals. However, the illiquidity makes it a high-risk environment.

