Regime Detection via Unsupervised Learning

- Ashwarya Jayesh Bhandarkar, IIT Kanpur

1. Introduction

This report presents the results of a market regime detection task using unsupervised learning on order book and volume data for RozReturns. The primary goal of this task was to detect different market regimes (trending vs mean-reverting, volatile vs stable, liquid vs illiquid) using a set of derived features such as liquidity, depth, volatility, and volume. The analysis employs clustering techniques to uncover hidden patterns in the data and provides insights into regime transitions.

2. Methodology

2.1 Data Preparation

The dataset contains multiple features representing different aspects of the market, such as:

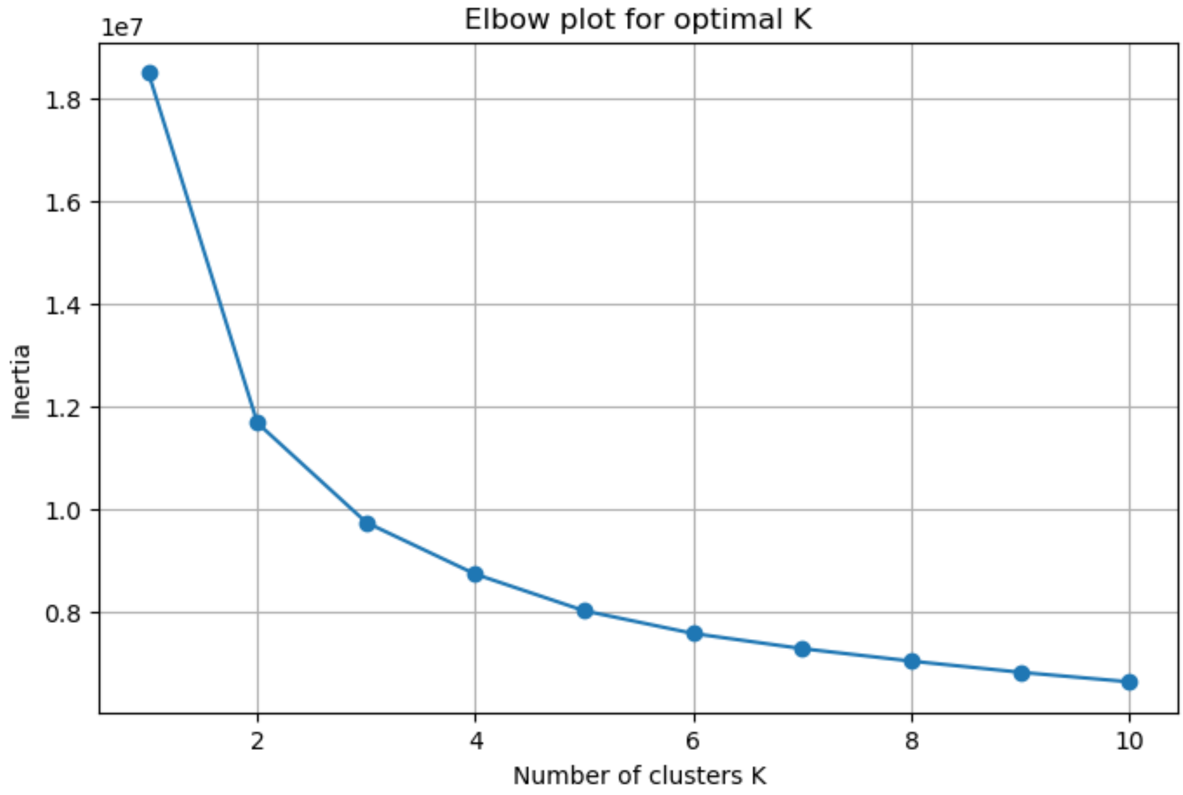
- **Liquidity Features**: Bid/ask spread, order book imbalance, volume imbalance.
- **Depth Features**: Cumulative depth, sloped depth.
- **Price Action Features**: VWAP shift, rolling mid-price return.
- **Volume Features**: Volume imbalance, market microprice.

These features were extracted and normalized using **StandardScaler**, ensuring comparability across different feature sets. The normalized features were grouped into five sets:

- **Feature Set 1 (volume-based)**
- **Feature Set 2 (price action-based)**
- **Feature Set 3 (depth-based)**
- **Feature Set 4 (imbalance-based)**
- **Feature Set 5 (microprice-based)**

2.2 Clustering Approach

Clustering was performed using the **K-Means algorithm**, which is effective for unsupervised learning in high-dimensional spaces. The number of clusters (k) was determined using the **Elbow Method** to balance between model complexity and clustering performance.



2.3 Evaluation Metrics

To evaluate the clustering performance, the following metric was used:

- **Silhouette Score**: Measures how similar each data point is to its own cluster compared to other clusters.
- Achieved a **clustering quality** of **0.22**

3. Results

3.1 Clustering Results

After applying the K-Means clustering algorithm, the market data was segmented into distinct regimes. Resulting clusters were analysed for their characteristics, few of which include:

- **Cluster 1 (Trending, High Liquidity)**: Associated with high volume and liquidity, exhibiting a strong upward or downward market trend.

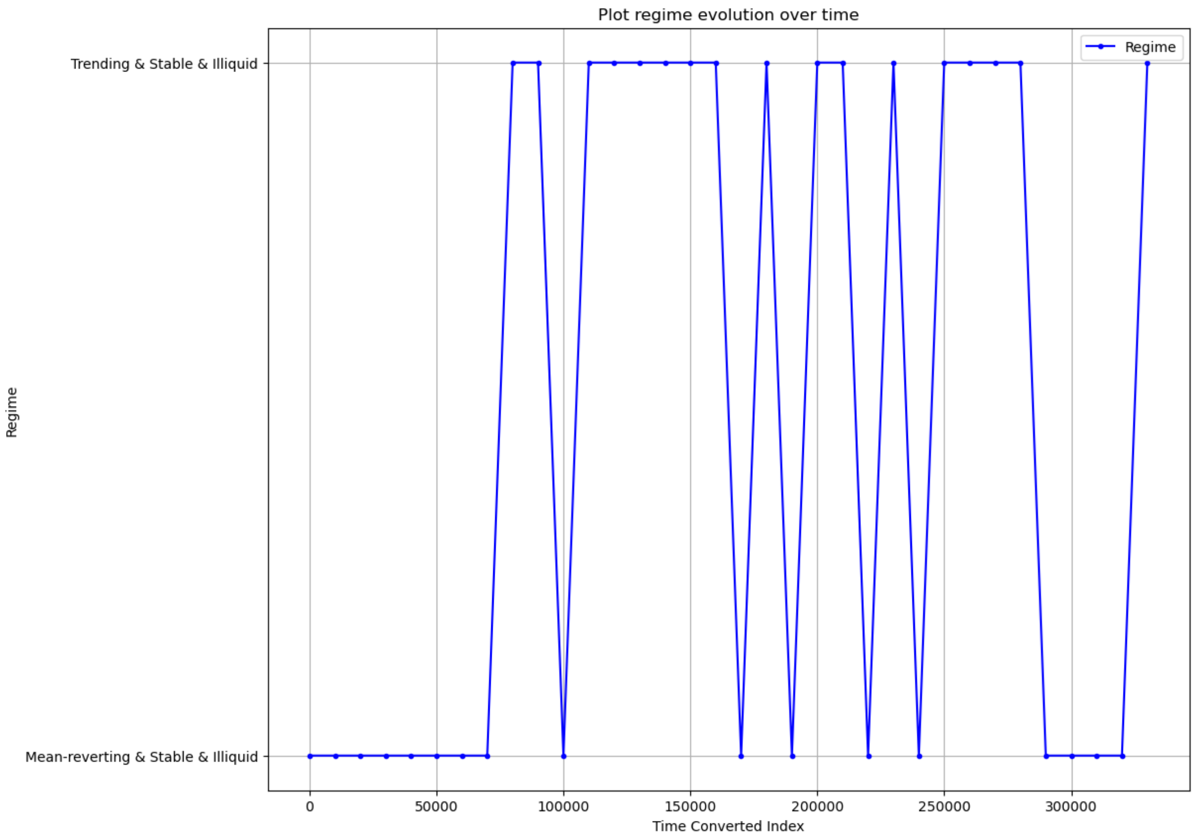
- **Cluster 2 (Stable, Low Volatility)**: Characterized by low volatility and stable prices.
- **Cluster 3 (Mean-Reverting, High Volatility)**: A volatile regime with price fluctuations but reverting to the mean.
- **Cluster 4 (Illiquid, Low Activity)**: Characterized by low trading volume and low market participation.

3.2 Regime Insights

- **Regime Transitions**: The clusters reveal how the market shifts between trending, stable, volatile, and illiquid phases.
- **Market Liquidity**: Liquidity and order book imbalance are strong indicators of trending and illiquid regimes.
- **Volatility**: Volatility features like the rolling mid-price return and VWAP shift help distinguish between stable and volatile regimes.

3.3 Visualizations

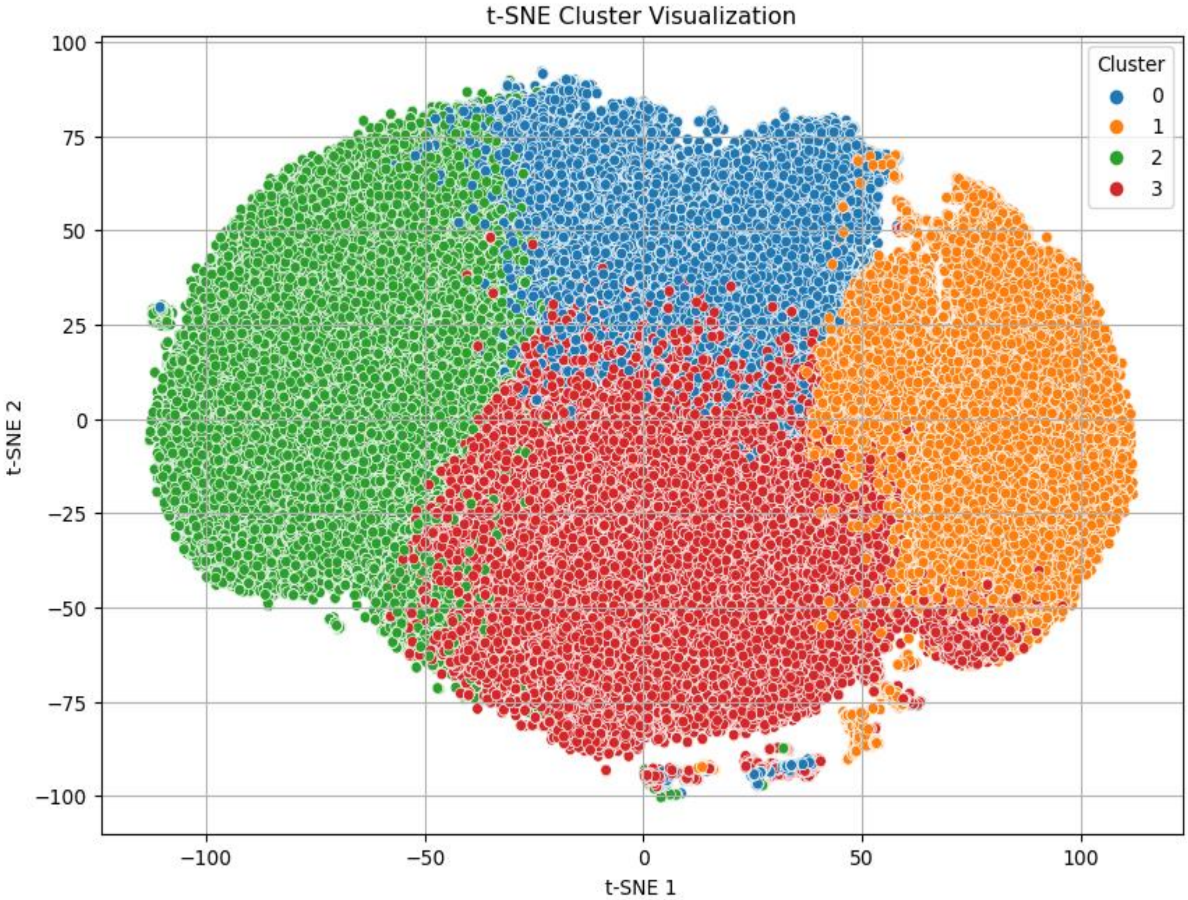
- **Graph Insights:**



The graph displays how market regime changes over time periods for all the five days included also, since the merged data had over 3 Lakh+ time stamps, they have been converted into indices for better understanding

- **Cluster Visualization (t-SNE):**

  To better understand the high-dimensional feature space, t-distributed Stochastic Neighbor Embedding (t-SNE) is used to reduce the data to two dimensions. This technique preserves the local structure of the data, making it ideal for visualizing clusters.



The visualization clearly shows separation between clusters obtained through K-Means, informing the model's ability to identify meaningful market regimes.

4. Conclusion

Successfully created the Regime Detection ML model via unsupervised learning which has a clustering quality of 0.22 and gives insights about market regimes (Mean-reverting, stable, liquid, etc)