**1. Introduction**

In this project, I used clustering techniques to identify and understand different market regimes in financial time series data. By analyzing features like volatility, bid-ask spread, and order flow, I aimed to group market conditions into distinct regimes and explore how these regimes transition over time. This can help traders and analysts anticipate market behavior.

**2. Custom Features**

To identify market regimes, I created a few custom features:

* **Volatility**: Measures how much prices fluctuate over time.
* **Bid-Ask Spread**: Reflects market liquidity by showing the difference between the highest bid and the lowest ask price.
* **Cumulative Bid and Ask Quantities**: Indicates the overall market sentiment and order flow.
* **Log Return**: Shows price changes over time, helping to understand price movements.
* **Microprice**: Represents the fair price of an asset at any given moment.

These features helped in capturing important aspects of the market behavior, like liquidity and volatility.

**3. Clustering Techniques and Metrics**

I used three clustering methods:

* **K-Means**: Groups data into K clusters based on feature similarity. It helped me understand the basic grouping of market states.
* **Gaussian Mixture Models (GMM)**: A more flexible model that assumes data is made of multiple Gaussian distributions. It provided a better understanding of overlapping regimes.
* **HDBSCAN**: This density-based method automatically finds clusters and noise, allowing for more nuanced regime detection.

To evaluate the clusters, I used metrics like the **Silhouette Score** and **Davies-Bouldin Index** to assess how well the clusters were formed.

**4. Clustering Results**

The clustering identified three main market regimes:

1. **Regime 0**: "Trending & Liquidity & Stable" – This is when the market shows a clear trend with stable liquidity.
2. **Regime 1**: "Mean Reverting & Illiquid & Volatile" – This regime is marked by volatility and low liquidity, with prices often reverting to the mean.
3. **Regime 2**: "Trending & Illiquid & Volatile" – Here, the market is volatile with a strong trend but low liquidity.

I also analyzed how these regimes transition over time, revealing patterns like a "Trending & Liquidity & Stable" regime often shifting into "Trending & Illiquid & Volatile".

**5. Regime Insights**

* **Volatility**: Regimes with high volatility, like "Mean Reverting & Illiquid & Volatile", had much higher average volatility than stable regimes.
* **Liquidity**: The bid-ask spread was wider in low liquidity regimes, which means higher trading costs.
* **Price Movements**: "Trending" regimes had more consistent price direction, while "Mean Reverting" regimes saw more frequent reversals.

**6. Visualizations**

To understand the regimes better, I created:

* **t-SNE Plot**: This 2D visualization showed how different market conditions are grouped.
* **Transition Matrix**: A heatmap that shows the likelihood of one regime following another.
* **Regime Evolution**: A timeline plot that showed how the regimes change over time, overlaid with price charts.

All graphs and figures are available in the notebook, and I encourage you to refer to them for detailed visualizations and insights.

**7. Conclusion**

This project was a deep dive into understanding how market conditions change over time. By using clustering techniques like GMM and HDBSCAN, I was able to uncover distinct market regimes and their transitions. The insights into volatility, liquidity, and price movement directionality can help traders make better decisions. I’m excited to continue exploring these patterns and applying them to real-world trading strategies!

### **A Personal Note**

I don’t know if anyone will actually read this part — but I truly hope you do. I just want to say thank you for the opportunity to work on this project.

Believe it or not, I started learning machine learning with one goal in mind: to apply it in high-frequency trading. Ever since I attended my first AI Club lecture in college, something just clicked. I’ve been fascinated by markets and algorithms ever since. I even began working on a similar project on my own — spending over a month trying to gather order book data. It was frustrating. Good data was nearly impossible to find, and the datasets I did manage to get were full of complex terminology I couldn’t make sense of at the time.

I reached out to professors, especially from IIMs, hoping for guidance or an internship — but I never got a reply. Eventually, I had to shift my focus to more “reachable” projects just to build a portfolio and apply somewhere. But even then, every time I opened the folder containing that unfinished HFT project, it sparked the same burning desire to give it another shot.

That’s why this project meant so much to me.

Getting access to clean, structured market data, along with a step-by-step guide to explore it — it was like someone handed me the very thing I’d been chasing for so long. I’ve tried to make the most of it, and even though I have my practical exams from tomorrow (bad timing, I know), I still gave it my best. With more time, I would’ve loved to polish everything further and experiment even more.

Still, I’m genuinely grateful. I’ve learned more in this short time than I ever imagined I could. And I truly believe that, with some guidance and direction, I can deliver much more — and much better.

Thank you again for this chance. It wasn’t just a project to me — it was something I’d dreamed of doing.