Final Project

Clustering Analysis on Gold Price (XAU/USD) Using K-Means

Arranged by:
Group 4

Salma Kamilah Rahma
1501224344

Akmal Muhammad Firdaus
1501224373

Resti Fresard
1501224384

Karina Ditya Amanda
1501224390

MD Touhidul Islam Kanon
1501224394

I INTRODUCTION

This project applies unsupervised machine learning using K-Means clustering to segment hourly gold price (XAU/USD) data. The goal is to uncover patterns in market behavior based on price movement, volatility, trading volume, and technical indicators.

Objective:
To explore and cluster hourly gold price data (XAU/USD) from May 1–19, 2025, using unsupervised machine learning (K-Means).

Tools Used:

- Python (pandas, matplotlib, seaborn, sklearn)
- Jupyter Notebooks

II DATASET DESCRIPTION

This dataset provides hourly historical price data for XAU/USD (Gold vs. US Dollar) from May 1, 2025, to May 19, 2025.

Each row represents a single hour and includes:
- Date and Time
- Open, High, Low, Close (OHLC) prices
- Tick Volume (Number of price changes)
- Spread

Our Dataset Link :

Dataset XAUUSD_Hourly.csv

Resulting File

Cleaned dataset saved as:
XAUUSD_Hourly_Cleaned.csv

III DATA CLEANING

1 Checked for Missing Values

To make sure there are no gaps or null values in important columns.
- df.isnull().sum() → Result: No missing values

2 Removed Irrelevant & Redundant Columns

To simplify the dataset and keep only useful information.
- Volume column: always 0 → no useful data → removed.
- Date and Time columns: already combined into datetime → now redundant → removed.

3 Datetime Standardization

To ensure the datetime column is in a usable format for time-based analysis.
- pd.to_datetime() to convert the datetime column.

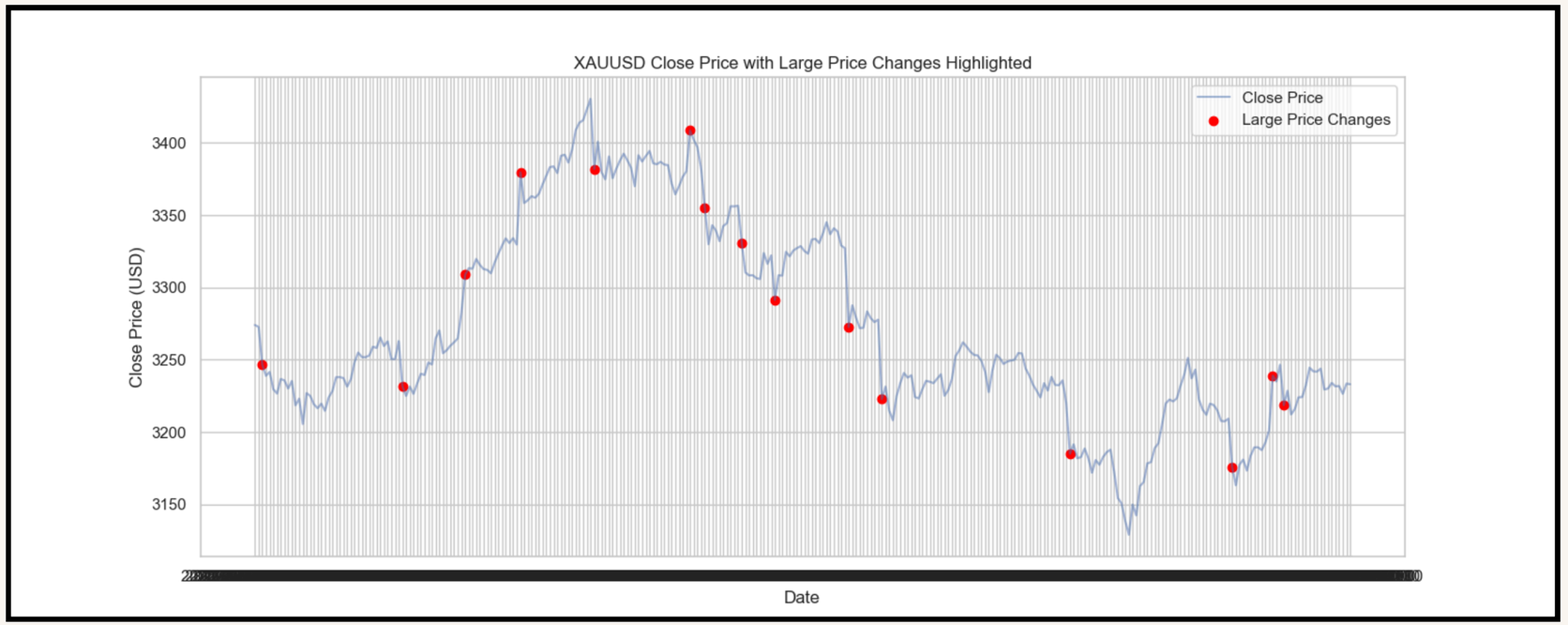
4 Spread Standardization

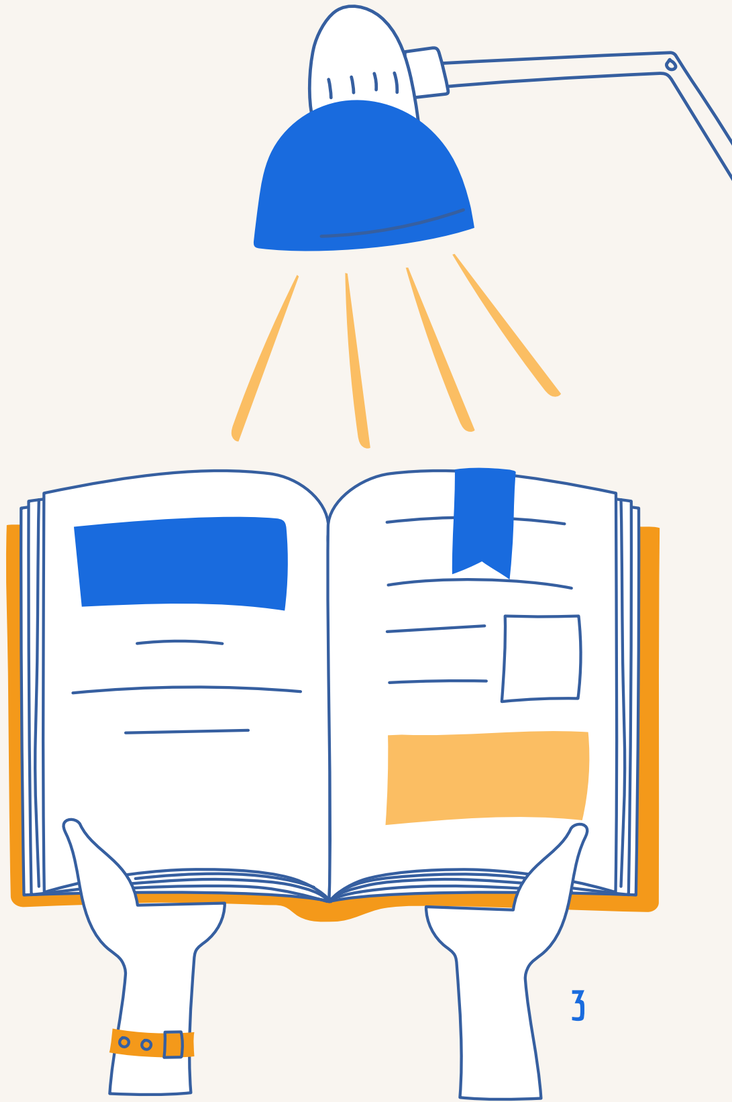
To make sure trading conditions in the dataset are consistent.
- The Spread column mostly had value 7, but some values were 6 or 8.
- To keep consistency, all spread values were replaced with the mode = 7.

5 Duplicate & Outlier Check

To remove any repeated or extreme data that could skew analysis.
- Used drop_duplicates() → No duplicates
- Outlier check using IQR → No extreme outliers found in price

IV EXPLORATORY DATA ANALYSIS (EDA)



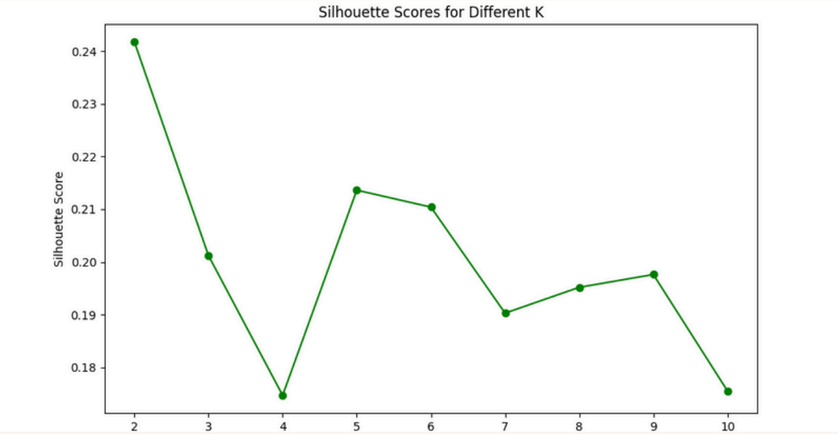


The Exploratory Data Analysis revealed key patterns in price behavior, volatility, trading activity, and technical indicators, offering insights into market trends and potential predictive signals.

- **Avg. close price: $3271.96 with moderate volatility**
- **High correlation across price columns**
- **15 major price swings linked to volatility peaks**
- **RSI and SMA indicate potential trend changes**

V UNSUPERVISED LEARNING – K-MEANS CLUSTERING

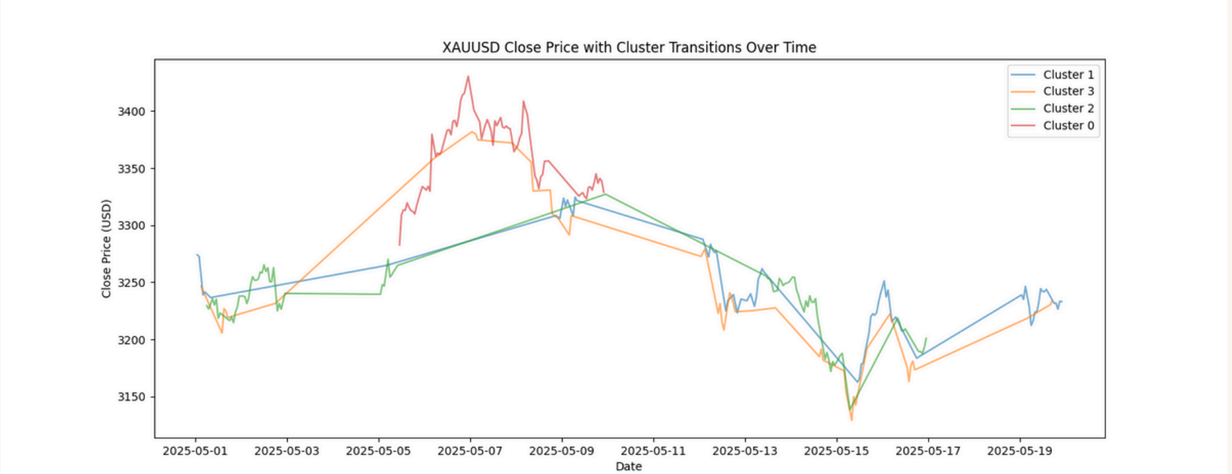
SILHOUETTE SCORE PLOT
BEST CLUSTERING QUALITY AT K = 2
WE USED K = 4 TO CAPTURE RICHER MARKET REGIMES.



CLUSTER	CHARACTERISTICS
0	High Price, High RSI → Bullish Trend
1	Mid Price, High Return, Mixed Momentum
2	Low Price, Low Volatility → Stable
3	High Volatility, Low RSI → Bearish/Correction Periods

Features Used:

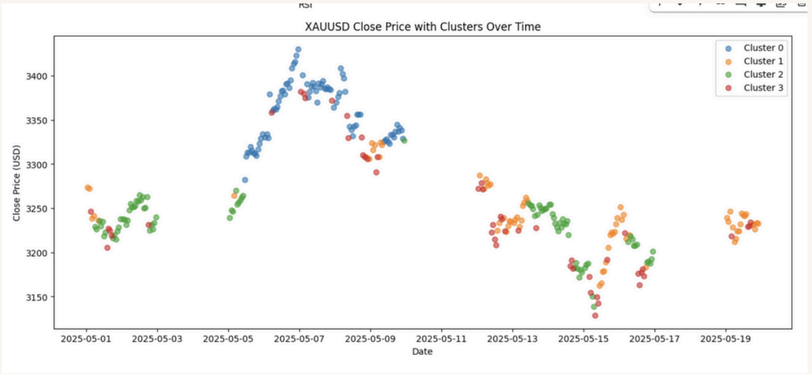
- **Close, Returns, Volatility, High_Low_Range, TickVolume, RSI**
- **All features scaled with StandardScaler.**

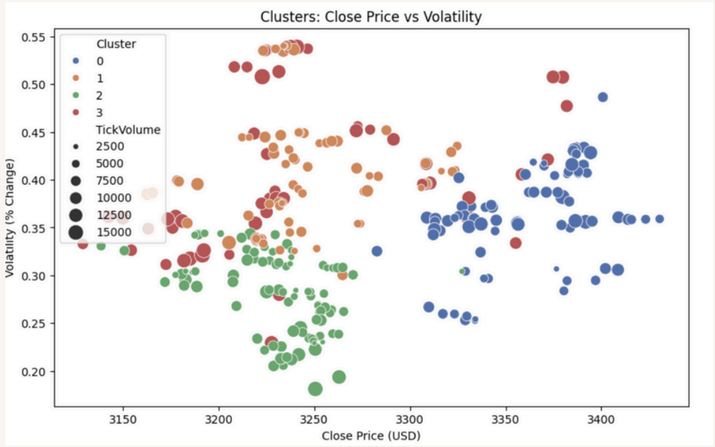


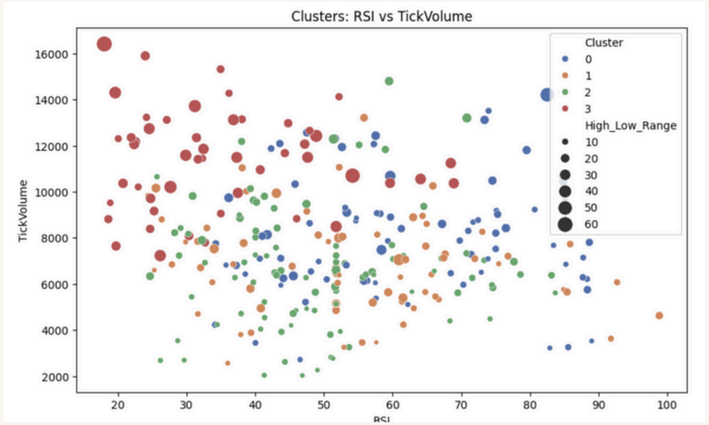
- Cluster Transition Plot (cluster_transitions_timeseries.png):
- Shows how the market transitioned between clusters over time.
- Clear periods of bullish (Cluster 0) and bearish (Cluster 3) trends.

VI CLUSTER SUMMARY

Cluster	Price Level	Volatility	RSI	Activity
0	High Price	Moderate	Bullish	High
1	High Price	Moderate	Bullish	High
2	High Price	Moderate	Bullish	High
3	High Price	Moderate	Bullish	High







- **CLUSTER 0:** HIGHEST PRICES, MODERATE VOLATILITY, STRONG TRADING ACTIVITY, BULLISH TREND.
- **CLUSTER 1:** MID PRICES, HIGHEST RETURNS, HIGHER VOLATILITY, MIXED MOMENTUM.
- **CLUSTER 2:** LOWER PRICES, STABLE RETURNS, LOWEST VOLATILITY, NEUTRAL MARKET.
- **CLUSTER 3:** WIDE PRICE RANGE, NEGATIVE RETURNS, HIGH VOLATILITY, BEARISH TREND.

VII BUSINESS / TRADING INSIGHT

We identified 4 distinct market clusters using unsupervised learning:

Cluster	Market Behavior	Suggested Strategy
0	High price, bullish, high RSI	Trend-following (Buy & Hold)
1	Mid price, high return, volatile	Short-term Momentum Trading
2	Low volatility, stable price	Low-risk or Hold strategy
3	High volatility, bearish, low RSI	Mean Reversion (Buy Low, Sell Fast)

Note: *Each cluster gives insight into the best action traders can take depending on market behavior.*

Key Insights & Real-World Use

Content:

📊 Indicators:

- RSI > 70 → Overbought → Sell
- RSI < 30 → Oversold → Buy
- High Volume & Volatility → Market reacting to news/events

🏦 Business Use:

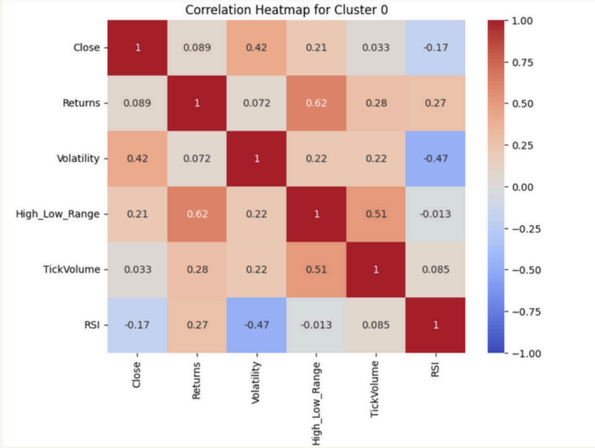
- Fast market classification
- Supports trading alerts & strategies
- Useful for fintech dashboards & trading bots

VIII CORRELATION ANALYSIS

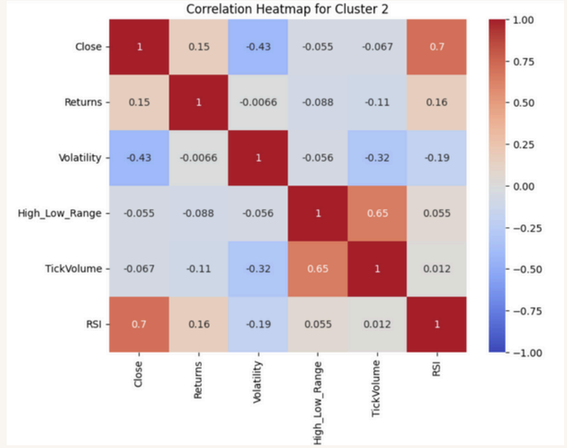
THE CORRELATION ANALYSIS REVEALED KEY RELATIONSHIPS BETWEEN MARKET INDICATORS, BOTH OVERALL AND WITHIN SPECIFIC CLUSTERS.

- **CLOSE AND RSI SHOW A MODERATE POSITIVE CORRELATION, LINKING PRICE WITH MOMENTUM**
- **TICKVOLUME STRONGLY CORRELATES WITH HIGH_LOW_RANGE, SUGGESTING ACTIVE MARKETS WIDEN SPREADS**
- **RETURNS AND VOLATILITY SHOW WEAK NEGATIVE CORRELATION**
- **CLUSTER HEATMAPS HIGHLIGHT UNIQUE INTRA-CLUSTER DYNAMICS FOR DEEPER INSIGHTS**

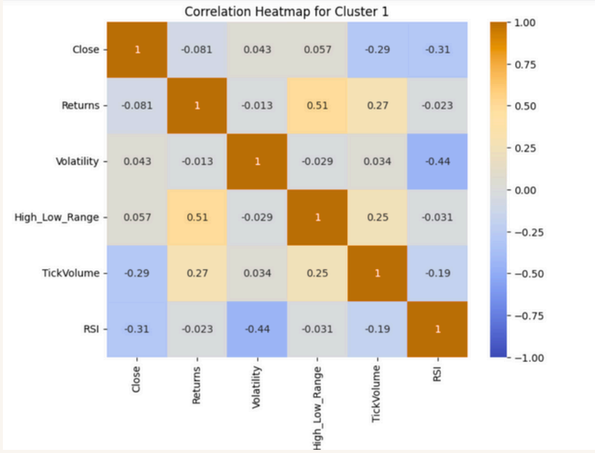
Cluster 0



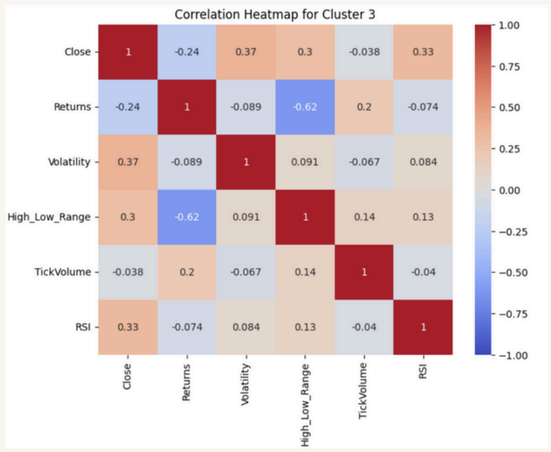
Cluster 2



Cluster 1



Cluster 3



VIII CONCLUSION & NEXT STEPS 🐸

This project showed that K-Means clustering is effective in analyzing hourly gold price data by identifying four distinct market patterns: bullish, bearish, stable, and volatile. The combination of data cleaning, EDA, and unsupervised learning helped simplify complex financial data and reveal valuable insights into price behavior and market momentum.

Next Steps:

- Extend the dataset for broader, long-term analysis
- Use Elbow and Silhouette methods to refine cluster accuracy
- Try other clustering methods (e.g., DBSCAN, HDBSCAN) for deeper insights
- Explore supervised models to predict market trends from cluster features

Welcome for any questions and

Thank you for Attention

Salma Kamilah Rahma
1501224344

Akmal Muhammad Firdaus
1501224373

Resti Fresard
1501224384

Karina Ditya Amanda
1501224390

MD Touhidul Islam Kanon
1501224394