Workflow of clustering

1.Data Import:
The notebook imports necessary libraries and reads three datasets: customers, transactions, and products.

2.RFM Calculation:
The function calculate_rfm computes Recency, Frequency, and Monetary (RFM) metrics for each customer based on their transaction history.
Recency: Days since the last transaction.
Frequency: Total number of transactions.
Monetary: Total spending.

3.Customer Metrics Calculation:
The function calculate_customer_metrics calculates additional metrics such as average order value and unique products purchased by each customer.
4.Clustering:
The function perform_clustering standardizes the features and applies KMeans clustering to segment customers into specified clusters.
It evaluates cluster quality using the Davies-Bouldin index and silhouette score.

5.Optimal Clusters:
A range of cluster numbers is tested (from 2 to 12), and the optimal number is determined based on the lowest Davies-Bouldin index.

6.Visualization:
The notebook generates plots to visualize the relationship between the number of clusters and clustering metrics.
Box plots display RFM metrics by cluster to illustrate differences among customer segments.

7.Results Summary:
Finally, it prints out clustering results, including optimal cluster count and profiles summarizing average RFM metrics for each segment.
Insights from the Analysis
8.Customer Segmentation: Identifying distinct customer segments allows for targeted marketing strategies.

9.Behavioral Patterns: Analyzing RFM metrics helps understand purchasing behavior, such as identifying high-value customers or those at risk of churn.

10.Data-Driven Decisions: Insights from this analysis guide inventory management, promotional campaigns, and customer relationship strategies.