CUSTOMER SEGMENTATION ANALYSIS REPORT

INTRODUCTION

This report summarizes the results of a customer segmentation analysis performed using the K-Means clustering algorithm. The analysis aimed to group customers into distinct segments based on their purchasing behavior and demographics.

METHODOLOGY:

1. **Data Preparation:** Customer and transaction data were merged, preprocessed, and relevant features were selected (Region, Product Frequency, and Average Price).
2. **Normalization:** Numerical features were normalized using StandardScaler to ensure equal weighting during clustering.
3. **Cluster Determination:** The Elbow Method was employed to determine the optimal number of clusters, which was found to be 4.
4. **Clustering:** The K-Means algorithm was applied with 4 clusters to group the customers.
5. **Evaluation:** Clustering performance was assessed using the Davies-Bouldin Index and Silhouette Score.

OVERVIEW

The customer base was segmented into **4 distinct clusters**, each representing unique buying behaviours based on product frequency, regional patterns, and average price.

Key Clustering Insights

1. Number of Clusters Formed:

- **4 Clusters**
- Each cluster reflects a unique set of attributes, such as product purchase frequency, regional preferences, and pricing behaviour.

2. Davies-Bouldin Index (DBI):

- **Value**: **1.1034530607074169**
- Interpretation: A DBI value closer to 1 indicates well-separated and compact clusters, suggesting an effective segmentation approach.

3. Clustering Metrics:

- **Silhouette Score**: **0.26672523790117947**

- **Cluster Sizes**:
    - Cluster 0 (Purple): Represents cost-conscious buyers with low frequency and average price.
    - Cluster 1 (Yellow): High-frequency buyers with moderate average prices, spanning multiple regions.
    - Cluster 2 (Blue): Moderate frequency buyers focused on premium goods.
    - Cluster 3 (Green): Regional-specific buyers, often from Europe, with targeted price and product habits.

DETAILED CLUSTER ANALYSIS

1. **Cluster 0 (Purple):**
    - **Characteristics**: Cost-conscious buyers.
    - **Product Frequency**: Low (1–3 products).
    - **Average Price**: Budget-friendly (<100).
    - **Regional Distribution**: North America and South America.
    - **Implication**: Customers may prefer economical options and infrequent purchases.
2. **Cluster 1 (Yellow):**
    - **Characteristics**: High-frequency buyers.
    - **Product Frequency**: High (8–10 products).
    - **Average Price**: Moderate.
    - **Regional Distribution**: Spans multiple regions, including Asia and Europe.
    - **Implication**: These are likely loyal or high-value customers who may respond well to loyalty programs.
3. **Cluster 2 (Blue):**
    - **Characteristics**: Premium buyers.
    - **Product Frequency**: Moderate.
    - **Average Price**: High (>300).
    - **Regional Distribution**: Concentrated in specific high-purchasing-power areas.
    - **Implication**: Focused marketing on luxury and exclusive products.
4. **Cluster 3 (Green):**
    - **Characteristics**: Regional-specific buyers.
    - **Product Frequency**: Varies.
    - **Average Price**: Targeted pricing.
    - **Regional Distribution**: Primarily Europe.
    - **Implication**: Campaigns should focus on localized preferences and offerings.

REGIONAL PATTERNS

1. **Asia and Europe**:
    - Dominate certain clusters, reflecting similarities in buying patterns or pricing trends.
2. **North America and South America**:
    - Form distinct clusters, likely driven by regional preferences and economic factors.

PRODUCT FREQUENCY TRENDS

- **High Frequency** (8–10 products): Indicates loyal or frequent buyers.
- **Low Frequency** (1–3 products): Suggests occasional or budget-conscious buyers.

AVERAGE PRICE BEHAVIOUR

- **Low Average Price (<100)**: Cost-sensitive markets.
- **High Average Price (>300)**: Premium or luxury market segments.

MARKETING STRATEGY RECOMMENDATIONS

1. **Cluster 0 (Purple)**:
    - Focus on budget-friendly options.
    - Offer discounts and value-for-money promotions.
2. **Cluster 1 (Yellow)**:
    - Promote mid-range products.
    - Enhance loyalty programs to retain frequent buyers.
3. **Cluster 2 (Blue)**:
    - Emphasize premium or exclusive product lines.
    - Highlight luxury branding.
4. **Cluster 3 (Green)**:
    - Tailor localized campaigns.
    - Address specific regional preferences.

OTHER RELEVANT CLUSTERING METRICS:

- **Within-Cluster Sum of Squares (WCSS):** WCSS was used in the Elbow Method to find the optimal number of clusters. It measures the total variance within each cluster. The Elbow Method helps to identify the point where increasing the number of clusters no longer significantly reduces WCSS.

- **Visualization:** 2D and 3D scatter plots were generated to visualize the clusters and their distribution based on the selected features. These visualizations provide insights into the characteristics of each customer segment.

CONCLUSION

The clustering analysis has effectively segmented customers into meaningful groups, enabling targeted marketing strategies and personalized approaches. The low DBI value confirms the quality and distinctiveness of the clusters, ensuring reliable insights for decision-making.