<u>Customer Clustering Analysis Report</u>

1. Overview:

The clustering analysis was performed to segment customers based on their transaction and profile information. The goal was to group customers with similar behaviours and characteristics into distinct clusters, enabling targeted marketing strategies and personalized offerings.

2. Methodology

1. Data Preparation:
    - Transactional attributes (e.g., total spend, total quantity purchased, number of transactions) were aggregated for each customer.
    - Profile attributes (e.g., region, signup days) were included to incorporate demographic information.
    - Categorical variables like region were encoded using one-hot encoding, and numerical features were standardized to ensure fair clustering.
2. Clustering Algorithm:
    - K-Means Clustering: The algorithm was chosen for its efficiency and interpretability. Multiple cluster sizes (from 2 to 10) were tested to identify the optimal number of clusters.
    - Davies-Bouldin Index (DB Index): This metric was used to evaluate the quality of clustering. A lower DB Index indicates better-defined clusters.
3. Dimensionality Reduction:
    - Principal Component Analysis (PCA) was used to reduce the feature dimensions to 2 for visualization while retaining the maximum variance.

3. Results:

1. Number of Clusters:

    - The optimal number of clusters determined using the DB Index was 4.
2. Davies-Bouldin Index:
    - The DB Index for the optimal clustering was 0.84. This value reflects well-separated and compact clusters.
3. Cluster Characteristics:
    - Each cluster was analyzed based on the mean values of the input features:
        - Cluster 0: High spenders with frequent transactions, mostly from specific regions.
        - Cluster 1: Moderate spenders with diverse product interests.
        - Cluster 2: Low-frequency shoppers with lower transaction values.
        - Cluster 3: Customers with above-average spending but fewer unique product categories.
4. Other Metrics:
    - Inertia (Within-Cluster Sum of Squares): This measure confirmed that increasing the number of clusters beyond 4 resulted in diminishing returns.
    - Silhouette Analysis: A silhouette analysis could be conducted in future iterations to provide further validation.

4. Visualization:

The clustering results were visualized using a PCA scatter plot:

Customers were plotted on two principal components (PCA1 and PCA2).

Distinct clusters were clearly visible, with minimal overlap between groups, indicating well-separated clusters.

5. Implications:

1. **Targeted Marketing**:
   - High-value clusters (e.g., Cluster 0) can be targeted with loyalty programs or premium offerings.
   - Low-frequency shoppers (e.g., Cluster 2) could benefit from promotional campaigns to increase engagement.
2. **Product Recommendations**:
   - Cluster-specific preferences can guide personalized product recommendations.
3. **Customer Retention**:
   - Understanding the characteristics of high-spending clusters can help retain valuable customers through focused engagement strategies.

6. Conclusion:

The clustering analysis successfully segmented the customer base into 4 distinct clusters, each with unique profiles and behaviours. The use of the Davies-Bouldin Index ensured high-quality clustering, and the results can directly inform business decisions related to marketing, inventory management, and customer experience optimization.

Next Steps:

- Further refine clustering by incorporating additional features like product-level preferences or time-based behaviours.
- Validate clusters using additional metrics like silhouette scores.
- Deploy the clustering results into a recommendation system or CRM for actionable insights.