Report on Clustering Results

INTRODUCTION:

The goal of this clustering task was to segment customers into meaningful groups using transactional and demographic data. Features such as spending behaviour, purchase recency, product diversity, and seasonality metrics were engineered to enrich the dataset.

NUMBER OF CLUSTERS FORMED:

- Optimal Number of Clusters: 5
- Method Used to Determine Clusters:

1. Elbow Method (Inertia)
2. Silhouette Score
3. Davies-Bouldin Index (DB Index)

CLUSTERING EVALUATION METRICS:

Davies-Bouldin Index

- **DB Index Value**: **1.0703**
  A lower DB Index indicates better clustering quality. The observed value of **1.0703** reflects reasonably well-separated and compact clusters.

Silhouette Score

- **Silhouette Score**: **0.2710**
  This score indicates moderately separated clusters. While it shows room for improvement, the clusters are distinguishable.

Calinski-Harabasz Index

- **Calinski-Harabasz Index**: **113.5291**
  A higher Calinski-Harabasz Index suggests well-defined clusters. The value of **113.5291** reflects some degree of separation.

CLUSTER PROFILES:

Cluster	Avg Spending	Avg Recency(days)	Avg CLTV	Churn Risk(%)	Cluster Size
0	$6,357.03	43.45	$1,318.87	0%	20
1	$2,532.07	77.50	$187.18	0%	71
2	$997.03	218.84	$14.12	8%	25
3	$3,559.83	130.84	$122.42	0%	38
4	$4,952.86	87.78	$448.42	0%	45

CLUSTER INSIGHTS:

Cluster 0:

- **Description**: High-value customers with **low recency** (recent purchases) and **high CLTV**. These are loyal and active customers.
- **Strategy**: Reward them with personalized offers or premium loyalty programs.

Cluster 1:

- **Description**: Moderate spenders with **moderate recency** and **low CLTV**. They form a large portion of the customer base but are not as engaged.
- **Strategy**: Offer targeted promotions to increase frequency and spending.

Cluster 2:

- **Description**: Low spenders with **high recency** (inactive customers) and **very low CLTV**. Some churn risk is present (**8%**).
- **Strategy**: Focus on retention campaigns, discounts, and personalized re-engagement strategies.

Cluster 3:

- **Description**: Moderate-to-high spenders with **higher recency** and **low CLTV**. These customers may have infrequent but higher-value purchases.
- **Strategy**: Use seasonal campaigns or incentives to make their purchases more frequent.

Cluster 4:

- **Description**: High spenders with **moderate recency** and **medium CLTV**. These customers are valuable but may need some attention to retain.
- **Strategy**: Provide exclusive offers to sustain spending and loyalty.

VISUALIZATIONS:

Elbow Curve

The elbow curve helped identify the optimal number of clusters.

PCA Visualization

Clusters visualized in 2D using Principal Component Analysis (PCA).

t-SNE Visualization

Clusters represented using non-linear t-SNE dimensionality reduction.

CONCLUSION:

The clustering process successfully segmented customers into **5 meaningful groups**.

- The **DB Index value (1.0703)** confirms reasonably good cluster quality.
- Additional metrics, including **Silhouette Score (0.2710)** and **Calinski-Harabasz Index (113.5291)**, further validate the clustering results.

RECOMMENDATIONS:

1. **Focus on Cluster 0** (high-value, loyal customers) by implementing VIP loyalty programs and personalized services.
2. **Address Cluster 2** (at-risk customers) with re-engagement campaigns and discounts.
3. **Leverage Cluster 3** with seasonal campaigns to boost repeat purchases.
4. **Maintain Cluster 4** customers by providing exclusive offers to retain their spending levels.