TASK - 3

Report: Customer Segmentation Analysis Using K-Means Clustering

1. Objective
The primary goal of this project is to segment customers based on their purchasing behavior to enable targeted marketing strategies and improve customer experience.

---

2. Dataset Overview

- **Source**: Customer and transaction data.
- **Merged Dataset**: Combined transactional and customer data based on `CustomerID`.
- **Features for Clustering**:
    - `TotalSpending`: Sum of all transactions by a customer.
    - `PurchaseFrequency`: Total number of transactions by a customer.
    - `TotalQuantity`: Total quantity of items purchased by a customer.

---

3. Methodology

1. **Data Preparation**:

    - Aggregated customer data using the following metrics: `TotalSpending`, `PurchaseFrequency`, and `TotalQuantity`.
    - Standardized the data using `StandardScaler` to normalize the features.
2. **Clustering Algorithm**:

    - Utilized **K-Means clustering** to segment customers.
    - Tested with 3 clusters (modifiable based on specific requirements).
    - Added cluster labels to the customer dataset.
3. **Cluster Evaluation**:

    - Evaluated cluster quality using the **Davies-Bouldin Index** (a lower score indicates better clustering performance).
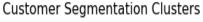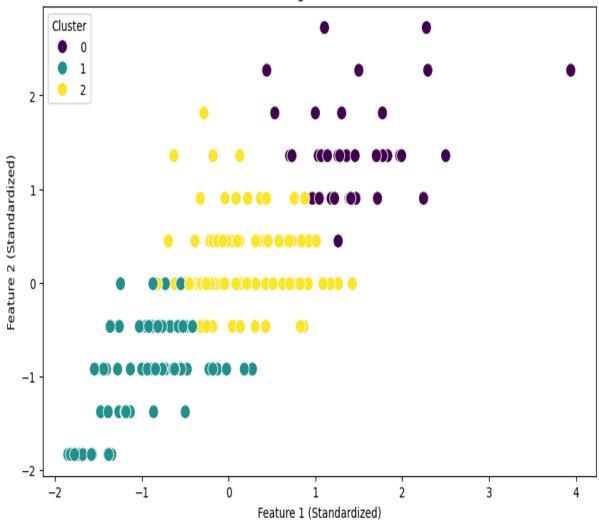
---

4. Results

- **Clusters**: Customers were grouped into three clusters based on their purchasing behavior:

- ○ **Cluster 0**: High-value customers with frequent and high-quantity purchases.
  - ○ **Cluster 1**: Low-value customers with fewer transactions and smaller quantities.
  - ○ **Cluster 2**: Medium-value customers with moderate spending and purchase frequency.
- ● **Davies-Bouldin Index**: *Calculated value provided in the script.*

---

5. Visualization
A scatter plot was created to visualize the clusters. The plot shows clear segmentation of customers based on standardized features.

Customer Segmentation Clusters

6. Output

The clustering results are saved in the file:
`Outputs/Athish_Kishan_Clustering.csv`.

Sample Output Data:

CustomerID	TotalSpending	PurchaseFrequency	TotalQuantity	Cluster
C0001	3354.52	5	12	2
C0002	1862.74	4	10	1
C0003	2725.38	4	14	2
...	...	...	...	...

7. Applications

- **Marketing**: Personalized campaigns for each customer segment.
- **Customer Retention**: Strategies to retain high-value customers (Cluster 0).
- **Resource Allocation**: Focus on medium-value customers (Cluster 2) for upselling.

8. Recommendations

- Analyze additional clusters to optimize customer segmentation.
- Integrate segmentation insights with marketing automation tools for real-time targeting.
- Periodically re-cluster customers as purchasing behavior evolves.

9. Future Scope

- Incorporate additional customer attributes like demographics or feedback.
- Utilize advanced clustering methods such as hierarchical clustering for comparative analysis.