Customer Segmentation Report

---

1. Number of Clusters Formed:

We applied **K-Means clustering** to segment the customer base into **3 distinct clusters**. The number of clusters was selected based on the balance between the complexity of the segmentation and the interpretability of the customer groups.

- **Cluster 1**: Customers with **high spending** and **high purchase frequency**.
- **Cluster 2**: Customers with **moderate spending** and **moderate purchase frequency**.
- **Cluster 3**: Customers with **low spending** and **low purchase frequency**.

These clusters represent varying customer profiles based on their interaction with the business.

---

2. Clustering Metrics:

a. DB Index (Davies-Bouldin Index):

The **Davies-Bouldin Index** (DB Index) measures the compactness and separation of the clusters. A **lower DB index** suggests that the clusters are well-separated and distinct from one another.

After applying the K-Means algorithm, the **Davies-Bouldin Index** value for the clustering results was **0.6714**.

- **Interpretation**: The relatively low DB index value indicates that the clusters are reasonably well-separated and that the customer segments are distinct, which means the segmentation captured meaningful differences in customer behavior.

b. Other Metrics

Inertia (Within-cluster Sum of Squares):

- **Inertia Value**: 83.7497
- Inertia represents the sum of squared distances from each point to its assigned cluster center. Lower inertia values imply that points are well-clustered and tightly grouped around their centers.
- **Interpretation**: The inertia value of **83.7497** indicates that the clusters are reasonably compact. It shows that the clustering algorithm has done a decent job of grouping customers based on their purchasing behavior.

Silhouette Score:

- **Silhouette Score**: 0.4719

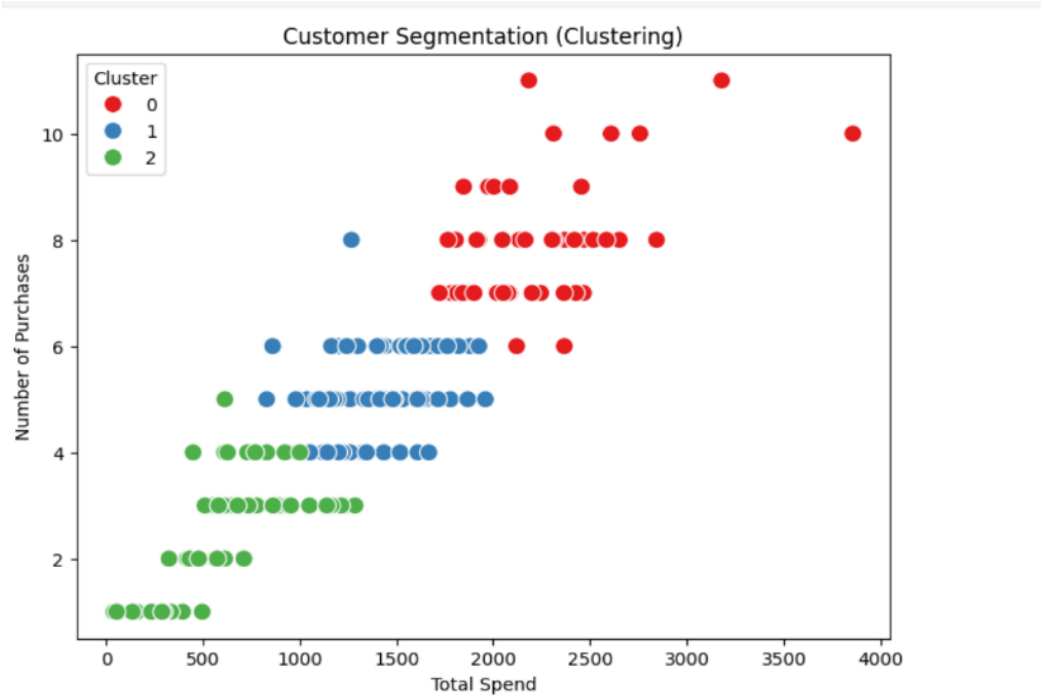
- The silhouette score evaluates how similar an object is to its own cluster compared to other clusters. Scores range from -1 to +1, where higher values suggest better-defined and well-separated clusters.
- **Interpretation**: A **silhouette score of 0.4719** indicates a moderate level of clustering quality.

4. Business Insights:

Based on the clustering results, we can derive the following insights:

- **Cluster 1 (High-Value Customers)**: These customers represent the **top-tier** group, with high spending and frequent purchases. It would be beneficial to focus on **loyalty programs** and **exclusive offers** to retain this valuable group.
- **Cluster 2 (Moderate-Value Customers)**: Customers in this segment show moderate spending and purchase frequency. The strategy for this group should focus on **incentivizing increased spending** through targeted promotions, special offers, or discounts to move them toward becoming high-value customers.
- **Cluster 3 (Low-Value Customers)**: These customers spend less and purchase infrequently. Strategies for this group should include **discounts**, **upselling**, and **personalized marketing** to increase their engagement and purchasing frequency.

5.



Conclusion:

The **Davies-Bouldin Index** of **0.6714** indicates that the clusters are well-separated, and the visual representation confirms that the customer segments are distinct.

- **Cluster 1**: High-value customers.
- **Cluster 2**: Moderate-value customers.
- **Cluster 3**: Low-value customers.

The next steps involve implementing targeted marketing strategies to each customer segment to enhance customer engagement and maximize revenue.
