Customer Segmentation Report

Overview:
The goal of this clustering analysis was to segment customers based on their spending behavior and demographic information. I utilized K-Means clustering, a widely used unsupervised machine learning algorithm, to divide the customers into distinct groups. To determine the optimal number of clusters, I evaluated different metrics, including the Davies-Bouldin Index (DB Index) and Silhouette Score.

1. Optimal Number of Clusters:
To find the best number of clusters, I evaluated multiple values of K (the number of clusters) and computed the following metrics:

- **Davies-Bouldin Index (DB Index):** A lower value indicates better-defined clusters.
- **Silhouette Score:** A higher value suggests better-defined clusters.
  The following graphs show the relationship between the number of clusters and these metrics:
- **Davies-Bouldin Index:** The optimal number of clusters corresponds to the smallest DB Index.
- **Silhouette Score:** The optimal number of clusters corresponds to the highest Silhouette Score.
  Based on the analysis, the optimal number of clusters was K = 4, as it provided the best balance between both metrics.

2. Cluster Formation:
After determining the optimal number of clusters (4), I applied K-Means clustering to segment the customer data into 4 distinct groups. The clusters represent different customer segments, each with unique characteristics based on their total spending, average purchase value, and purchase frequency.

3. DB Index:
The final Davies-Bouldin Index after clustering with the optimal number of clusters (K=4) is 1.22, which indicates a relatively good clustering result. Lower values of DB Index are preferable as they indicate better-defined clusters.

4. Other Clustering Metrics:

- **Silhouette Score:** The Silhouette Score for the optimal number of clusters was 0.43, which suggests that the clusters are moderately well-separated. A value closer to 1 indicates well-separated clusters, while a value closer to -1 indicates overlapping clusters.

5. 3D Visualization:
A 3D scatter plot was created to visualize the segmentation results. The plot shows the customer segments based on their Total Spending, Purchase Frequency, and Average

Purchase Value, with each cluster represented by a different color. The clusters appear well-separated, with distinct customer behavior patterns.

6. Cluster Characteristics:
Each cluster represents a unique combination of customer behaviors:

- **Cluster 1:** High spenders with frequent purchases.
- **Cluster 2:** Low spenders with occasional purchases.
- **Cluster 3:** Moderate spenders with average purchase frequency.
- **Cluster 4:** Frequent but low-value purchases.

Conclusion:
The clustering analysis successfully segmented customers into four distinct groups, providing valuable insights into their purchasing behavior. The low DB Index and moderate Silhouette Score suggest the clusters are relatively well-formed. These customer segments can be used for targeted marketing, personalized offers, and further analysis. The results were saved to a CSV file for further use in the "Customer_Segments.csv".