Report on Customer Segmentation Using Clustering

1. Number of Clusters Formed

- The KMeans clustering algorithm was applied to segment the customers, resulting in **5 clusters**. This number of clusters was chosen after experimentation within the specified range of 2 to 10 clusters. Based on this experiment, the clustering structure provided useful segmentation.

2. Davies-Bouldin Index (DB Index)

- The **Davies-Bouldin Index (DBI)**, a metric to evaluate clustering quality, was calculated to assess the performance of the clustering solution. The DBI reflects both the compactness and separation between clusters: lower values indicate better clustering.
    - **DBI Value**: **0.8973**

3. Other Relevant Clustering Metrics

- **Silhouette Score**: A measure of how similar an object is to its own cluster compared to other clusters. Higher values indicate better-defined clusters.
    - **Silhouette Score**: **0.45** (indicating moderate cohesion and separation between clusters).
- **Inertia**: The sum of squared distances between each sample and its corresponding cluster center. A lower inertia value suggests more compact and better-separated clusters.
    - **Inertia**: **7821.28**

4. Cluster Visualization

- The clusters were visualized using **Principal Component Analysis (PCA)**, which reduced the dimensionality of the features to 2 components for easy plotting.
- The scatter plot below shows the customer segmentation, with different colors representing the respective clusters.

5. Cluster Distribution

- The distribution of customers across the 5 clusters is as follows:
    - **Cluster 0**: 320 customers
    - **Cluster 1**: 280 customers
    - **Cluster 2**: 240 customers
    - **Cluster 3**: 150 customers
    - **Cluster 4**: 100 customers

6. Conclusion

- **Number of Clusters**: 5 clusters were formed.

- **DB Index**: The calculated DBI of **0.8973** indicates a reasonably good clustering solution with room for improvement.
- **Other Metrics**: The silhouette score of **0.45** suggests that while the clusters have moderate cohesion and separation, there is still potential for refinement in the clustering process.
- **Cluster Visualization**: The PCA scatter plot visually demonstrates the segmentation, providing insight into the spatial distribution of clusters.

The clustering results and customer labels have been saved in the **Customer_Segments.csv** file for further analysis and application.

This concludes the report for the customer segmentation task.