Task 3: Customer Segmentation / Clustering

Algorithm Used

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Advantages of DBSCAN for This Case

1. **Handles Noise**: Effectively identifies and excludes noisy data points, improving cluster quality.
2. **No Need for Predefined Cluster Count**: Determines clusters dynamically without requiring a predefined number.
3. **Supports Arbitrary Shapes**: Identifies clusters with non-spherical shapes better than traditional algorithms like K-Means.

Clustering Results

- **Number of Clusters Formed**: 3
- **DB Index Value**: 0.67 (indicating good cluster separation and compactness)
- **Silhouette Score**: 0.5326 (suggesting moderately well-defined clusters)

Metrics Summary

Davies-Bouldin Index (DBI)

- Lower value (0.67) indicates better clustering with minimal overlap between clusters.

Silhouette Score

- Value of 0.5326 implies that the clusters are reasonably cohesive and well-separated.

Conclusion

The DBSCAN algorithm effectively identified three distinct clusters in the dataset, demonstrating its strength in handling noise and detecting clusters with arbitrary shapes. The clustering quality is validated by a low DB Index value and a moderately high Silhouette Score.