1. Number of Clusters Formed:

The optimal number of clusters for the customer data was determined to be **4**. This conclusion was based on the analysis of clustering metrics such as the **Davies-Bouldin Index (DB Index)** and **Silhouette Score**, where **4 clusters** showed the lowest DB Index and the highest Silhouette Score. These metrics suggest that 4 clusters balance cluster cohesion and separation effectively, making it the most optimal choice.

2. Davies-Bouldin Index (DB Index):

The **DB Index** measures the quality of clustering, where a lower value indicates better-defined clusters. For 4 clusters, the **DB Index value was 0.90**, which is the optimal value for this dataset. The graph of DB Index versus the number of clusters showed that when the number of clusters exceeds 4, the index increases, implying that additional clusters lead to poorer quality and over-segmentation.

3. Silhouette Score:

The **Silhouette Score** is a measure of how well-separated the clusters are, with a score closer to 1 indicating better clustering. The **Silhouette Score for 4 clusters** was 0.34, which aligns with the DB Index observation. Visual trends from the Silhouette Score graph also confirmed that 4 clusters provide the best balance between compactness (how tightly grouped the points are) and separation (how distinct the clusters are from each other).

4. Visual Representation of Clusters:

Several visualizations were created to illustrate the clustering results:

- **PCA Visualization**: A plot of the data reduced to two dimensions using **Principal Component Analysis (PCA)** showed that the 4 clusters were distinctly separated, validating that they are well-defined in the reduced feature space.
- **t-SNE Visualization**: A **t-SNE plot** provided another perspective, focusing on local relationships between data points. This visualization reinforced the idea that the clusters are well-separated and consistent with the PCA results.
- **Pie Chart**: A pie chart displayed the proportional distribution of customers across the 4 clusters, revealing that the customer populations varied significantly between segments, with some clusters being larger than others.

5. Insights:

- **Optimal Clustering**: The best clustering solution was found with 4 clusters, as indicated by the DB Index and Silhouette Score.
- **Distinct Customer Groups**: The visualizations from PCA and t-SNE confirmed the presence of distinct customer segments, which likely represent groups with similar behaviors or characteristics.
- **Cluster Sizes**: The pie chart highlighted an imbalance in cluster sizes, with some clusters having more customers than others. The larger clusters may represent more common customer types, while the smaller clusters could correspond to niche groups.

- **Actionable Applications**: The distinct customer segments can be used to **tailor marketing strategies**, **personalize product recommendations**, and enhance **customer service** for each group. Further analysis of each cluster's characteristics (such as demographics or spending habits) could offer deeper insights.

6. Additional Insights:

- **Cluster Characteristics**: Investigating the specific characteristics of each cluster—such as customer demographics and buying behaviour could help refine customer profiles and target them more effectively.
- **Business Applications**: Understanding the clusters allows businesses to customize their offerings and campaigns to meet the unique needs of each group, improving engagement and sales.
- **Imbalance in Cluster Sizes**: The uneven distribution across clusters may suggest that businesses should focus their efforts on the larger clusters while still addressing the niche needs of the smaller segments.

In conclusion, by using these clustering results, businesses can make informed decisions on marketing, product strategies, and customer relationship management to maximize engagement and revenue.