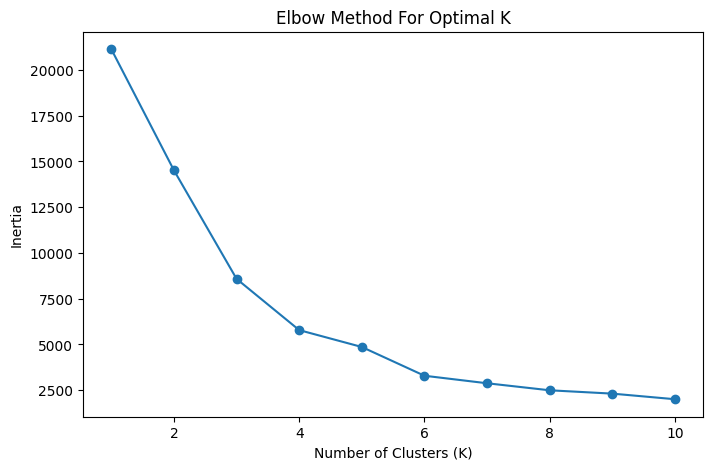
**1. Objective:**

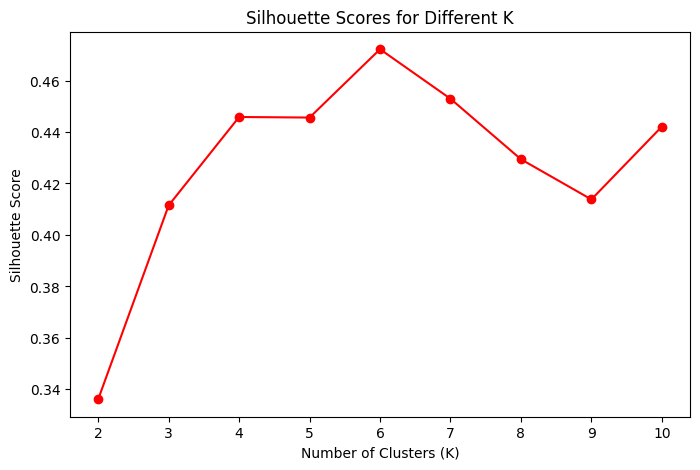
The clustering analysis was conducted to segment a set of customers into distinct groups based on their behavior or attributes using the **K-Means clustering algorithm**. The primary goal is to identify clusters of customers with similar features, allowing for personalized marketing strategies or improved customer management.

**2. Optimal Number of Clusters:**

* The **Elbow Method** was used to determine the optimal number of clusters by plotting the inertia (sum of squared errors within clusters) against the number of clusters (K). The curve showed an elbow point at **K = 4**, indicating diminishing returns in reducing inertia beyond 4 clusters.



* Additionally, **Silhouette Scores** were calculated for different values of K, and the scores reinforced the selection of **4 clusters** as a good fit. A reasonable silhouette score at K = 4 indicated well-defined clusters with distinct separation.

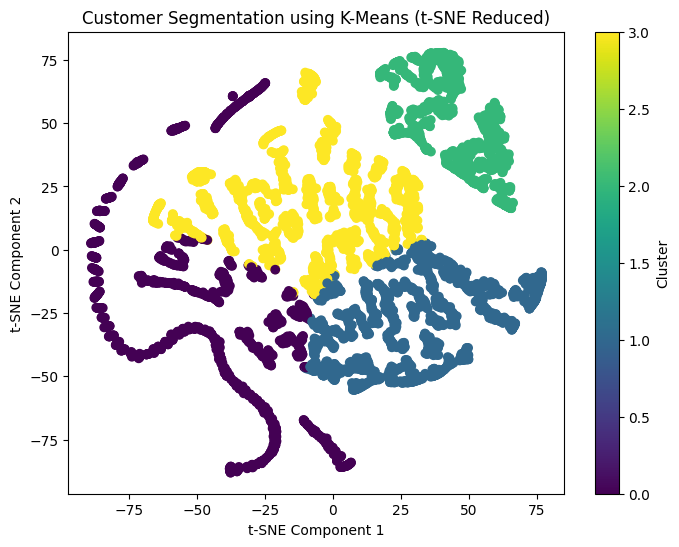


**3. Cluster Formation:**

The K-Means algorithm was trained with **K = 4**, and customers were assigned to four distinct clusters. These clusters were identified based on similar feature values in the scaled feature space. The model created well-separated groups of customers, with each cluster likely representing a different type of customer based on their data attributes.

**4. t-SNE Visualization:**

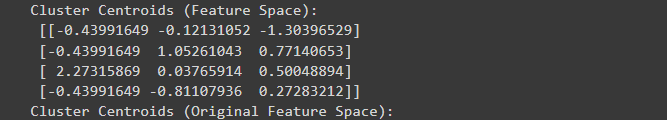
* **t-SNE** (t-distributed Stochastic Neighbor Embedding) was used to visualize the clusters in a 2D space, reducing the multi-dimensional data into two components for easier interpretation.
* The visualization clearly shows **four distinct clusters**, each represented by a different color. This confirms that the K-Means clustering algorithm successfully grouped customers into visually separable segments.



**5. Cluster Insights:**

The centroids of the clusters in both the scaled and original feature spaces provide insights into the characteristics of each cluster:

* **Cluster 0**: Likely contains customers with moderate values across the key attributes.
* **Cluster 1**: Customers in this group exhibit higher values in certain features compared to others.
* **Cluster 2**: Represents customers with the highest feature values.
* **Cluster 3**: This cluster groups customers with lower values, distinguishing them from other groups.



**6. Saving the K-Means Model**

To preserve the trained K-Means model for future use, it was saved using the pickle library. This allows the model to be reloaded without needing to retrain it, saving computational resources and time.