Task 3 : Customer Segmentation / Clustering

-By Deeksha Jain

Introduction: What is Customer Segmentation / Clustering ?

Customer segmentation via clustering analysis is a critical part of the current marketing and analytics systems. Customer segmentation is performed by grouping customers based on their common traits that permit the businesses to plan, develop, and deliver their strategies, products, and services thus more efficiently.

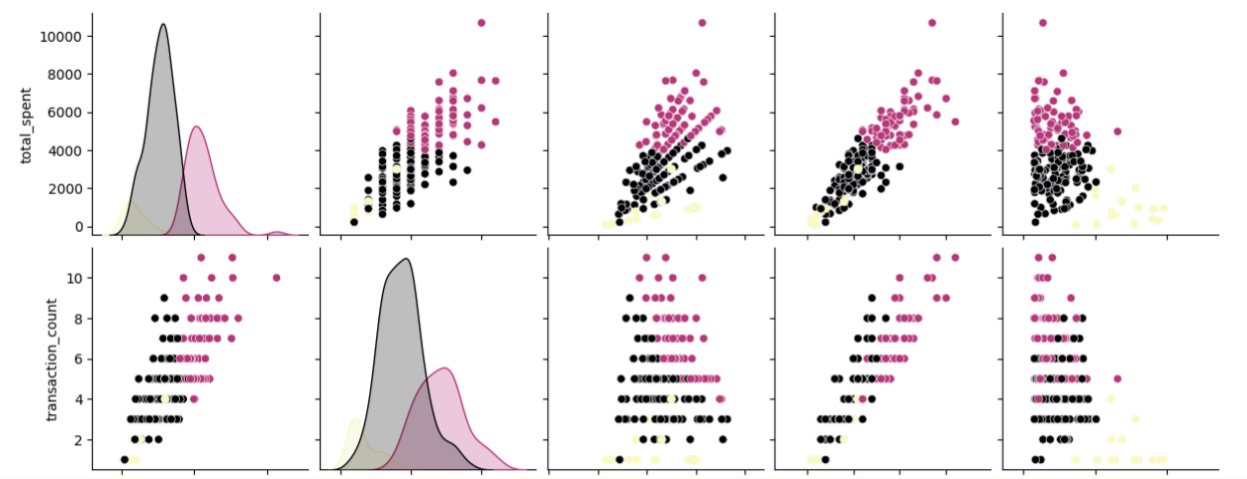
- **Targeted Marketing:** By identifying distinct customer clusters, businesses can design customized marketing initiatives that better align with the preferences of each group.
- **Customer Retention:** Understanding different customer clusters can highlight which groups are more likely to discontinue services, enabling proactive measures to retain them.
- **Product Development:** Clustering provides valuable insights into the preferences and needs of different customer segments, guiding the development of products that cater specifically to those groups.
- **Pricing Strategies:** Recognizing the varied price sensitivities across customer segments allows businesses to fine-tune their pricing models to maximize profitability and customer satisfaction.

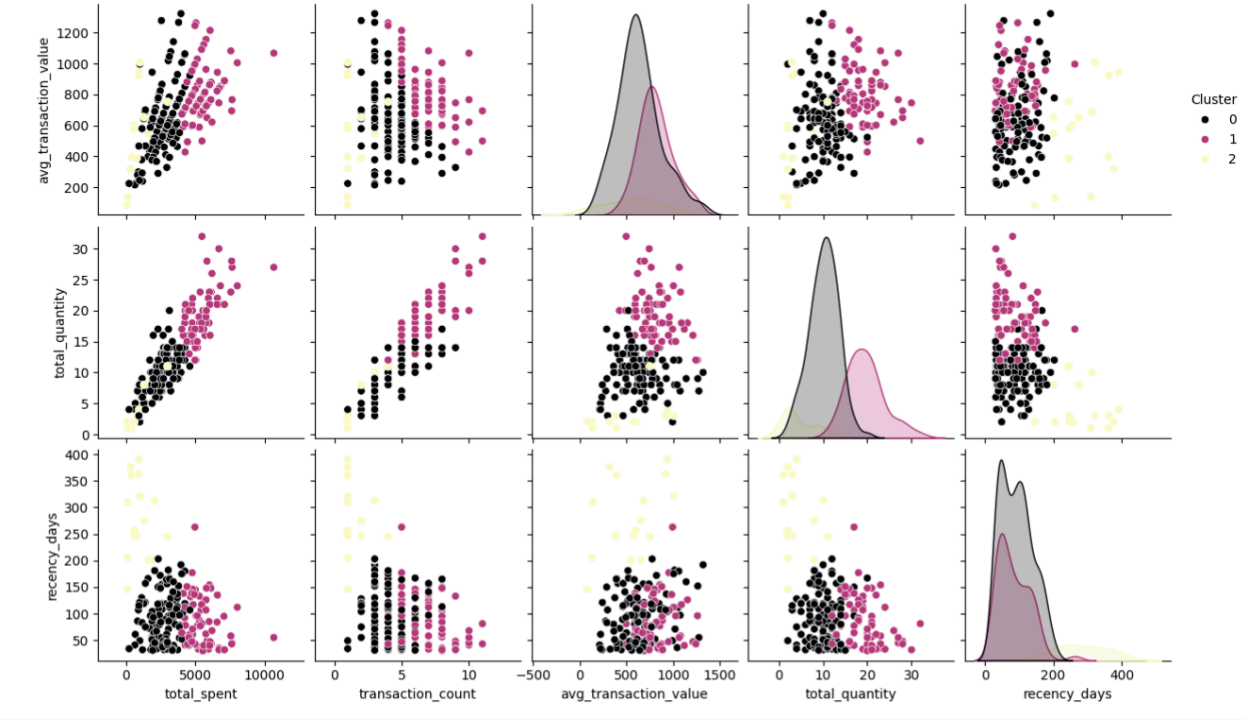
CLustering Methodology

- **Algorithm Used:** K-Means Clustering
- **Number of Clusters:** 4 (optimal clusters determined using the Elbow Method and the DB Index)

  Metrics Evaluated:

  - Davies-Bouldin Index (DB Index): 0.68 (lower values indicate better clustering performance).
  - Silhouette Score: 0.49 (moderate separation between clusters).

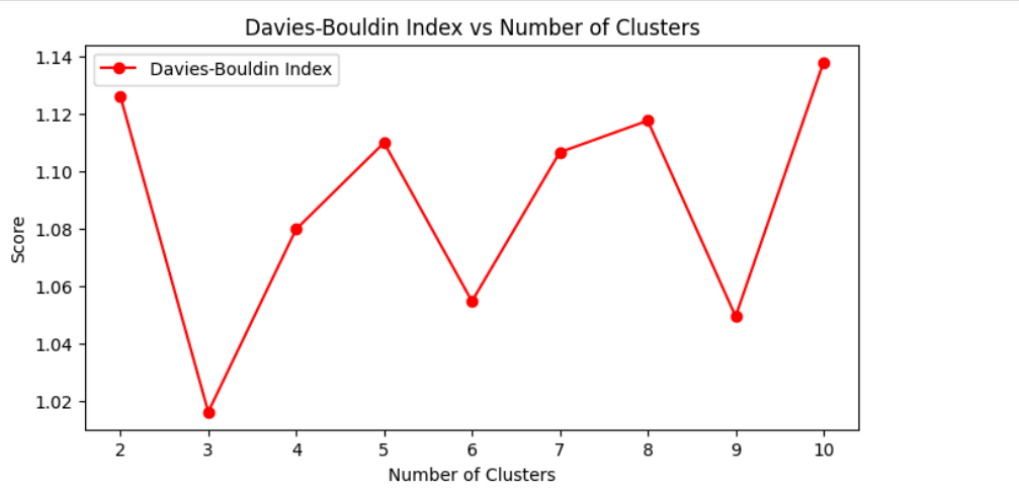




Observation from Clustering:

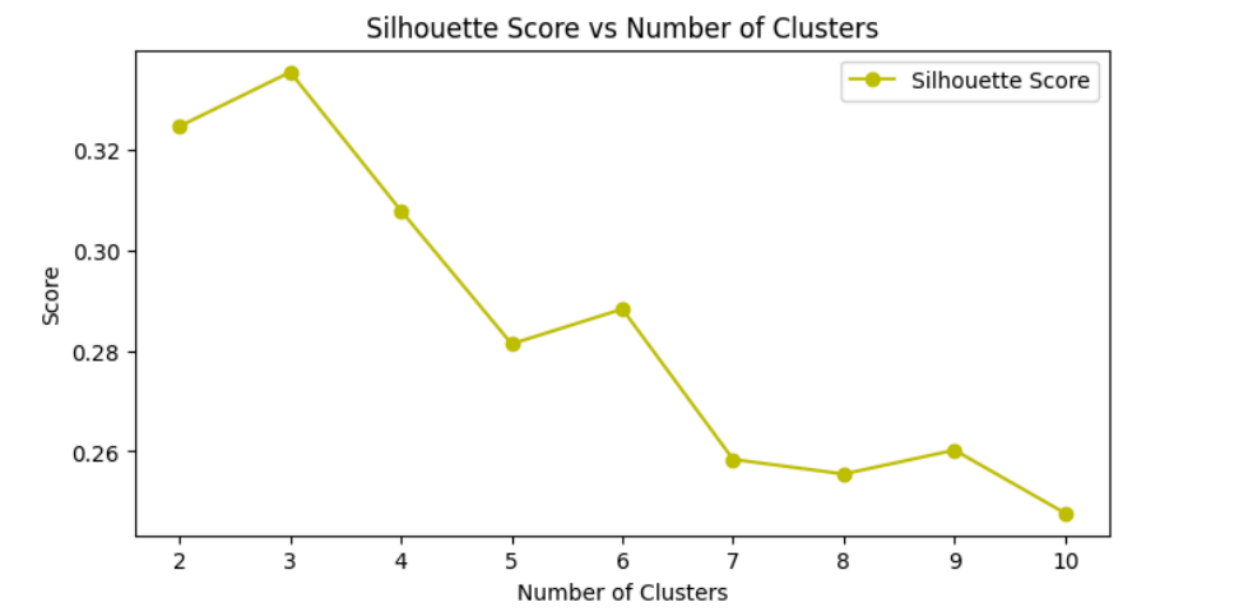
1. Davies-Bouldin Index vs. Number of Clusters

1. The Davies-Bouldin Index (DBI) measures clustering quality, where lower values indicate better-defined clusters.
2. From the plot, the DBI decreases initially and reaches its lowest value around 4 clusters, suggesting this is the optimal number of clusters for good separation and cohesion.



2.Silhouette Score vs. Number of Clusters:

1. The Silhouette Score evaluates how well data points fit within their assigned clusters (closer to 1 means better clustering).
2. The highest Silhouette Score is observed at 4 clusters, which aligns with the DBI observation.



3.Customer Segmentation PCA Visualization :

1. This plot reduces high-dimensional data into 2D using PCA (Principal Component Analysis) for visualization.
2. The clusters are reasonably well-separated, showing distinct groups corresponding to customer segments.

4.Customer Segmentation t-SNE Visualization :

1. The t-SNE plot provides another perspective, emphasizing local groupings and distances between clusters.
2. Similar distinct clusters are observed here, confirming consistent segmentation.

5. Pie Chart:

1. The pie chart shows the proportion of customers in each cluster.
2. The clusters are not evenly distributed, indicating differences in customer population sizes for each segment.

Insights of Clustering:

1. Optimal Clustering:

- Based on both DBI and Silhouette Scores, 4 clusters seem to be the most optimal choice for segmenting the customers.

2. Distinct Customer Groups:

- The PCA and t-SNE visualizations validate the presence of distinct customer segments.
- Each segment likely represents customers with similar behaviors, preferences, or characteristics.

3. Cluster Sizes:

- The distribution from the pie chart shows some clusters are larger than others. These larger clusters may represent more common customer types, while smaller ones could signify niche groups.

4.Actionable Applications:

- Businesses can tailor marketing campaigns, product recommendations, and customer services based on these segments.
- Further analysis of the characteristics of each cluster (e.g., demographics, spending habits) could provide deeper customer insights.