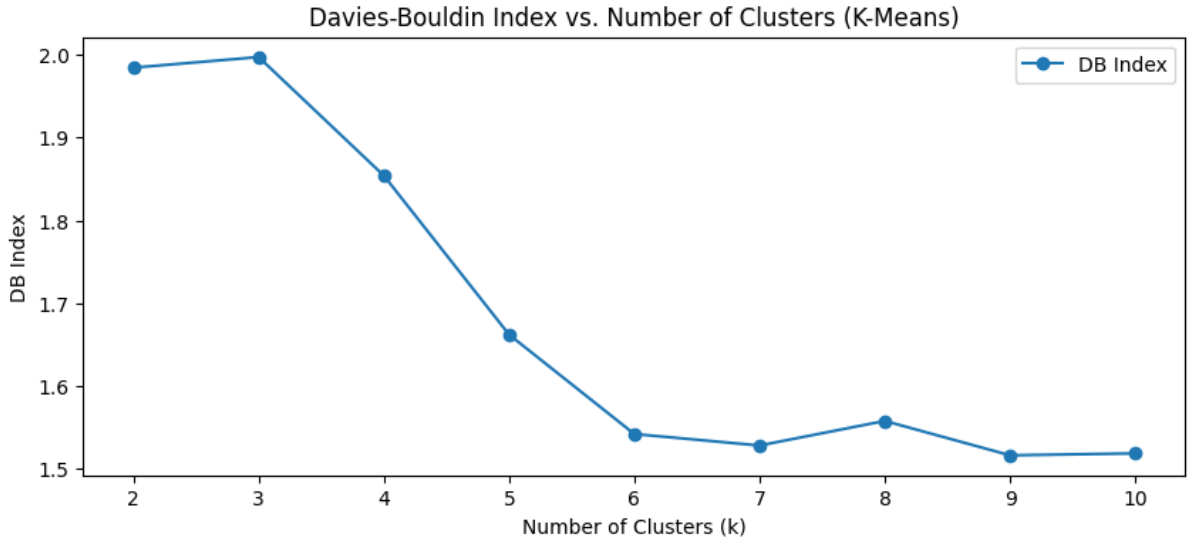
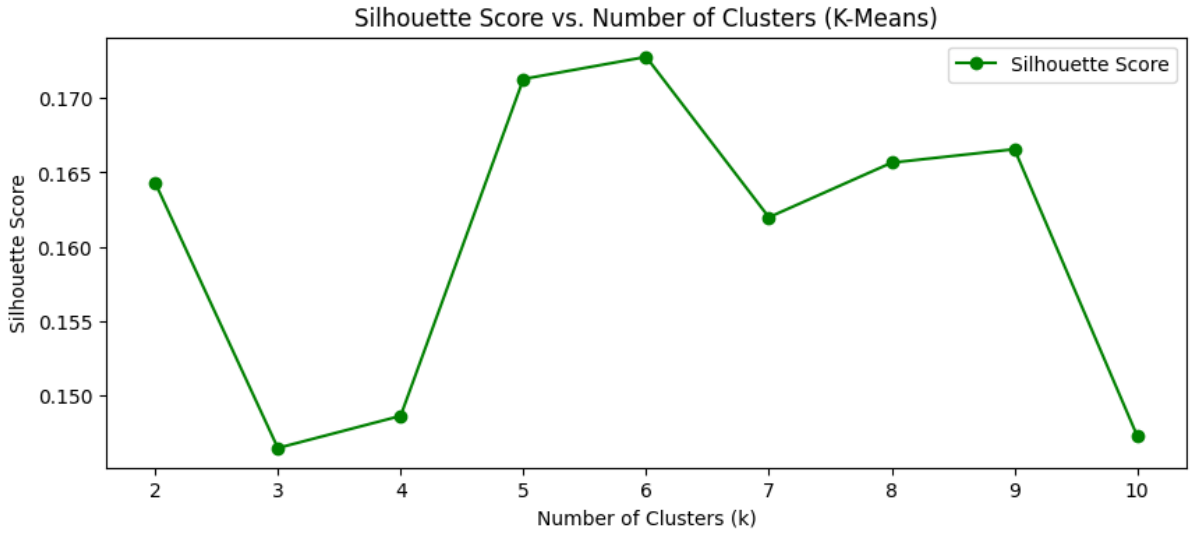
Clustering

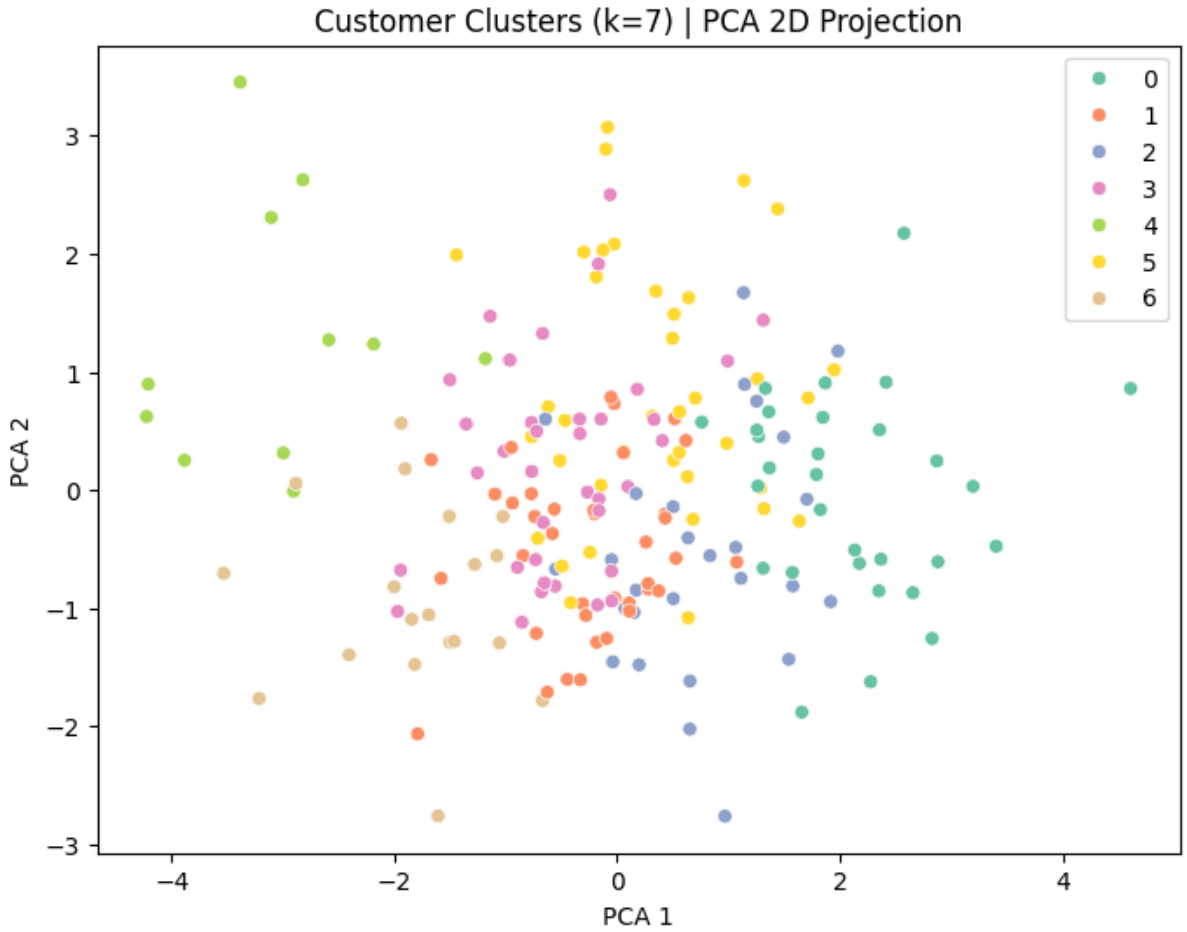
1. Why did we choose 7 cluster?

We tested different number of clusters as per requirement say from 2-10 and measured the DB Index and Silhouette Score to see how well the data separates. We chose 7 because it gives the gradual lowest DB index then other options, while maintaining a reasonable Sil Score





2. **Cluster Overview**



A total of 199 customers were grouped into 7 segments, where the following are as given:

1. Cluster 0: 30
2. Cluster 1: 37
3. Cluster 2: 27
4. Cluster 3: 37
5. Cluster 4: 11
6. Cluster 5: 38
7. Cluster 6: 19

Each cluster has distinct average traits, ranging from TotalSpend, AvgSpend, TxnSpend, etc For example, Cluster 0 has higher total spend (6284) and does around 8 transactions on average, but the average recency is 44 days since their last purchase meaning the’re faily active, While if you look at Cluster 4 has fewer transaction (around 1.6) and a large average recency of around 284 days, suggesting these customers rarely purchase and haven’t done so in a while.

Overview

1. Active High Spenders like cluster 0 and parts of cluster 2 and 5 could be prime for loyalty or membership programs or upselling.
2. Low purchase, high recency segments like cluster 5 might need re reengagement campaigns to spark interest.
3. Moderate buyers like cluster 3 and 1 can be targets for cross selling or product recommendations.
4.