TASK-3

Report on Clustering Results

Number of Clusters

From the elbow plot, I found that the best number of clusters to use is **4**. This was the point where adding more clusters didn't improve the model much, as shown by the smaller drops in the curve.

DB Index Value

The **Davies-Bouldin Index (DB Index)**, which measures how good the clusters are, came out to be **0.865**. A lower DB Index means the clusters are well-separated and tightly grouped, which is a good result.

Clustering Process

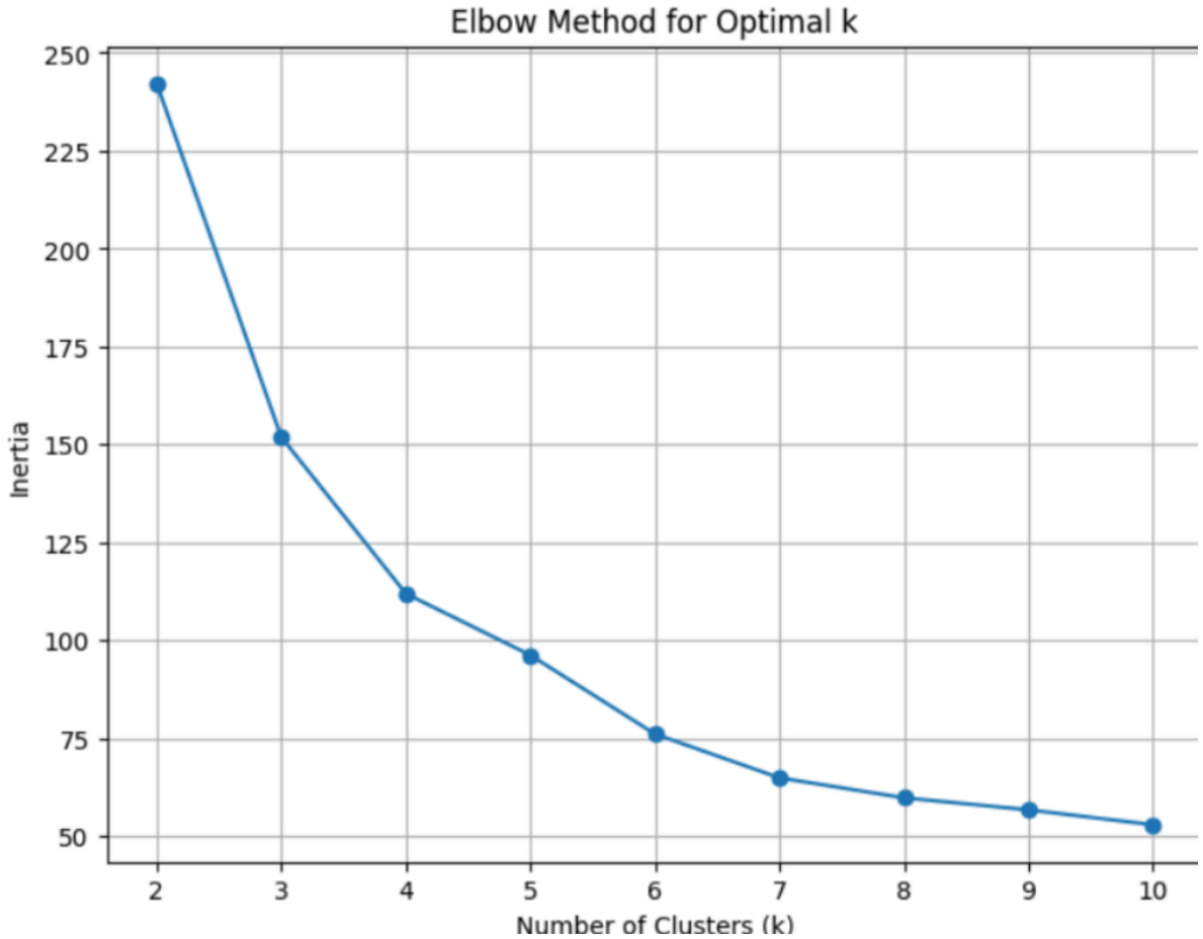
- **Method Used**: I used **K-Means clustering** to group customers based on their transaction history and purchase behavior.
- **Scaling the Data**: I standardized the data (made all features equally important) to improve clustering accuracy.
- **Visualization**: I used **PCA (Principal Component Analysis)** to reduce the data to two dimensions for easy visualization. The scatter plot shows that the customers are divided into four distinct groups.

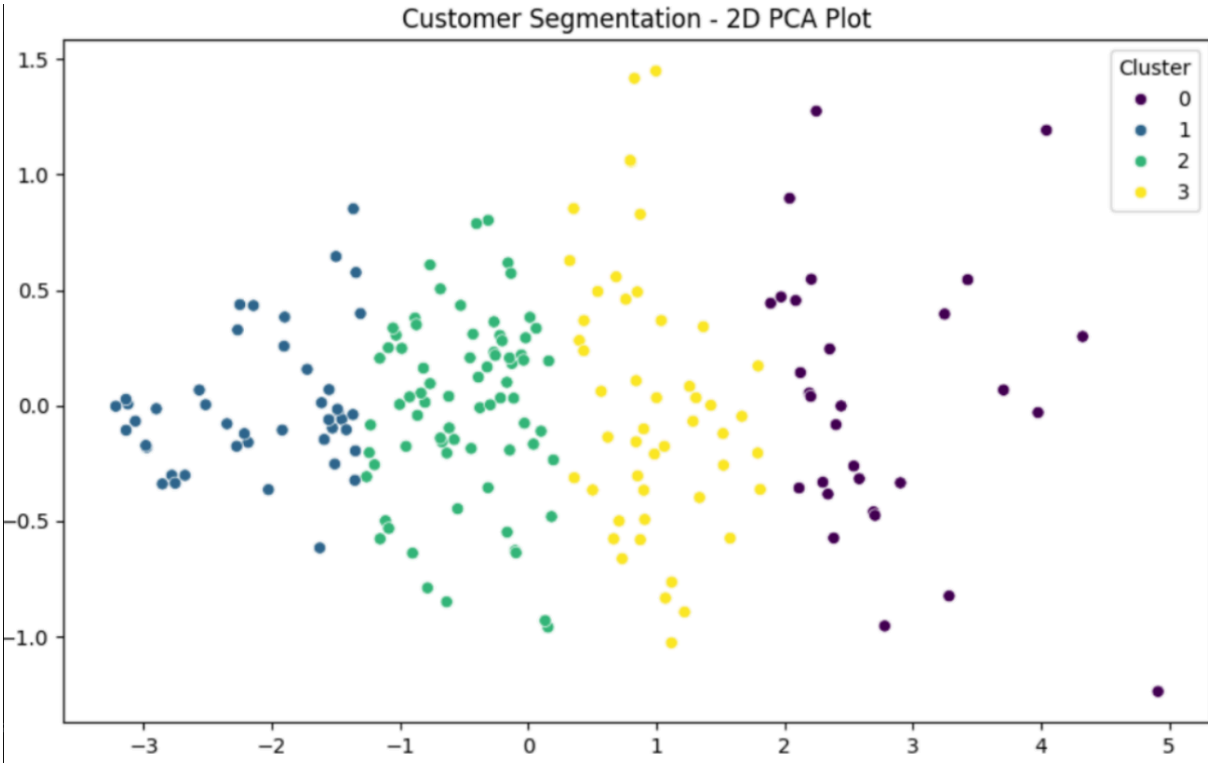
Visuals

1. **Elbow Plot**: This plot helped me decide the number of clusters. At 4 clusters, the curve flattened, which means adding more clusters wouldn't make much difference.
2. **Cluster Plot**: The PCA plot shows how the customers are divided into four clear groups, proving that the clusters are meaningful.

Conclusion

I successfully grouped the customers into four clusters based on their shopping patterns. These groups can help the company design better offers, target the right customers, and improve sales strategies.


Elbow Method for Optimal k
Inertia
Number of Clusters (k)
250
225
200
175
150
125
100
75
50
2
3
4
5
6
7
8
9
10


Customer Segmentation - 2D PCA Plot
Cluster
0
1
2
3
1.5
1.0
0.5
0.0
−0.5
−1.0
−3
−2
−1
0
1
2
3
4
5