1: What did you implement?

I implemented a K-Means clustering customers based on their Annual Income and Spending Score

1. Scaling: Used StandardScaler income and spending
2. SSE loop: Run K-Means with values of k = 1 to 11 check inertia SSE for the elbow method
3. Clustering: Selected k = 6 clusters and trained the model
4. Evaluation metrics: Calculated Silhouette Score and Davies-Bouldin Index
5. Labeling: Assign each customer a cluster and saved results new dataset
6. Cluster centers: Converted cluster centers back to original units

2. Choosing K

I tested K 1 to 11 and compared SSE Elbow method Silhouette score, and DBI

Therefore, I chose K = 6 because it balances interpretability and cluster quality

3. Cluster Interpretation

Based on the center income and spending score, I interpreted the clusters as:

1. Low Income, Low Spending = Customers with low purchasing
2. Medium Income, Medium Spending = Average customers
3. High Income, High Spending = Best customers VIP

4. Limitations & Next Steps

The segmentation used only Annual Income and Spending Score While useful, it ignores other factors that influence customer behavior

1. Age could reveal generational spending patterns
2. Visits per month could separate frequent visitors from occasional shoppers
3. Online Purchases could highlight digital vs in-store buyers

Next Step:

A concrete improvement would be to extend the clustering with 3 features e.g., add Age, visits and compare the results, additionally experimenting with DBSCAN could help discover non-spherical clusters and outlier customers that K-Means may miss