

## Step 4: Comparative Analysis and Discussion

Your Name

February 12, 2025

### Step 4 (3 marks): Comparative Analysis and Discussion

#### a. 3D Visualization Plots

Include two 3D visualization plots below:

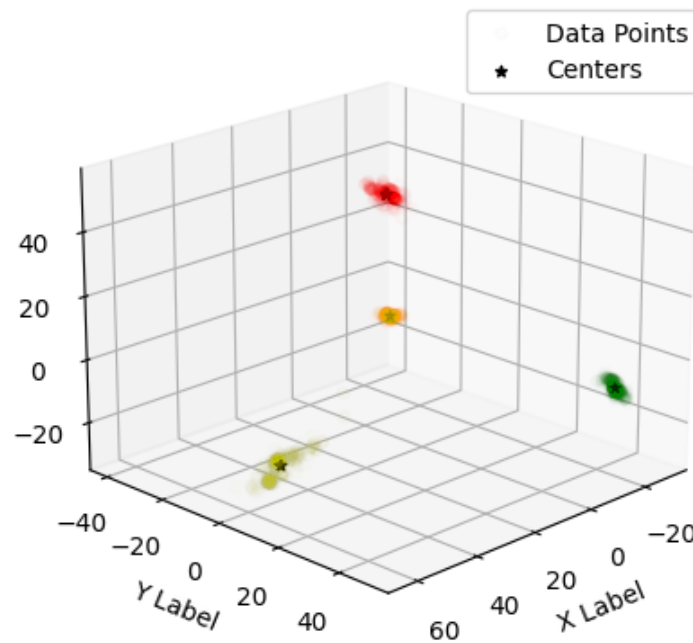


Figure 1: 3D Visualization of Clusters from K-Means ( $K = 4$ )

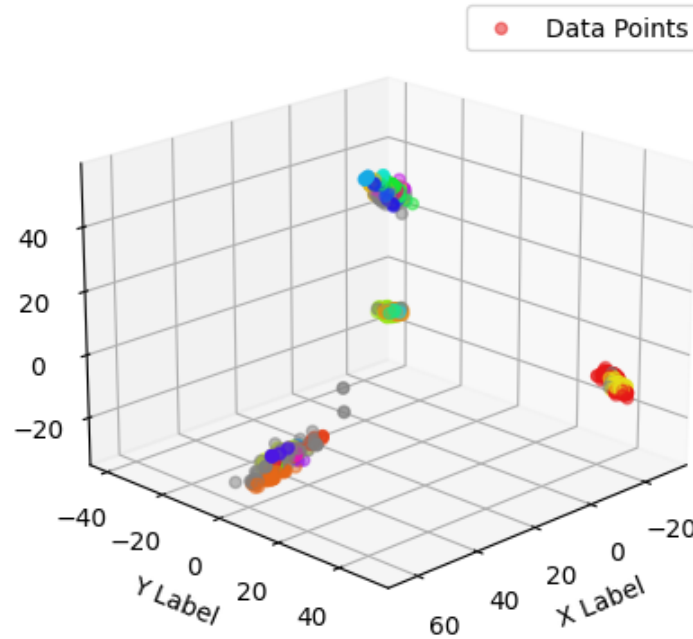


Figure 2: 3D Visualization of Clusters from DBSCAN (run#1)

### b. DBSCAN Clustering Analysis

- How does the number of clusters identified by DBSCAN vary between the two parameter settings (run#1 and run#2)?
- The run#1 generates 31 clusters in total, while the run#2 generates 22 clusters.
- Why does increasing `minPts` affect the number of detected clusters?
- As the `minPts` increase, only points with more neighbours would be considered a density, which decrease the numebr of clusters, more points would be considered noise.

### c. Cluster Membership and Separation

- How many data points belong to each cluster for K-Means ( $K = 4$ ) and DBSCAN (run#2)?

Value	Counter 1	Counter 2	Counter 3
-1	0	99	139

Value	Counter 1	Counter 2	Counter 3
0	375	203	203
1	399	23	23
2	425	100	100
3	350	117	117
4	0	135	134
5	0	192	192
6	0	117	117
7	0	294	294
8	0	43	43
9	0	5	65
10	0	6	13
11	0	65	11
12	0	13	8
13	0	4	14
14	0	11	6
15	0	5	25
16	0	8	9
17	0	14	8
18	0	6	6
19	0	10	6
20	0	25	10
21	0	4	6
22	0	9	0
23	0	8	0
24	0	4	0
25	0	6	0
26	0	7	0
27	0	4	0
28	0	4	0
29	0	4	0
30	0	4	0

- Which algorithm appears to provide better cluster separation for this dataset? Explain why.
- Considering the PCA, which takes the most important three features, the 3D plot shows 4 clusters in total, which is quite the same as the kmeans output, and the 4 center points are obviously the core points of the clusters. Thus, k-means is doing a better job.

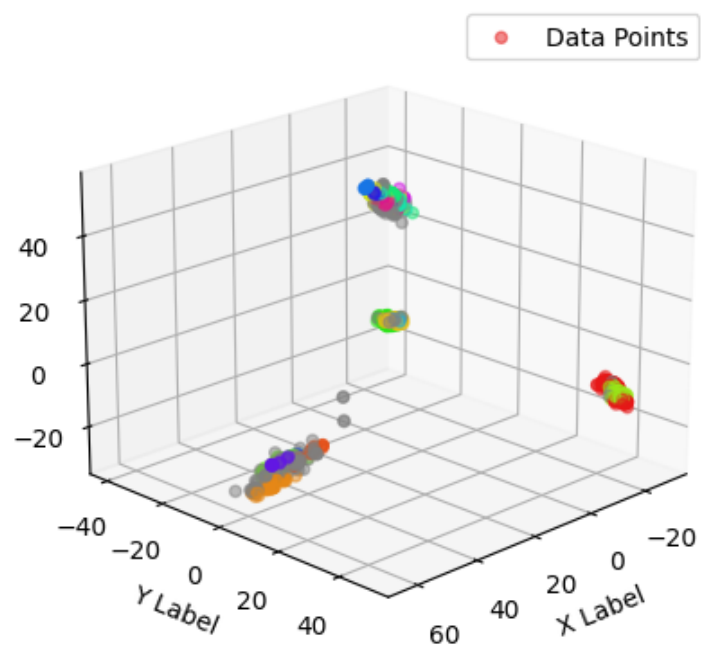


Figure 3: 3D Visualization of Clusters from DBSCAN (run#2)