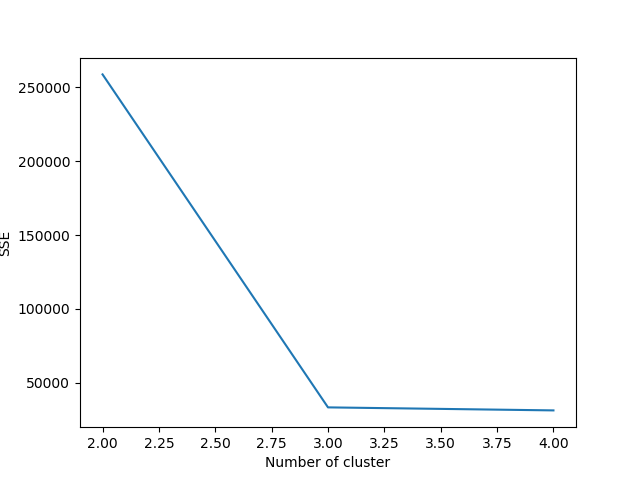
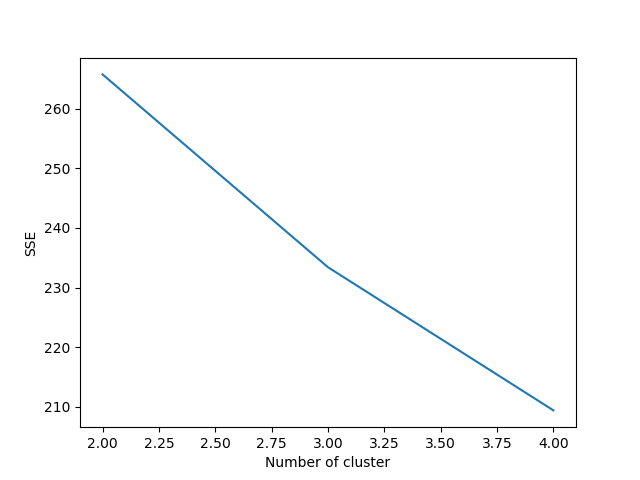
HW3 part2 Deepak Kanuri(G01070295)

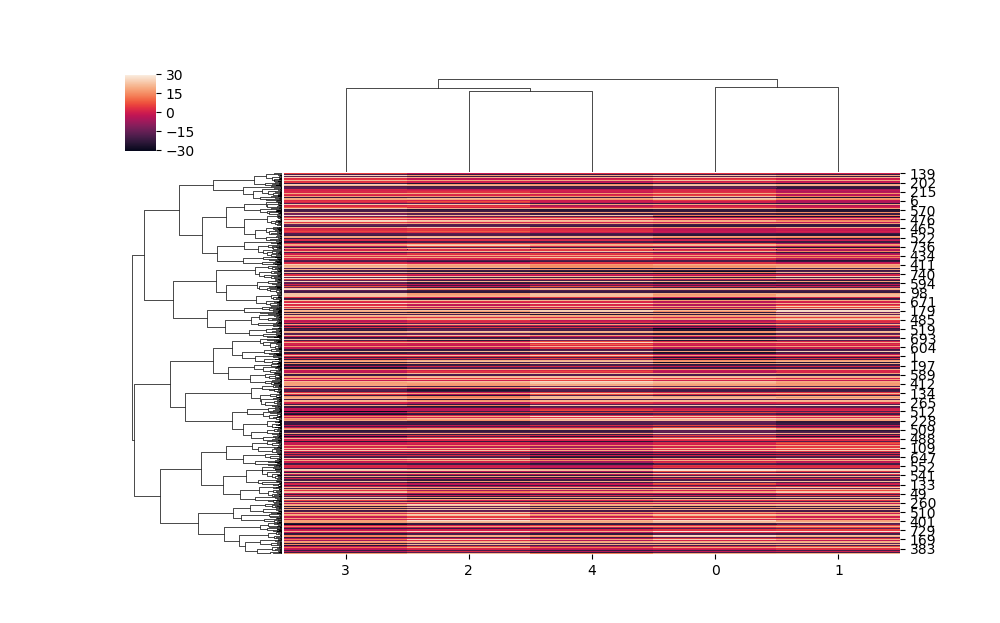
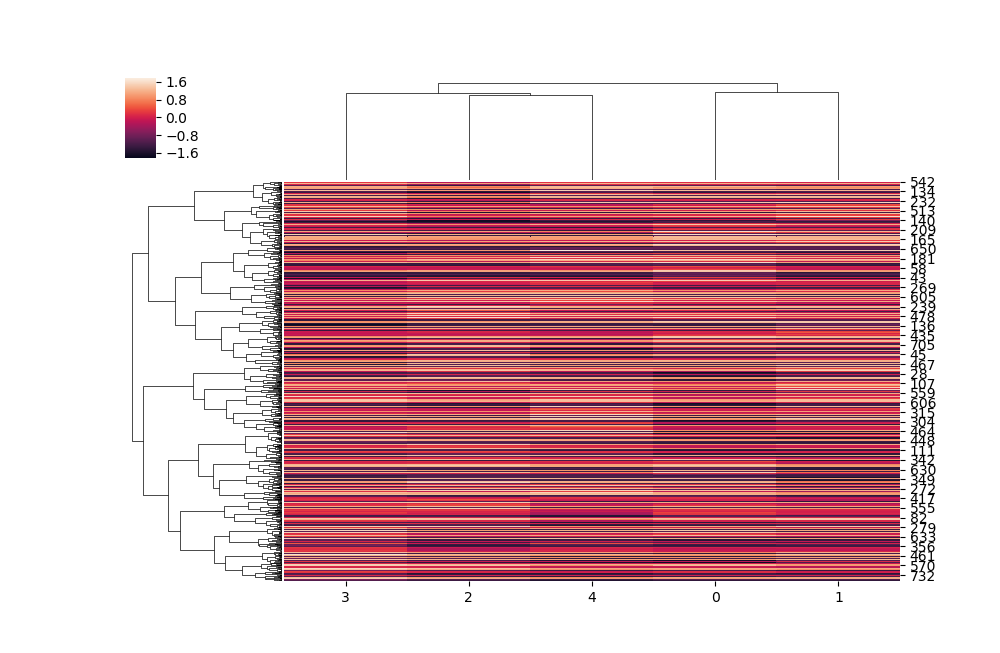
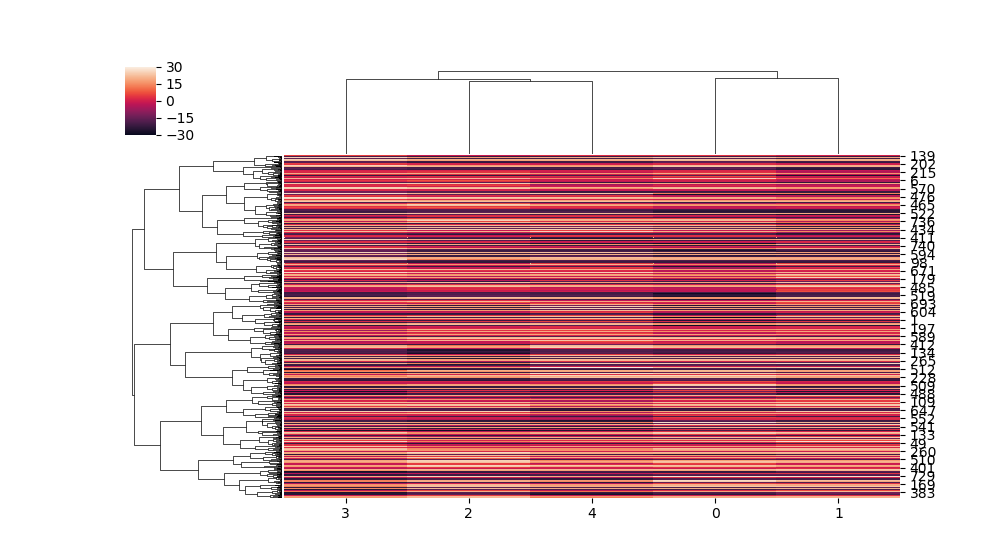
Using SSE the best value of ‘K’ is 3 for both the datasets. When comparing with silhouette score of DBScan, K=3 is the best.



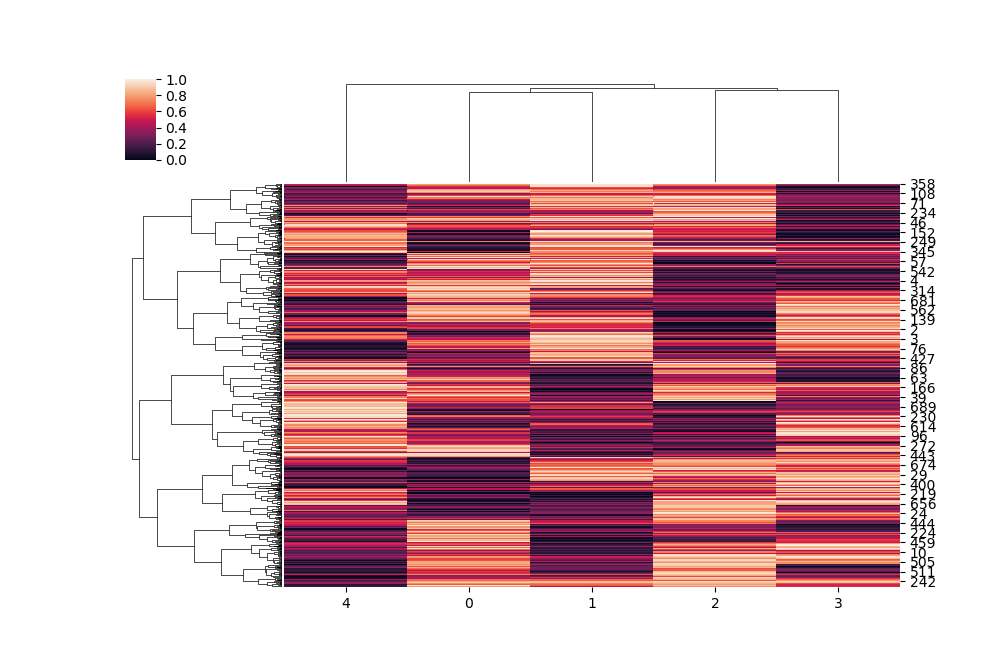


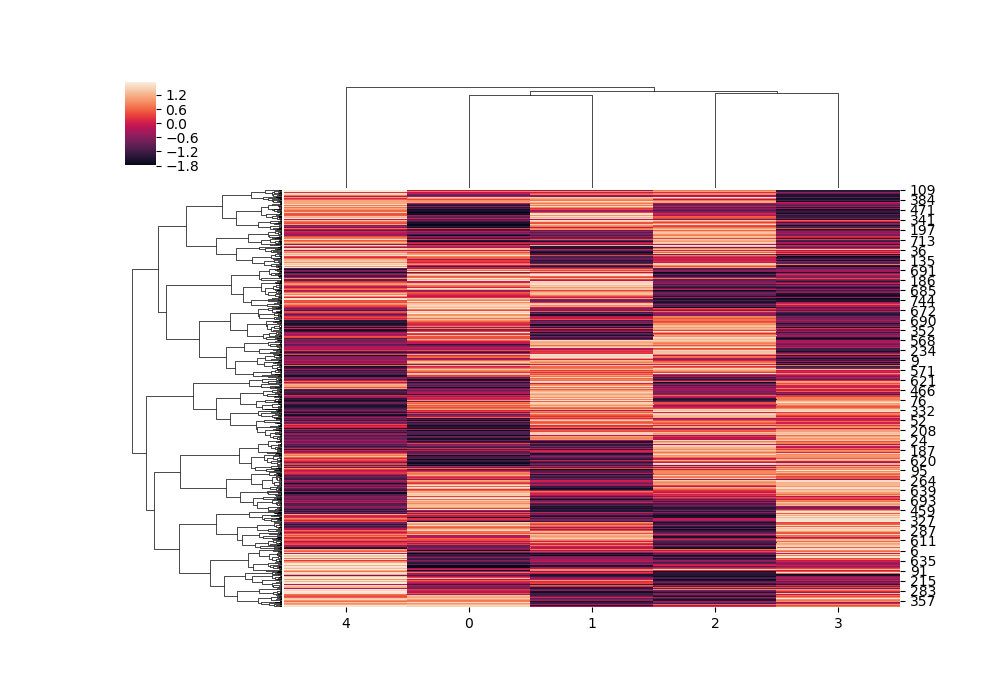
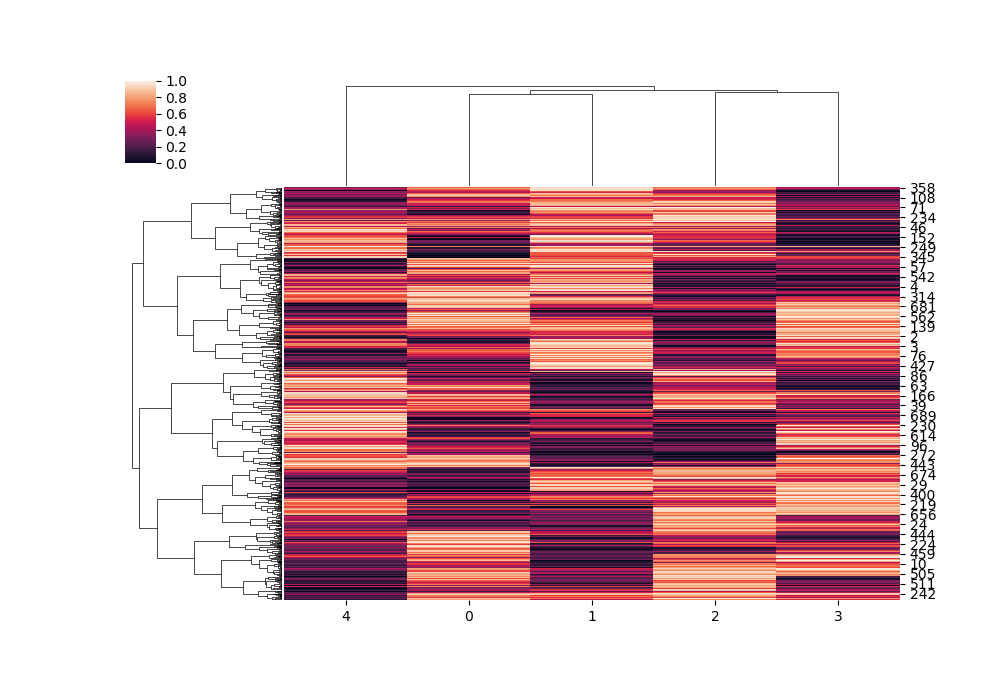
|  |  |  |  |
| --- | --- | --- | --- |
| silhouette coefficient | Kmeans | DBscan | EM |
| Dataset1 | 0.79850 | 0.79856 | 0.79850 |
| Dataset2 | 0.1530 | -0.30 | 0.1303 |

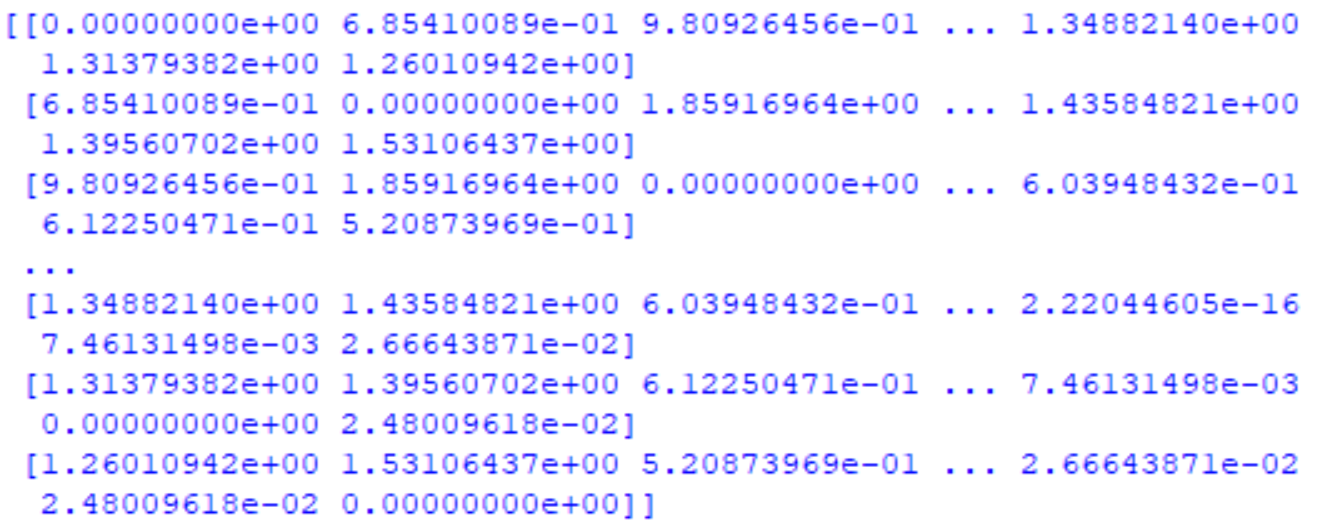
The clustering validity for dataset1 for kmeans, DBScan and EM respectively are



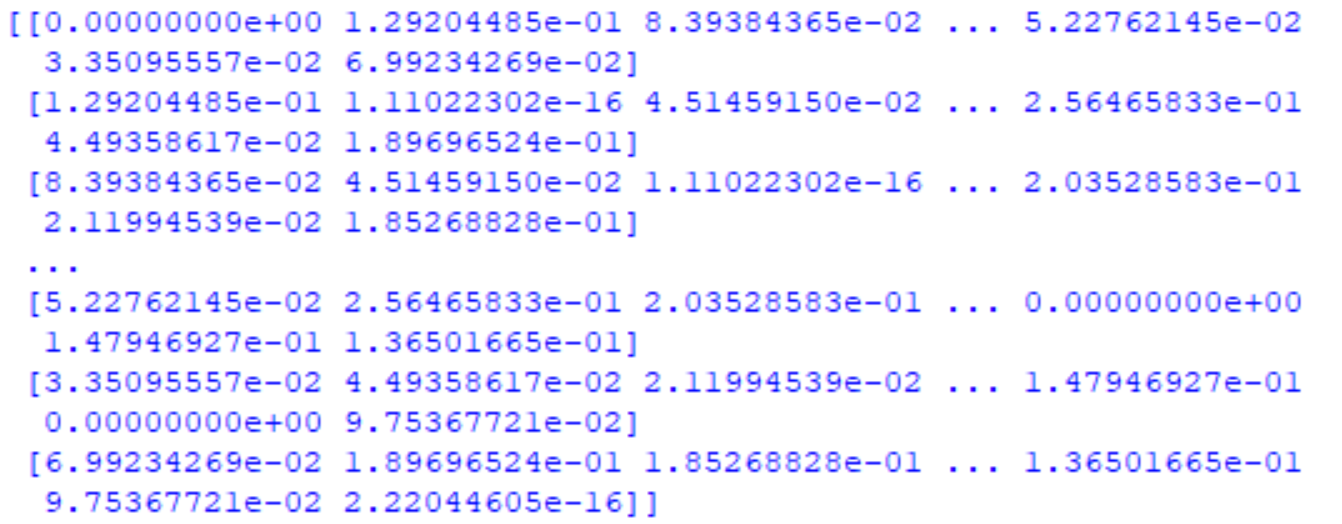
The clustering validity for dataset2 for kmeans, DBScan and EM respectively are



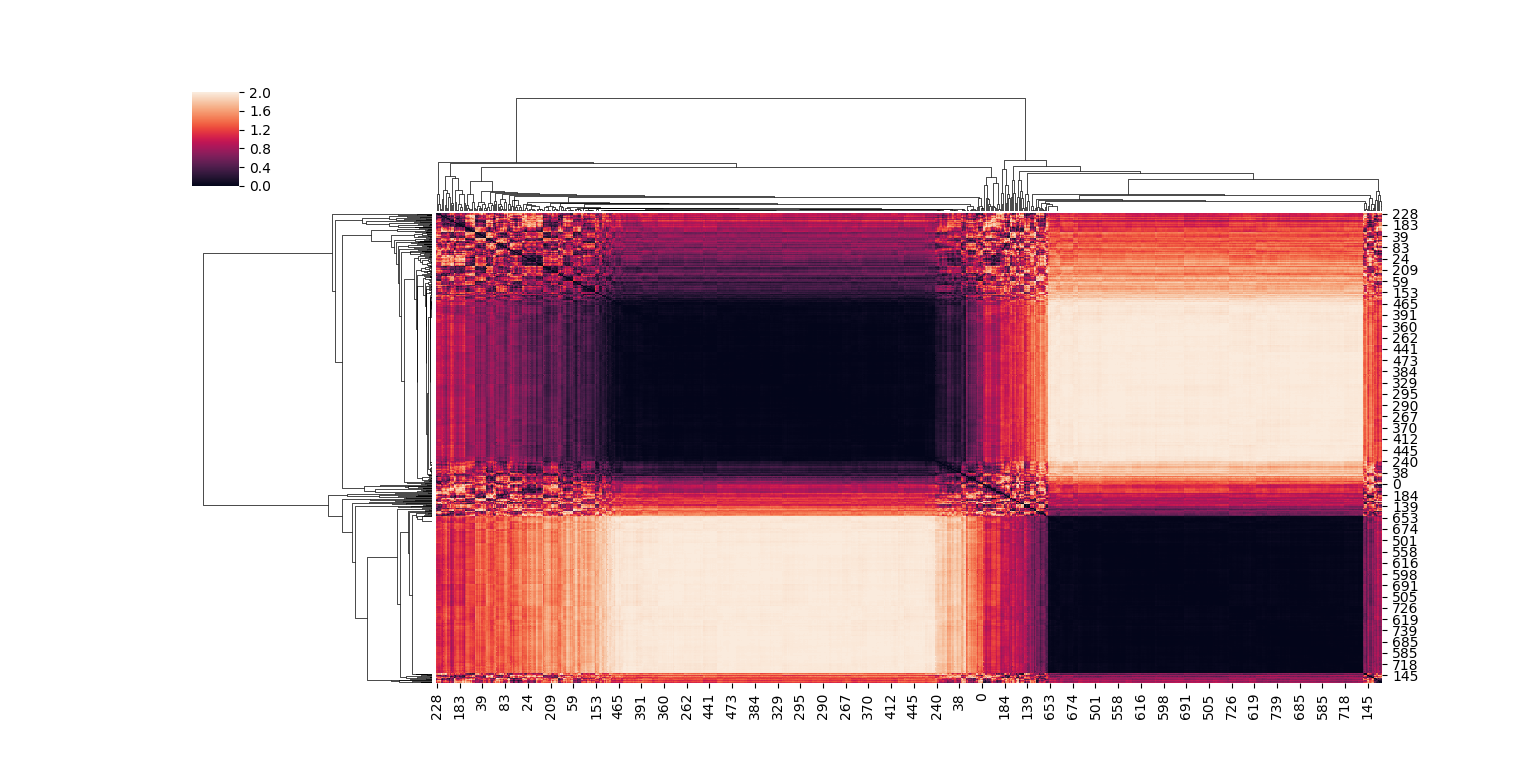
 

The pairwise distance matrix for 1st dataset is

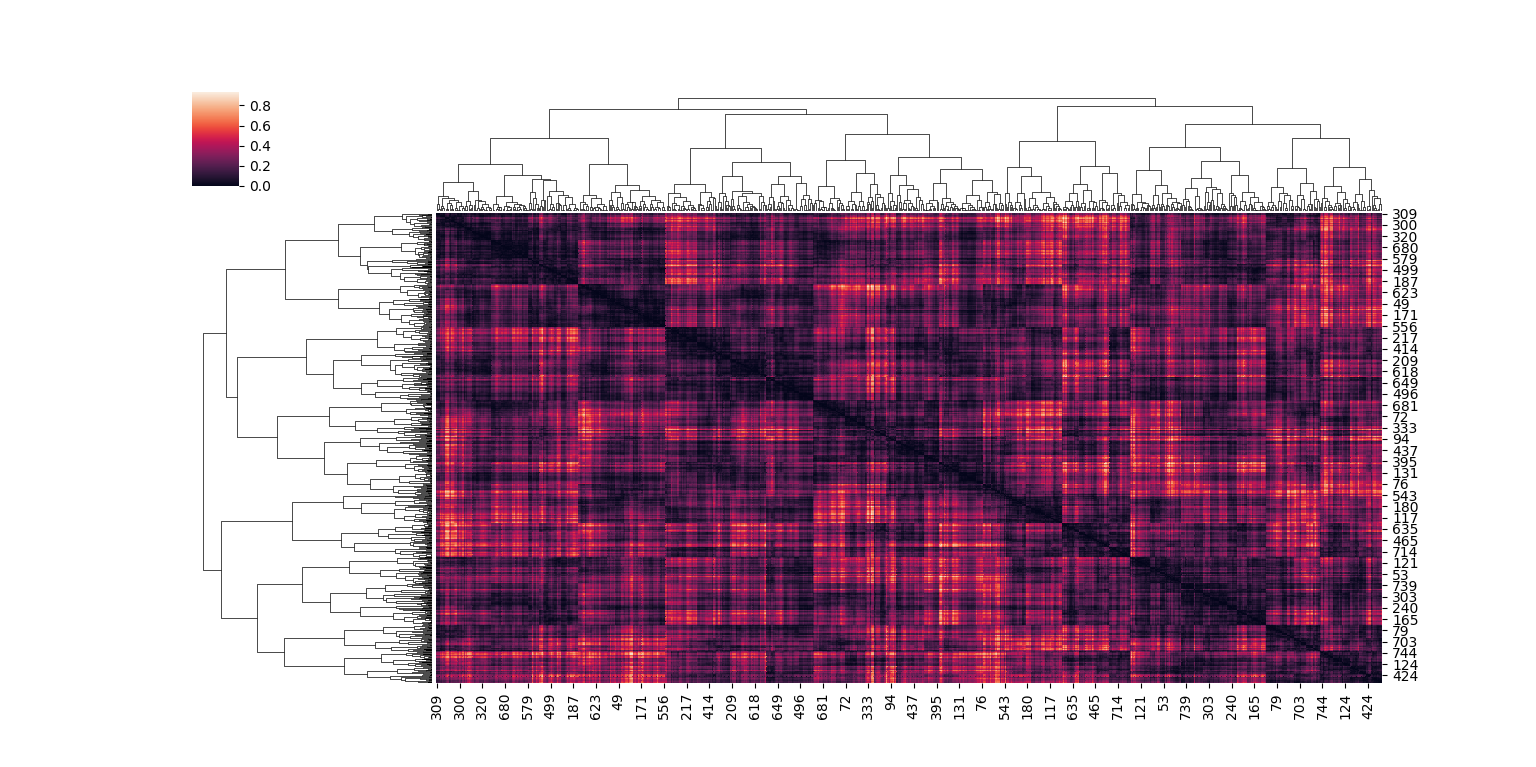
The pairwise distance matrix for 2nd dataset is



The heatmap for the first pairwise distance is



The heatmap for the second pairwise distance is



According to me the 1st dataset REALLY exhibits the clustering structure.