Analysis Report

1. Discuss any discrepancies observed between 4(a), 4(b), or 4(c). Use the silhouette score approach to choose the number of clusters for 4(a), 4(b), and 4(c). [10 points]

Between the three different version of metrics: Euclidean, Minkowski, and Cosine, I have prepared a silhouette score, a scatter plot, and a dendrogram for your viewing.

From the Silhouette scores, it can be seen that the Minkowski silhouette score has the highest value followed by the Euclidean and then the Cosine scores. As for the average scores, there is a clear trend of decrease from 2 to 5 clusters, with a slight uptick at 6 clusters.

As for the Scatter plots, it can be seen that each one has a different layout of coloring. In the Dendrograms the Euclidean one has the cleanest spread, whereas the Minkowski has the most sporadic spread, with the Cosine having the most biased spread.

Silhouette Scores:

Euclidean Silhouette Score: 0.082989365

For n\_clusters = 2 The average silhouette\_score is : 0.15156598

For n\_clusters = 3 The average silhouette\_score is : 0.12555625

For n\_clusters = 4 The average silhouette\_score is : 0.09605395

For n\_clusters = 5 The average silhouette\_score is : 0.09553286

For n\_clusters = 6 The average silhouette\_score is : 0.10251042

Minkowski Silhouette Score: 0.13071483

For n\_clusters = 2 The average silhouette\_score is : 0.15243226

For n\_clusters = 3 The average silhouette\_score is : 0.12537062

For n\_clusters = 4 The average silhouette\_score is : 0.11209189

For n\_clusters = 5 The average silhouette\_score is : 0.095036134

For n\_clusters = 6 The average silhouette\_score is : 0.101691455

Cosine Silhouette Score: 0.06333944

For n\_clusters = 2 The average silhouette\_score is : 0.15156598

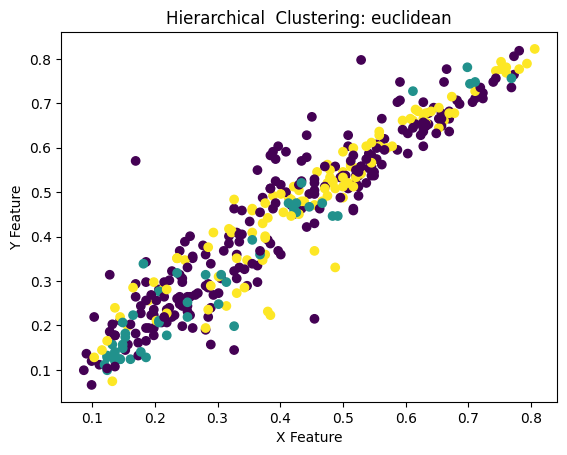
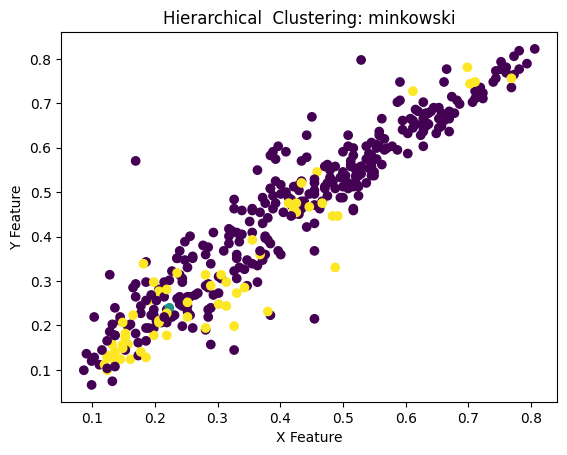
For n\_clusters = 3 The average silhouette\_score is : 0.12555625

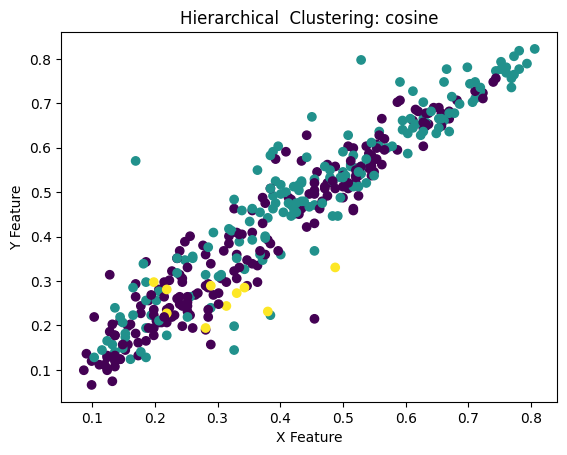
For n\_clusters = 4 The average silhouette\_score is : 0.11302761

For n\_clusters = 5 The average silhouette\_score is : 0.08503014

For n\_clusters = 6 The average silhouette\_score is : 0.08810151

Hierarchical Clustering:



Hierarchical Clustering Dendrograms:

