Milo Davis

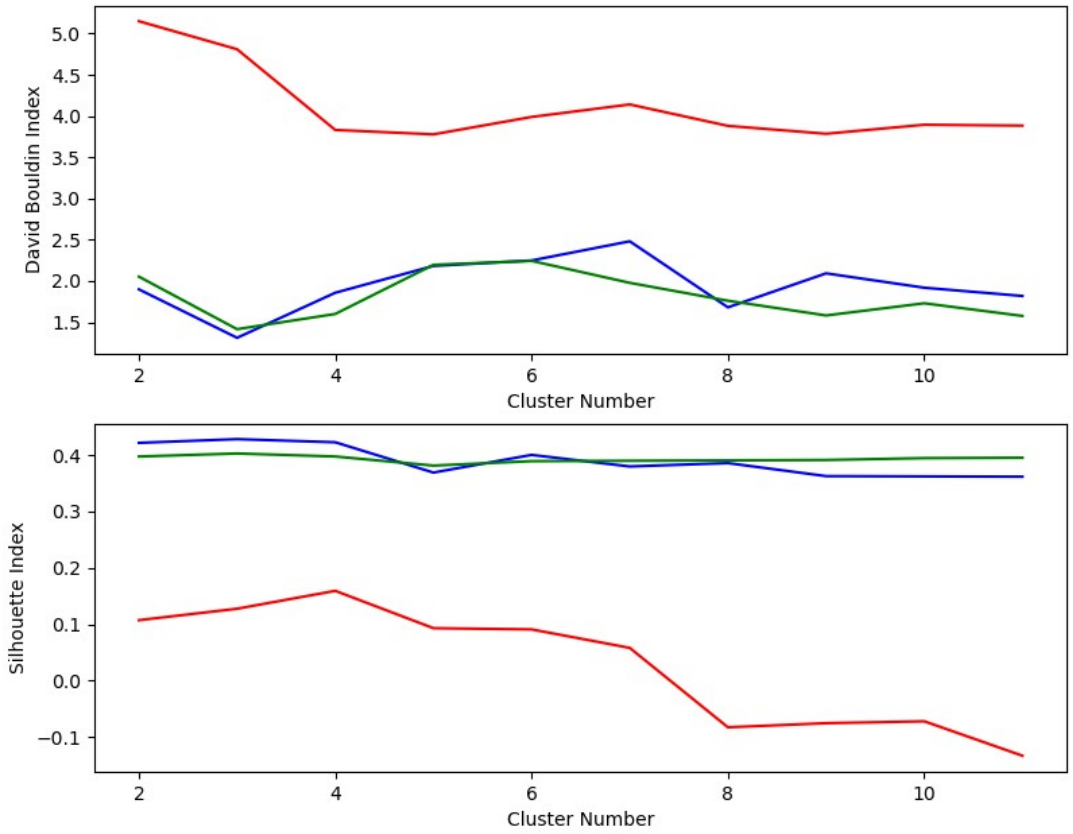
CS 4435

May 13th, 2020

Project 3 Report

a. Below is a plot of the clustering quality scores, David Bouldin Index and Silhouette index, against the number of clusters being used.

- K-means
- K-medoids
- Hierarchical Clustering (Agglomerative)



A lower Davies Bouldin Index is better quality, a higher Silhouette Index is better quality.

b. The performance of k-means and hierarchical agglomerative clustering were pretty similar in performance, while k-medoids was the biggest outlier. For both Davies Bouldin Index and Silhouette Index, k-means and agglomerative clustering are about the same, with k-means being better by a small margin with less clusters and agglomerative being better by a small margin with more clusters. For both metrics, however, k-medoid clustering had the worst quality overall.

c. The best clustering quality for agglomerative and k-means clustering was at 3 clusters, and the best quality for k-medoid was at 4 clusters.

- **Running Times at 3 clusters**
  - k-means: 3.191948175430298
  - k-medoids: 2.1889808177948
  - hierarchical (agglomerative): 2.903336763381958
- **Running Times at 4 clusters**
  - k-means: 3.4838945865631104
  - k-medoids: 2.201429605484009
  - hierarchical (agglomerative): 2.87555384635925