Discrepancies:

(Due to size of data, all the scores are low, which means not too accurate, however, if the train set is increased, for example, to increase it to the whole X, the scores will be higher.)

After train and test using different cluster, the principle of more clustering gives higher accuracy can be found. To compare each of distance calculations, with using Agglomerative Clustering, the one using euclidean Distance gives lowest Standard Deviation while the one with cosine gives highest mean scores of k-fold validation.

Euclidean Distance:

scores: [0.01492537 0.01515152] Mean: 0.015038444142921756 Standard deviation: 0.000113071008593397

best cluster set is: 120 -0.08277093

Cosine:

scores: [0.02985075 0.01515152] Mean: 0.022501130710085936 Standard deviation: 0.007349615558570782

best cluster set is: 120 -0.09102282

Manhattan:

scores: [0. 0.01515152] Mean: 0.007575757575757576 Standard deviation: 0.007575757575757576

best cluster set is: 120 -0.08277093