

Cluster data for problem 5

Num of elements	Homogeneity [pos, neg]	Completeness [pos, neg]
25	[0.88, 0.81]	[0.8, 0.9]
50	[0.6666, 0.6363]	[0.6, 0.7]
100	[0.5384, 0.5714]	[0.7, 0.4]
500	[0.5555, 0.5454]	[0.5, 0.6]
1000	[0.5, 0.5]	[0.4, 0.6]

After doing stopwords, punctuation, and capitalization

Num of elements	Homogeneity [pos, neg]	Completeness [pos, neg]
100	[0.6666, 0.6363]	[0.6, 0.7]