

Summary from the last meeting:

make some figures from the experiments

What has done this week:

1. Do clustering on the dataset

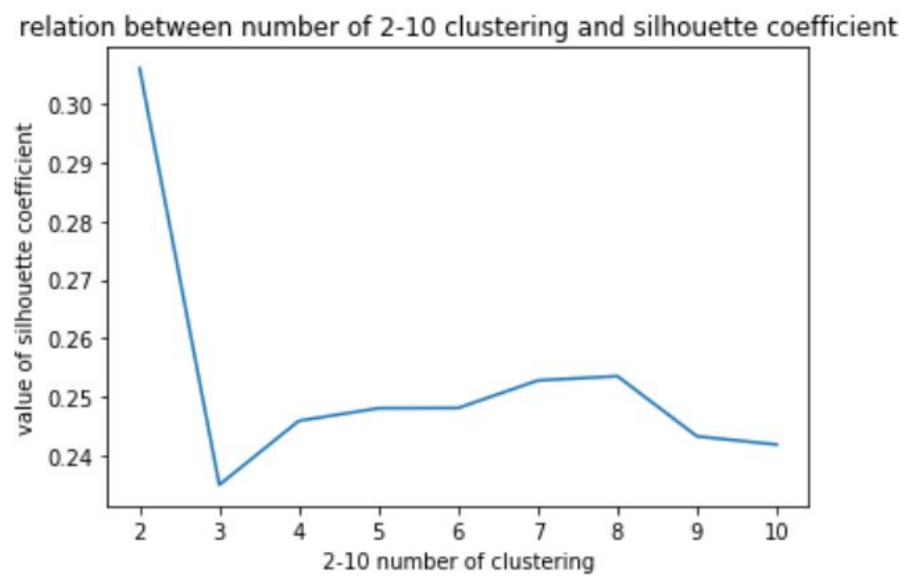
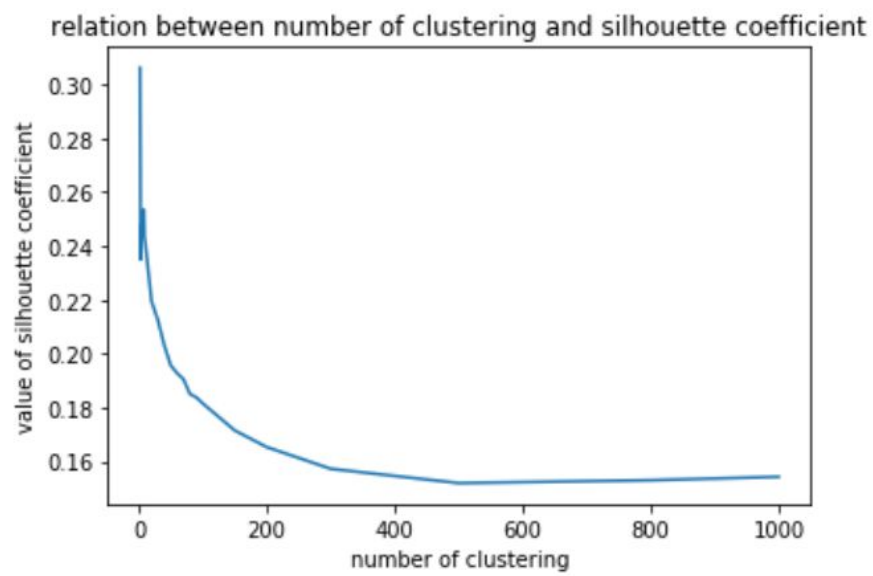
a. DBSCAN

Estimated number of clusters: 940

Estimated number of noise points: 52446

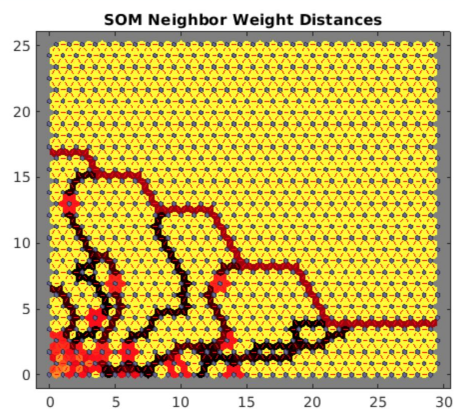


b. kmeans

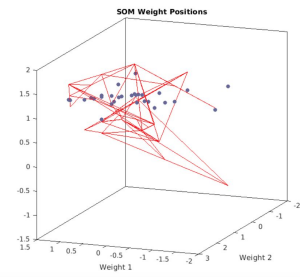


c. SOM

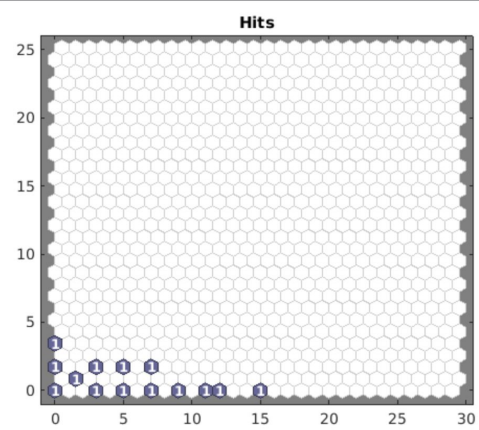
Neural Network Training SOM Neighbor Distances (plotsomnd), Epoch 200, Maximum ...



SOM Weight Positions (plotsompos)



Neural Network Training SOM Sample Hits (plotsomhits), Epoch 200, Maximum epoch ...

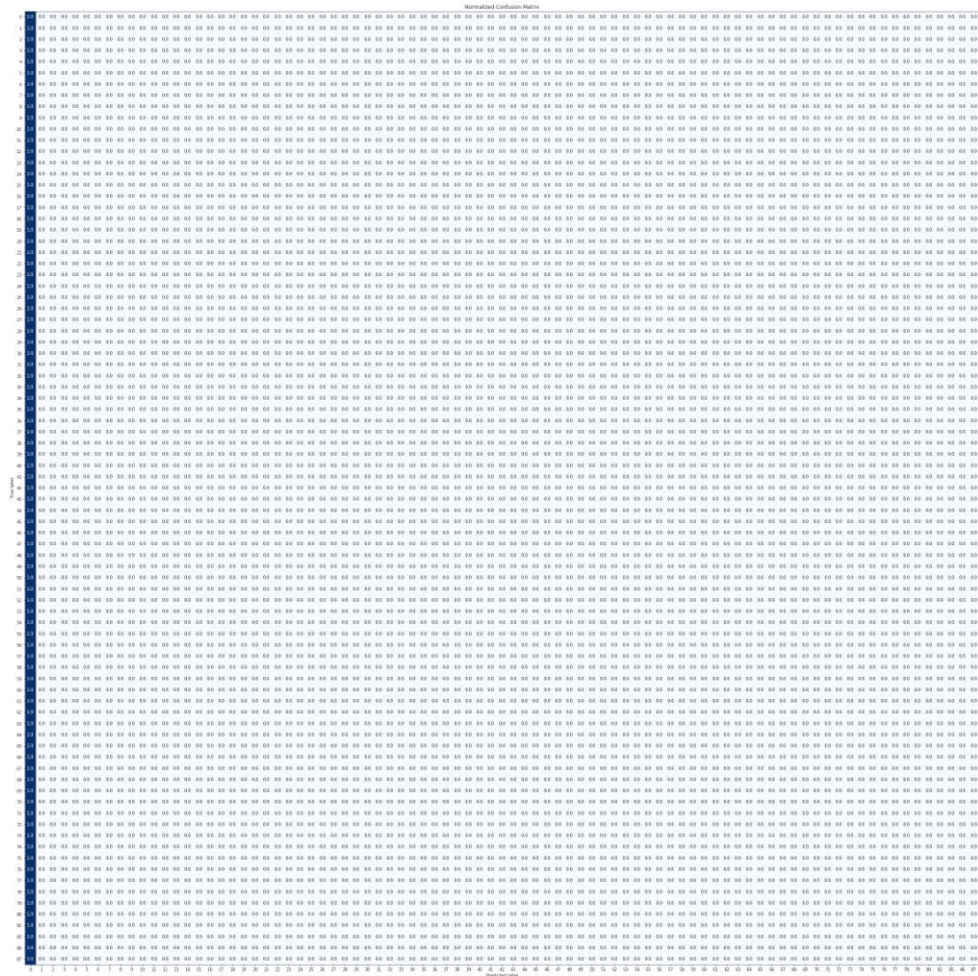


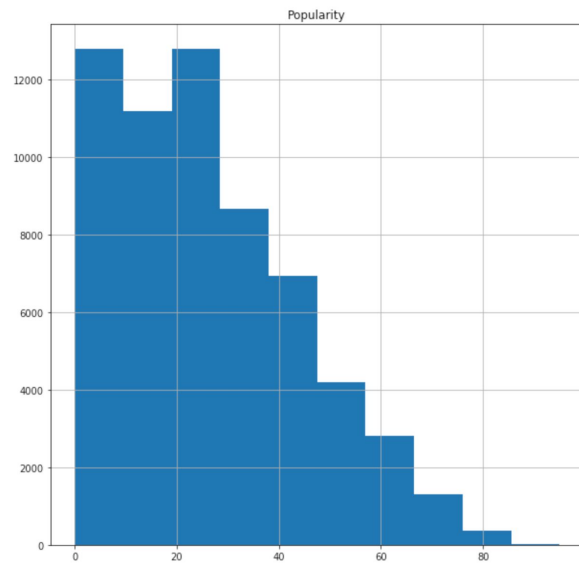
2. Do MLP classifier on the dataset

accuracy score of training: 0.03554827480290775

accuracy score of testing: 0.033906633906633905

accuracy score of validating: 0.03227391874180865





Problems:

1. how to bind the two purpose together:
predict popularity in spotify \

corona virus **model** / _____ how the situation change people's listening habit
how do we predict how coronavirus has affected popularity in the top list?

how should we map the rows if we just add a new column into the spotify dataset?

Next goals:

1. merge spotify and corona to do prediction