**NAME :** Luu Gia An  
**ID :** USTHBI7-003

**MACHINE LEARNING**

Report on K-mean Variance

**I.Algorithm:**

-My algorithm is based on the work of 2 authors on this website: <http://www.kpubs.org/article/articleMain.kpubs?articleANo=MTMDCW_2014_v17n10_1160#r027> (Please copy paste)

**II.K-mean same size code:**

-The code is contained in DoIt.m

-Instruction:

+Before running the code, please install package IO in Octave

+Make sure DoIt.m, kMeans.m, Group\_Data\_modified.xlsx is stored in the same folder.

+Run the script.m for the performance.

+The result is stored on cluster\_same\_size\_returned: a 30x5 matrix. Each three columns is the coordinate of the point. All Point (indexed by column) belong to the same cluster with the first 4 columns being the points and the last column be the centroid.

-Notes: because of random initialization of centroids, sometimes the algorithm will “stuck” , however, just run the script.m again and it will work.

**III.Evaluate the algorithm:**

-Cost: 2-2.5; (evaluating the (1/m)\*sum of Euclidean distance squared)

-