CIS 635 Data Mining Name \_Kritika Sijapati\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Homework 9

Hand in sheet

Write in or copy and paste the answers for the following:

### Part 1

Descriptions for each cluster:

1. The first cluster has the youngest group of people whose annual spending on housing seems to be very close towards the lower range in that category.
2. People belonging to the second cluster have the least annual spending on housing; could probably be backed up by the fact that they are the rather younger group of people.
3. The third cluster has people who spend the most on food and transport, both over 6000 annually.
4. The fourth cluster people are the ones spending the highest amount on housing annually despite having average age of 35 years.
5. The fifth cluster people seem to spend the least on media while spending rather higher amount on pets.
6. The sixth cluster has people with the most versatile attributes; they spend the most on media while the least on pets and are younger aged group people.
7. Cluster seven has the youngest group of people.
8. Cluster eight has people with the highest average age and who spend the most amount on pets annually.

### Part 2

Paste your code for the main kmeans function (only) below:

>>

data\_matrix=read.table("hw09data1.txt", header=TRUE)

my\_Kmeans=function(data\_matrix,cluster\_num) {

set.seed(300)

temp\_var=0

initial\_centers=initialCentroids(data\_matrix,cluster\_num)

org\_clusters=assignClust(data\_matrix,initial\_centers)

new\_centers=initial\_centers

while (sum(temp\_var)!=sum(new\_centers)){

temp\_var=new\_centers

print(temp\_var)

new\_clusters=assignClust(data\_matrix,new\_centers)

sse=sse(data\_matrix,temp\_var,new\_clusters)

new\_centers=calcCent(data\_matrix,new\_clusters)

print(sse)

}

return(new\_centers)

}

my\_Kmeans(data\_matrix,8)