Clustering – practice list

1. Download data files: s1.txt, s2.txt, s3.txt, s4.txt and check their structure. These files contains data of points in Cartesian coordinate system.
2. Write a parser. As an effect you should have a data stored in some accessible way for further analysis.
3. What are the differences between data saved in each file? How it will affect clustering process.
4. Use K-means centroid algorithm to cluster points for each of 4 cases. Compare effects of using different parameters like initialization etc.
5. Show the effect of clustering including centroids location. Analyse the outcome as a human. What do you think about the effects? Compare clustering quality between 4 cases. What are the differences between data in each of 4 cases?
6. Check also the spiral.txt. Change your parser to be able to read this file. Use K-means on that data assuming there are 3 clusters. Show the results of clustering and also show the dataset assuming that third column have information about clusters. Compare effects and explain.