**Homework 4**

**Instructions:**

The data file for this homework is Cereals.xls, which is to be downloaded from Canvas. Create a new Word document and save it as HW4Answers\_X (where X is your team number). Where required, write your answers or paste screenshots into this Word document. Your response should not exceed 100 words for each below question. Write every member’s full name and participation on the first page of the Word document as follows. You need to submit this Word document and XLMiner solution.

|  |  |  |  |
| --- | --- | --- | --- |
| Participant | Complete the Assignment before the Meeting (Y/N) | Percentage of Contribution | Justification |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

The dataset Cereals.xls includes nutritional information, store display, and consumer ratings for 77 breakfast cereals.

Tasks:

1. Data Preprocessing. Remove all cereals with missing values.
2. Apply hierarchical clustering to the data using Euclidean distance to the normalized measurements. Compare the dendrograms from single linkage and complete linkage, and look at cluster centroids. Comment on the structure of the clusters and on their stability. Hints: (1) To obtain cluster centroids for hierarchical clustering, apply Excel’s Pivot Table to the “Predicted Clusters” table. (2) Running hierarchical clustering in XLMiner is an iterative process – run it once with a guess at the right number of clusters, then run it again after looking at the dendrogram, adjusting the number of clusters if needed.
3. Which method leads to the most insightful or meaningful clusters?
4. Choose one of the methods. How many clusters would you use? What distance is used for this cutoff? (Look at the dendrogram.)
5. The elementary public schools would like to choose a set of cereals to include in their daily cafeterias. Every day a different cereal is offered, but all cereals should support a healthy diet. For this goal you are requested to find a cluster of “healthy cereals”. Should the data be normalized? If not, how should they be used in the cluster analysis?

**Important submission instructions**

Save your Word file and XLMiner solution. Use the link “Homework 4” to upload these files. **Due by 11.59 P.M. April 15, 2019.**