**Assignment #9: Clustering Using R**

|  |
| --- |
| **Submission Instructions**   * Submit the following **five** files through the form:  1. The completed, working **R script** that produced the analysis for the **15 cluster** scenario. 2. The **three output files**: “ClusteringOutput.txt” “ClusteringPlots.pdf” and “ClusterContents.csv” for the **15 cluster** scenario. 3. The completed **answer sheet** provided on the last page.   **Evaluation**  Your submission will be graded based on the correctness of the completed answer sheet, with other files as supporting documents. |

**Before you start**

For this assignment, you’ll be working with the **Jeans.csv** file and the **Clustering.r** script (which we used in ICA #13). This file has data from 689 stores that sell four different types of jeans: leisure, fashion, stretch, and original. The marketing division of the company wants to identify groups of stores that sell a similar mix of product so that they can roll out promotions specific to those stores.

The data file contains the following fields:

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| **StoreID** | Store identification number |
| **Fashion** | The number of pairs of “fashion” style jeans sold last month |
| **Leisure** | The number of pairs of “leisure” style jeans sold last month |
| **Stretch** | The number of pairs of “stretch” style jeans sold last month |
| **Original** | The number of pairs of “original” style jeans sold last month |
| **TotalSold** | The total number of jeans sold last month |

**Guidelines**

1. You’ll need to modify the **Clustering.r** script from ICA #13 with the following information to perform the analysis:

* Set the input filename (INPUT\_FILENAME) to the store’s dataset (i.e., “Jeans.csv”).
* **Set the number of clusters to create (NUM\_CLUSTER) to 5**.
* Set the variable list (VAR\_LIST) to use the Fashion, Leisure, Stretch, and Original variables by changing it to the following:

VAR\_LIST <- c("Fashion","Leisure","Stretch","Original")

1. Once you finish modifying the script, you can set the working directory and run the script.
2. Based on your script output, answer Questions 1-5 in the answer sheet at the end of this document.
3. Now rerun the script, this time with **15 clusters**. Then answer Questions 6-11 in the answer sheet.