Answer Sheet for Assignment: Clustering Using R

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Fill in the answersheet below based on the output from R/RStudio:*

|  |  |  |
| --- | --- | --- |
|  | **Question** | **Answer** |
| **5 clusters**  Based on your script output with 5 clusters, answer Questions 1-7 below. | | |
| 1 | Which cluster is the largest (write the number of the cluster)? |  |
| 2 | How many stores are in the largest cluster (i.e. what is the cluster size)? |  |
| 3 | Describe the sales of cluster 1 for each type of jeans (compared to the overall population average across all stores)? (write one or two sentences) |  |
| 4 | Describe the sales of cluster 5 for each type of jeans (compared to the overall population average across all stores)? (write one or two sentences) |  |
| 5 | In which of the five clusters of stores do original jeans sell the best? |  |
| 6 | What is the range of withinss errors (i.e. within-cluster SSE) for the five clusters? | Lowest: \_\_\_\_\_\_\_\_\_\_  Highest: \_\_\_\_\_\_\_\_\_\_ |
| 7 | What is the *average* betweenss error (i.e. average between-cluster SSE) for all five clusters? |  |
| **20 clusters**  Now rerun the script, this time with 20 clusters. Then answer the following questions: | | |
| 8 | Describe the sales of cluster 15 for each type of jeans (compared to the overall average across all stores)? (write one or two sentences) |  |
| 9 | Describe the sales of cluster 20 for each type of jeans (compared to the overall average across all stores)? (write one or two sentences) |  |
| 10 | What is the range of withinss errors for the 20 clusters? | Lowest: \_\_\_\_\_\_\_\_\_\_  Highest: \_\_\_\_\_\_\_\_\_\_ |
| 11 | What is the *average* betweenss error for all 20 clusters? |  |
| **5 Clusters versus 20 Clusters** | | |
| 12 | Which scenario (5 clusters or 20 clusters) produces clusters with better cohesion? |  |
| 13 | Which scenario (5 clusters or 20 clusters) produces clusters with better separation? |  |
| 14 | Besides cohesion and separation, what other advantage does the 5 cluster scenario have over the 20 cluster scenario? (write one or two sentences) |  |