Answer Sheet for Assignment: Clustering Using R

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

|  |  |  |
| --- | --- | --- |
|  | **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)?  How many stores are in the largest cluster (i.e. what is the cluster size)? | Cluster 5  151 stores |
| 2 | Describe the average sales of cluster 1 for each type of jeans (compared to the overall population average across all stores)? (write one or two sentences) | Stores in cluster 1 sell **less** fashion and original jeans than population average, and sell **more** leisure and stretch jeans than population average |
| 3 | In which of the 5 clusters of stores do fashion jeans sell the best, on average? | Cluster 3 |
| 4 | What is the range of withinss errors (i.e. within-cluster SSE) for the 5 clusters? | Lowest: 142.59  Highest: 273.70 |
| 5 | What is the *average* betweenss error (i.e. average between-cluster SSE) for all 5 clusters? | 266.56 |
| **15 clusters**  Now rerun the script, this time with 15 clusters. Then answer the following questions: | | |
| 6 | Describe the average sales of cluster 1 for each type of jeans (compared to the overall average across all stores)? (write one or two sentences) | Stores in cluster 1 sell **less** fashion, leisure, and original jeans lower than population average, and sell **more** stretch jeans than population average |
| 7 | In which of the 15 clusters of stores do fashion jeans sell the best, on average? | Cluster 15 |
| 8 | What is the range of withinss errors for the 15 clusters? | Lowest: 4.49  Highest: 68.70 |
| 9 | What is the *average* betweenss error for all 15 clusters? | 124.63 |
| **5 Clusters versus 15 Clusters** | | |
| 10 | Which scenario (5 clusters or 15 clusters) produces clusters with better cohesion? | **The 15 cluster scenario has better cohesion** |
| 11 | Which scenario (5 clusters or 15 clusters) produces clusters with better separation? | **The 5 cluster scenario has better separation** |