SEIS 632 Data Analytics and Visualization
Assignment 3
Due 10/7/21

Conducting Cluster Analysis

The **DUNGAREE** data set gives the number of pairs of four different types of dungarees sold at stores over a specific time period. Each row represents an individual store. There are six columns in the data set. One column is the store identification number, and the remaining columns contain the number of pairs of each type of jeans sold.

Name	Model Role	Measurement Level	Description
STOREID	ID	Nominal	Identification number of the store
FASHION	Input	Interval	Number of pairs of fashion jeans sold at the store
LEISURE	Input	Interval	Number of pairs of leisure jeans sold at the store
STRETCH	Input	Interval	Number of pairs of stretch jeans sold at the store
ORIGINAL	Input	Interval	Number of pairs of original jeans sold at the store
SALESTOT	Rejected	Interval	Total number of pairs of jeans sold (the sum of **FASHION**, **LEISURE**, **STRETCH**, and **ORIGINAL**)

- Create a new diagram in your project. Name the diagram **Jeans**.
- Define the data set **DUNGAREE** as a data source.
- Determine whether the model roles and measurement levels assigned to the variables are appropriate.
- By examining the distribution of the variables make sure that there are no unusual data values or missing values.
- Assign the variable **STOREID** the model role ID and the variable **SALESTOT** the model role Rejected. Make sure that the remaining variables have the Input model role and the Interval measurement level.
- Add the Data Source node to the diagram workspace.
- Add a **Cluster** node to the diagram workspace and connect it to the Data Source node.
- Select the **Cluster** node. Leave the default setting as **Internal Standardization** ⇨ **Standardization**.
- Run the diagram from the Cluster node and examine the results.

Question 1: How many clusters do you get?

Specify a maximum of six clusters and rerun the Cluster node.

- Use the Segment Profile node to summarize the nature of the clusters.

Question 2: Evaluate the output from the Segment Profile and explain the results in detail using the blue and red histograms and variable worth.

Submission:

1) Answers to the questions asked above with justification.

2) In the diagram, as a last node, add a Reporter node from the Utility tab. Change the Nodes property of the Reporter node to All. Now right click on the Reporter node and select Run. This will generate a pdf.

You should submit the above 2 files on Canvas.