SAS - CLUSTERING

Use “powerusagel.xlsx” dataset. This dataset contains measurements of electric power consumption in one household with a one-minute sampling rate for a day of a typical household.

This dataset has 5 attributes as described below:

1. **Time** in format hh:mm:ss
2. **global_active_power**: household global minute-averaged active power (in kilowatt)
3. **global_reactive_power**: household global minute-averaged reactive power (in kilowatt)
4. **voltage**: minute-averaged voltage (in volt)
5. **global_intensity**: household global minute-averaged current intensity (in ampere)

Using this data, user desires to divide data into different clusters to find further insights about power consumption statistics.

a) Perform K-means clustering on powerusage data with ***k=6*** between global_active_power, global_reactive_power and global_intensity variables

b) How many iterations are performed? What are initial and final centroids for Cluster 2?

c) Report “Cubic Clustering Criterion” of this clustering? Does this value suggest a good quality Clustering? If not, suggest what can be done to improve the quality of clustering with proper reasoning.

d) Perform K-means clustering with K=6 (after steps suggested in c) and comment on the quality of clustering.