

CIND 119
Introduction to Big Data Analytics
Lab 4

1. Download **iris.csv** from your D2L course shell, under Course Materials, in the “Lab CSV Files” folder. Complete the following tasks:
 - a. Read the file in SAS and display the contents using **import** and **print** procedures.
 - b. Explore the data properties using a **scatter plot** procedure.
2. Using the IRIS dataset, complete the following tasks:
 - a. Apply standardization to your dataset using **stdize** procedure.
 - b. Apply k-means clustering using **fastclus** procedure of SAS. **Scatter plot** your cluster labels to visualize and compare with the original data labels. Assuming that you do not know the exact number of clusters in the dataset, try k=2, 3, 4, 5 and evaluate the solutions.
 - c. Open **Results View** to evaluate the solution by examining the summary of clustering results.

End of CIND119 Lab 4

Ceni Babaoglu, PhD