**Assignment 7**

**Github:** [**https://github.com/DouglasBui/GCU/tree/main/DSC-540/Assignment7**](https://github.com/DouglasBui/GCU/tree/main/DSC-540/Assignment7)

This assignment mixes theory and application, in the form a complex challenge. Perform the tasks described in each. Note that this (and other) assignments includes a few challenging research-related tasks. They are aimed at gradually building your capacity to tackle complex topics, familiarize yourself with academic discourse, and provide context and practice for the skills you will eventually need when working on your capstone thesis or project.

More specifically, this assignment challenges you to think about interesting questions to ask about data, taking you one step further to becoming a data scientist.

Access the "DSC-540 Data Sets" document found in the class resources, focusing on data sets best suited for unsupervised cluster detection.

Select one of the 90+ datasets as the data source for your project.

Familiarize yourself with the KMeans package in Python and its use in a Jupyter notebook by utilizing the Learn KMeans resource within the topic resources.

Referring to the chosen dataset, formulate three questions worth asking about the data, and explain why it is important and beneficial to ask these questions. Ask yourself: “What can I learn by grouping similar data points together and discovering underlying patterns?”  
While the readings for this topic are comprehensive, you may use the GCU digital library to find additional articles describing the use of kmeans.

Perform a Kmeans cluster analysis, plot the results, and interpret them.

Use the results above and answer the questions you previously formulated. For each question write the answer mathematical/quantitative terms. Explain the results and the meaning of the patterns you uncovered. Use additional plots to support your arguments.

Since you are accessing a public data repository, discuss the ethical aspects of data management, such as privacy, archiving, and responsible data use.

APA style is expected, as well as formal and rigorous scientific writing, using appropriate mathematical notation and references.

This assignment uses a rubric. Review the rubric prior to beginning the assignment to become familiar with the expectations for successful completion.  
  
You are not required to submit this assignment to LopesWrite.