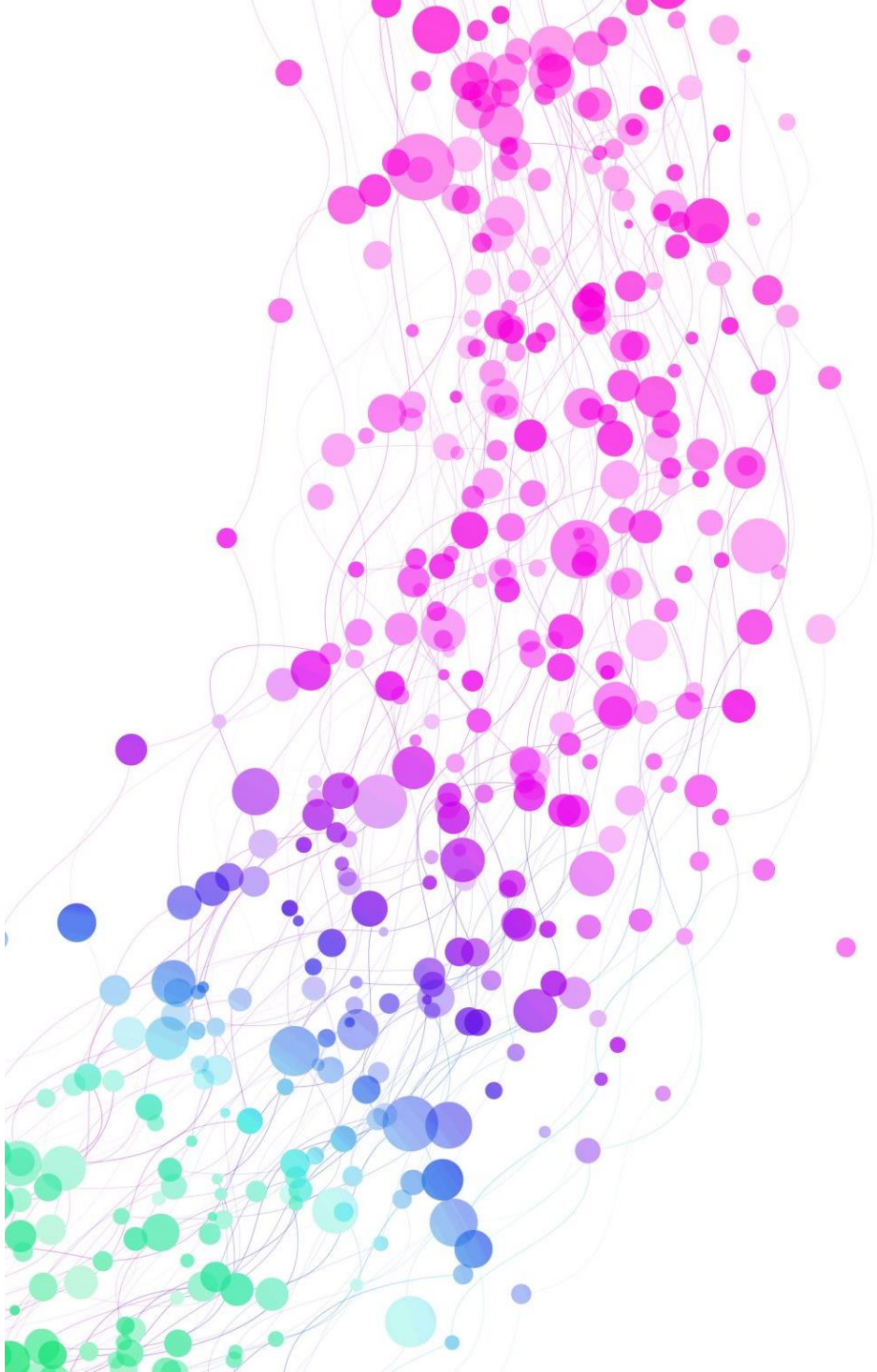
Competitiveness of Local Government Units

ASSIGNMENT #6 – EXERCISE FOR CLUSTER ANALYSIS



Background

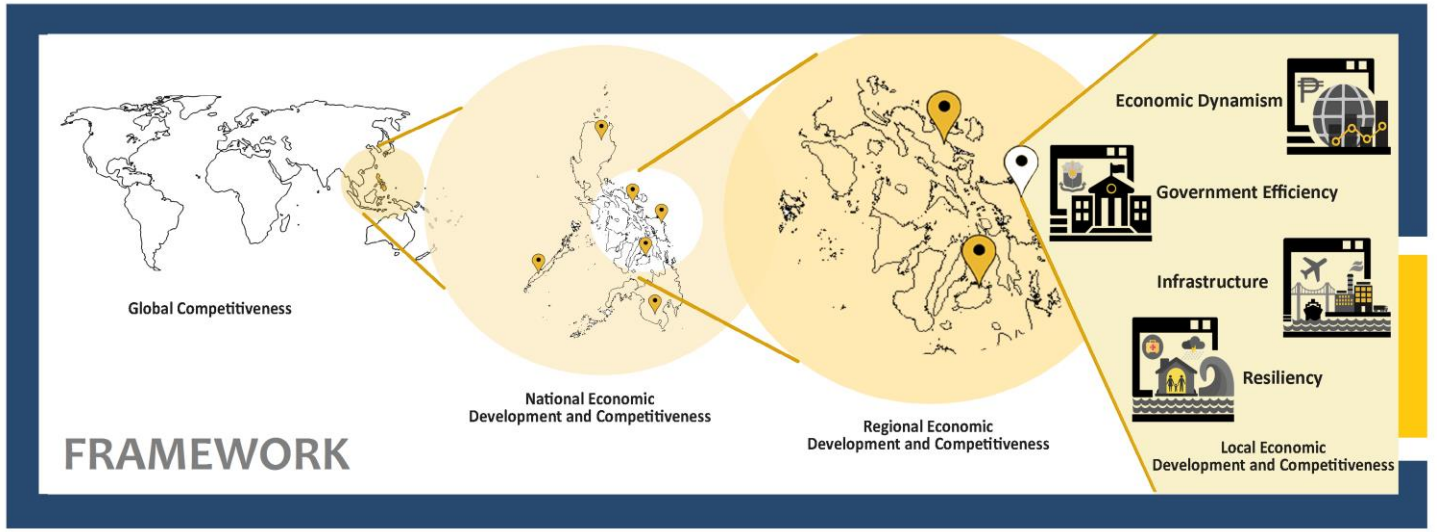
DTI commissioned a study to develop a competitiveness index for cities/municipalities of the Philippines. The study identified four pillars of competitiveness. These are economic dynamism, government efficiency, infrastructure, and resiliency. This year, a fifth pillar was introduced – "innovation". We will not yet include innovation in our assignment.

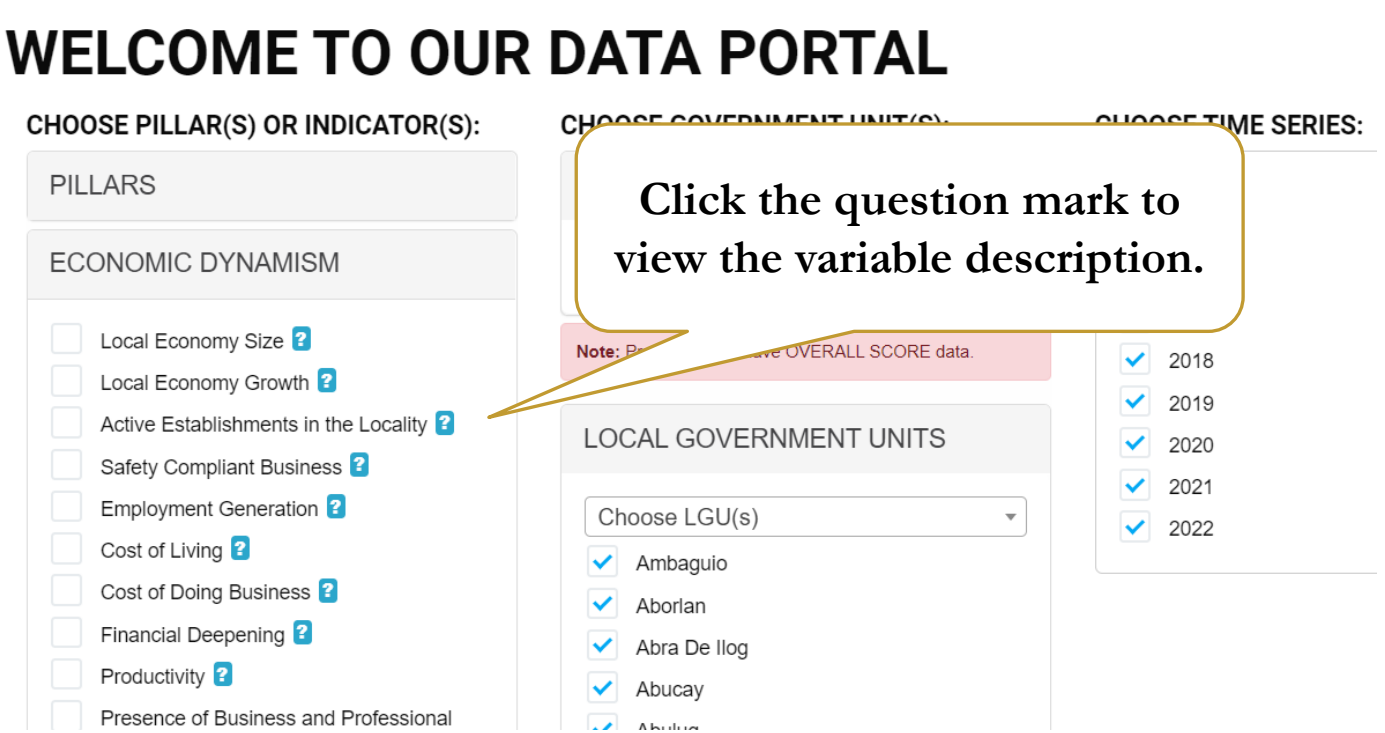
For each pillar, there are 10 identified indicators which are measured through private and public efforts. The value of each indicator were normalized such 100 is given to the observed maximum and 0 is given to the observed minimum. Then, the score for each pillar is computed by taking the weighted sum of the normalized indicators. Each indicators has a weight of 2.5% (resulting to each pillar having 25%). **In this assignment, your task is to cluster similar LGUs with respect to their scores in the four pillars of competence.**

Variables

For the details of each indicator/pillar, kindly refer to this link: ***https://cmci.dti.gov.ph/data-portal.php***

Please note that you may expand the analysis of this data in your research paper. Last academic year, data about innovation is not yet available.





Tasks and questions

1. For this task, we shall limit our method to the ward's linkage method on standardized values. A table of validation metrics is already provided. **What is the best number of clusters that you can form based on the various metrics presented? Explain your answer. (Don't forget to check the distribution of observations over the clusters)**

2. Conduct the K-means clustering and utilize the "best" clustering from hierarchical method as a starting point. Please note that cluster assignments for the different possible clustering solutions are already available in the excel file. **Does K-means result in better clustering of observations? Explain your answer.**

Tasks and questions

3. Identify your final clustering solution and describe each cluster. Report appropriate descriptive statistics and use visualizations to effectively differentiate the clusters. It is always useful to state the size of each cluster. **How will you describe and label each cluster?**

**Don't forget to bring a copy of your data and outputs during our workshop sessions. We'll also analyze other interesting insights. We did not remove outliers in this analysis, so that's another procedure that can be added if you wish to work on this data for your research paper*

Required outputs and deadline

- Required outputs:
  - Answers to each question in PowerPoint format.
  - Codes for the K-means clustering part and descriptive statistics.
  - Submit outputs via UVLe. This is a group assignment.
- Deadline:
  - 9 December 2022, 10:00 AM.