Rebecca Ubalde

Professor Kling

GEOG-485

3 October 2023

Final Project Proposal

**GEOG-485 Final Project Proposal:**

**Performing Density-Based Clustering on a Given Dataset**

For my final project, the task I intend to accomplish with a Python script is to perform density-based clustering (spatial analysis) on a given dataset, specifically focusing on identifying hotspots for botanical seed collections or rare plant species. Density-based clustering is a useful spatial analysis technique that helps identify and visualize regions with higher concentrations of points based on a specified density threshold. I aim to identify areas where seed collections or rare plant species occur in high concentrations, which will then indicate potential priority areas for future seed collections or habitat conservation efforts. I plan to utilize and reference ArcGIS Pro’s Mapping Clusters toolset, specifically the [Density-based Clustering](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/densitybasedclustering.htm), throughout this endeavor.

My proposed solution will enhance efficiency and accuracy by automating the entire spatial analysis process using a Python script, which will save time and reduce human error. A Python script also has an advantage over “out-of-the-box” tools because the parameters can be further customized based on each specific dataset and research needs. Finally, both the script and script tool allow for flexibility and usability; thus, the same analysis or repeat analyses can be performed on different datasets without needing to recreate the entire workflow.

For my final project deliverables, I aim to deliver a Python script that performs the density-based clustering analysis using a given dataset. I also aim to provide a script tool with input parameters and tooltips that can be accessed from ArcGIS Pro. Finally, I plan to deliver a sample dataset that the instructor can use to test the script and script tool.

Overall, the data and processing steps will include (but may not be limited to) the following:

1. **Data cleanup and preparation** – create a clean dataset with the seed collection points or rare plant points with the necessary attributes (i.e., geographic coordinates and species names).
2. **Configuring the tool parameters and deciding on the clustering method** – defining the relevant clustering parameters, such as search distance or search time interval, and choosing which cluster method to use in the analysis
3. **Analyzing the cluster results** – analyze the results of the clusters, identify density hotspots, and understand occurrence and spatial patterns
4. **Saving the output results** – generate and save the output data into a geodatabase