Identifying New York Counties with Food Insecurity using Rust Graphs

In this project, I utilized the US Food Access Research Atlas from the USDA to create a derivative dataset for counties in New York. The dataset contains census tracts for each county and columns that record demographic features, population, number of families without supermarket access within a mile, number of families using SNAP, and whether the tract is urban or rural. In this write-up, I will delve into my methodology, results, and conclusions.

Census tracts are small, relatively permanent statistical subdivisions of a county that are defined by the US Census Bureau. They are designed to be relatively homogeneous with respect to population characteristics, economic status, and living conditions. Census tracts are typically used to study and track population demographics and geographic data. In this project, I used census tracts as nodes to create a graph representation of the data. To create the graph, the program first reads the CSV file and stores the data in a vector of vectors. It then creates a HashMap to store the nodes and iterates through each record in the data. For each record, the program calculates the food insecurity score for the node and creates a new node using the census tract ID as the key in the HashMap. The program then adds each node to the Hashmap.

To create nodes and edges, I defined each census tract as a node and included features such as food insecurity score(discussed below), county, urban or rural status, and demographic information. I felt as though these features made the most sense since accessibility to food is lower in rural regions, poverty level is a good indicator of food insecurity (due to financial instability), and not having a supermarket near your home makes it more difficult to access food. I then created edges between nodes based on geographical proximity, shared demographics, and other relevant factors. For instance, nodes that share the same county and urban/rural status as well as supermarket access were connected by an edge. Additionally, nodes that were adjacent based on tract ID were also connected by an edge.

Sample output of one node:

Node: 36091062100, Edges: ["36091061401", "36091062508", "36091060702", "36091061200", "36091061800", "36091061404", "36091062501", "36091062503", "36091062406", "36091062405", "36091060400", "36091062200", "36091062506", "36091062700", "36091060102", "36091061100", "36091062300", "36091061702", "36091061302", "36091061403", "36091062509", "36091062403"]

The reasoning behind finding the degree centrality or the most central node based on a food insecurity score was to identify the census tract that would benefit the most from increased food accessibility efforts. I created a food insecurity score by combining factors such as poverty rate, lack of supermarket access, and number of families using SNAP, a governmental food assistance program . The node with the highest degree centrality and food insecurity score is the census tract that is most connected to other tracts in the network and has the highest need for increased food access. In this case, the node with the highest degree centrality was located in Ontario County, New York, which is a rural region that has a 10% poverty rate. Here is the output I received from my Rust code:

Number of Vertices: 1616
Node with the highest degree centrality: 36069050102
County: Ontario County
Degree Centrality: 0.011145510835913313
Food Insecurity Score: 10699.8

The output of the program contained all the nodes and their corresponding edges. Additionally, the output identified the census tract with the highest degree centrality, its degree centrality score, and its food insecurity score. The results revealed that the census tract with the highest degree centrality and food insecurity score was located in Ontario County, New York, which is a rural region. This suggests that increased food access efforts should be prioritized in this area to improve food security for its residents. For future use cases, features from my derivative dataset could be tested to find the most influential features, allowing us to perform k-means clustering to find a cluster of areas that need governmental attention.