## **Mapping Food Insecure Areas (Food Deserts) in Brooklyn – A Synopsis**

According to the USDA, a **food desert** is defined as a census tract that meets both low-income and low-access criteria, including:

1. A poverty rate greater than or equal to 20 percent, **or** median family income not exceeding 80 percent of the statewide (rural/urban) or metro-area (urban) median family income.
2. At least 500 people or 33 percent of the population located more than 1 mile (urban) or 10 miles (rural) from the nearest supermarket or large grocery store.

Our analysis of the Food Access Research Atlas 2019 aimed to identify census tracts that meet this definition of food deserts (LILA zones). However, the dataset did not reveal any census tracts classified as food deserts.

To delve deeper, we explored various sources such as community blog posts, research papers, and news articles to understand how these census tracts are identified and categorized as food or non-food deserts. While the Food Access Research Atlas provided limited insights, other sources pointed us toward key features to consider when classifying a census tract as a food desert. Factors like SNAP benefits, poverty rates, and income levels frequently appeared in areas recognized as food deserts.

To create a comprehensive dataset, we explored the repository of datasets provided on the NaNDA (National Neighborhood Data Archive) website, which included demographic characteristics, socioeconomic characteristics, grocery level, etc along with the Food Access Research Atlas. After experimenting with various combinations of variables, we selected a set of variables to input into clustering algorithms like K-Means, Gaussian Mixture and DB Scan.

**Clustering Algorithms and Model Selection**

The selected variables were normalized within a range of 0-100 before being processed by these algorithms. Among them, DB Scan emerged as the most effective clustering model, with a silhouette score of 0.56.

The variables included in the final model were:

1. **SNAP Benefits**: The proportion of households using SNAP benefits to purchase food.
2. **Population Earning Less Than $40K**.
3. **Proportion of Population with Less Than a High School Diploma**.
4. **Food Index**: A derived variable representing food accessibility.

The Food Index was calculated by combining the number of supermarkets, coffee shops, fast food restaurants, and the poverty rate. We used a weighted average, assigning weights of +0.4 to supermarkets, +0.1 to coffee shops, and -0.5 to fast-food restaurants. These were then combined with the poverty rate to assess healthy food accessibility across Brooklyn's census tracts. The negative weight for fast food restaurants reflects their status as less healthy food options compared to supermarkets and coffee shops.