**The SNAP Gap**

The SNAP Gap, a nonprofit startup, will connect communities with low income and/or low access (LILA) to local farms. They first need to identify the most at-risk areas where they should focus their work. The initial step was to identify features that identify these communities from known data, which is on the county-level. Once the classification model is developed, cities with a high need can be predicted from that model. The initial exploratory data analysis discovered the following areas to be strong predictors of LILA counties: high proportion of minorities, extremely low population, and low household income.

**Design:**

On average, 40% of all food is wasted in the United States each year, while many families struggle to put food on the table. The SNAP Gap wants to close the gap by connecting local farmers with those with low income or low access to healthy food. The major hurdles are identifying communities with a high proportion of LILA residents, farms close to those communities, developing business agreements, and the logistics of implementing the program. The goal of this project is to explore the data to determine features to focus on in future data science models.

**Data:**

The dataset used in this phase of the overall project is The Food Environment Atlas obtained through the USDA’s website that contains data on every county in the United States regarding their population of low income and/or low access residents. The dataset also includes demographics, local farm information, and stores.

**Algorithms:**

1. Added a FIPS (county identifier) to the dataset using an additional dataset and VLOOKUP since the one included in the original dataset did not have a unique number for each county. Combined county and state names to match the datasets.
2. Removed columns that would not be used in the analysis.
3. Found trends in the data using tools within Excel, such as VLOOKUP, pivot tables, and charts.
4. Lastly, I used Tableau to map the data with filters from the trends found in my initial analysis and to create charts that further explained those trends.

**Tools:**

As explained in the algorithms section, I used both Excel and Tableau to analyze, aggregate, and visualize the data. I summarized my findings in a Tableau Story with multiple interactive dashboards.

**Communication:**

I will be presenting my slides and visuals during a presentation on Zoom with my classmates. Afterwards, all of my material will be posted on my personal blog on GitHub.