**ETL Project Write Up**

Emma Szczesiul

Bill White

Ash Mewada

**Project Overview**

**Extract:**

We extracted our data from the following sources:

1. USDA (<https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/>). This comprehensive data source provides information by census tract for the entire United States, including population, income, and demographic information.
2. Data World (<https://www.data.world/cityofchicago/nearby-cook-county-grocery-store-chains>). This data set provides location and information of chain grocery stores in Cook County by zip code.
3. Data World (<https://www.data.world/cityofchicago/nearby-independent-cook-county-grocery-stores>). This data set provides location and information of independent (non-chain) grocery stores in Cook County by zip code, and includes information identifying ethnic specialties, food counter types (deli, butcher, seafood, etc.).

**Transform:**

1. The two CSV files from Data World were first combined into a single CSV using only common columns.
2. Using Jupyter Notebook, the USDA data CSV file was loaded. Following this, the census tracts were converted to zip code so that the data can be merged with the Cook County grocery data sets.
3. All transform work was completed in Jupyter Notebook.
4. Once merging was complete, we re-named and re-ordered the columns into a concise table.

**Load:**

1. Postgres was used to create our database
2. Table was created to load data from Jupyter Notebook into PG Admin
3. Database connection was established
4. Dataframes were loaded, tested and completed.