**Problem Statement**

**Organization Name:** U.S. Department of Agriculture (USDA)

**Dataset Name:** Packaged Meals

**Contact Information:** Ying Li, liyingfe@gmail.comA close up of a logo

Description automatically generated

**Difficulty:**

Level 2: Participants with basic data analysis knowledge.

The problem statement is open-ended yet straightforward. The dataset has a standard structure suitable for beginners. Creative and interdisciplinary solutions are welcomed.

# Background

Over the past several decades, Americans have grown to rely on the convenience of foods prepared outside of the home. While restaurant foods are still popular, consumers have other options, such as pre-packaged frozen or refrigerated meals. There have not been many studies about what are in the pre-packaged meals Americans consume, and what we can learn from these ingredients.

USDA released a branded food product database (BFPD) in 2018. The most recent version is dated Oct 2020 and the database is downloadable from [USDA FoodData Central](https://fdc.nal.usda.gov/index.html). The BFPD is the result of a Public-Private Partnership, whose goal is to enhance public health and the sharing of open data by complementing USDA Food Composition Databases with the nutrient composition of branded foods and private label data provided by the food industry. A subset of data was prepared from the BFPD to include packaged meals from various categories, such as “Frozen Breakfast Sandwiches, Biscuits & Meals”, “Vegetable Based Products / Meals - Not Ready to Eat (Frozen)” and “Ready-Made Combination Meals”. We will use this data for the UMD Data Challenge 2021.

# Questions

This problem is open-ended, so you can pose and answer any question involving the data. For example, you might consider one of the following three questions:

* What are the most popular ingredients in each meal category and for all meals overall?
* What combinations have appeared together most often?
* Which are the healthiest meal options in the restricted database?

Answering the first two questions will help agencies focus their efforts to study nutrient stability over a prolonged storage period. For instance, popular combinations of ingredients can be prioritized for packaged meals. Manufacturers can also benefit by understanding which combinations of food items are popular. They can then create new products with these ingredients or modify their supply chain to provide these combinations of ingredients at minimal cost.

Don’t feel limited to these questions. Come up with better ones on your own!

Data Considerations

Use the portion of the BFPD provided as part of the Data Challenge. You may want to supplement with additional data downloadable from the USDA FoodData Central website (<https://fdc.nal.usda.gov/download-datasets.html>). For instance, the food\_nutrient, nutrient, and fndds\_nutrient\_ingredient\_value tables may all be helpful. These will enable you to compute the nutrients in each meal. Each nutrient is identified by its nutrient\_nbr, also called its nutrient code. Each food item is identified by its fdc\_id number. You may find the full data dictionary (which references additional files found on the USDA website) helpful. It is contained in the pdf titled, Download & API Field Descriptions Oct 2020.pdf