DC21 - Team 21 Project Abstract Allie Cahanin & Katie Toren

In 2012, the World Health Organization defined a methodology known as the Quality of Life assessment to measure an individual's perception of their position in life. One way in which QOL is heavily impacted is proper nutrition. As the UN states in its Sustainable Development Goals, the need to "achieve food security and improved nutrition and promote sustainable agriculture" is paramount. In the US, the USDA is also working, via partnerships, to support improved nutrition by publicly sharing data from its Food Composition Databases.

The goal of this project is to leverage a subset of data from the USDA database to glean insights and make suggestions about sourcing healthy ingredients, providing meal options in support of a balanced diet, and the cultivation of healthier and more sustainable living. This project seeks to discover the most commonly used ingredients in popular meals, combinations of ingredients appearing together commonly, and ingredient nutrition information in order to make the above-mentioned recommendations.

To accomplish these goals, this project will use R to parse, analyze and visualize data regarding packaged meals by combining files which include information about ingredients and nutritional values. Data cleaning will be intensive, involving: parsing and unnesting ingredients and nutrients and using regular expressions and formulas from various R packages. Performing primary component analysis (PCA) on both the nutrient values of and ingredients in packaged meals will reduce the dimensionality of the data sets and enable easier identification of which nutrients and which ingredients contribute the most to overall variance. UMAP visualization of these nutrients and ingredients allows an alternative view to potentially show variable clustering based on similarity. These similar and influential nutrients and ingredients can then be visualized to identify patterns within the overall branded food categories and support additional analyses, visualization, and ultimately the generation of recommendations.

Abstract submission link: https://tinyurl.com/DC21-TeamAbstracts

Our Abstract Outline

Notes

- GitHub
    - Code
- Data cleaning
- PCAs

KT Brainstorming

- Intro/Background:
    - Support of leveraging data to cultivate better living
    - improving the lives of individuals via nutrition
        - perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns
    - Nutrition focus of UN SDG #2: “End hunger, achieve food security and improved nutrition and promote sustainable agriculture”
- Project goal:
    - Answer the question of: What are the most popular ingredients in each meal category and for all meals overall?
        - To help agencies focus their efforts to study nutrient stability over a prolonged storage period
        - 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.
    - Answer the question of: What combinations (ingredients) have appeared together most often?
- Data:
    - 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. 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.
    - Meals + ingredients
    - Nutrition information
- Methods:
    - Data cleaning - allllllll the cleannnnnnnggggggg
    - Summaries/visualizations
        - PCAs
        - Histograms to visualize top ingredients
        - ?? nutrients??
- Results:
    -
- Conclusions:
    -
- Next Steps:
- Our Learnings?

Judging Criteria

Cultivate Better Living

- Project adheres to the theme of Data Exploration to Cultivate Better Living
- World Health Organization Quality of Life (WHOQOL) is understood and acknowledged
- Project has good scope to improve everyday life
- Project demonstrates a carefully articulated path on how the idea can be implemented

Innovation

- Project demonstrates creativity
- Project demonstrates uniqueness
- Project includes additional information to enhance outcome
- Project provides new understandings about the dataset

Quality

- Project is well designed and constructed
- Implementation is relevant and complete enough to show important aspects of the project
- Process of project analysis is logical
- Project presentation demonstrates syntactic and visual clarity and consistency

Results

- Results are clearly stated
- Results address the problem statement
- Results demonstrate an effective use of the dataset
- Results demonstrate a high-quality solution

Presentation

- Project is demonstrated in appropriate detail
- All team members were aware of all aspects of project
- Responses to judges' questions demonstrate knowledge of the problem domain
- Presentation was completed within the time limit provided

From DC21: Packaged Meals (reference)

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. 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