Team: **Nutrition Nuts**

Executive Summary

Project Goal

Using the United States Department of Agriculture (USDA)'s Food and Nutrient Database, create an algorithm where a user inputs a food item, and the program spits out a few suggestions for side dishes that would make a balanced meal.

Data

The USDA's Food and Nutrient Database includes the Standard Reference version 28 (published 2015). This data set includes nutrient values per 100 grams edible portion of food for over 8,000 foods eaten by United States residents. Foods are grouped into 25 groups, such as Grains & Pasta or Baby Foods.

Reference Daily Intakes (RDI)

The Food and Drug Administration publishes recommendations for healthy daily intake levels of nutrients. We used ⅓ of these values as the targets to optimize our algorithm (⅓ represents the nutrient target for one meal).

Methods

As our project's target audience is adults, we removed the Baby Foods group from our data. We also removed the group of spices and herbs, to focus on foods (final n = 7,062 foods). We converted nutrient values per 100 grams of edible portion to nutrient values per common serving size. We limited the features to 25 nutrients for which there is an RDI.

Exploratory data analysis revealed United States residents eat many foods with near-zero values for required nutrients. There wasn't much correlation between the target nutrients. This was confirmed by principal component analysis, as the first principal component explained only 25% of the variance. Fifteen components were required to explain 90% of the variance, and almost 20 required to explain 95%.

We tested k-means and hierarchical clustering on the entire data set of 7,062 foods. Both methods yielded clusters with over 3,000 foods, so foods were grouped into USDA food groups, then clustered within each food group by k-means (which performed slightly better than hierarchical clustering on the whole data set). The result was 145 clusters.

Clusters were paired, and their mean nutrient values summed; however, no two clusters' nutrient profiles met the nutrient targets when summed. Groups of 3 and 4 clusters were compared to the nutrient targets to yield a list of balanced meal combinations.

Result

Using the PyQT package, we created a desktop application that allows the user to input a food item. The input is mapped to the most popular food item in the data that matches the input (e.g. "butter" maps to "BUTTER, WITH SALT"). The application then recommends the most popular foods from 2 or 3 other clusters, when paired with the input, will form a balanced meal. Users can then pick foods from the same clusters.

Implications

With the Nutrition Nuts application, people can plan their meals around their favorite foods. People think a healthy diet means giving up the things they like. The Nutrition Nuts application will show them they can have their cake and eat it, too.