**Project Proposal – Fast-food Nutritional Database**

**Project Members:** Kathy Manthey, Manuela Muñoz, Nicole Muscanell

**Purpose**

The purpose of this project is to extract, transform, and load into a final database information about food nutrition for several national (U.S.) fast food chains. The main question that the database could be used to answer is "What is the nutritional content or value of specific food items at fast food restaurants?". The database could further be used to compare nutritional values between the restaurants in the database. We will focus on three major fast-food restaurants: McDonalds, Subway, and Starbucks.

The final production database will be relational.

**Data Sources**

* Nutrition data for McDonalds: https://www.kaggle.com/mcdonalds/nutrition-facts
* Nutrition data for Subway: https://www.kaggle.com/davinm/subway-restaurant-nutrition-data
* Nutrition data for Starbucks: https://www.kaggle.com/starbucks/starbucks-menu

**Data Cleanup & Analysis**

For the transformation phase, we need to do some basic cleaning so that the files can be linked to each other, in addition to deciding what columns to include/exclude, and perform some joins. Below are some specific steps that our group plans to complete during this process.

Steps [subject to change as project progresses]:

Pre SQL & Pandas Phase:

* Create an ERD
* Decide final scheme for tables/final database
* Add categories to Starbucks foods.

SQL & Pandas Phase:

* Change column names across data sets to fit the scheme.
* Add id to each file.
* Drop caffeine from Starbucks drinks.
* Get rid of precent signs in Starbucks drink file (for last 4 columns: vitamin through iron).
* Add new column that adds together beverage item + beverage prep info together (for Starbucks drinks).
* Join the two Starbucks files together.
* Create a junction table, that links id with restaurant name.
* Finalize ERD (with cardinality)