Course Project – Extracting and Analyzing Data

Shaun Pritchard

Rasmussen College

QMB4100

Darice Corey

May 02, 2021

**Course Project – Extracting and Analyzing Data**

To declare data acquisition for the new proposed digital service product Allrecipes.com Meal planner feature. I will be allocating data from several sources including data extraction using custom python scripts I have personally developed. I will transcend the data to provide the information needed to facilitate the proposal concept. These are the following sources and datasets that I have built and acquired.

Planning meals will allow users to see how much they are actually eating while tracking fundamental requirements of thier diet and health. This also prevents them from overeating at restaurants, helps save money and finances, saves time, avoid wasting food and allows people to maintain many other benefits. This service would provide a user experience to gain all these benefits while facilitating a marketing endpoint to utilize more recipes, engage recipe and content creators, and to capitalize on marketing strategies and gains for Allrecipes.com (Beaumount, 2021).

**Data Sources:**

**Source**: Allrecipes.com

**URL**: https://github.com/shaungt1/actor-allrecipes-scraper.git

**Datatype**: CSV, JSON

For this data set I created my own pipeline and implementation using my own python scripts to web scrap recipe and user data. Most of the data was acquired using Apify, API, docker, Scrapy, Request, Beautiful Soup, and Selenium. This data was extracted for recipe data, URL data, Nutrition data, user data, and ranking data of recipes. With this data I can poll and analyze the top categories of data while cross referencing the nutritional data with data sets acquired from other sources such as Data.gov, Esha Research center (which is partner vendor with All recipes), USDA department of Agriculture Food Composition Database, and food and agriculture Organization (FAO) for dietary information.

**Source**: ESHA Research center API

**URL**: https://esha.com/products/nutrition-database-api/

**Datatype**: CSV, JSON

ESHA Meal Planning & Client Dietary Analysis API is an extensive food database and nutritional standards. This is where Allrecipes.com retrieves data for the recipes submitted to its platform. This is where I will compare meal plan data with the current recipes used on Allrecipes.com. It could also be used for the product digital service moving forward.

**Source**: USDA Department of Agriculture Food Composition Database

**URL**: https://ndb.nal.usda.gov/

**Datatype**: CSV, JSON

USDA Department of Agriculture Food Composition Database contains data for various types of food including the amounts of different vitamins and minerals found in the foods as well as macronutrient percentages. The food covered spans a large variety of foods.

**Source**: Food and Agriculture Organization (FAO)

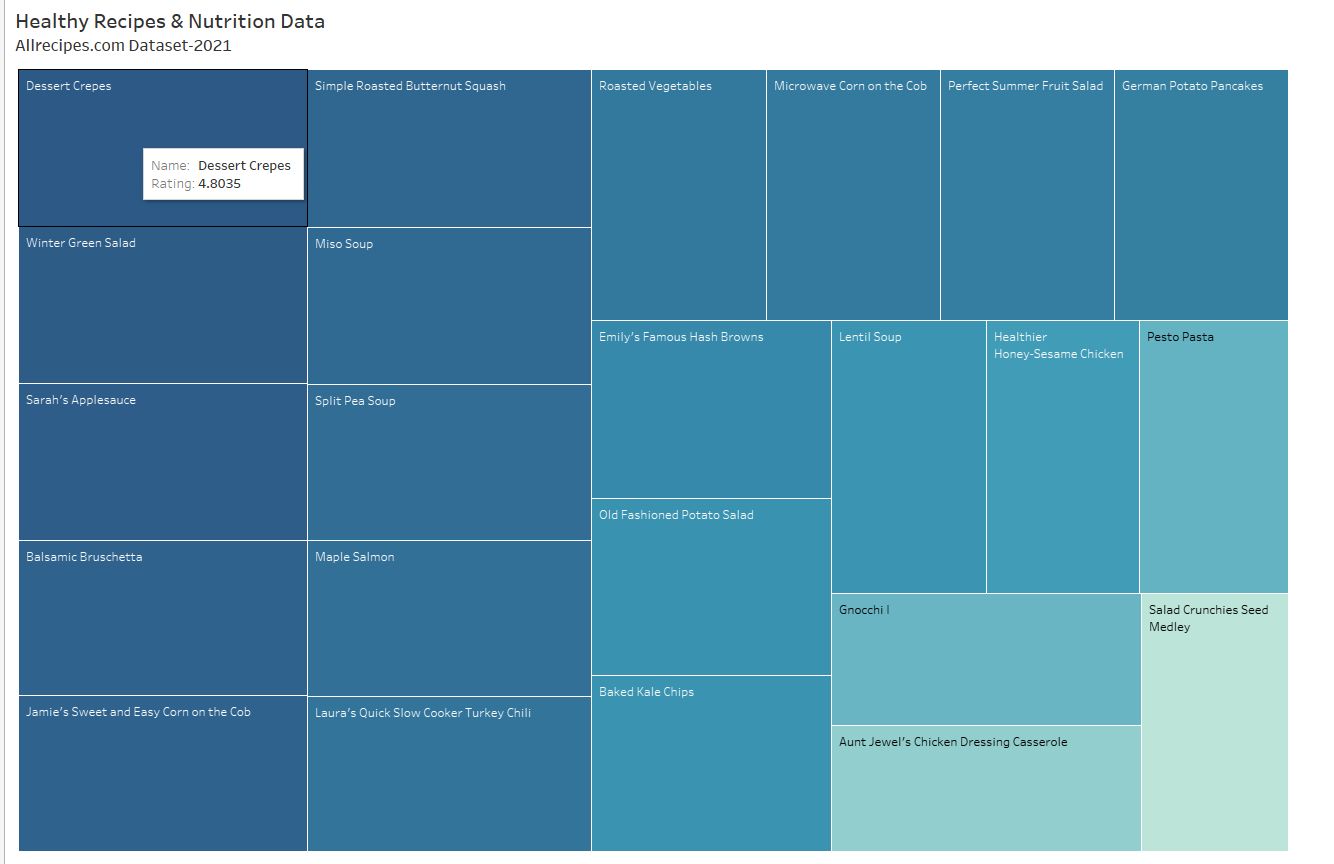
**URL**: http://www.fao.org/faostat/en/#data/FBS

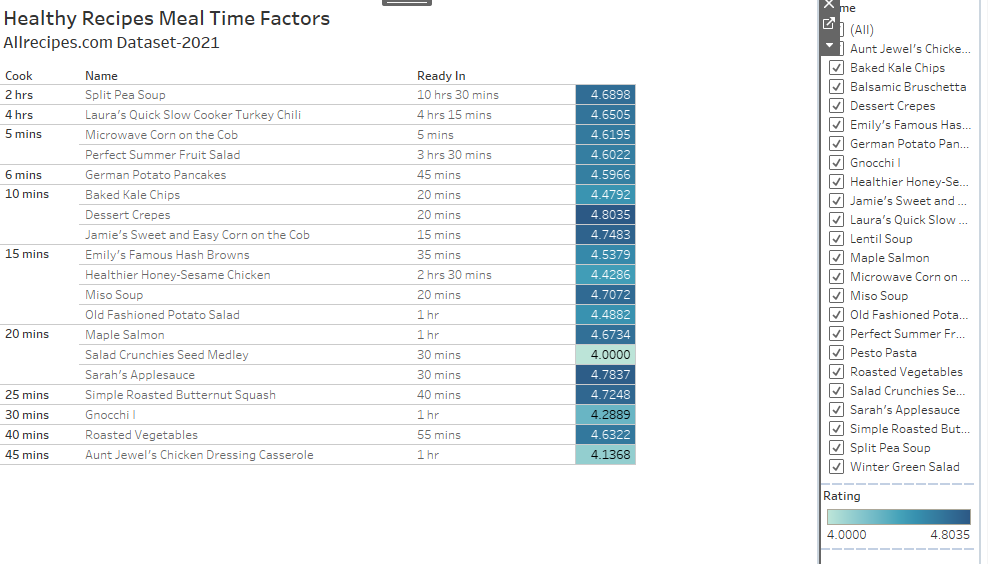
**Datatype**: CSV, JSON XML

This is another credited source of data that I want to use to compare different Mill and nutritional requirements that could be used for the application moving forward.

**Data Snapshots:**

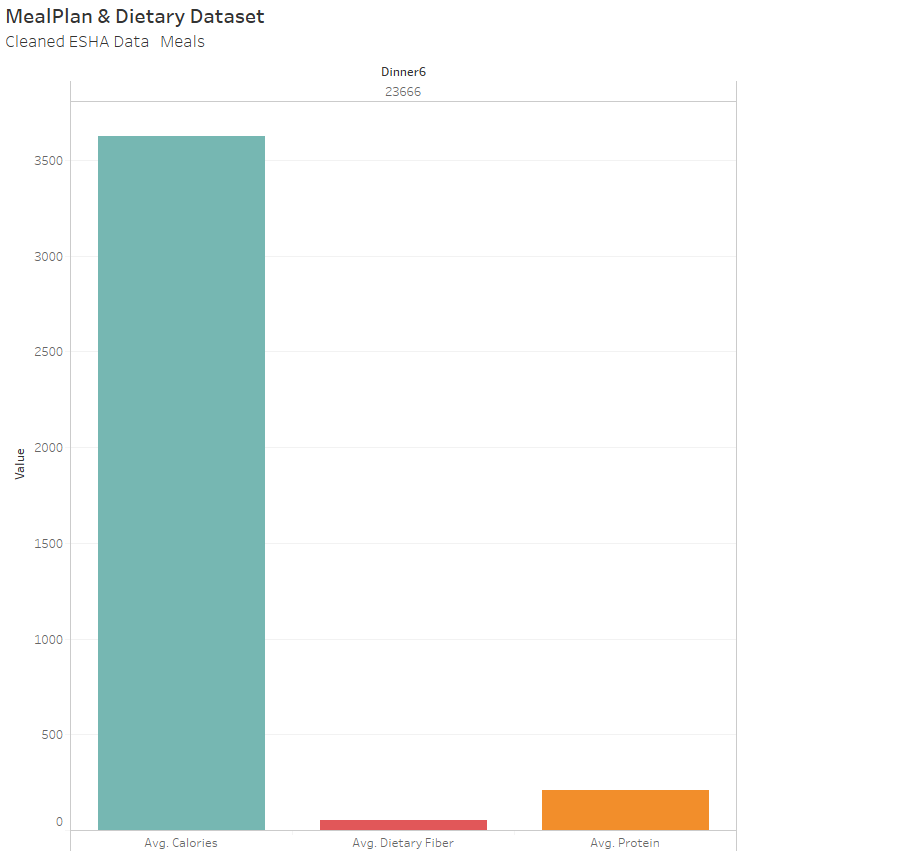
* Allrecipes.com

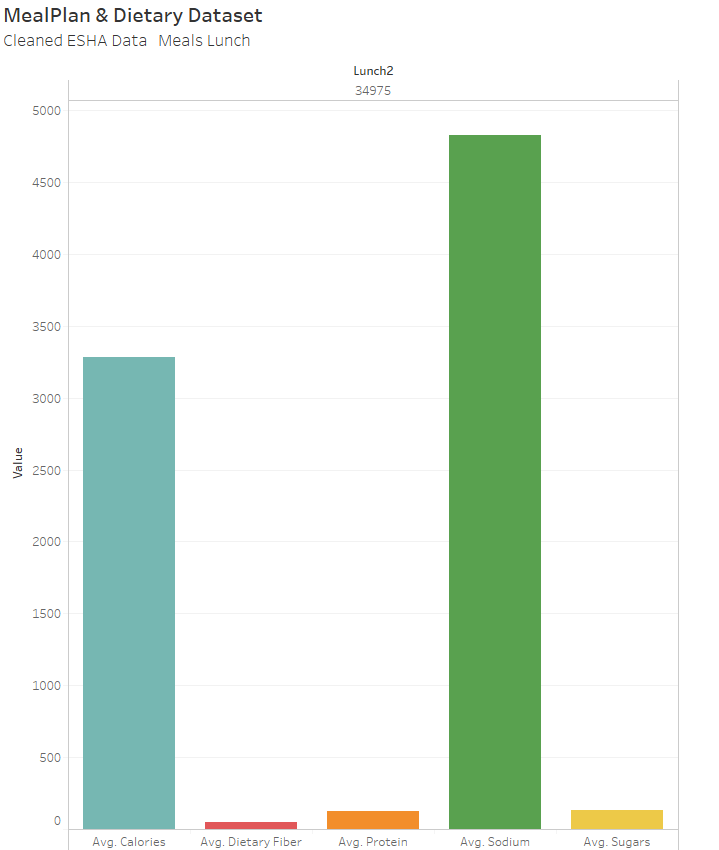




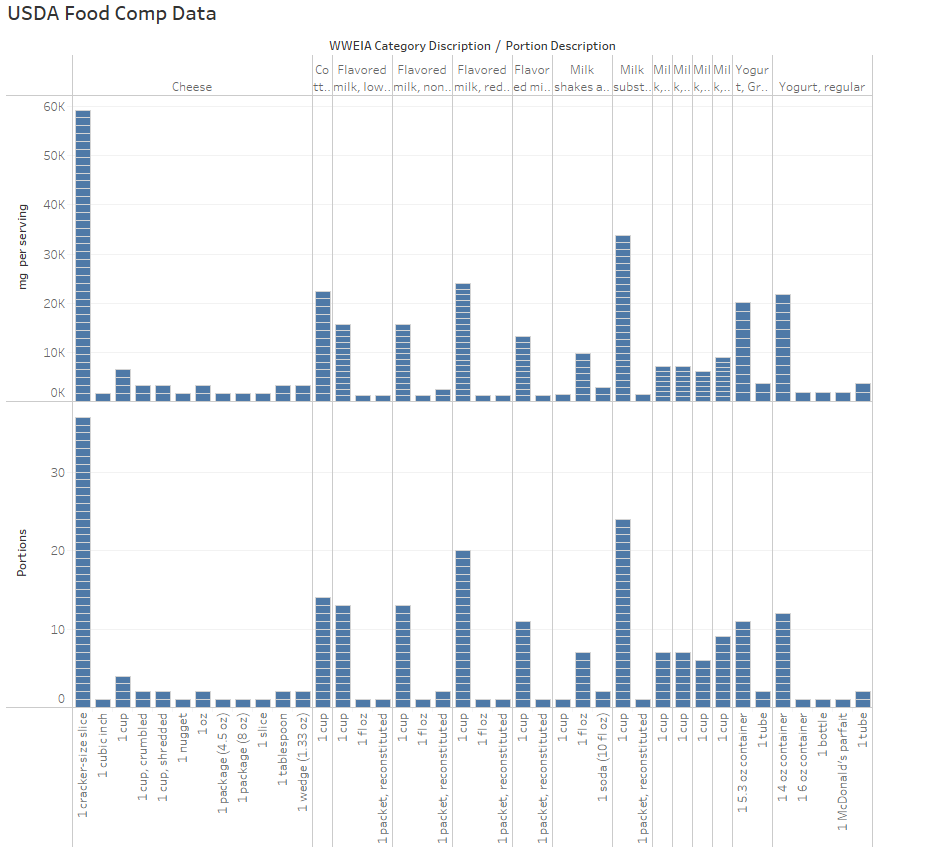


* ESHA Meal Planning & Client Dietary Analysis

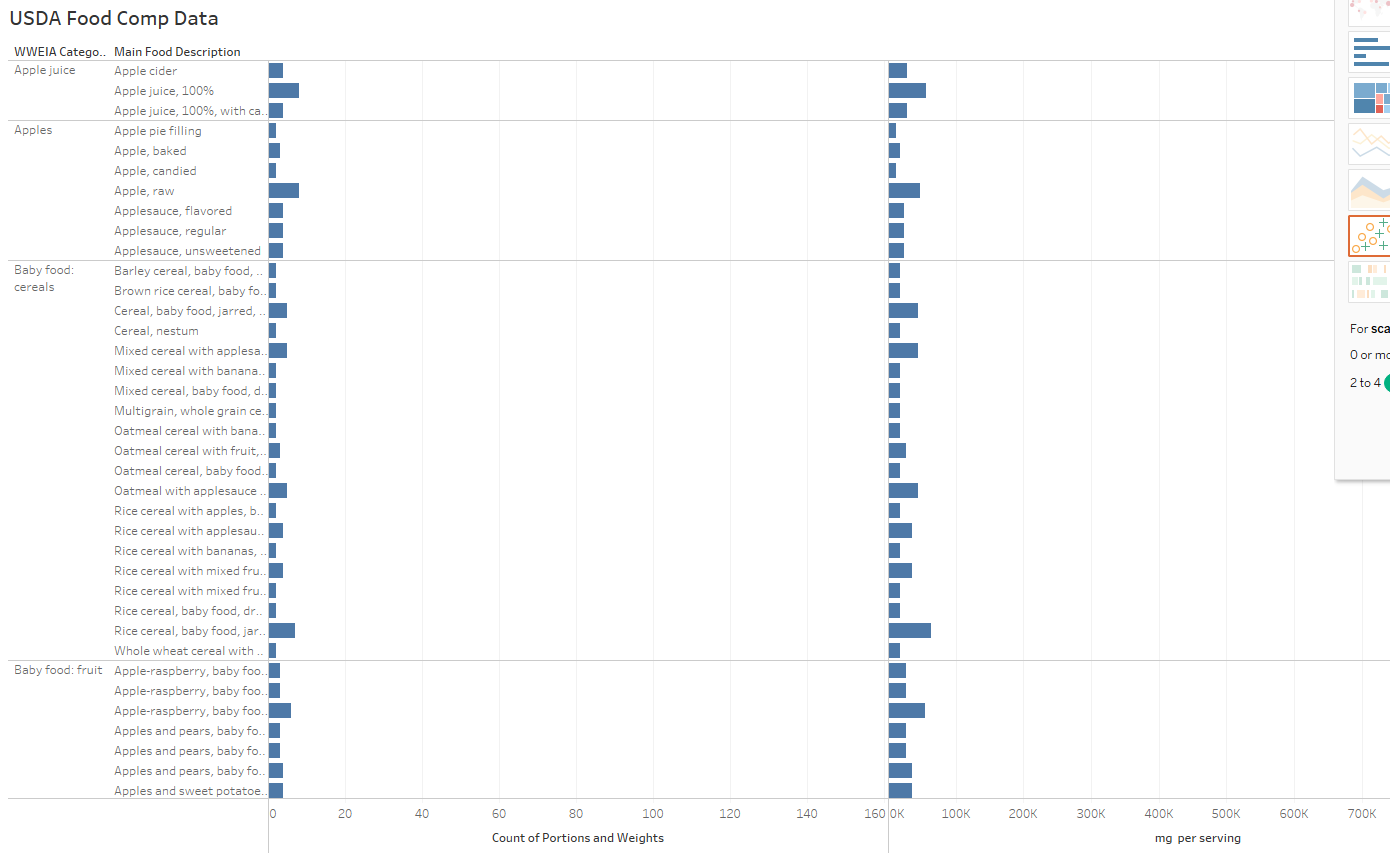


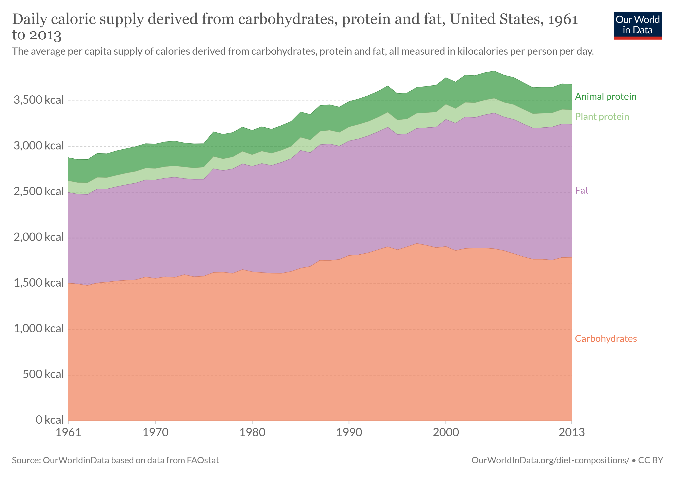
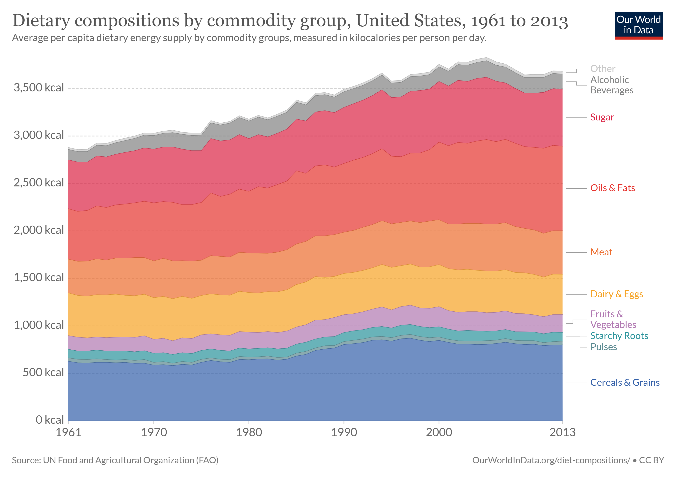


* USDA Department of Agriculture Food Composition Database



* Food and Agriculture Organization (FAO)





**Filters:**

The only filters I applied where NULL filters and filters to limit specific sections of the data set values.