

ISYS 412 - Application Development for Data Analytics

Project Part 1: Dataset Selection and Analysis Plan

Group Members:

- Thomas Cowart
 - Antonio Moore Jr.
 - Evgeni Resh
 - Matthew Riley
-

1. Dataset Information

Dataset Name:

Daily Food & Nutrition Dataset

Dataset Link:

[Kaggle - Daily Food & Nutrition Dataset](#)

Dataset Shape:

- **Columns:** 14
- **Rows:** 10,000

Description of the Data:

This dataset tracks **daily food intake and nutritional values** for individuals, helping in **diet analysis and machine learning applications** related to **health and wellness**. Each record represents a **food item consumed by a user on a specific date**, including its **nutritional breakdown**.

Field Descriptions:

- **Date:** The date when the food item was consumed (YYYY-MM-DD).
- **User_ID:** A unique identifier for each individual to track dietary habits.
- **Food_Item:** The name of the food consumed (e.g., "Apple," "Chicken Breast," "Rice").
- **Category:** The food group classification (Fruits, Vegetables, Dairy, Meat, Grains, Beverages, Snacks).
- **Calories (kcal):** The total energy content of the food item.

- **Protein (g):** The amount of protein in grams, essential for muscle growth and repair.
- **Carbohydrates (g):** The total carbohydrate content, including starches, sugars, and fiber.
- **Fat (g):** The total fat content, including healthy and unhealthy fats.
- **Fiber (g):** The dietary fiber content, supporting digestion and gut health.
- **Sugars (g):** The total sugar content, including natural and added sugars.

Reason for Choosing This Dataset:

We selected this dataset because of its relevance to **nutrition and health**, a field of growing interest in **dietary planning, fitness, and public health**. Several group members have personal interest in **food science, diet optimization, and nutritional analysis**, making this dataset ideal for data-driven exploration.

Personal Reasons for Choosing This Dataset:

- **Antonio Moore:** I love tweaking my diet to gain better nutritional value for less calories, cost, and volume.
- **Thomas Cowart:** I think it would be interesting to see which foods in each category are the most and least healthy and nutrient rich. I think that analyzing this dataset will allow me to gain a more thorough understanding of nutrition and how the foods I eat every day affect me.
- **Evgeni Resh:** It seems to be pretty fascinating to explore dietary trends and values to better understand nutrition.
- **Matthew Riley:** After I finish my Business degree, I am going for a degree in nutrition and dietetics, which has always been an interest of mine.

2. Driving Question(s)

Primary Research Questions:

1. "How do different nutritional factors impact daily calorie intake, and what trends can be observed across different food groups?"
2. "Which food categories contribute the most to overconsumption of nutrients such as sodium, sugar, or fat?"
3. "Can patterns in nutrient intake help identify potentially unhealthy eating behaviors across users?"
4. "How do users' dietary habits align with recommended daily intake levels for essential nutrients?"

These are options for questions that will guide our analysis by helping us uncover **patterns in food consumption, caloric intake, and nutrient distribution**. The goal is to **understand how diet composition varies and what insights can be drawn to improve dietary recommendations**.

3. Specific Analysis Questions

To break down our primary question, we will analyze the dataset using the following key questions:

1. Which food groups contribute the most to daily calorie intake?
2. What is the average daily intake of macronutrients (carbs, proteins, fats) across different food groups?
3. Are there specific food groups that exceed recommended daily sodium or sugar intake?
4. Which nutrients tend to be under-consumed or over-consumed compared to daily recommendations?
5. How do calorie and nutrient intake differ between different meal times (breakfast, lunch, dinner)?
6. Are there any interesting trends in the consumption of certain foods?
7. What are the most and least "healthy" foods in each category?
8. What's the average and median daily caloric intake per user? Are there any outliers?
9. What are the most commonly consumed foods among low, average, and high caloric intake users?
10. Can we predict a user's next meal type based on previous consumption?
11. Can we classify users based on their dietary habits?
12. Can we identify food choices that lead to higher calorie intake over time?
13. Is there a pattern in how users consume sugar throughout the day?
14. Can we detect potential unhealthy eating behaviors from the data?
15. What is the relationship between fiber intake and sugar consumption?
16. Are users meeting recommended daily nutritional intake (e.g., protein or fiber goals)?
17. Does higher sodium intake correlate with higher cholesterol consumption?
18. How does water intake vary by meal type or day of the week?
19. Which users have the most balanced macronutrient intake based on health guidelines?

*(The first five questions will serve as the **primary focus** of our analysis, with the remaining questions providing additional areas of exploration.)*

4. Initial Data Exploration Plan

To ensure a structured approach, our analysis will begin with the following steps:

Data Preparation & Cleaning

- Handle **missing values** and ensure consistency in nutrient measurement units.
- Convert categorical data into structured formats if necessary.

Exploratory Data Analysis (EDA)

- **Summary Statistics:** Calculate **mean, median, and distributions** of key nutritional components.
- **Trend Analysis:** Identify patterns in **food consumption and macronutrient intake**.

Correlation & Relationship Analysis

- Investigate relationships between different nutrients (e.g., **protein vs. calorie intake**).
- Analyze **calorie intake patterns based on food groups and meal times**.

Visualization & Insights

- Use **graphs and charts** to illustrate:
 - **Dietary patterns**
 - **Common food choices**
 - **Nutritional trends**
 - **Excessive or deficient nutrient consumption**

By following this structured approach, we will **ensure meaningful insights from the dataset** and prepare for **further in-depth analysis in later project phases**.