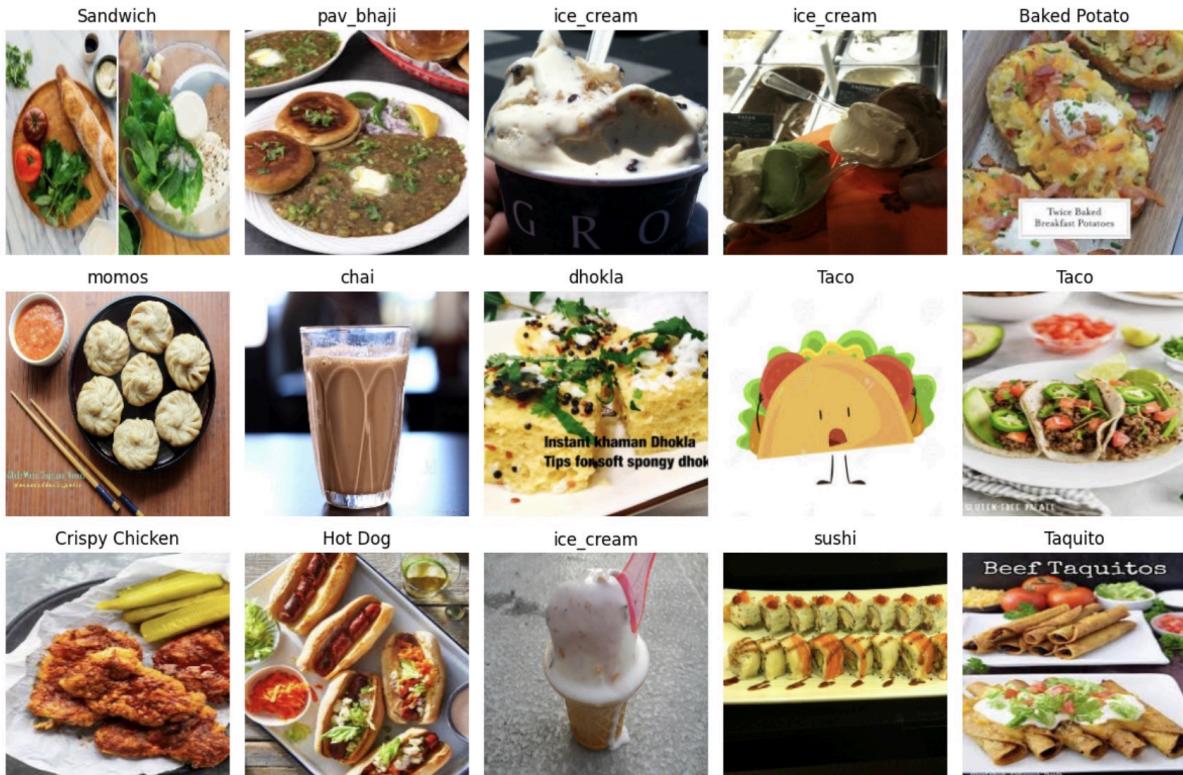


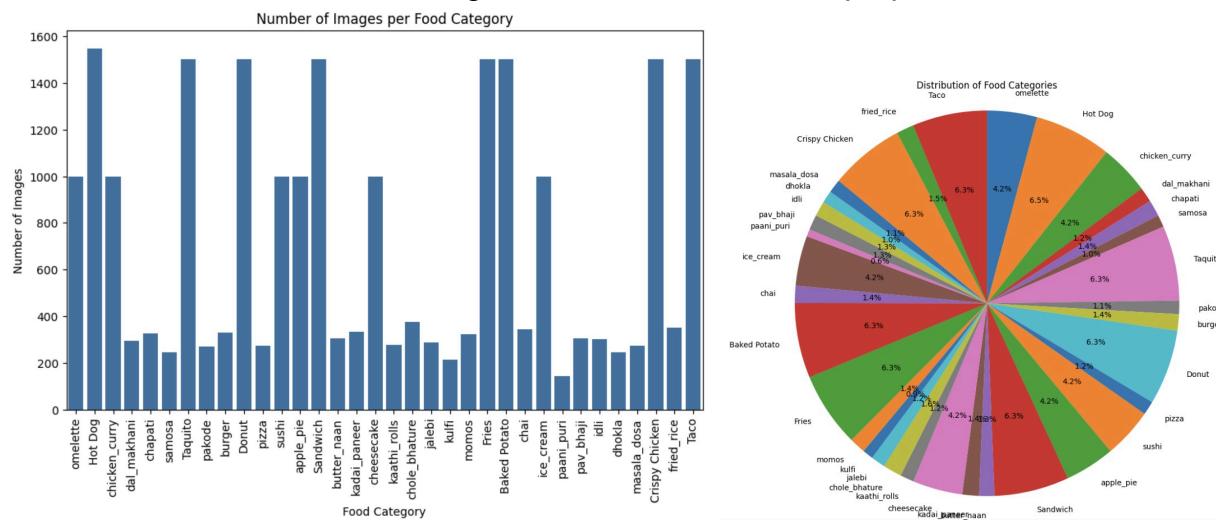
AC215 - SmartEats - EDA

Dataset1 - Food image dataset:

The Food Image Classification Dataset contains 24K unique images in 34 categories obtained from various Google resources. The total size of the dataset is 1.68GB. The images were resized to 224*224 pixels. We show 15 random images from the dataset.



We visualized the number of images of each food and their proportions as follows.



Dataset2 - USDA FoodData Central

USDA FoodData Central contains comprehensive food composition data, including the following data types:

	Foundation Foods	Experimental Foods	Food and Nutrient Database for Dietary Studies (FNDDS)	Branded Foods	SR Legacy
Definition	Data and metadata on individual samples of commodity/commodity-derived minimally processed foods with insights into variability	Data on food published in peer-reviewed journals supported by or in collaboration with USDA	Data on nutrients and portion weights for foods and beverages reported in What We Eat in America, NHANES	Data from labels of national and international branded foods collected by a public-private partnership	Historic data on food components including nutrients derived from analyses, <small>Table comparing the data cal USDA Food Data Central, published literature</small>
Data Source	USDA: based on analytically derived values	Researchers: based on scientific publications	USDA: compiled based on values from FDC data types	Manufacturers: based on food label information	USDA: based on Standard Reference originally available via the USDA National Nutrient Database (NNDB)
Update Frequency	April and October of each year	April and October of each year as data are available	Every two years in concert with WWEIA, NHANES release	Monthly	Final release April 2018

(from <https://fdc.nal.usda.gov/data-documentation.html>)

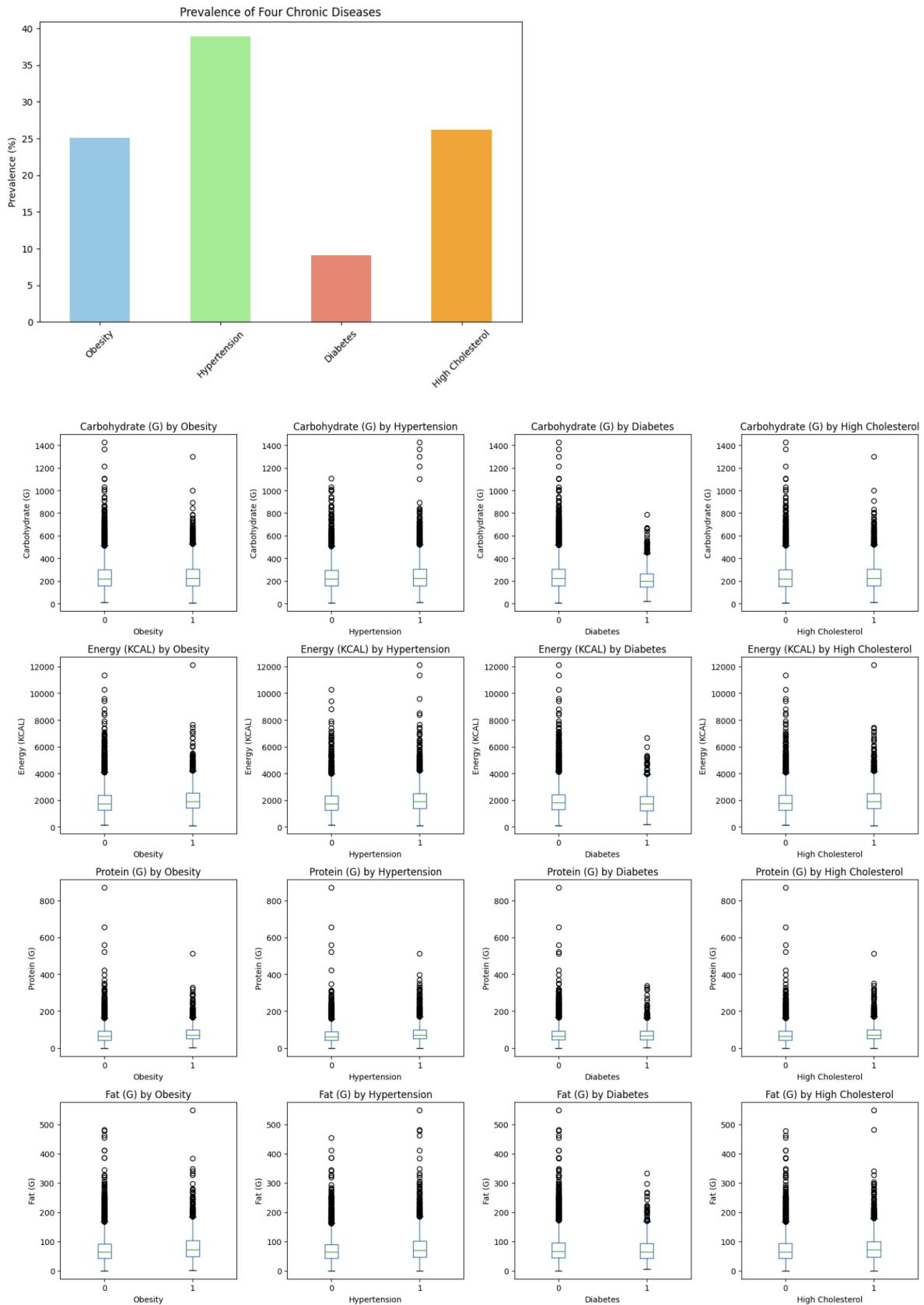
We currently use the Food Details endpoint of the FoodData Central API to extract nutritional components information for specific food items.

Dataset3 - NHANES

Data source:

<https://www.kaggle.com/datasets/cdc/national-health-and-nutrition-examination-survey/data>

The dataset I used was NHANES (National Health and Nutrition Examination Survey) taken in 2013-2014, which is a study designed to assess the health and nutritional status of adults and children in the United States. The survey combines interviews and physical examinations with a focus on a variety of health and nutrition measurements.



Dataset4 - LLM Fine-tuning Dataset

Dataset sample

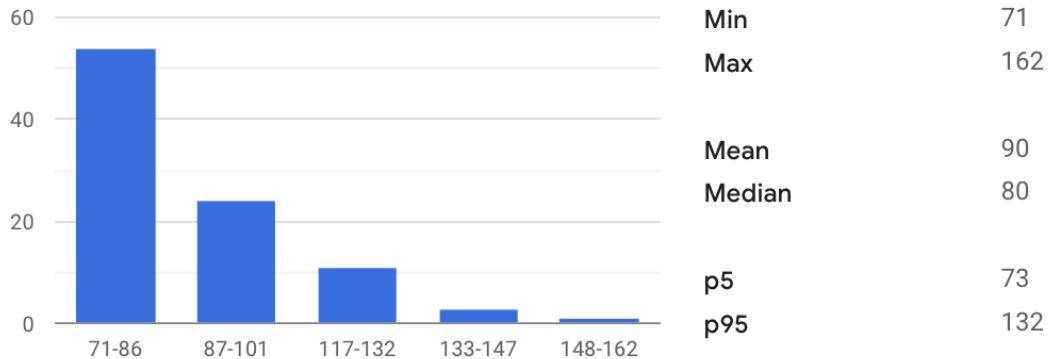
Displaying the first 3 rows of the tuning dataset (samples may be truncated for preview)

Message ?

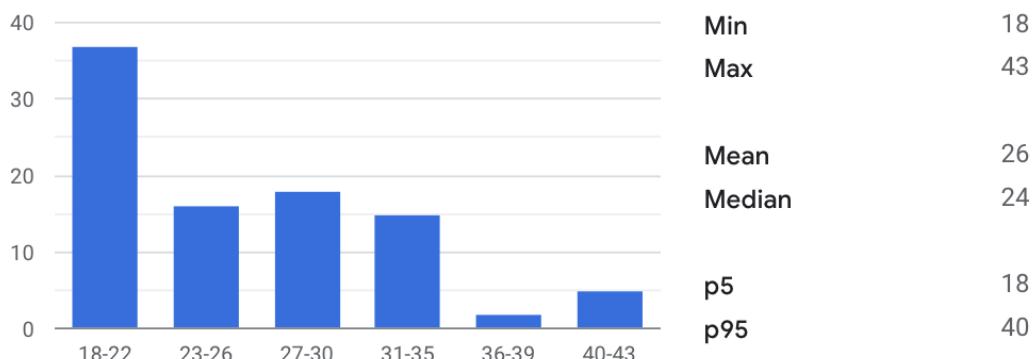
- Recommend if the product is suitable for consumption based on user allergies, preferences, and exceptions. ("user_allergies": ["peanuts", "dairy"], "user_preferences": ["vegan"], "user_exceptions": [], "product_ingredients": ["soybeans", "rice", "coconut milk", "cocoa powder", "dates"], "ingredient_details": {})
💡 This product is recommended for you. It does not contain any peanuts or dairy ingredients, aligning with your vegan preference.
- Recommend if the product is suitable for consumption based on user allergies, preferences, and exceptions. ("user_allergies": [], "user_preferences": ["vegetarian"], "user_exceptions": ["fish occasionally"], "product_ingredients": ["chicken breast", "carrots", "celery", "onions", "chicken broth"], "ingredient_details": {})
💡 This product is recommended for you based on your vegetarian preference with occasional fish exceptions. The product contains chicken, but does not contain any fish ingredients.
- Recommend if the product is suitable for consumption based on user allergies, preferences, and exceptions. ("user_allergies": ["peanuts", "soy"], "user_preferences": ["organic", "gluten-free"], "user_exceptions": [], "product_ingredients": ["quinoa", "spinach", "sweet potatoes", "garlic", "olive oil", "sodium benzoate (E211)", "sulfur dioxide (E220)", "ascorbic acid (E300)", "xanthan gum (E415)...
💡 This product is recommended for you. It is organic and gluten-free, and it does not contain peanuts or soy, aligning with your preferences and allergies.

Data distribution

Number of input token per example



Number of output token per example



Dataset5 - LLM RAG Dataset:

We compiled an expansive dataset from two authoritative sources in the field of nutrition: the "Dietary Guidelines for Americans, 2020-2025," encompassing 160 documentations, and "Wikibooks: Fundamentals of Human Nutrition," from 80 detailed URLs.