Data Analysis Project
P3: Data Analysis and Visualization
Correlation Between Dietary Habits and Nutritional Intake
(Group 1)

Yaowen Rui and Patrick Weber

University of Basel
Databases (CS244) course
Autumn Semester 2022

1 Introduction

In an era of rapid globalization, people have access to various food options from the whole world, ranging from organic and pollution-free choices to highly processed fast food and convenience meals. With this considerable number of options, questions arise about how factors such as education level, family influence, and personal preference affect individuals' dietary choices and whether these choices are truly nutritious.

This project aims to explore the correlation between individual dietary habits, food consumption patterns, and actual nutritional intake by combining two comprehensive data sources: Open Food Facts and the Food Consumption and Preferences Dataset. The main concern is to assess whether the foods people consider as healthy align with their actual nutritional value, and to identify whether so-called convenience foods are as unhealthy as commonly believed.

2 Analysis Goals

Analysis Goals:

1. Compare individuals' self-perception of health and their reported "healthy meals" with their actual nutritional intake.

   (a) Expectation:
   Identify whether people's perception of their diet aligns with their real nutritional intake.
   (b) Potential needed Data:
   $Diet\_current$, $Healthy\_feeling$, $Ideal\_diet$, $Healthy\_meal$ and nutritional information for each specific reported food.

2. Compare the nutritional quality of food consumed by individuals who frequently cook by themselves versus those who eat out often.

   (a) Expectation:
   Identify whether home-cooked meals lead to better nutrient intake (e.g., lower sodium, higher vitamins) compared to restaurant or fast food meals.
   (b) Potential needed Data:
   $Eating\_out, Comfort\_Food, Preferred\_drinks\_cuisines$ and detailed nutritional ingredients.

3. Determine how individual food choices and habits (e.g., breakfast consumption, eating out frequency, comfort food preferences) correlate with nutritional intake (e.g., including calories, protein, fats, vitamins, minerals).

   (a) Expectation:
   Identify whether specific dietary patterns lead to healthier or less healthy nutritional outcomes.
   (b) Potential needed Data:
   $Favorite\_Cuisine, Eating\_Changes, Favorite\_Food, Ideal\_Diet$ and detailed nutritional information for each food item.

3 Datasets

Used datasets

- Open Food Facts
- 9.6 GB
- CSV

- Food Consumption and Preferences Dataset
- 7.3 KB
- CSV

3.1 Open Food Facts

Open Food Facts dataset is a free, open, collaborative database of food products from around the world, with ingredients, allergens, nutrition facts and all the information we can find on product labels.

3.2 Food Consumption and Preferences Dataset

The Food Consumption and Preferences Dataset offers a comprehensive overview of individual food habits, preferences, and lifestyle choices. It includes detailed data such as daily calorie intake, specific food item consumption (e.g., coffee, fruits), eating habits (e.g., breakfast, comfort food), and beverage preferences. Additionally, this dataset captures lifestyle information like exercise frequency, employment status, education levels, and cooking habits.

4 Entity Relation Diagram per Data Source

Dataset 1 - Open Food Facts ER Diagram:
The attributes of all entities are not displayed in the ER Diagram because of limited space. The attributes related to the entities are listed below:

Products:
ProductID, code, url, product_name, abbreviated_product_name, generic_name, brands, brands_tags, brand_owner

ContributionHistory:
ContributionHistoryID, creator, created_t, created_datetime, last_modified_t, last_modified_datetime, last_modified_by, last_updated_t, last_updated_datetime

PackagingInfo:
PackagingInfoID, quantity, product_quantity, packaging, packaging_tags, packaging_en, packaging_text, states, states_tags, states_en

Labeling:
LabelingID, labels, labels_tags, labels_en

Categories:
CategoriesID, categories, categories_tags, categories_en, main_category, main_category_en

Origins:
OriginsID, origins, origins_tags, origins_en

ManufacturingDetails:
ManufacturingDetailID, manufacturing_places, manufacturing_places_tags, emb_codes, emb_codes_tags, first_packaging_code_geo

PurchasesInfo:
PurchasesInfoID, cities, cities_tags, purchase_places, stores, countries, countries_tags, countries_en

Ingredients:
IngredientsID, ingredients_text, ingredients_tags, ingredients_analysis_tags

AllergensAndTraces:
AllergensAndTracesID, allergens, allergens_en, traces, traces_tags, traces_en

NutritionMeta:
NutritionMetaID, serving_size, serving_quantity, no_nutrition_data, nutriscore_score, nutriscore_grade, nova_group, pnns_groups_1, pnns_groups_2, food_groups, food_groups_tags, food_groups_en

Additives:
AdditivesID, additives_n, additives, additives_tags, additives_en

EnviromentalInfo:
EnviromentalInfoID, ecoscore_score, ecoscore_grade, nutrient_levels_tags, carbon-footprint_100g, carbon-footprint-from-meat-or-fish_100g

ImageInfo:
ImageInfoID, image_url, image_small_url, image_ingredients_url, image_ingredients_small_url, image_nutrition_url image_nutrition_small_url, last_image_t, last_image_datetime

MiscellaneousInfo:
MiscellaneousInfoID, owner, data_quality_errors_tags, unique_scans_n, popularity_tags, completeness

EnergyAndFat:
EnergyAndFatID, energy-kj_100g, energy-kcal_100g, energy_100g, energy-from-fat_100g

FattyInfo:
FattyInfoID, fat_100g, saturated-fat_100g,butyric-acid_100g, caproic-acid_100g, caprylic-acid_100g, capric-acid_100g, lauric-acid_100g, myristic-acid_100g, palmitic-acid_100g, stearic-acid_100g, arachidic-acid_100g, behenic-acid_ 100g, lignoceric-acid_100g, cerotic-acid_100g, montanic-acid_100g, melissic-acid_100g, unsaturated-fat_100g, monounsaturated-fat_100g, omega-9-fat_100g, polyunsaturated-fat_100g, omega-3-fat_100g, omega-6-fat_100g, alpha-linolenic-acid_100g, eicosapentaenoic-acid_100g, docosahexaenoic-acid_100g, linoleic-acid_100g, arachidonic-acid_100g, gamma-linolenic-acid_100g, dihomo-gamma-linolenic-acid_100g, oleic-acid_100g, elaidic-acid_100g, gondoic-acid_100g, mead-acid_100g, erucic-acid_100g, nervonic-acid_100g, trans-fat_100g, cholesterol_100g

Carbohydrates:
CarbohydratesID, carbohydrates_100g, sugars_100g, added-sugars_100g, sucrose_100g, glucose_100g, fructose_100g, lactose_100g, maltose_100g, maltodextrins_100g, starch_100g, polyols_100g, erythritol_100g

Proteins:
ProteinsID, proteins_100g, casein_100g, serum-proteins_100g, nucleotides_100g

Fiber:
FiberID, fiber_100g, soluble-fiber_100g, insoluble-fiber_100g, beta-glucan_100g

Minerals:
MineralsID, salt_100g, added-salt_100g, sodium_100g, iron_100g, magnesium_100g,zinc_100g, copper_100g, calcium_100g, iodine_100g, silica_100g, phosphorus_100g, man-

ganese_100g, selenium_100g, potassium_100g, chloride_100g, fluoride_100g, chromium_100g, molybdenum_100g, bicarbonate_100g

Vitamins:
VitaminsID, vitamin-a_100g, beta-carotene_100g, vitamin-d_100g, vitamin-e_100g, vitamin-k_100g, vitamin-c_100g, vitamin-b1_100g, vitamin-b2_100g, vitamin-pp_100g, vitamin-b6_100g, vitamin-b9_100g, vitamin-b12_100g

VitaminDerivatives:
VitaminDerivativesID, biotin_100g, pantothenic-acid_100g, folates_100g, phylloquinone_100g, choline_100g

NutritionalCompositionAndScores:
NutritionalCompositionAndScoresID, fruits-vegetables-nuts_100g, fruits-vegetables-nuts-dried_100g, fruits-vegetables-nuts-estimate_100g, fruits-vegetables-nuts-estimate-from-ingredients_100g, collagen-meat-protein-ratio_100g, cocoa_100g, chlorophyl_100g, nutrition-score-fr_100g, nutrition-score-uk_100g

BioactiveCompounds:
BioactiveCompoundsID, caffeine_100g, taurine_100g, nitrate_100g, sulphate_100g, carnitine_100g, inositol_100g,alcohol_100g

PhysicalAndChemicalProperties:
PhysicalAndChemicalPropertiesID, ph_100g, glycemic-index_100g , water-hardness_100g, acidity_100g

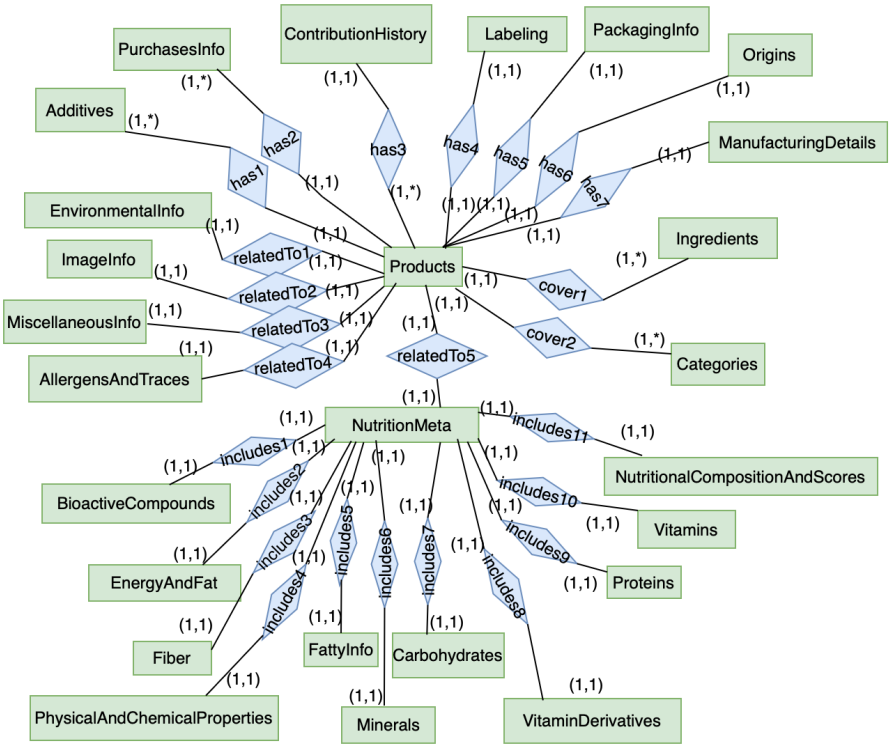

Fig. 1: Open Food Facts ER Diagram

Dataset 2 - Food Consumption And Preferences ER Diagram: The attributes of all entities are not displayed in ER Diagram because of limited space. The attributes related to the entities are listed below:

IndividualProfile:
IndividualProfileID, GPA, Gender, income, grade_level, weight, employment, on_off_campus, marital_status

FamilyBackground:
FamilyBackgroundID, mother_profession, mother_education, father_education, father_profession, parents_cook

FoodConsumption:
FoodConsumptionID, breakfast, coffee,veggies_day, soup, fruit_day, fries

CalorieIntake:
CalorieIntakeID, calories_chicken, calories_day, calories_scone, waffle_calories, tortilla_calories, turkey_calories

FoodPreferences:
FoodPreferencesID, comfort_food,comfort_food_reasons, comfort_food_reasons_coded, comfort_food_reasons_coded.1, cuisine, drink, indian_food, italian_food, ethnic_food, fav_food, fav_cuisine,fav_cuisine_coded, greek_food, persian_food, thai_food, food_childhood

Behaviour:
BehaviourID, cook, nutritional_check, vitamins, healthy_meal, eating_changes, eating_changes_coded, eating_changes_coded1, eating_out, exercise, pay_meal_out, sports, type_sports, meals_dinner_friend

SelfPerception:
SelfPerceptionID, healthy_feeling, life_rewarding, self_perception_weight

Diet:
DietID, ideal_diet, ideal_diet_coded, diet_current, diet_current_coded

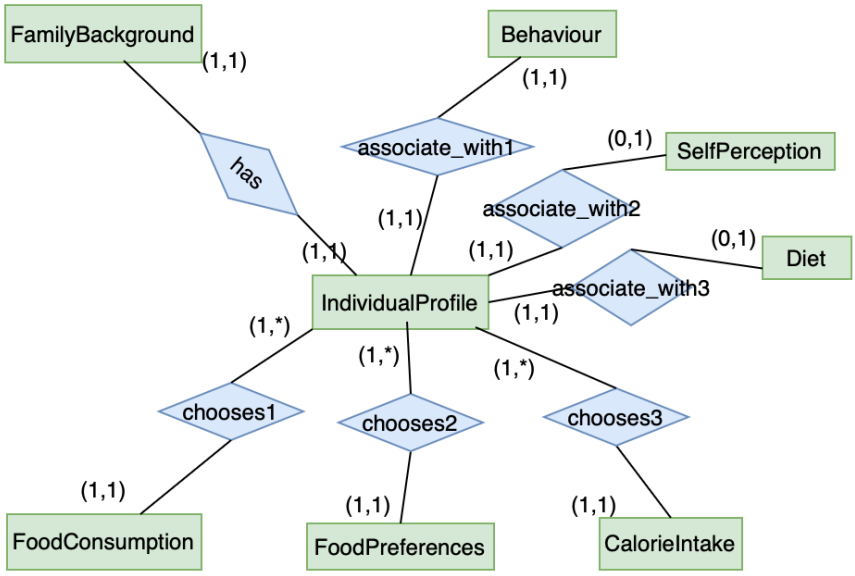

Fig. 2: Food Consumption And Preferences ER Diagram

5 Relational Schema

Logical, relational schema of the integrated schema:

5.1 Logical relational schema

Dataset 1 - Open Food Facts Relational Schema

Products(ProductID, code, url, product_name, abbreviated_product_name, generic_name, brands, brands_tags, brand_owner, LabelingID, PackagingInfoID, OriginsID, ManufacturingDetailsID, ContributionHistoryID, PurchasesInfoID, AdditivesID, EnvironmentalInfoID, ImageInfoID, MiscellaneousInfoID, AllergensAndTracesID, NutritionMetaID, CategoriesID, IngredientsID)

ContributionHistory(ContributionHistoryID, creator, created_t, created_datetime, last_modified_t, last_modified_datetime, last_modified_by, last_updated_t, last_updated_datetime)

PackagingInfo(PackagingInfoID, quantity, product_quantity, packaging, packaging_tags, packaging_en, packaging_text, states, states_tags, states_en)

Labeling(LabelingID, labels, labels_tags, labels_en)

Categories(CategoriesID, categories, categories_tags, categories_en, main_category, main_category_en)

Origins(OriginsID, origins, origins_tags, origins_en)

ManufacturingDetail(ManufacturingDetailID, manufacturing_places, manufacturing_places_tags, emb_codes, emb_codes_tags, first_packaging_code_geo)

PurchasesInfo(PurchasesInfoID, cities, cities_tags, purchase_places, stores, countries, countries_tags, countries_en)

Ingredients(IngredientsID, ingredients_text, ingredients_tags, ingredients_analysis_tags)

AllergensAndTraces(AllergensAndTracesID, allergens, allergens_en, traces, traces_tags, traces_en)

NutritionMeta(NutritionMetaID, serving_size, serving_quantity, no_nutrition_data, nutriscore_score, nutriscore_grade, nova_group, pnns_groups_1, pnns_groups_2, food_groups, food_groups_tags, food_groups_en, BioactiveCompoundsID, EnergyAndFatID, FiberID, PhysicalAndChemicalPropertiesID, FattyInfoID, MineralsID, CarbohydratesID, VitaminDerivativesID, ProteinsID, VitaminsID, NutritionalCompositionAndScoresID)

Additives(AdditivesID, additives_n, additives, additives_tags, additives_en)

EnviromentalInfo(EnviromentalInfoID, ecoscore_score, ecoscore_grade, nutrient_levels_tags, carbon-footprint_100g, carbon-footprint-from-meat-or-fish_100g)

ImageInfo(ImageInfoID, image_url, image_small_url, image_ingredients_url, image_ingredients_small_url, image_nutrition_url image_nutrition_small_url, last_image_t, last_image_datetime)

MiscellaneousInfo(MiscellaneousInfoID, owner, data_quality_errors_tags, unique_scans_n, popularity_tags, completeness)

EnergyAndFat(EnergyAndFatID, energy-kj_100g, energy-kcal_100g, energy_100g, energy-from-fat_100g)

FattyInfo(FattyInfoID, fat_100g, saturated-fat_100g,butyric-acid_100g, caproic-acid_100g, caprylic-acid_100g, capric-acid_100g, lauric-acid_100g, myristic-acid_100g, palmitic-acid_100g, stearic-acid_100g, arachidic-acid_100g, behenic-acid_ 100g, lignoceric-acid_100g, cerotic-acid_100g, montanic-acid_100g, melissic-acid_100g, unsaturated-fat_100g, monounsaturated-fat_100g, omega-9-fat_100g, polyunsaturated-fat_100g, omega-3-fat_100g, omega-6-fat_100g, alpha-linolenic-acid_100g, eicosapentaenoic-acid_100g, docosahexaenoic-acid_100g, linoleic-acid_100g, arachidonic-acid_100g, gamma-linolenic-acid_100g, dihomo-gamma-linolenic-acid_100g, oleic-acid_100g, elaidic-acid_100g, gondoic-acid_100g, mead-acid_100g, erucic-acid_100g, nervonic-acid_100g, trans-fat_100g, cholesterol_100g)

Carbohydrates(CarbohydratesID, carbohydrates_100g, sugars_100g, added-sugars_100g, sucrose_100g, glucose_100g, fructose_100g, lactose_100g, maltose_100g, maltodextrins_100g, starch_100g, polyols_100g, erythritol_100g)

Proteins(ProteinsID, proteins_100g, casein_100g, serum-proteins_100g, nucleotides_100g)

Fiber(FiberID, fiber_100g, soluble-fiber_100g, insoluble-fiber_100g, beta-glucan_100g)

Minerals(MineralsID, salt_100g, added-salt_100g, sodium_100g, iron_100g, magnesium_100g,zinc_100g, copper_100g, calcium_100g, iodine_100g, silica_100g, phosphorus_100g, manganese_100g, selenium_100g, potassium_100g, chloride_100g, fluoride_100g, chromium_100g, molybdenum_100g, bicarbonate_100g)

Vitamins(VitaminsID, vitamin-a_100g, beta-carotene_100g, vitamin-d_100g, vitamin-e_100g, vitamin-k_100g, vitamin-c_100g, vitamin-b1_100g, vitamin-b2_100g, vitamin-pp_100g, vitamin-b6_100g, vitamin-b9_100g, vitamin-b12_100g)

VitaminDerivatives(VitaminDerivativesID, biotin_100g, pantothenic-acid_100g, folates_100g, phylloquinone_100g, choline_100g)

NutritionalCompositionAndScores(NutritionalCompositionAndScoresID, fruits-vegetables-nuts_100g, fruits-vegetables-nuts-dried_100g, fruits-vegetables-nuts-estimate_100g, fruits-vegetables-nuts-estimate-from-ingredients_100g, collagen-meat-protein-ratio_100g, cocoa_100g, chlorophyl_100g, nutrition-score-fr_100g, nutrition-score-uk_100g)

BioactiveCompounds(BioactiveCompoundsID, caffeine_100g, taurine_100g, nitrate_100g, sulphate_100g, carnitine_100g, inositol_100g,alcohol_100g)

PhysicalAndChemicalProperties(PhysicalAndChemicalPropertiesID, ph_100g, glycemic-index_100g , water-hardness_100g, acidity_100g)

Dataset 2 - Food Consumption And Preferences Relational Schema

IndividualProfile(IndividualProfileID, GPA, Gender, income, grade_level, weight, employment, on_off_campus, marital_status, FamilyBackgroundID, FoodConsumptionID, FoodPreferencesID, CalorieIntakeID, SelfPerceptionID, DietID, BehaviourID)

FamilyBackground(FamilyBackgroundID, mother_profession, mother_education, father_education, father_profession, parents_cook)

FoodConsumption(FoodConsumptionID, breakfast, coffee,veggies_day, soup, fruit_day, fries)

CalorieIntake(CalorieIntakeID, calories_chicken, calories_day, calories_scone, waffle_calories, tortilla_calories, turkey_calories)

FoodPreferences(FoodPreferencesID, comfort_food,comfort_food_reasons, comfort_food_reasons_coded, comfort_food_reasons_coded.1, cuisine, drink, indian_food, italian_food, ethnic_food, fav_food, fav_cuisine,fav_cuisine_coded, greek_food, persian_food, thai_food, food_childhood)

Behaviour(BehaviourID, cook, nutritional_check, vitamins, healthy_meal, eating_changes, eating_changes_coded, eating_changes_coded1, eating_out, exercise, pay_meal_out, sports, type_sports, meals_dinner_friend)

SelfPerception(SelfPerceptionID, healthy_feeling, life_rewarding, self_perception_weight)

Diet(DietID, ideal_diet, ideal_diet_coded, diet_current, diet_current_coded)

6 Database Setup

6.1 Raw dataset1: Open Food Facts

- Utilize Pandas to preprocess the large raw CSV file. Treat the first row as the header and ensure all subsequent rows align with it. Save the cleaned output to a new CSV file for further processing.
  See *preprocess-dataset1.py*

- Classify each column in the dataset to its corresponding entity using a predefined entity-to-attributes mapping. Generate a separate CSV file for each entity, splitting the large dataset into manageable CSVs.
  See *dataset1-to-csv.py*

- Create database tables for each entity and load data from the CSV files into the respective tables.
  See *dataset1-CSVs-toDB.py*

- Once all tables are populated, define and add foreign key constraints.
  See *add-foreign-keys-inDB.py*

6.2 Raw dataset2: Food Consumption And Preferences Dataset

The processing steps for this dataset are similar to Dataset 1, with one key difference: Foreign key constraints are added during table creation and data insertion.
See *preprocess-dataset2-toDB.py*

6.3 Integration

We need the CategoriesID that is linked to the corresponding products. Extract the usefulCategories, as not every product has associated categories; we only require products that have categories.
The Categories table (3,500,000 rows) from Dataset 1 has a relationship with, for example, the FoodPreference table from Dataset 2. However, there is a significant size difference between these two tables. Therefore, we need to refine the Categories table to ensure relevance and efficiency.
The FoodConsumption, CalorieIntake, and Diet tables are not relevant, as the required categories are already specified in the header. Retrieving product information is more straightforward using SQL queries.

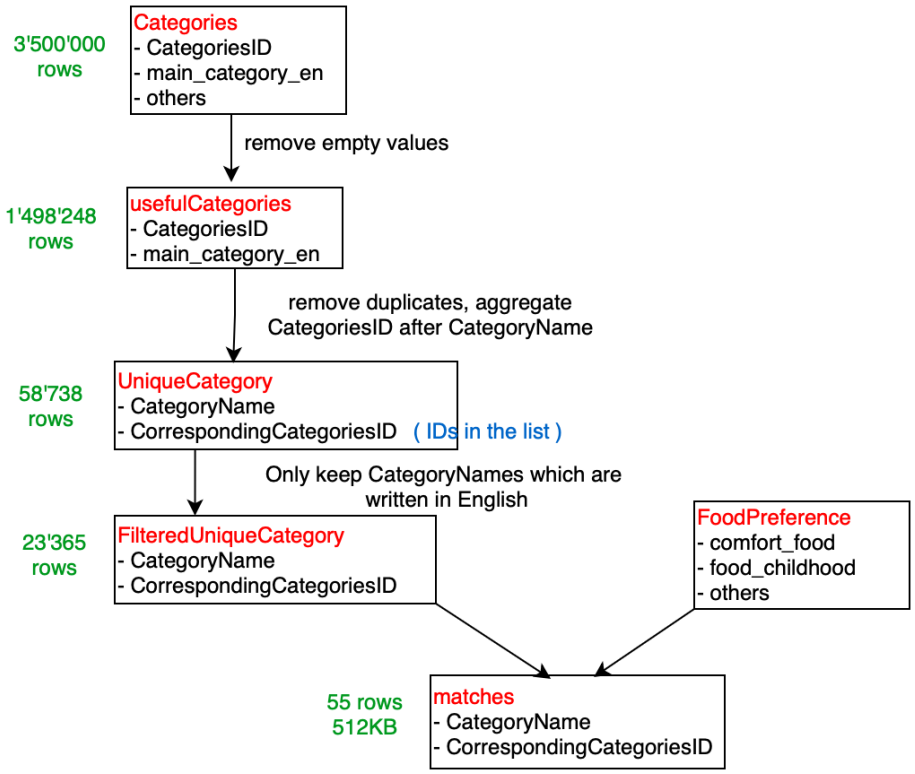

Fig. 3: refine the table Categories

To find the matches between the FilteredUniqueCategory and FoodPreference tables:

- Extract category-related values from the FoodPreference.
- Compare these values against CategoryName in the FilteredUniqueCategory
- Identify matching records and their corresponding IDs from FilteredUniqueCategory.
- Only matches table is needed to be inserted into database.

See *extract-categories.py* and *matches-toDB.py*

6.4 Integrated schema

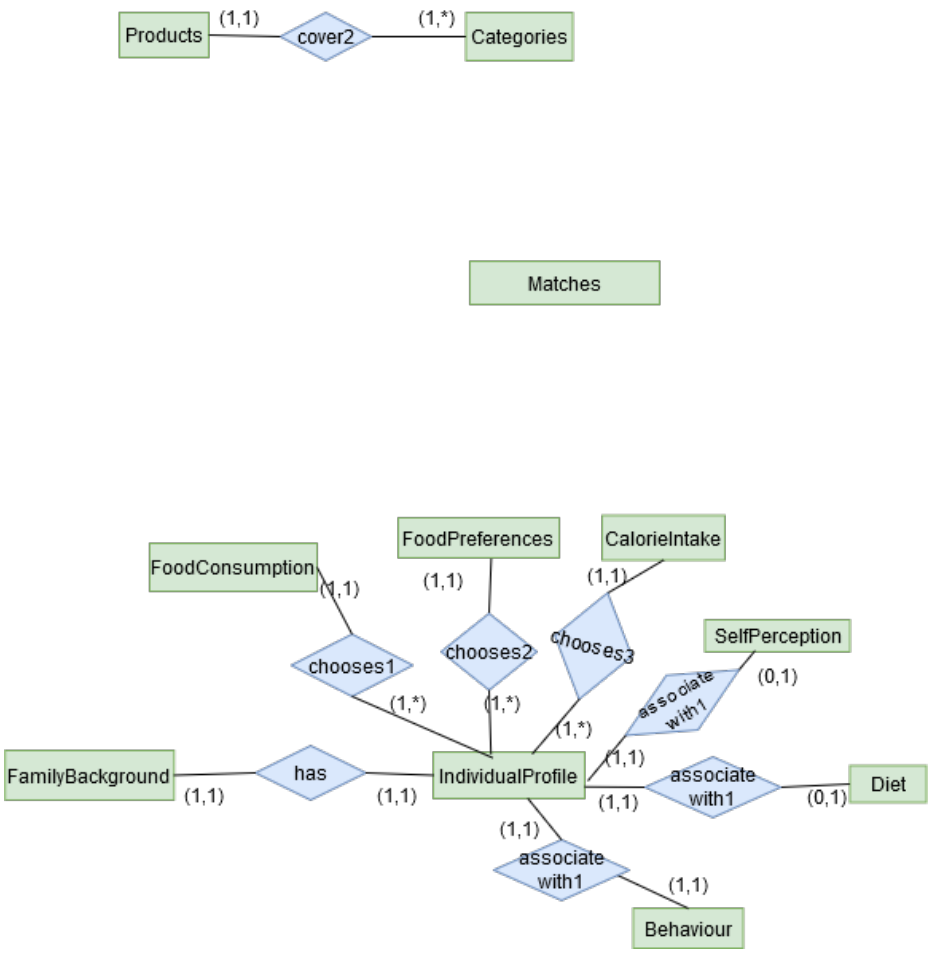

Fig. 4: Integrated ER Diagram

The integrated schema contains an additional entity:
Matches(CategoryName, CorrespondingCategoriesID)
This entity is not connected to any other entity, but is rather used to make queries based on the Food and Consumption Preferences Dataset.

6.5 Set up DB

We use a Docker container to run the PostgreSQL database. command:

```
docker run --name database_project
-e POSTGRES_USER=user -e POSTGRES_PASSWORD=password
-e POSTGRES_DB=database_project -p 127.0.0.1:54321:5432
-d postgres
```

6.6 Data Access

The SQL-dump can be downloaded here: *SQL-dump*
The SQL-dump can be imported to the previously created docker container with the following command:

```
docker exec -i database_project pg_restore
-U user -d database_project < [SQL-dump location]
```

7 REST API for table energyAndFat

A new Docker container is used to package the REST API and database together. The frontend-only website(website.html) provides a simple dashboard to handle GET, POST, PUT and DELETE operations via the REST API. The API (Flask App in file add-REST-API.py)runs locally and handles the database operations and sends responses.

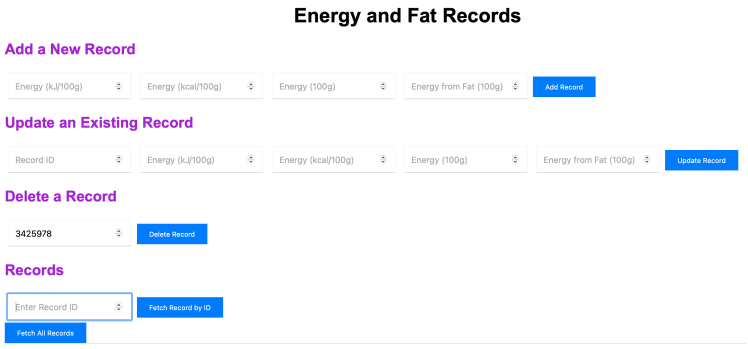

Fig. 5: website

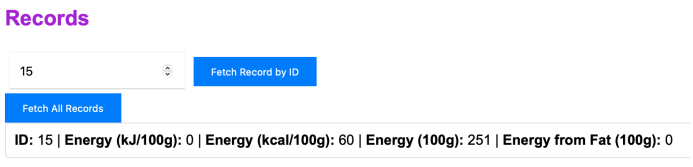

Fig. 6: fetch a record by ID

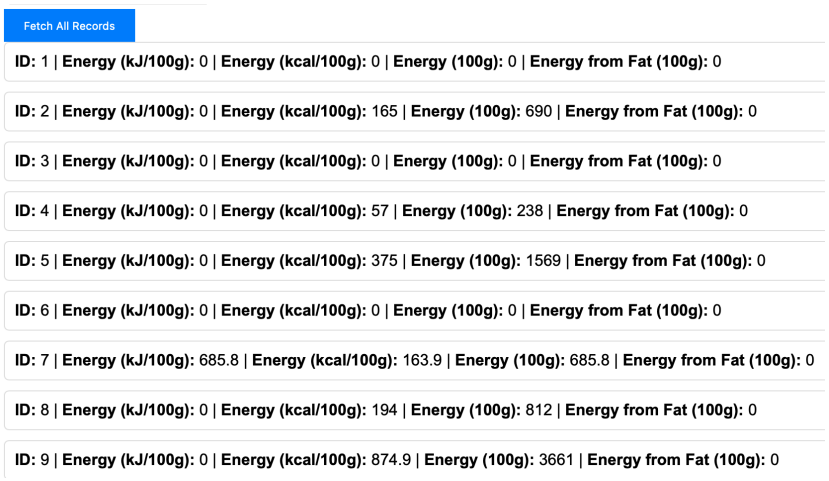

Fig. 7: fetch all records

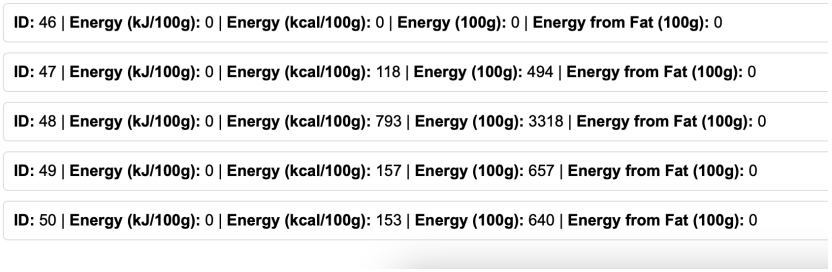

Fig. 8: fetch all records with max. 50 records per page

Fig. 9

Fig. 10

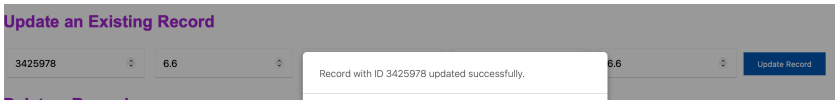

Fig. 11

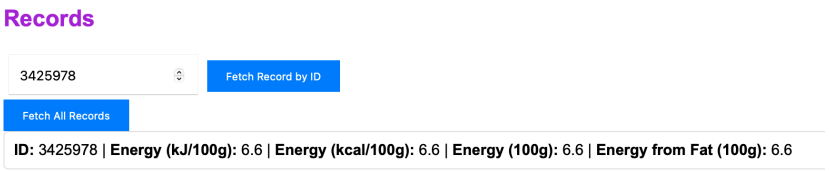

Fig. 12: fetch newly updated record

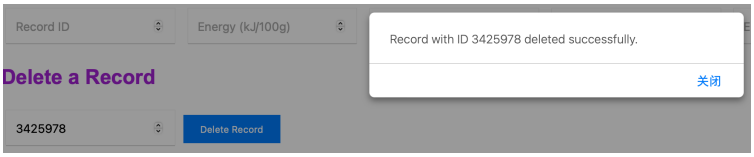

Fig. 13: delete

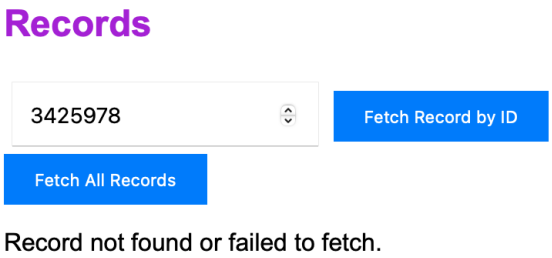

Fig. 14: fetch a deleted record

8 Analysis Goal 1

Goal 1: Compare individuals' self-perception of health and their reported "healthy meals" with their actual nutritional intake.

The **ideal_diet** from the **diet** table records the healthy foods people consider ideal, which include high-protein and lean protein sources, fruits, fresh vegetables, organic food, chicken, fish, nuts, eggs, green beans, red and white meats, seafood, while avoiding processed products, excessive sugar, high-carb items and snacks, and reducing carbohydrate intake. **Almonds** and **quinoa**, selected as two of representatives, will be discussed below.
This query extracts the nutritional data for representative products that belong to the specified categories (Almonds, Beans, etc.).
The visualized results of that query can be seen in Fig. 15-20.

```
SELECT
        p.product_name,
        p.categoriesid,
        eaf.energy_100g,
        fi.fiber_100g,
        fi.soluble_fiber_100g,
        fi.insoluble_fiber_100g,
        fai.fat_100g,
        fai.saturated_fat_100g,
        fai.omega_3_fat_100g,
        fai.trans_fat_100g,
```

```
12          fai.cholesterol_100g,
13          mi.salt_100g,
14          mi.sodium_100g,
15          mi.iron_100g,
16          mi.calcium_100g,
17          mi.magnesium_100g,
18          mi.zinc_100g,
19          mi.selenium_100g,
20          ca.carbohydrates_100g,
21          ca.sugars_100g,
22          ca.fructose_100g,
23          ca.lactose_100g,
24          ca.maltose_100g,
25          pr.proteins_100g,
26          pr.casein_100g,
27          vi.vitamin_a_100g,
28          vi.vitamin_c_100g,
29          vi.vitamin_d_100g,
30          vi.vitamin_e_100g,
31          vi.vitamin_b1_100g,
32          vi.vitamin_b6_100g,
33          vi.vitamin_b9_100g
34  FROM
35      products p
36  JOIN
37      nutritionmeta nm ON p.nutritionmetaid = nm.nutritionmetaid
38  LEFT JOIN
39      energyandfat eaf ON nm.energyandfatid = eaf.energyandfatid
40  LEFT JOIN
41      fiber fi ON nm.fiberid = fi.fiberid
42  LEFT JOIN
43      fattyinfo fai ON nm.fattyinfoid = fai.fattyinfoid
44  LEFT JOIN
45      minerals mi ON nm.mineralsid = mi.mineralsid
46  LEFT JOIN
47      carbohydrates ca ON nm.carbohydratesid = ca.carbohydratesid
48  LEFT JOIN
49      proteins pr ON nm.proteinsid = pr.proteinsid
50  LEFT JOIN
51      vitamins vi ON nm.vitaminsid = vi.vitaminsid
52  WHERE
53      p.categoriesid = %s;
```

Quinoa's Actual Nutritional Profile:
Quinoa dominates the carbohydrate spectrum: 63.9g/100g. It ranks high in energy(calories) content (1,703 kJ/100g) which aligns with its carbohydrate den-

sity. Quinoa contributes a moderate amount of sugar(3.7g/100g). Quinoa holds the third-highest position for iron content among the analyzed foods and it's iron content covers approximately 15% of the recommended daily intake per 100g, making it particularly beneficial for vegetarians or individuals with iron deficiency.
While quinoa provides essential nutrients like iron, its extremely high carbohydrate content may not align with the ideal diet's focus on reducing carbohydrate intake. Quinoa should be consumed in moderate portion size (around 90g) that provides a balanced nutrient profile without overwhelming daily carbohydrate limits.

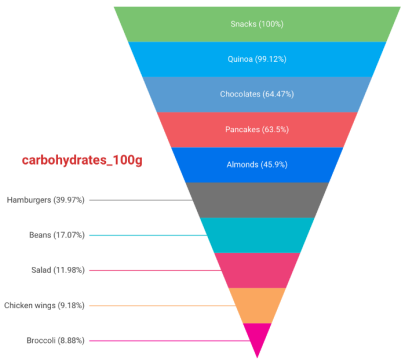

Fig. 15: carbohydrates

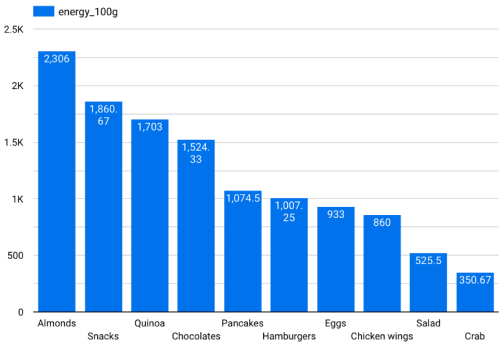

Fig. 16: energy

Almonds' Actual Nutritional Profile:
Almonds dominate in energy content, providing an impressive 2,306 kJ/100g, making them an energy-dense food suitable for high-energy needs. With a moderate carbohydrate content of 45.9%/100g and high fiber content (11.1g/100g), almonds support digestion. Almonds also enrich in protein which, along with fiber, increase the feeling of fullness. Although almonds are high in total fat, they stand out for their relatively low saturated fat content (3.01%/100g). This makes them a heart-healthy option, as saturated fat intake should not exceed 10% of daily caloric intake to minimize cardiovascular risks. Almonds provide one of the highest calcium contents among the analyzed foods, making them an excellent choice for bone health. They also rank highly in iron content, contributing to anemia prevention and improved oxygen transport.
While almonds provide significant health benefits, their high caloric density necessitates portion control to avoid excessive energy intake. With their nutrient-dense profile, almonds align well with the perception of a healthy food and are an ideal option for snacking or as part of a balanced diet.

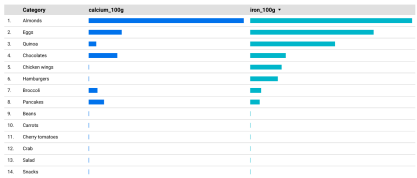

Fig. 17: calcium and iron

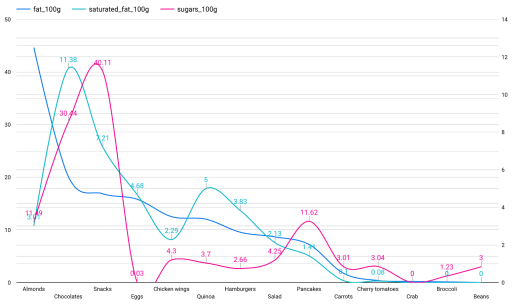

Fig. 18: fat, saturated-fat and sugar

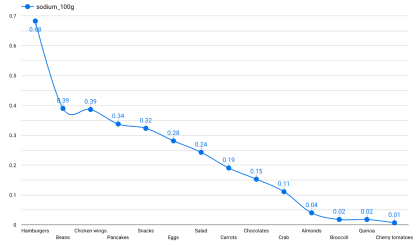

Fig. 19: sodium

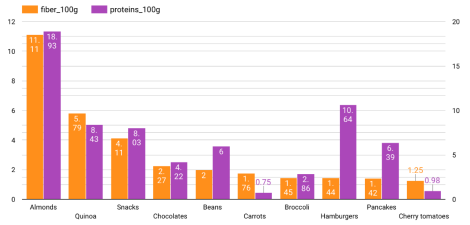

Fig. 20: fiber and protein

9 Analysis Goal 2

Goal 2: Compare the nutritional quality of food consumed by individuals who frequently cook by themselves versus those who eat out often.

For this analysis we have some interesting attributes, like **cook** and **eating_out** from the **behaviour** table. To be able to analyse these values, the **fav_cuisine** from the **foodpreferences** table can be compared with the weighted sum of the **cook** and **eating_out** attributes. For this the following query was used:

```
1 SELECT LOWER(TRIM(f.fav_cuisine)) AS normalized_cuisine,
      SUM(b.cook) AS weighted_cook, SUM(b.eating_out) AS
      weighted_out
2 FROM foodpreferences f
3 JOIN individualprofile i ON i.foodpreferencesid = f.
      foodpreferencesid
4 JOIN behaviour b ON i.behaviourid = b.behaviourid
5 GROUP BY normalized_cuisine
6 ORDER BY weighted_cook desc
```

The result was that italian food is the most popular for both so it would be useless to analyse that. The second most popular is for both Mexican, but if the results get merged by similarity (for example there are people who wrote "Mexican" and people who wrote "Mexican Food"), there is a clear difference which can be analysed. People who frequently cook at home prefer Chinese food, while people who frequently eat out prefer Mexican food. The following information was gathered with the following query:
The visualized results of that query can be seen in Fig. 21-26.

```
SELECT
        mat.categoryname,
    AVG(NULLIF(eaf.energy_100g, 0)) AS avg_energy_100g,
    AVG(NULLIF(fi.fiber_100g, 0)) AS avg_fiber_100g,
    AVG(NULLIF(fai.fat_100g, 0)) AS avg_fat_100g,
    AVG(NULLIF(fai.saturated_fat_100g, 0)) AS
        avg_saturated_fat_100g
    AVG(NULLIF(mi.salt_100g, 0)) AS avg_salt_100g,
    AVG(NULLIF(mi.sodium_100g, 0)) AS avg_sodium_100g,
    AVG(NULLIF(mi.iron_100g, 0)) AS avg_iron_100g,
    AVG(NULLIF(mi.calcium_100g, 0)) AS avg_calcium_100g,
    AVG(NULLIF(ca.carbohydrates_100g, 0)) AS
        avg_carbohydrates_100g,
    AVG(NULLIF(ca.sugars_100g, 0)) AS avg_sugars_100g,
    AVG(NULLIF(pr.proteins_100g, 0)) AS avg_proteins_100g
FROM
    products p
JOIN
    nutritionmeta nm ON p.nutritionmetaid = nm.
        nutritionmetaid
LEFT JOIN
    energyandfat eaf ON nm.energyandfatid = eaf.
        energyandfatid
LEFT JOIN
    fiber fi ON nm.fiberid = fi.fiberid
LEFT JOIN
    fattyinfo fai ON nm.fattyinfoid = fai.fattyinfoid
LEFT JOIN
    minerals mi ON nm.mineralsid = mi.mineralsid
LEFT JOIN
    carbohydrates ca ON nm.carbohydratesid = ca.
        carbohydratesid
LEFT JOIN
    proteins pr ON nm.proteinsid = pr.proteinsid
LEFT JOIN
    vitamins vi ON nm.vitaminsid = vi.vitaminsid
```

```
32 JOIN
33     matches mat ON p.categoriesid = ANY(SELECT unnest(mat
           .correspondingcategoriesid))
34 WHERE
35     mat.categoryname IN ('Chicken wings', 'Crab', '
           Dumplings', 'Rice', 'Beans', 'Burritos', 'Stuffed
           peppers', 'Tacos')
36 GROUP BY
37     mat.categoryname
38 ORDER BY
39     mat.categoryname;
```

Mexican Food nutritional profile:
Same as with the Chinese food, representative categories will be chosen to analyse the nutritional profile of Mexican food. The following categories will be taken a closer look at: Beans, Burritos, Stuffed Peppers and Tacos.
Mexican food dominates in almost all categories, but is this actually good thing? Having a higher energy density (1174kJ/100g) can easily lead to overeating if one is not careful. Having more fiber (4.7g/100g) is positive though, since that supports digestion and helps regulating blood sugar. It is very high in fat (13g/100g) and saturated fat (4.7g/100g) which once again increases energy density and the flavor of the food. Saturated fats can lead to cardiovascular risk when consumed in excess. The carbohydrate (32.1g/100g) and sugar (3.2g/100g) density is also high, which leads to energy boosts but also to high blood sugar levels. It also has more salt (1.6g/100g) and slightly more sodium (0.7g/100g), which enhances the taste of the food. Lastly, Mexican food has a lot more iron (0.11g/100g) and calcium (0.36g/100g), which is beneficial for oxygen transportation in the blood and for bone and teeth health.
Mexican food provides a lot of energy which is a good option for active lifestyles. However, since it also has more fat, sugar and sodium, this can lead to health issues if too much food is consumed. Mexican food contains the most important nutrients for the body, which means it is very reliable in keeping one's body healthy but eating too much can be unhealthy as well, so one has to be careful about the amounts they're consuming.

Chinese Food nutritional profile:
Since it is hard to take a look at Chinese food directly and the Open Food Facts dataset has categories from which a category mapping has already been created, representative categories from that were chosen. The following categories will be taken a closer look at: Chicken Wings, Crab, Dumplings and Rice.
Chinese food has less nutrients than Mexican food in any category apart from proteins, where it dominates with an average of 12.7g/100g. But this is not necessarily a bad thing, since proteins are helpful for keeping the immune system healthy and for some other nutrients it's actually better to have less as discussed in the Mexican food section.
Chinese food has higher amounts of protein, which are in general good for build-

ing and maintain muscles. However it has less fiber and has a higher reliance on sauces for flavor, since it doesn't have as much sodium or fats. Chinese food has almost no iron or calcium, which is another important nutrient for the body, which means one can't rely solely on this kind of food, although it is very good for keeping a healthy diet.

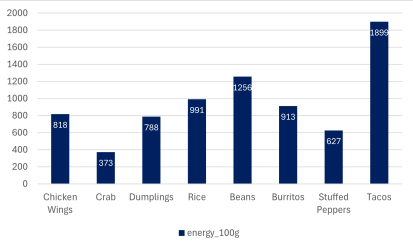

Fig. 21: Energy

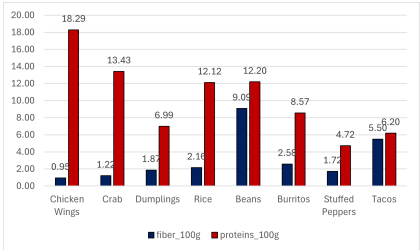

Fig. 22: Fiber and Proteins

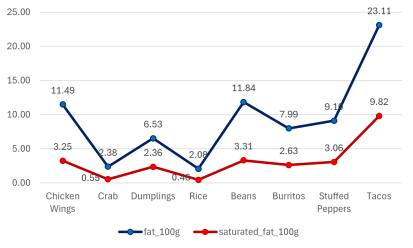

Fig. 23: Fat and Saturated Fat

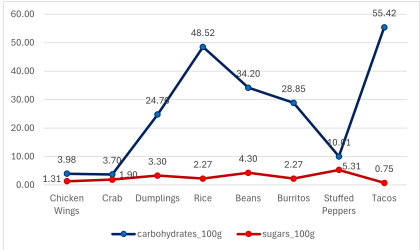

Fig. 24: Carbohydrates and Sugars

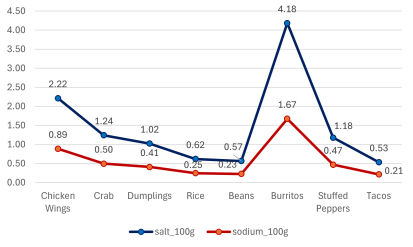

Fig. 25: Salt and Sodium

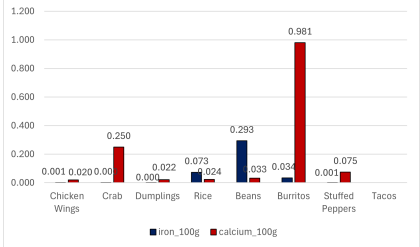

Fig. 26: Iron and Calcium

10 Analysis Goal 3

Goal 3: Determine how individual food choices and habits correlate with nutritional intake.

This analysis can be done by taking a look at the **comfort_food** attribute from the **foodpreferences** table. The data of the comfort_food usually consist of multiple categories split by a comma. This means that these values have to be split first and can then be added up to see, which categories appear the most. Afterwards, these categories can be taken a closer look at. The following query was used to extract the categories:

```
SELECT category_normalized, count(*)
FROM
(
        SELECT LOWER(TRIM(category)) AS
            category_normalized
        FROM foodpreferences f, unnest(string_to_array(f.
            comfort_food, ',')) AS category
)
GROUP BY category_normalized
ORDER BY count DESC
```

The result of the query shows, that the most popular comfort foods are Ice Cream, Pizza, Chocolate, Chips and Cookies. But since there are no category matches for Pizza and Cookies, only the other three categories will be analysed further. The analysis will compare the nutritional values of the chosen categories with various other categories.
A more general query than the one for analysis 2 was used, which won't be shown here because it is basically the same.
The visualized results of that query can be seen in Fig. 27-32.

Ice Cream nutritional profile:
Ice cream doesn't stand at the top of any nutritional value. It has a relatively low energy density (972 kJ/100g) and contains barely any salt (0.15g/100g), sodium 0.06g/100g) or iron (0.0004g/100g). The only thing where it stands out is it's sugar density with 21.41g/100g, which is one of the highest out of all categories, even though it's carbohydrate density isn't that high with 26.98g/100g.
From this it can be seen that ice cream doesn't really play a major role in contributing to any nutritional value. Well, it's mostly made out of water so that was to be expected. Nevertheless, in order to still have some kind of outstanding flavor it contains a lot of sugar. This is exactly what makes it so unhealthy for us, not contributing to any important nutritional values and only filling us up with sugar.

Chocolate nutritional profile:
Chocolate is very rich in energy (2208 kJ/100g) and also has a very high fat density (33.01g/100g) and especially saturated fat density (18.91g/100g). The most outstanding result is it's sugar density, with an amazing 44.24g/100g is it by far the highest out of all considered categories, although the carbohydrate density isn't as high as others with 50.46g/100g. This just means that almost all carbohydrates are actually sugar. It doesn't rank that high in any other nutrient category though. Especially for iron (0.004g/100g) and calcium (0.16g/100g) it doesn't fare that well.
This shows that chocolate is very good at providing energy and therefore an understandable popular comfort food, but it is not healthy at all. The sugar and saturated fat values are straight up worrying and it doesn't really provide any healthy and necessary nutrients. But these values also explain the strong flavor that chocolate has and why everyone loves it.

Chips nutritional profile:
Chips is rich in energy (1774 kJ/100g) and also stands in the top ranks for its fat density (21.27g/100g) but not as high as chocolate. What makes chips unique is their amount of salt (11.49g/100g) and sodium (4.6g/100g). Other than that, chips don't have that high nutritional values, also barely contributing to the iron (0.002g/100g) and calcium (0.09g/100g) intake. Compared to the previous two categories, chips rank high in carbohydrates with 49.44g/100g but they don't consist nearly as much sugar with only 4.65g/100g.
Just like the previous two categories, chips aren't good at contributing to any important nutrients. But unlike them, chips aren't unhealthy because they have a lot of sugar but rather because of their insane amount of salt and sodium, which is clearly way above what the daily intake should be. Eating chips is therefore unhealthy and can lead to an excessive amount of salt and sodium intake.

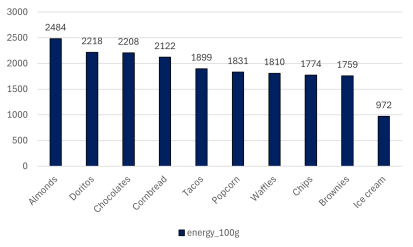

Fig. 27: Energy

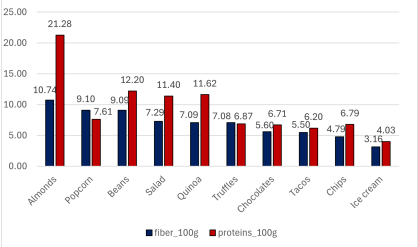

Fig. 28: Fiber and Proteins

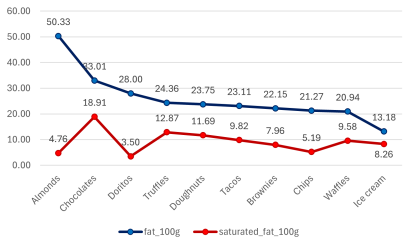

Fig. 29: Fat and Saturated Fat

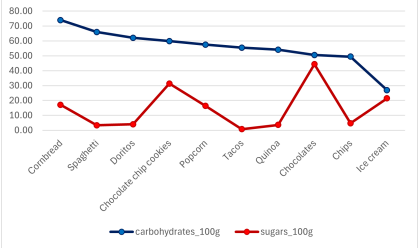

Fig. 30: Carbohydrates and Sugars

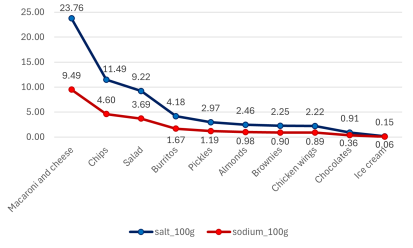

Fig. 31: Salt and Sodium

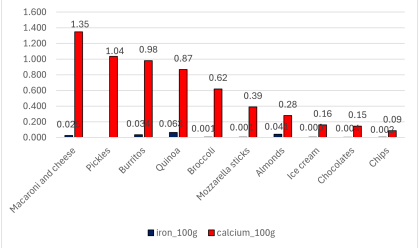

Fig. 32: Iron and Calcium