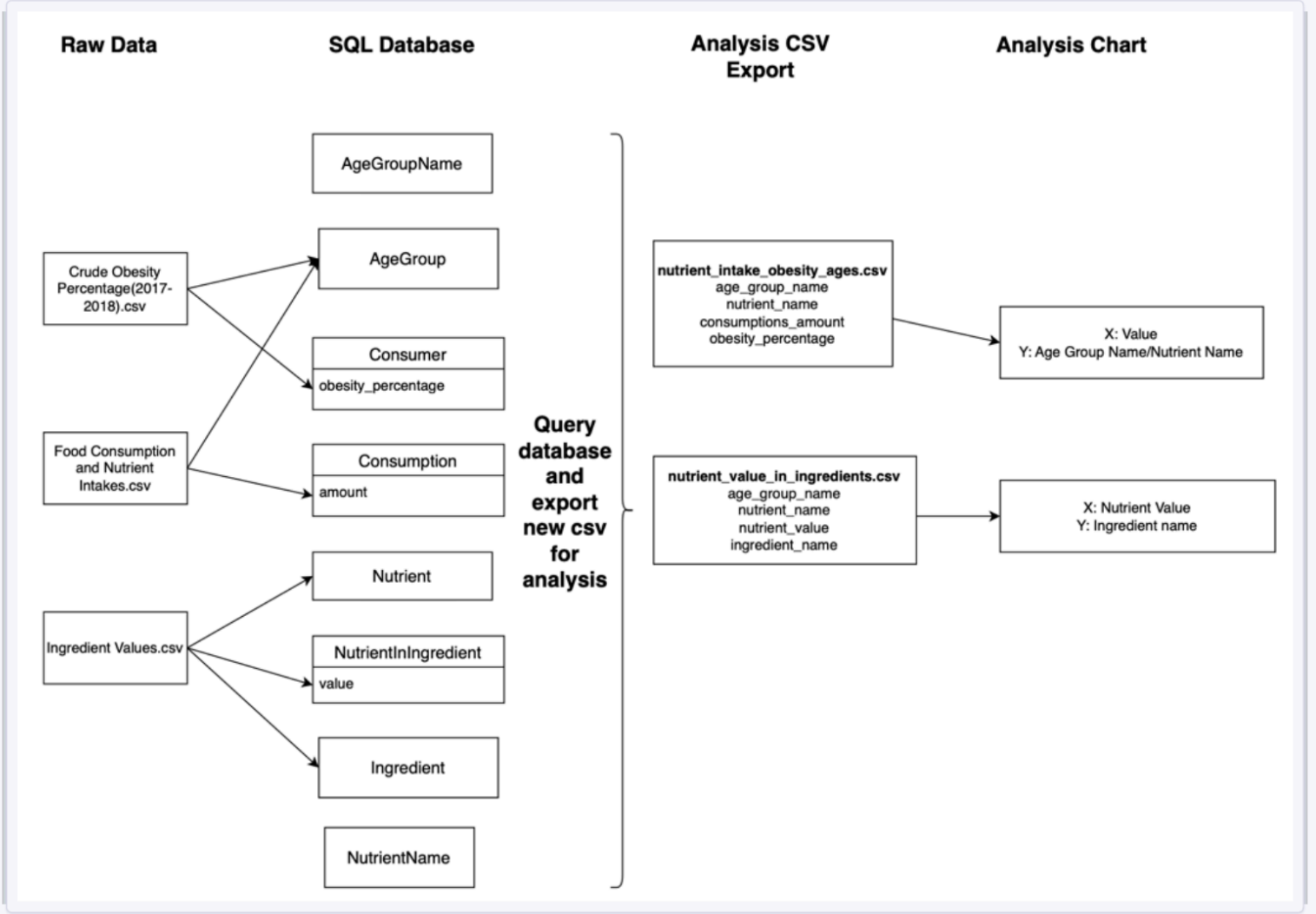
Project Workflow



Data Collection

- Download and convert each .csv datasets from the web.
- Upload to datacamp workplace for access.
- Clean and process the original .csv files by removing unnecessary columns and rows. Save the refined data into a new .csv file for further actions using Python code.
- Manully create new age_groups.csv, age_group_names.csv and nutrient_names.csv files to manage synonyms.

SQL Database Initialization

- Execute SQL commands via Python to create the necessary tables and attributes according to our ER diagram.

Parsing Data and Loading into SQL Database

- Use Python to read 'nutrient_names.csv' and write nutrient names synonyms and correspond nutrient id into **nutrient_names table**
- Read 'ingredient_values_processed.csv' and write ingredient names and id into **ingredients table.**
- Read 'ingredient_values_processed.csv' and write nutrient names and id into **nutrients table.**
- Read 'age_groups.csv' and write in new age group names we created and id into **age_groups table.**
- Read 'age_group_names.csv' and write all age_group_name synonyms into names column and their corresponding age_group_id in **age_group_names table.**
- Read 'age_group_names.csv' and 'obesity_age_processed'. Replace the age_group in obesity_age_processed.csv to age_group_id from age_group_names.csv and create **consumer table.**
- Read 'ingredient_values_processed.csv' write ingredient_id, nutrient_id and value in **nutrient_in_ingredients table.**
- Read 'age_group_names.csv' and 'nutrient_names.csv' to create look up in order to replace the original synonyms. Read 'nutrient_intake_processed.csv' and write amount, age_group_id and nutrient_id into **consumptions table.**

Export for Analysis

- Query age_group_names, age_groups, consumers, consumptions tables to compile standard age_group_name, nutrient_name, consumptions_amount and obesity_percentage in **'nutrient_intake_obesity_ages.csv' file.**
- Query age_group_names, age_groups, consumptions, nutrients, nutrient_names, nutrient_in_ingredients, ingredients table to compile standard nutrient name, nutrient_value and ingredient_name. Filter 'Adult' age_group and create **'nutrient_value_in_ingredients.csv' file.**

Analysis Chart

- Upload .csv files to Tableau and produce charts.