Food category, major nutrients and obesity

Project description
This idea aims to analyze the nutrients contents like fat, saturated fat and cholesterol in dairy products so as to determine if their consumption leads to obesity. By doing this analysis we can make better food choices next time picking diary products. This analysis will help cut down on the consumption of dairy products and contribute towards a healthier lifestyle.

Project goals
Which ingredients in dairy products are the least healthy and most related to obesity?

Intended analysis
First we plan to get the nutrients value of the dairy products(butter, cheese, cream, yogurt, ice cream), then we are going to check the annual consumption of those nutrients in an age demographic. Lastly, we are going to see the obesity rate corresponding to age.

Datasets
For this project, three datasets were used:

1.(NIH)https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity(2017-2018)
This dataset was downloaded as a PDF. We converted into a CSV using an online converter(https://www.zamzar.com/convert/pdf-to-csv/).

```
SELECT *
  FROM 'obesity_ages.csv';
```

	column0	column1	column2	column3	column4
0	Data Brief 360. Prevalence of …	null	null	null	null
1	Data table for Figure 1. Prevale…	null	null	null	null
2	age: United States, 2017–2018	null	null	null	null
3	null	null	Age group	null	null
4	Sex	20 and over	20–39	40–59	60 and over
5	null	null	Percent (standard error)	null	null
6	Total	42.4 (1.8)	40.0 (2.6)	44.8 (1.9)	42.8 (2.5)
7	Men	43.0 (2.7)	40.3 (3.8)	46.4 (3.2)	42.2 (3.3)
8	Women	41.9 (2.0)	39.7 (2.7)	43.3 (2.7)	43.3 (3.0)
9	NOTES: Estimates for adults a…	null	null	null	null

2.https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nut rition-research-center/food-surveys-research-group/docs/fndds-download-databa ses/
This dataset was downloaded as an excel, we converted into a CSV file by an online converter(https://cloudconvert.com/xlsx-to-csv)

```
SELECT * FROM 'ingredient_values.csv'
```

	column0 ⌄	column1 ⌄	column2 ⌄	column3 ⌄	column4 ⌄
0	Ingredient Nutrient Values 2019-2020 Food and Nutrient ...	null	null	null	null
1	Ingredient code	Ingredient description	Nutrient code	Nutrient description	Nutrient v...
2	1001	Butter, stick, salted	203	Protein	0.85
3	1001	Butter, stick, salted	204	Total Fat	82.2
4	1001	Butter, stick, salted	205	Carbohydrate	0.06
5	1001	Butter, stick, salted	208	Energy	743
6	1001	Butter, stick, salted	221	Alcohol	0
7	1001	Butter, stick, salted	255	Water	15.8
8	1001	Butter, stick, salted	262	Caffeine	0
9	1001	Butter, stick, salted	263	Theobromine	0

Table   Chart   « < 1 of 500 > »

Rows per page 10 ⌄   122,332 rows   ⚠ truncated to 5,000 ⤓

3.https://www.ers.usda.gov/data-products/food-consumption-and-nutrient-intakes.aspx
This dataset was downloaded as an excel, we converted into a CSV file by an online converter(https://cloudconvert.com/xlsx-to-csv)

```
SELECT * FROM 'nutrient_intakes.csv'
```

	column00 ⌄	column01 ⌄	column02 ⌄	column03 ⌄	column04 ⌄	column05 ⌄
0	Nutrient intake by food sourc...	null	null	null	null	null
1	null	null	null	null	null	Away from h...
2	Nutrient group	Total	At home	Total	Restaurant	Fast food
3	null	null	null	null	2015-16	null
4	Energy (calories)	null	null	null	null	null
5	Total population1	2048.24	1369.60	678.64	178.61	330.94
6	Children age 2–19	1868.78	1180.14	688.64	112.21	291.52
7	Adults age 20–642	2187.00	1444.17	742.83	208.91	396.53
8	Seniors age 65 and above2	1798.29	1374.88	423.41	164.91	144.65
9	Household income < 185%...	1992.09	1373.99	618.10	133.58	292.65

Table   Chart   « < 1 of 8 > »   Rows per page 10 ⌄ 72 rows ⤓

For the nutrients_intake table - This dataset has multiple CSV files combined into one. (Each nutrient was present in a separate csv and then its consumption is mentioned for different age groups). We are going to check for the heading for each nutrient (Eg: Energy) and look if the consecutive cells are empty or not. If they are empty then we are going to extract that heading value and add it as a column with the same value in each row at the end. We are going to manage this process via python code.

Conceptual ER Diagram
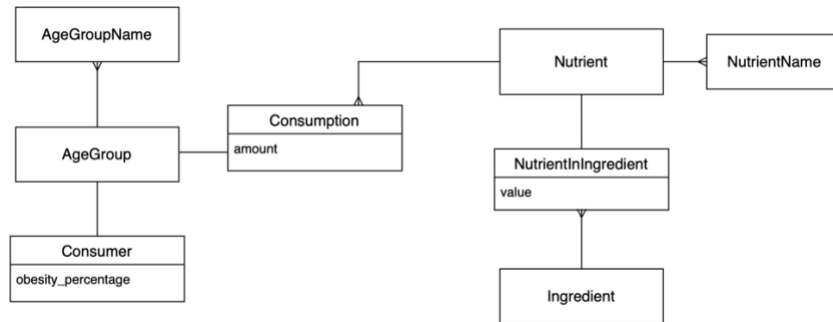
Managing Synonyms

We manage the synonyms for different nutrient names in separate tables by creating another Nutrient entity, and turning different nutrient names into a distinct nutrient id.
As for the age that was divided into different groups and thus couldn't match in analysis. We manage that by creating a new AgeGroup entity that describes the age range into names, and turning them into age group ids.
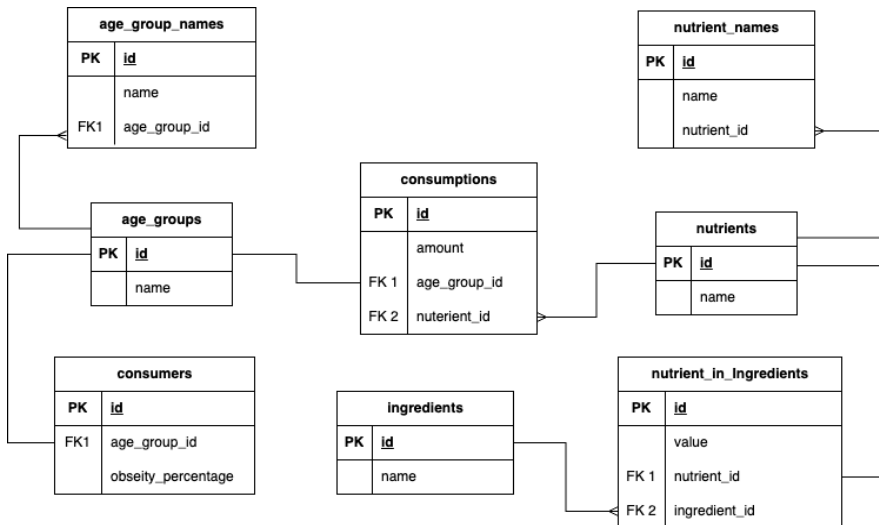
Physical ER Diagram

Table Sketch

consumers

id	age_group_id	obesity_percentage
1	1	42.4
2	2	40.0
3	2	44.8
4	3	42.8

age_groups

id	name
1	Children
2	Adults
3	Seniors

age_group_names

id	name	age_group_id
1	20-64	2
2	20-39	2
3	40-59	2
4	Adults	2
5	2-19	1
6	<20	1
7	Children	1
8	>60	3
9	>65	3
10	Seniors	3

consumptions

id	nutrient_id	age_group_id	amount
1	601	1	225.12
2	606	1	25.72
3	204	1	73.10
4	601	2	312.65
5	606	2	28.3
6	204	2	86.86

nutrients_in_ingredients

id	ingredient_id	nutrient_id	value
1	1001	601	82.2
2	1001	606	45.6
3	1001	204	235

ingredients

id	name
1001	Butter, stick, salted
1002	Butter, whipped, with salt
1003	Butter oil, anhydrous

nutrients

id	name
601	Cholesterol
606	Saturated Fat
204	Total Fat

nutrient_names

id	name	nutrient_id
1	Cholesterol	601
2	Fatty acids, total saturated	606
3	Fatty acids, total monounsaturated	606
4	Fatty acids, total polyunsaturated	606
5	Saturated Fat	606
6	Total Fat	204