CS306 Project

Project Repository: SleepyCoffy/CS-306-BasedData-Project: CS 306 Database project that will use data from www.ourworldindata.com and MySQL as a main tool. Project Group name: BasedData. (github.com)

Group Name: **BasedData**

Group Members:

Name-Surname	Student ID	E-mail
Emir Aikan	28614	aikan@sabanciuniv.edu
Safa Abdullah Söğütlügil	29214	safasogutlugil@sabanciuniv.edu
Yunus Emre Şencan	29562	sencanyunus@sabanciuniv.edu
Yusuf Erkut Bayram	28976	yusuferkut@sabanciuniv.edu
Yusuf Kılıç	29431	yusufk@sabanciuniv.edu

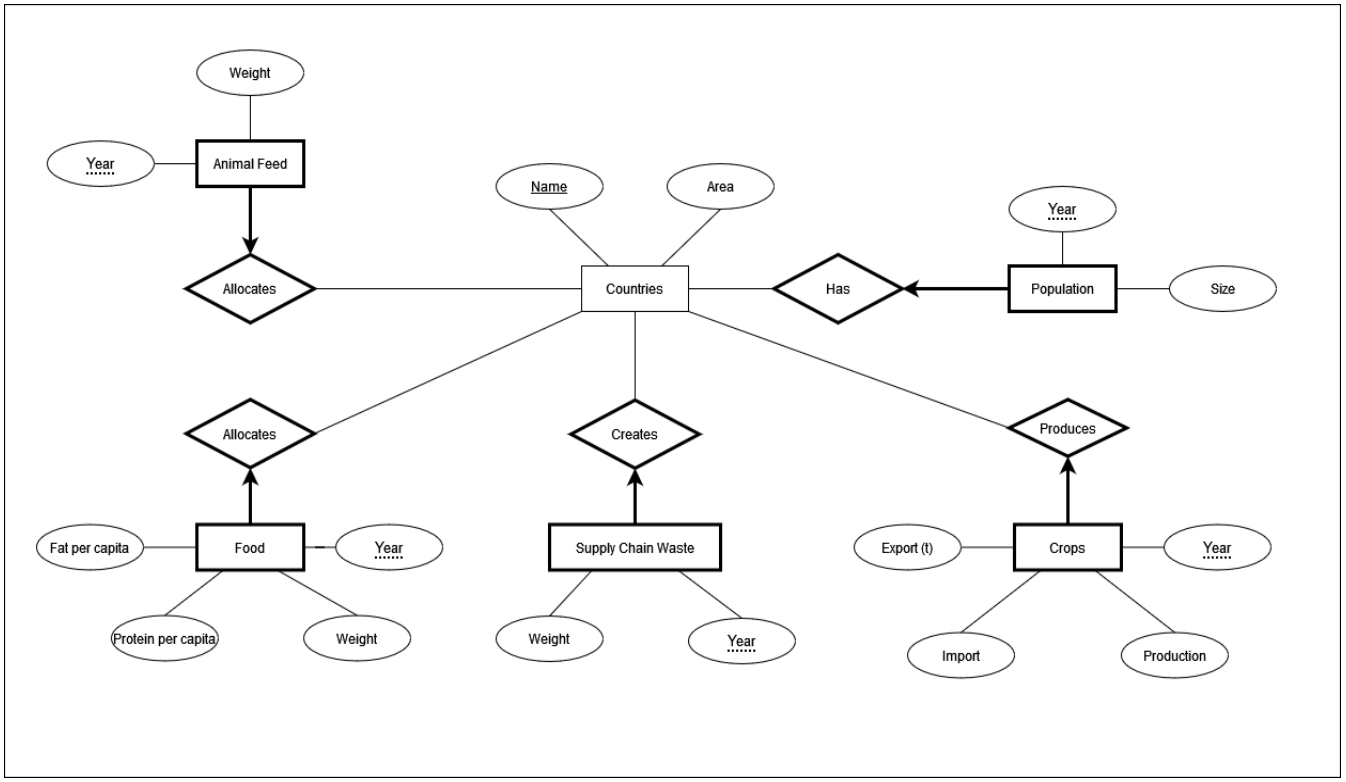
Agriculture and Food Supply Database

Project Description

This database application will be about Agricultural status, Food Supply-Wastage and food sufficiency of all countries in the World. The database will store information about countries and their annual figures regarding food production, food wastage, population, usable area, animal food, food's quality, etc. This project addresses Global Famine, Global Food Wastage and sustainability of Food Production, which are some of the most important main global problems in modern times.

The data is taken as a CSV file from "Agricultural Production" under the "Food and Agriculture" topic from https://ourworldindata.org/agricultural-production

Entity Relationship Diagram



Explanation of the Entity Relationship Diagram

1. **Identification of CSV files**
    a. ‘Countries’ entity is identified as Country_Entity.csv, which includes attributes Country_names and Area. Country_names attribute contains unique names of each state, no country names are equal. Area attribute refers to countries’ land sizes in hectare scale.
    b. ‘Population’ entity is identified as Population_Entity.csv, which includes attributes Year and Size. Year attribute contains sequential annual report dates for each country. All figures in Year attribute for a specific country will be unique. Size attribute contains numbers of population of each country.
    c. ‘Food’ entity is identified as Food_Entity.csv, which includes Year, Food, Protein and Fat attributes. Food attribute refers to total produced crops that were allocated as human food in a given year. Protein and Fat attributes correspond to nutrition values of produced food in terms of grams per capita per day.

d. ‘Animal Feed’ entity is identified as Animal_Feed.csv, which includes attributes Weight and Year. Weight refers to total produced crops that were allocated as animal feed in a given year. Year attribute contains sequential annual report dates for each country.
e. **‘**Crops’ entity is identified as Crops_Entity.csv, which includes attributes Year, Production (t), Imports (t), Exports (t). Production attribute refers to crops produced during that year according to the Year attribute. Imports(t) and Exports (t) attribute refers to the goods imported and exported during that year.
f. ‘Supply Chain Waste’ entity is identified as Supply_Chain_Waste.csv, which includes Year and Weight attributes. Weight attribute refers to the total weight of products that were wasted in the supply chain in terms of tonnes in a given year.

2. Relationship of Entities

a. ‘Countries’ entity is the main entity in this project, it has no Unary relationship with itself.
b. ‘Animal Feed’ entity has a many-to-one relationship with the ‘Countries’ entity called ‘Allocates’, additionally the relationship includes participation constraint from the ‘Animal Feed’ entity.
c. ‘Population’ entity has a many-to-one relationship with the ‘Countries’ entity called ‘Has’, additionally the relationship includes participation constraint from the ‘Population’ entity.
d. ‘Crops’ entity has a many-to-one relationship with the ‘Countries’ entity called ‘Produces’, additionally the relationship includes participation constraint from the ‘Crops’ entity.
e. ‘Supply Chain Waste’ entity has a many-to-one relationship with the ‘Countries’ entity called ‘Creates’, additionally the relationship includes participation constraint from the ‘Supply Chain Waste’ entity.

f. ‘Food’ entity has a many-to-one relationship with the ‘Countries’ entity called ‘Allocates’, additionally the relationship includes participation constraint from the ‘Food’ entity.

3. **Constraints**
    a. For all entity sets except ‘Countries’, Country Name and Year will together be primary keys. Primary key for the entity set ‘Countries’ will be Country Name.
    b. Name and Year will be set to NOT NULL ON DELETE CASCADE

Explanation of Data Cleansing Steps

1. We removed the rows which included general regions in their ‘Country’ column, such as, continents, world, organizations, EU, income related regions etc. We only kept countries. This allowed us to delete 2.500 lines that we found redundant from over 13.000 entries.

2. We deleted the year 2020 from the dataset as it was empty for all countries.

3. We also tried deleting the duplicates, however since we used year and country information together to form an identifying key, no duplicates were found.