Data Analytics

Waste Less, Taste More

A Deep Dive into Food Waste in Europe &
Using Machine Learning to Cook Up Solutions

Kranta Gaterojn

April, 2024

Table of content

Introduction

In an era where sustainability and resource conservation have become paramount, the issue of food waste stands out as a significant challenge. Each year, a staggering amount of food is discarded globally, contributing to environmental degradation and economic inefficiency. In light of this, my project, titled "Waste Less, Taste More: A Deep Dive into Food Waste in Europe & Using Machine Learning to Cook Up Solutions," aims to tackle the pervasive problem of food waste, with a primary focus on household wastage within European countries.

Aligned with the United Nations Sustainable Development Goals (SDGs), particularly Goal 12: "Responsible Consumption and Production," my project is dedicated to addressing Target 12.3: "By 2030, halve per capita global food waste at the retail and consumer levels and reduce food losses along production and supply chains, including post-harvest losses."

Food waste is a pressing issue with far-reaching implications for both society and the environment. Despite the abundance of resources devoted to food production, distribution, and consumption, a significant portion ends up in landfills, contributing to greenhouse gas emissions, resource depletion, and economic inefficiency. Therefore, I aim to contribute to the global effort to reduce food waste by focusing on household wastage within European countries.

Utilizing data from authoritative sources such as Eurostat and the United Nations, I seek to quantify the extent of food waste and identify key drivers and patterns across the food supply chain. By doing so, I hope to shed light on the root causes of food waste and develop targeted interventions and strategies for waste reduction, in line with the principles of sustainable consumption and production outlined in Goal 12.

Through my project, I aspire to demonstrate the transformative potential of data analytics and machine learning in addressing complex societal challenges. By harnessing the power of data-driven insights, I aim to empower stakeholders with actionable information and innovative solutions to achieve the ambitious targets set forth by the United Nations, paving the way for a more sustainable and resilient future for generations to come.

Business Case:

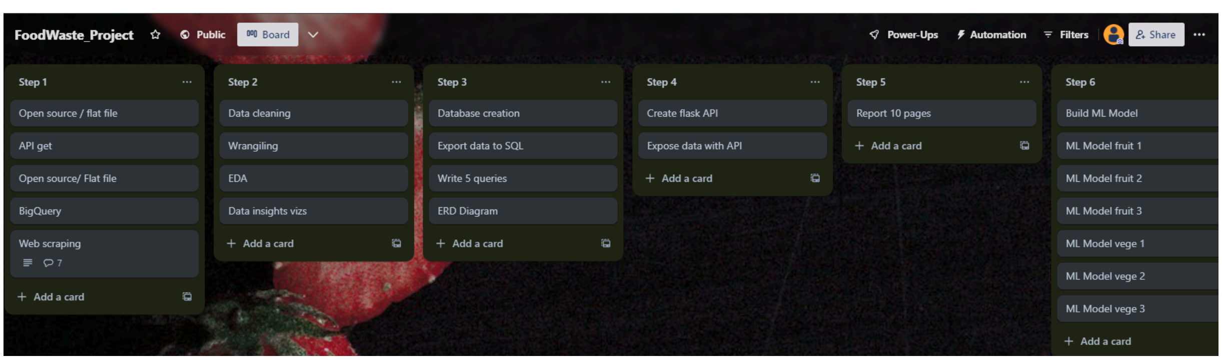
The economic and environmental implications of food waste are substantial, with billions of dollars lost annually and significant greenhouse gas emissions associated with decomposing food in landfills. By focusing on household food waste, I aim to conserve valuable resources, mitigate environmental pollution, and contribute to the global effort towards sustainability. Moreover, addressing household food waste presents an opportunity to instill responsible consumption habits and promote a culture of sustainability within communities. Through my project, leveraging data analytics and machine learning, I seek to showcase the effectiveness of data-driven approaches in reducing food waste, inspiring broader adoption of sustainable practices worldwide.

Goal:

My aim is to reduce food waste, emphasizing households and addressing waste across various stages of the food supply chain. Leveraging data analytics and machine learning techniques, I'll identify key drivers and patterns of food wastage within households, develop targeted interventions, and utilize machine learning to create recipes for leftover food. These efforts aim to promote sustainable consumption practices and contribute to building a more efficient food system for future generations.

Project Plan:

- Planning the project on Trello
- Collecting data using various methods such as web scraping, flat files, APIs, and BigQuery
- Cleaning the collected data
- Creating a database and Entity Relationship Diagram (ERD) using MySQL
- Aggregating the data within MySQL
- Creating APIs with Swagger documentation to expose the collected data
- Processing data for machine learning purposes
- Training and testing models



Data and data sources

Flat file:

Eurostat

In my project, I begin by gathering comprehensive data on food waste across European countries for the years 2020 and 2021. Utilizing datasets from authoritative sources such as Eurostat ensures the reliability and accuracy of my analysis.

For a deep dive into food waste, I further categorize it into specific categories, including food production, manufacturing, distribution, services, and total household activities. This granular analysis allows me to pinpoint key areas with the highest levels of waste generation.

Eurostat is a primary source of statistical information on waste generation, consumption, and related indicators. Specifically, I rely on Eurostat's data on food waste and food waste prevention by NACE Rev. 2 activity - tonnes of fresh mass to enrich my analysis and provide valuable insights into food waste trends.

Eurostat's data on food waste and food waste prevention by NACE	
NACE_R2 (Labels)	Country name
Total	Total of 5 type of food waste subcategory
Primary production of food - agriculture, fishing and aquaculture	Food waste from production process
Manufacture of food products and beverages	Food waste from manufacturing process
Retail and other distribution of food	Food waste from retail or distribution
Restaurants and food services	Food waste from restaurant
Total activities by households	Food waste from household

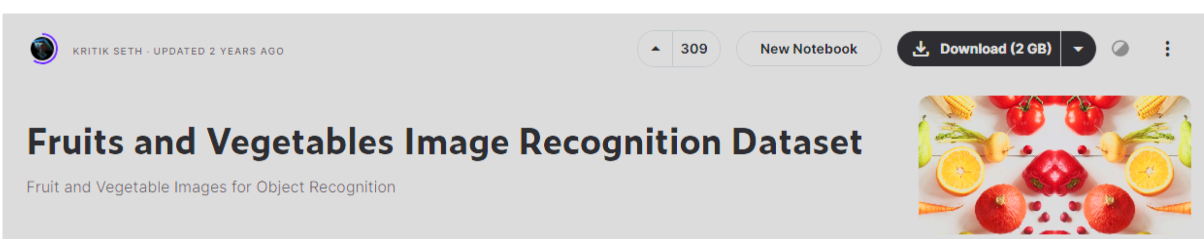
Kaggle

In line with the objectives outlined in the report, I am integrating image data of fruits and vegetables sourced from Kaggle to train and test the machine learning model. This dataset comprises approximately 100 images for training, 10 images for testing, and 10 images for validation, encompassing 36 different types of fruits and vegetables.

By leveraging these images and employing machine learning techniques, I aim to identify key drivers and patterns of food wastage within households. Additionally, I will develop targeted interventions and utilize machine learning to generate recipes for utilizing leftover food effectively.

Through these efforts, I seek to promote sustainable consumption practices and contribute to the establishment of a more efficient and resilient food system for future generations.

Fruits and Vegetables Image Recognition Dataset contains images of	
Vegetables	cucumber, carrot, capsicum, onion, potato, lemon, tomato, raddish, beetroot, cabbage, lettuce, spinach, soy bean, cauliflower, bell pepper, chilli pepper, turnip, corn, sweetcorn, sweet potato, paprika, jalepeño, ginger, garlic, peas, eggplant.
Fruits	banana, apple, pear, grapes, orange, kiwi, watermelon, pomegranate, pineapple, mango.
The images are divided into 3 separate folders:	Train (100 images each)
	Test (10 images each)
	Validation (10 images each)

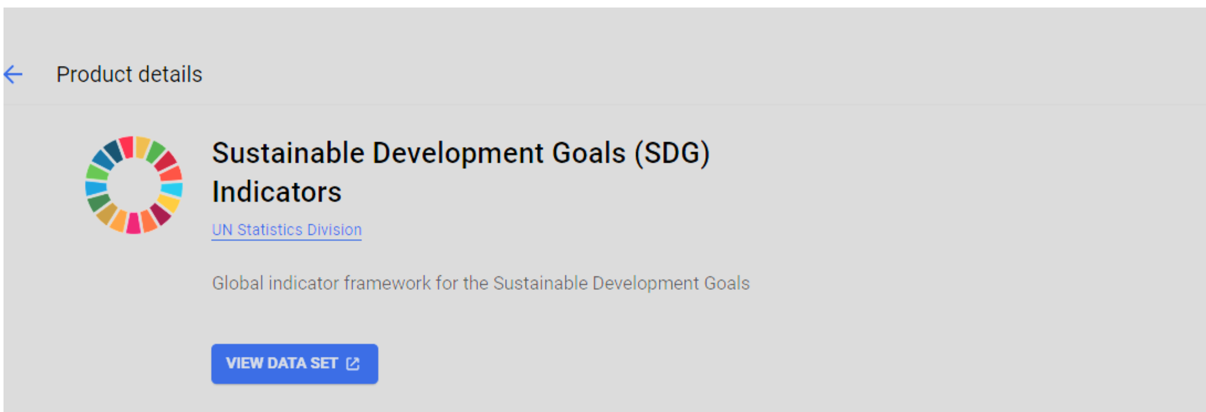


BigQuery:

In my quest to understand food waste and consumption patterns, I tap into various valuable data sources. One of my go-to sources is the United Nations Sustainable Development Goals (SDGs) data repository on Google BigQuery. This treasure trove of information houses a wealth of indicators related to food waste, giving me insights into global trends and how people eat.

With BigQuery's help, I can search through this massive database using keywords like waste, food, hunger, and consumption. This lets me dig out the most relevant data points to enrich my analysis and get a better understanding of what's going on worldwide with food waste.

However, there's a little hiccup when it comes to exporting data directly to my computer in CSV or JSON format. It just doesn't work like that. So, I found a workaround—I export the data to CSV format on Google Drive and then download it from there. It's a bit of a detour, but it gets the job done, and I can keep diving into the data to uncover more insights for my project.



Sustainable Development Goals (SDG) Indicators	
Goal	Goal: One of the 17 broad objectives set by the UN General Assembly to address global challenges and achieve sustainable development by 2030.

Target	A specific, measurable objective or aim within a broader goal, as defined in the United Nations Sustainable Development Goals (SDGs).
Indicator	Each target has one or more indicators to measure progress towards achieving the goal
SeriesCode	A unique code assigned to a specific series of data or indicator
SeriesDescription	Description or name of the series of data
GeoAreaCode	Code representing the geographical area for which the data is recorded
GeoAreaName	Name of the geographical area
TimePeriod	Time period during which the data was recorded
Value	The actual value or measurement of the data being recorded
Time_Detail	Additional detail about the time period, such as specific month or quarter
Source	Source or organization responsible for collecting the data
Footnote	Additional information or clarification related to the data
Nature	Nature or type of the data (e.g., survey data, administrative data).
Age	Age group to which the data pertains
Bounds	Bounds or range associated with the data
Cities	Cities related to the data (if applicable)
Education_Level	Level of education associated with the data
Freq	Frequency of data collection or reporting
Hazard_Type	Type of hazard (if applicable)
IHR_Capacity	International Health Regulations (IHR) capacity related to the data

Level_Status	Status or level associated with the data
Location	Location associated with the data
Migratory_Status	Migratory status of individuals (if applicable)
Mode_of_Transportation	Mode of transportation associated with the data
Name_of_International_Institution	Name of the international institution related to the data
Name_of_Non_Communicable_Disease	Name of non-communicable disease (if applicable)
Sex	Gender or sex associated with the data
Tariff_Regime_Status	Tariff regime status associated with the data
Type_of_Mobile_Technology	Type of mobile technology (if applicable)
Type_of_Occupation	Type of occupation associated with the data
Type_of_Product	Type of product associated with the data
Type_of_Skill	Type of skill associated with the data
Type_of_Speed	Type of speed associated with the data
Units	Units of measurement for the data

Data collection

API Data:

Leveraging the UN SDG API for Target 12.3

In my project report, I highlight the utilization of the United Nations Statistics Division's API, tailored specifically for Sustainable Development Goal (SDG) indicators, including Target 12.3 addressing food waste reduction. This API, known as the UN SDG API, serves as a pivotal tool in my endeavor to combat food waste.

With the UN SDG API, I have direct access to real-time and historical data on various metrics related to food waste, empowering me to monitor progress, identify trends, and make informed decisions regarding waste reduction strategies. Through personalized Python scripting, I fetch data from the API, allowing me to seamlessly integrate this valuable information into my analysis.

In one Python code snippet from my report, I illustrate how I iteratively retrieve JSON data from the UN SDG API, aggregating it into a cohesive DataFrame using the powerful Pandas library. This consolidated dataset serves as the foundation for my subsequent data analysis and modeling efforts.

By harnessing the capabilities of the UN SDG API, I gain valuable insights into food waste metrics across different regions and time periods. This personalized approach to data collection not only enhances the accuracy of my analysis but also aligns with my commitment to leveraging cutting-edge technology to address pressing societal challenges, such as food waste reduction.

Leveraging the UN SDG API for Target 12.3	
Goal	The broad objective or aspiration of the Sustainable Development Goals (SDGs). Each goal represents a thematic area of sustainable development.
Target	A specific, measurable objective or aim within a broader goal. Targets provide more detailed milestones for achieving the goals.
Indicator	A quantitative or qualitative variable used to measure progress toward achieving a specific target. Indicators help assess the performance and impact of policies and interventions.
Series	A group or category of related data or statistics. Series may include multiple indicators or variables that share common characteristics or themes.
Series Description	Description or name of the series of data. It provides additional context or information about the series.
Series Count	The number of data series or indicators within a particular category or group.
Geo Area Code	Code representing the geographical area for which the data is recorded.
Geo Area Name	Name of the geographical area.
Time Period Start	The starting point of the time period for which the data is reported.
Value	The actual value or measurement of the data being recorded.
Value Type	The type or format of the value (e.g., numerical, categorical).
Time Detail	Additional detail about the time period, such as specific month, quarter, or year.
Time Coverage	The extent or duration of the time period covered by the data.
Upper Bound	The upper limit or maximum value associated with the data.
Lower Bound	The lower limit or minimum value associated with the data.

Base Period	The reference period or baseline against which changes or trends are measured.
Source	Source or organization responsible for collecting and providing the data.
Geo Info URL	URL or link providing additional geographical information or context.
Footnotes	Additional information or clarifications related to the data.
Attributes	Additional attributes or metadata associated with the data.
Dimensions	Dimensions or factors influencing the data, such as demographic characteristics, geographic regions, or time periods.

Web scraping

Web scraping techniques are essential for me to extract targeted data from online sources. These sources include consumer surveys, scientific studies, and lists of national fruits for each country. They provide valuable insights into consumer preferences, regional variations in food waste, and cultural factors influencing consumption patterns.

By scraping consumer surveys and reports on the top 20 unwanted fruits and vegetables, I gain insights into potential sources of food waste and consumer behavior. Additionally, scientific studies detailing country-specific food waste generation profiles offer crucial data for understanding regional variations in waste.

Despite encountering challenges with inaccessible PDFs and restricted sites, such as the 403 forbidden error, I persist in enriching my dataset with valuable information. Incorporating data on national fruits allows me to understand cultural influences on consumption patterns and identify unique sources of food waste in each region.

This knowledge empowers me to tailor waste reduction strategies to specific cultural contexts, facilitating more effective interventions and promoting sustainable consumption practices. Overall, the inclusion of national fruit data adds depth to my dataset, enriching my analysis and decision-making processes.

List 20 unwanted fruit and vegetable	Salad leaves (bagged), Bananas,Lettuce (whole), Cucumber, Tomatoes, Carrots, Mushrooms, Potato, Grapes, Strawberries, Spinach, Apples, Oranges, Broccoli, Avocado, Celery, Raspberries, Onions, Cabbage, Blueberries

List of national fruit	A "national fruit" is a fruit that is officially recognized as a symbol of a particular country or nation. It holds cultural, historical, or symbolic significance and may be designated as such through official channels such as government declarations or cultural institutions.
Grown and thrown: Exploring approaches to estimate food waste in EU countries	Food waste composition which are subdivided into these categories: Meat, Fish, Dairy, Eggs , Cereals, Fruits, Vegetables, Potatoes, Sugarbeets, Oil crops, Total

Data cleaning and Exploratory data analysis

Data Cleaning

During the data cleaning process, I undertook several steps to refine the dataset and prepare it for analysis, focusing on food waste patterns within European countries. Here are the key steps:

1. **Drop Empty Columns:** I removed empty columns to streamline the dataset and eliminate unnecessary information, ensuring a more focused analysis.
2. **Standardization of Country Names:** To maintain consistency across the dataset, I standardized country names, replacing variations with a uniform format. For instance, "United Kingdom of Great Britain and Northern Ireland" was replaced with "United Kingdom."
3. **Filtering for European Countries:** I filtered the dataset to include only European countries, aligning with the project's scope and objectives.
4. **Column Removal:** Irrelevant columns were dropped from the dataset to simplify it and focus solely on essential information relevant to food waste analysis.
5. **ID Generation:** I generated unique identifiers ('id') for each dataset entry based on country name and time period, facilitating efficient data management and analysis.
6. **Rename Columns:** Column labels were renamed to provide clearer and more descriptive names, enhancing the dataset's interpretability and ease of use.
7. **Target Data Filtering:** I filtered and retained data related to the project's target variables, ensuring that only relevant information was included in the dataset.
8. **Data Segmentation:** The dataset was segmented into separate DataFrames based on specific conditions, allowing for more focused analysis on each subset of data and enabling deeper insights into food waste patterns.
9. **Handling Missing Values:** Columns with missing or NaN values were identified and dropped from the dataset to maintain data integrity and accuracy.

10. **Standardization and Formatting:** Column names were standardized by replacing spaces with underscores and converting all letters to lowercase, ensuring consistency and facilitating smoother data manipulation and analysis.

By meticulously following these cleaning steps, I refined the dataset, ensuring that it contained only relevant and consistent information essential for investigating food waste trends within European households. This meticulous preparation laid a solid foundation for subsequent analysis and deriving actionable insights to address food waste challenges effectively.

Original shape

```
Shapes of original DataFrames:
df_kilo_capita: (54, 14)
df_composition: (6, 13)
df_api_12_3: (4744, 21)
df_national_fruit_o: (70, 2)
df_bq_keywords_eu: (62679, 10)
```

Shape after cleaning

There are more dataframe after cleaning because df_bq_keywords_eu from bigquery is subdivided into 4 different dataframe based on its main target

```
Shapes of clean DataFrames:
df_kilo_capita_c: (54, 9)
df_composition_c: (6, 12)
df_12_3_eu: (927, 10)
df_country_region: (49, 2)
df_national_fruit: (22, 2)
df_severe_percent: (75, 10)
df_moderate_food_percent: (75, 10)
df_severe_population: (75, 10)
df_moderate_population: (75, 10)
```

Original data type for df_kilo_capita_c

```
NACE_R2 (Labels)                                                         object
Total                                                                    object
NaN                                                                      object
Primary production of food - agriculture, fishing and aquaculture        object
NaN                                                                      object
Manufacture of food products and beverages                               object
NaN                                                                      object
Retail and other distribution of food                                    object
NaN                                                                      object
Restaurants and food services                                            object
NaN                                                                      object
Total activities by households                                           object
NaN                                                                      object
year                                                                      int64
dtype: object
```

Cleaned data type for df_kilo_capita_c

```
id                    object
country               object
total                float64
food_production      float64
food_manufacture     float64
food_distribution    float64
food_services        float64
households           float64
year                   int64
dtype: object
```

Example: Original columns for df_kilo_capita_c

	NACE_R2 (Labels)	Total	NaN	Primary production of food - agriculture, fishing and aquaculture	NaN	Manufacture of food products and beverages	NaN	Retail and other distribution of food	NaN	Restaurants and food services	NaN	Total activities by households	NaN	year
0	Belgium	250	NaN	3	NaN	161	NaN	6	NaN	8	NaN	71	NaN	2020
1	Bulgaria	108	NaN	9	NaN	19	NaN	7	NaN	18	NaN	55	NaN	2020
2	Czechia	91	e	3	NaN	9	NaN	6	e	4	e	69	e	2020
3	Denmark	221	NaN	11	NaN	102	NaN	17	NaN	11	NaN	79	NaN	2020
4	Germany	131	NaN	2	NaN	19	NaN	9	NaN	22	NaN	78	NaN	2020
5	Estonia	125	NaN	18	NaN	24	NaN	15	NaN	8	NaN	61	NaN	2020
6	Ireland	154	NaN	11	NaN	44	NaN	14	NaN	36	NaN	48	NaN	2020
7	Greece	191	NaN	35	e	35	NaN	14	d	21	e	87	NaN	2020
8	Spain	90	NaN	18	e	30	NaN	7	e	4	e	30	d	2020
9	France	129	NaN	18	e	26	NaN	10	NaN	16	NaN	60	e	2020
10	Croatia	71	NaN	10	d	2	NaN	1	NaN	4	NaN	53	NaN	2020

Cleaned columns for df_kilo_capita_c

	id	country	total	food_production	food_manufacture	food_distribution	food_services	households	year
0	belgium2020	Belgium	250	3	161	6	8	71	2020
1	bulgaria2020	Bulgaria	108	9	19	7	18	55	2020
2	czechia2020	Czechia	91	3	9	6	4	69	2020
3	denmark2020	Denmark	221	11	102	17	11	79	2020
4	germany2020	Germany	131	2	19	9	22	78	2020
5	estonia2020	Estonia	125	18	24	15	8	61	2020
6	ireland2020	Ireland	154	11	44	14	36	48	2020
7	greece2020	Greece	191	35	35	14	21	87	2020

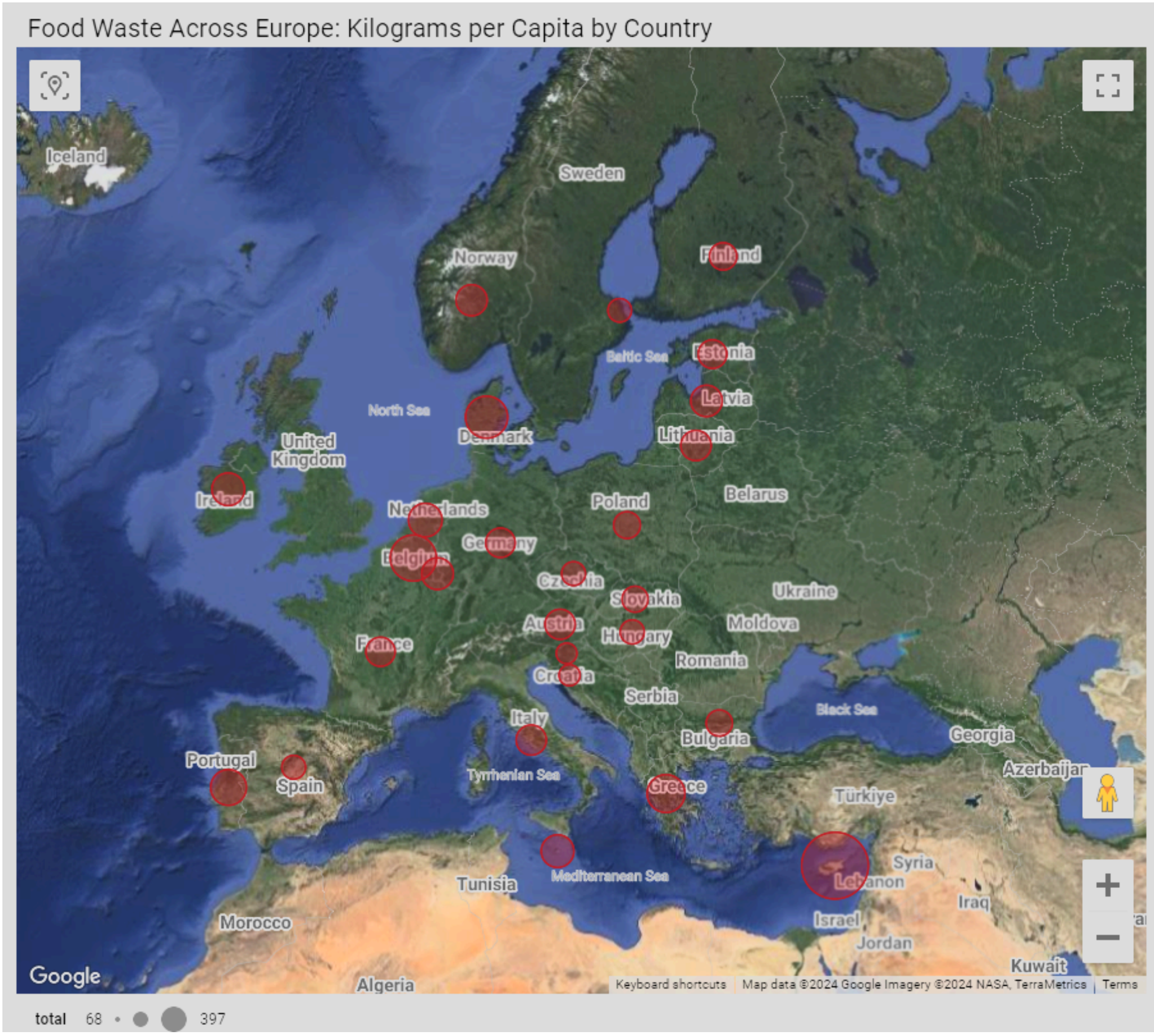
Describe for df_kilo_capita_c

	total	food_production	food_manufacture	food_distribution	food_services	households	year
count	48.000000	48.000000	48.000000	48.000000	48.000000	48.000000	54.000000
mean	139.145833	12.687500	32.645833	10.666667	13.083333	69.916667	2020.500000
std	57.933560	10.567224	43.105259	8.048611	9.948449	20.199888	0.504695
min	68.000000	0.000000	2.000000	1.000000	1.000000	30.000000	2020.000000
25%	105.250000	3.000000	9.000000	7.000000	4.750000	59.000000	2020.000000
50%	131.000000	11.000000	19.000000	9.000000	13.000000	66.000000	2020.500000
75%	148.500000	18.000000	29.250000	12.250000	18.000000	82.250000	2021.000000
max	397.000000	49.000000	190.000000	56.000000	45.000000	124.000000	2021.000000

Describe for df_composition

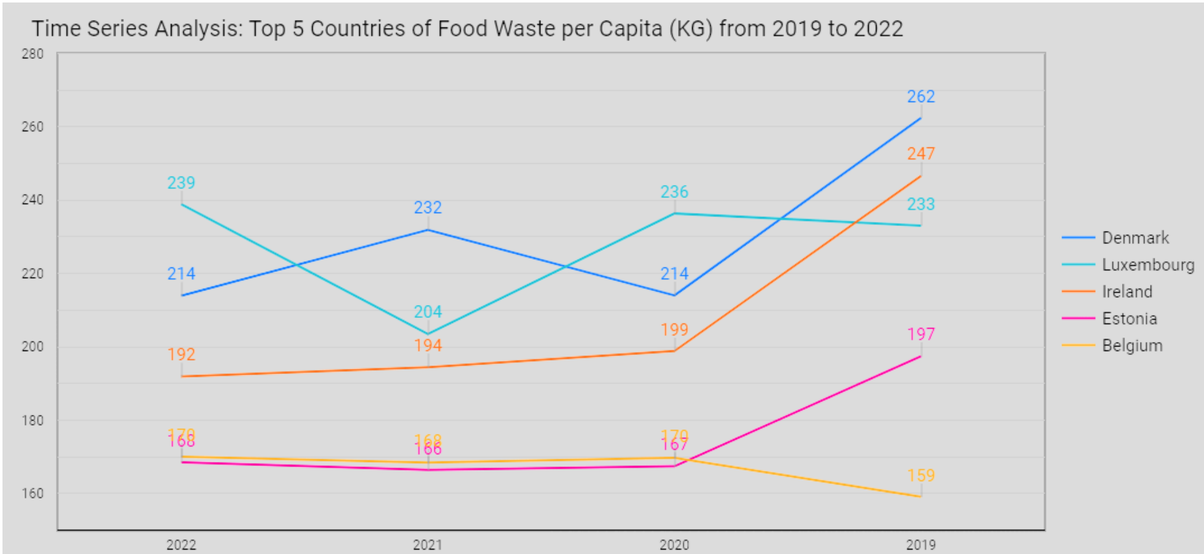
	meat	fish	dairy	eggs	cereals	fruits	vegetables	potatoes	oilcrops	total
count	6.00000	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000
mean	58.39500	67.838333	89.851667	10.405000	56.701667	289.280000	348.735000	35.480000	12.725000	969.045000
std	84.14882	38.652710	77.323586	7.075571	78.395937	359.599126	326.955547	77.068168	23.533318	742.054399
min	0.00000	27.890000	14.900000	1.350000	2.060000	2.020000	24.050000	0.000000	0.000000	148.730000
25%	0.00000	42.827500	35.995000	4.455000	17.062500	30.670000	87.530000	0.000000	0.180000	302.080000
50%	10.06500	62.260000	66.395000	14.005000	31.850000	117.355000	301.630000	4.725000	0.965000	983.605000
75%	112.69500	76.892500	127.147500	14.675000	44.972500	543.745000	510.982500	10.590000	11.695000	1603.112500
max	186.69000	136.500000	218.160000	16.900000	212.970000	811.300000	868.170000	192.460000	59.230000	1813.810000

Data Visualization



The data on food waste across Europe reveals significant variations in kilograms per capita by country:

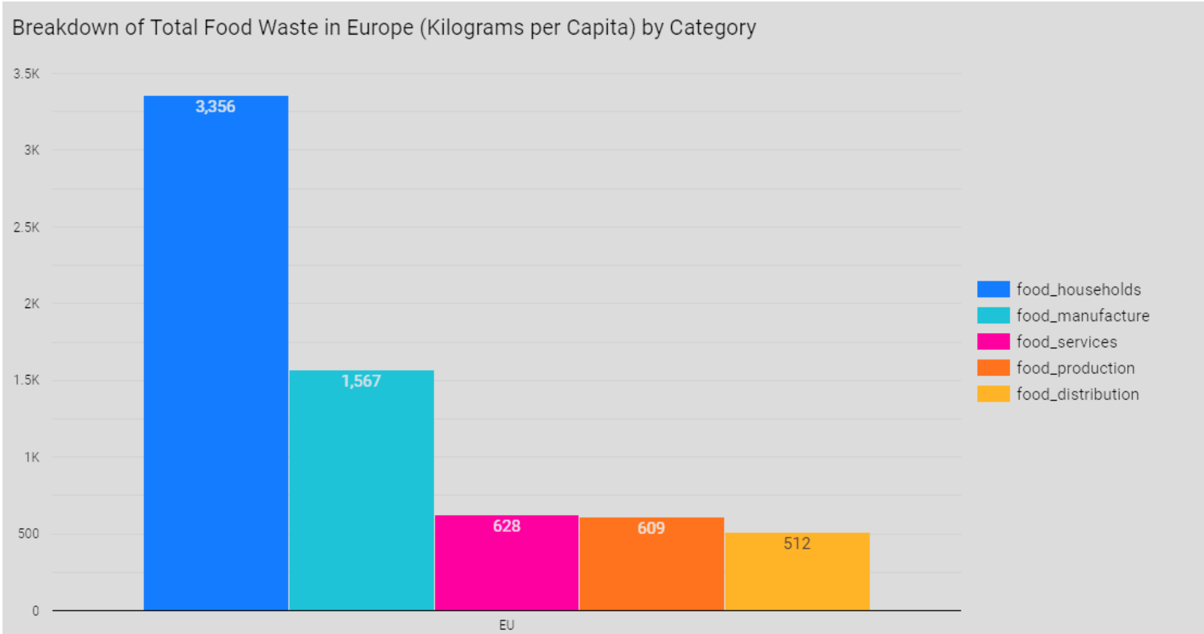
1. Cyprus exhibits the highest food waste level, with 397 kilograms per capita, indicating a substantial challenge in waste management.
2. Several countries, including Belgium, Denmark, and Ireland, also demonstrate high food waste levels, ranging from 221 to 250 kilograms per capita.
3. Moderate food waste levels, ranging from 108 to 191 kilograms per capita, are observed in countries like Bulgaria, Germany, Greece, Italy, and Portugal.
4. Countries such as Slovenia, Croatia, and Slovakia have comparatively lower levels of food waste, ranging from 68 to 106 kilograms per capita.
5. Regional trends show similarities in food waste levels among neighboring countries, emphasizing the importance of tailored waste reduction strategies at both national and regional levels.



The analysis of the top countries is as follows:

1. Denmark: Denmark shows fluctuations in food waste per capita over the years. It had the highest value in 2019 at approximately 262 KG, followed by a decrease to around 214 KG in 2020. However, there was a notable increase in 2021 to about 232 KG. In 2022, the value slightly decreased to approximately 214 KG.
2. Ireland: Ireland exhibits a declining trend in food waste per capita from 2019 to 2021, with values decreasing from around 247 KG to about 194 KG. There is a slight increase in 2022 to approximately 192 KG.
3. Luxembourg: Luxembourg demonstrates fluctuations in food waste per capita over the years, with the highest value in 2022 at approximately 239 KG, followed by values of about 233 KG in 2019, 236 KG in 2020, and 204 KG in 2021.
4. Belgium: Belgium shows relatively stable food waste per capita over the years, with values ranging from approximately 159 KG in 2019 to around 170 KG in both 2020 and 2022, with a slight increase to about 168 KG in 2021.
5. Estonia: Estonia's data shows relatively consistent food waste per capita across the years, with values ranging from around 197 KG in 2019 to approximately 166 KG in both 2021 and 2022.

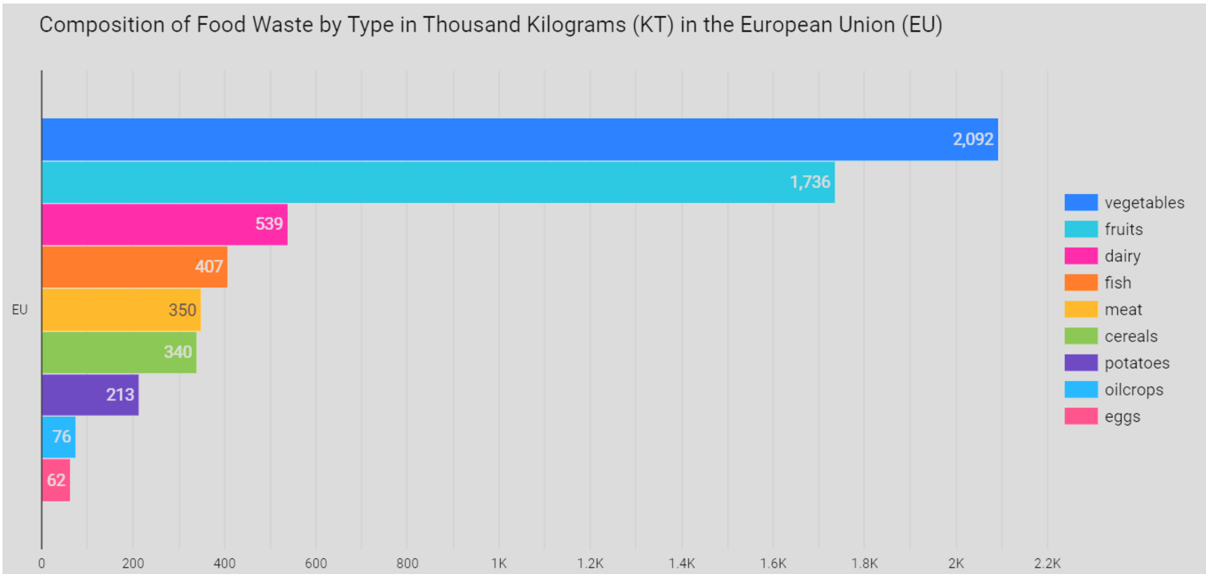
Overall, Denmark and Luxembourg show fluctuations in food waste per capita, while Ireland exhibits a declining trend. Belgium demonstrates relatively stable waste levels, and Estonia maintains consistent values over the years.



Histogram Analysis: Breakdown of Total Food Waste in Europe (Kilograms per Capita) by Category

1. Food Households: Food households contribute the highest and most substantial portion of food waste, with 3,356 kilograms per capita. This underscores the critical role of households in overall food wastage, highlighting the urgent need for targeted interventions to reduce wastage at the consumer level and promote sustainability in food consumption practices.
2. Food Production: Food production contributes 609 kilograms per capita to food waste, indicating a significant portion of waste generated during the production phase.
3. Food Manufacture: Food manufacture accounts for the highest proportion of food waste, with 1,567 kilograms per capita. This highlights the substantial waste generated during the manufacturing process.
4. Food Distribution: Food distribution contributes 512 kilograms per capita to food waste, indicating a notable portion of waste occurring during the distribution stage.
5. Food Services: Food services contribute 628 kilograms per capita to food waste, emphasizing the significant waste generated within the service industry.

Overall: The histogram provides a comprehensive overview of the distribution of food waste across different stages of the supply chain, with households contributing the most significant share. This analysis emphasizes the importance of addressing household food waste through effective strategies and awareness campaigns to achieve sustainable consumption practices.



Histogram Analysis: Composition of Food Waste by Type in Thousand Kilograms (KT) in the European Union (EU)

1. Vegetables: Leading the composition, vegetables contribute approximately 2,092 KT, indicating substantial wastage in this category.
2. Fruits: Following closely, fruits account for around 1,736 KT, underscoring the significant portion of fruit wastage in the EU.
3. Dairy: Dairy products contribute 539 KT to food waste, representing a notable portion of discarded items.
4. Fish: Fish waste amounts to 407 KT, highlighting the need for strategies to minimize seafood wastage.
5. Meat: Meat waste totals 350 KT, indicating a significant but slightly smaller proportion compared to other categories.
6. Cereals, Potatoes, Oilcrops, and Eggs: These categories collectively contribute to food waste, albeit in smaller quantities compared to the top five categories.

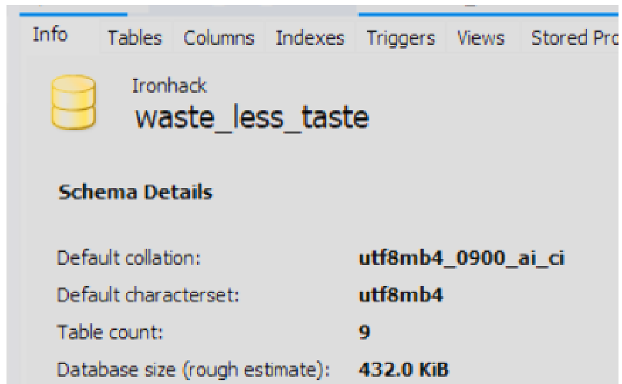
Overall, the histogram provides insights into the distribution of food waste types in the EU, emphasizing the importance of targeted efforts to reduce wastage across various food categories.

Database type selection

Database Creation

The food waste database was established in MySQL Workbench to encompass 14 tables capturing various aspects of food waste in Europe. Data collection involved sourcing information from flat files, APIs, BigQuery, and web scraping. The tables within the food waste database include:

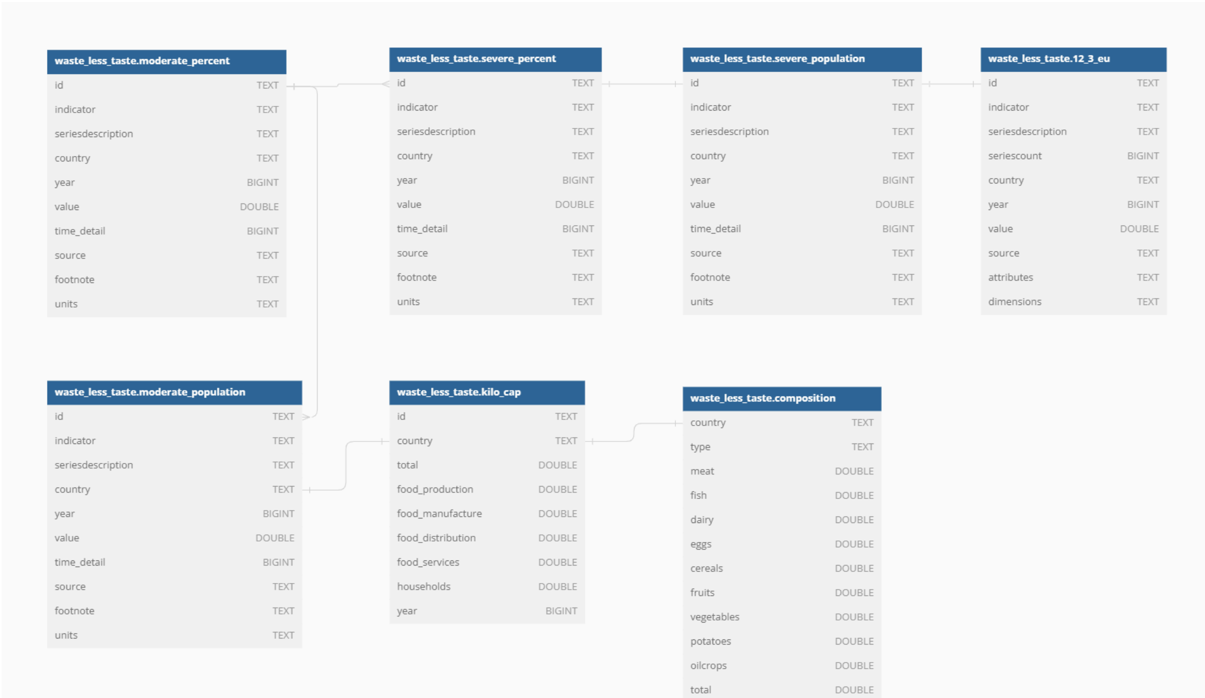
1. df_kilo_cap.csv
2. df_composition.csv
3. df_severe_percent.csv
4. df_moderate_percent.csv
5. df_severe_population.csv
6. df_moderate_population.csv
7. df_12_3_eu.csv
8. df_national_fruit.csv
9. df_top_wasted.csv



Info | Tables | Columns | Indexes | Triggers | Views | Stored Procedures | Functions | Grants | Events

Name	Engine	Version	Row Format	Rows	Avg Row Length	Data Length
12_3_eu	InnoDB	10	Dynamic	927	265	240.0 KiB
composition	InnoDB	10	Dynamic	6	2730	16.0 KiB
kilo_cap	InnoDB	10	Dynamic	54	303	16.0 KiB
moderate_percent	InnoDB	10	Dynamic	75	655	48.0 KiB
moderate_population	InnoDB	10	Dynamic	75	218	16.0 KiB
national_fruit	InnoDB	10	Dynamic	22	744	16.0 KiB
severe_percent	InnoDB	10	Dynamic	75	655	48.0 KiB
severe_population	InnoDB	10	Dynamic	75	218	16.0 KiB
top_wasted	InnoDB	10	Dynamic	20	819	16.0 KiB

Entities ERD



SQL Query

```sql
-- Query 1 |Top 10 countries with the highest average foodwaste in tonnes --
SELECT
country, seriesdescription AS target_name,
ROUND(AVG(value), 0) AS average_value
FROM 12_3_eu
GROUP BY country_id, country, seriesdescription
ORDER BY AVG(value) DESC
LIMIT 10;
```

country	target_name	average_value
Turkey	Food waste (Tonnes)	6315670
Germany	Food waste (Tonnes)	4447899
Italy	Food waste (Tonnes)	3449377
France	Food waste (Tonnes)	3262447
Spain	Food waste (Tonnes)	1803699
Poland	Food waste (Tonnes)	1512827
Romania	Food waste (Tonnes)	962189
Portugal	Food waste (Tonnes)	787754
Greece	Food waste (Tonnes)	732924
Switzerland	Food waste (Tonnes)	633276

```sql
-- Query 2 | Top 5 countries with highest aveage percentage and kilo per capita of
food insecurity  --
SELECT
    per.country, per.year,per.seriesdescription AS target_name,
    ROUND(SUM(per.value) / 3,2) AS average_percentage_value,
    ROUND(SUM(cap.value) / 3,2) AS average_capita_value
FROM
    moderate_percent AS per
INNER JOIN
    moderate_population AS cap ON per.id = cap.id
GROUP BY
    per.country_id, per.country, per.year, per.seriesdescription
LIMIT 5;
```

country	year	target_name	average_percentage_value	average_capita_value
Switzerland	2015	Prevalence of moderate or severe food insecurity in the adult population (%)	12.9	1078.9
Germany	2015	Prevalence of moderate or severe food insecurity in the adult population (%)	11.1	9053.7
Portugal	2015	Prevalence of moderate or severe food insecurity in the adult population (%)	43.8	4545
Ireland	2015	Prevalence of moderate or severe food insecurity in the adult population (%)	29.4	1385.6
Slovakia	2015	Prevalence of moderate or severe food insecurity in the adult population (%)	18.6	1005.9

```sql
-- Query 4 What is the most common national fruit in Europe and from which countries --
SELECT country,
    common_name
FROM
    national_fruit
WHERE
    common_name = (
        SELECT
        common_name as national_fruit
        FROM
            national_fruit
        GROUP BY
            common_name
        ORDER BY
            COUNT(*) DESC
        LIMIT 1);
```

country	national_fruit
Austria	Apple
Belgium	Apple
Bulgaria	Apple
Germany	Apple
Netherlands	Apple
Poland	Apple
Portugal	Apple
Romania	Apple
Sweden	Apple
Switzerland	Apple
United Kingdom	Apple

```sql
-- Query 5 Filtering only France from 4 United Nation tables--
SELECT id, indicator, seriesdescription, country, year, value, time_detail, source,footnote, units
FROM moderate_percent
WHERE country = 'France'
UNION ALL
SELECT
   *
FROM moderate_population
WHERE country = 'France'
UNION ALL
SELECT id, indicator, seriesdescription, country, year, value, time_detail, source, footnote, units
FROM severe_percent
WHERE country = 'France'
UNION ALL
SELECT *
FROM severe_population
WHERE country = 'France';
```

id	indicator	seriesdescription	country	year	value	time_detail
france2015	2.1.2	Prevalence of moderate or severe food insecurity in the adult population (%)	France	2015	5.7	2015
france2015	2.1.2	Prevalence of moderate or severe food insecurity in the adult population (%)	France	2015	8.1	2015
france2015	2.1.2	Prevalence of moderate or severe food insecurity in the adult population (%)	France	2015	6.9	2015
france2015	2.1.2	Adult population in moderate or severe food insecurity (thousands of people)	France	2015	5265.3	2015
france2015	2.1.2	Adult population in moderate or severe food insecurity (thousands of people)	France	2015	3673.1	2015
france2015	2.1.2	Adult population in moderate or severe food insecurity (thousands of people)	France	2015	4469.2	2015
france2015	2.1.2	Prevalence of severe food insecurity in the adult population (%)	France	2015	1.4	2015
france2015	2.1.2	Prevalence of severe food insecurity in the adult population (%)	France	2015	2.1	2015
france2015	2.1.2	Prevalence of severe food insecurity in the adult population (%)	France	2015	0.7	2015
france2015	2.1.2	Adult population in severe food insecurity (thousands of people)	France	2015	505.8	2015
france2015	2.1.2	Adult population in severe food insecurity (thousands of people)	France	2015	1358	2015
france2015	2.1.2	Adult population in severe food insecurity (thousands of people)	France	2015	931.9	2015

API

I created APIs with Swagger documentation to expose the collected data. Using Flask and Flasgger, I set up endpoints to access food waste information. The **/foodwastes** endpoint provided all food waste data, while **/foodwastes/<country_name>** allowed retrieval of data specific to a country. Additionally, I implemented an endpoint **/fruit/<country_name>** to retrieve the national fruit for a given country. My **top_wasted** endpoint returned a list of the top food wasted data. With this setup, users could easily access and interact with my food waste database, enabling informed decision-making and analysis.

Machine Learning

Recognizing Fruits and Vegetables: Leveraging advanced machine learning algorithms, the project aims to develop a sophisticated system capable of accurately recognizing and categorizing various types of fruits and vegetables commonly found in European households. By analyzing images and patterns, the system will efficiently identify produce items, enabling users to streamline their ingredient identification process while reducing the risk of misclassification or wastage.

Personalized Recipe Suggestions: An intuitive mobile application will be developed, allowing users to scan the fruits and vegetables available in their kitchen. Based on the scanned ingredients and user preferences stored in the app, the system will generate personalized recipe suggestions tailored to individual dietary requirements, cooking skill levels, and cultural backgrounds. These tailored recommendations not only promote sustainable cooking practices by encouraging the use of available ingredients but also enhance user engagement and satisfaction by offering relevant and appealing meal ideas.

Conclusion

"Waste Less, Taste More" is a comprehensive initiative addressing food waste, particularly focusing on household wastage within Europe, in line with UN sustainability objectives. Leveraging data from sources like Eurostat and Kaggle, the project meticulously analyzes food waste patterns, employing visualization techniques and API integration to make insights accessible. Through machine learning applications like image recognition for identifying national fruits and recipe generation, the project aims to catalyze sustainable consumption practices and reduce food waste, contributing to a more efficient and resilient food system.

The project's multifaceted approach encompasses data collection, cleaning, and analysis to uncover key insights into food waste trends across Europe. By providing stakeholders with actionable information and innovative solutions, including API access and machine learning-based interventions, "Waste Less, Taste More" seeks to empower individuals and organizations to make informed decisions and drive meaningful change towards a sustainable future.

GDPR

In compliance with GDPR regulations, "Waste Less, Taste More" ensures the protection of personal data throughout the project lifecycle. Any personal information collected, such as user data for API access or data used for machine learning training, is handled with the utmost confidentiality and security measures. Users are informed about the purpose of data collection, their rights regarding their data, and how their data will be used and stored. Additionally, robust data encryption and access controls are implemented to prevent unauthorized access or disclosure. Regular audits and reviews are conducted to ensure ongoing compliance with GDPR requirements and to address any potential data security risks effectively.

References

Trello: :
https://trello.com/invite/b/iBCQVz5W/ATTI38c64c788a6940d2be0bf86912bac023FCC875AC/foodwasteproject

Flat file:
https://ec.europa.eu/eurostat/databrowser/view/env_wasfw/default/table?lang=en
https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition
https://champions123.org/target-123

API
https://unstats.un.org/sdgapi/swagger/#/

Web Scraping
https://swnsdigital.com/uk/2023/02/these-are-the-top-20-unwanted-fruit-and-vegetables-from-bagged-salad-leaves-to-bananas/
https://en.wikipedia.org/wiki/List_of_national_fruits
https://www.sciencedirect.com/science/article/pii/S0921344921000331?ref=pdf_download&fr=RR-2&rr=878c54440e916f0c

GitHub Repository (in progress)
https://github.com/krantagat/food_waste_project