Environmental Impact of Food Production Documentation

---

1. Business Understanding

Objective

The goal of this project is to analyze and understand the **environmental impact of different food products** across the supply chain. The project seeks to:

- Compare **animal-based vs plant-based foods** in terms of emissions and resource use.
- Identify **which foods contribute most** to greenhouse gas (GHG) emissions, land use, water consumption, and eutrophication.
- Examine **supply chain stages** (Farm, Land Use Change, Animal Feed, Processing, Transport, Packaging, Retail) to identify the key drivers of environmental impact.
- Cluster food items by environmental footprint (high, medium, low impact).
- Provide a **scenario-based What-If analysis** (e.g., replacing X% of one food with another) to evaluate potential reductions in emissions.

Business Value

- Inform policymakers and researchers on the foods with the highest environmental burden.
- Provide insights for food companies, restaurants, and consumers seeking to reduce environmental footprints.
- Serve as a decision-support tool in sustainability planning and dietary guidance.

---

2. Data Understanding

Dataset Overview

The dataset contains metrics on food production and its environmental impact. Each row represents a food product with the following metrics:

- **Supply Chain Emissions**: Land Use Change, Animal Feed, Farm, Processing, Transport, Packaging, Retail, Total Emissions.

- **Emissions Normalized by Nutrition**:
    - Eutrophying emissions (per 1000 kcal, per kilogram, per 100g protein).
    - GHG emissions (per 1000 kcal, per 100g protein).
- **Resource Use**:
    - Freshwater withdrawals (per kcal, per 100g protein, per kg).
    - Land use (per kcal, per 100g protein, per kg).
    - Scarcity-weighted water use (per kcal, per 100g protein, per kg).

Food product	Land use change	Animal Feed	Farm	Processing	Transport	Packging	Retail	Total_emissions	Eutrophying emissions per 1000kcal (gPO
Wheat & Rye (Bread)	0.1	0	0.8	0.2	0.1	0.1	0.1	1.4	
Maize (Meal)	0.3	0	0.5	0.1	0.1	0.1	0	1.1	
Barley (Beer)	0	0	0.2	0.1	0	0.5	0.3	1.1	
Oatmeal	0	0	1.4	0	0.1	0.1	0	1.6	
Rice	0	0	3.6	0.1	0.1	0.1	0.1	4	
Potatoes	0	0	0.2	0	0.1	0	0	0.3	
Cassava	0.6	0	0.2	0	0.1	0	0	0.9	
Cane Sugar	1.2	0	0.5	0	0.8	0.1	0	2.6	
Beet Sugar	0	0	0.5	0.2	0.6	0.1	0	1.4	
Other Pulses	0	0	1.1	0	0.1	0.4	0	1.6	
Peas	0	0	0.7	0	0.1	0	0	0.8	
Nuts	-2.1	0	2.1	0	0.1	0.1	0	0.2	
Groundnuts	0.4	0	1.4	0.4	0.1	0.1	0	2.4	
Soymilk	0.2	0	0.1	0.2	0.1	0.1	0.3	1	
Tofu	1	0	0.5	0.8	0.2	0.2	0.3	3	
Soybean Oil	3.1	0	1.5	0.3	0.3	0.8	0	6	
Palm Oil	3.1	0	2.1	1.3	0.2	0.9	0	7.6	
Sunflower Oil	0.1	0	2.1	0.2	0.2	0.9	0	3.5	
Rapeseed Oil	0.2	0	2.3	0.2	0.2	0.8	0	3.7	

Search
- _measures
- food_production_impact
    - Σ Animal Feed
    - Σ Eutrophying emissions per 1.
    - Σ Eutrophying emissions per 1.
    - Σ Eutrophying emissions per ki
    - Σ Farm
    - Food Process
    - Food product
    - Food product (impact levels)
    - Food Type
    - Σ Freshwater withdrawals per 1
    - Σ Freshwater withdrawals per 1
    - Σ Freshwater withdrawals per k
    - Σ Greenhouse gas emissions p.
    - Σ Greenhouse gas emissions p.

Grouping Frameworks

Food Type	Source Type	Major Nutrient	Food Process
Cereals & Staples	Plant-based	Carb-rich	Processed
Cereals & Staples	Plant-based	Carb-rich	Unprocessed
Drinks	Plant-based	Mixed	Processed
Cereals & Staples	Plant-based	Carb-rich	Unprocessed
Cereals & Staples	Plant-based	Carb-rich	Unprocessed
Cereals & Staples	Plant-based	Carb-rich	Unprocessed
Cereals & Staples	Plant-based	Carb-rich	Unprocessed
Sugars	Plant-based	Carb-rich	Unprocessed
Sugars	Plant-based	Carb-rich	Unprocessed
Pulses & Legumes	Plant-based	Protein-rich	Unprocessed

1. **By Food Type / Category**
    - Cereals & Staples, Sugars, Pulses, Nuts, Oils, Vegetables, Fruits, Drinks, Confectionery, Animal Products.
2. **Animal-based vs Plant-based**
    - Animal-based: Meat, Dairy, Eggs, Fish, Shrimps.
    - Plant-based: All others.
3. **Nutrient-Dominant Groups**

○ Protein-rich, Carb-rich, Fat-rich, Mixed-nutrient.

---

3. Data Preparation

Cleaning & Transformation Steps

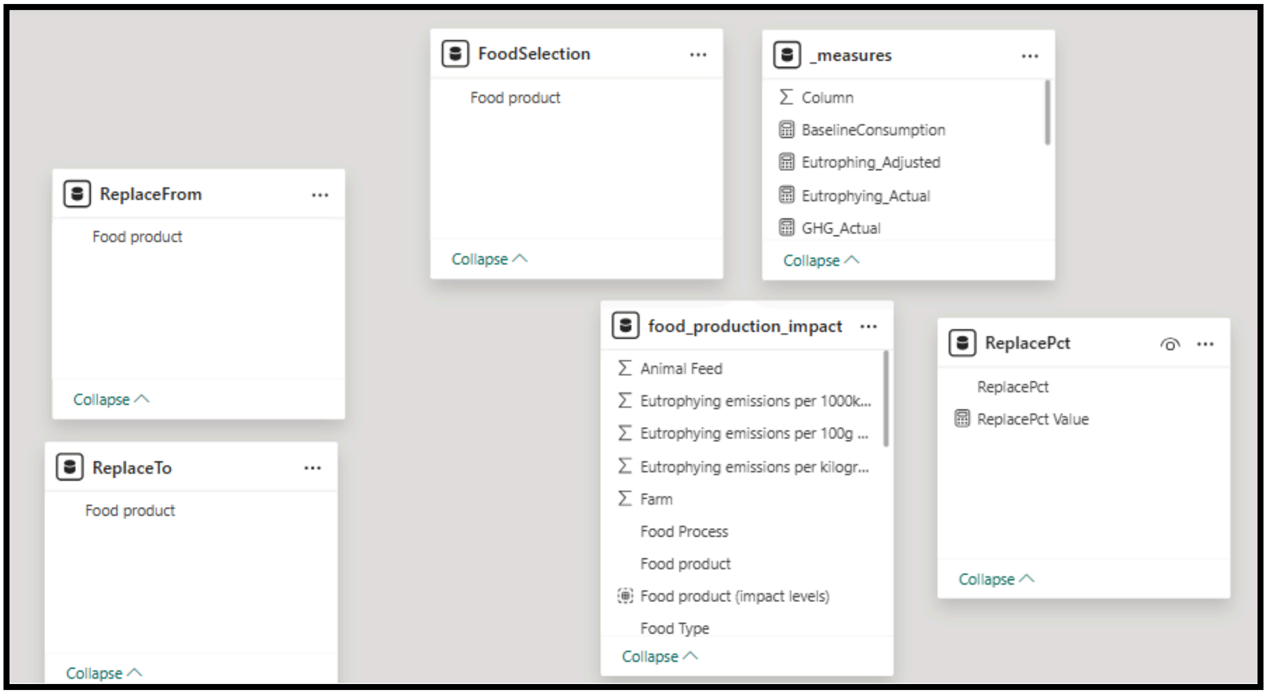
- **Protein-related Nulls**: Inputted `0` for foods with no protein when metrics were protein-normalized.
- **External Nutritional Values**: Sourced from *foodstruct.com* and *eatthismuch.com* (maize meal) to fill selected missing values.
- **Remaining Nulls**: Left as-is to avoid methodological inconsistencies.
- **Grouping**: Organized food items into meaningful categories for comparison.
- **Tools Used**:
  - Python (Jupyter Notebook) → Data cleaning and preprocessing.
  - Power BI → Visualization and What-If analysis setup.
- **Measures & Parameters**:
  - Created a dedicated measure table.
  - Designed parameters (e.g., %Replacement) for scenario analysis.

```
In [11]: #Cane Sugar, Beet Sugar, Soybean Oil, Palm Oil, Sunflower Oil, Rapeseed Oil, Olive Oil do not contain protein so I would input 0
         dup_data.loc[dup_data["Food product"] == "Cane Sugar", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"] =
         dup_data.loc[dup_data["Food product"] == "Beet Sugar", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"] =

         dup_data.loc[dup_data["Food product"] == "Soybean Oil", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"]
         dup_data.loc[dup_data["Food product"] == "Palm Oil", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"] = 0
         dup_data.loc[dup_data["Food product"] == "Sunflower Oil", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"
         dup_data.loc[dup_data["Food product"] == "Rapeseed Oil", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"]
         dup_data.loc[dup_data["Food product"] == "Olive Oil", "Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"] =

In [12]: #confirming inputted values
         dup_data.loc[dup_data["Eutrophying emissions per 100g protein (gPO₄eq per 100 grams protein)"].isnull(), "Food product"]

Out[12]: 0     Wheat & Rye (Bread)
         1            Maize (Meal)
         2          Barley (Beer)
         13              Soymilk
         14                 Tofu
         24     Other Vegetables
         29                 Wine
         30          Other Fruit
         42      Shrimps (farmed)
         Name: Food product, dtype: object
```



---

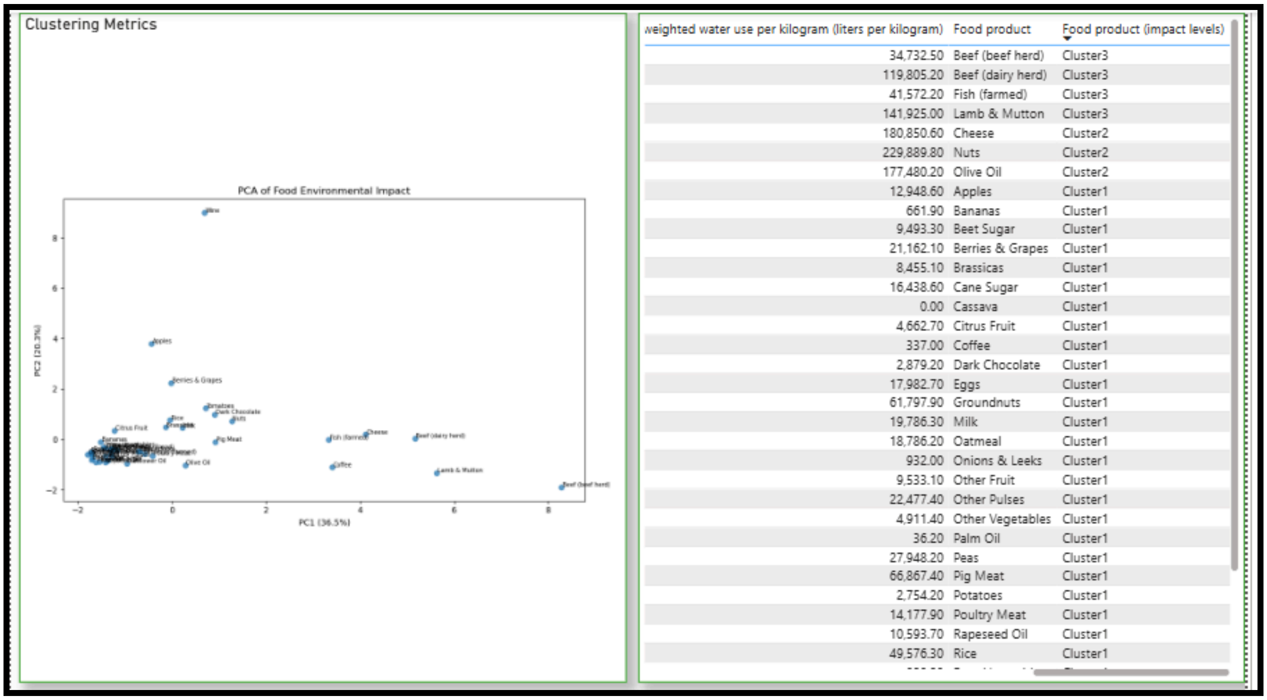
4. Modeling

Analysis Approaches

- **Comparative Analysis**: Animal vs Plant-based foods, processed vs unprocessed.
- **Supply Chain Stage Analysis**: Identified which stages drive most environmental impact.
- **Clustering**: Grouped foods into high, medium, and low environmental impact.
- **Scenario Modeling**: What-If analysis to simulate replacements of food consumption.

Clustering Results

- **High Impact**: Beef, Farmed Fish, Lamb & Mutton.
- **Medium Impact**: Cheese, Nuts, Olive Oil.
- **Low Impact**: All other foods.

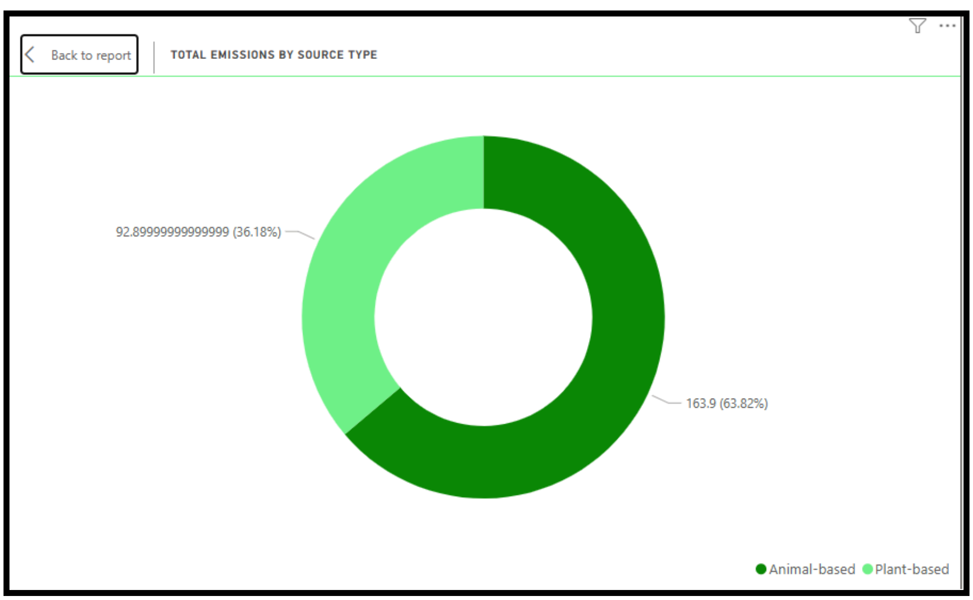


weighted water use per kilogram (liters per kilogram)	Food product	Food product (impact levels)
34,732.50	Beef (beef herd)	Cluster3
119,805.20	Beef (dairy herd)	Cluster3
41,572.20	Fish (farmed)	Cluster3
141,925.00	Lamb & Mutton	Cluster3
180,850.60	Cheese	Cluster2
229,889.80	Nuts	Cluster2
177,480.20	Olive Oil	Cluster2
12,948.60	Apples	Cluster1
661.90	Bananas	Cluster1
9,493.30	Beet Sugar	Cluster1
21,162.10	Berries & Grapes	Cluster1
8,455.10	Brassicas	Cluster1
16,438.60	Cane Sugar	Cluster1
0.00	Cassava	Cluster1
4,662.70	Citrus Fruit	Cluster1
337.00	Coffee	Cluster1
2,879.20	Dark Chocolate	Cluster1
17,982.70	Eggs	Cluster1
61,797.90	Groundnuts	Cluster1
19,786.30	Milk	Cluster1
18,786.20	Oatmeal	Cluster1
932.00	Onions & Leeks	Cluster1
9,533.10	Other Fruit	Cluster1
22,477.40	Other Pulses	Cluster1
4,911.40	Other Vegetables	Cluster1
36.20	Palm Oil	Cluster1
27,948.20	Peas	Cluster1
66,867.40	Pig Meat	Cluster1
2,754.20	Potatoes	Cluster1
14,177.90	Poultry Meat	Cluster1
10,593.70	Rapeseed Oil	Cluster1
49,576.30	Rice	Cluster1

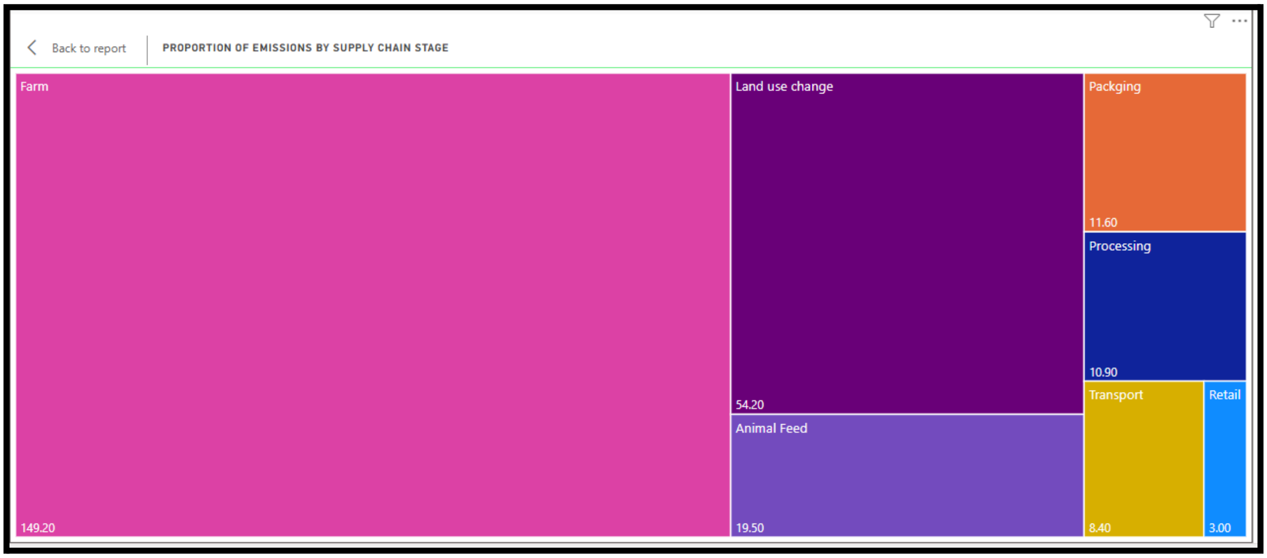
5. Evaluation

Key Insights

- **Animal-based foods**: Higher emissions compared to plant-based.



- **Beef (beef herd)**: Highest single contributor to environmental impact.
- **Unprocessed foods**: Surprisingly, they generate more emissions than processed foods.
- **Agriculture (farm, land use, feed)**: Primary driver of environmental impact.



- **Food-specific insights**:
  - Oils and Drinks → High packaging emissions.
  - Sugars → Highest transport emissions.
  - Coffee → Highest packaging emissions.
  - Nuts, Olive Oil, Wine, Citrus Fruits → Regenerative land effects.

---

6. Deployment

Tools & Outputs

- **Power BI Dashboard**: Interactive visuals for emissions, clustering, and What-If analysis.
- **What-If Analysis**: Slider parameter enabling scenario testing (e.g., replacing Beef with Pulses).

Recommendations

- Replace high impact foods with low impact foods of the similar nutrotional value.
- Increase production of land replenishing food products.
- Encourage diets that minimize or eliminate meat, especially beef.

Next Steps

- Expand sourcing of missing nutritional data with verified consistent methodology.
- Extend What-If analysis to multiple replacements simultaneously.
- Explore **time-based trends** (e.g., impact changes over years).