Programming Studio 1
COSC2803 | Semester 1 2024
Food Loss and Waste
Requirements Document

1 Overview

This document outlines the Sub-Tasks for the "Food Loss and Waste" social challenge. You can think of this document as the **product idea** (from a UX/UI perspective) and the **business case** (from the database perspective).

1.1 Challenge: Investigating Food Loss and Waste

Food loss and waste is defined as food products that end up not being eaten for a variety of reasons. Food loss occurs along the food supply chain from harvest/catch/slaughter up to, but not including, the sales level. Food waste occurs at the retail and consumption level. Reasons for loss and waste can vary during each stage. For example, consumption by pests, and spoilage/decay due to improper temperature control, can both occur during the storage stage. Food lose and waste both have a significant impact on many other current global issues, including climate change, food security (availability), resource usage (land and water) as well an impact on biodiversity among others.

Your challenge is to develop a web-application to **help various key parties in the supply chain (including consumers and policy makers) explore _unbiased_ information on the sources of food loss and waste over an extended time period.** You will need to present statistics and calculated information about the types of food loss/waste. You will need to:

- Present this information in an informative, respectful, and unbiased manner.
- Cater for a diverse range of users who are seeking to become more informed on this topic.
- Provide diverse types of information and functionality, including both high-level summaries, and enabling an in-depth analysis of the data.

By using your website, users should be able to become well-informed of the levels of food waste/loss across the world.

1.2   Notes on the provided Datasets

You have been provided datasets that relate to historic (yearly) food loss and waste statistics for various food products across different geographic locations. In addition, you are provided a dataset of various codes, and descriptions for different food products and groups.

The food loss and waste statistics are from **1966 – 2022** at a **country** and **region** level of granularity.

Not all countries have regional data, and not all countries have data for each year in the range above.

Note that several fields have been removed and some values aggregated (averaged) to simplify the data.

2   Project Level Details (All Groups)

'A' & 'B' Sub-Tasks for each level are outlined in this section. **All groups** must use these Sub-Tasks.

2.1   **Level 1 (GREEN) "Big Picture" Content**

Level 1 sub-tasks capture the attention of the users and show how to use your website. It is up to you to decide the best way to show the users the required information, that is, complete the UX/UI design.

Sub-Task A: Landing Page

This is the first page users see when they visit your website. This should:

- Capture the attention of all users of your website.
- Highlight the topics targeted by your website.
- Present a snapshot of the data covered by your website, specifically:
  - The year range (first and last year) for available data
  - The maximum single year loss percentage of any commodity
  - Identify the commodity (or commodities) that correspond with the maximum loss identified above
- These values should be calculated from your database (in Milestones 2-4), thus you should not store these values in your database or ER Model.

Sub-Task B: Mission Statement

This presents the overall purpose of your website. This should:

- Present your perspective on how your website addresses the social challenge.
- Describe how your site can be used.
- Present the Personas that your website targets. **These must be stored in and retrieved from your database.**
- List the names and student numbers of all team members. **These must be stored in and retrieved from your database.**

2.2   **Level 2 (ORANGE) "Shallow Glance" of the data**

Level 2 sub-tasks explore, at a shallow level, the issue of food loss and waste for various geographic locations, food types/categories and stages. You should *carefully consider* your UX/UI designs and the ER model. The sub-tasks include UX/UI design and ERM design challenges.

Sub-Task A: Focused view of loss/waste change by Country

Present information about food loss and waste changes for **user selected** Countries. You should:

- Show each selected countries name.
- Enable the user to choose a start year and end year.
- Display the change in food loss/waste between the start year and end year (where available) for each Country. This must be computed from the database using SQL queries. If data is not available for some years, your website should handle this is a suitable manner.
- Allow the user to filter which fields are / aren't displayed (commodity, activity, food supply stage, cause of loss). The result should not display duplicate rows (loss/waste values must be appropriately aggregated i.e. averaged to handle the filtered fields).
- Display the changes between the start year and end year using raw values, ie the loss/waste percentage values of the country on the start and end years.
- Give the user the option to list data for all years that fall into the year range (rather than just data for the start and end year).
- Provide a way for the user to sort by the result by the loss/waste value (during the start year or during the end year), in both ascending and descending order. In Milestones 2-4 this sorting should be done using SQL queries. That is, we require you to show knowledge of SQL sorting methods and cannot sort using Java or JavaScript

Sub-Task B: Focused view of loss/waste change by Food Group

Present information about food loss and waste changes for **user selected** Food Groups. You should:

- Show each selected groups name (eg Cereals, Live animals etc).
- Enable the user to choose a start year and end year.
- Display the change in food loss/waste between the start year and end year (where available) for each group. This must be computed from the database using SQL queries. If data is not available for some years, your website should handle this is a suitable manner.
- Allow the user to filter which fields are / aren't displayed (activity, food supply stage, cause of loss). The result should not display duplicate rows (loss/waste values must be appropriately aggregated i.e. averaged to handle the filtered fields).
- Display the changes between the start year and end year using proportional values, ie the difference in percentage values of the Food Group between the start and end years.
- Provide a way for the user to sort by the result by the loss/waste value (the difference), in both ascending and descending order. In Milestones 2-4 this sorting should be done using SQL queries. That is, we require you to show knowledge of SQL sorting methods and cannot sort using Java or JavaScript

2.3   Level 3 (RED) "Deep-dive" of the data

Level 3 sub-tasks explore a deeper view of the issue of food loss and waste for various geographic locations, food types/categories and stages. You should *carefully consider* your UX/UI designs and the ER model. These sub-tasks include complex UX/UI design and ERM design challenges. For top-marks you will need to take a *very well-considered approach*.

Sub-Task A: Identify locations with similar food loss/waste percentages

Enable a user to find countries that have similar rates of food loss/waste. You should:

- Enable the user to select a year and a geographic location (country, or region), then find other countries (or regions) that are "the most similar" to the chosen combination on a set of user selected properties, such that users may:
    - Select similarity in terms of the foods products they have in common
    - Select similarity in terms of the overall percentage of food loss/waste
    - Select similarity in terms of both common foods products and their loss/waste percentage (for country only)
    - The user can select to determine similarity in terms of:
        - the absolute values - ignoring food products that are not common between locations
        - the level of overlap - food products that are not common to the selected country impact (reduce) the similarity score
- From the user selection, your system must find the <x> most similar results, identifying the region, the year, the values used to determine its similarity to the selected region, as well as the similarity score, using the user's selection of constraints, where the user chooses the number of regions <x> to locate:
    - For example, the user selects the year 2013, the region 'Australia'. The user selects similarity by the food products alone based on the absolute value, and to find the 5 most similar regions. Since the user has selected a country (Australia), the system should then find other countries that have similar food product, then display the 5 most similar results.
- Present the selected region and most similar regions, along with the relevant data that was used to determine the similarity.
- Sort the regions that are found by the 'most similar' to the 'least similar'.
- Retrieve the *exact* information from your database using as few queries as possible:
    - Ideally, use only a single query (with column selections and JOIN operations)
    - Unnecessary information should not be returned in the SQL queries.
    - Performing the sorting through SQL queries if possible.
    - Note that this is challenge in making suitable SQL queries.

Sub-Task B: Exploring Food Commodities and Groups

Enable a user to explore food commodities and groups to find foods that have similar rates of food loss/waste. You should:

- Enable the user to select a food commodity, determine its group and then identify the food groups that are "the most similar" to the chosen group on a set of user selected properties, such that users may:
    - Select similarity in terms of the ratio of food loss to food waste (percentage average)
    - Select similarity in terms of the highest percentage of food loss/waste product
    - Select similarity in terms of the lowest percentage of food loss/waste product
- From the user selection, your system must find the <x> most similar groups, identifying the group, and the values used to determine its similarity to the selected group, as well as the similarity score, using the user's selection of constraints, where the user chooses the number of groups <x> to locate:

- o For example, the user selects the food commodity 'Tomatoes, similarity by the highest percentage of food loss/waste and the 5 most similar groups. Tomatoes belong to the group 012 "Vegetables". The system determines the highest percentage for any product in the Vegetable Group. Then determines what the highest percentage is for each of the groups. It then returns the 5 closest groups to the Vegetable group.
- Present the selected commodity, its group (including the name), the closest groups (including name) along with the relevant data that was used to determine the similarity.
- Sort the groups that are found by the 'most similar' to the 'least similar'.
- Retrieve the *exact* information from your database using as few queries as possible:
  - o Ideally, use only a single query (with column selections and JOIN operations)
  - o Unnecessary information should not be returned in the SQL queries.
  - o Performing the sorting through SQL queries if possible.
  - o Note that this is challenge in making suitable SQL queries.

3   Project Level Details (Groups of 3 ONLY)

'C' Sub-Tasks for each level are outlined in this section. **ONLY groups of 3** may use these Sub-Tasks.

Level 1: Sub-Task C: Summary of the Dataset

Provide an overview of the dataset. For each dataset in the database, identify the name of the collection, its attributes and a summary of the purpose of each field (note that this information is available in the Description.txt file).

Present an informative summary of every attribute across all Datasets. This should:

- Provide an informative summary of each Dataset used to power your website.
- List all attributes using a "natural language" label. **These labels must be stored in and retrieved from your database.**
- Provide a "natural language" description of each label. **These descriptions must be stored in and retrieved from your database.**

Level 2: Sub-Task C: Focused view of loss/waste change by Stage

Present information about food loss and waste changes for **user selected** Food Supply Stages.

You should:

- Show each selected Food Supply Stage name (eg Harvest, Storage, Transport etc).
- Enable the user to choose a start year and end year.
- Display the average loss percentage of food loss/waste <u>for all years</u> between the start year and end year (where available) for each selected stage. This must be computed from the database using SQL queries. If data is not available for some years, your website should handle this is a suitable manner.
- Allow the user to optionally group by the activity for each stage as well.
- Allow the user to also display the percentage value of loss vs waste separately. Recall that waste occurs at the Retail and Consumption (Households) stages.
- Provide a way for the user to sort by the result by the loss/waste value (for each Stage/Activity), in both ascending and descending order. In Milestones 2-4 this sorting should be done using SQL

queries. That is, we require you to show knowledge of SQL sorting methods and cannot sort using Java or JavaScript

Level 3: Sub-Task C: Exploring stages and causes

Enable a user to investigate each stage to determine the major causes of loss/waste. You should:

- Enable the user to select a stage, and a cutoff value <x>.
- For the chosen stage identify the average percentage impact of each of the major causes of loss.
- Display the top x causes of loss for the stage, including its percentage impact, ranked from highest to lowest.
- For each identified cause in the chosen stage, identify the commodity that is most adversely affected by the cause of loss, the activity that was being performed when the loss occurred, and the country with the most number of years where the loss has occurred.
- Retrieve the *exact* information from your database:
    - Use as few queries as necessary (with column selections and JOIN operations), without retrieving unnecessary information.
    - Perform the sorting through SQL queries if possible.

---

4   Database Business Case Summary

This section provides a summary of some of the Business Case for your Database. You should consult with the project level details (and the Dataset descriptions) for full details.

- Personas are identified by a name and a picture. A Persona description contains multiple sub-elements, including (but not limited to) a list of Attributes, a Background, a list of Needs, a list of Goals, a list of Skills & Experience. Each list for each Persona can be of variable length.
- Students (that is the developers of the website) are identified by a student number. A student may be described by a name, along with other information.
- Each country has a unique m49code.
- For some countries no regions are recorded, while other have several regions, and some have entries for both the country in general as well as for specific regions.
- The CPC code is made up of a group, class and subclass which can have prefix of '0' and decimal points – commodities in the same group/class share properties
    - For example, all commodities in group 012 are vegetables.
    - Some of the classes within this group include 0121 Leafy or stem vegetables
    - Within this we have 01211 Asparagus, 01212 Cabbages etc
- The dataset contains a lot of missing values, so your design will need to take this into consideration.
- The cause of loss contains a lot of mixed data – sentences, references, combined stats etc.