**Assignment (GitHub Exercise) #1 – Data Collection: Kitchen Ecosystem**

**Course:** Ashoka Horizons: Applied Data Science with ML and AI  
**Instructors:** Rintu Kutum, Gautam Ahuja  
**Student:** Advay  
**Date:** 30th May 2025

**Overview of Collected Data**

To comprehensively capture the activity in the kitchen, I created three datasets:

1. **dish\_log.csv** – a record of dishes cooked and their key features.
2. **ingredients\_log.csv** – a catalogue of ingredients, their characteristics, and storage.
3. **nutrition\_log.csv** – nutritional values per 100g for commonly used ingredients.

Each dataset focuses on a different aspect of kitchen data, allowing clearer organization and potential for later analysis.

The dish log serves as the main data list tracking the activity in the kitchen (aka cooking), which applies the information collected in the ingredients list (the information of the individual units of kitchen). The ingredients list records the important information of the units of cooking (the main activity in kitchens) in accordance to its location, shelf life, type of nutrition, and unit of measurement. The nutrition list acts as an aid to the ingredients list to individually record the effect of the ingredients which go into cooking the dishes.

**Why I Chose These Fields**

**a. dish\_log.csv**

Captures real-time cooking events. I included:

* **Date** – to analyze cooking patterns over time.
* **Dish Name** – identifies what was cooked.
* **Ingredients Used & Quantity** – central to recipe-level data.
* **Cooking Time** – useful for time-optimization analysis.
* **Appliance Used** – may reveal energy/resource usage patterns.

**b. ingredients\_log.csv**

Helps track inventory and shelf life. Fields included:

* **Ingredient** and **Type** – for categorization.
* **Unit** – for consistency across dishes.
* **Shelf Life** – useful for planning and waste reduction.
* **Storage Location** – relates to accessibility and freshness.

**c. nutrition\_log.csv**

Supports health-conscious analysis. For each food item, I recorded:

* **Calories, Protein, Fat, Carbohydrates** per 100g.  
  These values were sourced from online nutritional databases, including Kaggle.

**Connection**

The dish log serves to show the relationship between combination of the individual units in the kitchen, the ingredients, appliances, etc. The ingredients list contains the inventory of the building blocks of dishes, which serves as a deciding factor on what to cook, based on its freshness, shelf life, and with the help of the nutrition log which elucidates upon the nutritional value of each ingredient. This can help create a meal/dish suited to their likings, and helps analyse the allocation of resources in the kitchen.

**Why I Used Multiple Files**

Splitting data into multiple files improves clarity, modularity, and scalability.

* The dish log can grow daily with new entries.
* The nutrition log may not change often rather can be reused.
* The ingredient log helps bridge the two datasets and enables applications like inventory management.

**Software Setup Process**

As per the course requirements, I completed the following installations and configurations:

**a. Miniconda**

* Downloaded and installed Miniconda from the official website.
* Verified installation via conda --version.

**b. Conda Environment**

* Created a virtual environment using:
* conda create -n horizons25
* conda activate horizons25
* This environment ensures package consistency for future Python-based analysis.

**Conclusion**

This assignment helped me understand the practical aspects of data sourcing, formatting, and management. I realized that even everyday environments like a kitchen hold valuable data that can be structured and analysed in meaningful ways.