Horizons-25-A1-GE

Write-up: Kitchen Ingredients Data Collection and Environment Setup

Name - Tushar Joshi | Week - 1 | Assignment 1

Choice of Data Fields

For this assignment, I chose to focus on the following fields to capture comprehensive and meaningful data about the cooking process and nutritional aspects of the dishes:

- **Ingredients:** Listing all ingredients used gives a detailed insight into the components of each dish.
- **Quantity:** Quantifying ingredients is important for standardizing recipes and understanding portion sizes.
- **Spices Used and Quantity:** Spices significantly affect flavor and nutrition; recording their quantity helps analyze their impact.
- **Water Consumption (Purpose) and Volume of Water Consumed:** Water usage is critical both for cooking and sustainability analysis.
- **Cooking Time:** This reflects preparation effort and affects nutrient retention.
- **Appliances Used:** Different appliances can influence cooking efficiency and energy consumption.
- **Utensils Used:** Utensils affect the cooking method and sometimes the taste or texture.
- **Calorie, Carbohydrate, Protein, and Fat Contents:** Nutritional content is essential for assessing the health impact of the dishes.

These fields were selected to provide a holistic dataset that covers ingredient specifics, cooking logistics, and nutritional information, enabling multiple angles of analysis such as health, sustainability, and efficiency.

Why Separate Files Were Created

I decided to organize the data into separate files to maintain clarity, modularity, and ease of access. Each file can focus on a particular aspect — for example, one for ingredients and their

quantities, another for water consumption details, and another for nutritional content. This separation helps in:

- Avoiding overly large and complex datasets that are difficult to manage.
- Simplifying data processing, as one can load and analyze relevant parts without the overhead of unrelated data.
- Enhancing collaboration and version control by isolating different data aspects.
- Facilitating future updates and additions without affecting unrelated data sections.

Process and Tools Used

 # Installing Windows Subsystem for Linux (WSL)

To create a versatile and powerful working environment, I installed the Windows Subsystem for Linux (WSL) on my Windows machine. WSL allows running a Linux environment directly on Windows without a virtual machine, providing access to Linux command-line tools and software.

Steps taken:

1. **Opened PowerShell as Administrator.**
2. **Ran the command:**
   ```
   wsl --install
   ```
3. **After installation, I restarted the system.**
4. **Launched the Ubuntu terminal from the Start menu.**
5. **Updated packages using:**
   ```
   sudo apt update && sudo apt upgrade
   ```

WSL provided a smooth, native-like Linux environment for my data processing needs.

 # Installing Miniconda and Creating a Conda Environment

To manage Python packages and dependencies efficiently, I installed Miniconda inside the WSL Ubuntu environment.

Steps:

1. **Downloaded the latest Miniconda installer for Linux using**

   `wget`.

2. **Ran the installer script:**
   `bash Miniconda3-latest-Linux-x86_64.sh`
3. **Followed the prompts to complete the installation.**
4. **Restarted the terminal or sourced the** `.bashrc` **to update paths.**
5. **Created a dedicated conda environment for the project to keep dependencies isolated:**
   `conda create -n horizon25 python=3.10`
6. **Activated the environment:**
   `conda activate horizon25`
7. **Installed necessary packages such as** `numpy`, `pandas`, **and others as needed using:**
   `conda install numpy pandas matplotlib scikit-learn seaborn ipykernel`

Using a conda environment ensured that the required packages were installed without conflicts and that the project environment could be replicated easily.