Data Engineering Intern Assignment

This Assignment comprises two tasks. Your submission should include a link to a GitHub repository; the repository should be organized into two subfolders, each corresponding to one of the tasks. Each subfolder must contain a README.md file that briefly explains the solution and provides the necessary documentation for fellow data engineers to work with the solution.

Task 1: Data Scraping and Consolidation

This Task involves two sections. The first requires building a scraper, and the second requires consolidating datasets into one.

1. Scrape the MacroNutrient(Table View) and MicroNutrient(Table View) data at district level for all states for 2023-24 from https://soilhealth.dac.gov.in/piechart and save them in a folder. To accomplish this, create a script named **get_raw_data.py** that saves the downloaded files to a data directory. Organize the files within the data directory into subfolders for each state. Within each state's subfolder, create additional subfolders for the reports, each containing files for individual tables.
2. Create a script titled **consolidate_table.py** that consolidates Micro and Macro Nutrient data into a single CSV file.

Task 2: PDF Parsing and Consolidation

The provided PDF file contains data from the India Tourism Statistics. Your task is to create a script named **process_foriegn_tourism.py** that parses the PDF file, extracts the tables, and consolidates it into a CSV file, giving a complete view of Foreign Tourist Arrivals. The following table needs to be parsed and consolidated-

1. Table 2.3.4 Nationality-wise and Quarter-wise Distribution of FTAs in India, 2021
2. Table 2.6.2 Nationality-wise Gender wise Distribution of FTAs in India, 2021
3. Table 2.7.2 Distribution of Nationality-wise FTAs in India according to Age- Group, 2021
4. Table 2.8.1 Distribution of Nationality-wise FTAS in India according to Purpose 2021
5. Table 2.9.2 Nationality-wise Percentage Distribution of FTAs in India According to Average Duration of Stay Groups, 2021

Please note the following guidelines:

1. Utilize popular Python tools from the SciPy ecosystem, such as pandas, polars, scrapy, bs4, selenium, etc. Document any less common tools used in the README file for the respective task.
2. Ensure your code is well-structured, follows best practices, and includes comments for clarity and maintainability.

Your submission should demonstrate your ability to work with data engineering tasks, including data consolidation, transformation, and building scrapers.