Concurrent Weather Data Extraction Tool

Overview

This script enables efficient and concurrent extraction of weather data for multiple cities by leveraging multithreading. The tool reads city data from a CSV file, processes each city's weather information using the `WeatherDataExtractor` class, and saves the results in a structured format.

Key Features

- **Concurrent Processing:** Uses multithreading to process multiple cities in parallel.
- **Dynamic Directory Management:** Automatically creates dated directories to store extracted data.
- **Logging:** Provides detailed logs for tracking progress and troubleshooting errors.
- **Customizable:** Adjusts the number of threads and city selections dynamically based on system capabilities and input data.

Approach

1. **City Selection:** Filters cities based on timezone differences to target specific time slots for data extraction.
2. **Task Queue:** Uses a queue to manage city processing tasks and distribute them across worker threads.
3. **Multithreading:** Spawns multiple worker threads, each responsible for processing a portion of the city list.
4. **Weather Data Extraction:** Employs the `WeatherDataExtractor` class to scrape and save weather data for each city.
5. **Graceful Shutdown:** Ensures all threads complete or terminate safely after processing.

Add - Ons

- `recurrent.py` can be used to rescrape data for the datapoints which have already been scraped in case that some data is added later on. Although the script has been used at the start of the project, it might have become obsolete with the new time zone approaches we took.