How to Prepare Data for Data Analysis

Follow the steps below to prepare the data and run the analysis using Python:

Step 1: Download and Install Required Tools

1. **Download the following files:**
    - **CzechFOI-DA** (project folder)
    - **DB Browser for SQLite** (from the official website)
    - **Vesely_106_202403141131.tar.xz** (CSV data)
2. **Install DB Browser for SQLite:**
    - Extract the contents of DB Browser for SQLite to C:\github\CzechFOI-DA\DB.Browser.for.SQLite.
    - Run DB Browser for SQLite.exe from this folder.
3. **Extract Vesely Data:**
    - Extract Vesely_106_202403141131.tar.xz to a CSV file: Vesely_106_202403141131.csv.

Step 2: Create and Set Up the Database

1. **Create a New SQLite Database:**
    - Open DB Browser for SQLite.
    - Go to **Menu: File > New Database**.
    - Name the database czechFO and save it to C:\github\CzechFOI-DA\DB\czechFOI.db.
2. **Connect to the Database:**
    - Open the newly created database by going to **Menu: File > Open Database** and select czechFOI.db.
3. **Import CSV Data into Database:**
    - Go to **Menu: File > Import > Table from CSV File**.
    - Select the Vesely_106_202403141131.csv file and name the imported table czech.
4. **Verify Imported Table:**
    - Go to the **Database Structure tab**.
    - Right-click on the czech table and select **Browse Table**.
    - Verify that the table has 11,028,372 rows.
5. **Save the Database:**
    - Go to **Menu: File > Write Changes** and ignore any errors that appear.
    - Close DB Browser for SQLite and save changes when prompted.

Step 3: Execute SQL Queries for Data Preparation

1. **Open the Database:**
    - Reopen the database: **Menu: File > Open Database** and select czechFOI.db.
2. **Run SQL Queries:**
    - Select the **Execute SQL tab**.
    - Click the yellow **Open SQL Query** icon and open C:\github\CzechFOI-DA\SQLQueries\All SQL Time.sql.
    - Click the arrow to **Execute all/selected SQL**.
    - Wait for the queries to complete (approximately 30 minutes).
3. **Pivot Tables and Views:**
    - After execution, the database will have 22 pivot views.
    - Export these views into CSV files, which will be used for further analysis in Python.

Step 4: Export Data Views as CSV

1. **Export Views:**
    - In **DB Browser for SQLite**, go to the **Database Structure** tab.
    - Expand the **Views** section to see 22 views.
    - Right-click each view and select **Export as CSV File**.
    - Use the following export settings:
        - **Column Names in First Row**: Yes
        - **Field Separator**: Comma (,)
        - **Quote Character**: Double Quote (")
        - **Newline Character**: Windows CR + LF (\r\n).
2. **Export All Views at Once:**
    - In the export UI, you can select all 22 views at once.
    - Choose the **TERRA Folder** to export the files to.
3. **Verify Export:**
    - After exporting, confirm that the 22 CSV files are in the **TERRA Folder**.
4. **Close DB Browser for SQLite** when done, ensuring changes are saved.

Step 5: Run Data Analysis Using Python

1. **Download and Install Required Software:**
    - Download and install **Microsoft Visual Studio Code** and **Python**.
    - Open **Visual Studio Code**.
2. **Open Python Script in VS Code:**

- Open the Python script AH) 2D 6-Axis age-compare rolling-mean significance-1D-2D same-scale.py located at C:\github\CzechFOI-DA\Py Scripts\.

3. **Install Required Python Libraries:**
    - Open a terminal in Visual Studio Code (View > Terminal).
    - Run the following pip install commands to install the necessary Python libraries:

```
bash
Code kopieren
pip install pandas
pip install plotly
pip install numpy
pip install xarray
pip install scipy
pip install matplotlib
```

4. **Run the Python Script:**
    - With the script open in Visual Studio Code, click on the **Run Python File** button (the green triangle at the top-right).
    - If prompted, trust the script by clicking **Trust**.
5. **Check for Output:**
    - The script will generate interactive HTML plots in the folder C:\github\Plot Results\.
    - The plots will contain calculated data for further analysis.
6. **Modify Calculations and Plots:**
    - If needed, you can modify the calculation or plotting settings in the plot_config section of the script.

Step 6: Customize Your Analysis

- You can customize the Python script to calculate derivatives, rolling correlations, or any other analysis you need by adjusting the code.
- If you're unsure about how to make changes, you can always ask **ChatGPT** for help with Python code explanations or modifications.

Conclusion:

By following these steps, you'll prepare the data, import it into a database, run SQL queries to create necessary views, export those views to CSV files, and finally run the Python script to perform data analysis. You can customize the script for your own analysis needs as well!