



Global Ecological Footprint Data Analysis

Summary of Data Extraction, Cleaning, Storage, and
Visualization

BY DAMILARE OMOBORIOWO AND CATHERINE MATTHEWS

Summary of Data Extraction, Cleaning, Storage, and Visualization

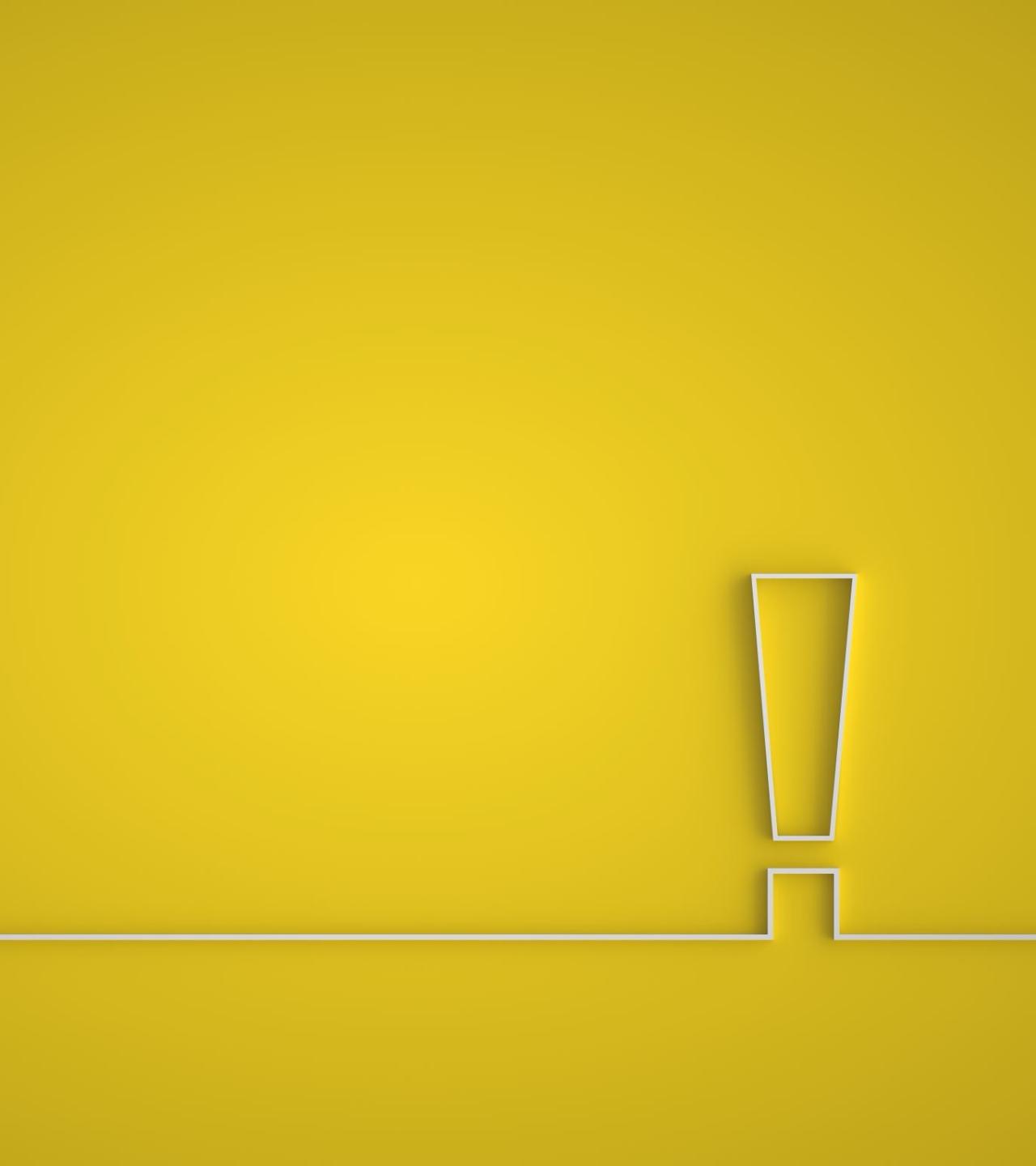
- ▶ Introduction:
- ▶ This presentation provides an overview of the process of extracting, cleaning, storing, and visualizing global ecological footprint data. The goal is to make the data easily accessible and presentable for analysis.



► Data Loading:

- We utilized a CSV file to load the Global Ecological Footprint data into a pandas Data Frame for further processing.





► Data Cleaning:

- Sanitized column names to ensure they are valid Python identifiers using regular expressions.
- Converted non-numeric values to numeric types and handled missing values appropriately.
- Renamed columns to correct any typographical errors

► **Database Storage:**

- Utilized SQL Alchemy to create a dynamic database model based on the structure of the Data Frame.
- Stored the cleaned data in an SQLite database to ensure persistence and ease of access.
- Defined a SQL Alchemy model dynamically to accommodate the Data Frame's structure and inserted the data into the SQLite database.



► **Data Visualization:**

- Created a Flask web application to serve as the interface for data extraction and visualization.
- Generated various plots using matplotlib and seaborn to visualize the ecological footprint data, including histograms, box plots, scatter plots, and heatmaps.
- Provided a route in the Flask application to download a PowerPoint presentation summarizing the data extraction, cleaning, storage, and visualization processes. The presentation included slides detailing each step of the process and a sample data table.

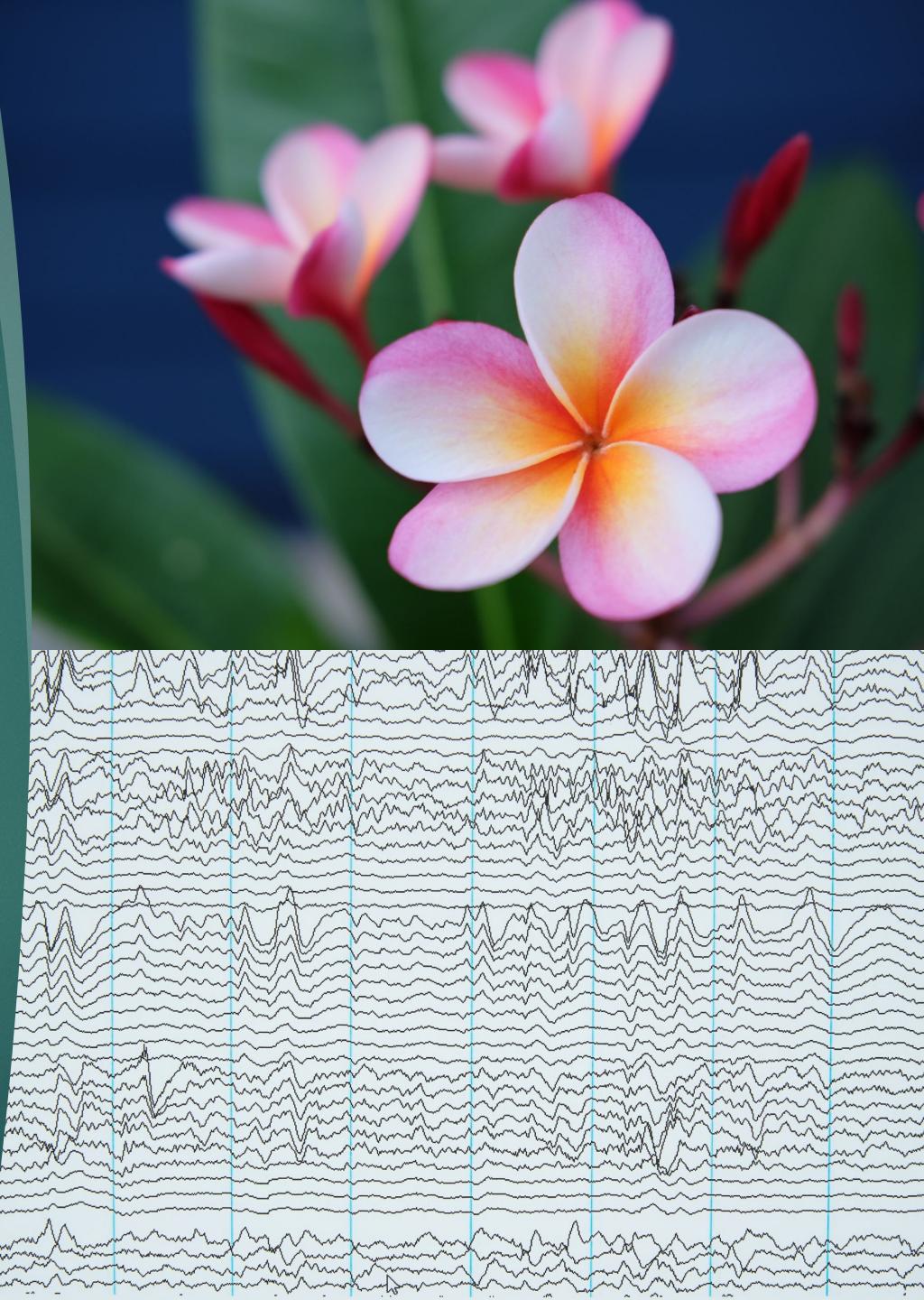


Table Relationships

