

Dataset Title: World Greeter: A Multimodal Cultural Dataset



Dataset Documentation

Dataset Name

World Greeter: A Multimodal Cultural Dataset

Version

1.0

Last Updated

April 20, 2025

Authors

Aditya Kumar

License

- **CSV Data:** [CC BY 4.0](#)
 - **Images:** Sourced from [Unsplash](#) (Free for commercial and noncommercial use under the [Unsplash License](#)). Attribution is recommended but not legally required.
-

Overview

The **World Greeter** dataset is a unique, multilingual, and multicultural resource designed to foster cross-cultural understanding through language, symbols, and visual storytelling. It comprises structured country data (greetings, flags, language, region) alongside high-quality flag and landmark images. This dataset is ideal for developers, educators, and machine learning practitioners looking to build culturally aware applications or multilingual interfaces.

Folder Structure

```
/world-greeter/
├── index.html      → A sample interface to display greetings.
├── data/
│   └── countries.csv → Structured country data (see schema below).
├── images/
│   ├── flags/      → Flag images for each country (PNG).
│   └── landmarks/   → Landmark images for each country (JPG).
```

Schema: countries.csv

Column Name	Type	Description
code	String	One-letter country code (A–Z)
name	String	Full country name
greeting	String	Common local greeting in the native language
language	String	Most widely spoken language in the country
region	String	Continent or geographical region
flag_image	Path	Relative path to the country's flag image (PNG)
landmark_image	Path	Relative path to the country's landmark image (JPG)

Usage Examples

✔ Educational Tools

- Teach kids about greetings and cultures in different countries.
- Used in geography or language apps.

✔ UI/UX Development

- Build a country selector with greetings and visuals for onboarding.
- Language learning platforms can use this for immersive experiences.

✔ Machine Learning

- Train a basic multimodal classifier that associates visuals with metadata.
- Use landmark images for classification or image captioning tasks.

✔ Data Visualization

- Create engaging maps or cultural dashboards using D3.js, React, or Tableau.

How to Use

1. Load the `countries.csv` using pandas or JavaScript CSV parsers.
2. Display each record with its associated flag and landmark image.

3. Customize the UI using `index.html` as a template for rendering greetings.

Attribution for Images

Images used in the `/images/flags/` and `/images/landmarks/` folders are downloaded from [Unsplash](https://unsplash.com), a copyright-free image platform. While attribution is not legally required, it's good practice to credit photographers.

Sample attribution (optional but recommended):

Photos from Unsplash. Flag and landmark visuals by various contributors. Visit unsplash.com for more.

Potential Improvements

- Add ISO 3166 country codes.
 - Include audio files for pronunciation.
 - Provide latitude and longitude for each landmark.
 - Extend support for more than one language per country.
-

Citation

If you use this dataset in your work, please cite it as:

```
@misc{worldgreeter2025,  
  title={World Greeter: A Multimodal Cultural Dataset},  
  author={Your Name},  
  year={2025},  
  url={https://your-github-or-dataset-link.com},  
  note={Version 1.0}  
}
```