Data dictionary of the Infodengue Sprint dataset

This data dictionary has been developed for the **Infodengue Sprint 2024: Predictive Modeling of Dengue in Brazil.** It contains detailed documentation of the tables available for the Sprint, providing essential information for data analysis and modeling.

Disease data

Period: epiweek 201001 to epiweek 202424[1]
Aggregation: cases aggregated by the epidemiological week of dengue symptom onset; municipality
File: dengue.csv.gz
Sources: from SINAN and IBGE, organized by Infodengue

Table 1. Description of the columns in dengue.csv.gz

Column name	Type	Description
date	YYYY-mm-dd	First day of the epiweek (Sunday)
year	int(YYYY)	Year
epiweek	int (YYYYWW)	Epidemiological week defined by the date of symptom onset
geocode	int	municipality code (source IBGE)
casos	int	Number of cases per week, classified as probable dengue cases[2].
Rt	float	Point estimate of the reproductive number of dengue, provided by Infodengue
p_rt1	float	Probability (Rt > 1), provided by Infodengue
regional[3]	str	Health district
regional_geocode	int	Health district code
macroregional	str	Health macroregion
macroregional_geocode	int	Health macroregion code
uf	str	Federative Unit (state)
train_1	bool	data for the first training (pre season 22/23)
train_2	bool	data for the first validation (season 22/23)
target_1	bool	data for the second training (pre season 23/23)
target_2	bool	data for the second validation (season 23/24)

---

[1] Note that the last weeks are subject to update as cases are still being reported
[2] Case definition: Probable cases = Suspected cases - discarded cases
[3] Regional and Macroregional are the subdivisions used by the Ministry of Health.

Climate - reanalysis

Reanalysis hour data from ERA5, summarized by week by the Mosqlimate project. It covers the period from 201001 to 202423.

> **Period:** epiweek 201001 to epiweek 202423[4]
> **Aggregation:** temperature, humidity and precipitation, originally by hour, was first aggregated by day (min, max, mean), and these daily measures were aggregated by epidemiological week (mean).
> **File:** climate.csv.gz.
> **Sources:** Copernicus ERA5, organized by Mosqlimate.

Table 2. Description of the columns of climate.csv.gz. *Atmospheric pressures are given as if the place was at sea level.

Parameter name	Type	Description
date	YYYY-mm-dd	First day of the epiweek (Sunday)
epiweek	int (YYYYWW)	Epidemiological week
geocode	int	municipality code
temp_min	float (°C)	Minimum temperature
temp_med	float (°C)	Mean temperature
temp_max	float (°C)	Maximum temperature
precip_min	float (mm/h)	Minimum precipitation rate
precip_med	float (mm/h)	Average precipitation rate
precip_max	float (mm/h)	Maximum precipitation rate
precip_tot	float (mm)	Total precipitation
pressure_min	float (atm)	Minimum daily sea level atmospheric pressure*
pressure_med	float (atm)	Average atmospheric pressure*
pressure_max	float (atm)	Maximum atmospheric pressure*
rel_humid_min	float (%)	Minimum relative humidity
rel_humid_med	float (%)	Average relative humidity
rel_humid_max	float (%)	Maximum relative humidity
thermal_range	float (°C)	Difference between daily maximum and minimum temperature averaged by week
rainy_days	int	Number of days in the week which `precip_tot > 0.03`.

---

[4] Note that the last weeks are subject to update as cases are still being reported

Ocean temperature and level oscillations

These are organized in three files:

Files	Description
enso.csv.gz	**El Niño-Southern Oscillation.** is a climate pattern in the Pacific Ocean that has two phases: El Niño and La Niña. In a normal year, in the Pacific Ocean, the trade winds blow westward along the Equator and push warm surface waters near Australia and Indonesia. On the other side of the Pacific Ocean, nutrient-rich cold waters come up off the coast of central and South America, creating favorable conditions for fishing. During an El Niño event, the trade winds weaken and warm, nutrient-poor waters are not pushed anymore by the winds, and sea level rises in the eastern tropical Pacific and falls in the western tropical Pacific. La Niña is the opposite phase of El Niño with warm water piling up in the western Pacific and colder water in the eastern Pacific. This causes higher sea level in the western tropical Pacific and lower sea level in the eastern tropical Pacific.
iod.csv.gz	**The Indian Ocean Dipole.** is a climate pattern affecting the Indian Ocean. During a positive phase, warm waters are pushed to the Western part of the Indian Ocean, while cold deep waters are brought up to the surface in the Eastern Indian Ocean. This pattern is reversed during the negative phase of the IOD.
pdo.csv.gz	**The Pacific Decadal Oscillation PDO.** is a long-term (10-20 year) oscillation of the Pacific Ocean in response to the changes in the atmosphere. During a warm (positive) phase, the response of the ocean to low atmospheric pressure over the Aleutian Islands causes ocean currents to bring warm waters in the Eastern Pacific Ocean and along the coast of North America and cool nutrient-rich waters in the western Pacific Ocean. This leads to higher sea levels along the coastlines of the Northeast Pacific. During a cool (negative) phase the Eastern Pacific Ocean becomes cooler and the Western Pacific Ocean becomes warmer. This leads to lower sea levels along the coastlines of the Northeast Pacific.

Environmental data

Environmental characteristics of the municipalities. Other variables can be aggregated as necessary.

> **Period:** 2010
> **File:** environ_vars.csv.gz.
> **Sources:** IBGE, Embrapa

Table 3. Description of the columns of environ_vars.csv.gz

Parameter name	Type	Description
geocode	int	IBGE's municipality code
muni_name	str	Municipality's name
altitude	int	Altitude
koppen	str	main climate type
biome	str	main biome type

Demographic data

Table 4. Population data. Files with population by city and year (2010 - 2021) in POPTBR.csv.gz

Parameter name	Type	Description
MUNIC_RES	int	Same as geocode in other tables
ANO	int	Year (YYYY)
POPULACAO	int	Population of the city

Table 5. Municipalities that are regions of influence (REGIC)

Parameter name	Type	Description
geocode	int	IBGE's municipality code
UF	str	Federative units Brazil
name_muni	str	Municipality's name
hierarquia	str	Municipality's Influence