

# NUTS2 Documentation

Sebastian Hanika

2024-09-20

## Purpose

This is a teaching dataset designed for introductory courses in quantitative methods and statistics. It includes a variety of socio-economic and demographic variables for NUTS2 regions. The aim is to create a straightforward dataset with variables that are not overly complicated and are easy for students to understand, allowing them to explore descriptive and simple inferential statistics. It's important to note that not every student may be familiar with the concept of NUTS regions, so explanations about them and their differences should be provided in the instructions.

## Source

The data is obtained from Eurostat through the *eurostat* package. The script for generating the dataset can be accessed [here](#). Utilizing this script, it should be straightforward to recreate and customize the dataset according to your preferences.

## Variables

Table 1: Variables

Variable	Description	Note
geo	Code of the NUTS2 region	
nuts_name	Name of the NUTS2 region	
country	Name of country	
cntr_code	Code of Country	
region	Name of region	
area	Area in square kilometer	
landlocked	Is the region landlocked?	1 = yes, 0 = no
pop	Total population	
median_age	Median age of the population in years	
mig_rate	Crude rate of net migration	The crude rate of net migration is the ratio of net migration to the average population.
pop_grw_rate	Population growth rate	Population growth rate from 2018 to 2019.
le_t	Total life expectancy in years	
le_gap	Gap between male and female life expectancy in years	It was calculated by subtracting male life expectancy from female life expectancy
gdp	Gross domestic product in Million Euros (PPS) <sup>1</sup>	
sh_trade_services	Share of workers employed in trade and service jobs	
sh_industry	Share of workers employed in Industry jobs	
sh_knowledge	Share of workers employed in knowledge-intensive	
total_jobs	Total number of persons employed	
hrs_t	Average number of hours worked	
hrs_gap	Gap between male workers and female workers	It was calculated by subtracting female hours worked from male hours worked

<sup>1</sup>PPS (Purchasing Power Standard) for GDP is a metric used to compare and standardize the economic output of different countries by accounting for differences in price levels and cost of living. It allows for a more meaningful comparison of GDP between nations by adjusting for variations in purchasing power

Variable	Description	Note
sh_unemp	Share of unemployed persons	The unemployment rate is the number of people unemployed as a percentage of the labor force.

## Data Preparation

All variables are based on the year 2019, striking a balance between data completeness and avoiding unintended developments related to Covid-19. Future updates will be based on more timely data as they become available. If you wish to include the influence of Covid-19, adjusting the script for the desired year is a simple process.

I excluded some countries from the dataset due to missing data for key variables. Most of them are either ETFA or EU-neighboring countries, with the exception of Croatia, which was excluded despite being an EU country.