

# Data Visualization Main Project

---

## Overview

This is the report file requested as part of the CA for Data Visualization module. It's named 'readme.md' so GitHub can show its content as the repository cover.

**This file is part of the following GitHub repository:**

[https://github.com/gsonego/DataViz\\_MainProject\\_Public](https://github.com/gsonego/DataViz_MainProject_Public)

## Summary

This project is the Data Visualization main project which has the objective to implement an dashboard using Shiny for Python, in combination with supporting libraries such as Pandas, Matplotlib, Seaborn, Scikit-learn, and Plotly.

The decision behind this dataset was due to the number of features it has, which describes demographic, economic, social, and infrastructural characteristics of Brazilian municipalities. The other option was a dataset (Hurricanes between 2015–2025), which offered date series information but just some features.

The final dashboard was structured by using a global sidebar with shared filters (region, population range, HDI range, capital vs. non-capital cities) and ten thematic tabs: Home, Overview, Demographics, Economic Indicators, Human Development (HDI), Infrastructure & Services, Agriculture, Cluster Analysis, Regression Analysis, and Regional Structure & Inequality.

The main purpose of this dashboard is to allow users to explore regional differences, development patterns, and also structural inequalities across Brazilian cities, states and regions from multiple analytical perspectives.

## Background Research

A initial research focused on identifying the main indicators which are commonly used to describe regional development and inequality, such as population size, GDP per capita, Human Development Index (HDI).

Special attention was given to understanding Brazil's administrative structure (regions, states, municipalities) and the well-documented socioeconomic heterogeneity across its territory.

This background indicated the selection of variables and the decision to emphasize regional comparisons and normalized views, which seem adequate for a country with strong structural disparities.

## Data source and description

The dataset used in this project is Brazilian Cities, available on Kaggle:

<https://www.kaggle.com/datasets/crisparada/brazilian-cities/>

Brazil is the world's fifth-largest country by area and population and is administratively divided into 26 states, one Federal District, and 5,578 municipalities.

This dataset contains multiple publicly available sources in just one single table, which covers from demographic, economic, social, agricultural, to infrastructural indicators at the municipal level.

After dataset selection, a data verification and preparation phase was initiated. This included the creation of a region attribute based on official administrative divisions, also the correction of cities labeled wrongly as state capitals, and other minor consistency adjustments across a few variables.

Overall, the dataset appears coherent and reliable, with only a small number of isolated inconsistencies that do not significantly affect the analyses presented.

## A short criticism which provides an honest review of the work

When analysing the final work, it's clear to see that while the dashboard provides a broad and multidimensional view of Brazilian municipalities, the absence of a temporal dimension limits the ability to analyze trends or causal dynamics over time.

Some of the selected analysis are just exploratory rather than fully validated models, and their results should be interpreted as descriptive insights rather than definitive conclusions.

The project succeeds in its primary goal: demonstrating how thoughtful data visualization and interaction design can reveal meaningful patterns in high-dimensional data, but maybe some charts could be more clear on its purpose.

Finally, dataset looks interesting but some numbers are outdated and some features had noticeable disparities, like GDP\_CAPITA for some cities.

## A link to a server hosted

A suitable solution to publish the dashboard on a public server could not be identified with the project constraints.

For this reason, I'm providing all the data and source code into this public repository so it can be downloaded (git clone) and the dashboard can be executed locally.

I'm also submitting all the content of this work with the CA submission page.

All the instructions on how to run this dashboard can be found on the **setup.md** file

## Any additional information

Nothing to be reported on this section.

## A basic version history file

### 20/11/2025

The dataset is selected. I was considering two options:

- A global hurricane historical data from 2015 to 2025, which is an interesting topic but contains a limited number of features, like: Date, Time, City, Country, LatLong, Magnitude, Depth and Score.
- A compilation of several publicly available information about Brazilian Municipalities containing 70+ features.

**Note:** The decision was in favour of the Brazilian cities dataset because it contains more feature to be explored, although in this case we don't have date series to be analysed.

### 22/11/2025

Clean-up work. There was not much work involved as there were no missing values and all data types look cohesive. Also, it was time to decide about which columns to keep and those ones to drop.

### 23/11/2025

Create a new column "Region" to cluster states in smaller groups as per Brazil's administrative Region definition. After column creation, I checked if all regions were assigned correctly. Also, found some cities wrongly flagged as capitals.

### 25/11/2025

I tried to use separate files to implement all visual elements but faced many issues, so I opted by using Shiny Express mode with just one file (app.py)

### 25/11/2025

I decided creating the dashboard in just one file even if it ended up big, unfortunately I was wasting too much time trying to make it tidy, clean and component-wise but it was not possible.

At the end, it's better this way so the code will be in just one place for lecturer assessment.

## **26/11/2026**

I Started the creation of the dashboard layout, based on a sidebar to keep global filters and multi-tabs on the center.

Each tab will contain some visualisations for an specific area / topic.

## **27/11/2026**

Global filters added to sidebar:

- Regions
- Population Range
- HDI Range
- GDP per Capita
- Capital / Non Capital

## **28/11/2026**

First charts were added:

- Overview Tab
  - Key Performance Indicators
  - Interactive Brazil Map
  - Regional Comparison

A unit problem was detected, so I multiply some fields by 1000 for better data representation.

## **10/12/2025**

Additional sections/charts created:

- Demographics
- Economic Indicators
- Human Development

## **11/12/2025**

Additional sections/charts created:

- Infrastructure and Services
- Agriculture

## **15/12/2025**

Additional sections/charts created:

- Cluster Analysis
- Regression Analisys

## **16/12/2025**

Additional sections/charts created:

- Regional Structure & Inequality

## **17/12/2025**

- Fix some chart glitches with division by zero operation.

## **05/01/2026**

- Landing page created
- All layouts revised, colours adjusted for a better look and feel.

## **10/01/2026**

Final revision for all charts and report.

## **11/01/2026**

Final version released and submitted