

## Data Cleaning and Preparation Summary

In the provided zip folder, you'll find a well-processed and clean dataset available in both CSV and Stata (.dta) formats. Additionally, the folder contains two files: `main.py` and `main.do`. These files include the code, processes, and techniques applied to ensure that the data is organized and ready for meaningful analysis.

Key points regarding the data cleaning process:

- **Column Reduction:** The dataset initially had over 350 columns, which have been reduced to 20 for more effective analysis and exploration. This refinement was focused on variables of particular interest to the project.
- **Column Merging:** Some columns were combined to create more meaningful insights. However, this merging can easily be reversed if necessary, depending on specific analysis needs.
- **Python-Driven Cleaning:** The dataset was primarily cleaned using Python. A copy of the cleaned dataset was also converted into a Stata file for compatibility and flexibility in further analysis.
- **Final Dataset:** The finished dataset is provided in both CSV and Stata (.dta) formats to accommodate different analytical preferences.

Additional modifications or cleaning can be carried out as needed based on the requirements of your project.