Master script of how to reproduce the work

Download dataset and import in (in Stata or Jupiter)

In Stata

1. Download Stata on your PC or use CEU VPN for a remote access (in my case first option).
2. Figure out the structure of you code in general (how many do files you will have, what will you do in them, etc.).
3. In my case I did 3 do files, first to understand folder structure on PC and do some cleaning as well as, loading dataset and handling missing values, second for data transformation and “job” analyze and the third one for creating table and graphs. Of course this is not a strict rule, it is just for me and other people to understand what is going overall (I did some cleaning in 2 do file for example, as well).
4. Along all this 3 do file I did something income, like I saved them and dataset, made useful comments for easy comprehension, as well as I used appropriate names for variables and files.

In python

1. Download Anaconda- Navigator.
2. Set up the environment (download required necessary libraries, like matplotlib.pyplot, openpyxl and etc.).
3. Loading dataset (we don’t need to make 3 do files like in Stata)
4. Convert date columns, creating new ones and making analysis
5. Summary statistics of specific variables and visualization (random chosen paraments, let’s say, given by out boss)
6. Saving the dataset for future analysis and observations.

Commit files into git version control and push a repository to GitHub.

1. Create a GitHub Repository in GitHub (Also need to create account in GitHub, if u haven’t done so)
2. Open Shell om PC (I had a new Mac, so I opened terminal and switch zsh from to bash)
3. Then you need to link your repository to GitHub.
4. Check the status of your repository
5. Add the files you want to commit.
6. Commit your changes with some meaningful message
7. Push to GitHub