

GOV3219 Tutorial exercise 1

LCH213, February 3, 2026

Please try to do the following analysis with Jupyter Notebook, do it independently.

1. Import the employee data: download the data from Github page, import to jupyter notebook, and name the original data as "data_original".
2. Create a new data that contains only 4 variables (educational level, beginning salary, minority, and gender), and name the new data as "data_analysis".
3. Check the first 15 rows, shape, and datatypes of the new data.
4. Describe the new data in two ways: 1) use "describe" function to see the distribution of each variable, *remember that codes for numerical and categorical variables are different; 2) use visualisation to understand the distribution of each variable, with histogram for numerical variables, and frequency graph for categorical variables (*in this case, gender).