Assignment: Working With BEA Data

Introduction

In this assignment, you are to use data from the U.S. Bureau of Economic Analysis (BEA) to summarize the connections between various industrial sectors in the United States. In this example, the BEA categorizes industries and commodities according to a coding system called the North American Industry Classification System (NAICS). The BEA records how much of each commodity each sector uses and how much of each commodity each sector produces. These are stored in files called `IOUseDetail.txt` and `IOMakeDetail.txt`, respectively. These are provided in the attached folder called `ndn0307`. Included in this directory is a README file that provides information to help you understand the included data. However, you will only need `IOUseDetail.txt` and `IOMakeDetail.txt` for this assignment.

Also, for convenience, I have included a Jupyter notebook called `Understanding_Revshare_Cust_and_Supp.ipynb` that provides examples.

Objectives

Create a Jupyter notebook that imports the BEA data and produces the following items.

- Produce a matrix called `share` that tells us what percentage of the total amount of a commodity an industry produces.
- Produce a matrix called `revshare` that is the matrix of dollar flows from the customer industry in column j to the supplier industry in row i.
- Produce a matrix called `cust` that records the percentage of industry i's sales that are purchased by industry j
- Produce a matrix called `supp` that records the percentage of industry j's inputs that are purchased from industry i. This is the cost share of input i in the production of output j.

For examples of these matrices, see the attached Jupyter notebook.

- Use the Python package called Pandas to handle the data.
- Do your best to use functions built-in to Pandas to process the data.
- Keep your code clean and well documented.
- Note that computing the above tables might not work out as cleanly as you would expect. Do your best given the limitations!