Edward Joseph

The data sets we combined were gotten from Kaggle. One contains information about food nutrition of items eaten in certain countries. The other contains quantities of food bought and consumed by category in certain countries. Combining the data allows for one to analyze the patterns of food consumption in certain countries by year.

To clean the data set first we turned the first data set into a csv from a tsv. Then we pulled the columns that we wanted from each data set. we dropped some columns with the ‘.drop’ function in pandas. We renamed a few columns. Then we finally merged the data sets on the column of ‘country’. Then that was exported to sql. We uploaded it to sql because it was a structured table that was already in a csv format.