**Extraction:**

Data consists of 4 datasets, one is the 2018 World Happiness Report data (WHR), another the World Development Indicators (WDI) from 1960-2019, the third is 2018 world life expectancy, and the final is 2018 world crime rate. One data set, the WDI set, was extracted in the form of an Excel spreadsheet. The remaining datasets were extracted as CSVs. The datasets were drawn from the World Bank (WDI, life expectancy, and crime rate) and Kaggle (WHR).

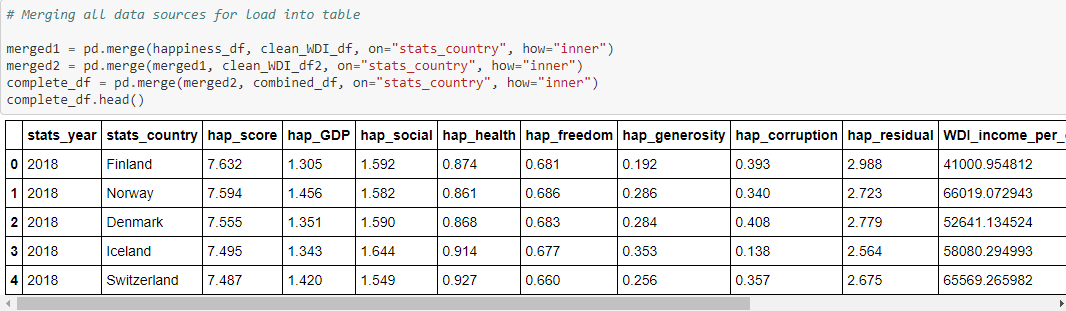
**Transformation:**

Once the data was downloaded, it was trimmed where necessary COUNTRIES (some manual, some via VBA) prior to loading into Python via Jupyter Notebook for transformation. The size of the WDI dataset slowed processes down, so only the variables that were going to be utilized were loaded into Python.

Data was cleaned and transformed using pandas in Jupyter Notebook. Filepaths were created and the files were read into the notebook. Unnecessary columns were removed from each dataframe and fields were renamed as appropriate. The four disparate dataframes were merged into a complete dataframe that could be loaded into Postgres.







**Loading:**

A table schema was created to hold our combined table.



Finally, a connection to the Postgres database was made in the notebook and an engine was created. The connection created the table as well. We also included code to replace the table if the code was re-run.

