Group 4 ETL Project Analysis and Report

Our team extracted our data from the three following resource: <https://www.kaggle.com/sudalairajkumar/undata-country-profiles#country_profile_variables.csv>, <https://www.kaggle.com/andytran11996/cost-of-living> and <https://www.kaggle.com/fkosmowski/nightlight-data-and-mpi-among-kiva-borrowers/data>. The files are in csv format. We also scraped <https://en.wikipedia.org/wiki/World_Happiness_Report> for some data.

In order to manipulate the data, we transformed the four different datasets into cleaned panda data frames. The process in which we did that was to manipulate the included setting an index column, white space stripping while dropping duplicates and NaNs. We also dropped columns from the scraped wiki data.

Next we created a database in MySQL named country\_db with four tables in it to hold the data from the cleaned csv data frames. We then connected to the MySql database through pandas and fed the information into the appropriate tables. Then we joined the tables using MYSQL.