ETL Project
August 3, 2020

Carrie Chunn
Kevin Eggert

We wanted to look at the World Happiness Index and compare it to the Economic Freedom Index to gain a broader view on a countries' overall quality of life. Using the information, we could see the best set of circumstances that gave their citizens the happiest and most economically stable opportunity to live. Using these two datasets, we were able to compare and contrast a country's overall happiness against their economic freedom, specifically utilizing categories such as social support, freedom to make life choices and perceptions of corruption against government integrity, judicial effectiveness and unemployment.

Merging these two datasets, we used both Pandas and Python to download the information and SQL to merge the indices together. Both datasets(csv) were found on Kaggle, the most recent Happiness Index from 2018 - 2019, and the Economic Freedom Index was also from 2019. We read them in using Pandas, created a data frame for both datasets, then pared down the categories so they were more relatable to the other dataset. Downloading the csv from Kaggle for the Economic Freedom Index, we had to change the read using 'encoding='ISO-8859-1'' to overwrite the lack of UTF-8.

We transformed the columns to export to SQL. While in SQL, we created tables using the countries as primary keys, and read the rest of the categories in as various characters. We also created code to show the data merged, using the country name as the common denominator. This allows individuals to be able to view an individual country's results.

These data frames are now loaded in Python as well as tables in SQL, to allow for easier joining. The final tables were econ_index and happiness_index; below is a screenshot of the join for these two datasets. We chose SQL for the export to see how the data worked together, plus it gives greater freedom to choose what information you want to include. Because the tables are what is permanent, the joins can be done over and over without loss of information in translation.