

Extract

Transform

**Extract:**

We took our original data sources from Kaggle and directly from the World Health Organization. All of the data sets we used were .csv files.

Human Freedom Index <https://www.kaggle.com/gsutters/the-human-freedom-index>

FIFA World Rankings <https://www.kaggle.com/tadhgfitzgerald/fifa-international-soccer-mens-ranking-1993now>

World Happiness Report <https://www.kaggle.com/unsdsn/world-happiness>

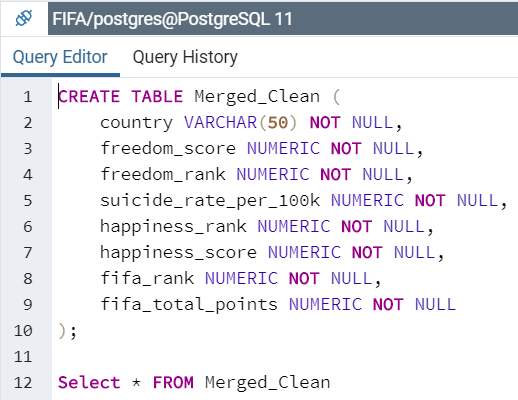
World Health Organization <https://www.who.int/mental_health/prevention/suicide/suicideprevent/en/>

**Transform:**

Transforming the data is what took the longest. Deleting the unused columns was fairly easy using Pandas and Numpy, but the real time sink was in making the data uniform. While each of the data sets had most of the same countries, the naming conventions were not consistent. We had to rename dozens of row names in each table to make sure the data joined correctly.

**Load:**

Once the tables were joined using Jupyter Notebook, we loaded the tables into PostgreSQL because it was the system that we were the most familiar with, and ease of use was a big factor in our decision.



CREATE TABLE Merged\_Clean (

country VARCHAR(50) NOT NULL,

freedom\_score NUMERIC NOT NULL,

freedom\_rank NUMERIC NOT NULL,

suicide\_rate\_per\_100k NUMERIC NOT NULL,

happiness\_rank NUMERIC NOT NULL,

happiness\_score NUMERIC NOT NULL,

fifa\_rank NUMERIC NOT NULL,

fifa\_total\_points NUMERIC NOT NULL

);

Select \* FROM Merged\_Clean