**Final Report**

***Extract:***

We used two CSV’s found on Kaggle to extract into python. The CSV’s where chosen since they both had country columns that could later be used to join them in SQL. Both CSV’s where stored as data frames in python.

Links to both data sets:

<https://www.kaggle.com/unsdsn/world-happiness#2016.csv>

<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016#master.csv>

***Transform:***

Transformation for both data sets occurred using python. The first step for both data sets was to drop the columns that weren’t needed for the final query. Most columns had to also be renamed to eliminate spaces that existed in the original names so they wouldn’t cause issues in SQL.

Suicide data needed to be filtered to 2016 data only since the data for the “Happiness” report was for only 2016.

***Load:***

Finally, the cleaned data was loaded into Postgres. Postgres was used so the data sets could be joined for final analysis. Both tables where joined on the country column that existed in both tables. A where statement was used to filter out any countries that existed in one table but not the other.

\* \*\*L\*\*oad: the final database, tables/collections, and why this was chosen.