# H

# ETL project: World Happiness versus Suicide Rate

# Resources: Kaggle.com

1. World Happiness Report by Sustainable Development Solutions Network.
2. Suicide Rates 2015

# Data Types:

# CSV: master.csv

# SQLITE: happiness. SQLite

**Why did we choose CSV and SQLITE?**

SQLite is often faster than many data base engines. It doesn’t have dependency on other servers. Data Analysis and transfer format can be done easily.

**Extract:** The data was formatted in CSV and SQLITE (csv to SQLite).

**Transform: what data cleaning or transformation was required.**

We did not have to do much to the world happiness data. We extracted the Country and Happiness Score columns.

For the Suicides data we had to first group by country and year, then we had to get all the data from 2015. Then we had to get a sum of the suicides for that year for each country.

**Load: the final database, tables/collections, and why this was chosen.**

Once both datasets were ready, we merged on Country and exported our data into a mongoDB database.