**The sources of data that you will extract from.**

We are using data from the following two links:

* <https://www.kaggle.com/unsdsn/world-happiness>
* <https://github.com/iancoleman/cia_world_factbook_api#data>

The world happiness data is delivered in 3 .csv files, one for each of the years from 2015-2017. These files each contain the following columns:

* Country (Name of the country.)
* Region (Region the country belongs to.)
* Happiness Rank (Rank of the country based on the Happiness Score.)
* Happiness Score (A metric measured in 2015 by asking the sampled people the question: "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest.")
* Standard Error (The standard error of the happiness score.)
* Economy (GDP per Capita) (The extent to which GDP contributes to the calculation of the Happiness Score.)
* Family (The extent to which Family contributes to the calculation of the Happiness Score.)
* Health (Life Expectancy) (The extent to which Life expectancy contributed to the calculation of the Happiness Score.)
* Freedom (The extent to which Freedom contributed to the calculation of the Happiness Score.)
* Trust (Government Corruption) (The extent to which Perception of Corruption contributes to Happiness Score.)
* Generosity (The extent to which Generosity contributed to the calculation of the Happiness Score.)
* Dystopia Residual (The extent to which Dystopia Residual contributed to the calculation of the Happiness Score.)

The CIA Factbook data is delivered in a JSON format and contains a large amount of data points about every country in the world. Because this data set is so large, we chose to select a few interesting data points to bring into our database.

**The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).**

In order to pare down the World Factbook we filtered the large JSON document to a smaller data frame and joined this to the World Happiness data.

**The type of final production database to load the data into (relational or non-relational).**

MySQL database with one table directly mirroring our data frame.

**The final tables or collections that will be used in the production database.**

Table columns: