|  |  |
| --- | --- |
| **ETL TECHNICAL REPORT** |  |

# Team Members: Christin Whitlock, Isha Iqbal

February 13, 2020

**EXTRACT**

The data sources utilized in our ETL project were found on Kaggle.com. Originally both datasets were from the census bureau and provide international health and population metrics of countries since 1950 and projections through 2050. estimates of country populations since 1950 and projections through 2050. Multiple data sources were provided but for our purposes we chose the following:

**DATA SOURCES**

1. Age Specific Fertility Rates
   1. Format: CSV File
   2. Link: <https://www.kaggle.com/census/international-data#age_specific_fertility_rates.csv>
2. Mortality Life Expectancy
   1. Format: CSV File
   2. Link: <https://www.kaggle.com/census/international-data#mortality_life_expectancy.csv>

**TRANSFORM**

Similar processes were used for the transformation of both CSV files. The steps were as follows:

1. CSV file was loaded onto jupyter notebook
2. Dataframe was created from the CSV file through pandas
3. Filtered dataframe was created through removal of columns
4. Columns in filtered dataframe was renamed

**LOAD**

A relational database, (PostgreSQL), was chosen to store the data we transformed in python. SQL Alchemy was utilized in order to create the engine to connect the jupyter notebook to PostgreSQL for the final creation of our tables.