ETL PROJECT - Vaccination Rates (Japan and Canada)

By Rebecca & Petra

MOTIVATION

The reason we chose this data source was to look at flu vaccines and the child vaccines in Japan and Canada. Due to time constraints we selected two countries. We selected Japan and Canada as they had more data to work with. We wanted to collect this data so that it can be used in the future to see if there is a relationship between child vaccination rates and flu vaccination rates in these two countries.

EXTRACT

We extracted 2 datasets from the public platform OECD. Both files were CSV files. The sources for our datasets are as follows:

- https://data.oecd.org/healthcare/child-vaccination-rates.htm
- https://data.oecd.org/healthcare/influenza-vaccination-rates.htm#indicator-chart

TRANSFORM

Our steps for transforming are as follows:

- for each of our datasets, we dropped the ‘Flag codes’ columns, as there was no explanation to what the data related to and also there were multiple NA values
- filtered our data by the two chosen countries
- filtered our child vaccination data down to Measles to simplify the datasets for a better comparison
- merged the data based on “location” and “time” (year) for both datasets

LOAD

The last step was to load our final dataframe into a Database in SQL. We used SQL as we had structured data to load and store in our database. We created a table to hold the merged data frame as one table for easy analysis.