Technical Report

For our extraction we used a CSV file from Kaggle that focused on the 2021 Happiness Report. This file was put into a Resources folder in our repository. Our other source was from Data.World on the 2020 Happiness Report and it was also a CSV file that we placed in the Resources Folder. We used Quick DBD to create a schema. We quickly discovered that it was not taking our column names. With a little assistance, we discovered our problem was that we need underscores between each word.

The next step was to transform the data. We began this process by working in Jupyter Notebook. We loaded all of our dependencies into Jupyter Notebook as well as created an engine for sql. We then created data frames in the notebook.

Finally, we loaded the data. We merged the tables in PgAdmin. We had some difficulty getting the data to load in PgAdmin. We recognized that we had not loaded the database into Jupyter Notebook. So we had to go back into Jupyter Notebook to do that. Because we used, Quick DBD, we then needed to rename our columns to match. We did this by going into Microsoft Excel and renaming each column in the CSV. When that was unsuccessful, we opened up our DBD sql file and edited it, removing all the quotes around column names. The tables then ran successfully in PgAdmin. We then decided to form a union of the 2020 and 2021 data in PgAdmin.