Technical Report

Sources of Data

From Kaggle.com we sourced CSV files for the 2020 World Happiness Report, and five other world datasets scraped from Wikipedia by a third party.

Types of transformations

A schema was set up in Postgres to load all the spreadsheets into one database and join them together into one dataset. Country name was used as the primary key to join the data together. It was then loaded into Pandas to verify the data was clean, i.e. searching for null values, ensuring that there were no duplicate values, etc.

Type of final production database

There was no need to save the data into a relational database, so we used MongoDB.