**ETL Project**

By Elizabeth Overman, Jennifer Borgogno and John Wittenborn

**Summary**

In a time where good news seems hard to come by, we sought to build a database of happiness! Using the results of the World Happiness Report survey of citizens of 156 countries and their perception of happiness, we created a database of additional geographic and social factors that could also be used as determinants of overall happiness perception.

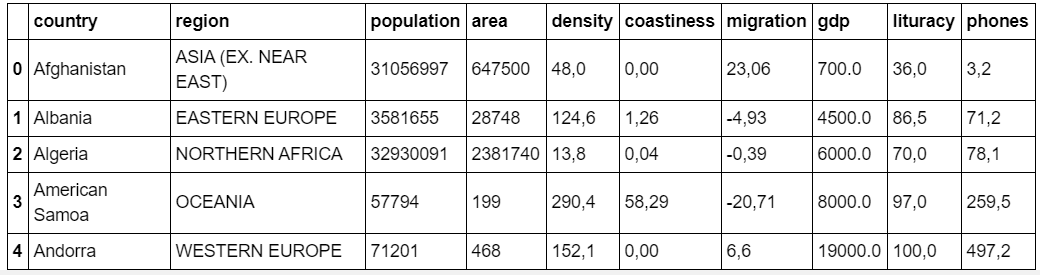
**Extraction**

We used a dataset from the public platform Kaggle and data available from the World Happiness Report. We selected data as recent as possible by using the WHR data from the 2020 report and additional country data compiled to 2017. Data was extracted as .csv from the following:

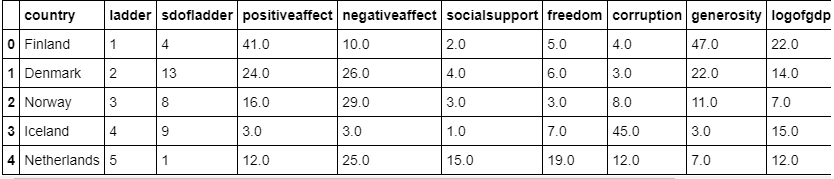


**Transformation**

Our first steps in cleaning the data sets included selecting variables from the Kaggle countries database. From the database of 18 variables, we selected and renamed the variables we found most informative in determining additional factors of country happiness:



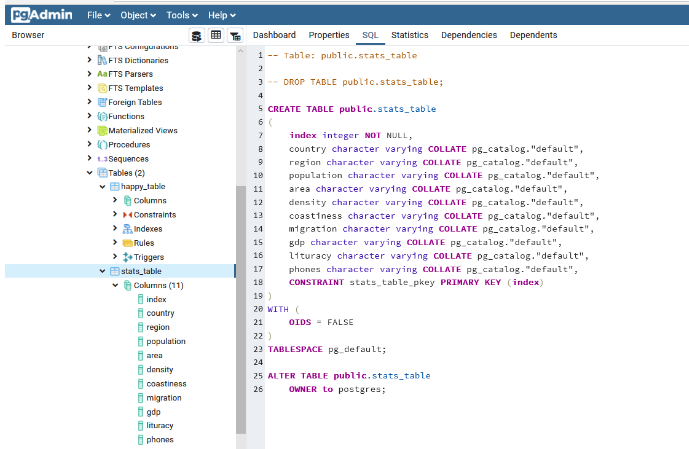
Additionally, we imported and renamed columns of the World Happiness report for clarity:

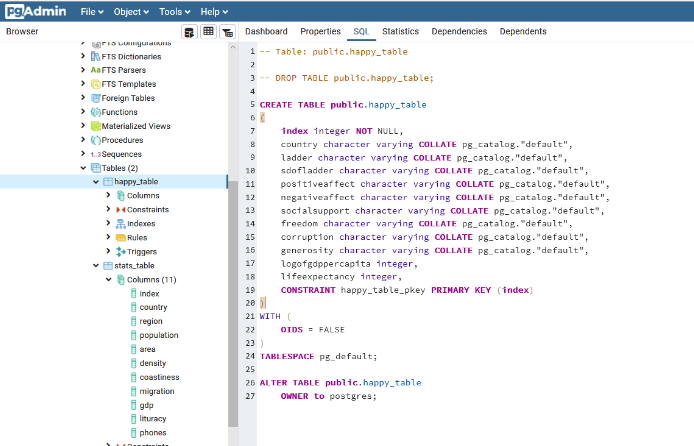


Each dataframe was indexed by country to ensure that the final dataframes could be merged by country if required for further analysis.

**Load**

The last step was to transfer our final output into a database. We created a database and respective tables to match the columns from the final pandas dataframe using pgAdmin and then connected to the database using SQLAlchemy and loaded the result. Here we were able to perform multiple queries to suit a desired search criteria.





Finally, we queried the sql database to ensure that the data was added. Now we have a database that allows for the comparison of geographic and social determinants of happiness! Enjoy!