**ETL Project - Covid Vaccination by Country**

**Members:**

William Powell

Jungkyu Kim

Kaixin Yang

Hannah Slay

Alex Gainer

1. Extract: the original data sources and how they were formatted
   1. Covid World Vaccination Progress:
      1. https://www.kaggle.com/gpreda/covid-world-vaccination-progress
   2. World Happiness Data:
      1. https://www.kaggle.com/unsdsn/world-happiness
   3. UN Country Statistics:
      1. <https://www.kaggle.com/sudalairajkumar/undata-country-profiles>
   4. All three files were found on Kaggle and were csv format
   5. Used pandas to read CSVs files
2. Transform: what data cleaning or transformation was required
   1. Dropped unnecessary columns
   2. Retrieve most recent date from individual datasets using the .max() function
   3. Merged most recent date data frame with individual dataset to isolate latest date
   4. Set index to country
   5. Changed formatting on erroneous records using .replace()
3. Load the final the database, tables, and collections. Why this was chosen
   1. Created ERD using Quick DB Diagrams
   2. Created tables and relationships in Postgres
   3. Loaded cleaned tables into Postgres using .to\_sql in our Jupyter Notebook
   4. We decided to build a relational database as we wanted to join the tables based on country.