## ETL Project Report

**Team Members:**

Allyssa

Emily

Mrun

Nireesha

**Project Name:** Data Integration of Ecological Footprint and Human Freedom Index Datasets

**Datasets to Extract:**

1. National Footprint Accounts 2018

The Ecological Footprint of 196 Countries

Link: <https://www.kaggle.com/footprintnetwork/national-footprint-accounts-2018>

1. The Human Freedom Index

A global measurement of personal, civil, and economic freedom

Link: <https://www.kaggle.com/gsutters/the-human-freedom-index>

**ETL Steps:**

1. Extraction & Initial Data Profiling:
   1. Determine Database Tool:
      1. We selected Postgres as the database tool for importing data from csv files.
   2. Pandas/Manual Extraction steps:
   3. Identify data attributes:
2. Data Cleanup & Transformations:
   1. Remove duplicates
   2. Standardize/Cleanup: (such as column names, datatypes)
      1. Renamed columns by replacing spaces between words with underscore
      2. Changed the datatype of GDP Per Capita to populate as float data type in the tables
   3. Aggregations/Joins:
3. Load:
   1. Database name:
      1. Database name is “freedom\_footprint\_db”
   2. Tables/Columns/Datatypes:
      1. Two tables “freedom” and “footprint” are created using Pandas Dataframes
      2. Columns and data types in the database are created based on the Pandas Dataframes
   3. Filter Queries (if any):