Demetria

Jethro Torczon

Xiaobing Peter Shang

**Purpose**

The database includes the data we collected from different data sources of the economic and governance outlook and GDP grow for all countries. The ranges from various ranking on economic freedom, country GDP growth, government regulation, labor market, real property, and much more. The database enables end users to access all information needed in one place and compile them for various purposes, includes, but not limit to, analyze specific country, region, and world-wide economic development, market freedom, and business environments to make business, economic, and political decisions.

The ETL process was used to create a database in PostgreSQL containing 6 datasets. Each dataset has a unique identifier, and a variable to relate to at least one other table in the database via join/merge or query/subquery. The purpose of building database is to exercise and demonstrate the benefit of database. Proper design and administration of database to improve the data store, information share, quick access of information, and increase productivity.

Below three sections lay out the detail information on how we completed to finish data extraction, cleaning/transform, and loading process.

**Extract**

* Download economic freedom of the world csv file from kaggle.com/datasets
* Download GDP annual growth file (data source: <https://query.data.world/s/ygxz7jq3k2thay5keu7vbpne6ss2bc>)
* Extract relevant information into different CSV files to ensure dataset focus on specific factors:

1. Economic freedom
2. Country legislation and political influence,
3. Country government investment and military spending,
4. Country market regulation
5. Country currency policy

**Transform**

* Convert dataset into Pandas DataFrame
* Cleaning each dataset to drop null information.
* Create output files to store clean data and ready for data loading

**Load**

The procedure for loading all the datasets into PostgreSQL using Python is as follows:

* Create database in PostgreSQL
* Create a connection in python file to PostgreSQL using SQLAlchemy
* Read the dataset from Pandas DataFrame
* Clean database as needed to make SQL friendly names
* Commit table as a new table in the PostgreSQL database.
* Confirm data has been added by querying the tables in Python and PostgreSQL.

The final product is a database with 6 unique tables. Each table has a unique identifier, and all tables can be referenced in another table by a common identifier. Below are examples of how the database looks like , and demonstrate how information can be retrieved from Python and PostgreSQL:



