World Development Indicator:  
PostgreSQL ETL Documentation

Data Sets: Data is extracted from the wdi.worldbank.org/table as CSV files. A total of 3 data sets; Size of the economy wv1.csv, Global goals: ending poverty and improving lives wv2.csv, and Global goals: promoting sustainability wv3.csv.

ETL Objectives:

* Create new PostgreSQL database in PGAdmin called world\_development\_indicator\_db with 3 tables: economy\_size, ending\_poverty, promoting\_sustainability
* Extract all 3 CSV files to a different Pandas Dataframe in Jupyter Notebooks.
* Transform each Dataframe to match to Schema for PostgreSQL world\_development\_indicator\_db database.
* Load transformed Dataframes to PostgreSQL world\_development\_indicator\_db, via Pandas to\_sql function in Jupyter Notebook.

# Data Extraction

* Jupyter Notebook created for Project.
* Dependencies:
  + Pandas
    - Dataframe Functions
  + Columns Header combine and format
    - Combining multi rows of header name
    - Formatting header
    - Removing some notes at the bottom of each data file
    - Replace invalid data “..” in date fields with “NaN”
  + SQL (create\_engine)
    - Load Dataframes into
  + Config (File)
    - Contains PGAdmin Password

# Size of the economy wv1.csv Transformation

* Extracted CSV files via Pandas read\_csv function

Graphical user interface, text, application, chat or text message

Description automatically generated

* Combine headers of multiple rows

A picture containing qr code

Description automatically generated

* Formatting headers: lower case, replace “ “ with “\_”, “,” with “”, “-“ with “\_”, “&” with “”

Text

Description automatically generated

* Formatting headers: rename some column header

Text

Description automatically generated with medium confidence

# Global goals: ending poverty and improve lives wv2.csv Transformation

* Extracted CSV files via Pandas read\_csv function

Scatter chart

Description automatically generated

* Combine headers of multiple rows

Scatter chart

Description automatically generated

* Formatting headers: lower case, replace “ “ with “\_”, “,” with “”, “-“ with “\_”, “&” with “”

Graphical user interface, application, Word

Description automatically generated

* Formatting headers: rename some of column headers

Text

Description automatically generated

# Global goals: Promoting sustainability wv3.csv Transformation

* Extracted CSV files via Pandas read\_csv function

Text

Description automatically generated with low confidence

* Combine headers of multiple rows

Text

Description automatically generated

* Formatting headers: lower case, replace “ “ with “\_”, “,” with “”, “-“ with “\_”, “&” with “”

Graphical user interface, text

Description automatically generated with medium confidence

* Formatting headers: rename some of column headers

Graphical user interface, text, application

Description automatically generated

# More cleaning up of data fields

* Removing notes at the bottom of each data file

Graphical user interface, text, application

Description automatically generated

* Removing invalid data “..” with “NaN” in data field

Graphical user interface, text, application, chat or text message

Description automatically generated

# PostgreSQL Database

* Create Table Statements Text

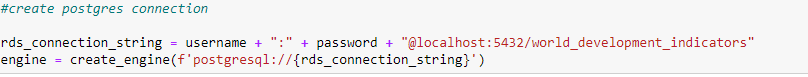
  Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Load to PostgreSQL Database

* Create Database Connection to PostgreSQL world\_development\_indicator\_db: Config file contains Password



* Confirm available tables in PostgreSQL world\_development\_indicator\_db



* Load Transformed Dataframes into PostgreSQL world\_development\_indicator\_db

A picture containing text

Description automatically generated

# SQL Queries

Table

Description automatically generated with low confidence