Population & Energy: A Comparison look at Canada & United States

By: Ashvin Rajput & Holly Clark

**Purpose:**

The purpose of this project was to extract data, transform the data to our liking and load the data into a database. Our project is a comparison between energy consumption in U.S. and Canadian by geopolitical regions and corresponding populations.

**Extract:**

By reading websites, such as Wikipedia, into our Jupyter Notebook we were able to extract data on U.S. Populations by state into pandas data frames. We gathered CSV files from U.S. Energy Information Administration on individual state energy consumption. Further, we gathered CSV files on the Canadian population and energy consumptions by province from Statistics Canada.

**Transform:**

Types of transformations included dismantling multi-index data frames from Wikipedia to simplify data clean up process. We converted data types, remove unwanted data, finally we merge all data frames to a select few data frames.

Challenges we experienced included working with multi-index data frames from HTML websites. First, we were trying to navigate how each data frame was indexed, we were able to solve the issue by deleting a level of the multi-index data frame, we also benefited by using iloc to manipulate the data frame to see certain columns.

**Load:**

We created a connection in our Jupyter Notebook using Pandas to load our data frame into Postgres. By doing so we were able to be store our data frame as the database, ‘ETL Population & Energy’ in Postgres.

We experienced challenges establishing the connection between Jupyter Notebook and Postgres, we received guidance from the Instructor and Teaching Assistants to successfully load our database into Postgres.

**Conclusion:**

Our database shows the relationship between energy consumption by regions in Canada and the United States in relation to the population of each state or province. The database allows users to see the total energy by pairing this data with total populations, a user can understand energy consumption patterns by geopolitical regions.