ETL PROJECT

1. Data used for this project:

* Diversity Index data and Median Income data.

Data Source:

* Kaggle -: [https://www.kaggle.com](https://www.kaggle.com/)
* Data World -: [https://data.world](https://data.world/)

Project Proposal:

1. Project title: ETL PROJECT

2. Project duration: One week.

3. Project Summary:

a. Extraction the data from different datasets and sources

b. Convert the data to csv file, if not already that specified format

c. Use pandas to Transformation the data:

1. Data Cleanup & Analysis

2. Data merging

3. Load the data to postgres

II. Project Report

1. Data extraction:

Kaggel and data world were used to obtain the data. Once obtained, the data were converted or save into a csv format for further analysis. It is to be noted that in the search for the data, it was important to not just look for the data but also think of the data from a relational viewpoint, how the datasets relate to each other, how the data will be transformed.

1. Data Transformation – during the transformation process, Jupyter was used to read the csv files. Pandas and sqlalchemy were imported as dependencies to transform the data. Each csv file was transformed into Pandas Dataframes to be further transformed. The dataframes went into a cleaning stage where the columns were being renamed for merging and loading into postgres.
2. Data Loading – the csv files were also uploaded into postgres where a connection was established between Jupyter notebook and Postgres. Once the connection was established, we checked that the tables were indeed uploaded into Postgres.