**Project Proposal**: **ETL\_project**

**Datasets**: In GitHub repository. (We might use a better dataset if we find one later, but their content will essentially be the same)

**Link:** <https://github.com/sabyasachi31/ETL_Project>

**Data Sources:** kaggle.com

**Purpose:** Correlating the average temperature of major world cities with parameters like population, latitude, and longitude.

**Importing into Jupyter:**

Both datasets are csv files. So, importing them would be just using “read\_csv” function of pandas.

**Scope of transformation:**

We don’t anticipate too much transformation as these are csv files and the columns are rows are pretty complete. However, we do need to make sure that the city names are formatted consistently across both tables.

**Storing Data:**

The data is quite uniform and consistent. So, we will be using SQL.