**Team Members:** Zouheir Elhalabi, Krystal Davis, Brian Knorr, Kashish Misbah

**Objective:** Extract, Transform and Load multiple sources of data sets pertaining to AirBnB listings in relation to various locations and populations in Chicago.

**Extract:**

**Airbnb listing information:** <https://www.kaggle.com/datasets/jinbonnie/chicago-airbnb-open-data?select=listings.csv> (CSV)

**Population and Demographic:** <https://www.kaggle.com/datasets/sergejnuss/chicago-community-areas-demographics> (CSV)

**Community Area Numbers:** "Resources/chicago-community-areas.csv" (CSV)

**Transform:**

**Source: Airbnb**

* Changed ID to index column
* Dropped neighborhood group column
* Dropped rows that had values of NaN
* Separated the data frame into two for data normalization purposes
  + First involving listing information named “listing\_info\_df”
  + Second involving location information called “location\_info\_df”

**Source: Population and Demographic**

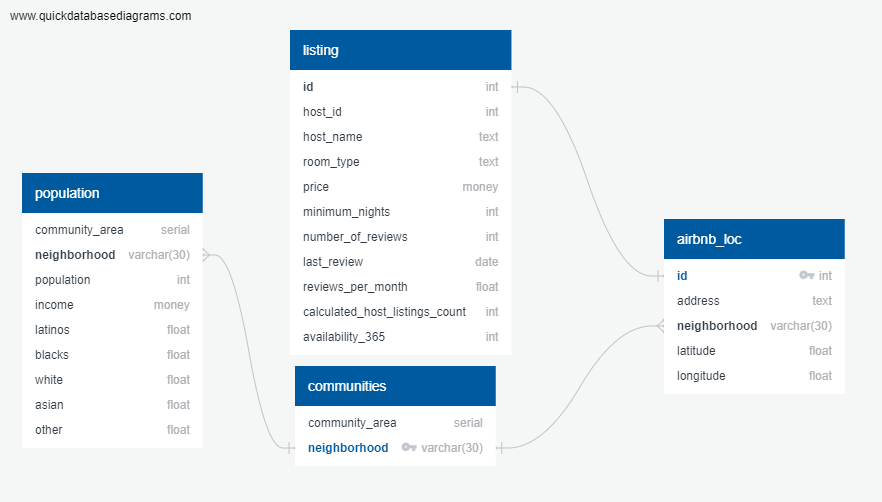
* Separated data by “,” in order to make data usable
* Set index as community area
* Transposed table to make it joinable in postgres

**Source: Community Area Numbers**

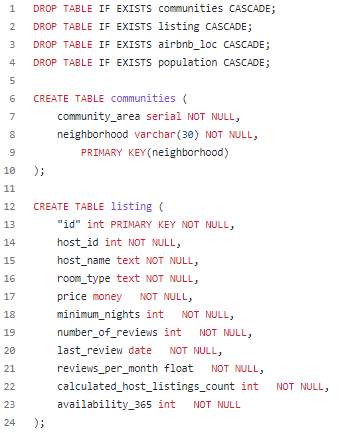
* Set index as community area
* Transpose the table to make data joinable with Airbnb data
* Renamed column headers to more easily load into postgres

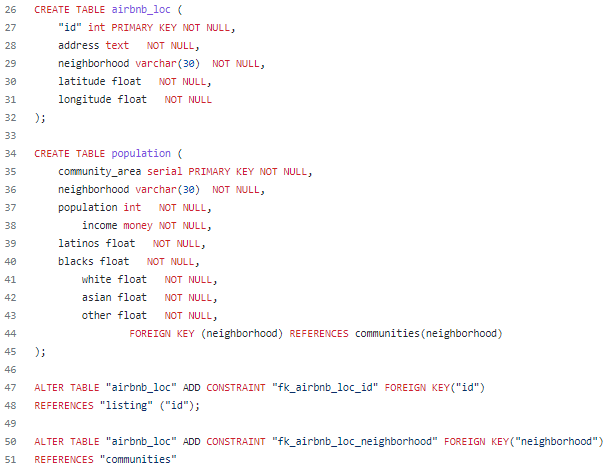
**Load:**

**Table Relationship Diagram**

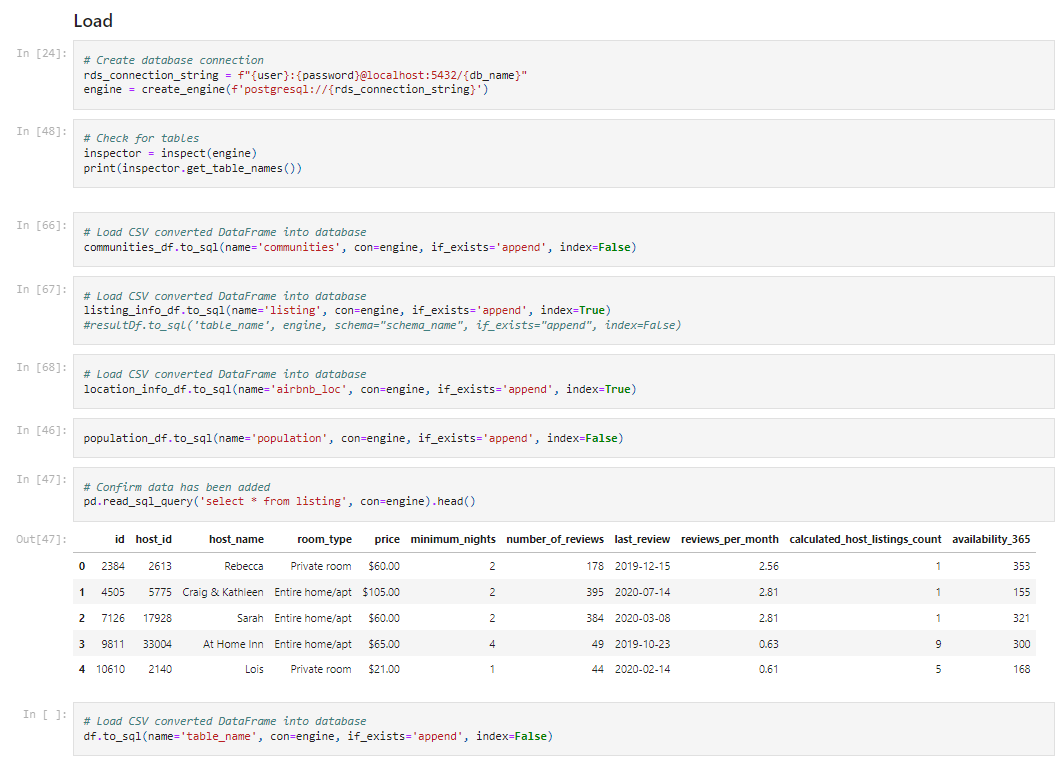
****

**Postgress Table Formatting**

****

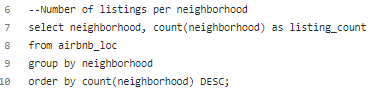
****

**VS Code**

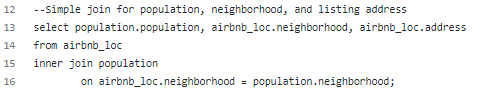
****

**Relationship Queries and Tables**

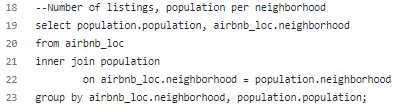
**Query 1**

****

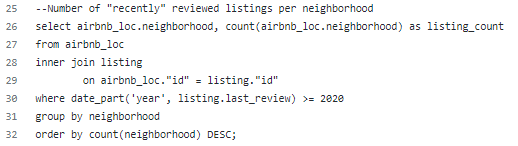
**Query 2**

****

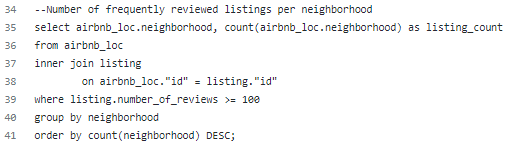
**Query 3**

****

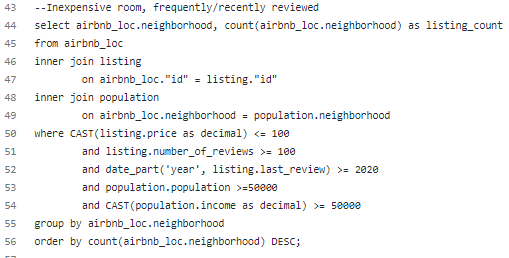
**Query 4**

****

**Query 5**

****

**Query 6**

****