**Introduction:**

To see if there is a profitability trend in the rental price and gross capitalization rate in U.S residential real estate sector.

**Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc)?**

The first data source is comprised of CSV files on rental price per square foot. These CSV files were obtained from Zillow.com. A total of 9 CSV’s were downloaded from Zillow.com. The second data source was a JSON file populated with latitude and longitude info. The JSON file was obtained by calling the Google Maps API. The purpose of the Google Maps API was used for mapping the rental price per square foot.

**Transform: what data cleaning or transformation was required?**

CSV files were checked to see if there were any missing values (NAN’s), and a dropna method was executed on the Data Frame.

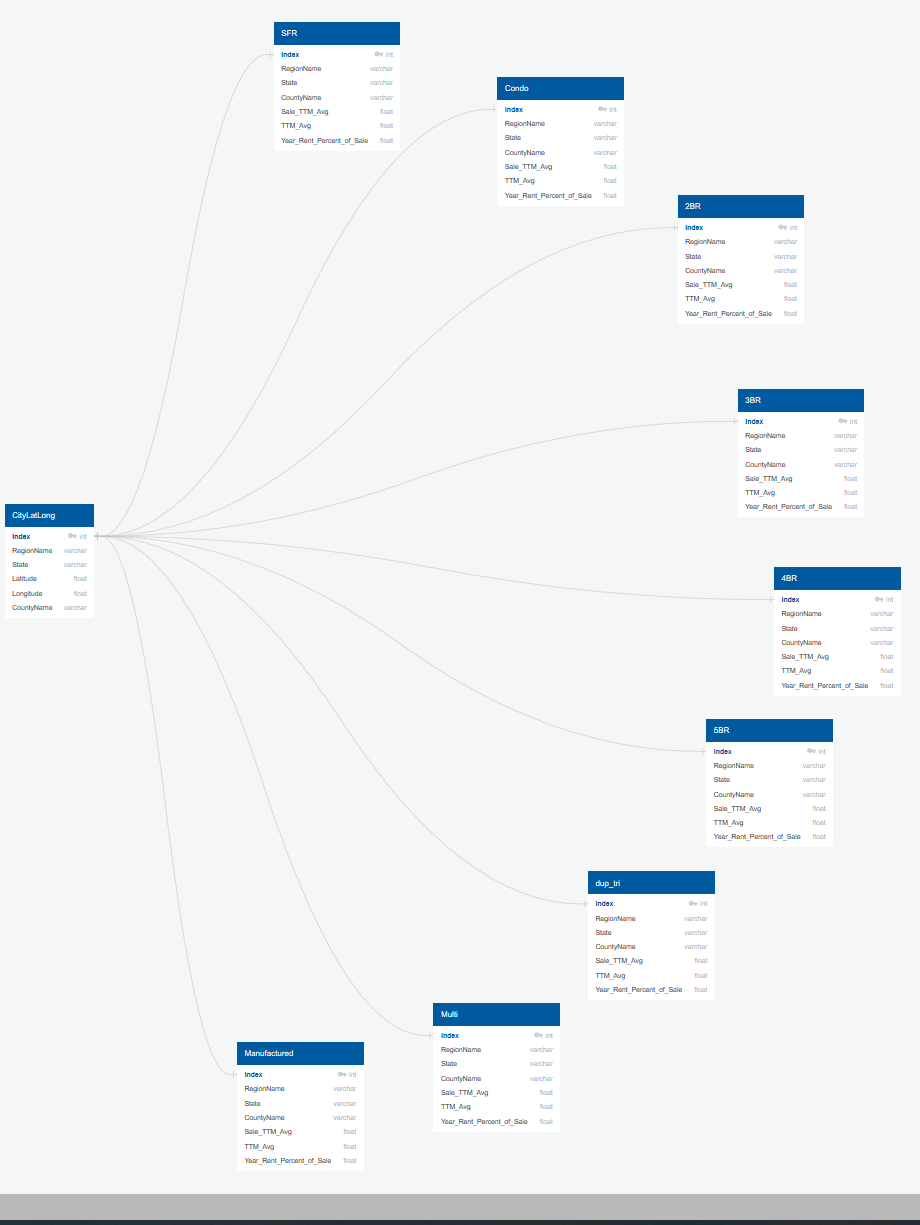
**3)Load: the final database, tables/collections, and why this was chosen.**

For the data retrieved from Zillow, each CSV was broken into their own table, an index was assigned to the tables, and finally primary keys were assigned each table. Similarly, the lat and long were assigned their own table with an id column.

We chose the final database to be structured, as opposed to non-structured, since we are dealing with quantitative data exclusively. Choosing a structured database also makes it simpler as each row has an id.

Finally, we used SQLAlchemy to stage and commit the data contained in the csv’s into the database.

**ERD DIAGRAM: (larger image can be viewed in GitHub directory)**



**Heat Map of Price Per Square foot > $1,500**

