Project: Airbnb Analysis

We downloaded two CSV files from the open source website data.world, one representing Airbnb data from Seattle and the other from Texas to begin our analysis, as this site was recommended in class by our instructor.

We used libraries from Anaconda in Jupyter Notebook to begin manipulation of our data. After importation of pandas and sqlalchemy to create and engine, we extracted the CSV data into a data frame on pandas. We encoded the data using the ISO-8859-1 encryption and added an id column to our data frame. After successfully extracting data, we proceeded to the next step to transform the data frame. We dropped many columns which we deemed to be irrelevant to our search and then we isolated the four columns of price, bedrooms, address/city, and id for both of our datasets. We then set the index of this new table to be bedrooms as we wanted to compare room sizes and prices between the two regions.

The next step was to create a database connection to prepare to upload the data into SQL. After establishing a connection and confirming the table names we then created a database in SQL, wrote schema for our tables and then proceeded to upload the information into the SQL tables after verifying the names of all the engines, databases, tables, and columns. The better part of our later work was spent resolving issues arising from dataset naming problems.