Justus Jarvis

Brandon Spruill

ETL Project

For our project we chose to work as consultants for a property management group looking to expand in major Airbnb markets. To start off the project we had to extract data from 2 different sources. The 2 data sources in which we extracted from were a csv and a website. The csv was then formatted and cleaned up to only contain the column/data that we wanted to merge with our other source. The website we used as a source required up to scrape data that contained the top most expensive cities to live in, and additional data about each city. After cleaning up the data we actually wanted use we then transformed it into 2 separate data frames using jupyter notebook. After the transformation of the data was complete, we then moved on to loading/merging into PGAdmin. After creating the database, imported the tables that contained the city stats and property stats. PGadmin allowed us to merge the 2 data frames into 1 table, which helped up look at the top cities in which Airbnb has properties and compare things like the property attributes (ie. Bedrooms, bathrooms,etc.), to the average cost of living in each city, the average price for a house in each city, and the average household income.