UCSD Data Science Bootcamp Project 2 Proposal, 2/19/20

AirBnB Rental Rates vs Local Income

Team Members:

* Grant Thompson
* Arundhati Chakraborty
* Alexis Perumal

Assignment

* Use two or more data sources from sources such as data.world, Kaggle.
* Perform ETL on that data: sources, transformations, DB type (relational, nosql), final tables put into production database.
* Technical report with above information and steps to reproduce the ETL process.

Project Objective

* Groom a dataset that a data analyst can use to evaluate AirBnB rental rates vs. local income and other census characteristics.

Project Description/Outline

* Extract 3 data sets that we can join
  + [AirBnB rental rates for New York City, includes Lat/Long](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data%20kaggle.com) (Kaggle)
  + [API to convert Lat/Long to Census Tract](https://geo.fcc.gov/api/census/) (geo.fcc.gov api)
  + [Census Income Data by census tract](https://www.kaggle.com/muonneutrino/new-york-city-census-data) (converted to census tract)
* Join the data sets to allow us to understand AirBnB rental rates and other attributes by census tracts.

Research Questions to Answer

* How do AirBnB rental rates correlate with local income levels.

Rough Breakdown of Tasks

* Pull down the three data sets into a Jupyter Notebook.
* Join them together.
* Put the joined results into our Postgres DB.
* Do some simple data exploration to convince ourselves that it is usable.
* Build a report.