UCSD Data Science Bootcamp Project 2 Report, 2/24/20

AirBnB Rental Rates vs Local Income

Team Members:

- Grant Thompson
- Arundhati Chakraborty
- Alexis Perumal

Project Objective

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

Extraction -

We pulled data from two data sets:

- AirBnB rental rates for New York City, includes Lat/Long (Kaggle)
- Census Income Data by census tract (converted to census tract)

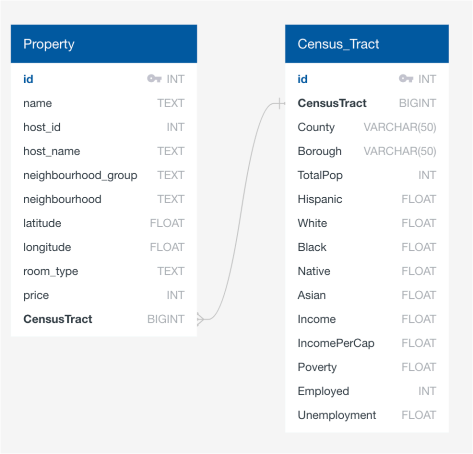
The AirBnb dataset describes the listing activity and metrics in NYC, NY. It includes county and township of each listing, the listing price, and latitude and longitude data for the listing.

The Census Data for New York includes information down to the Census Tract level, which includes approximately 8,000 people. It has information on income, income per capita, race, and employment statistics.

Both datasets are formatted in CSV. We used Pandas to extract and explore the data.

Transformation -

We wanted to join the two tables together with this schema.
Intended Schema



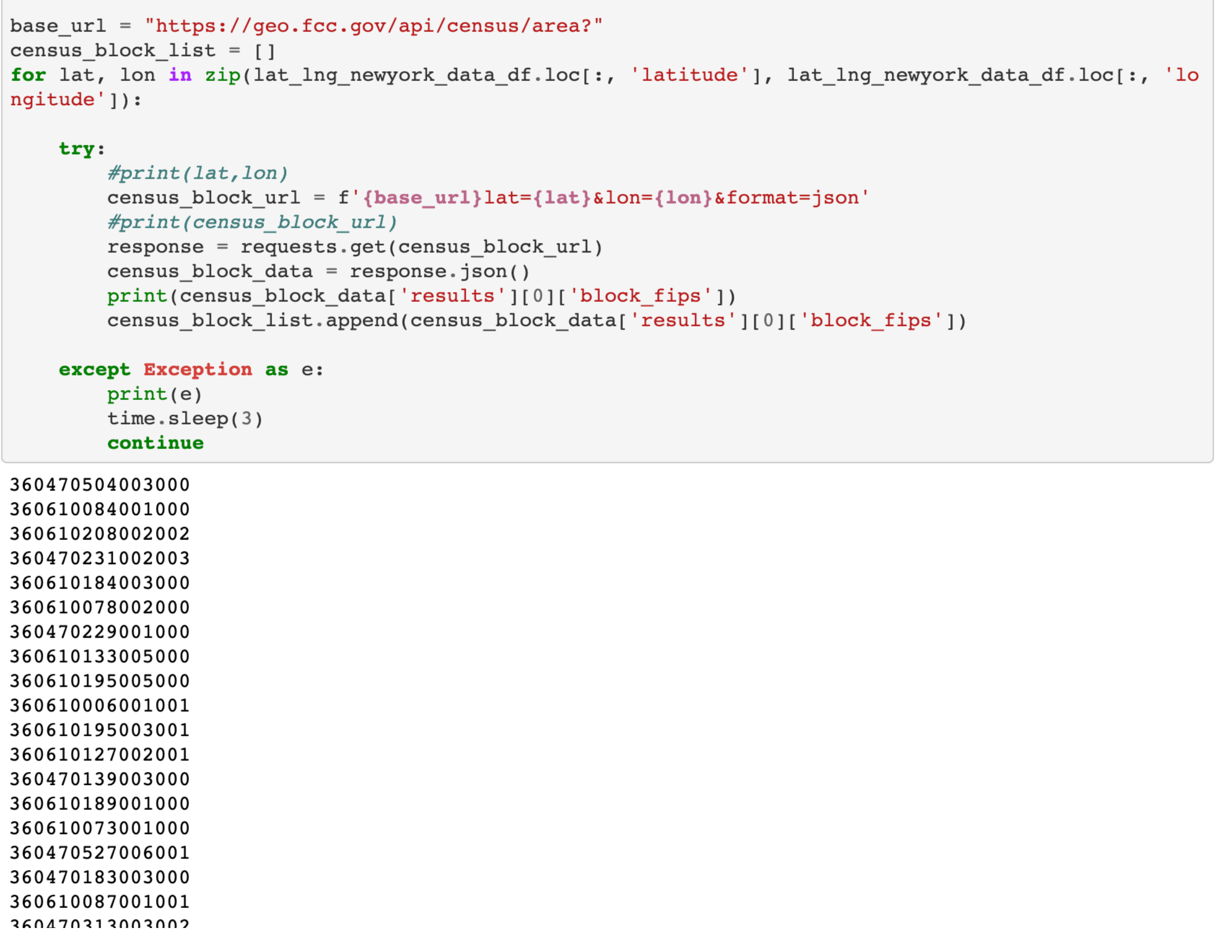
Property	
id	INT
name	TEXT
host_id	INT
host_name	TEXT
neighbourhood_group	TEXT
neighbourhood	TEXT
latitude	FLOAT
longitude	FLOAT
room_type	TEXT
price	INT
CensusTract	BIGINT

Census_Tract	
id	INT
CensusTract	BIGINT
County	VARCHAR(50)
Borough	VARCHAR(50)
TotalPop	INT
Hispanic	FLOAT
White	FLOAT
Black	FLOAT
Native	FLOAT
Asian	FLOAT
Income	FLOAT
IncomePerCap	FLOAT
Poverty	FLOAT
Employed	INT
Unemployment	FLOAT

However, since the two tables didn't have matching data for a join, we had to convert the latitude/longitudes of the table 1 AirBnb listings to the appropriate Census Tract codes to match table 2. To do so, we used the geo.fcc.gov API to convert Lat/Long to Census Tract. The API returns Census Block data, which is more granular than Census Tracts (~8,000 people compared to ~1,500 people). To fix this, we simply dropped the last four digits of the Census Block to get the Census Tract.

The API queries took about 5 hours to run to do 48,895 calls.

API Code and first few responses:



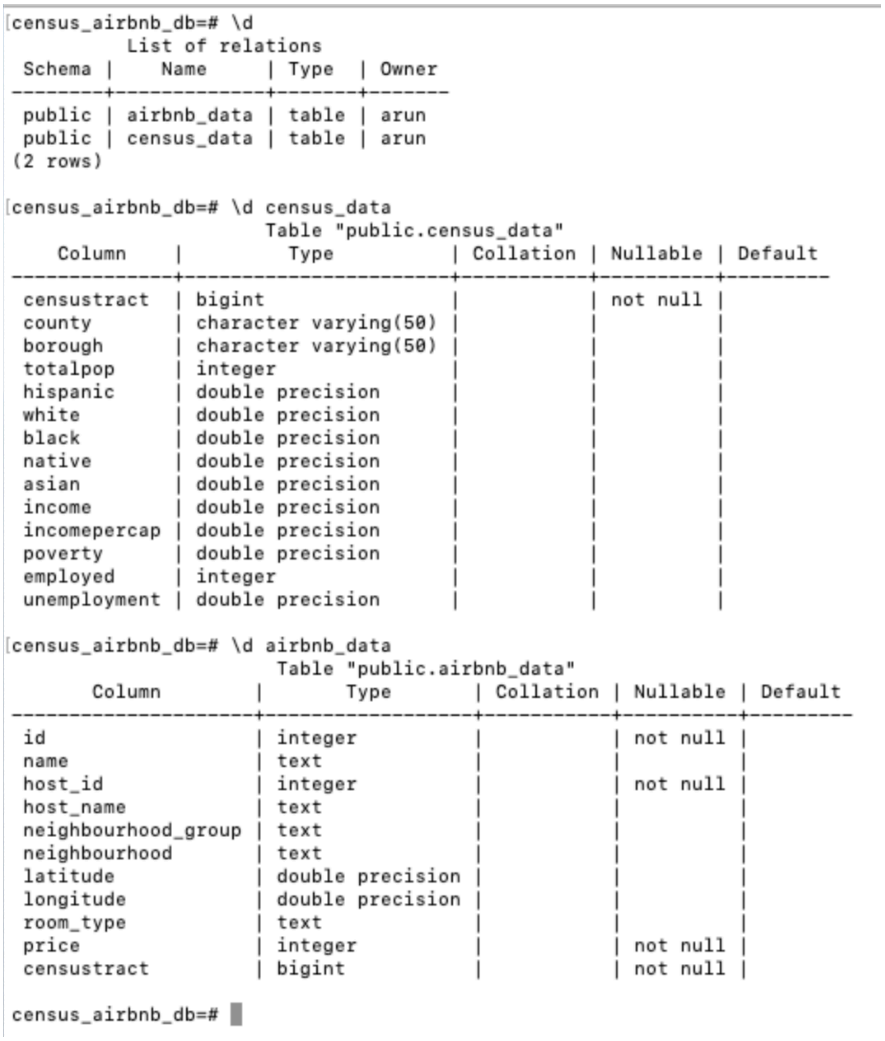
```
base_url = "https://geo.fcc.gov/api/census/area?"
census_block_list = []
for lat, lon in zip(lat_lng_newyork_data_df.loc[:, 'latitude'], lat_lng_newyork_data_df.loc[:, 'lo
ngitude']):

    try:
        #print(lat,lon)
        census_block_url = f'{base_url}lat={lat}&lon={lon}&format=json'
        #print(census_block_url)
        response = requests.get(census_block_url)
        census_block_data = response.json()
        print(census_block_data['results'][0]['block_fips'])
        census_block_list.append(census_block_data['results'][0]['block_fips'])

    except Exception as e:
        print(e)
        time.sleep(3)
        continue
```

```
360470504003000
360610084001000
360610208002002
360470231002003
360610184003000
360610078002000
360470229001000
360610133005000
360610195005000
360610006001001
360610195003001
360610127002001
360470139003000
360610189001000
360610073001000
360470527006001
360470183003000
360610087001001
360470313003002
```

This result included 4 low order decimal digits we didn’t need, so we dropped those. That then allowed us to create a modified version of table 1 with the census track data added to each row.

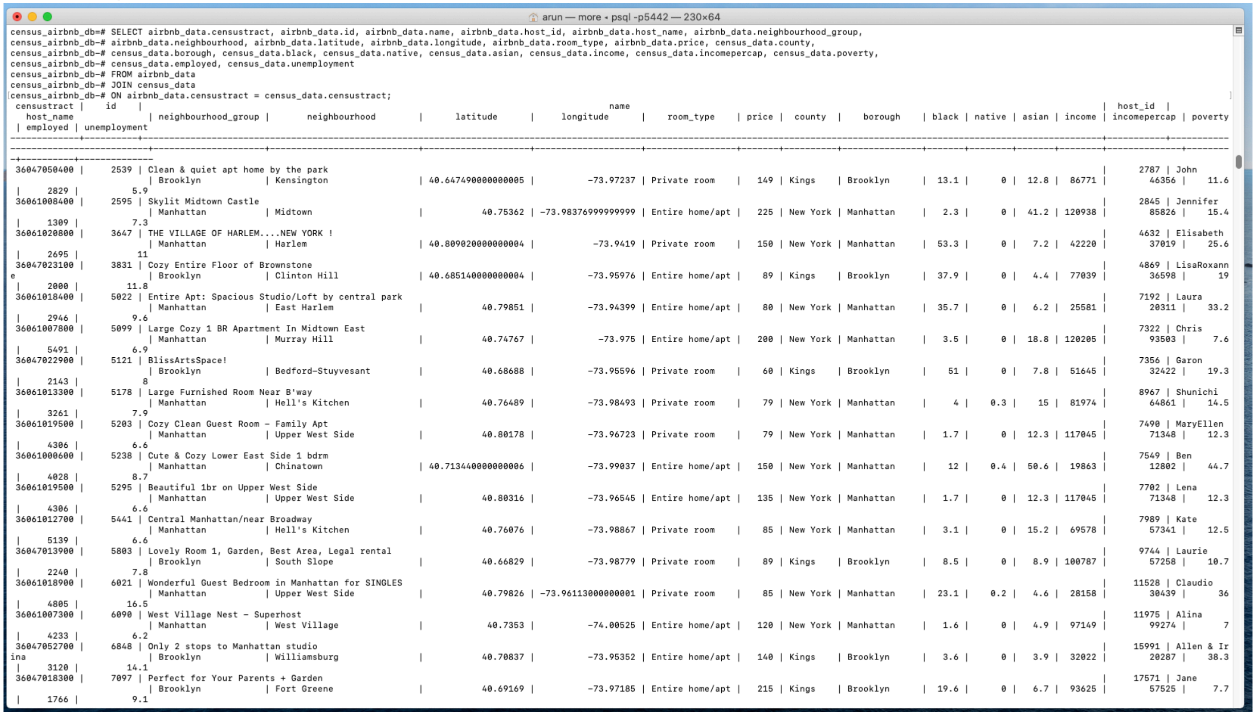
This can be seen in the schema doc.

Actual Schema: DB, Table 1 airbnb_data, Table 2 census_data



```
[census_airbnb_db=# \d
          List of relations
 Schema |    Name     | Type  | Owner
--------+-------------+-------+-------
 public | airbnb_data | table | arun
 public | census_data | table | arun
(2 rows)

[census_airbnb_db=# \d census_data
                      Table "public.census_data"
    Column    |         Type          | Collation | Nullable | Default
--------------+-----------------------+-----------+----------+---------
 censustract  | bigint                |           | not null |
 county       | character varying(50) |           |          |
 borough      | character varying(50) |           |          |
 totalpop     | integer               |           |          |
 hispanic     | double precision      |           |          |
 white        | double precision      |           |          |
 black        | double precision      |           |          |
 native       | double precision      |           |          |
 asian        | double precision      |           |          |
 income       | double precision      |           |          |
 incomepercap | double precision      |           |          |
 poverty      | double precision      |           |          |
 employed     | integer               |           |          |
 unemployment | double precision      |           |          |

[census_airbnb_db=# \d airbnb_data
                       Table "public.airbnb_data"
       Column        |       Type       | Collation | Nullable | Default
---------------------+------------------+-----------+----------+---------
 id                  | integer          |           | not null |
 name                | text             |           |          |
 host_id             | integer          |           | not null |
 host_name           | text             |           |          |
 neighbourhood_group | text             |           |          |
 neighbourhood       | text             |           |          |
 latitude            | double precision |           |          |
 longitude           | double precision |           |          |
 room_type           | text             |           |          |
 price               | integer          |           | not null |
 censustract         | bigint           |           | not null |

census_airbnb_db=# █
```

Resulting Joined Table:

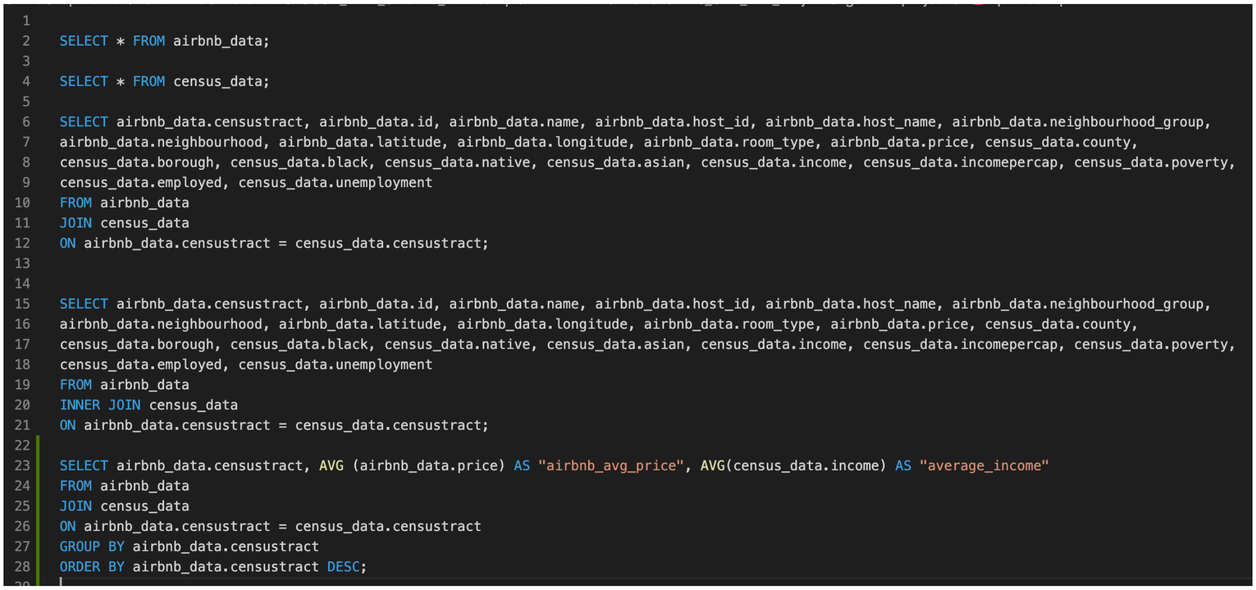


```
census_airbnb_db=# SELECT airbnb_data.censustract, airbnb_data.id, airbnb_data.name, airbnb_data.host_id, airbnb_data.host_name, airbnb_data.neighbourhood_group,
census_airbnb_db-# airbnb_data.neighbourhood, airbnb_data.latitude, airbnb_data.longitude, airbnb_data.room_type, airbnb_data.price, census_data.county,
census_airbnb_db-# census_data.borough, census_data.black, census_data.native, census_data.asian, census_data.income, census_data.incomepercap, census_data.poverty,
census_airbnb_db-# census_data.employed, census_data.unemployment
census_airbnb_db-# FROM airbnb_data
census_airbnb_db-# JOIN census_data
census_airbnb_db-# ON airbnb_data.censustract = census_data.censustract;
 censustract |    id    |                                         name                                                                  | host_id  |
   host_name              | neighbourhood_group |       neighbourhood        |      latitude      |     longitude      |    room_type     | price |  county  |    borough    | black | native | asian | income | incomepercap | poverty
 | employed | unemployment
-------------+----------+----------------------------------------------------------------------------------------------------------------+----------+----------
--------------------------+---------------------+----------------------------+--------------------+--------------------+------------------+-------+----------+---------------+-------+--------+-------+--------+--------------+--------
-+----------+--------------
 36047050400 |     2539 | Clean & quiet apt home by the park                                                                             |     2787 | John
                          | Brooklyn            | Kensington                 | 40.647490000000005 |          -73.97237 | Private room     |   149 | Kings    | Brooklyn      |  13.1 |      0 |  12.8 |  86771 |        46356 |    11.6
 |     2829 |          5.9
 36061008400 |     2595 | Skylit Midtown Castle                                                                                          |     2845 | Jennifer
                          | Manhattan           | Midtown                    |           40.75362 | -73.98376999999999 | Entire home/apt  |   225 | New York | Manhattan     |   2.3 |      0 |  41.2 | 120938 |        85826 |    15.4
 |     1309 |          7.3
 36061020800 |     3647 | THE VILLAGE OF HARLEM....NEW YORK !                                                                            |     4632 | Elisabeth
                          | Manhattan           | Harlem                     | 40.809020000000004 |           -73.9419 | Private room     |   150 | New York | Manhattan     |  53.3 |      0 |   7.2 |  42220 |        37019 |    25.6
 |     2695 |           11
 36047023100 |     3831 | Cozy Entire Floor of Brownstone                                                                                |     4869 | LisaRoxann
e                         | Brooklyn            | Clinton Hill               | 40.685140000000004 |          -73.95976 | Entire home/apt  |    89 | Kings    | Brooklyn      |  37.9 |      0 |   4.4 |  77039 |        36598 |      19
 |     2000 |         11.8
 36061018400 |     5022 | Entire Apt: Spacious Studio/Loft by central park                                                               |     7192 | Laura
                          | Manhattan           | East Harlem                |           40.79851 |          -73.94399 | Entire home/apt  |    80 | New York | Manhattan     |  35.7 |      0 |   6.2 |  25581 |        20311 |    33.2
 |     2946 |          9.6
 36061007800 |     5099 | Large Cozy 1 BR Apartment In Midtown East                                                                      |     7322 | Chris
                          | Manhattan           | Murray Hill                |           40.74767 |            -73.975 | Entire home/apt  |   200 | New York | Manhattan     |   3.5 |      0 |  18.8 | 120205 |        93503 |     7.6
 |     5491 |          6.9
 36047022900 |     5121 | BlissArtsSpace!                                                                                                |     7356 | Garon
                          | Brooklyn            | Bedford-Stuyvesant         |           40.68688 |          -73.95596 | Private room     |    60 | Kings    | Brooklyn      |    51 |      0 |   7.8 |  51645 |        32422 |    19.3
 |     2143 |            8
 36061013300 |     5178 | Large Furnished Room Near B'way                                                                                |     8967 | Shunichi
                          | Manhattan           | Hell's Kitchen             |           40.76489 |          -73.98493 | Private room     |    79 | New York | Manhattan     |     4 |    0.3 |    15 |  81974 |        64861 |    14.5
 |     3261 |          7.9
 36061019500 |     5203 | Cozy Clean Guest Room - Family Apt                                                                             |     7490 | MaryEllen
                          | Manhattan           | Upper West Side            |           40.80178 |          -73.96723 | Private room     |    79 | New York | Manhattan     |   1.7 |      0 |  12.3 | 117045 |        71348 |    12.3
 |     4306 |          6.6
 36061000600 |     5238 | Cute & Cozy Lower East Side 1 bdrm                                                                             |     7549 | Ben
                          | Manhattan           | Chinatown                  | 40.713440000000006 |          -73.99037 | Entire home/apt  |   150 | New York | Manhattan     |    12 |    0.4 |  50.6 |  19863 |        12802 |    44.7
 |     4028 |          8.7
 36061019500 |     5295 | Beautiful 1br on Upper West Side                                                                               |     7702 | Lena
                          | Manhattan           | Upper West Side            |           40.80316 |          -73.96545 | Entire home/apt  |   135 | New York | Manhattan     |   1.7 |      0 |  12.3 | 117045 |        71348 |    12.3
 |     4306 |          6.6
 36061012700 |     5441 | Central Manhattan/near Broadway                                                                                |     7989 | Kate
                          | Manhattan           | Hell's Kitchen             |           40.76076 |          -73.98867 | Private room     |    85 | New York | Manhattan     |   3.1 |      0 |  15.2 |  69578 |        57341 |    12.5
 |     5139 |          6.6
 36047013900 |     5803 | Lovely Room 1, Garden, Best Area, Legal rental                                                                 |     9744 | Laurie
                          | Brooklyn            | South Slope                |           40.66829 |          -73.98779 | Private room     |    89 | Kings    | Brooklyn      |   8.5 |      0 |   8.9 | 100787 |        57258 |    10.7
 |     2240 |          7.8
 36061018900 |     6021 | Wonderful Guest Bedroom in Manhattan for SINGLES                                                               |    11528 | Claudio
                          | Manhattan           | Upper West Side            |           40.79826 | -73.96113000000001 | Private room     |    85 | New York | Manhattan     |  23.1 |    0.2 |   4.6 |  28158 |        30439 |      36
 |     4805 |         16.5
 36061007300 |     6090 | West Village Nest - Superhost                                                                                  |    11975 | Alina
                          | Manhattan           | West Village               |            40.7353 |          -74.00525 | Entire home/apt  |   120 | New York | Manhattan     |   1.6 |      0 |   4.9 |  97149 |        99274 |       7
 |     4233 |          6.2
 36047052700 |     6848 | Only 2 stops to Manhattan studio                                                                               |    15991 | Allen & Ir
ina                       | Brooklyn            | Williamsburg               |           40.70837 |          -73.95352 | Entire home/apt  |   140 | Kings    | Brooklyn      |   3.6 |      0 |   3.9 |  32022 |        20287 |    38.3
 |     3120 |         14.1
 36047018300 |     7097 | Perfect for Your Parents + Garden                                                                              |    17571 | Jane
                          | Brooklyn            | Fort Greene                |           40.69169 |          -73.97185 | Entire home/apt  |   215 | Kings    | Brooklyn      |  19.6 |      0 |   6.7 |  93625 |        57525 |     7.7
 |     1766 |          9.1
```

Load

We created a SQL database called airbnb_db with two tables: airbnb_data and census_data. The schema for the database and tables is above. We used SQL because both data tables were very structured, and the result was conducive to structured queries of the joined data. To run an analysis, we simply perform a join on the Census Tract column. This allows us to run queries to find, for example, if correlations exist between the average cost of an Airbnb and the average income level for the same area, or an aggregation of average listing prices by census track.

Example Queries



```

SELECT * FROM airbnb_data;

SELECT * FROM census_data;

SELECT airbnb_data.censustract, airbnb_data.id, airbnb_data.name, airbnb_data.host_id, airbnb_data.host_name, airbnb_data.neighbourhood_group,
airbnb_data.neighbourhood, airbnb_data.latitude, airbnb_data.longitude, airbnb_data.room_type, airbnb_data.price, census_data.county,
census_data.borough, census_data.black, census_data.native, census_data.asian, census_data.income, census_data.incomepercap, census_data.poverty,
census_data.employed, census_data.unemployment
FROM airbnb_data
JOIN census_data
ON airbnb_data.censustract = census_data.censustract;


SELECT airbnb_data.censustract, airbnb_data.id, airbnb_data.name, airbnb_data.host_id, airbnb_data.host_name, airbnb_data.neighbourhood_group,
airbnb_data.neighbourhood, airbnb_data.latitude, airbnb_data.longitude, airbnb_data.room_type, airbnb_data.price, census_data.county,
census_data.borough, census_data.black, census_data.native, census_data.asian, census_data.income, census_data.incomepercap, census_data.poverty,
census_data.employed, census_data.unemployment
FROM airbnb_data
INNER JOIN census_data
ON airbnb_data.censustract = census_data.censustract;

SELECT airbnb_data.censustract, AVG (airbnb_data.price) AS "airbnb_avg_price", AVG(census_data.income) AS "average_income"
FROM airbnb_data
JOIN census_data
ON airbnb_data.censustract = census_data.censustract
GROUP BY airbnb_data.censustract
ORDER BY airbnb_data.censustract DESC;
```