

Towards an Automated Star Rating for Airbnb listings



Avikal Somvanshi

Fernando Melchor

Jonathan Pichot

HOTEL



Why?

- Help customers quickly assess a property's character
- Give Airbnb a new way to classify and sell 'experiences'

Data

Inside Airbnb: New York City Bramus

insideairbnb.com/index.html

Inside Airbnb: New York City
Adding data to the debate

Home About Behind Get the Data

[Like](#) [Share](#) 291 [Tweet](#) 177

27,392
out of 27,392 listings (100%)

Filter by: New York City

Listings per Host Only multi-listings

Some Airbnb hosts have multiple listings.
A host may list separate rooms in the same apartment, or multiple apartments or homes available in their entirety.

Hosts with multiple listings are more likely to be running a business without a license and not paying taxes, and if they are renting out an entire home or apartment and aren't present, are probably doing so **illegally**.

listings

29.3%
multi-listings
19,370 (70.7%)
single listings
8,022 (29.3%)
multi-listings

Show top hosts

Activity Only with recent and frequent reviews

Airbnb guests may leave a review after their stay, and these can be used as an indicator of airbnb

0.9
reviews/listing/month

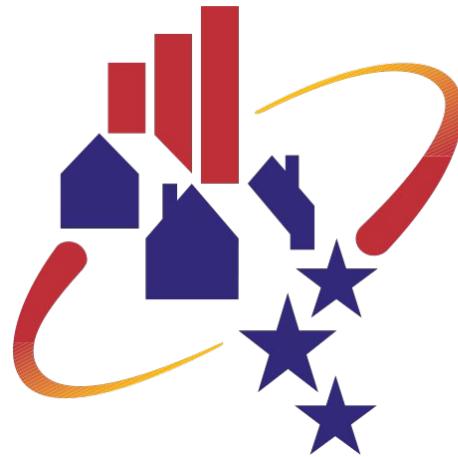
Leaflet | ©Mapbox | ©OpenStreetMap | Locations anonymized | Data: January 1-3 2015





New York City Subway





AMERICAN COMMUNITY SURVEY

U.S. CENSUS BUREAU



Feature Selection





Mapzen

10 minute walkshed

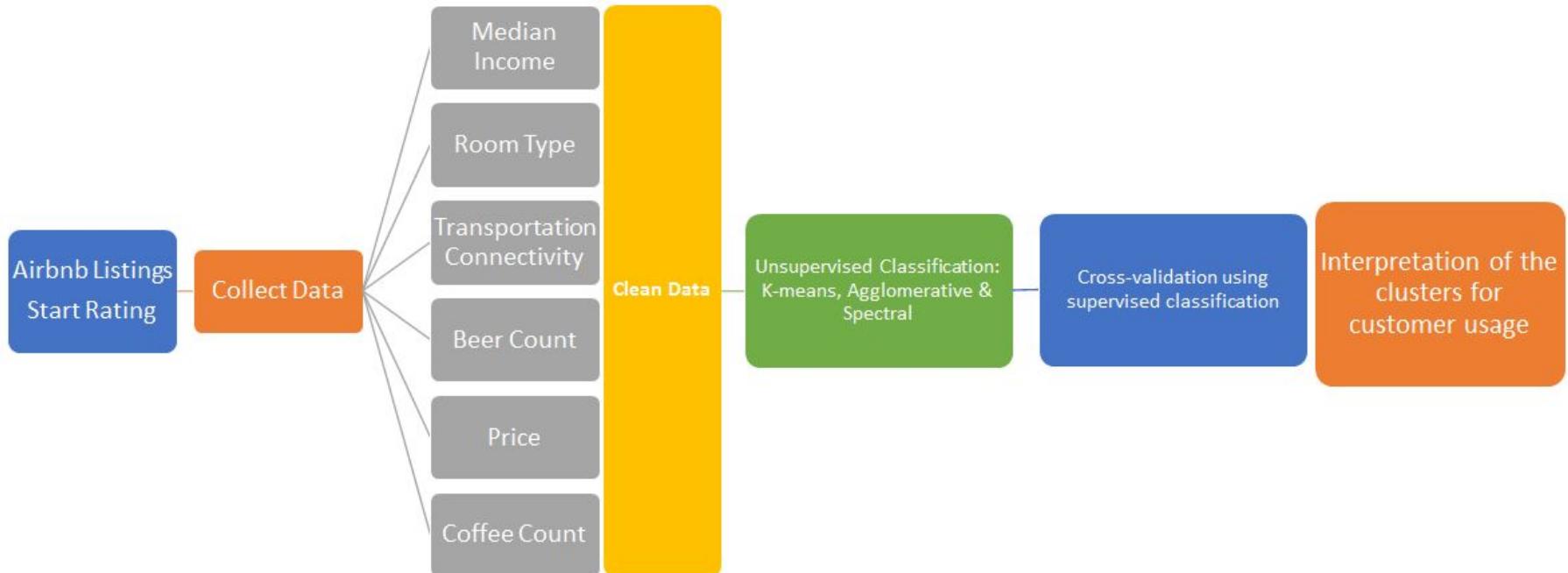


Subway Network

Using the average shortest path for each node we create the connectivity score for each listing. The score is assigned if the listing is within 10 minute walking distance from the station.



Methodology



Outliers

Removed listings with minimum stay > **7 days**

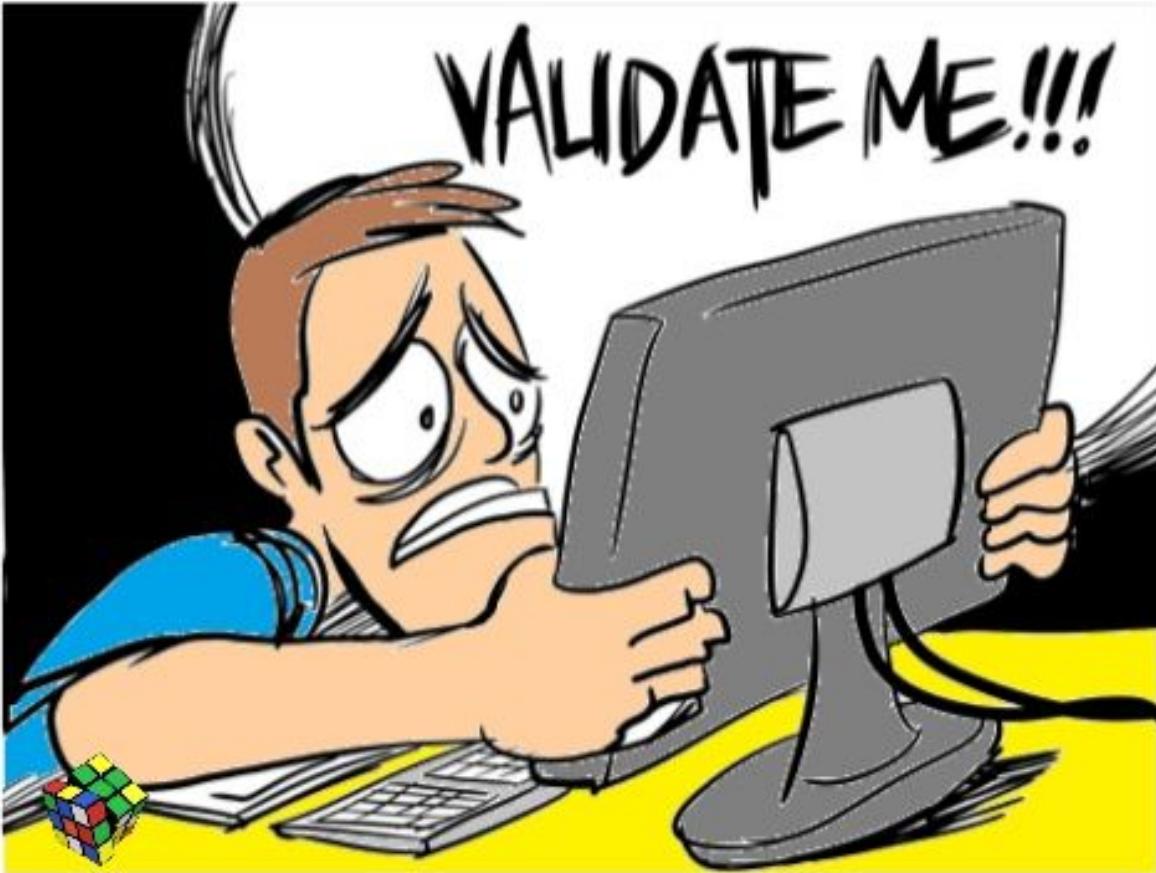
Removed listings with price > **\$500 a night**

Removed listings with **0 reviews**

Clustering Techniques

Kmeans

Agglomerative



Assessing the accuracy of the supervised classification techniques to predict the classes defined by K-means and Agglomerative. And using the results to decide which clustering is the best fit for our analysis.

Validation Results

(All Features)

Percent misclassified		
	Kmeans	Agglomerative
Decision Tree	0.00%	0.00%
Bagging	0.00%	0.00%
Adaboost	0.00%	0.00%
Gradient Boosting	0.00%	0.00%
Random Forest	0.0017%	0.0009%
Naïve Bayes	6.8755%	7.4523%

Validation Results

(Without Median Income)

Percent misclassified		
	Kmeans	Agglomerative
Decision Tree	25.38%	28.43%
Bagging	23.76%	27.10%
Adaboost	33.87%	36.27%
Gradient Boosting	22.35%	25.27%
Random Forest	24.45%	27.85%
Naïve Bayes	33.27%	36.35%

Clustering Techniques

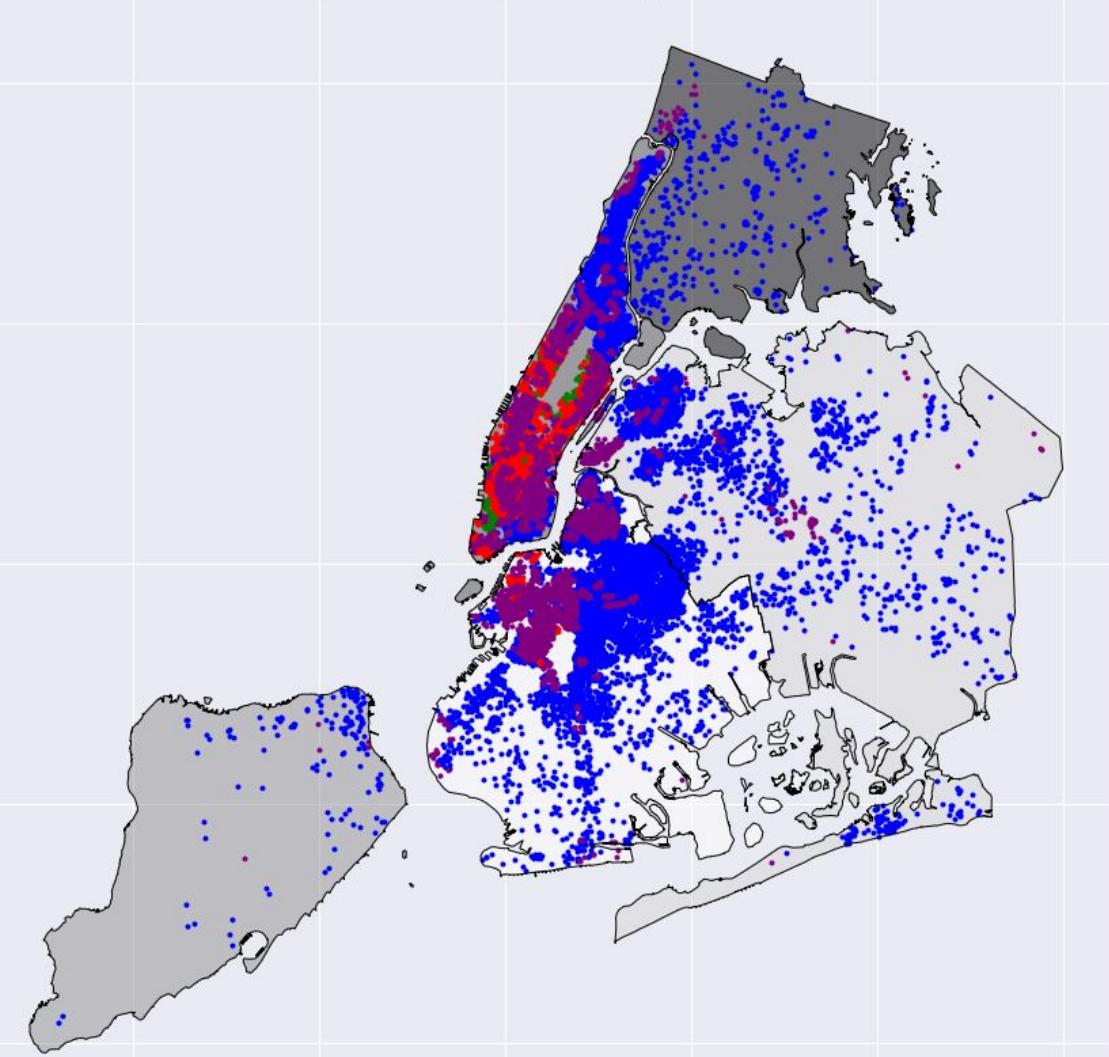
Kmeans

Agglomerative

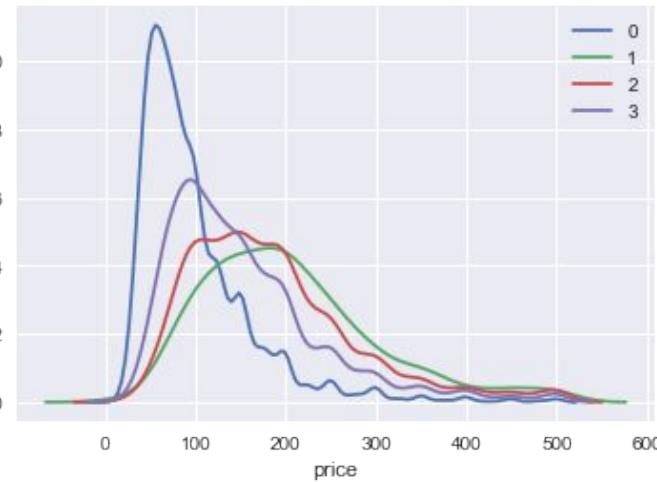
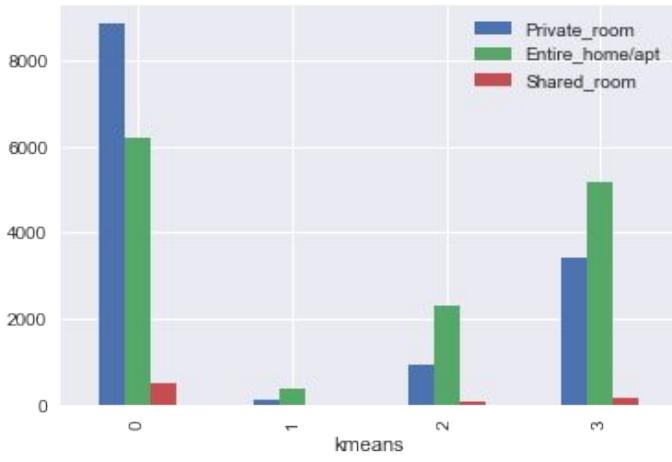
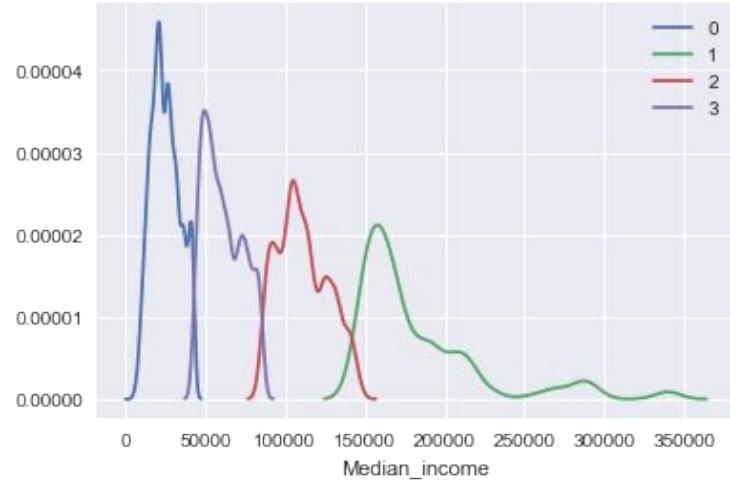
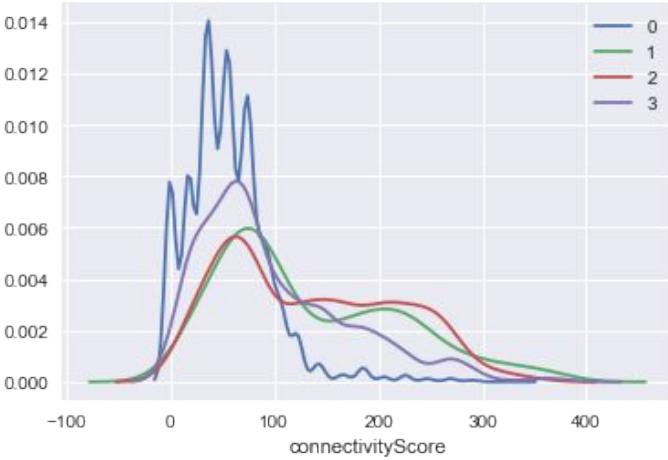


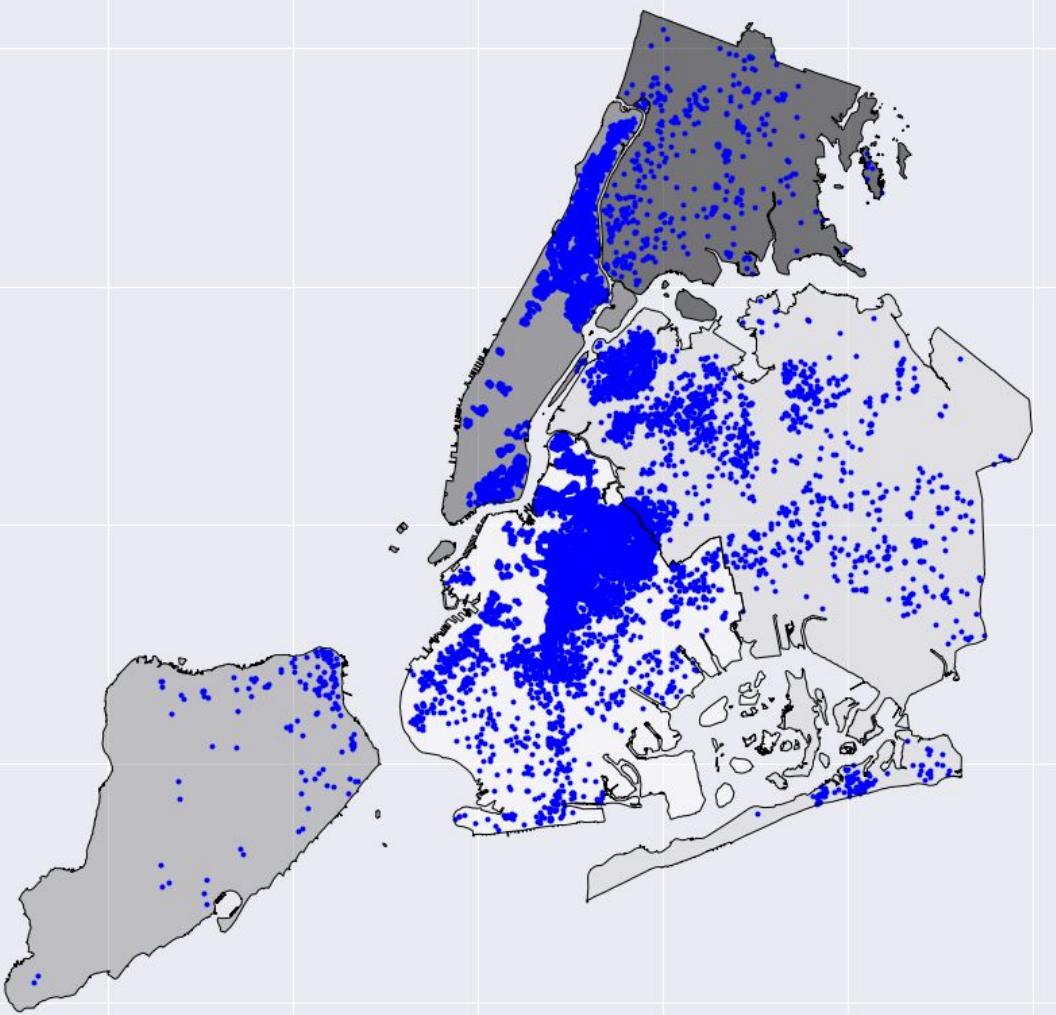
Inference

WTF!!!



- **Normal People**
- **The 2%**
- **Central Action**
- **Hip Kids**



A map of New York City with a grid overlay, showing the distribution of Airbnb listings. Darker blue dots represent higher concentrations of listings, primarily in Manhattan and Brooklyn.

Normal People

55% of listings

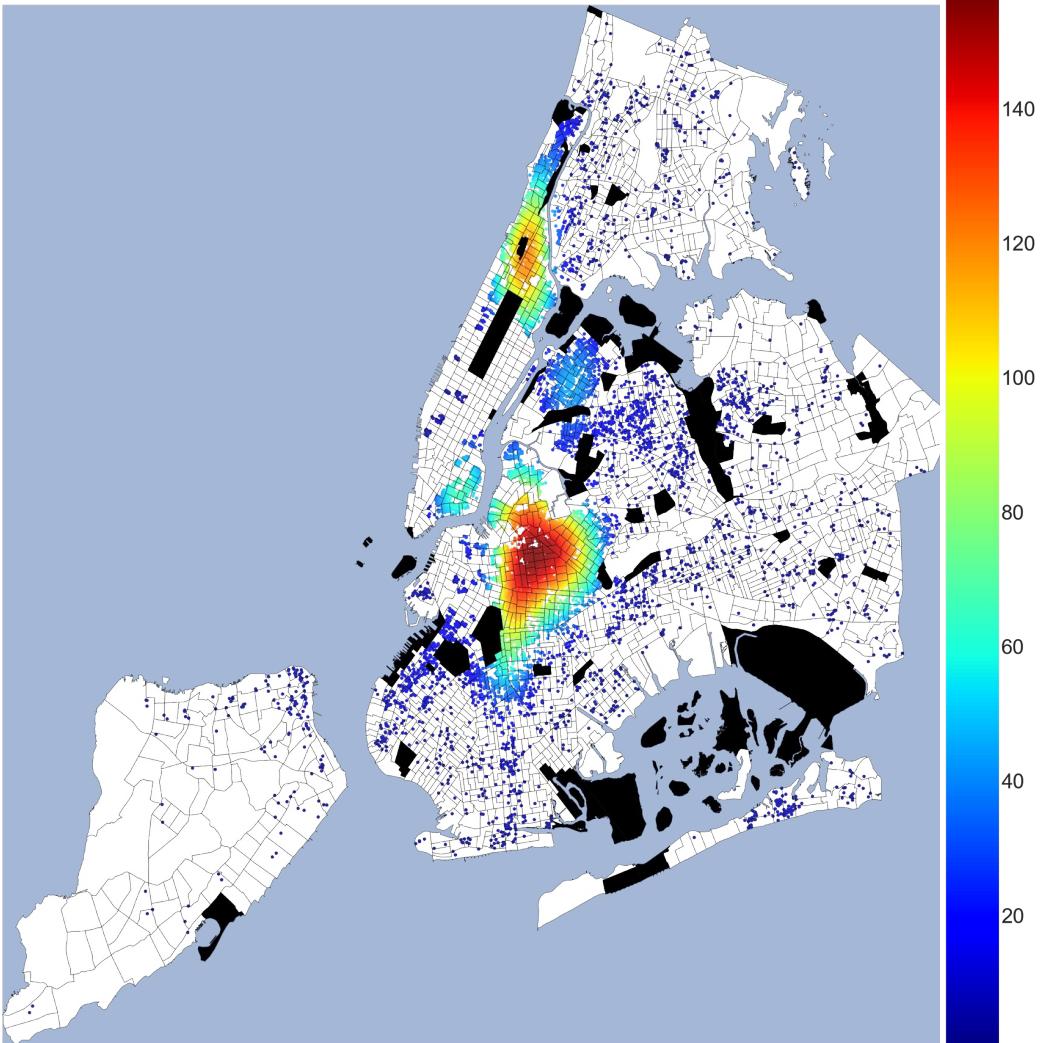
\$25,556 income (4th)
\$99 price (4th)

☕ - 4th

🍺 - 4th

🚌 - 4th

Largest share of
Shared Rooms



Normal People

55% of listings

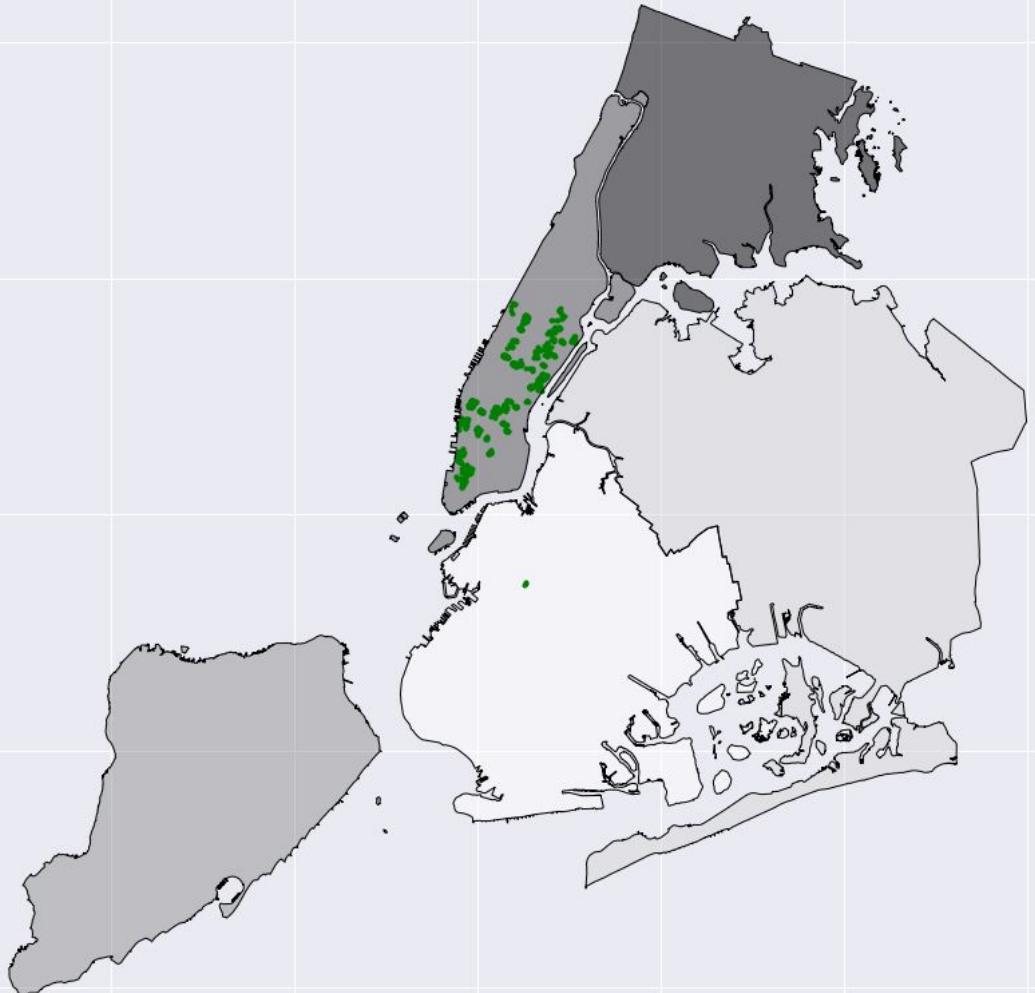
\$25,556 income (4th)
\$99 price (4th)

☕ - 4th

🍺 - 4th

🚌 - 4th

Largest share of
Shared Rooms



The 2%

2% of listings

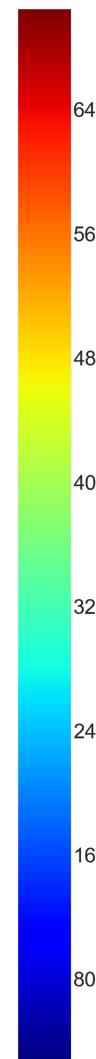
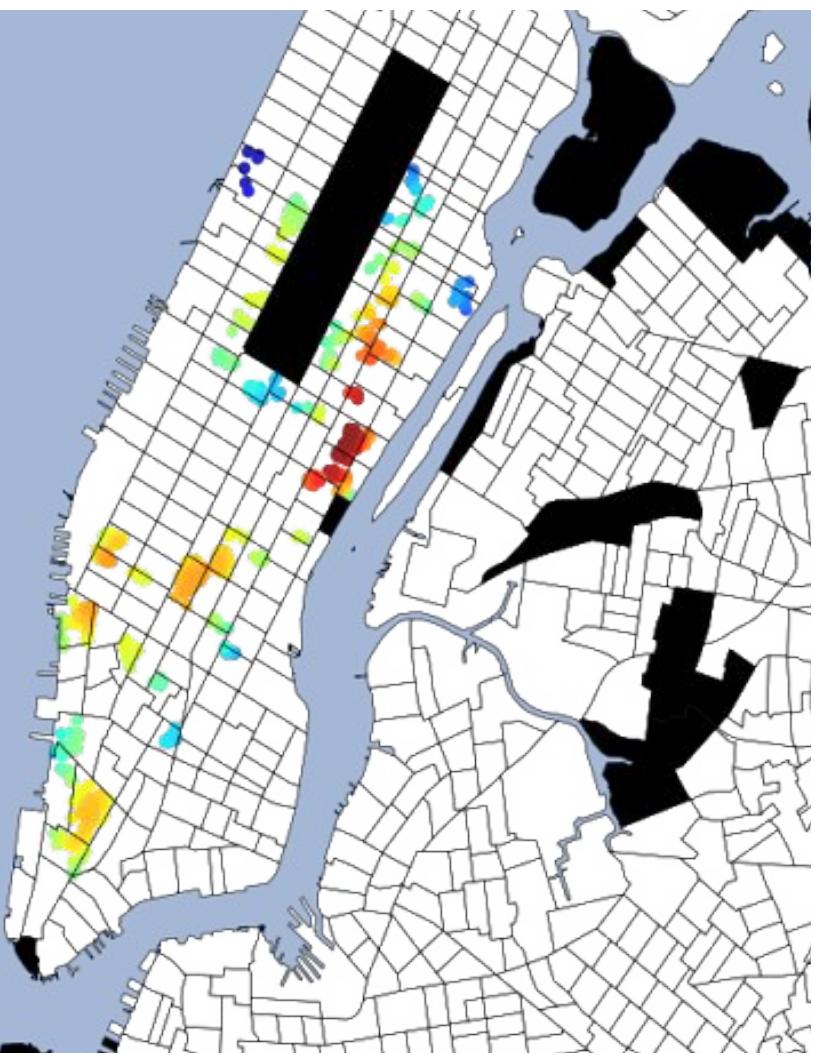
\$183,092 income (1st)
\$203 price (1st)

☕ - 2nd

🍺 - 3rd

🚌 - 2nd

Donald?



The 2%

2% of listings

\$183,092 income (1st)
\$203 price (1st)

- ☕ - 2nd
- 🍺 - 3rd
- 🚌 - 2nd

Donald?



Central Action

12% of listings

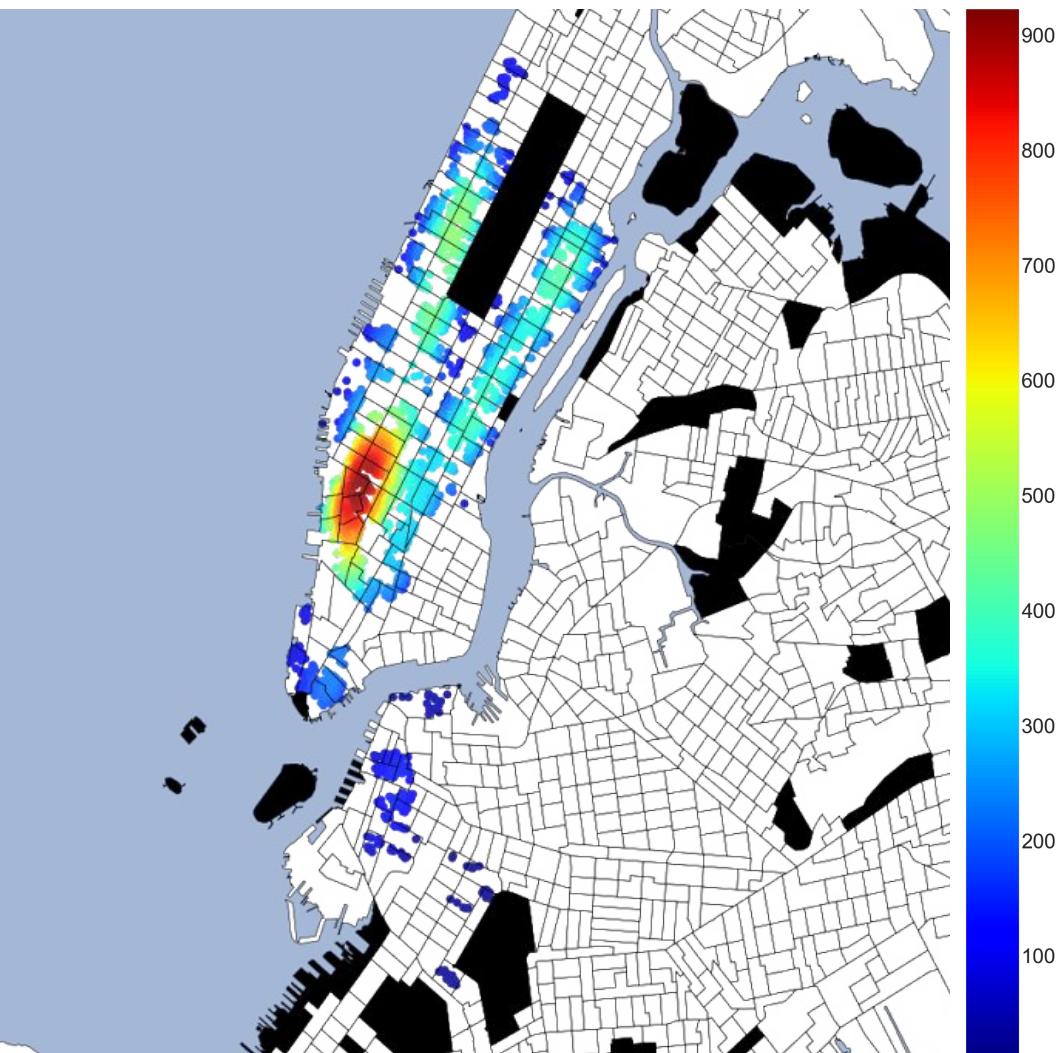
\$111,361 income (2nd)

\$183 price (2nd)

☕ - 1st

🍺 - 2nd

🚌 - 1st

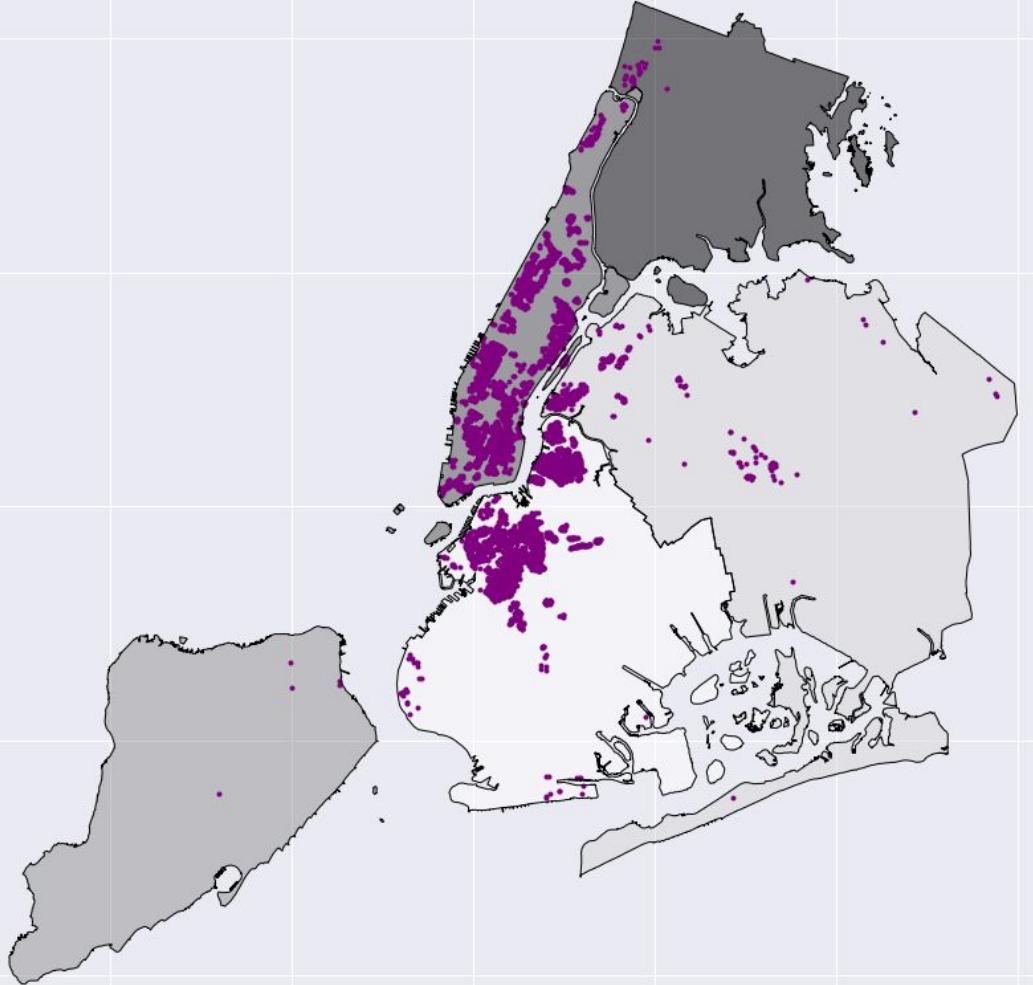


Central Action

12% of listings

\$111,361 income (2nd)
\$183 price (2nd)

- ☕ - 1st
- 🍺 - 2nd
- 🚌 - 1st



Hip Kids

31% of listings

\$61,686 income (3rd)

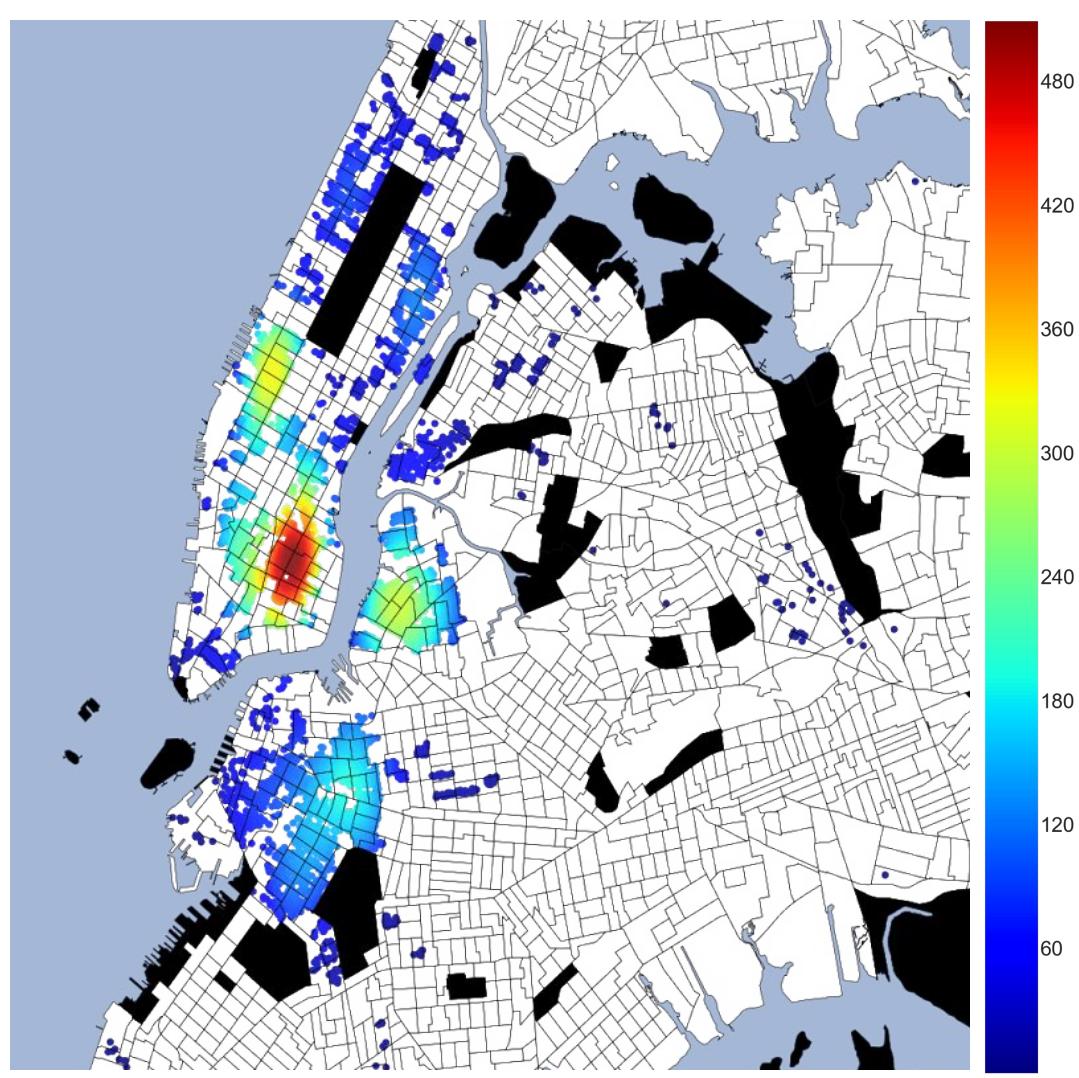
\$152 price (3rd)

☕ - 3rd

🍺 - 1st

🚌 - 3rd

Significantly higher 🍺 score than next closest



Hip Kids

31% of listings

\$61,686 income (3rd)
\$152 price (3rd)

☕ - 3rd

🍺 - 1st

🚌 - 3rd

Significantly higher 🍺
score than next closest

Everything is geography

