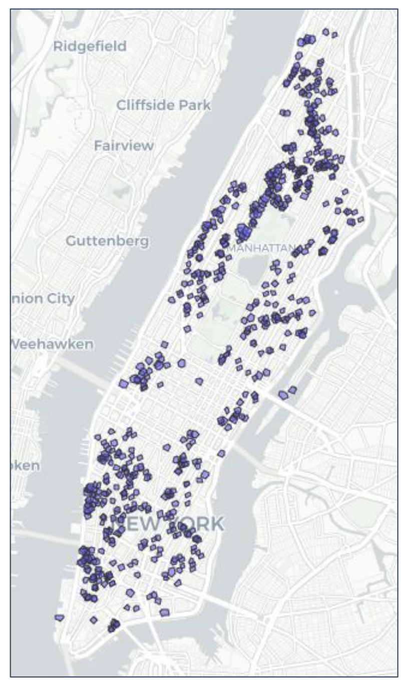
House Hunting *...made simpler*

Let's find your dream new place in **Manhattan, NY, USA**.

We used **advanced spatial analysis** (R) and **machine learning** techniques (Python).



Motivation

People shift to new houses all the time due to

1. Feng shui and Superstitions
2. $1M Lottery win!
3. Bored with current / annoying landlords
4. Job switch / promotion / lay off,
5. Family / new born child

People have different requirements like

1. Feng shui and Superstitions
2. Proximity to workplace / public transit
3. quiet, safe neighborhood
4. within specific communities
5. ocean view apartment / top floor of a high rise

It is difficult for a realtor to satisfy every needs of a client. We have built a system which accumulates data from a panoply of sources allowing simultaneous comparisons and generates meaningful insights.

Checkout an example conversation

Work and Results

We augmented basic features with those computed geospatially and generated an ML model (**accuracy of 88%**).

	Model	Precision	Recall	F1	Accuracy
0	Logit	0.795618	0.959184	0.878505	0.880734

Conclusion, Future Work

Given parameters, one can now predict whether a housing locality is affordable or not using our model. The geospatial analysis performed (point-polygon buffers, unions, intersections, joins) led us to many new features. We need to add more features to perform a comprehensive analysis and output holistic believable probabilities.

Datasets, References

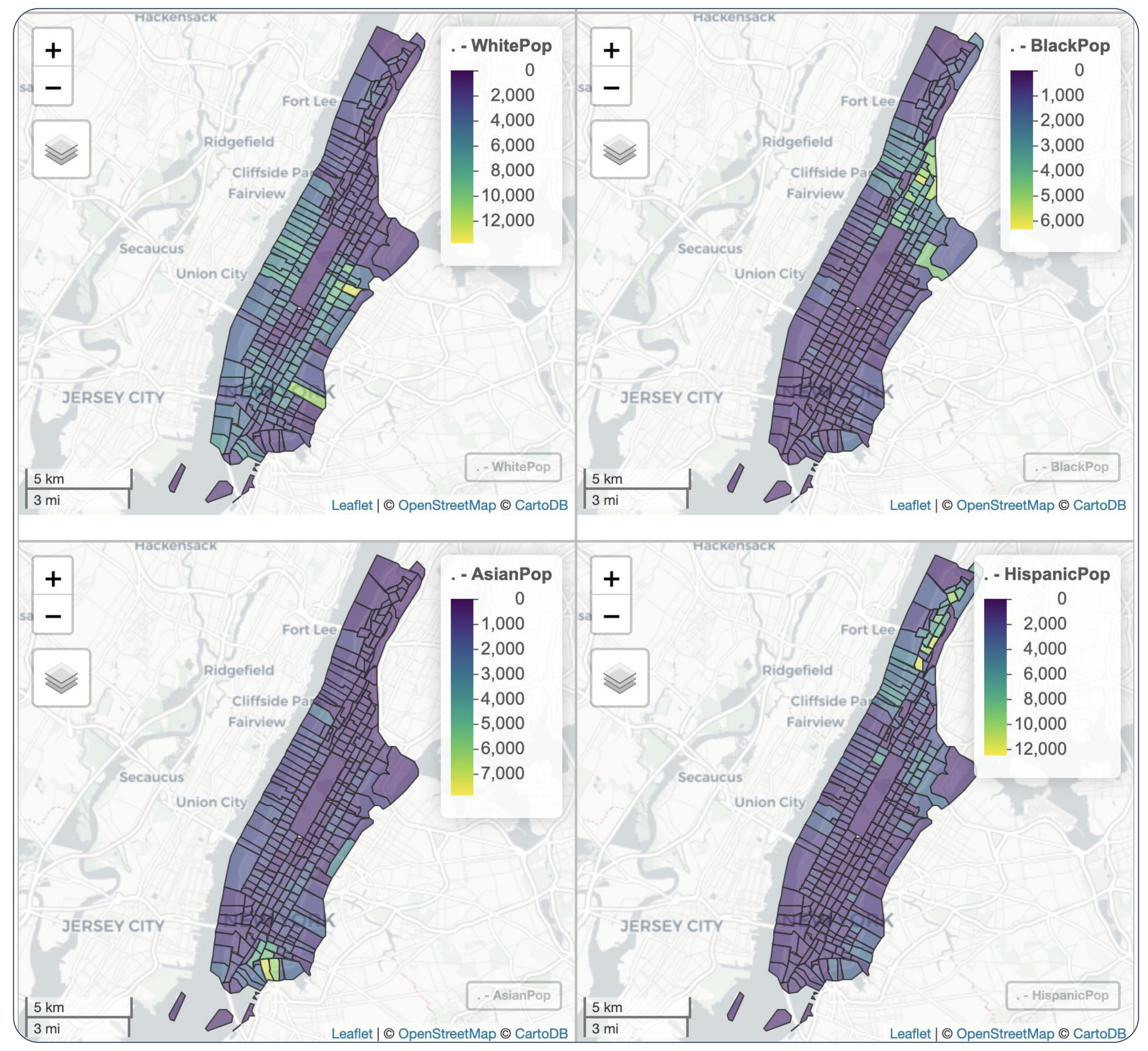
1. American Community Survey (ACS):
2. PLUTO (Properties, Tax lot data): https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page
3. Park Zone: https://data.cityofnewyork.us/City-Government/Parks-Zones
4. POIs: https://data.cityofnewyork.us/City-Government/Points-Of-Interest
5. Walk Score - https://www.walkscore.com
6. https://www.esri.com/about/newsroom/arcuser/house-hunting/
7. https://www.gislounge.com/gis-used-real-estate/

Arpit Narechania

MS CSE, 1st year
CP-6514 Advanced GIS,
Spring' 19

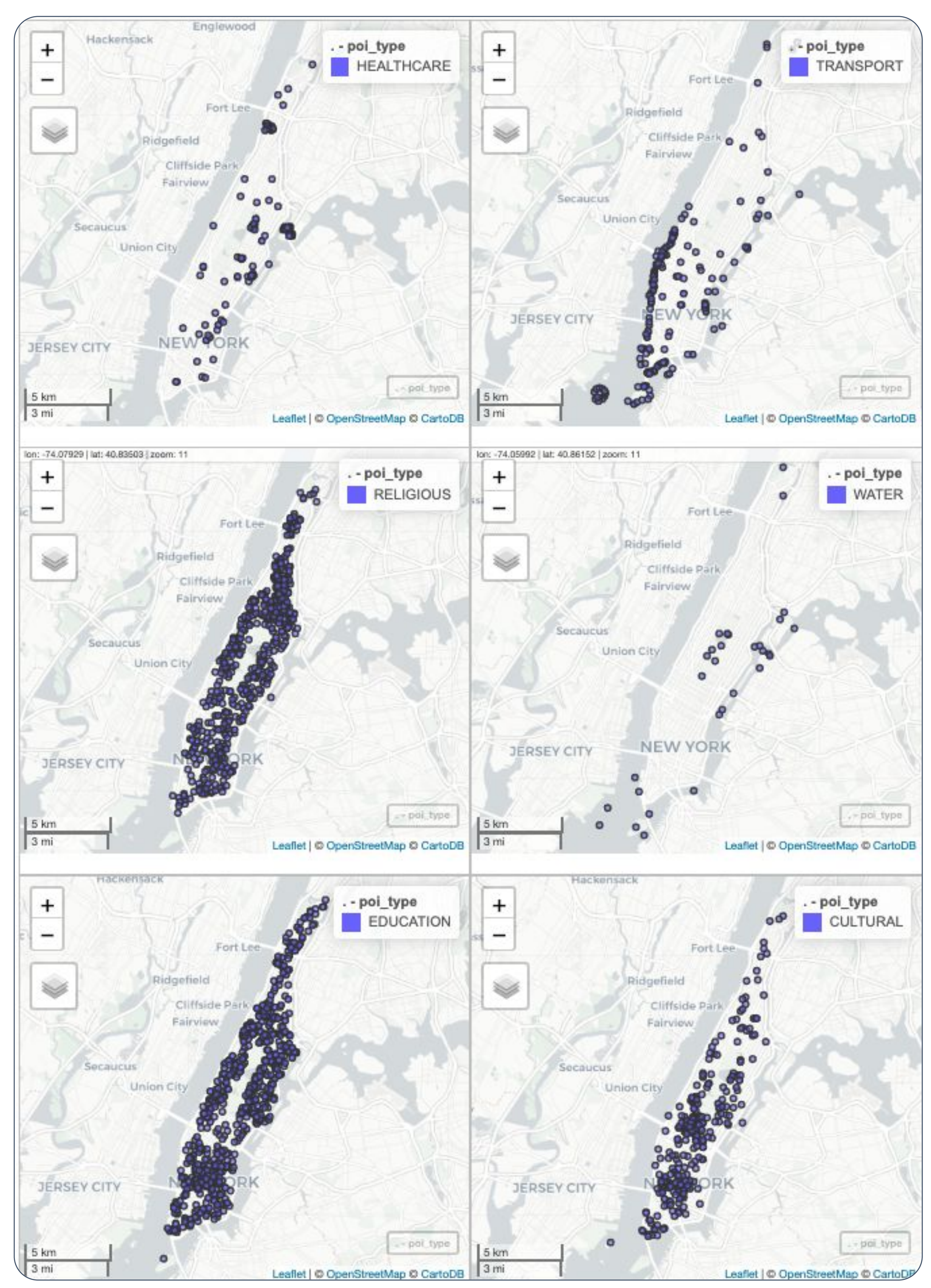
I'm shifting to New York and need to find an **affordable** place in a **community neighborhood**; it must be **close to my office** but also to places of leisure like **parks** and **cultural centers**. I have a small kid so will also need a **good school and health centers nearby** ...

Here's showing the **Demographics of New York by race (communities)**.



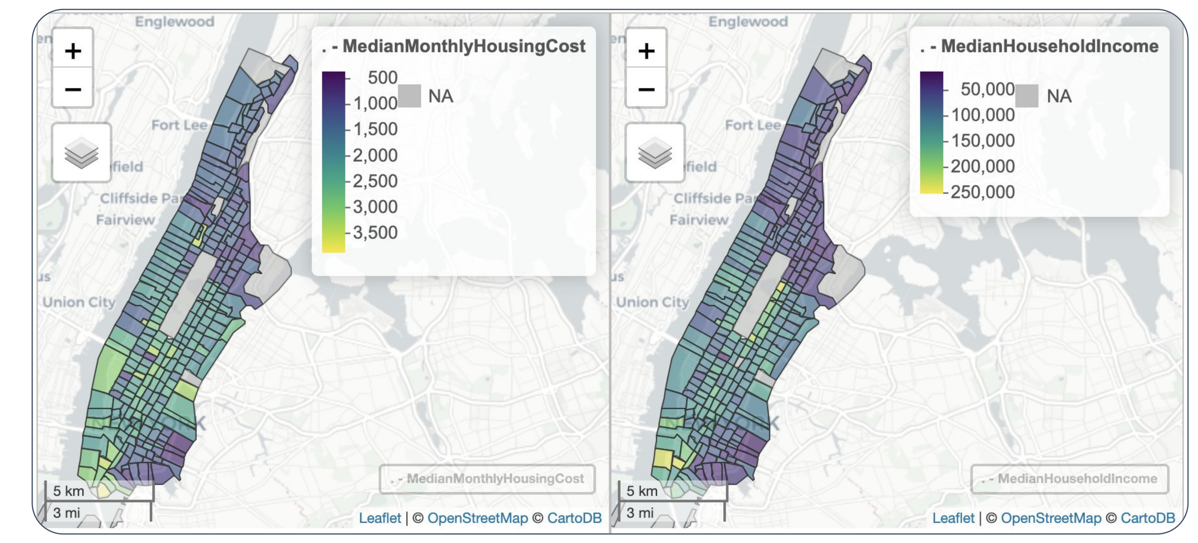
There is clear separation here! **Asians in abundance in SE Manhattan (Chinatown); Central Park separating the Whites from the Black (Harlem) and Hispanic**.

Places of Interest (POIs) scattered across Manhattan



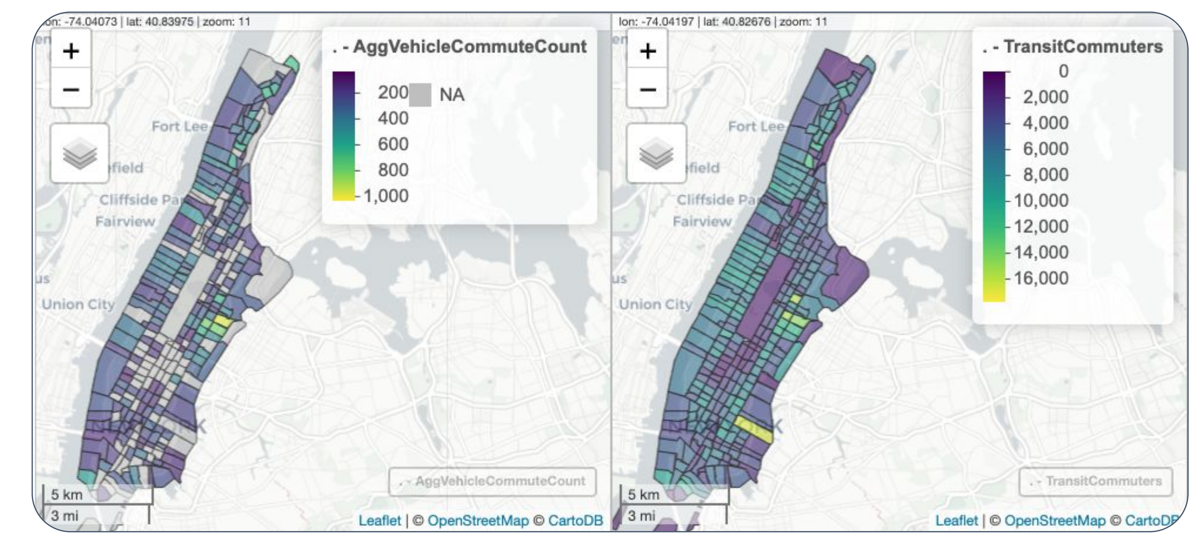
Transport, Healthcare and Cultural centers are abundant in downtown Manhattan; Education and Religious centers scattered across the city.

Here's showing the **Demographics of New York by ethnicity.**



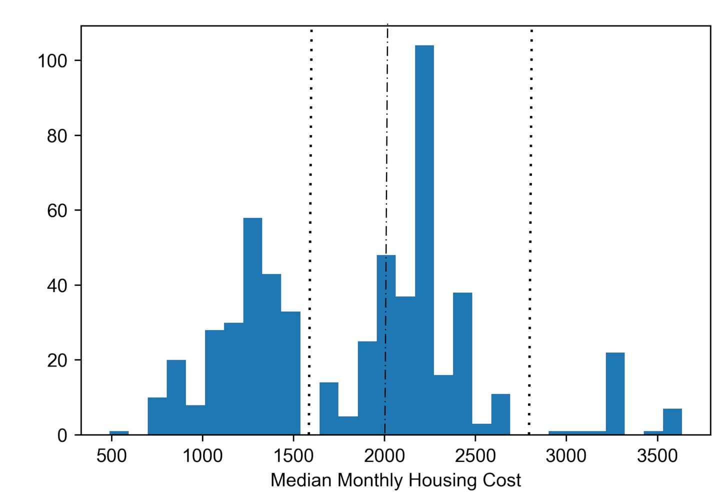
I expected correlation here **BUT** connecting this to the Demographics map earlier, it appears **Asians, Blacks, Hispanic communities are economically weaker than their White counterparts?**

Here's showing the **Commute Behavior** for New Yorkers!

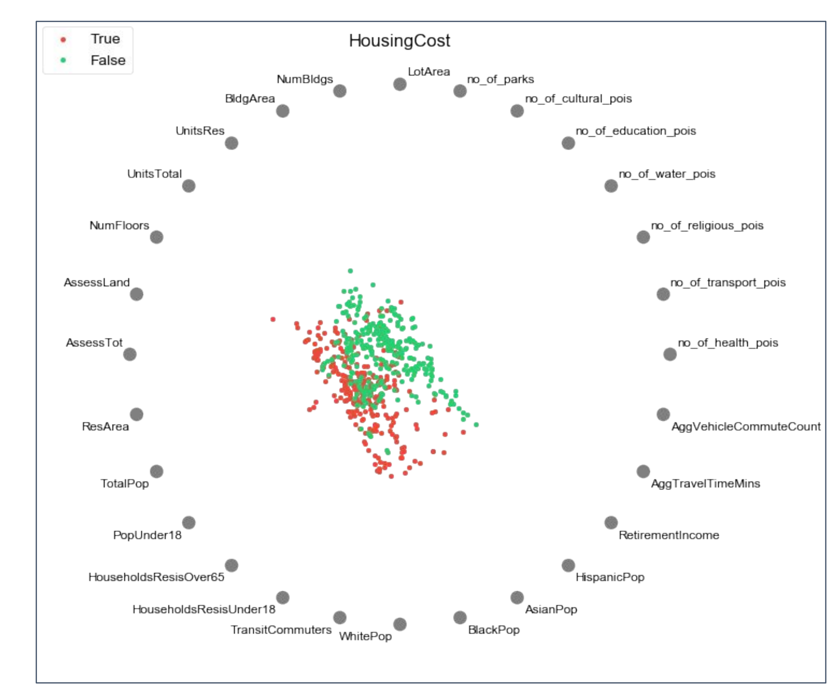


Along with the POI (transport) map, I **downtown Manhattan seems a commercial area** (**people generally take the public transit**); **Vehicle commute counts increase is we go towards north east.** Hmm...

Using an unsupervised ML technique, we came to a hypothesis that the **rich are driven by nearness to POIs, commute by vehicles** whereas the **poor are driven by the demographics, transit commute, housing variables**.



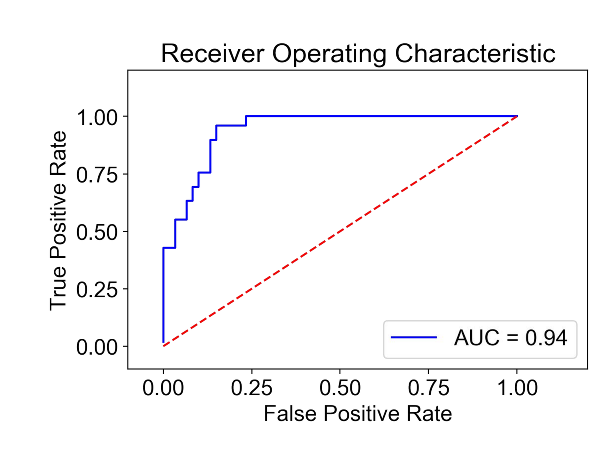
- **3 distinct economic classes** exist (dotted lines)
- For simplicity, the below definitions are used.
- Rich = [HousingCost = False, Median Housing cost > 2000]
- Poor = [HousingCost = True, Median Housing cost < 2000]



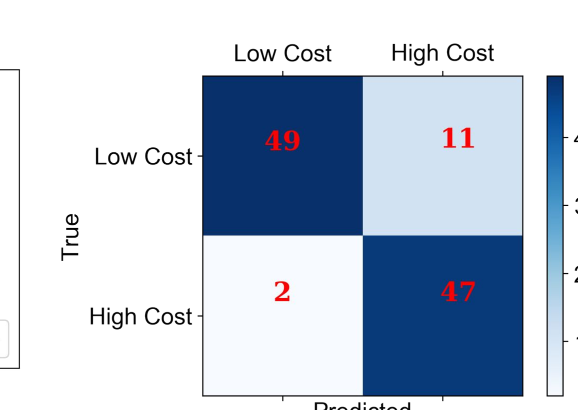
We built a supervised ML model, to predict "HousingCost" given the above set of parameters. **Low housing cost is indeed associated with Asians, area & no. of buildings in plot, cultural centers.**

Feature Name	Importance Score
no_of_cultural_pois	9.7052
AsianPop	8.6161
NumBldgs	5.3585
TotalPop	5.0086
LotArea	4.7814
no_of_water_pois	3.4913
LandUseCategory	3.2392
no_of_education_pois	3.0795
no_of_religious_pois	2.7027
BldgArea	2.4660

Feature Importance



ROC Curve



Confusion Matrix

You got an **accuracy of 89%,** impressive! **BUT**..can you add **more features** for me to consider? **Buying a house is a BIG DEAL for me.**