Prioritizing the Allocation of COVID-19 Resources to Vulnerable Neighborhoods

Based on Pre-Existing Health Conditions

Track E - Team 02

Tri, Tridib, Shaina, Kazeem

Methodology Overview

1. **Cluster Vulnerable Neighborhoods** based on pre-existing health conditions
2. **Identify Relationships** between clusters and COVID-19 cases per capita
3. **Identify Neighborhood Clusters** that should be the focus of greater testing/preventative public health measures.
4. **Determine High-Need Neighborhoods** based on the ratio between cases per capita and tests per capita.
5. **Prioritize** the allocation of COVID-19 resources.

Data: Pre-Existing Health Conditions

- **Dataset:** Environment & Health Data Portal
  - Run by the NYC Department of Health and Mental Hygiene (DOHMH)

- **Health Conditions:**
  - The following conditions were chosen because they have been demonstrated by the CDC to put people at higher risk for severe illness from COVID-19:
    - Heart Attack
    - Chronic obstructive pulmonary disease (COPD)
    - Asthma
    - Obesity

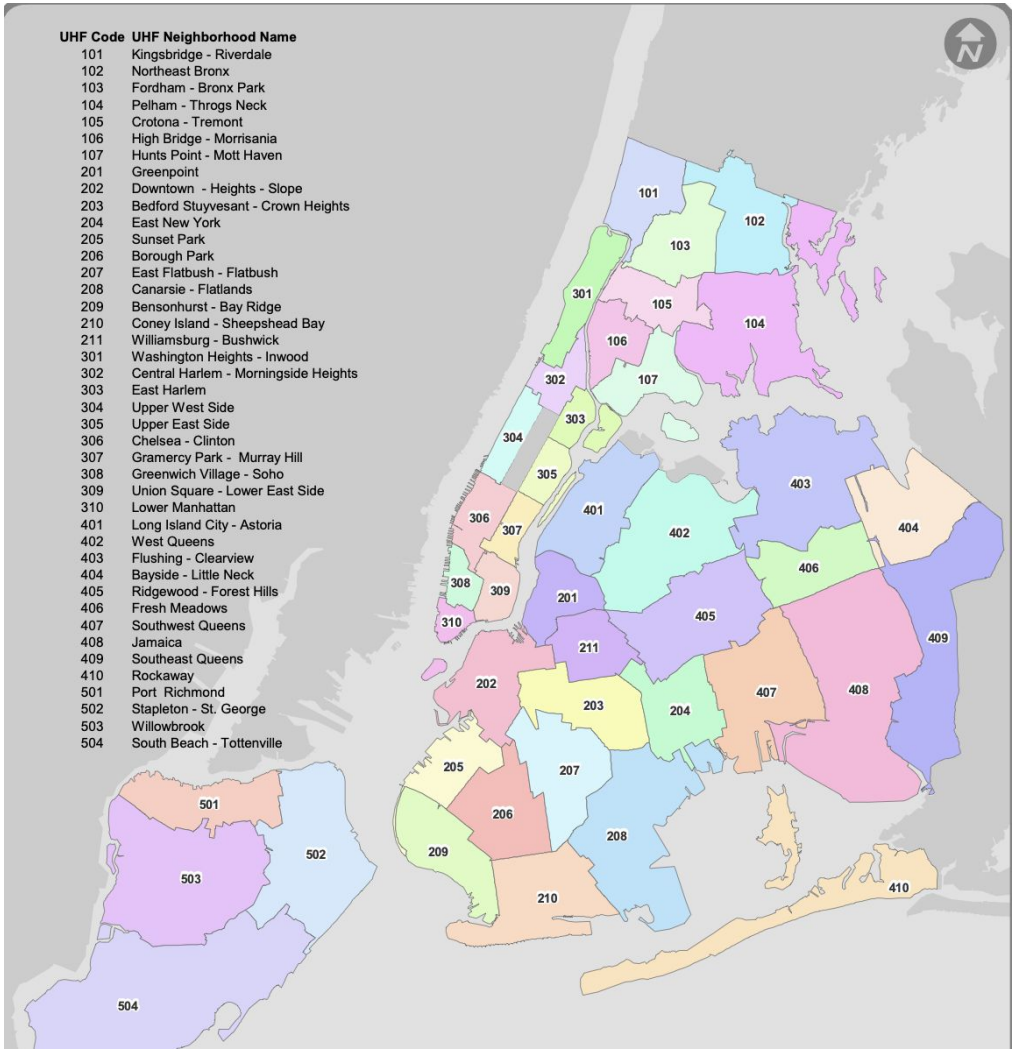
Reference:
https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html

Data: COVID-19

- **Dataset:** NYC Coronavirus (COVID-19) Data
  - Updated daily
  - Assembled by the NYC Department of Health and Mental Hygiene (DOHMH)

Data: Neighborhood Standardization

- Pre-Existing Health Condition dataset has neighborhood data (**UHF 34** or **UHF 42)**
- COVID-19 dataset has **ZIP code** data



MODZCTA Zip Codes ➔ UHF Neighborhoods

	MODZCTA	Positive	Total	zcta_cum.perc_pos	uhf34_neigh	uhf42_neigh
0	10001	321	1130	28.41	Chelsea-Village	Chelsea-Clinton
1	10002	931	2457	37.89	Union Square-Lower Manhattan	Union Square - Lower East Side
2	10003	414	1549	26.73	Union Square-Lower Manhattan	Union Square - Lower East Side
3	10004	28	108	25.93	Union Square-Lower Manhattan	Lower Manhattan
4	10005	54	256	21.09	Union Square-Lower Manhattan	Lower Manhattan

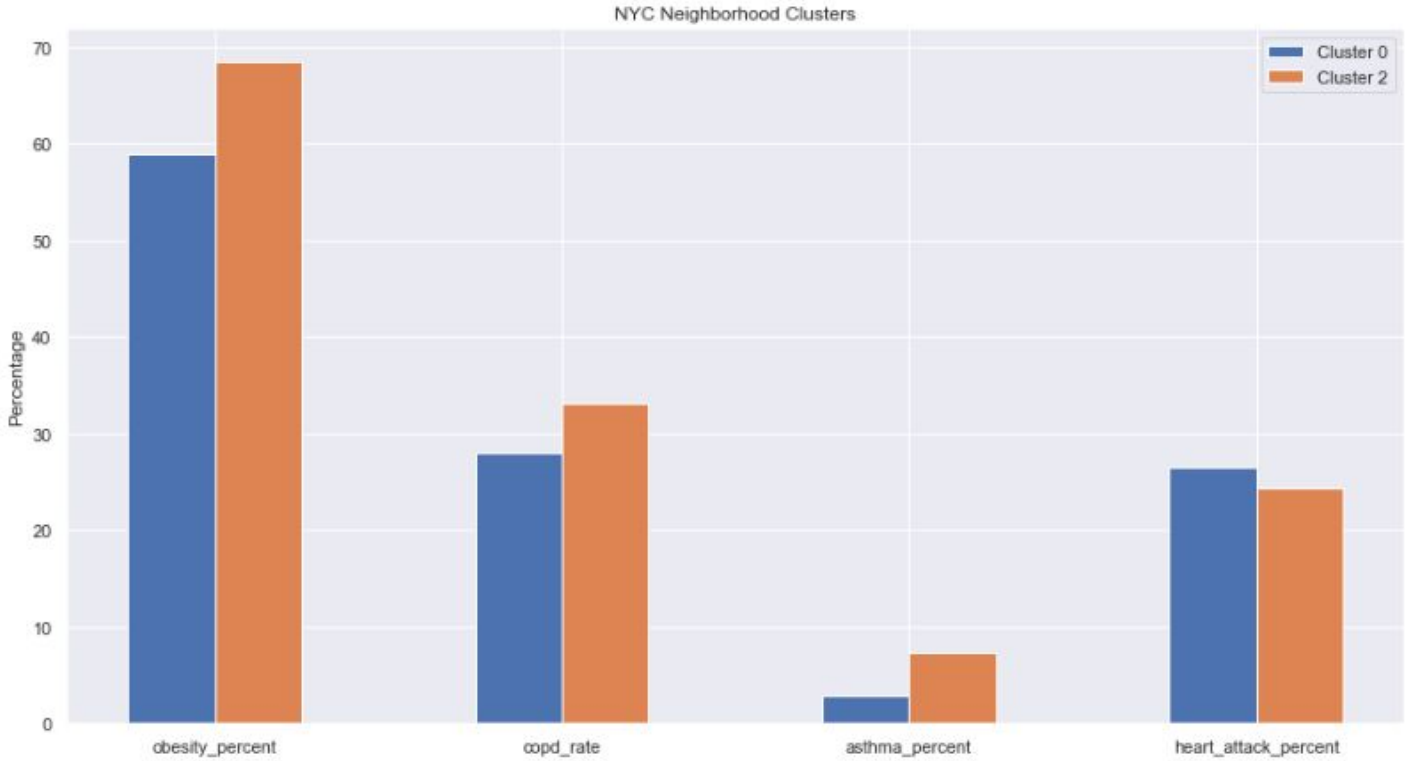
Aggregated by neighborhoods based on **AVERAGE** values

	borough	uhf34_neigh	uhf42_neigh	population	positive_rate	case_per_capita	test_per_capita	obesity_percent	copd_rate
0	Queens	Bayside Little Neck-Fresh Meadows	Bayside - Little Neck	87423	35.44	0.014	0.040	53.4	11.7
1	Queens	Bayside Little Neck-Fresh Meadows	Fresh Meadows	95537	39.06	0.022	0.055	53.4	12.0
2	Brooklyn	Bedford Stuyvesant - Crown Heights	Bedford Stuyvesant - Crown Heights	316269	39.09	0.017	0.044	67.2	21.4
3	Brooklyn	Bensonhurst - Bay Ridge	Bensonhurst - Bay Ridge	201541	37.38	0.015	0.040	50.3	40.8
4	Brooklyn	Borough Park	Borough Park	322563	38.58	0.025	0.065	56.3	15.5

Vulnerable Neighborhoods (K-Means Clustering)

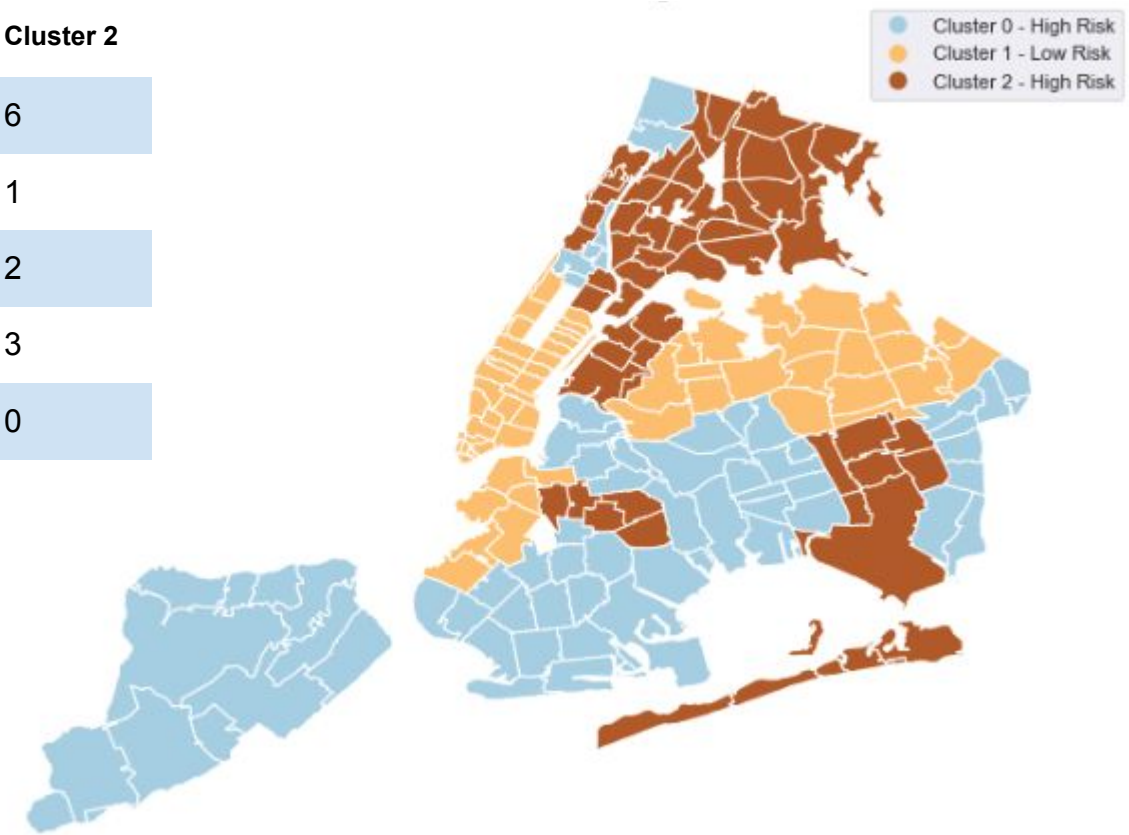
	Cluster 0	Cluster 1	Cluster 2
population	192680.235294	179754.461538	215331.250000
positive_rate	39.127059	33.063846	41.044167
case_per_capita	0.023176	0.013692	0.026167
test_per_capita	0.059294	0.040692	0.063417
obesity_percent	58.976471	43.592308	68.475000
copd_rate	27.947059	14.384615	33.083333
asthma_percent	2.777399	3.415385	7.333333
heart_attack_percent	26.523529	15.707692	24.366667

Most Vulnerables Neighborhoods



COVID-19 Cases & Community Health Conditions

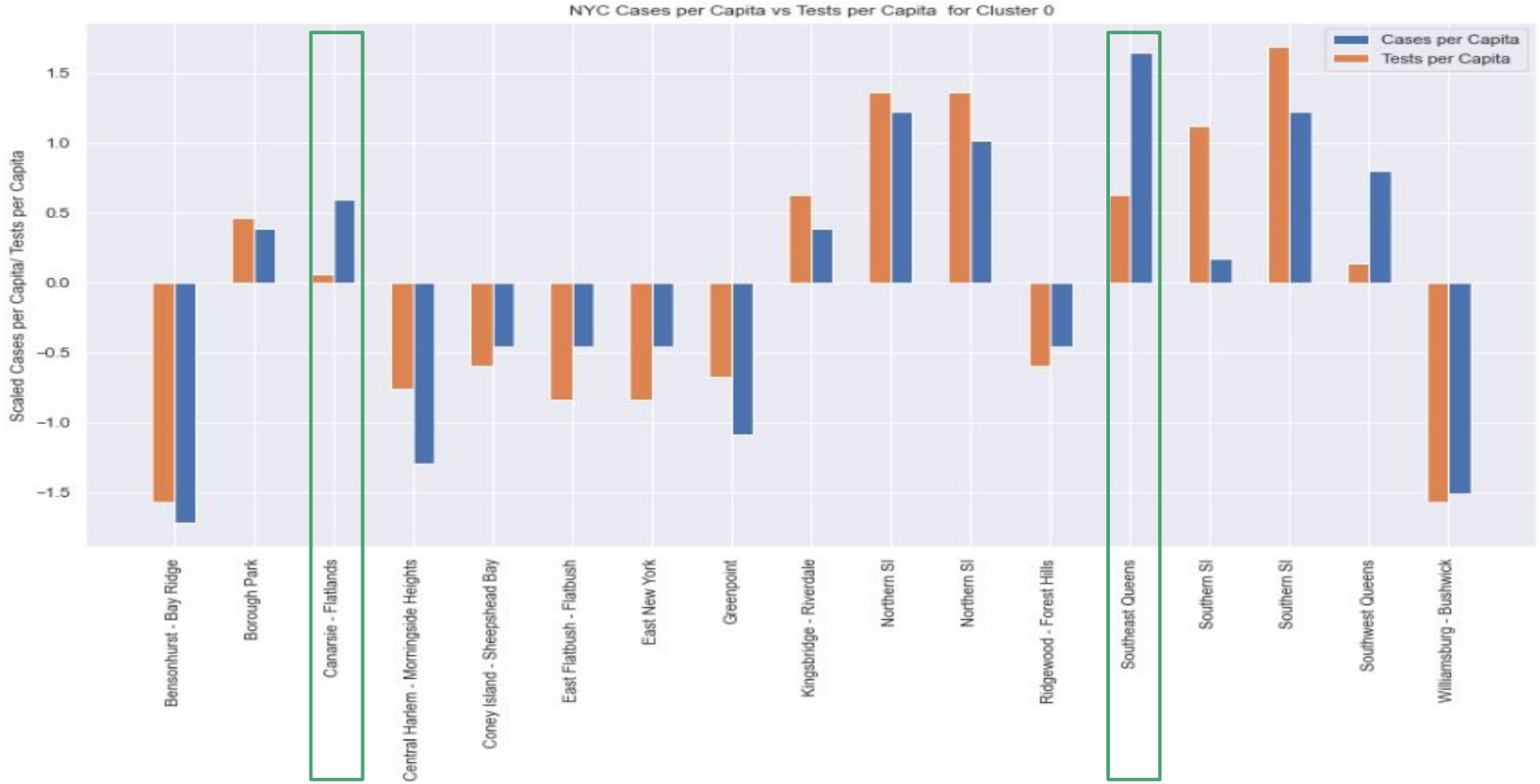
	Cluster 0	Cluster 1	Cluster 2
Bronx	1	0	6
Brooklyn	8	2	1
Manhattan	1	7	2
Queens	3	4	3
Staten Island	4	0	0



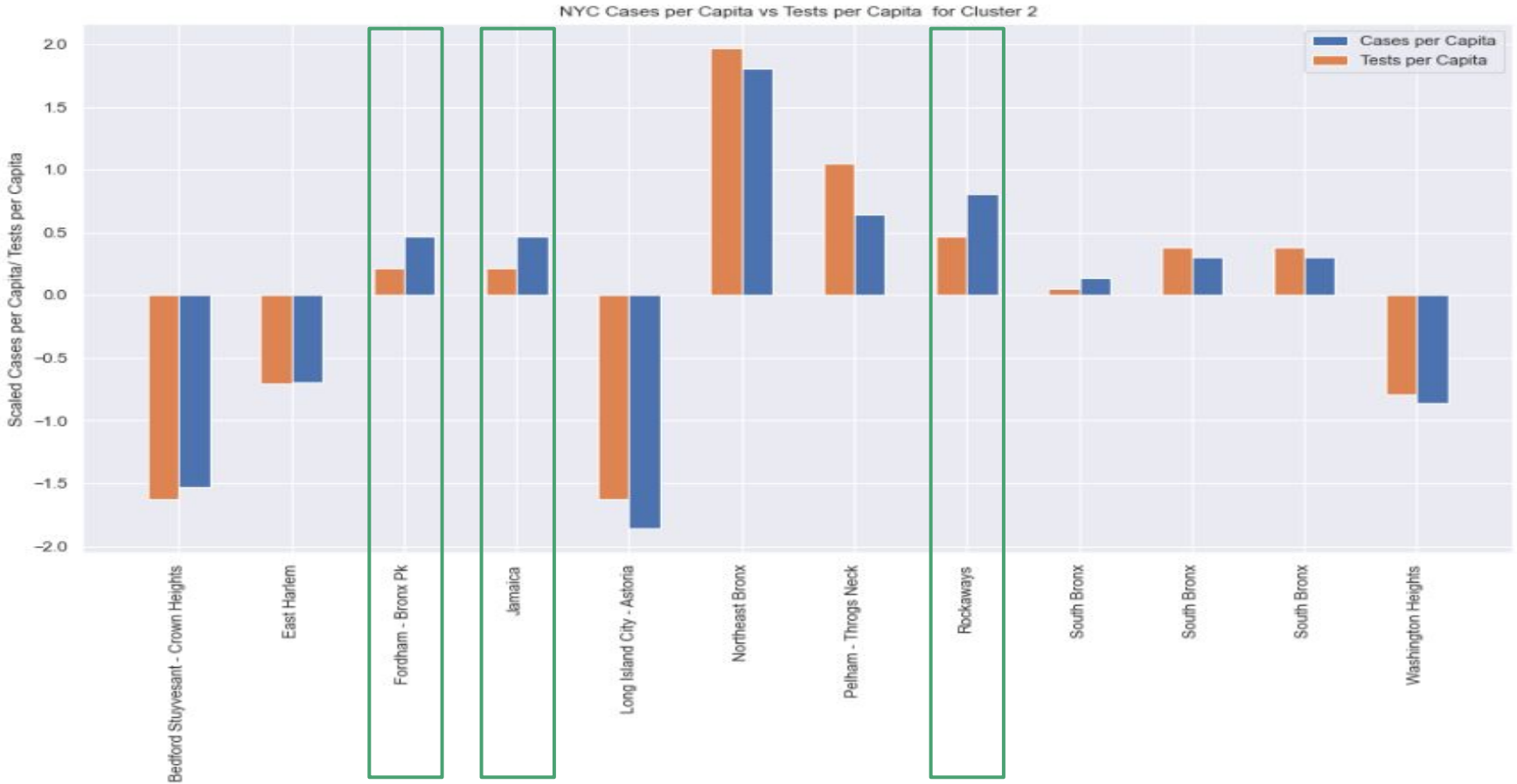
High-Need Neighborhoods

- With testing data on the basis of neighborhood we look at two features:
  - Number of Positive Cases
  - Number of Tests
- Neighborhood with **high** positive rates and **low** tests should potentially be the target of more testing.

Allocation of COVID-19 Resources - Cluster 0



Allocation of COVID-19 Resources - Cluster 2



Conclusion - Limitations

- Finding **datasets** with **consistent granularity**
- Finding a large number of **features** in **public datasets**

The above limitations can easily be overcome by **public health departments** with **more access to data**.

Conclusion - Next Steps

- **Build on our proof-of-concept** by evaluating **additional health conditions** and **additional cities**
- **Allocate resources** to **high risk areas**
- **Build regression models** that can potentially predict **COVID-19 case rate** based on underlying health conditions
- **Identify health factors** that drive **COVID-19 cases** in certain neighborhoods

GitHub Repo: https://github.com/nqtri/mit_covid19_challenge