Spread of COVID-19 in Mumbai

Balasubramanian Viswanathan

September 2020

Introduction

- Problem
  - Consistent growth in COVID-19 cases over the last many months in Mumbai, the commerical capital of India

- Goal
  - Arrive at observations and high level insights to fine tune the actions taken to combat the disease and help improve the Health and Socio-Economic welfare of the residents

- Objectives
  - Understand how the COVID-19 Positive cases are distributed across Mumbai
  - Understand the relationship between the distribution of Cases and
    - Geographical and administrative parameters
    - Population and related information
    - Location Data i.e. Venues of interest

Data Required for Analysis

- Population
  - Mumbai Wards & Districts: Population & Density by Sector 2001 from
    - http://www.demographia.com/db-mumbaidistr91.htm

- Location Data i.e. Venues of Interest
  - The Foursquare API to identify venues of interest from a given geographical location by querying https://api.foursquare.com

- Latitude and Longitude
  - OpenStreetMap for Geographical coordinates of a Ward and/or an Area

- COVID-19 Data
  - Municipal Corporation of Greater Mumbai's (**MCGM**) COVID-19 dashboard at
    - http://stopcoronavirus.mcgm.gov.in/assets/docs/Dashboard.pdf

Data Cleanup

- Removed initial rows of data scraped from www.demographia.com

- Columns renamed by using names extracted from the table

Ward	Area	Land Area (SKM)	Households	Population	Density per Square Kilometer	Density per Square Mile

- 'Density per Square Mile' is made redundant by 'Density per Square Kilometer'

- 'Households' is not used for modeling as it is directly related to 'Population'

- The 'Density per Square Kilometer' is not used for modeling as it is well represented by 'Population' and 'Land Area (SKM)' columns

- Ward level COVID-19 data is manually captured from COVID-19 Dashboard (rich pdf document) onto a spreadsheet and is read for further processing

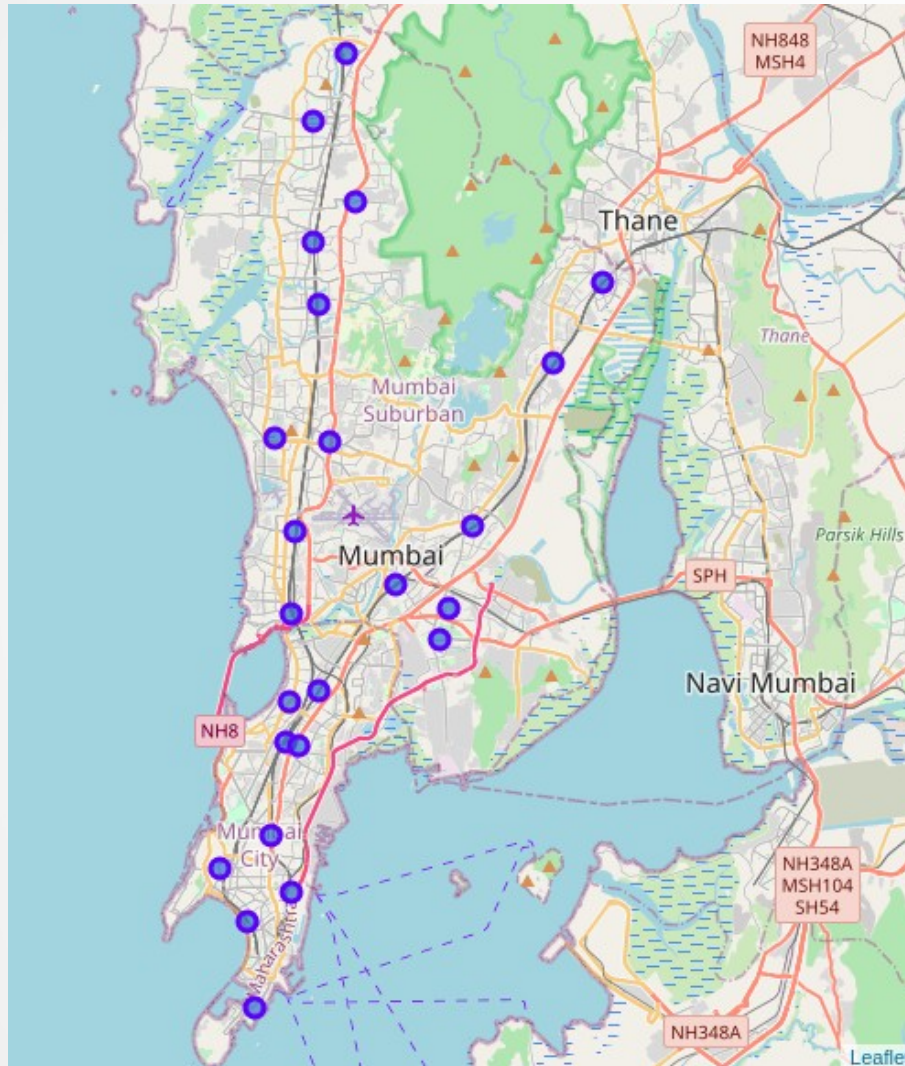- Retained Total Positive cases data while deaths & active cases are dropped / ignored

Ward	Total Positive

Approach

- Is similar to the exercise on Segmenting & Clustering Neighborhoods in NY City
- Data related to Mumbai like the Wards, Area (location within the Ward), Land Area in Square Kilo Meter and Population are gathered
- Latitude and Longitude of the Ward are looked up for the Wards by using the Nominatim library
- Wards visualized by using a Folium map to know more about their locations
- Foursquare APIs used to gather info on the venues of interest nearby a Ward by limiting the radius of the search as a function of the Land Area of the Ward
- Top 10 common category of venues identifed based on the Foursquare results the are merged and Wards split into clusters by using the KMeans algorithm
- Clusters are formed by pooling Wards with similar type of Venues.
- The number of clusters are determined by using the Elbow method and visualized by using a Folium map.
- Feature set enhanced by including Total Positive cases of COVID-19 in Mumbai, along with the Top common venues in each of the Wards
- Merged data is normalized using the Standard score method to ensure differences in the range of each of these features do not impact the model
- The resultant dataset is run by the KMeans algorithm to arrive at clusters of similar Wards with the Elbow method and visualized with a Folium map
- Graphs to visualize the similarities within a Cluster & differences between Clusters

Location of Wards in Mumbai

The Wards are spread primarily in the North to South line with an East to West fork
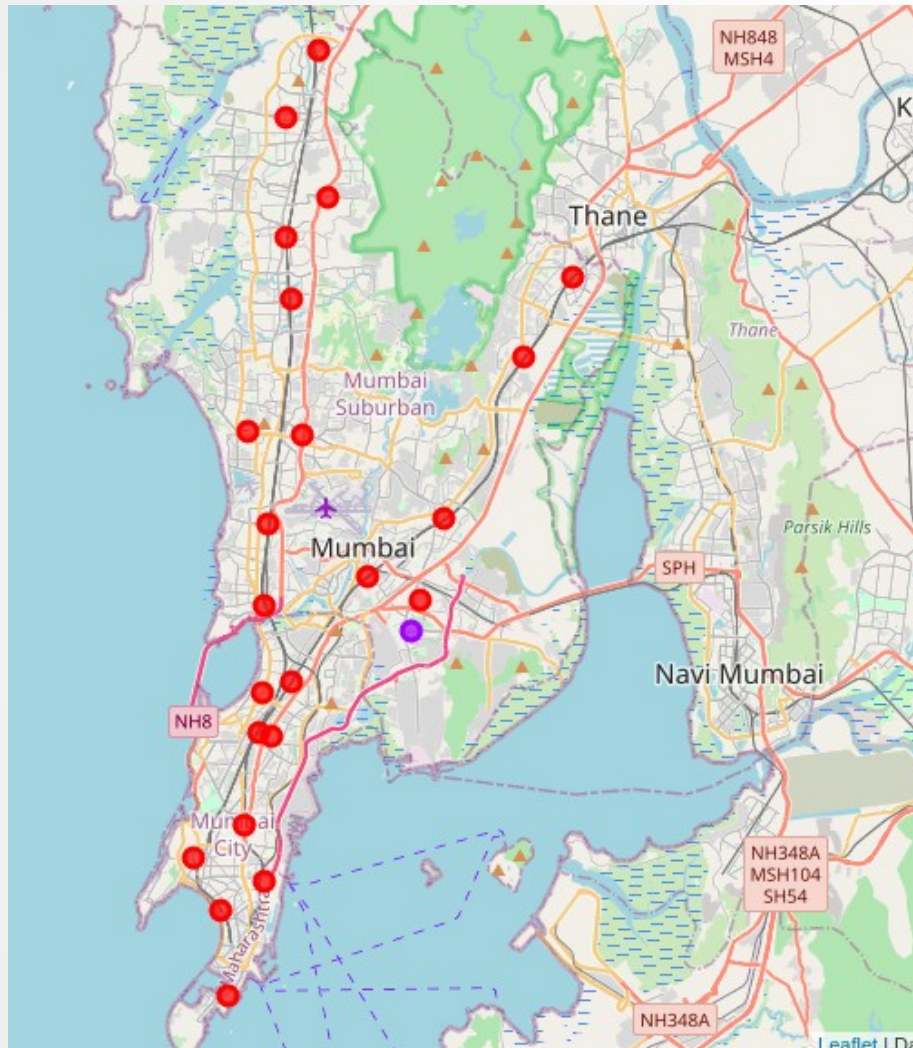

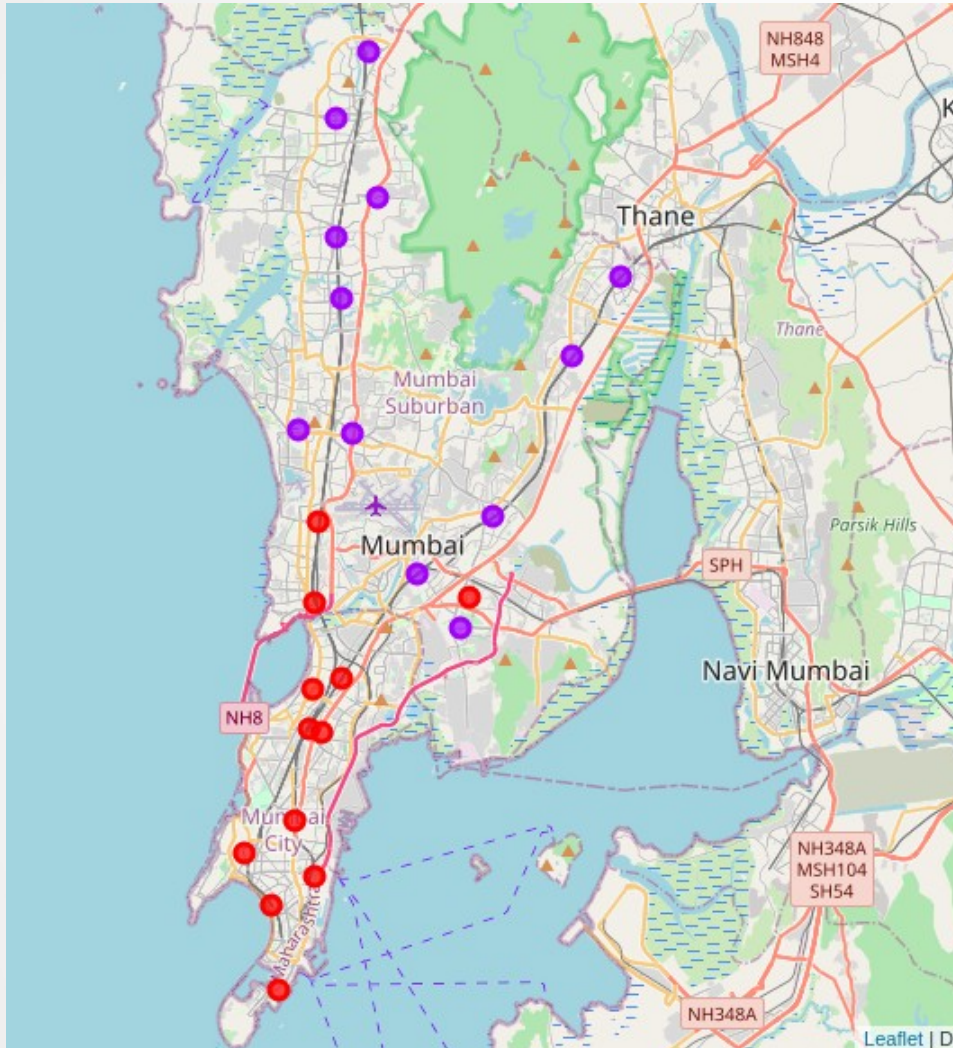|	Ward	Area	Land Area (SKM)	Population	Latitude	Longitude
0	A	Colaba	12.5	210847	18.915091	72.825969
1	B	Sanhurst Road	2.5	140633	18.956310	72.839838
2	C	Marine Lines	1.8	202922	18.945670	72.823781
3	D	Grant Road	6.6	382841	18.964447	72.813573
4	E	Byculla	7.4	440335	18.976622	72.832794
5	F South	Parel	14.0	396122	19.009482	72.837661
6	F North	Matunga	13.0	524393	19.027436	72.850147
7	G South	Elphinstone	10.0	457931	19.008427	72.842505
8	G North	Dadar/Plaza	9.1	582007	19.023767	72.839403
9	H East	Khar/Santacruz	13.5	580835	19.083966	72.841699
10	H West	Bandra	11.6	337391	19.054979	72.840220
11	K East	Andheri (East)	24.8	810002	19.115883	72.854202
12	K West	Andheri (West)	23.4	700680	19.117249	72.833968
13	L	Kurla	15.9	778218	19.065280	72.879380
14	M East	Chembur East	32.5	674850	19.046066	72.895473
15	M West	Chembur West	19.5	414040	19.057110	72.899388
16	N	Ghatkopar	26.0	619556	19.085954	72.908238
17	P South	Goregaon	24.4	437849	19.164803	72.850045
18	P North	Malad	19.1	796775	19.186719	72.848588
19	R South	Kandivalli	17.8	589886	19.201200	72.864332
20	R Central	Borivali West	50.0	513077	19.229456	72.847990
21	R North	Dahiser	18.0	363827	19.253479	72.860903
22	S	Bhandup	64.0	691227	19.143868	72.938433
23	T	Mulund	45.4	330195	19.172290	72.956469

Clusters of Wards based on Top 10 Venues

Cluster 0 has 23 of the 24 Wards that appear to be similar and hence pooled together
Cluster 1 has a mixed type of common venues including food, market & others

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A	Indian Restaurant	Café	Hotel	Bakery	Coffee Shop	Scenic Lookout	Cricket Ground	Italian Restaurant	Dessert Shop	Chinese Restaurant
1	B	Indian Restaurant	Dessert Shop	Ice Cream Shop	Café	Market	Restaurant	Chinese Restaurant	BBQ Joint	Garden	Cheese Shop
2	C	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Coffee Shop	Juice Bar	Gastropub	Pizza Place	Ice Cream Shop	Indie Movie Theater
3	D	Indian Restaurant	Restaurant	Bakery	Bar	Coffee Shop	Snack Place	Chinese Restaurant	Park	Fast Food Restaurant	Vegetarian / Vegan Restaurant
4	E	Indian Restaurant	Café	Lounge	Chinese Restaurant	Restaurant	Coffee Shop	Hotel	Bar	Italian Restaurant	Multiplex
5	F South	Indian Restaurant	Fast Food Restaurant	Ice Cream Shop	Café	Coffee Shop	Italian Restaurant	Snack Place	Dessert Shop	Juice Bar	Pizza Place
6	F North	Indian Restaurant	Chinese Restaurant	Movie Theater	Deli / Bodega	Café	Tea Room	Gym / Fitness Center	Breakfast Spot	Seafood Restaurant	Flower Shop
7	G South	Indian Restaurant	Chinese Restaurant	Harbor / Marina	Bakery	Flower Shop	Fast Food Restaurant	Miscellaneous Shop	Convention Center	Gym	Plaza
8	G North	Boat or Ferry	Vegetarian / Vegan Restaurant	Train Station	Pizza Place	Tea Room	Scenic Lookout	Café	Indian Restaurant	Donut Shop	Field
9	H East	Café	Seafood Restaurant	Lounge	Indian Restaurant	Ice Cream Shop	Dessert Shop	Coffee Shop	Bar	Bakery	Chinese Restaurant
10	H West	Lounge	Italian Restaurant	Indian Restaurant	Bar	Hotel	Scenic Lookout	Bagel Shop	Jewelry Store	Brewery	Brazilian Restaurant
11	K East	Indian Restaurant	Hotel	Italian Restaurant	Coffee Shop	Café	Lounge	Pizza Place	Multiplex	Brewery	Restaurant
12	K West	Indian Restaurant	Gym	Italian Restaurant	Pub	Beach	Bakery	Mexican Restaurant	Ice Cream Shop	Chinese Restaurant	Donut Shop
13	L	Indian Restaurant	Restaurant	Lounge	Café	Chinese Restaurant	Ice Cream Shop	Dessert Shop	Coffee Shop	Snack Place	Gym / Fitness Center
15	M West	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Juice Bar	Movie Theater	Plaza	Seafood Restaurant	Vegetarian / Vegan Restaurant	Deli / Bodega	Field
16	N	Indian Restaurant	Lounge	Italian Restaurant	Coffee Shop	Diner	Afghan Restaurant	Pub	Sandwich Place	Fast Food Restaurant	Brewery
17	P South	Indian Restaurant	Multiplex	Donut Shop	Chinese Restaurant	Shopping Mall	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Coffee Shop	Clothing Store	Breakfast Spot
18	P North	Fast Food Restaurant	Gym	Coffee Shop	Bookstore	Indian Restaurant	Seafood Restaurant	Multiplex	Pizza Place	Bar	Women's Store
19	R South	Ice Cream Shop	Fast Food Restaurant	Pizza Place	Café	Indian Restaurant	Sandwich Place	Electronics Store	Restaurant	Lounge	Historic Site
20	R Central	Indian Restaurant	Chinese Restaurant	Juice Bar	Water Park	Restaurant	Sandwich Place	Theme Park	BBQ Joint	Seafood Restaurant	Dessert Shop
21	R North	Café	Chinese Restaurant	Indian Restaurant	Fast Food Restaurant	Restaurant	Department Store	Gym	Ice Cream Shop	Coffee Shop	Gym / Fitness Center
22	S	Indian Restaurant	Fast Food Restaurant	Snack Place	Coffee Shop	Restaurant	Pub	Café	Gym	Dessert Shop	Italian Restaurant
23	T	Café	Indian Restaurant	Golf Course	Asian Restaurant	Scenic Lookout	Residential Building (Apartment / Condo)	Chinese Restaurant	Fast Food Restaurant	Dessert Shop	Pizza Place

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	M East	Chinese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Food	Women's Store	Flea Market	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant

Clusters of Wards based on multiple features

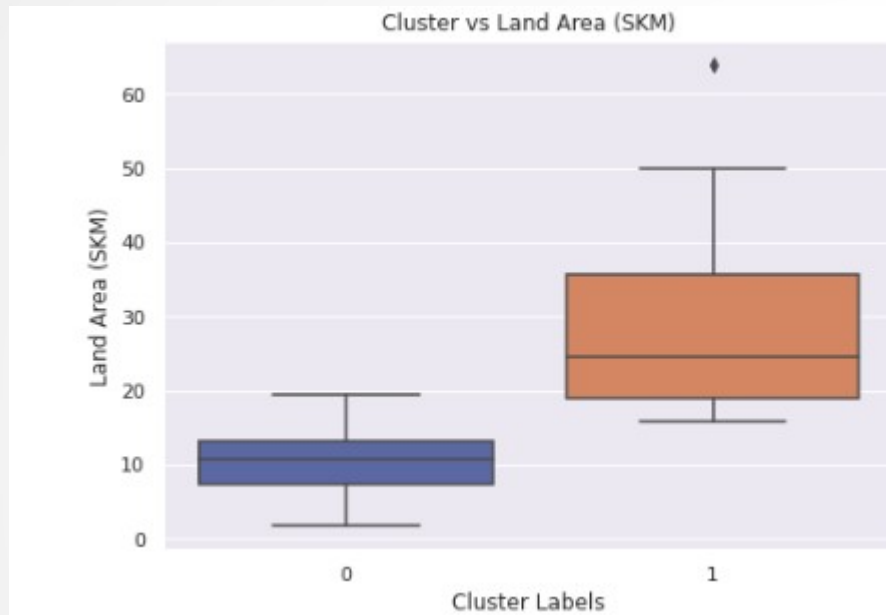Wards have been split into two groups in the North and South directions

	Cluster Labels	Ward	Area	Land Area (SKM)	Population	Latitude	Longitude	Total Positive
0	0	A	Colaba	12.5	210847	18.915091	72.825969	4367
1	0	B	Sanhurst Road	2.5	140633	18.956310	72.839838	1501
2	0	C	Marine Lines	1.8	202922	18.945670	72.823781	2498
3	0	D	Grant Road	6.6	382841	18.964447	72.813573	7848
4	0	E	Byculla	7.4	440335	18.976622	72.832794	6594
5	0	F North	Matunga	13.0	524393	19.027436	72.850147	6794
6	0	F South	Parel	14.0	396122	19.009482	72.837661	7657
7	0	G North	Dadar/Plaza	9.1	582007	19.023767	72.839403	9704
8	0	G South	Elphinstone	10.0	457931	19.008427	72.842505	7134
9	0	H East	Khar/Santacruz	13.5	580835	19.083966	72.841699	5715
10	0	H West	Bandra	11.6	337391	19.054979	72.840220	5420
15	0	M West	Chembur West	19.5	414040	19.057110	72.899388	5094

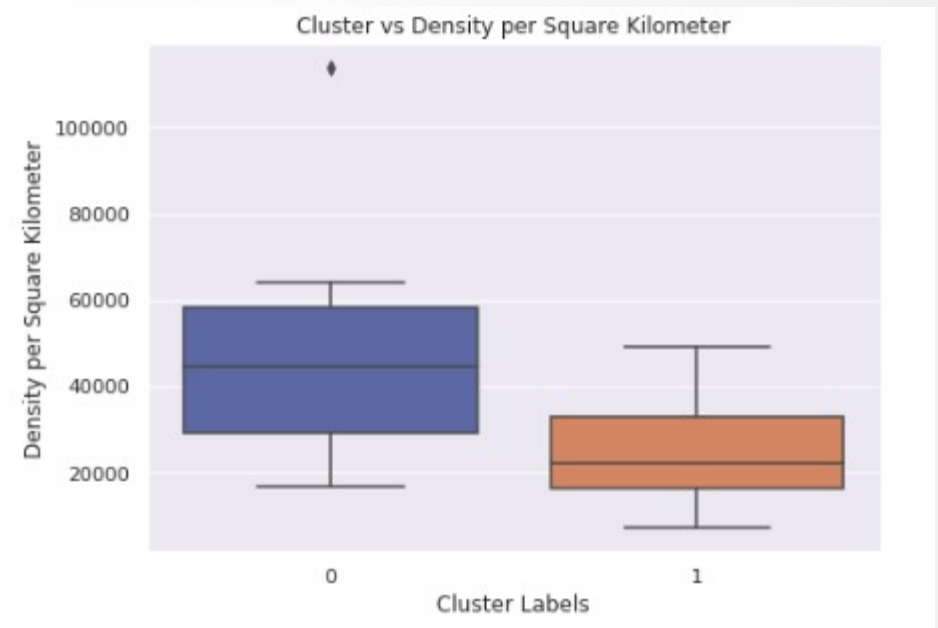|	Cluster Labels	Ward	Area	Land Area (SKM)	Population	Latitude	Longitude	Total Positive
11	1	K East	Andheri (East)	24.8	810002	19.115883	72.854202	10517
12	1	K West	Andheri (West)	23.4	700680	19.117249	72.833968	10475
13	1	L	Kurla	15.9	778218	19.065280	72.879380	6935
14	1	M East	Chembur East	32.5	674850	19.046066	72.895473	5792
16	1	N	Ghatkopar	26.0	619556	19.085954	72.908238	8871
17	1	P North	Malad	19.1	796775	19.186719	72.848588	10548
18	1	P South	Goregaon	24.4	437849	19.164803	72.850045	6282
19	1	R Central	Borivali West	50.0	513077	19.229456	72.847990	10927
20	1	R North	Dahiser	18.0	363827	19.253479	72.860903	5038
21	1	R South	Kandivalli	17.8	589886	19.201200	72.864332	9272
22	1	S	Bhandup	64.0	691227	19.143868	72.938433	9004
23	1	T	Mulund	45.4	330195	19.172290	72.956469	7987

Clusters and their relationships

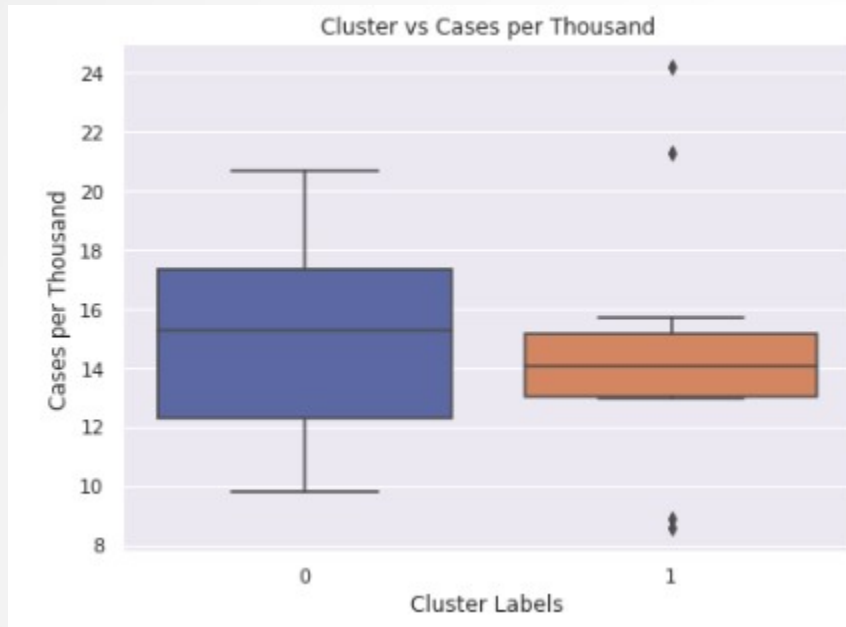Unearthing similarities within a Cluster and differences between Clusters

- Cluster 0 includes almost all the Wards that are smaller than in Cluster 1
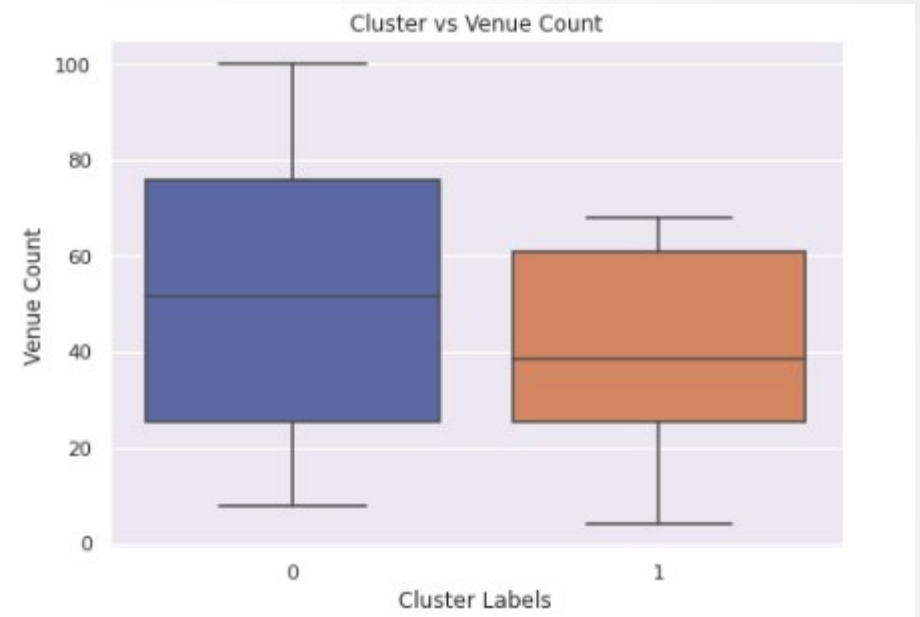- Largest Ward in Cluster 0 is marginally bigger than the smallest in Cluster 1

- Cluster 0 includes more of the denser Wards of the City
- Contrast this with the earlier observation of Cluster 0 comprising of smaller Wards

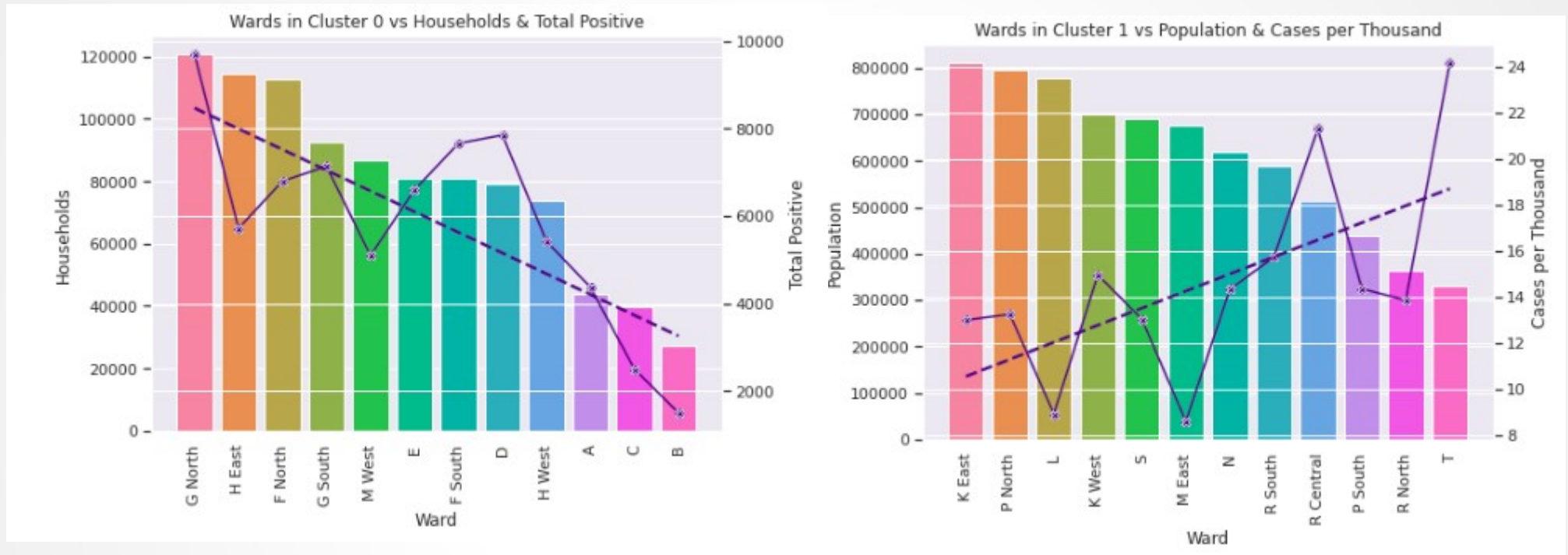Unearthing similarities within a Cluster and differences between Clusters

- Cluster 0 shows a wide variation in the # of Cases per Thousand across Wards when compared with Cluster 1 excluding the outliers

- Cluster 0 includes more Venues of interest than Cluster 1 despite being smaller in average size

Unearthing similarities within a Cluster and differences between Clusters

- Shows a direct relationship between the number of Households and the Total COVID-19 Positive Cases of the Wards in Cluster 0

- Shows an inverse relationship between the population and the # of COVID-19 Positive Cases per Thousand of the Wards in Cluster 1

Observations

- Smaller & the larger sized wards form Cluster 0 and 1 respectively

- Cluster 0 includes all the downtown locations with predominantly higher densities

- More venues of interest in Cluster 0, as it includes all the downtown areas

- Cases per Thousand per Ward in Cluster 0 is in the range of ~10 to ~21 (2X)

- Cases per Thousand per Ward in Cluster 1 is in the range 13 to ~16 (w/o outliers)

- Cluster 0: Total Positive per Ward relates directly to size of the Households

- Cluster 1: Cases per Thousand per Ward inversely relates to the size of the Population

- Foursquare seems to provide sparse info on the venues and their categories

- Sparse Venue Data appear to have had very little influence on the final outcome

- Avg # of Cases per Thousand per Ward is ~15 for Cluster 0 & ~14 for Cluster 1

Conclusions

- **Insights**
  - The spread of the virus is slightly more in Cluster 0 than Cluster 1

  - In Cluster 0, focus for combating the disease may be more along the lines of
    - Households in descending order i.e. priortize Wards with more over lesser

  - In Cluster 1, focus for combating the disease may be based on the
    - Population in an ascending order i.e. priortize Wards with lesser over more

  - A more broader strategy may need to be followed to contain the spread in Cluster 0 as the Cases per Thousand per ward is in a wide range (2X)

  - A homogeneous approach may work well in containing the spread in Cluster 1 with outliers handled as appropriate

- **Future Directions**
  - Recompute the Clusters by including Total Deaths per Ward as a new feature
  - Look at alternate source of Location Data to arrive at a better understanding of the relationship between Venues and the spread of COVID-19 in Mumbai

Thank You!