

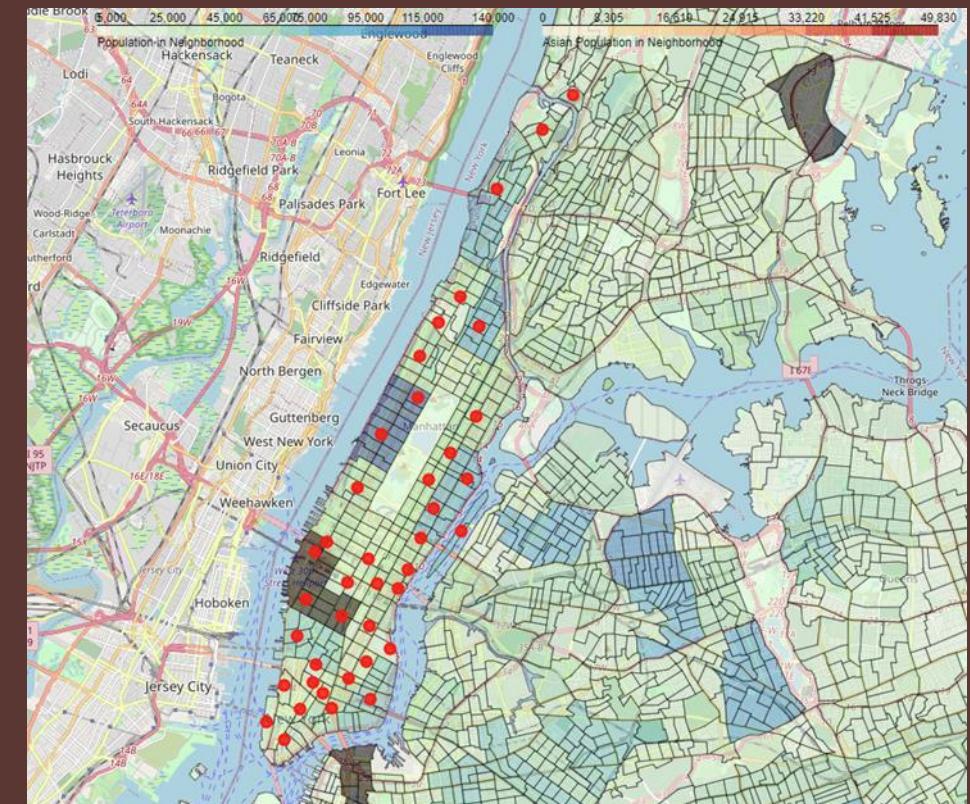


BATTLE OF THE NEIGHBORHOODS

TORONTO VS NEW YORK

Introduction

- Explore New York and Toronto for potential neighborhoods
 - to set up restaurants
 - For potential immigrants/students etc
- Business Problem
 - Understand the city and its neighborhoods
 - Current tastes and competition
 - Potential demand / market size
- Success Criteria
 - Identify potential neighborhoods



Toronto City

- Capital city of Ontario
- Most populous city in Canada, fourth most in North America
- Recognized as one of the most multicultural and cosmopolitan cities in the world
- Considered the financial and industrial capital of Canada.
- Greater Toronto Area has an estimated population of 6.4 million (census 2016), and is one of the most diverse cities in the world.



New York City



- Most populous/dense city in the United States
 - It's one of the world's most populous megacities
 - NYC is considered the cultural, financial and media capital of the world
 - Extremely influential
- 3 of the 10 most visited tourist attractions in the world.
- Approximately 37% of population is foreign born. Very diverse!

Business Environment

- Both cities are:
 - Highly populated
 - Economically active
 - Very competitive
- An initial understanding of how the neighborhoods and popular venues are distributed would help prospective businesses to understand the risks/rewards

Cuisines - Toronto

- Different ethnic neighbourhoods throughout the city focus on specific cuisines,
 - authentic Chinese and Vietnamese found in the city's Chinatowns
 - Korean in Koreatown,
 - Greek on The Danforth
 - Italian cuisine in Little Italy and Corso Italia
 - Indian/Pakistani in Little India

Cuisines - New York City

- In New York there are over 12,000 bodegas, delis, and groceries, and many among them are open 24 hours a day, 7 days a week.
 - Central and Eastern European immigrants - popularized bagels, cheesecake, hot dogs, knishes and delicatessens
 - Italian immigrants : New York-style pizza and Italian cuisine
 - Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous
 - Mobile food vendors : 4000 licensed by city, many immigrant owned
 - The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin.

Factors studied in this sprint

- Population
- Demographics : ethnicities distribution
- Venues distribution
 - Restaurant types and distribution
- More factors to be studied in later sprints

Data Description 1 - Location / Neighborhoods

- New York Neighborhood dataset has a total of 5 boroughs and 306 neighborhoods. This dataset includes the Boroughs, Neighborhood in them and the latitude and longitude coordinates of each neighborhood. We will initially focus only on the Manhattan borough.
link: https://geo.nyu.edu/catalog/nyu_2451_34572
- The data was downloaded by JSON and converted to dataframe format.
 - dataset filtered for Manhattan Borough
 - however, there were certain errors when trying to combine with the next data set (venues).
 - Proceed with only Manhattan

Data Description 1 - Location / Neighborhoods

- BeautifulSoup was used on the wikipedia page to obtain the Boroughs and Neighborhoods in Toronto and create a dataframe.
 - Used nominatim geolocator (geopy) addon, proceeded to get each neighborhoods Latitude and Longitude coordinates
- Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- The boroughs in the data set was manually cleaned, for example EtobicokeNorthwest was converted to Etobicoke Northwest with a manual replace line in the code

Data Description 2 - Venue

- Utilize the neighborhood location data and Foursquare API to explore venues nearby
 - Search parameters : radius of 500m and 100 venues limit per neighborhood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roselawn	43.7113	-79.4195	Ceiling Champions	43.713891	-79.420702	Home Service
1	Roselawn	43.7113	-79.4195	BAGEL TIME MONTREAL STYLE	43.709067	-79.415858	Fast Food Restaurant
2	Roselawn	43.7113	-79.4195	The Bewding	43.707047	-79.417756	Spa
3	Summerhill West, Rathnelly, South Hill, Forest...	43.6861	-79.4025	The Market By Longo's	43.686711	-79.399536	Supermarket
4	Summerhill West, Rathnelly, South Hill, Forest...	43.6861	-79.4025	LCBO	43.686991	-79.399238	Liquor Store

Data Description 3 - Population/ethnicity and Neighborhood boundaries

- New York
 - Population data : Taken from NYC open data website
<https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulation/swpk-hqdp>
 - Ethnicity (Asian, Hispanic): census 2010 data from the official website of the City of New York
 Downloaded PDF, converted to CSV and performed data shaping using excel. Then imported into Dataframe

	Borough	Year	FIPS County Code	NTA Code	NTA Name	Population
0	Bronx	2000	5	BX01	Claremont-Bathgate	28149
1	Bronx	2000	5	BX03	Eastchester-Edenwald-Baychester	35422
2	Bronx	2000	5	BX05	Bedford Park-Fordham North	55329
3	Bronx	2000	5	BX06	Belmont	25967
4	Bronx	2000	5	BX07	Bronxdale	34309
...
385	Staten Island	2010	85	SI37	Stapleton-Rosebank	26453
386	Staten Island	2010	85	SI45	New Dorp-Midland Beach	21896
387	Staten Island	2010	85	SI48	Arden Heights	25238
388	Staten Island	2010	85	SI54	Great Kills	40720
389	Staten Island	2010	85	SI99	park-cemetery-etc-Staten Island	0

	Borough	2010\nCensus FIPS County Code	Code	Name	Total\nPopulation	White	Black/\nAfrican American	American\nAlaska Native	Asian	Native\nHawaiian and Other Pacific Islander	Some\nOther Race	Total	Hispanic\nOrigin (of any race)
0	Bronx	5.0	BX01	Claremont-Bathgate	31078	370	13036	82	108	7	60	221	17194
1	Bronx	5.0	BX03	Eastchester-Edenwald-Baychester	34517	1111	24381	147	232	10	148	509	7979
2	Bronx	5.0	BX05	Bedford Park-Fordham North	54415	3637	9805	145	2727	29	490	623	36959
3	Bronx	5.0	BX06	Belmont	27378	5381	5059	38	620	18	84	249	15929
4	Bronx	5.0	BX07	Bronxdale	35538	5559	10594	102	1404	9	169	430	17271

Data Description 3 - Population/ethnicity and Neighborhood boundaries

- Toronto
 - Population data & Ethnicity data : 2016 census data : taken from toronto open data portal home
<https://open.toronto.ca/dataset/neighbourhood-profiles/>

Category	index	Population	Latin American origins	Asian origins
0	Agincourt North	29113.0	470.0	24305.0
1	Agincourt South-Malvern West	23757.0	480.0	17955.0
2	Alderwood	12054.0	315.0	2055.0
3	Annex	30526.0	765.0	6485.0
4	Banbury-Don Mills	27695.0	585.0	12025.0
...
135	Wychwood	14349.0	645.0	2500.0
136	Yonge-Eglinton	11817.0	370.0	2895.0
137	Yonge-St.Clair	12528.0	300.0	2330.0
138	York University Heights	27593.0	2055.0	12550.0
139	Yorkdale-Glen Park	14804.0	1025.0	4090.0

140 rows × 4 columns

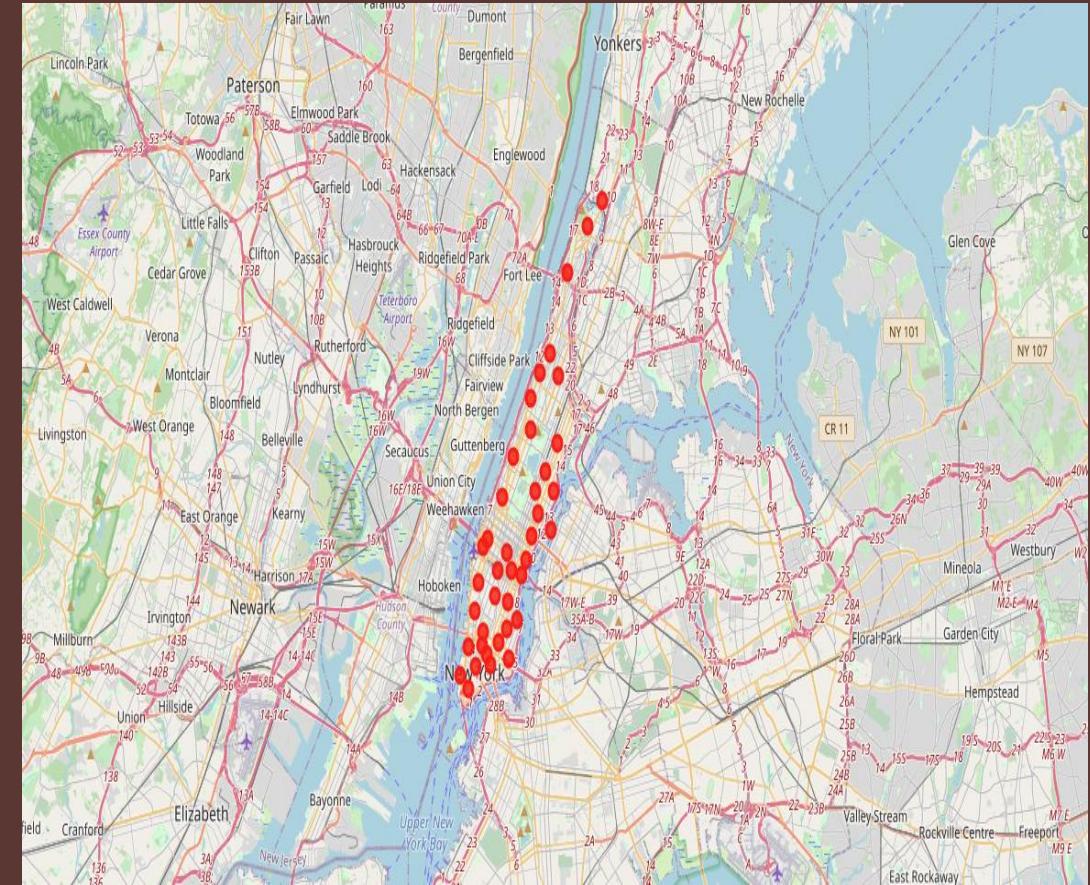
Methodology

- New York city has a total of 5 boroughs and 306 neighborhoods
 - Only focusing on Manhattan Borough
- Toronto has 5 districts and over 140 neighborhoods
 - Only focusing on neighborhoods with ‘Toronto’ in the designation
- Clustering is based on all venues from Foursquare to get an understanding of the neighborhood profiles
- Venue categories are filtered to ‘restaurants’ to explore the any differences between the two cities

Exploratory Data Analysis - Neighborhood

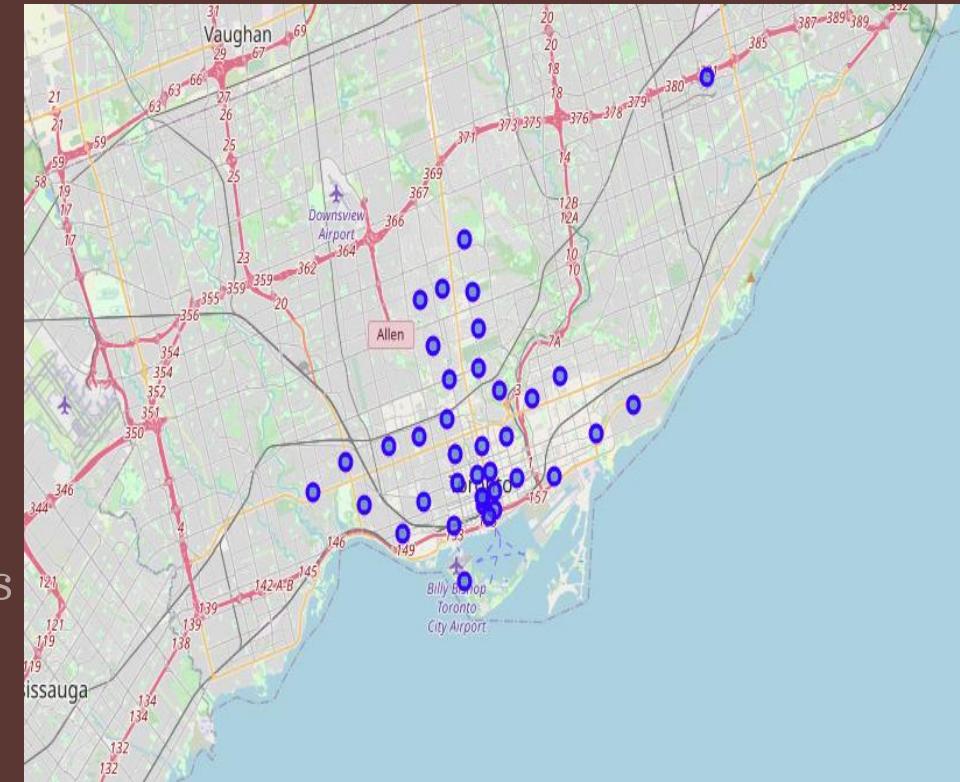
New York

- We load data from newyork_data.json file
- Transform data of nested python dictionaries into a pandas dataframe
- This dataframe contains the geographical coordinates of New York city neighborhoods
- Use the above coordinates to feed into the ‘get’ command using the Foursquare API
- Use Geopy and folium libraries to create map of New York city with neighborhoods superimposed



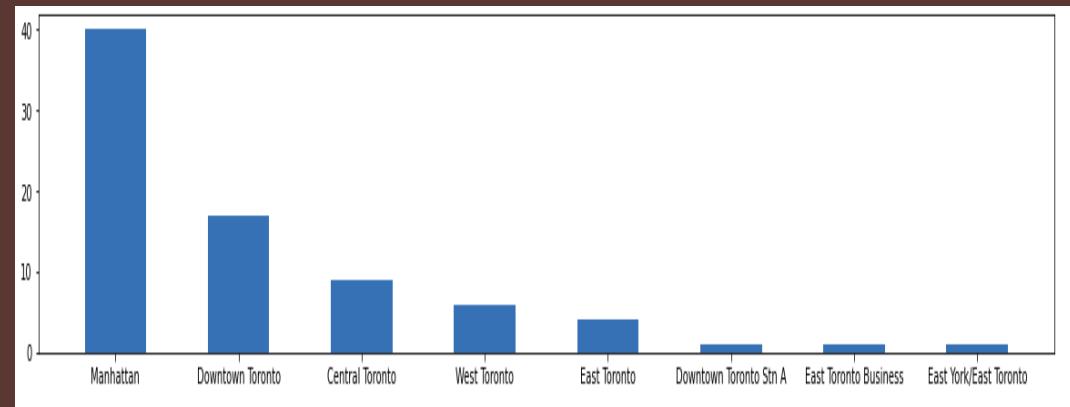
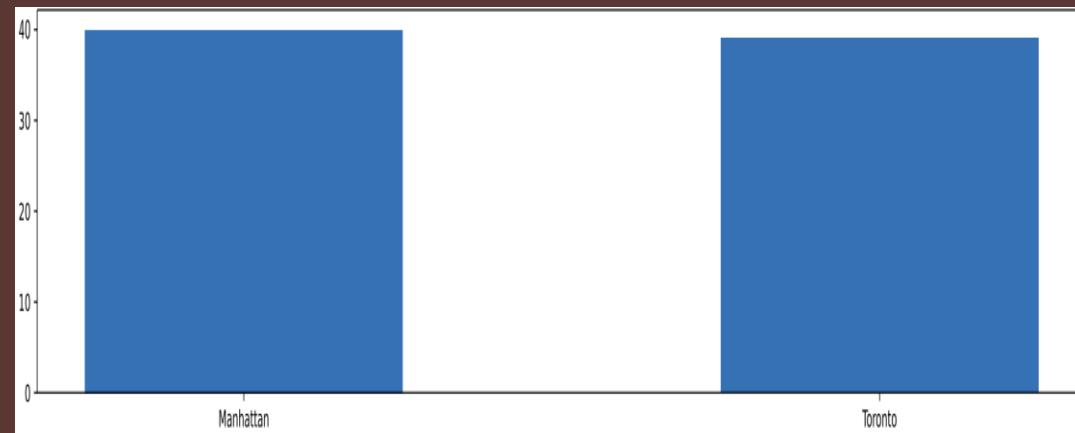
Exploratory Data Analysis - Neighborhood

Toronto



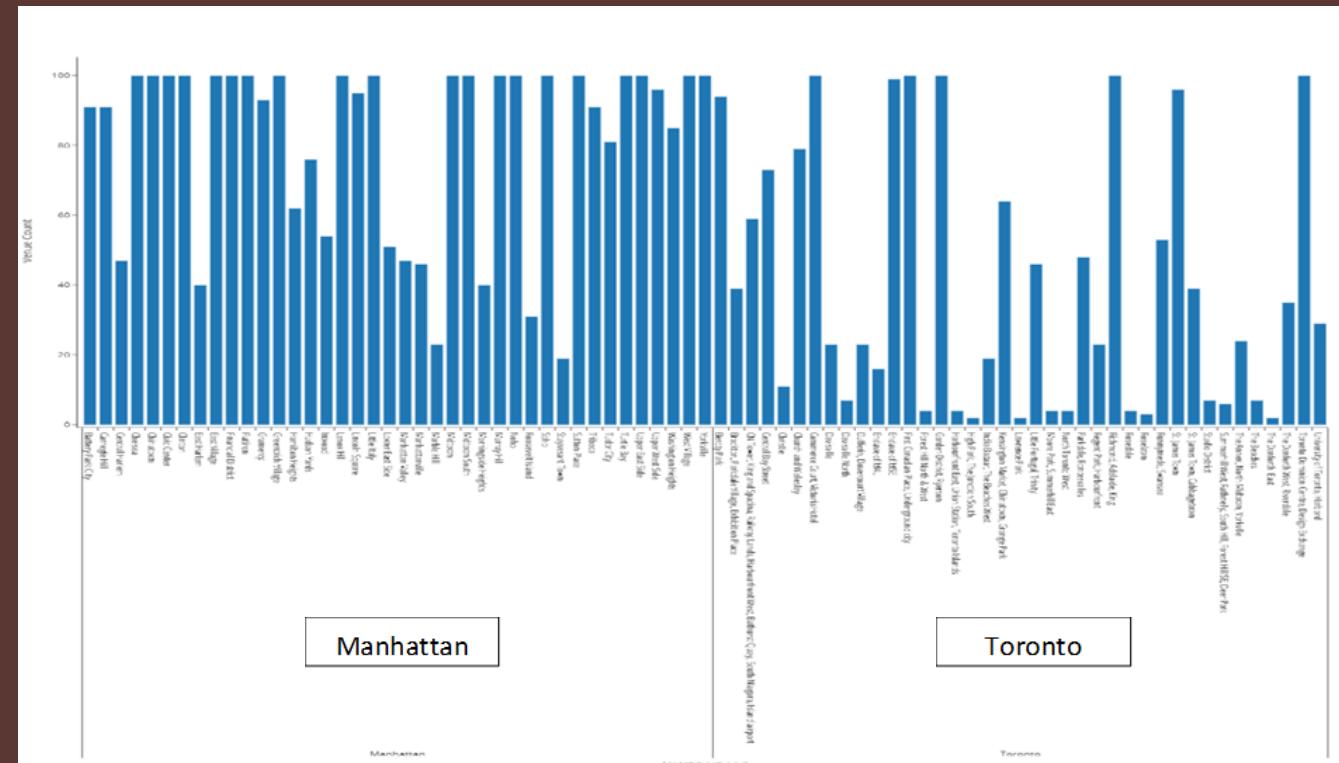
Exploratory Data Analysis - Neighborhoods

- A brief comparison of neighborhood data set
 - Similar number of neighborhoods for comparison (40 Manhattan, 39 Toronto)
 - Only Manhattan borough in NYC while 7 separate districts for Toronto



Exploratory Data Analysis - Foursquare venues

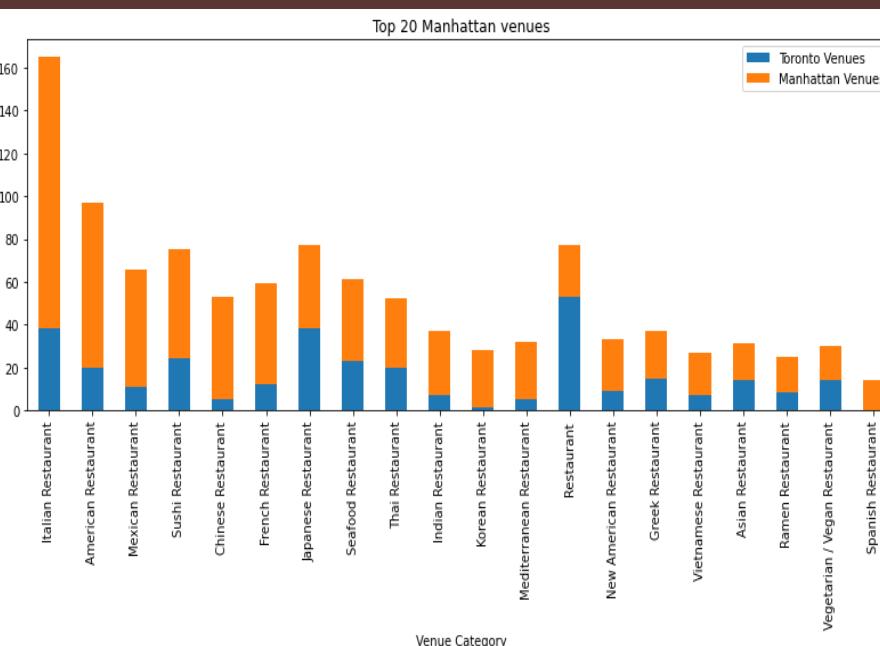
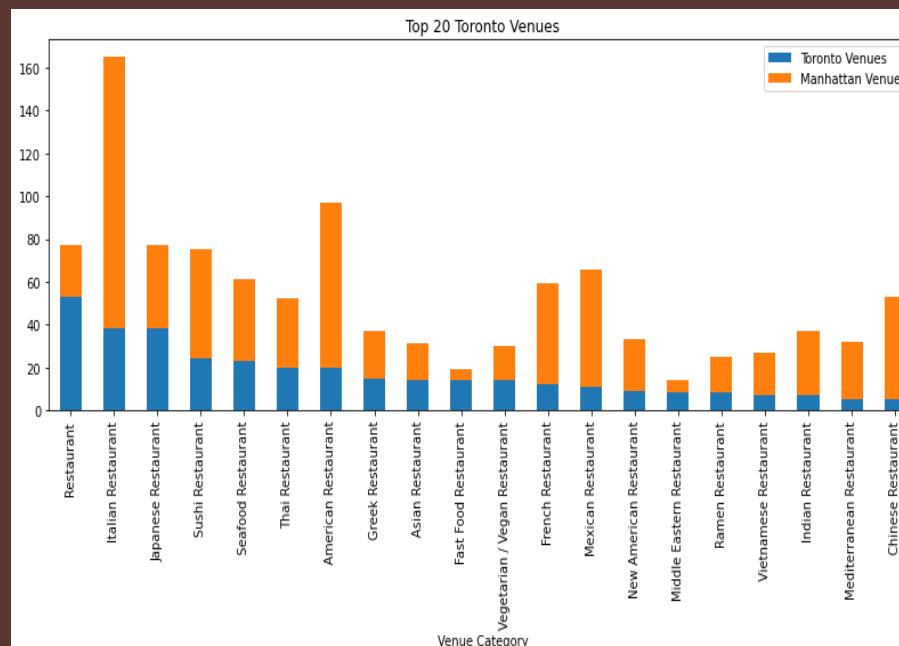
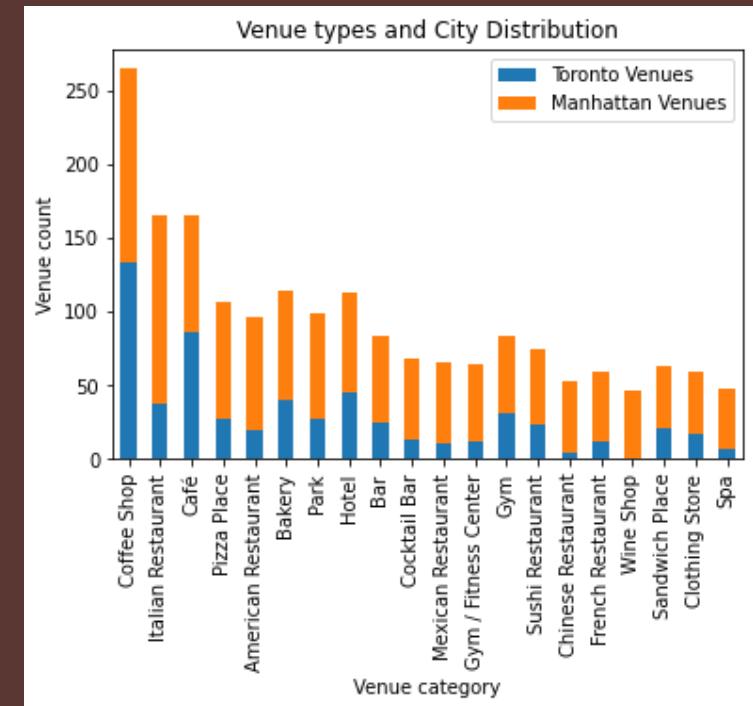
- A brief comparison of venues data set
 - Manhattan has far more venues than Toronto in this data set
 - More variation as well from the number of categories
 - Toronto dataset a mix of high venue density neighborhoods and some very low density ones
 - Low density venues : Highly likely residential areas
 - Manhattan mostly high density venue with a few medium density ones



	Neighborhood	Venue	Venue Category
city			
Manhattan	40	2787	332
Toronto	39	969	217

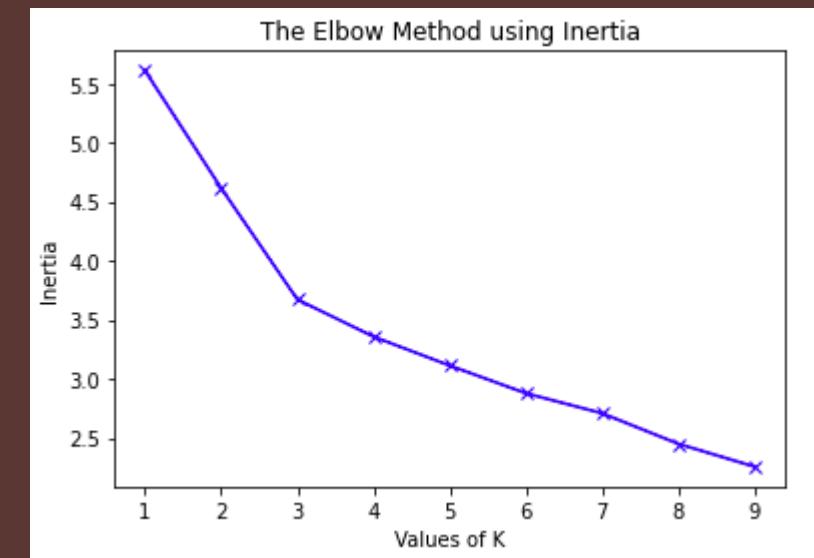
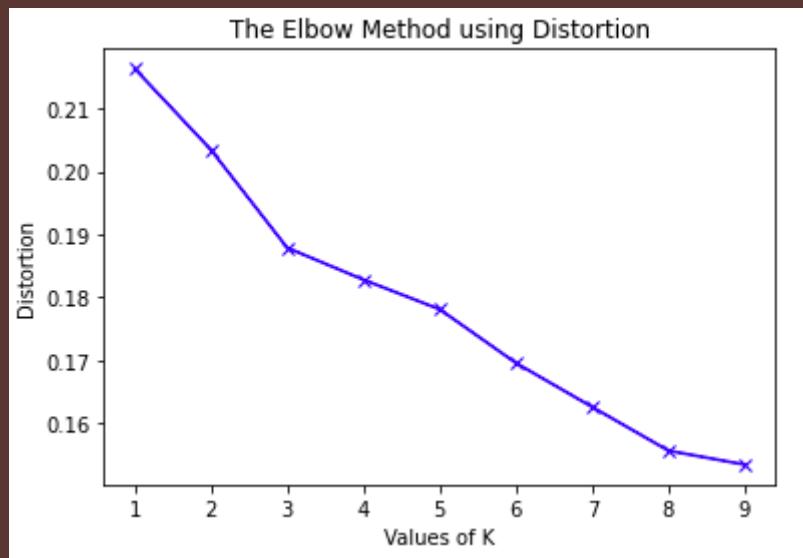
Exploratory Data Analysis - Foursquare venues

- A brief comparison of venues data set (Re)
 - In terms of quantity, Coffee shops and cafes are highly popular in both cities.
 - Restaurants
 - Italian , Asian cuisines are highly popular in both cities



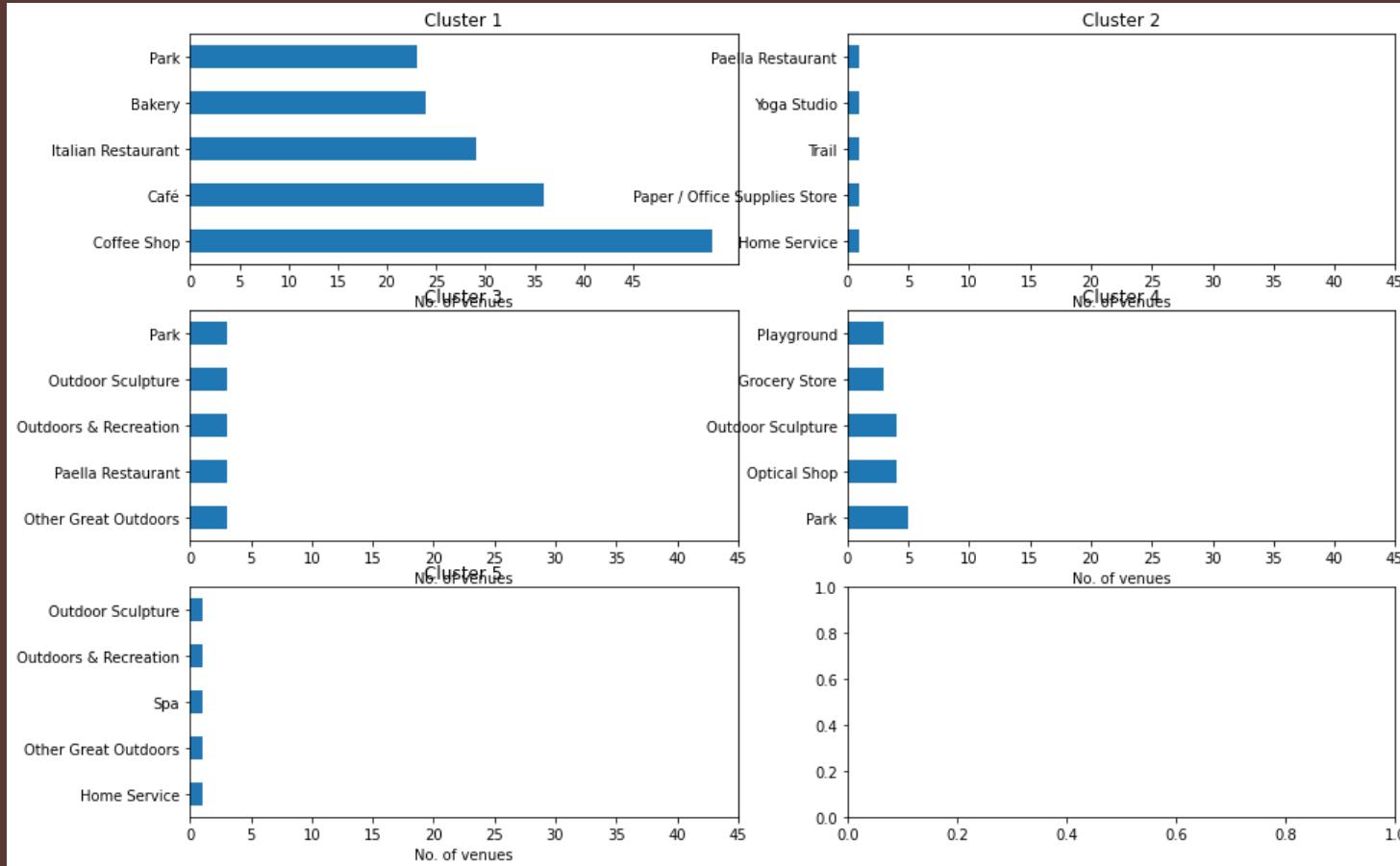
Results

- Clustering of neighborhoods based on mean occurrence of venue categories
 - Finding optimum value of K
 - Optimum k ~4
 - Our K set at 5



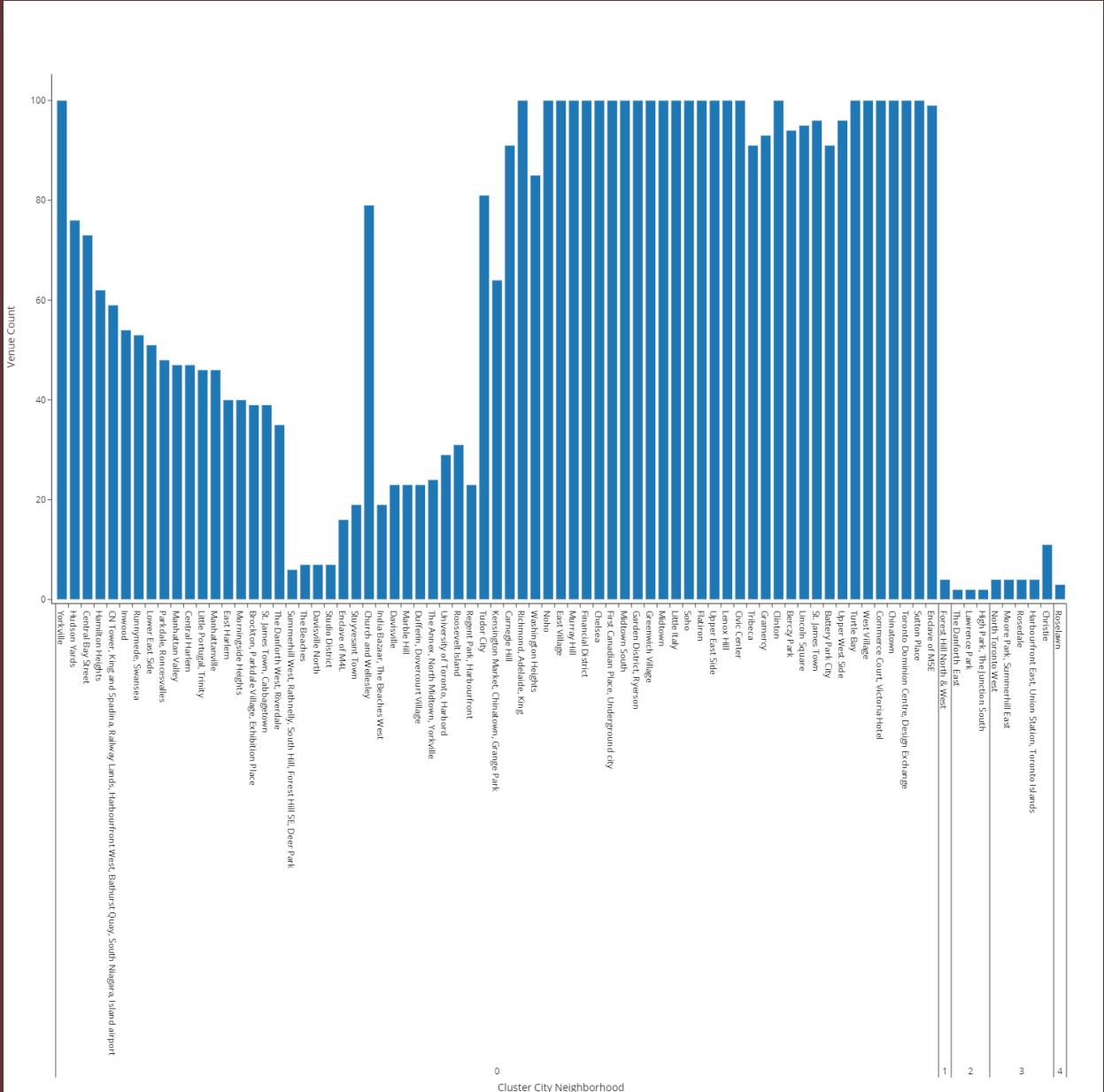
Results

- Clustering of neighborhoods based on mean occurrence of venue categories
 - Cluster 0 is the one with the highest number of neighborhoods.
 - composition does not differ much
 - coffee shops,cafes and eateries popular
 - other clusters have much smaller densities
 - Cluster 3 : consist mostly of residential neighborhoods
 - grocery parks, stores, and playgrounds being the popular venues.



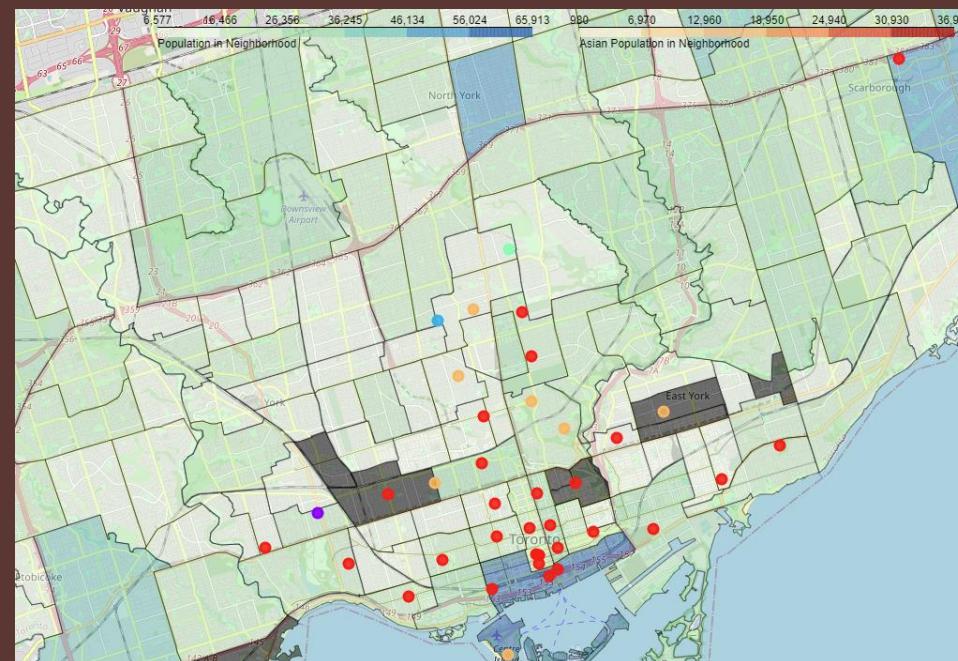
Results

- Clustering of neighborhoods based on mean occurrence of venue categories
 - Cluster 0 is the one with the highest number of neighborhoods.
 - composition does not differ much
 - coffee shops,cafes and eateries popular
 - other clusters have much smaller densities
 - Cluster 3 : consist mostly of residential neighborhoods
 - grocery parks, stores, and playgrounds being the popular venues.



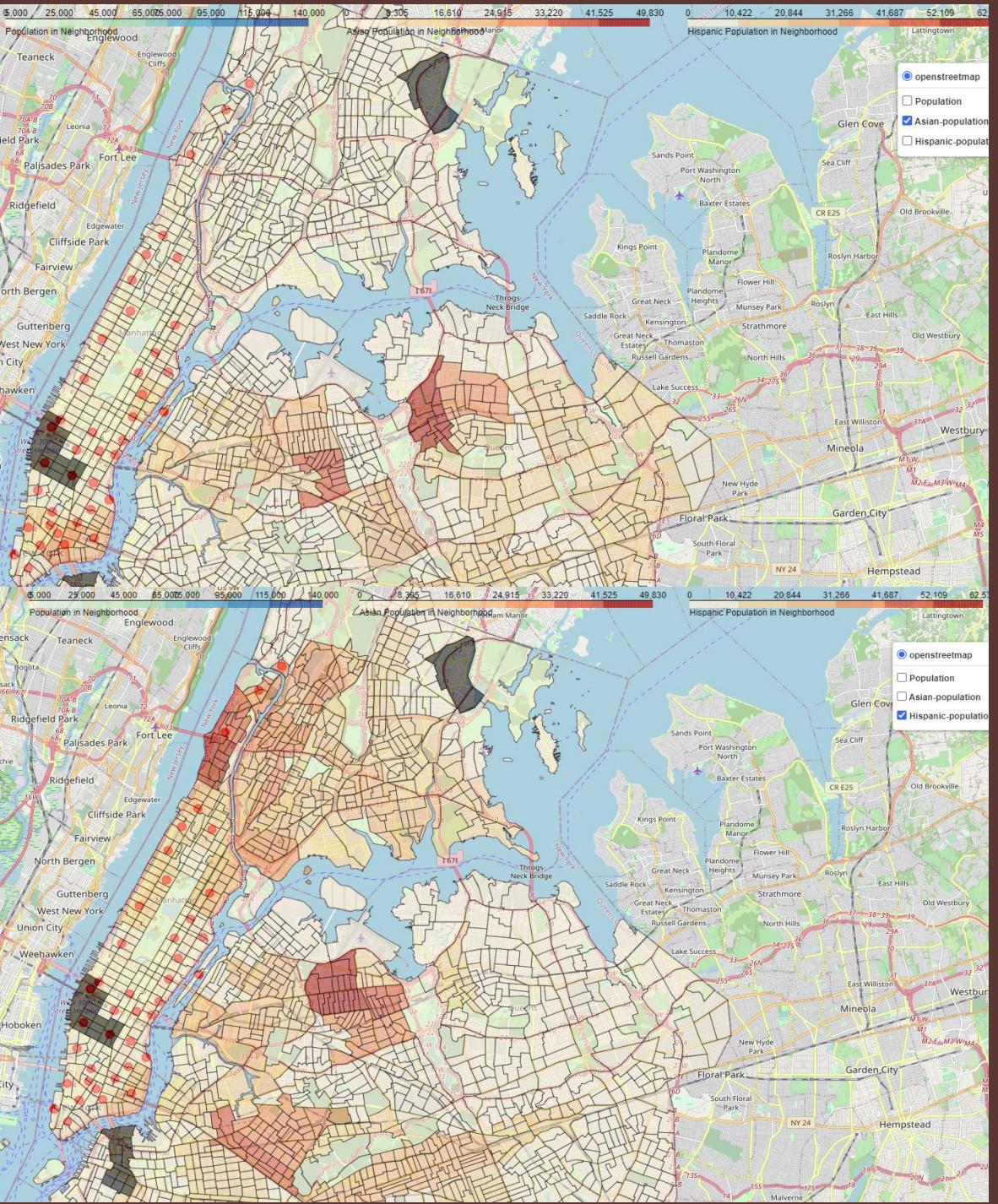
Results

- Clustering of neighborhoods based on mean occurrence of venue categories.
- Overlaid with population by neighborhood
 - Cluster 0 is the one with the highest number of neighborhoods.
 - composition does not differ much
 - coffee shops,cafes and eateries popular
 - other clusters have much smaller densities
 - Cluster 3 : consist mostly of residential neighborhoods
 - grocery parks, stores, and playgrounds being the popular venues.



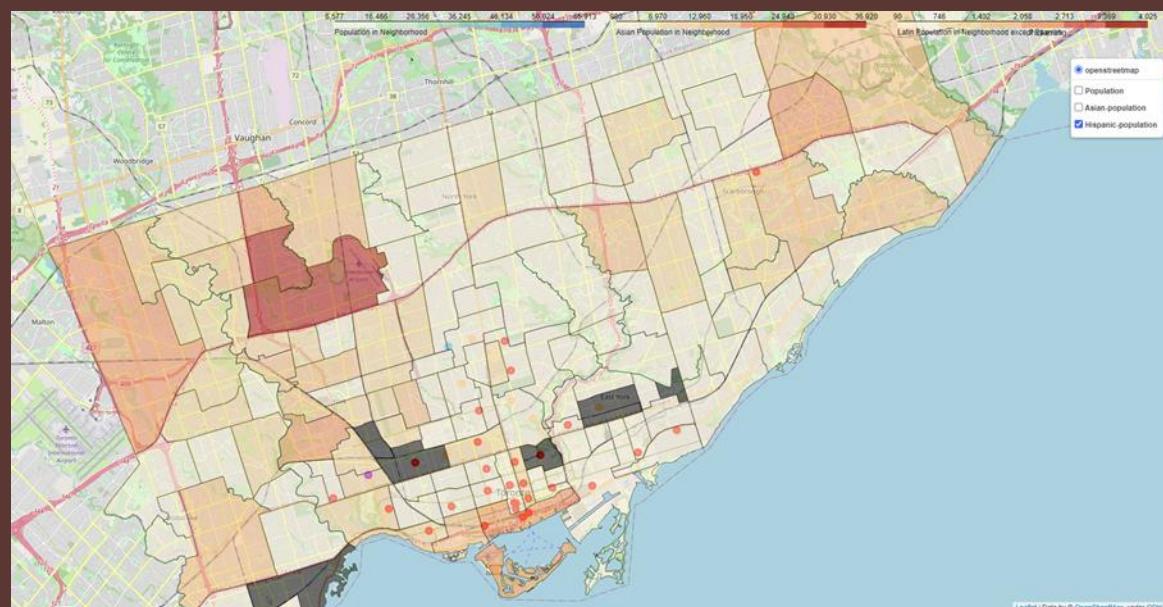
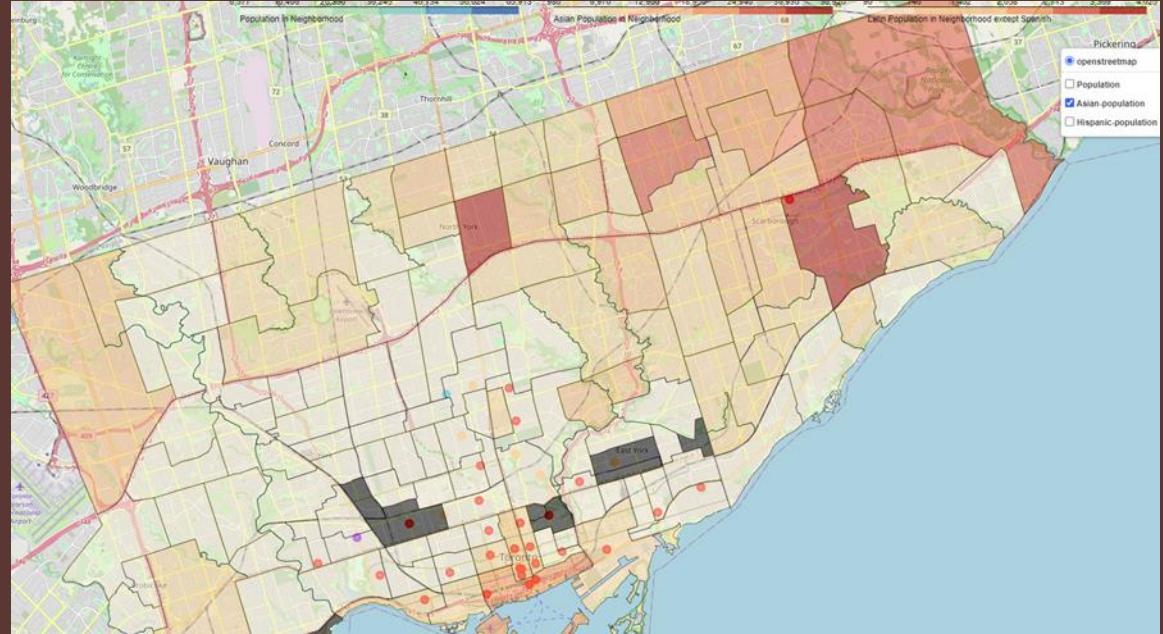
Results

- Choropleth of Asian and Hispanic populations
 - Asian and Hispanic population distribution are distinct
 - Asian population centers roughly
 - West Queens
 - Brooklyn
 - South East of Manhattan
 - Hispanic population centers
 - Northwest Manhattan
 - Queens



Results

- Choropleth of Asian and Hispanic populations
 - Asian and Hispanic population distribution are distinct except for downtown Toronto
 - Very mixed population
 - Asian population centers roughly
 - Scarborough
 - Hispanic population centers
 - Etobicoke
 - North York



Discussion

- For a prospective investor into the restaurant in either city, another useful information would be real estate prices either for rental or purchase
- There is much more competition in Manhattan but this is proportionate to the density of the population compared to Toronto
 - Need to check the other boroughs
- With the pandemic, a different approach might be considered. A restaurant does not need to be located exactly in the middle of a population dense area if the location can be easily serviced by food delivery riders.
- The data is inconclusive as the current data set is not yet fully explored. There is scope for
 - investigating on current popularity, pricing, menu of the venues using foursquare
 - investigate the status of more neighborhoods especially ones that have high contrast of ethnicity, tourist footfall etc.

Conclusion

- Analysis was performed on limited time and data
 - plenty more open access data available that can be utilized to give a better picture of the neighborhoods
 - Insufficient time to shape or utilize data for study
 - Foursquare : price, menu and rating data unused
 - Census data : family income per year etc ,age etc unused
 - Using the data in different ways
 - Could have attempted to cluster the neighborhoods on filtered venues to restaurants only for second layer after understanding the neighborhoods
- Areas under analysis limited
 - Much of the variation is on the outskirts of the areas under scrutiny