MOKHTAR-SLIMANE Ghaïs

IBM Project : Create a New business in Toronto

Table of Contents

1. Business understanding

1.1 Executive summary :

1.2 Introductory/background :

Setting up the problem for the reader who might be new to the topic, and he need the introduced the subject matter.

1.3 Interest of an audience

Background : News, review & History of the Topics

2. Data Section

2.3 Data required to resolve problem

2.3 How the data will be used to solve the problem

2.4 Mapping of Data

3. Methodolies

3.2 Data Wrangling

3.3 Exploratory Data Analysis

3.4 Clustering

3.5 Explore Clustering

4. Results

5. Conclusion

My stakeholders , a investor who want to create a business in specifying area, need answers to this question :

Where open a Restaurant in Old Toronto City, who minimized the competition of popular venues category ?

1.1 Executive summary:

Goals : The challenge to resolve is being able to find a best location in Toronto to open Restaurant.

That offers the best characteristics to maximize the chance of getting a benefits into create a profitable business in this area.

Objectives :

1) Look venues around each Neighborhood.
2) Segment the Borough and their Neighborhood within the Most common venues in each (populars spots).
3) Choose the not used categories (new shop category in this area)
4) Minimized the direct concurencies. b. Minimized the indirect concurencies

Others Options : Maximize the trendings venues(highest foot trafic).

IF Statement : Go More Data

1) Affinement of Data mining used for analysis & visualization

- Description of Borough characteristics. (Demographics, transportation, economics, Cultural...)
- Try too choose the best population living around (workers, tenant of this houses, tourism etc)
- Choose the Neighborhood who corresponding the best of the Business model.
- Maximize the trendings venues (highest foot trafic)

1) Change the question, Others way to choose the business location :

- Choose the most used categories (same shop with innovative features, i.a.e = Restaurant['French','Spanish...)] :
- . Maximize the trendings venues . Maximize the domain wihtin the venues categories of the Neighborhood. . Find innovative pattern in this category.

1.3 Interested Audience

I believe this is a relevant project for a person or entity considering investing to a major city in Canada..

Since the approach and methodologies used here are applicable in all cases.

The use of FourSquare data and mapping visualisation, combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

Background :

Toronto is a provincial capital of Ontario, and the the most populous in Canada, located on the northwestern shore of Lake Ontario.

This area of the city is an international Centre of Business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, business services, environmental innovation, food services, and tourism.

Toronto is a prominent centre for music, theatre, motion picture production, and television production, and is home to the headquarters of Canada's major national broadcast networks and media outlets.

Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities, attract over25 million tourists each year.

The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada.
More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.

2. Data Section :

Description of the data and its sources that will be used to solve the problem

2.1 Data Required to resolve the problem

In order to make a good choice to obtain a good venues in Downtown Toronto, the following minimum data is required:

1) **First table** with 2 steps :

Information on each Postal Code with the Borough/Neighborhoods from Toronto. Scrape the wikipedia page : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, in order to obtain the data that is in the table of postal code

b) Information of the Geocoordinates for each Costal codes (latitude and longitude). Just excluded the table not contains the ' Downtown Toronto' bourough.

2) **Second table**

With this table, I can used the Foursquare API with : venues/explore, venues/trendings/ venues/search and others methdologies of this API. .

I can used the Foursquare location, firstly with 'venues/explore', to obtain the 10th Most common venues (popular spots) for each Neighborhood, where are in the borough of Old Toronto.

2.3 How the data will be used to solve the problem

The data will be used as follows: Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups of poular sports.

After obtains this table, I can cluster the borough with common venues, and obtain segmented borough with Unsupervised machine learning Clustering = K means algorithms.

2.4 Mapping of Data

The following maps were created to facilitate the analysis and the choice of the place to invest for your Restaurant. Which correspond to the minimum direct concurrencies of your venues category.

Also, you can Create a map that depicts, the demographics incomes and others datas, to Describe the Neighborhood more within the population.

3.2 Exploratory Data Analysis

- The downtown Toroton contains 18 borough.

Most popular venues in Downtown return with FourSquare

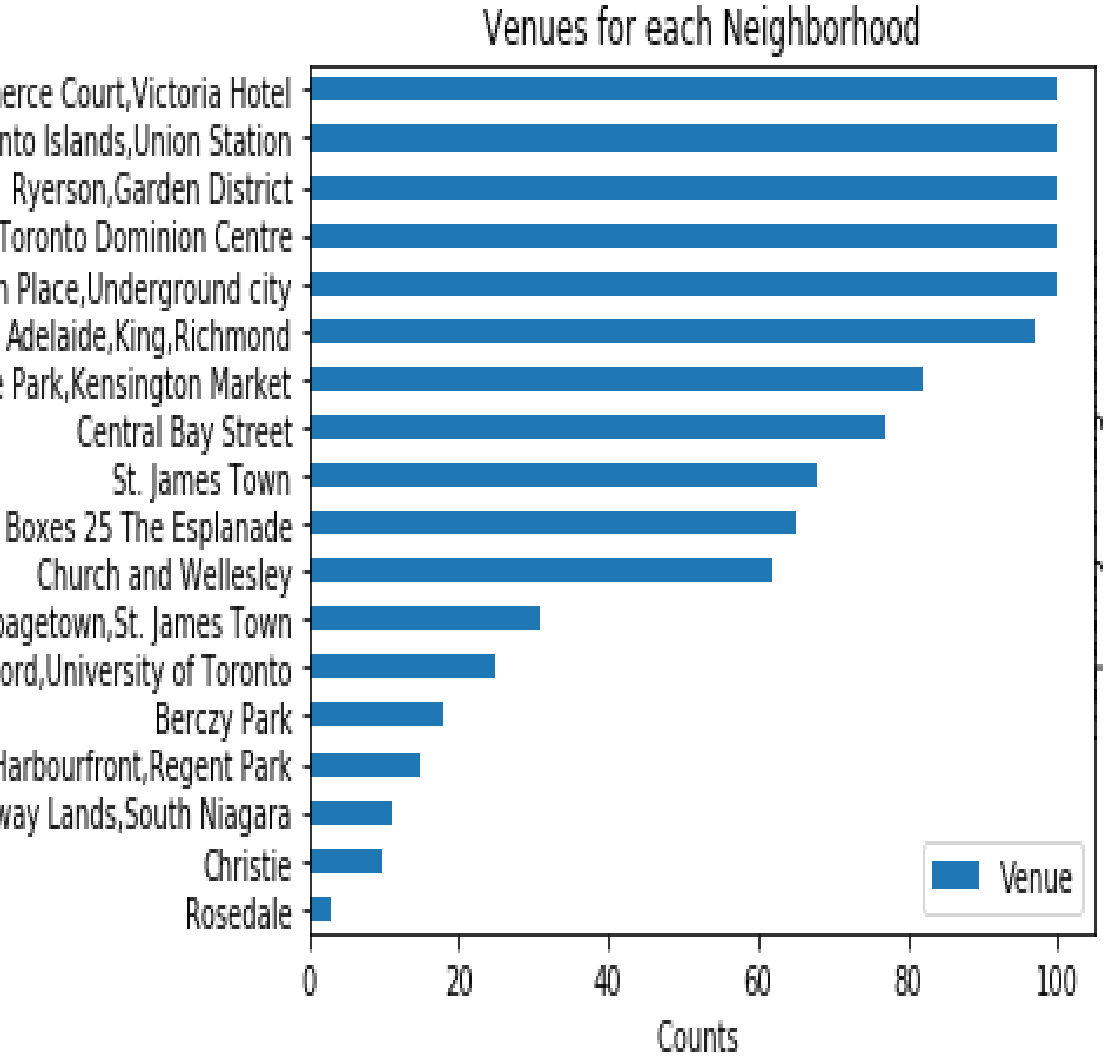
- I obtain 1057 venues and map the all venues in this plot.

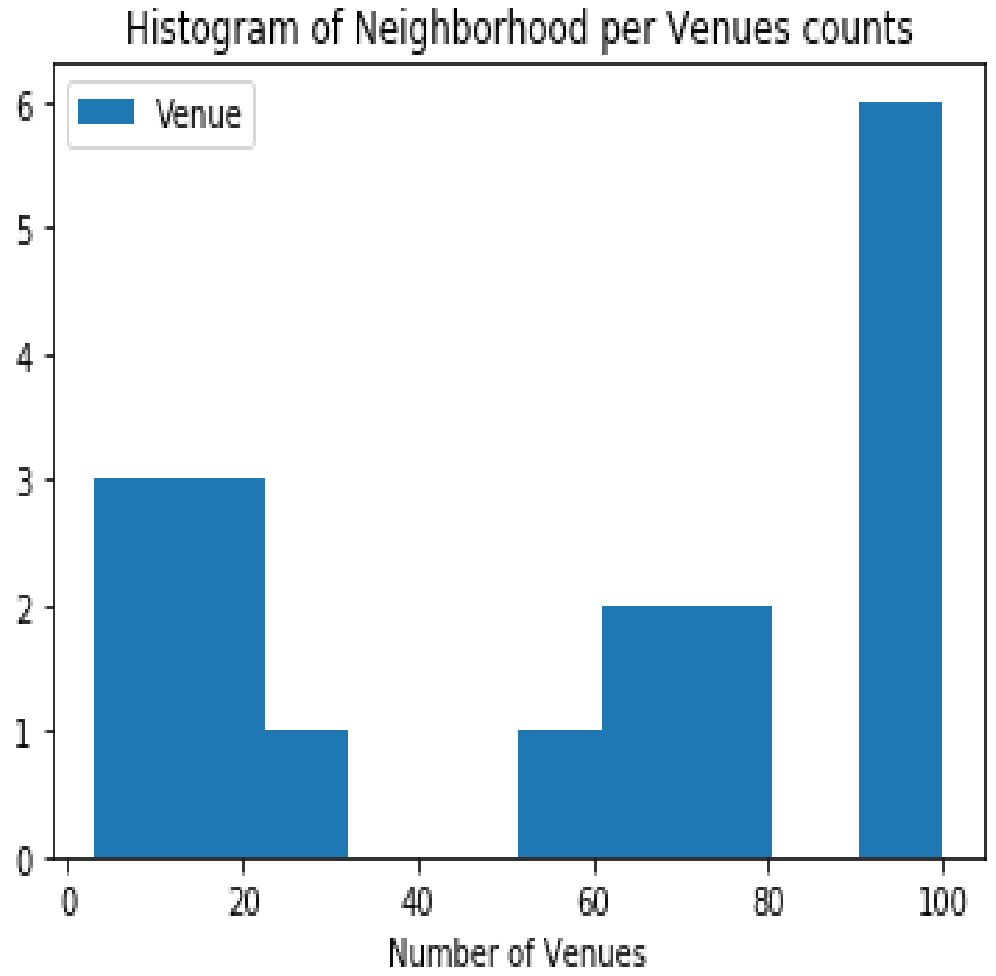
- Just with map, you can see which borough are satured which

  neighbourhoods are saturated with

Descriptive statistics of Venues per Neigh

- I have 185 unique categories & 659 uniques venues in this table and map.
- With this Graph, I can see the 6 borough who are satured for venues





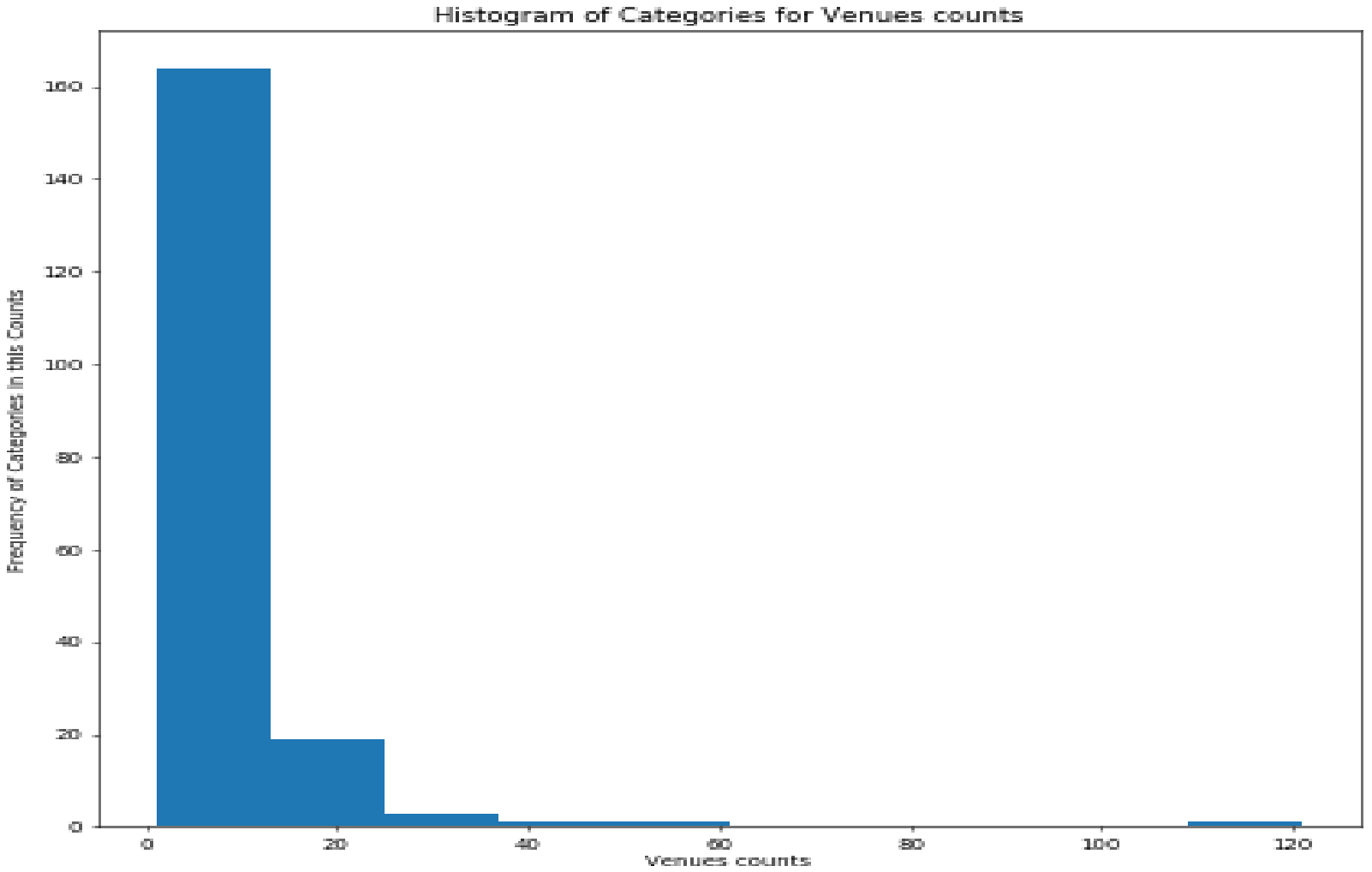
Most popular venues in Downtown return with FourSquare

- I obtain 1057 venues and map the all venues in this plot.

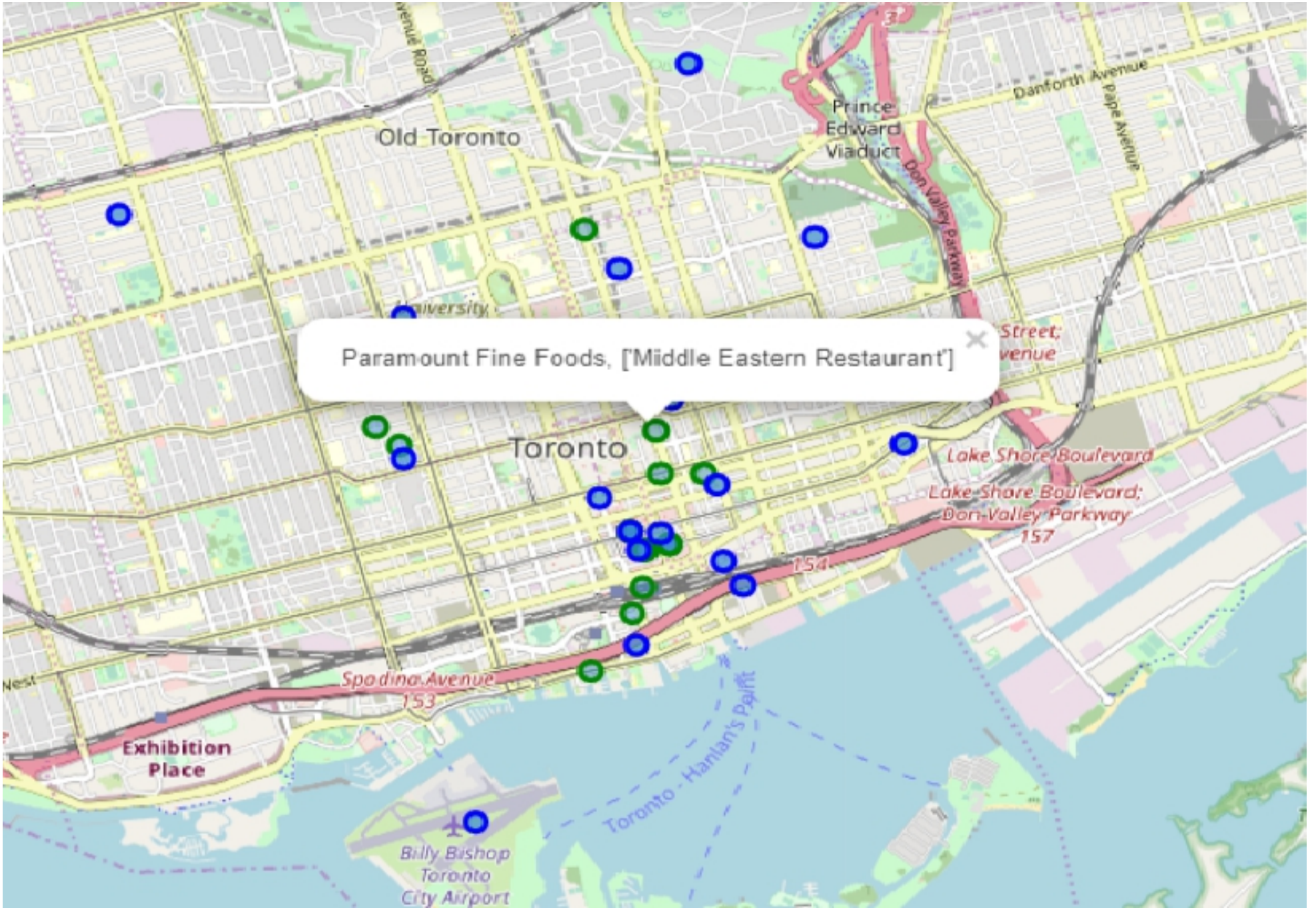
- I have 185 unique categories & 659 uniques venues in this table and map.

Descriptives stats : Venues & Categories

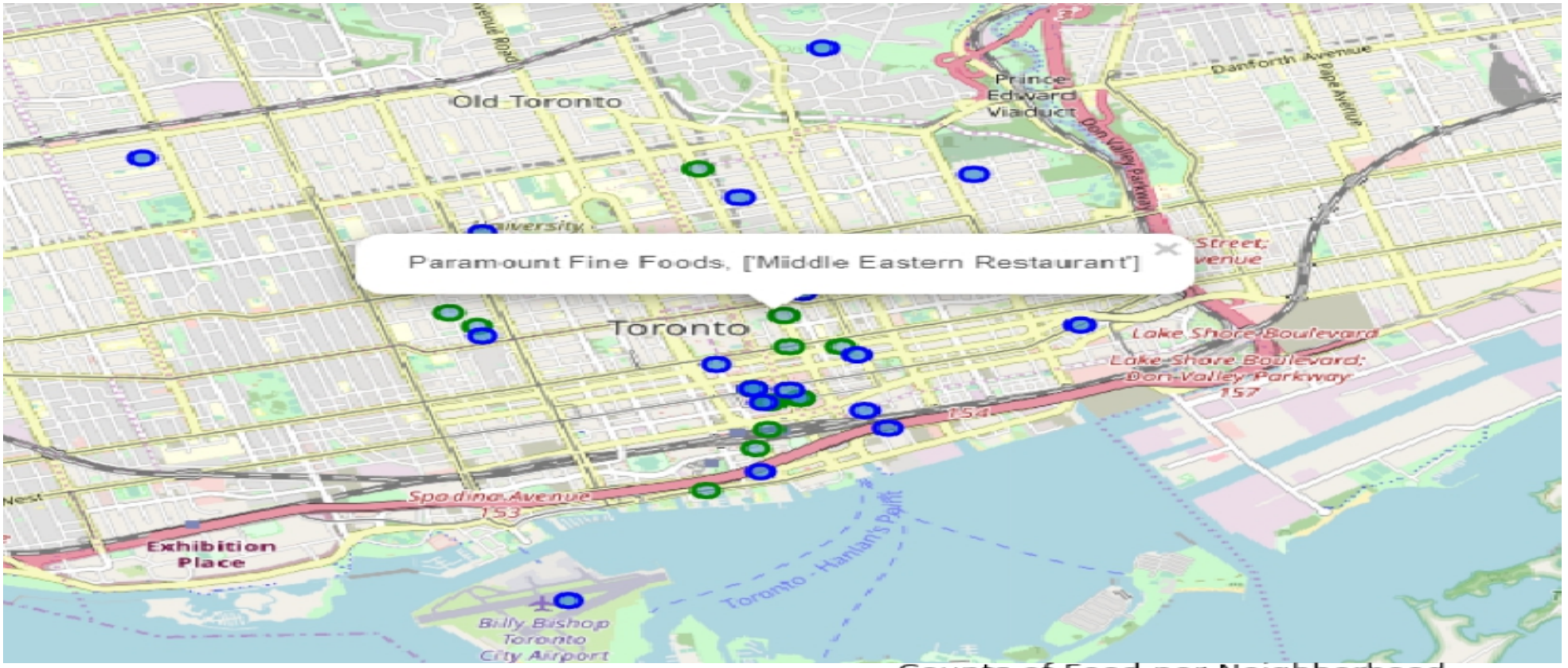
- You can see with this histogram, which categories has more lenght of venues. And one categories has 120 venues. This is a most popular categories of venues in Downtown !

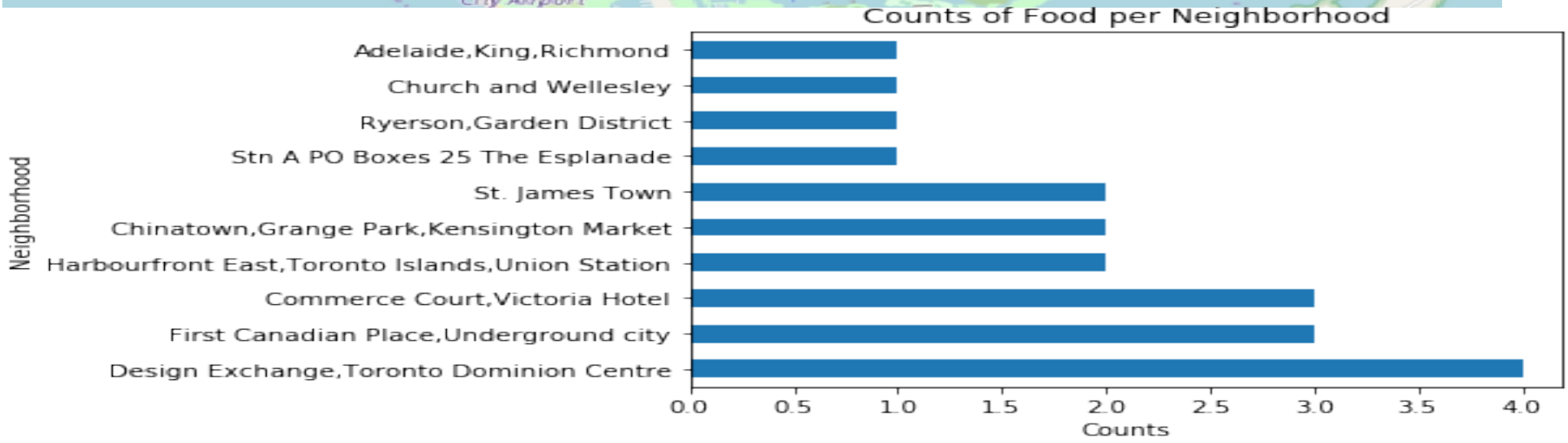


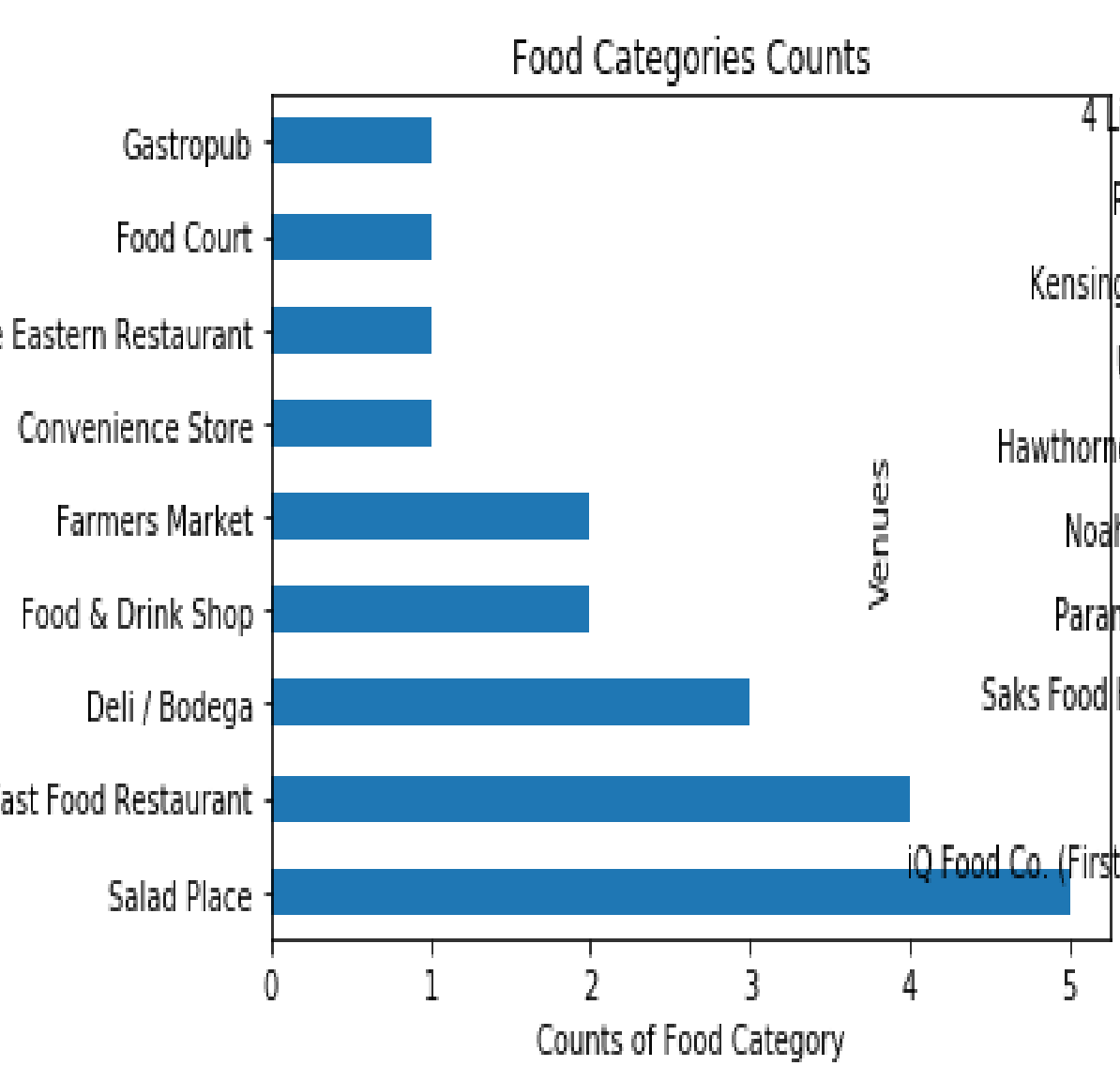
Most popular Eat venues (indirect/direct)



Most popular Eat venues (indirect/direct)

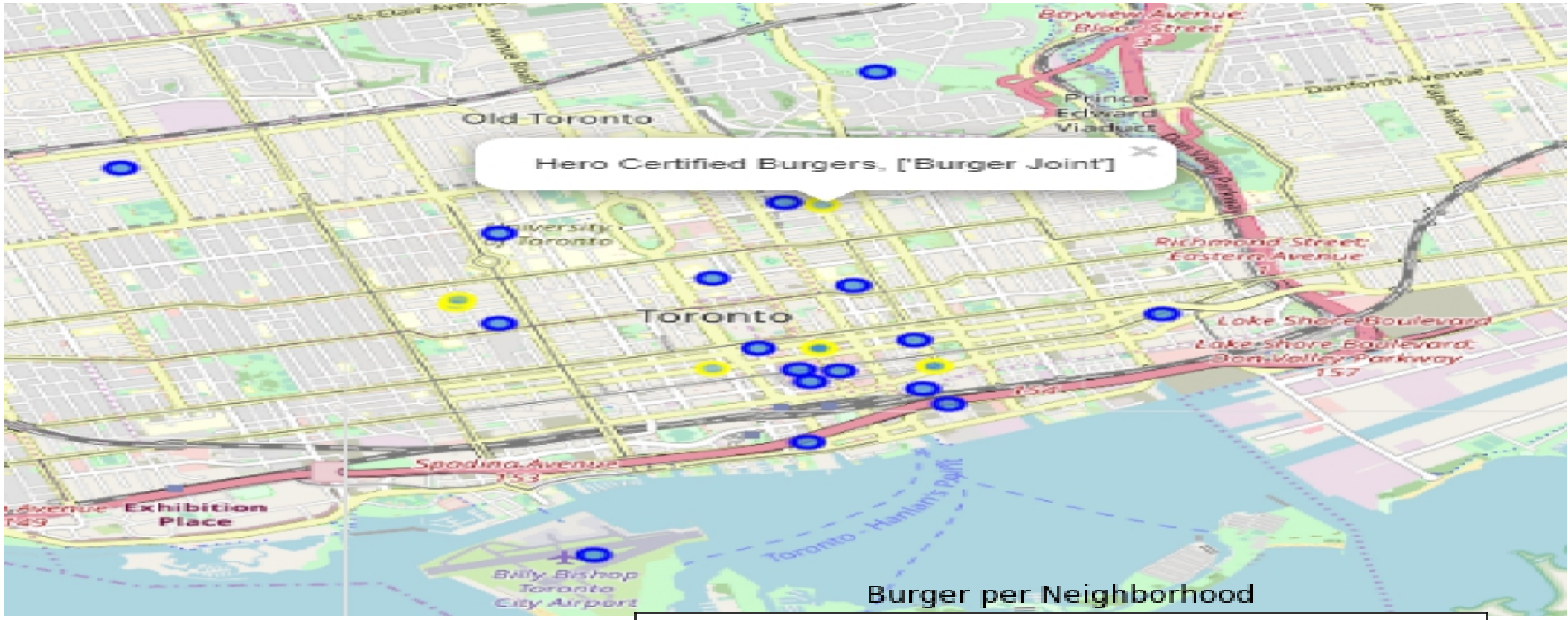


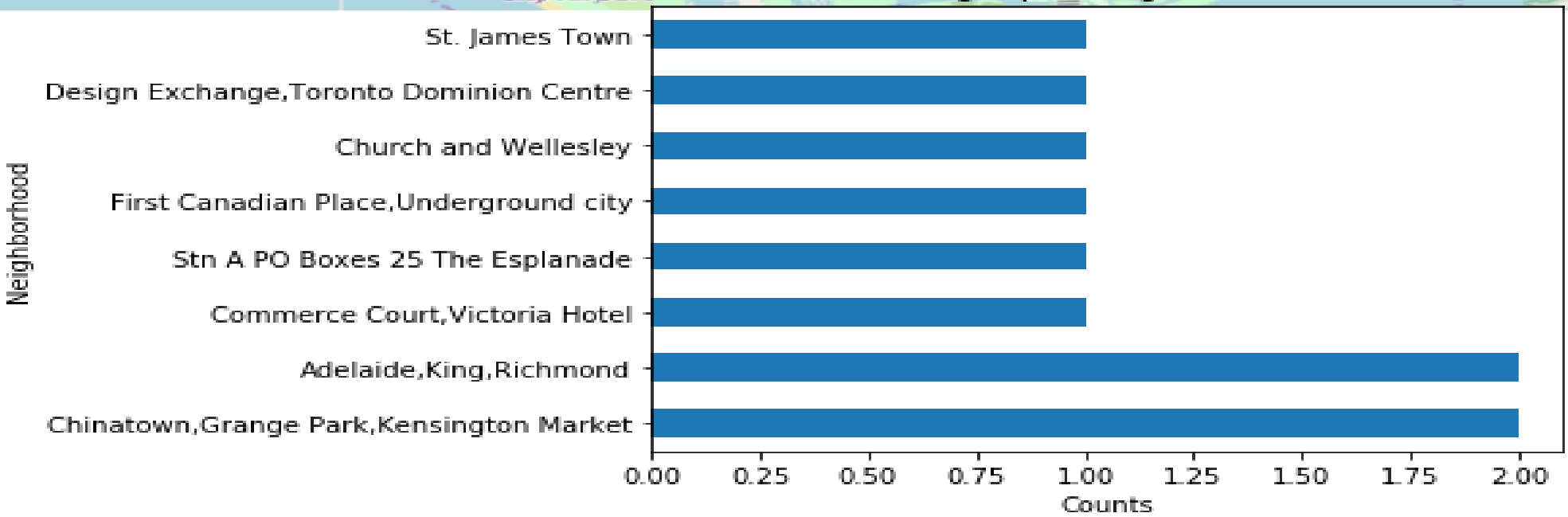



Food Categories Counts
Gastropub
Food Court
Eastern Restaurant
Convenience Store
Farmers Market
Food & Drink Shop
Deli / Bodega
ast Food Restaurant
Salad Place
Venues
0
1
2
3
4
5
Counts of Food Category

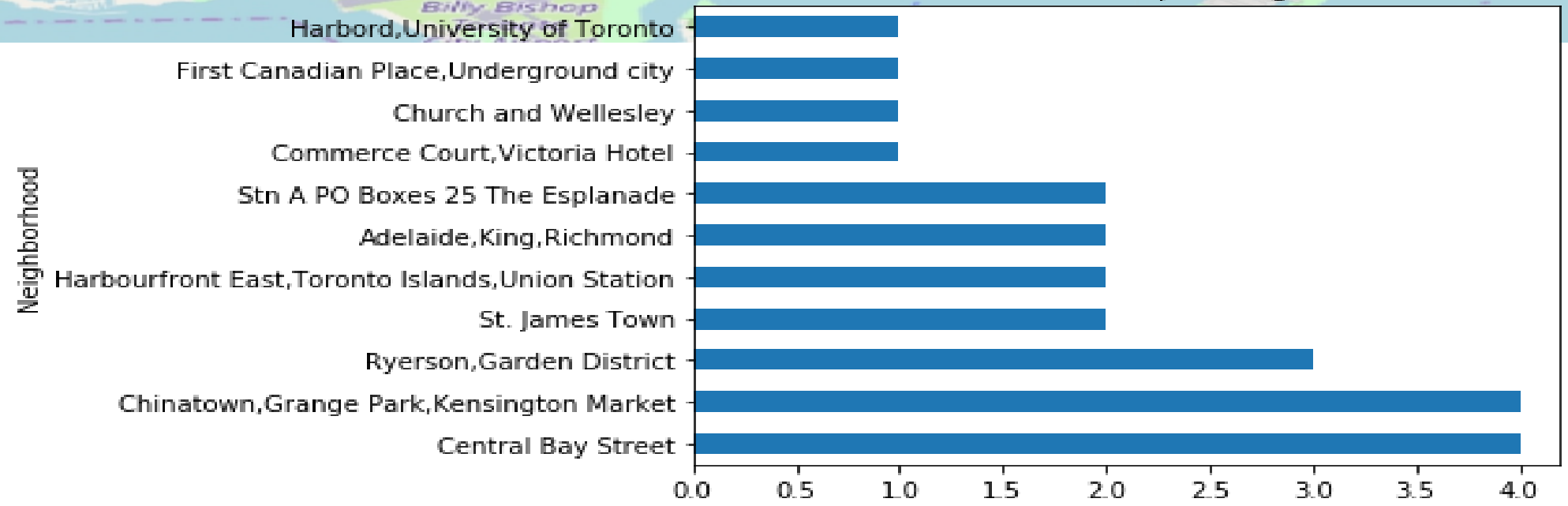
Counts of Venues Food
4 Life Natural Foods
Rabba Fine Foods
Kensington Food Market
Union Food Court
Hawthorne Food and Drink
Noah's Natural Foods
Paramount Fine Foods
Saks Food Hall by Pusateri's
McEwan Foods
iQ Food Co. (First Canadian Place)
iQ Food Co
0
1
2
3
Counts

Most popular Burger (indirect)

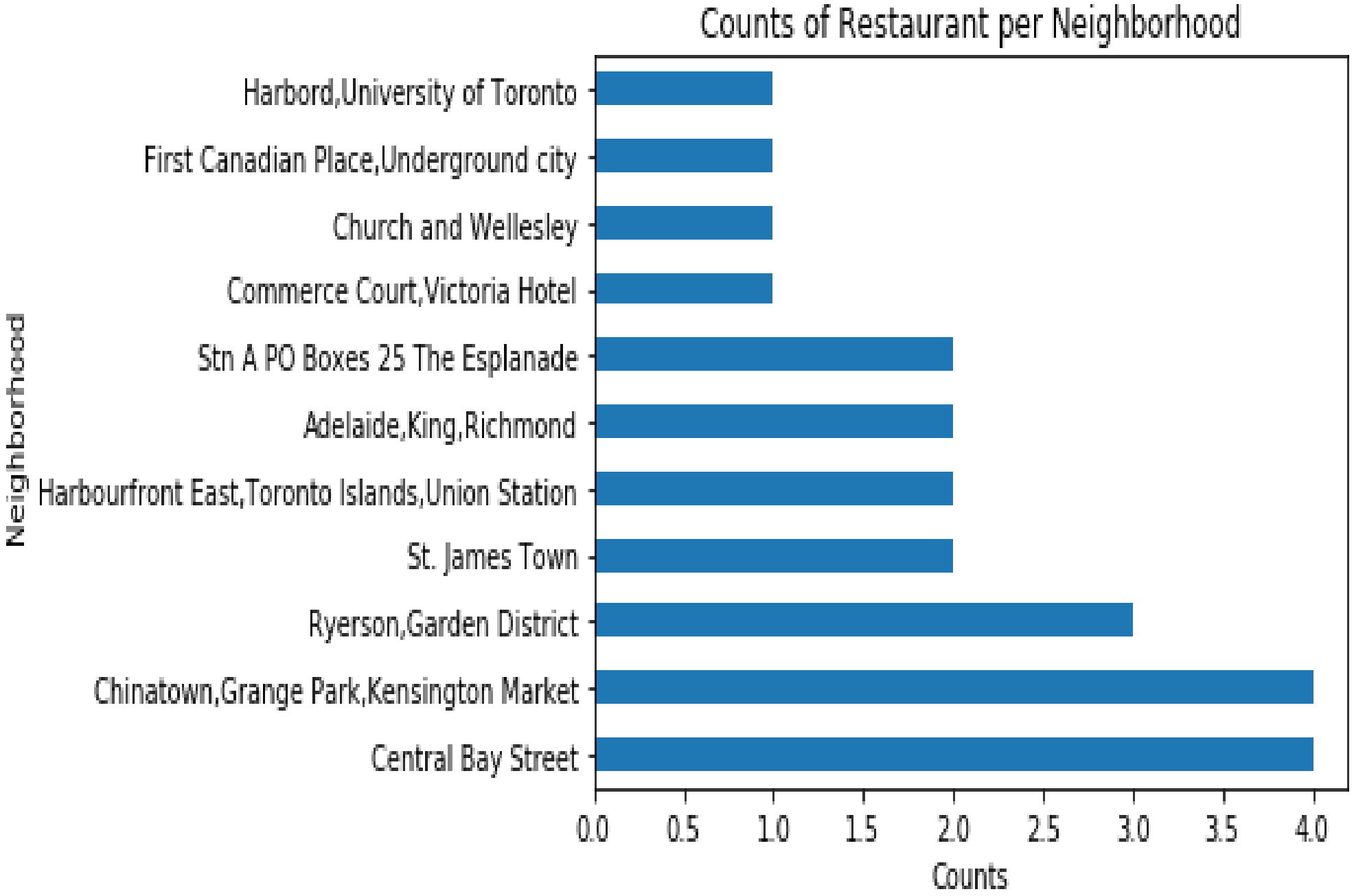




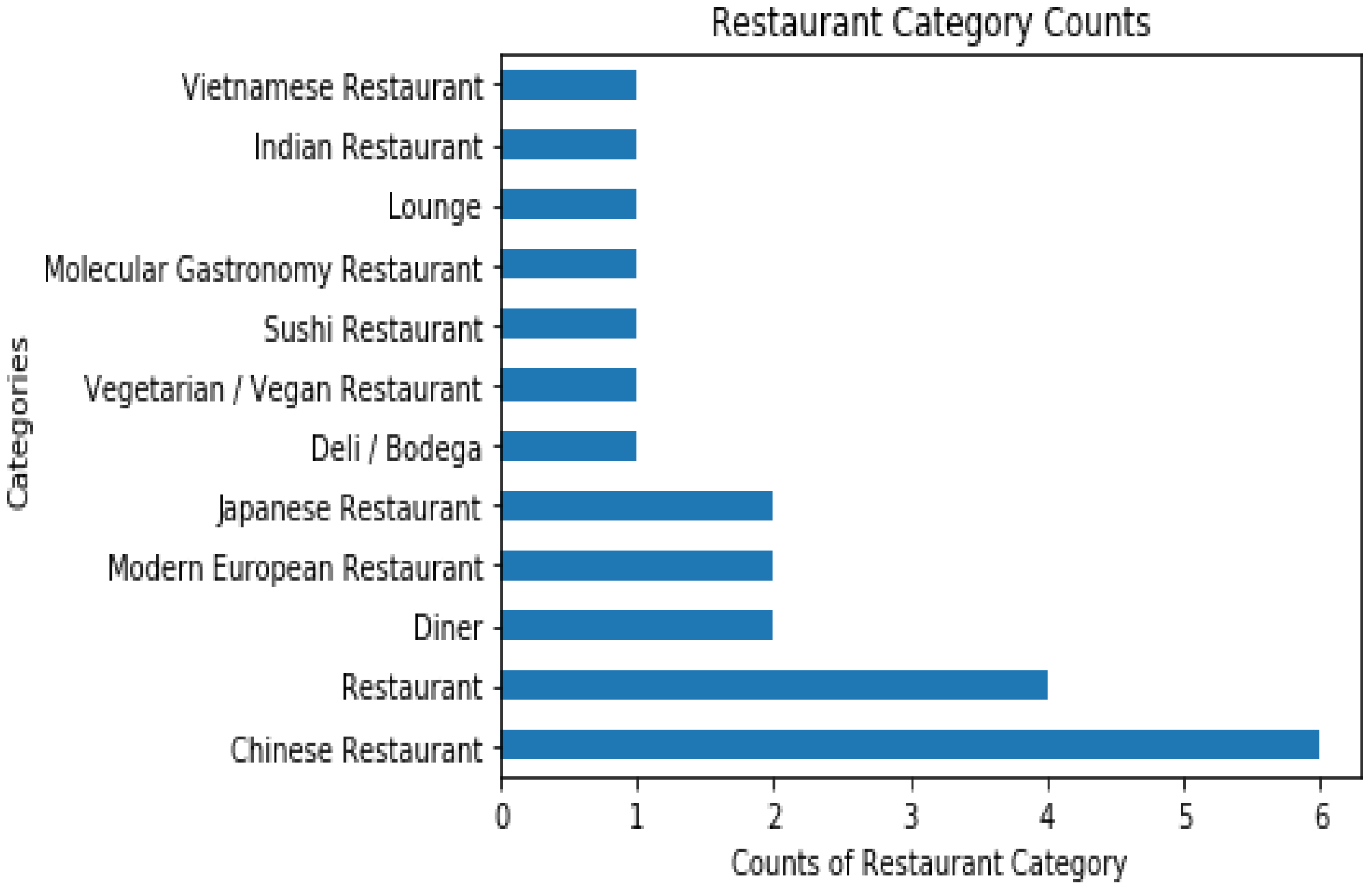
Restaurant (direct rivalry)



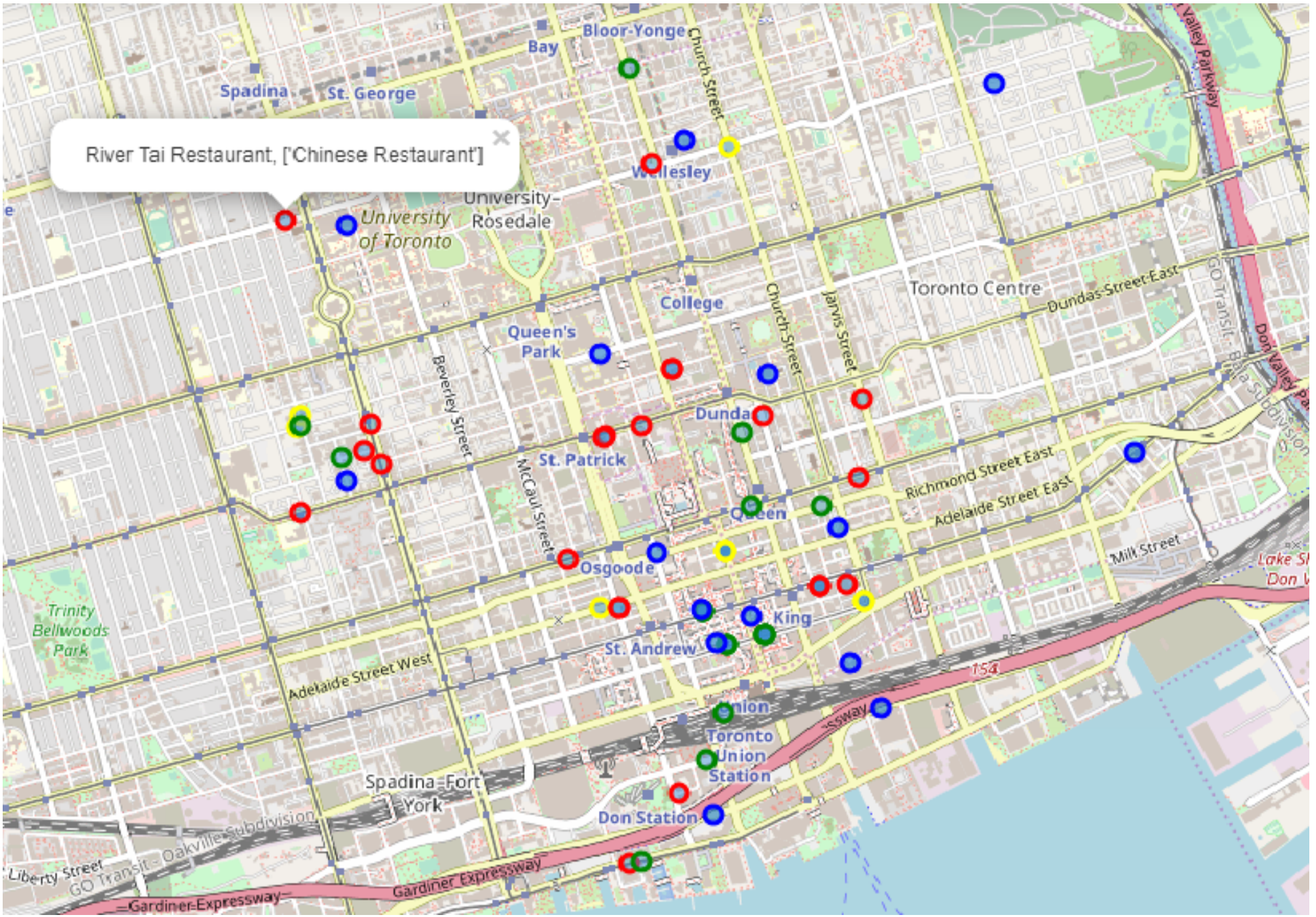
Bar of Restaurant per Neighborhood



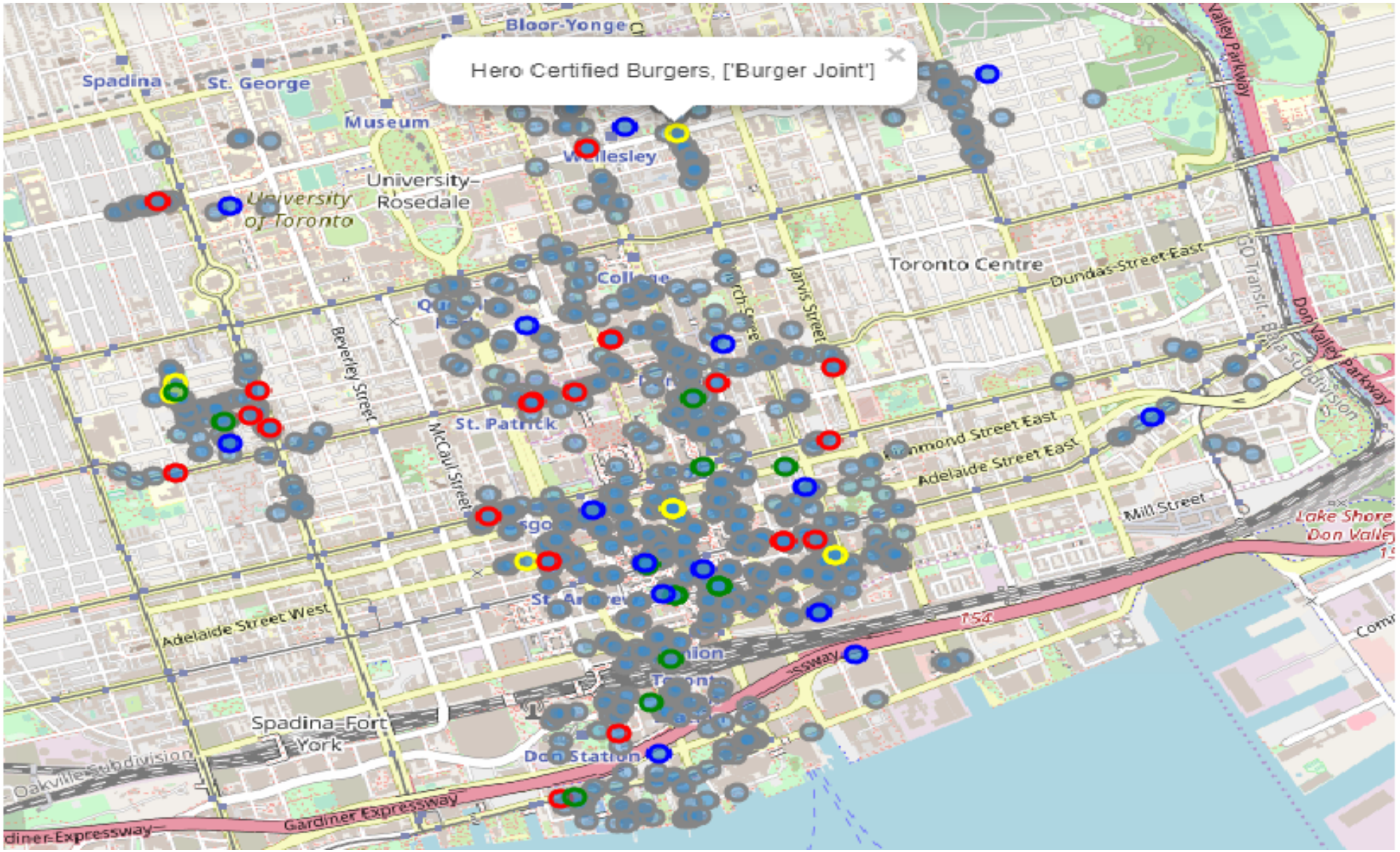
Restaurant categories



Map Burger + Food + Restaurant



Map Burger + Food + Restaurant + All venues



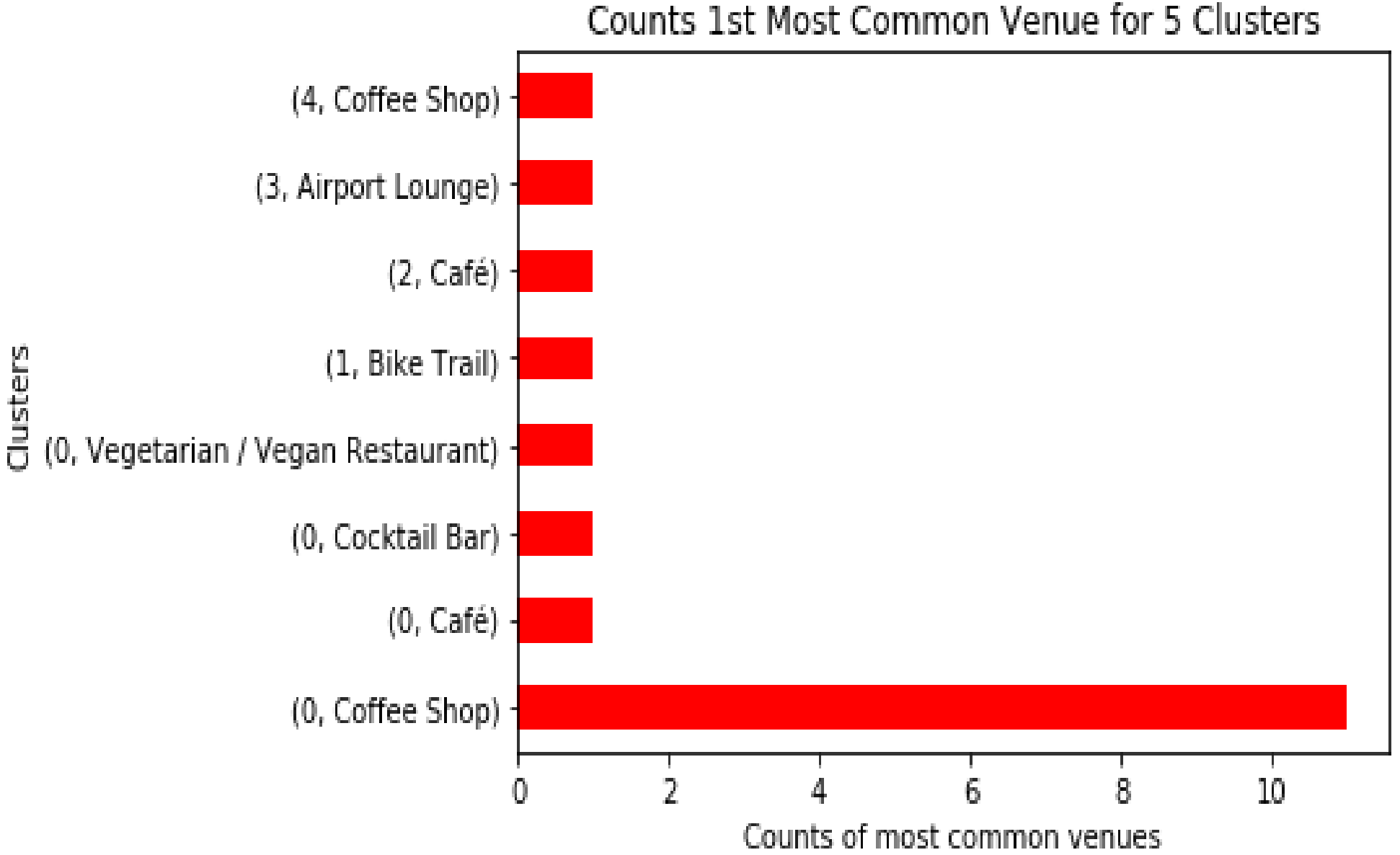
Clustering process

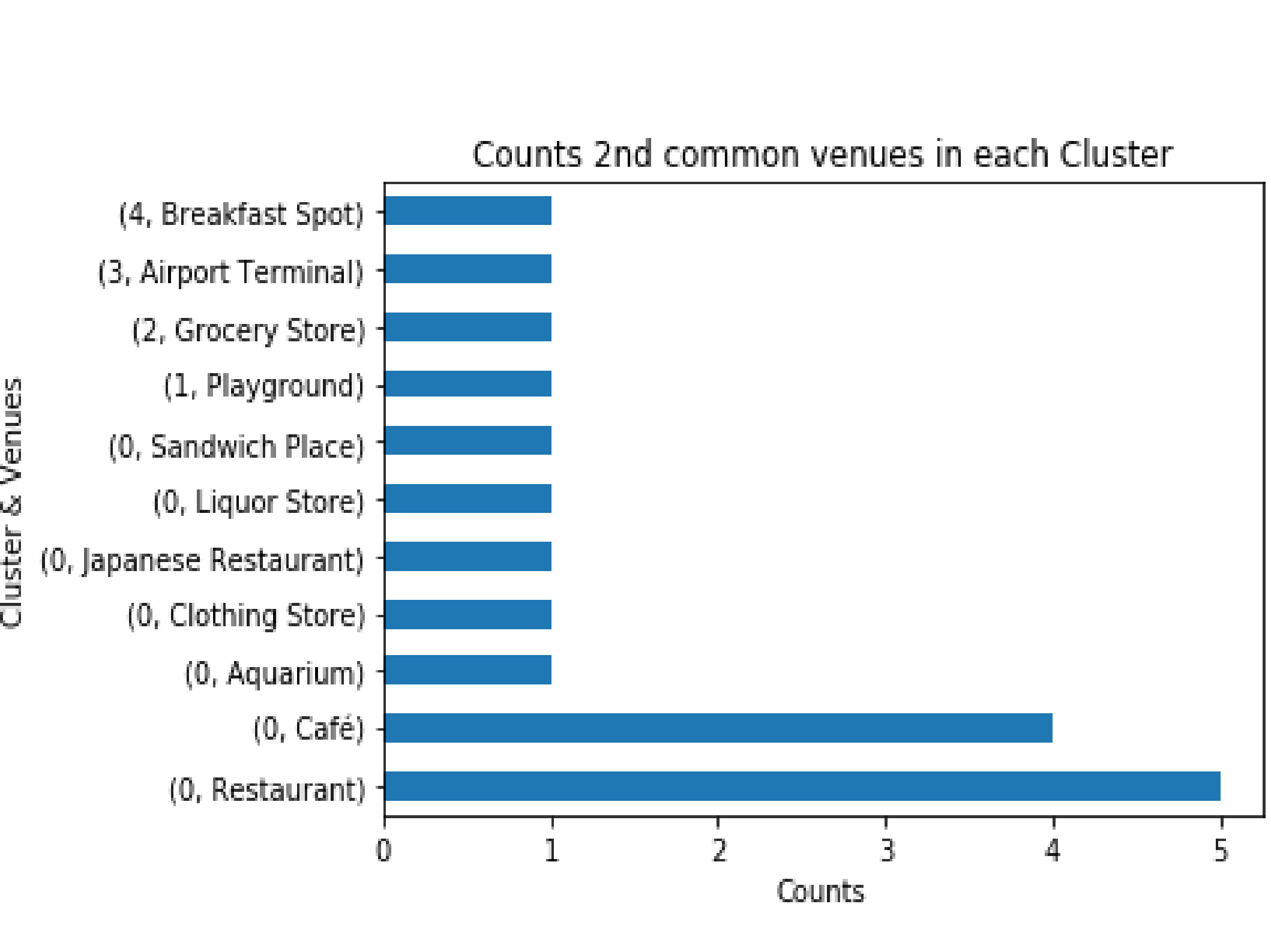
	Neighborhood	Wings Joint	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Art Gallery
0	Adelaide,King,Richmond	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.038462	0.000000
1	Berczy Park	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
2	CN Tower,Bathurst Quay,Island airport,Harbourf...	0.000000	0.1	0.1	0.1	0.2	0.1	0.1	0.000000	0.000000
3	Cabbagetown,St. James Town	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.040000	0.000000

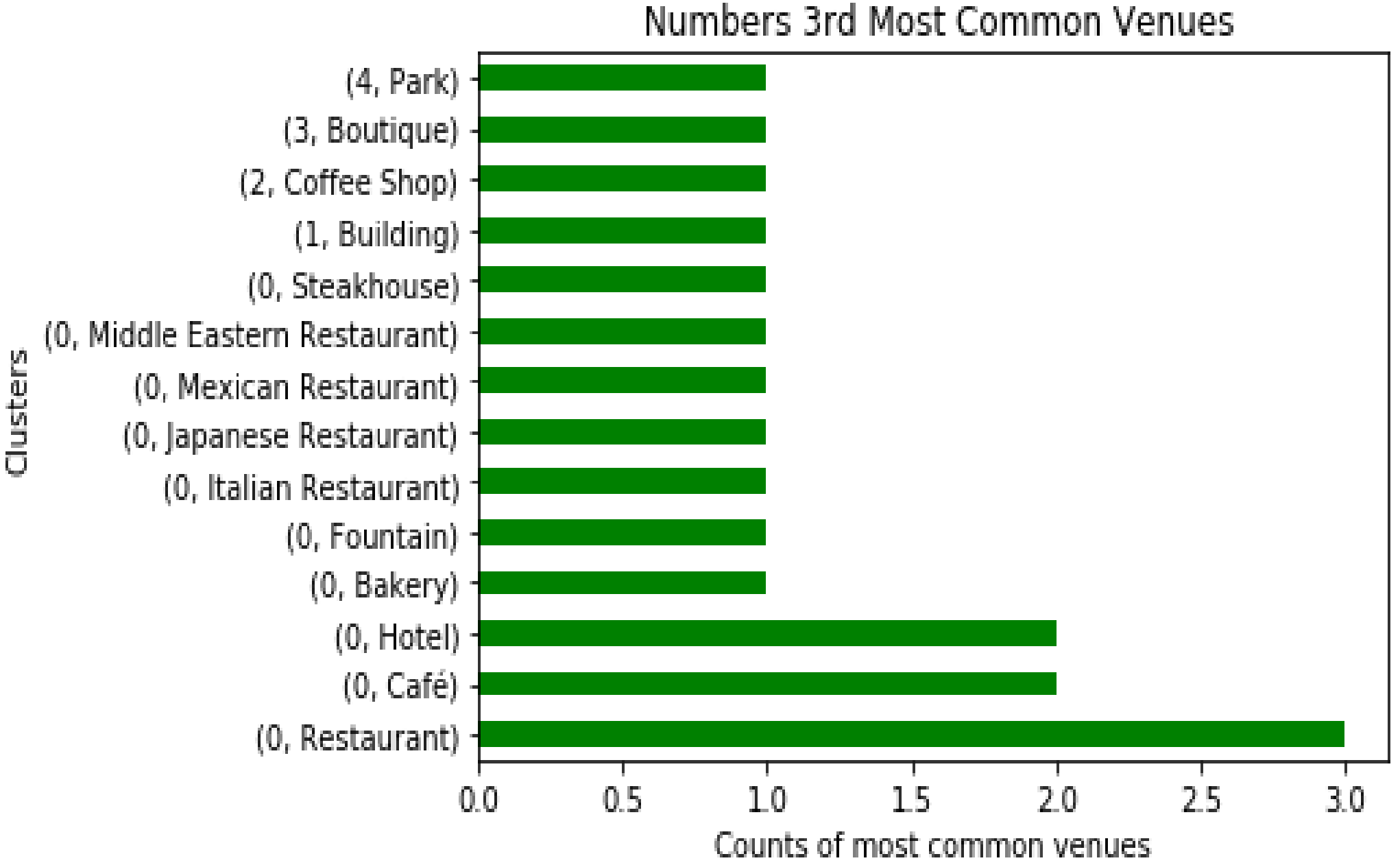
	Neigh	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
0	Adelaide,King,Richmond	Coffee Shop	Japanese Restaurant	Steakhouse	
1	Berczy Park	Cocktail Bar	Liquor Store	Fountain	
2	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Lounge	Airport Terminal	Boutique	Co
3	Cabbagetown,St. James Town	Coffee Shop	Restaurant	Café	Pi
4	Central Bay Street	Coffee Shop	Sandwich Place	Italian Restaurant	

Table with Clustering Labels

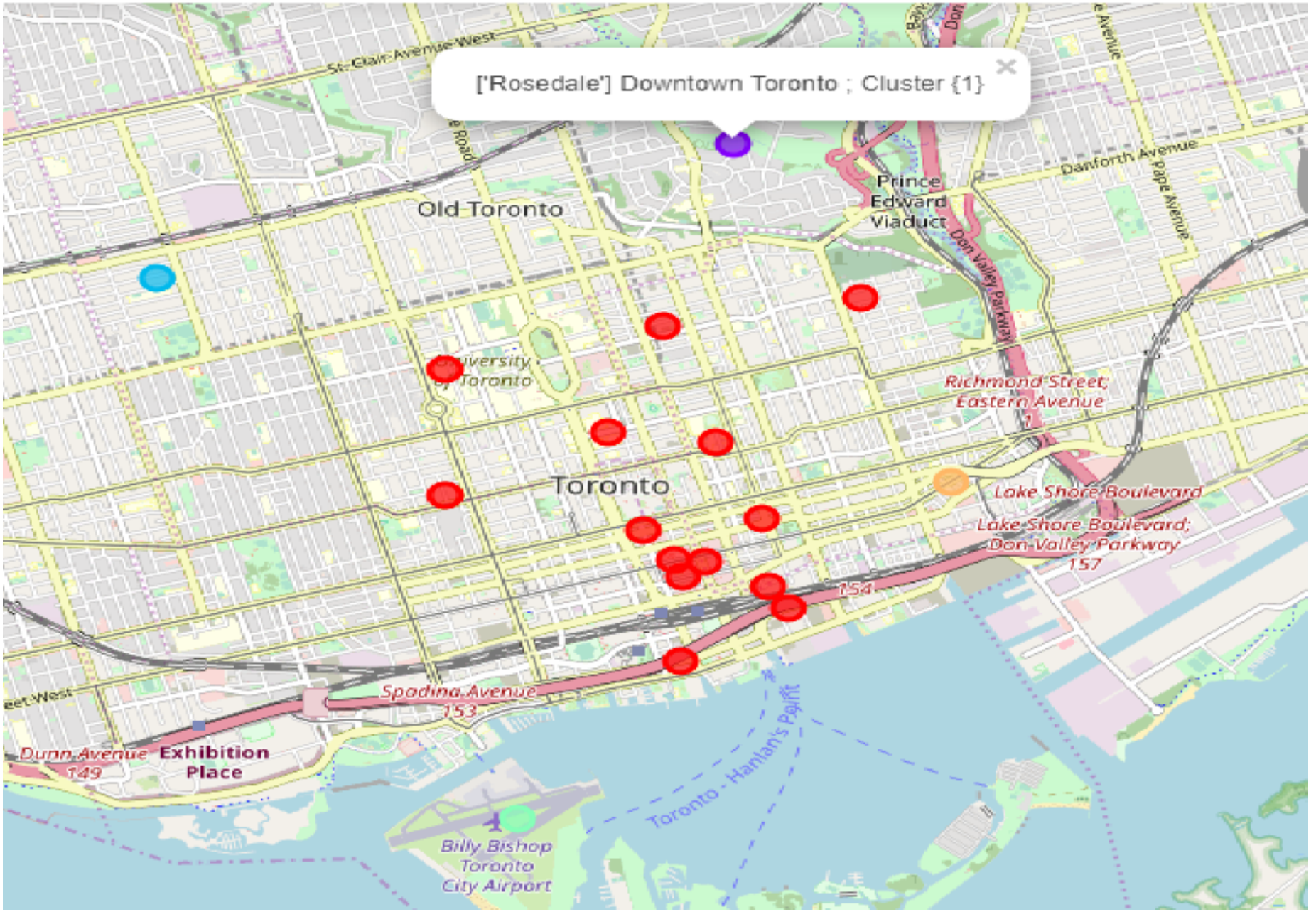
	Post	Borough	Neigh	Latitude	Longitude	ClustLabels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
50	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	1	Bike Trail	Playground	Building
51	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675	0	Coffee Shop	Restaurant	Café


Counts 1st Most Common Venue for 5 Clusters
(4, Coffee Shop)
(3, Airport Lounge)
(2, Café)
(1, Bike Trail)
(0, Vegetarian / Vegan Restaurant)
(0, Cocktail Bar)
(0, Café)
(0, Coffee Shop)
Clusters
0
2
4
6
8
10
Counts of most common venues

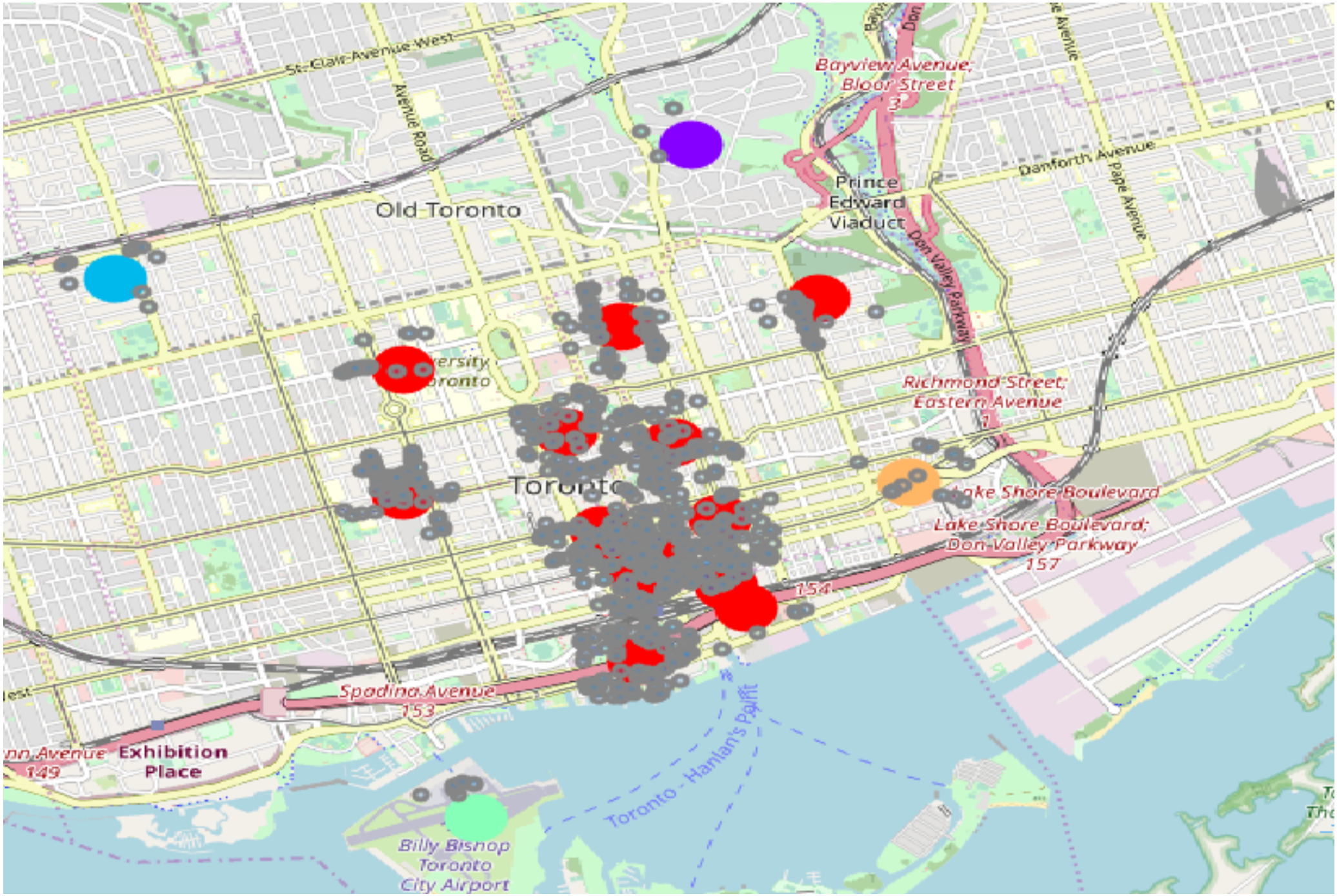

Counts 2nd common venues in each Cluster
(4, Breakfast Spot)
(3, Airport Terminal)
(2, Grocery Store)
(1, Playground)
(0, Sandwich Place)
(0, Liquor Store)
(0, Japanese Restaurant)
(0, Clothing Store)
(0, Aquarium)
(0, Café)
(0, Restaurant)
Cluster & Venues
0
1
2
3
4
5
Counts


Numbers 3rd Most Common Venues
Clusters
(4, Park)
(3, Boutique)
(2, Coffee Shop)
(1, Building)
(0, Steakhouse)
(0, Middle Eastern Restaurant)
(0, Mexican Restaurant)
(0, Japanese Restaurant)
(0, Italian Restaurant)
(0, Fountain)
(0, Bakery)
(0, Hotel)
(0, Café)
(0, Restaurant)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Counts of most common venues

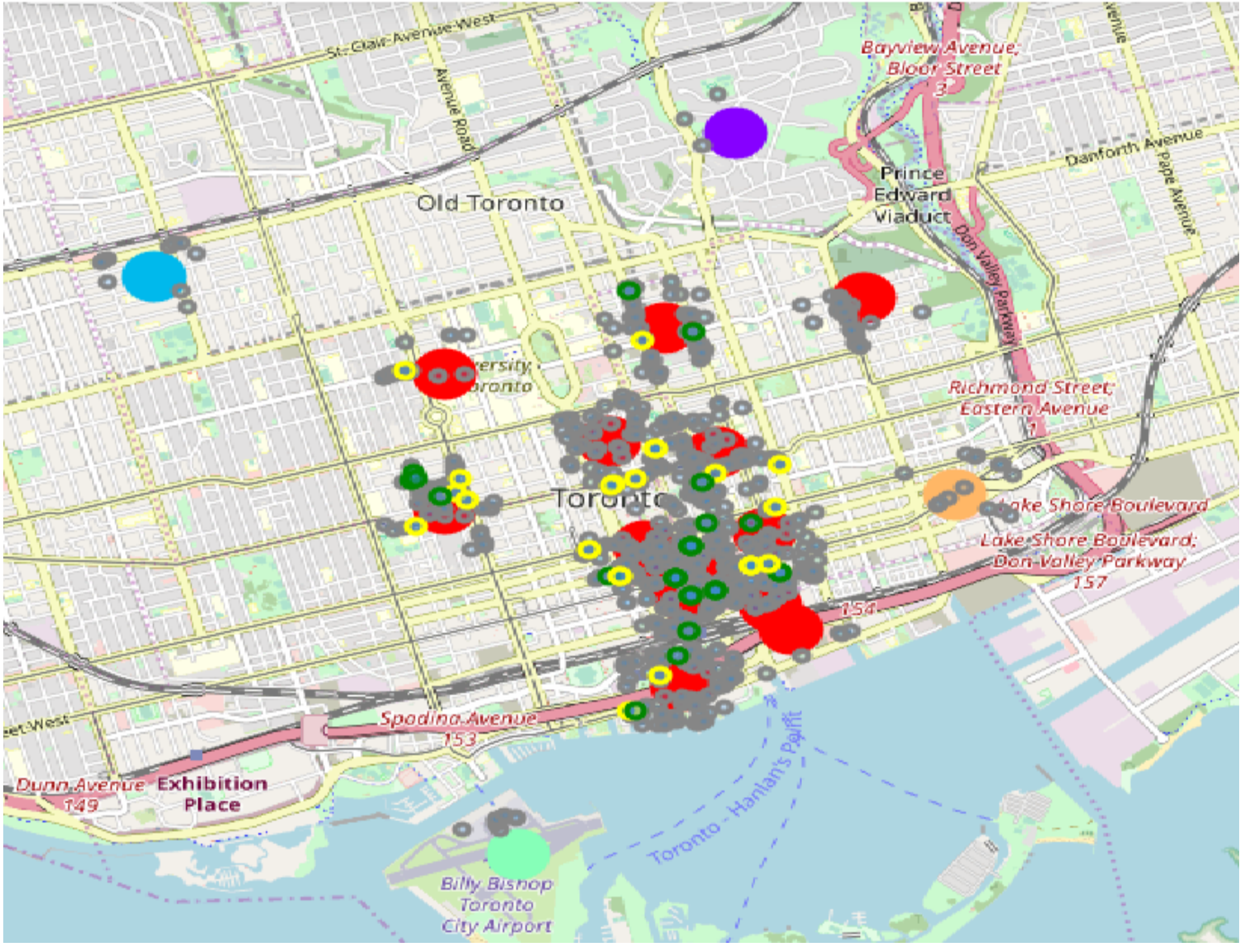
Map of Borough Clustering



Map Cluster + all venues



Map Cluster + Indirect & Direct rivalry + all venues



Explore the Cluster 1 : Shape cluster

	Borough	Neigh	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
count	14	14	14	14	14	14	14	14	14	14	14	14
unique	1	14	4	7	10	12	12	14	12	11	13	12
top	Downtown Toronto	Ryerson,Garden District	Coffee Shop	Restaurant	Restaurant	Café	Chinese Restaurant	Deli / Bodega	Hotel	Bubble Tea Shop	Seafood Restaurant	Beer Bar
freq	14	1	11	5	3	3	2	1	3	2	2	2

- This cluster contains 14 borough, located in the town center. He colored with red in the map.
- The cluster are the Coffee in Most Common Venue.after you see the top venues are Restaurant, Café and foreign(Italian, Mexican, Japanese)

  This segment is for primary commodities and needs
- With maps, you can see popular spots and eat venues are primarly near this cluster. This satured in venues, but the trendings and top venues are highest score ! Goals : rejected this cluster.

Explore others Cluster (each contains 1 borough)

- Cluster 2 : The most common venue is a Bike Trail and Playground

  This segment is for Sport !

- Cluster3 : This cluster it's for the services of all commodities and go to Night Club !

- Cluster 4 : This place are a local space in the Airport, with all Airport services !

- Cluster 5 : Most commons venues are a Coffe Shop, Breakfast Sport, Park, Spa..¶

  Place used for well-being near the town center and the Lake !

Rename the clusters with best venues frequency

- 1 = Commodities
- 2= Sport
- 3= Nigh Club
- 4= Airport
- 5= Well place

Discussion and recommendations

If you want to minimize the concurencies :

- You can choose the 4 Clusters (Sport, Night Club, Airport, Well place) for open you Restaurant.

Arguments :

1) You don't have popular restaurant venues and indirect concurencies surroundings area. And all venues are next to cluster 1.
2) Also, the environnement of this Cluster are more quiet, surroundings Lake, Park, Nature..

Discussion

If not choose the Cluster 1, for more advice for open shop in others Cluster, I can :

- Recolted more venues for this Borough
- Trendings venues around it.
- Search eat venues, ratings, tips...

- More Data for explore the local population and tourism of this Borough (demographics, education, cultural...)