MOKHTAR-SLIMANE Ghais

IBM Coursera Capstone Project

Create a new business in Toronto

PS :

Firslty, I want to give my gratitude, for the Data Scientist who create the IBM Data Science Professionnal Certificate specialization, and contenues of 9 courses for obtain that.

The methodology, pedagogies, & explanation of each courses are very values to obtain theory & skills. Each videos & Lab are appreciables, motivating, inspiring and full of interest !!!

Thanks you very much !!

Table of Contents

Create a new business in Toronto.

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. Business understanding

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 : Others decisions in choose for what the stakeholders want

1. Look venues around each Neighborhood.
2. Segment the Borough and their Neighborhood within the Most common venues in each (populars spots).

Choose the not used categories (new shop category in this area)

1. Minimized the direct concurencies. b. Minimized the indirect concurencies

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

If statement :

If after this objectives accomplish I'm not able to find the best Neighborhood, I can mining & explore more data, variables & features for obtain the best conclusion.

A) 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)

B) 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.

Old Toronto is the Cardinal Borough of the City : East, West, Downtown, and Central Toronto.

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 over 25 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 :

a) 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. To obtain a table like this :

PostCode	Borough	Neighborhood	latitude	longitude
M5S	Downtown Toronto	Harbord,University of Toronto	43.6627	-79.4
M5T	Downtown Toronto	Chinatown,Grange Park...	43.6532	-79.4

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.

Neigh	1st Most Common Venue	2nd Most Common Venue	3rd Most Common
Adelaide,King..	Coffee Shop	Café	Thai Restaurant
Berczy	Coffee Shop	Park Bakery	Cocktail Bar
CN Tower..	Airport Service	Airport Lounge	Airport Terminal

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

2.3 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.

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. Methodologies

3.1 Data Wrangling

Table 1

- Firstly , I have drop the Borough with « Not Assigned ».

For the 'Neighborhood' missing values, I have select the 'Borough' name in the same index.

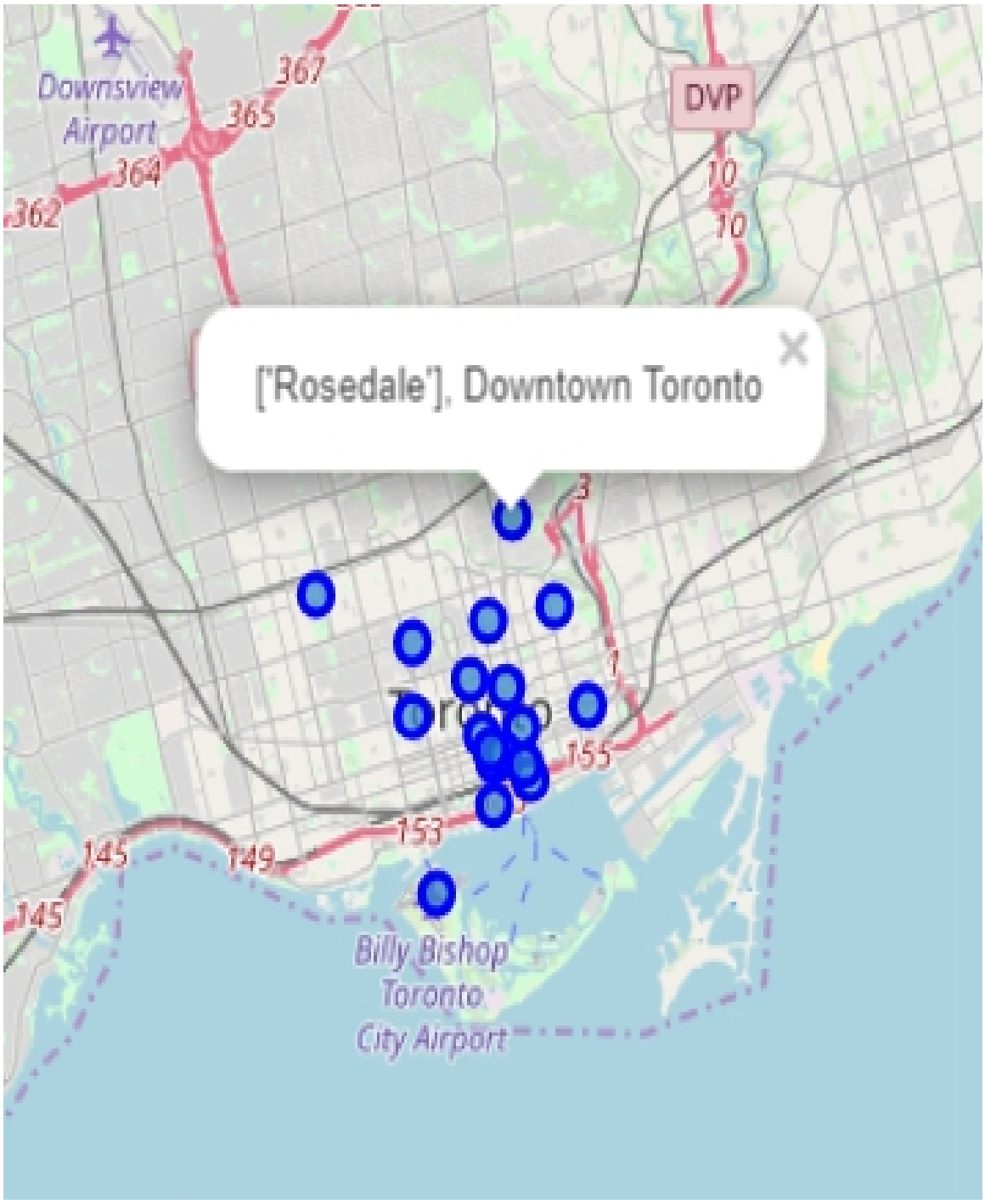
- After, I have group by Postal Code & Borough, and join each Neighborhood with ','.join .

And I obtain the first table with shape (103 rows, 3 columns)

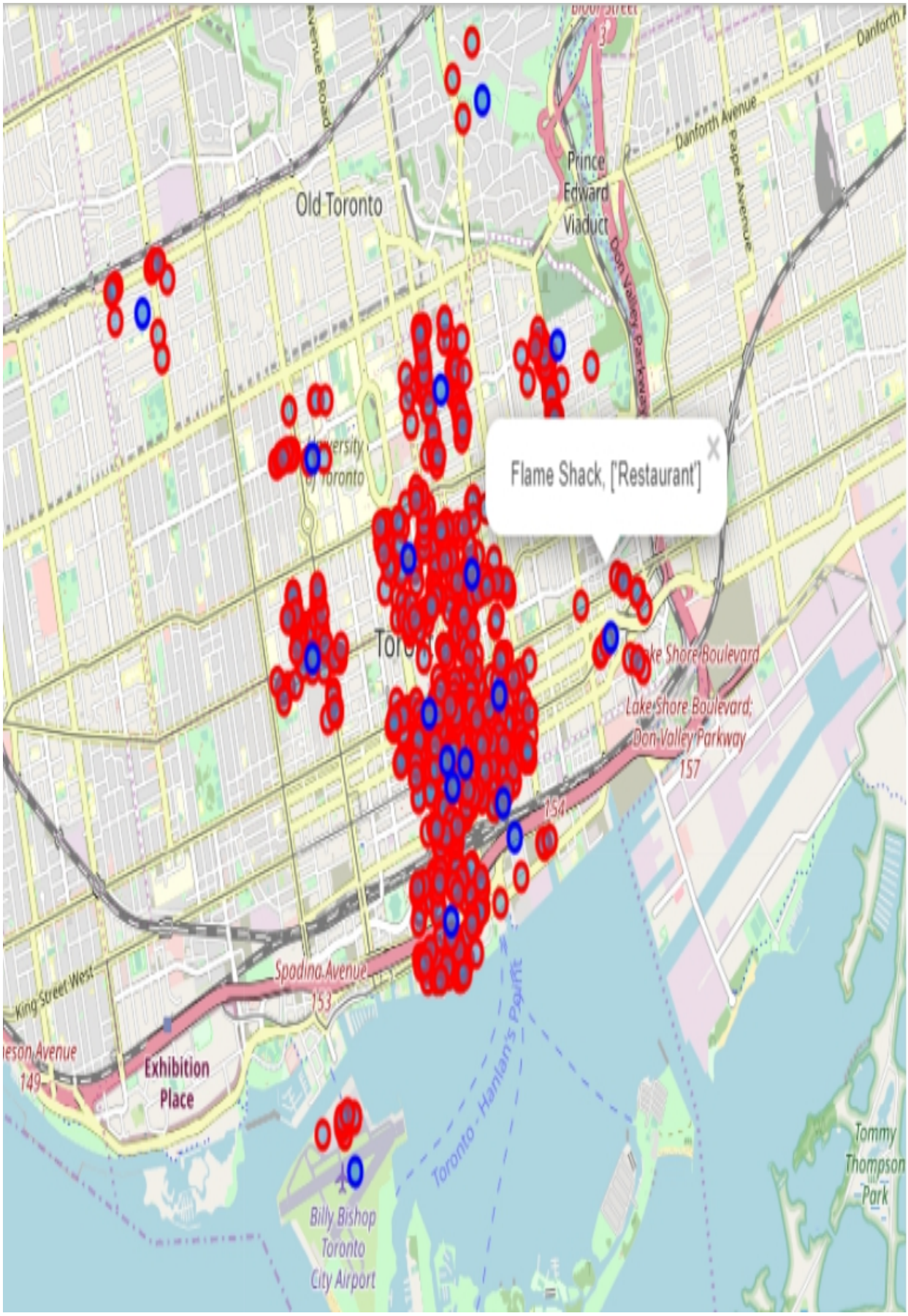
- I have download the csv of coordinates Toronto, and concat the tables of obtain the Longitutes and Latitudes for each Borough
(103,5)

3.2 Exploratory Data Analysis

Map of Downtown Toronto Borough

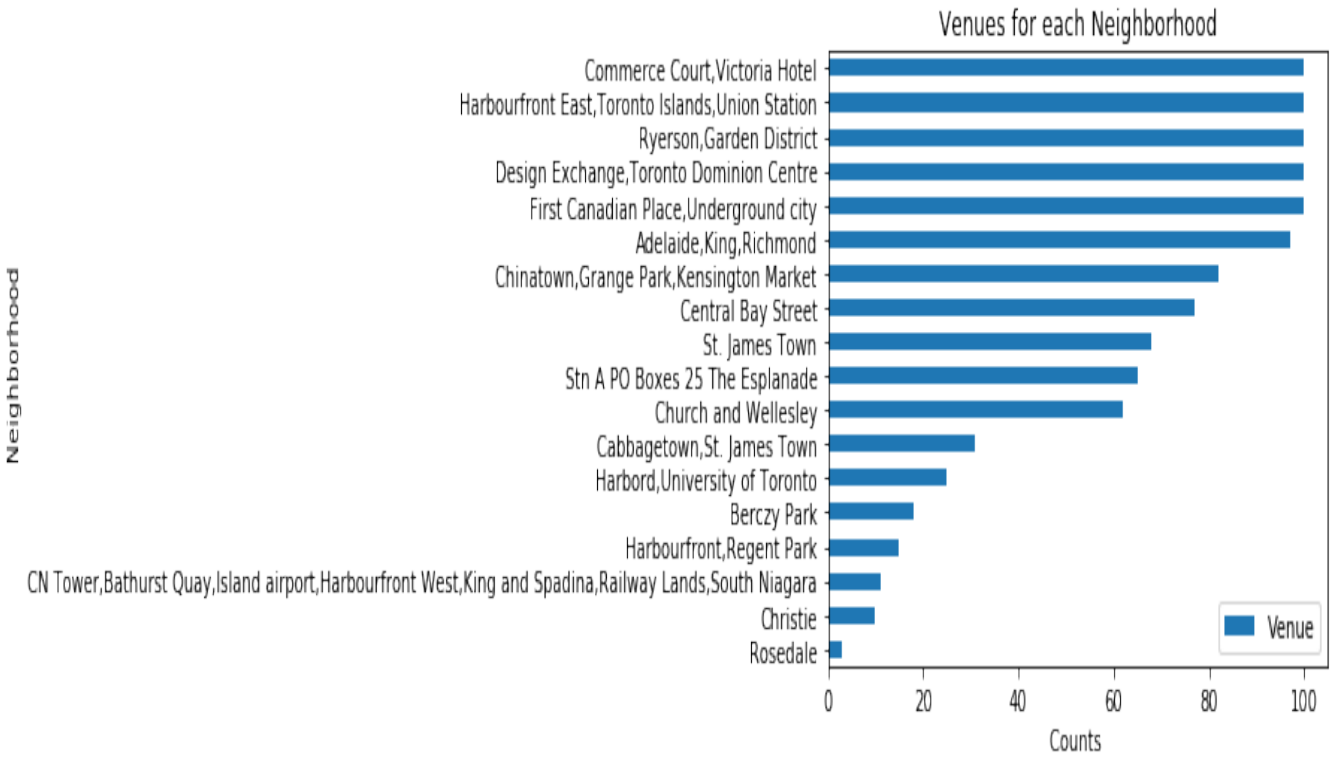


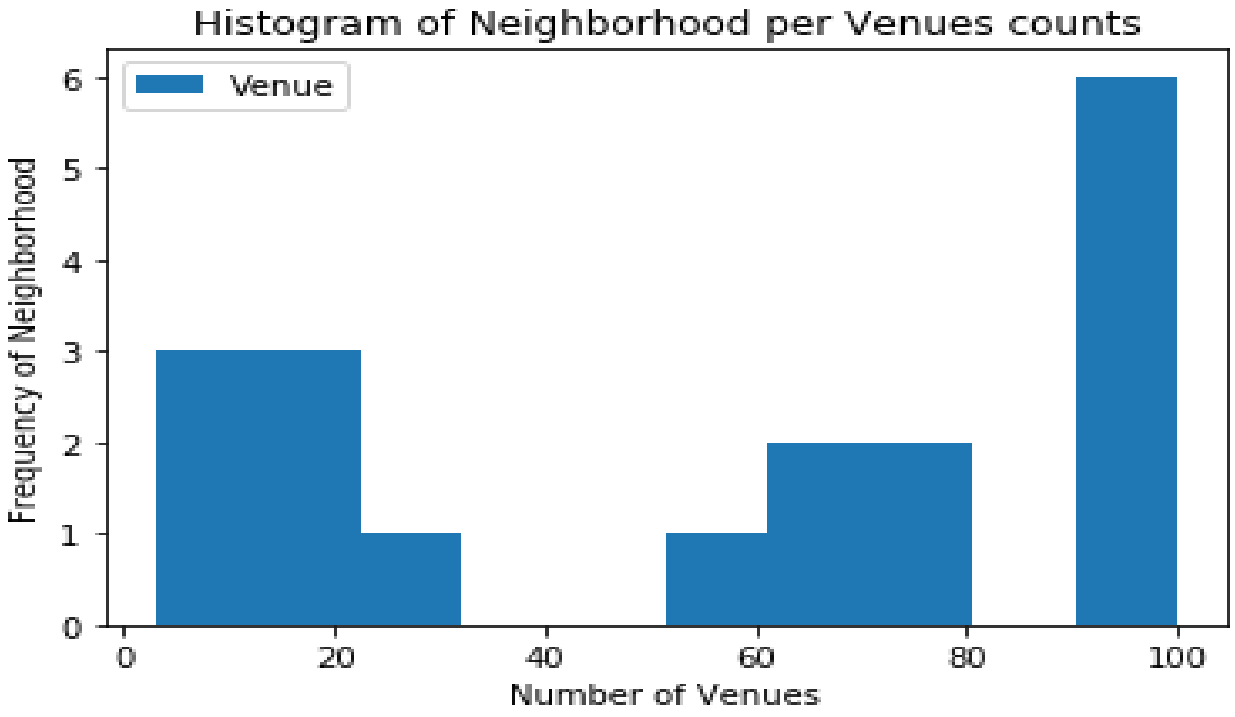
Map of the All Popular venues around 500m for each Downtown Toronto.

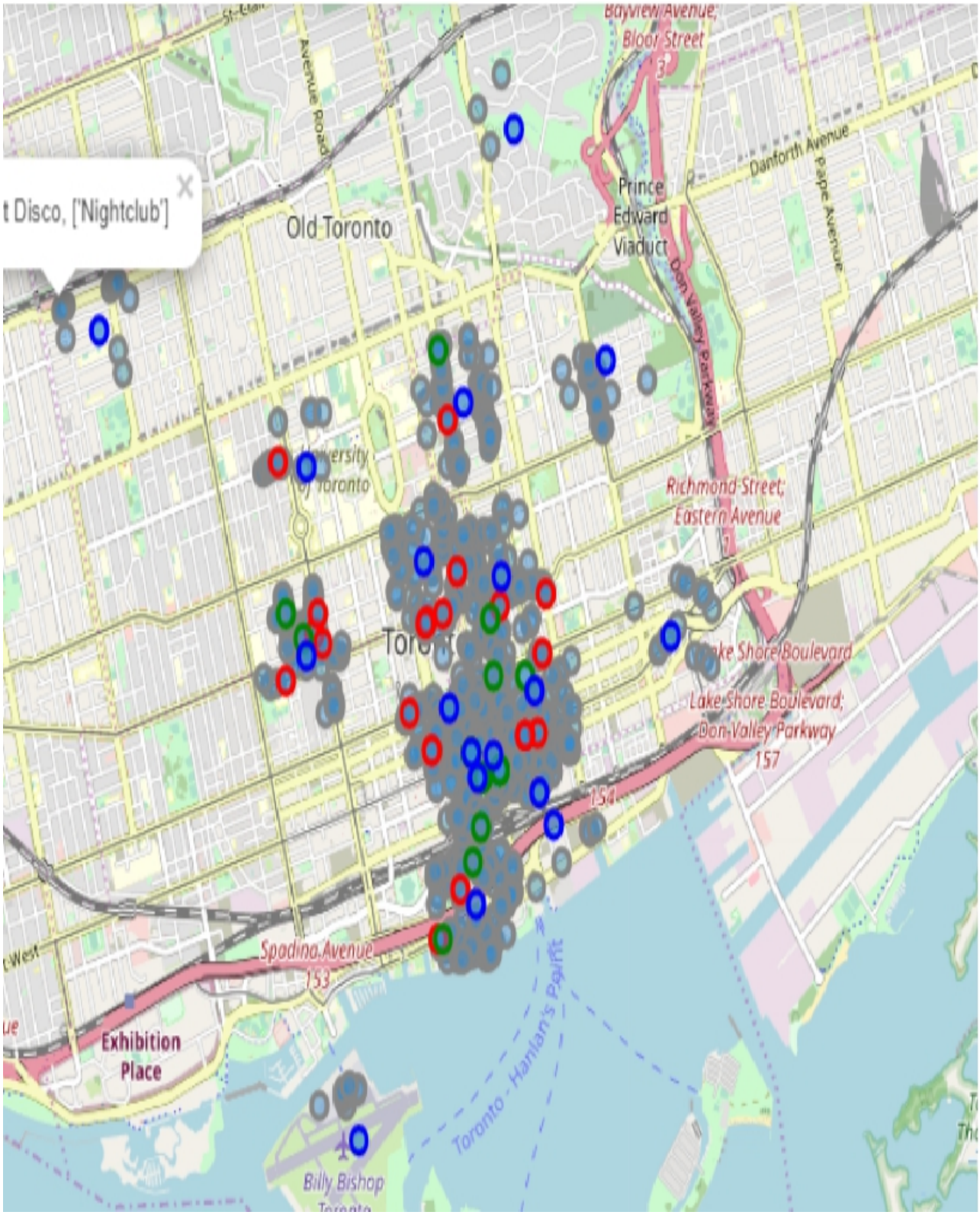


I have 185 unique categories & 659 uniques venues in this table and map.

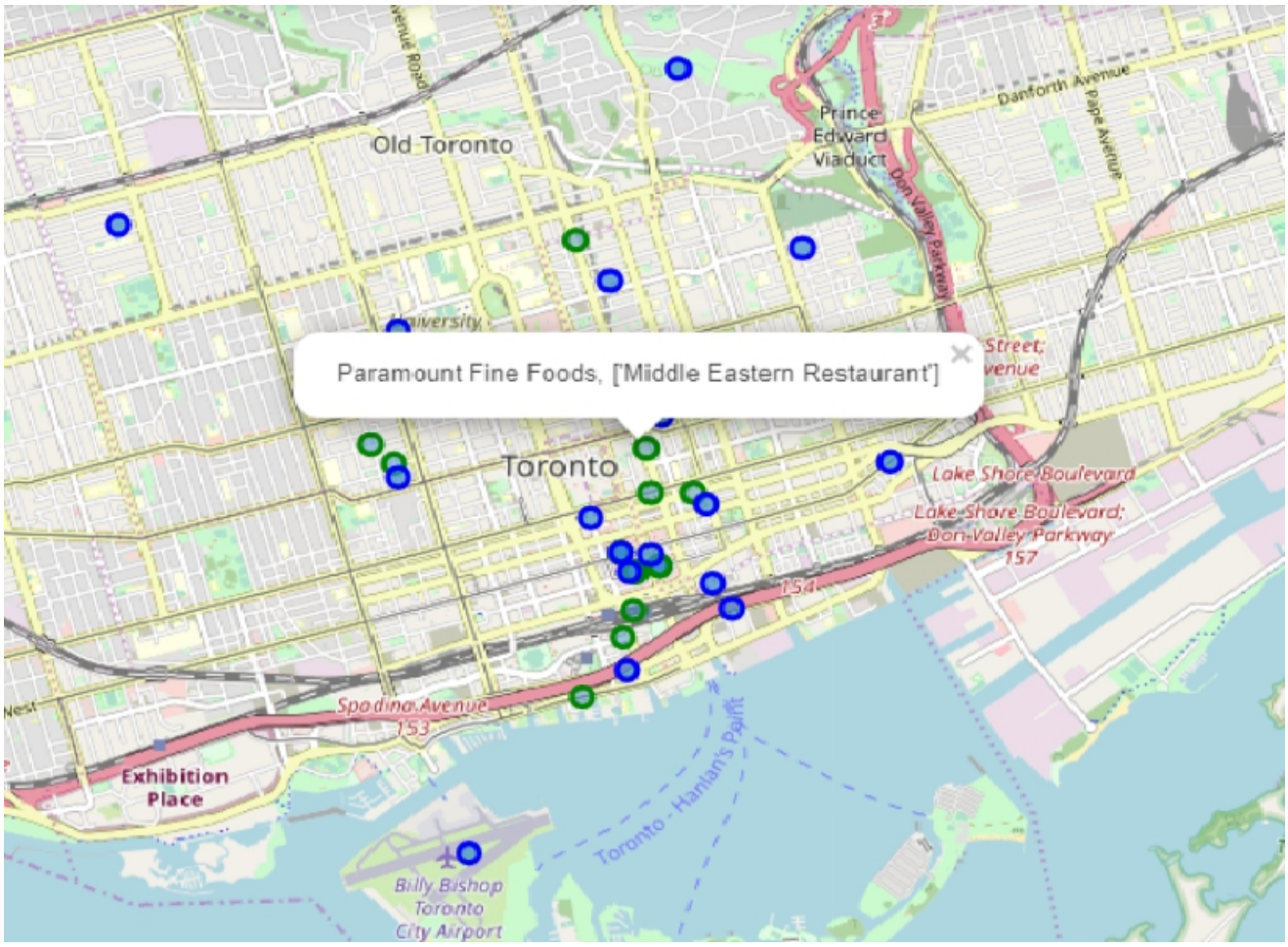
With this Graph, I can see the 6 borough who are satured for venue.

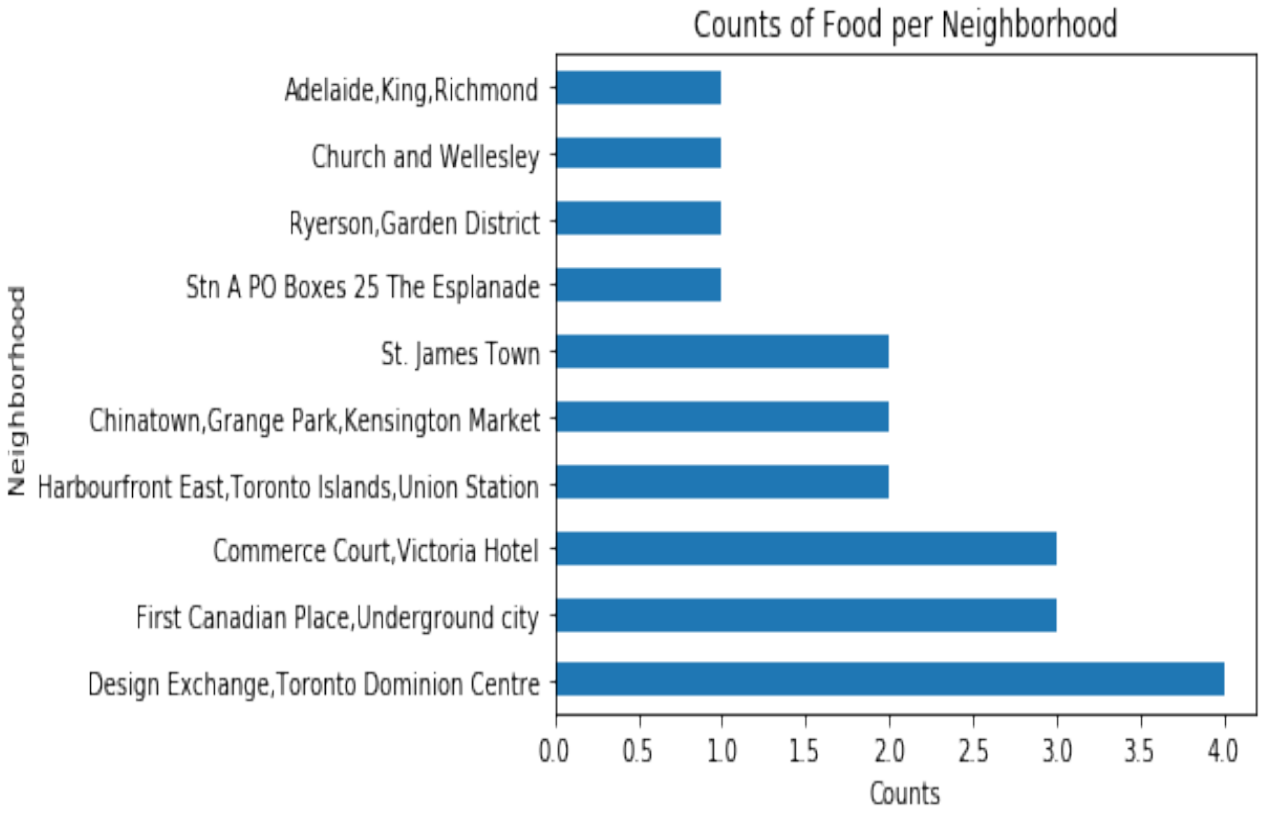


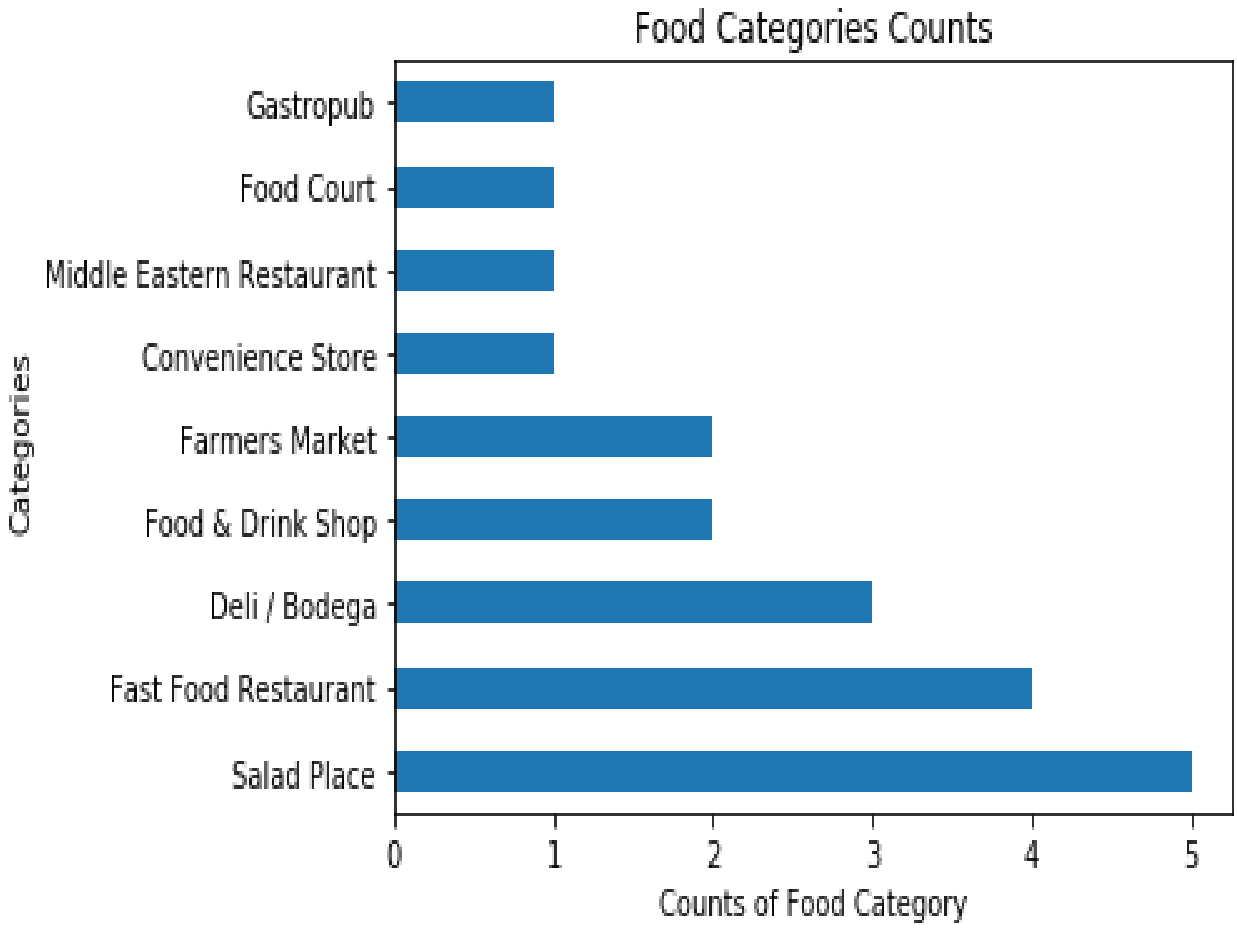


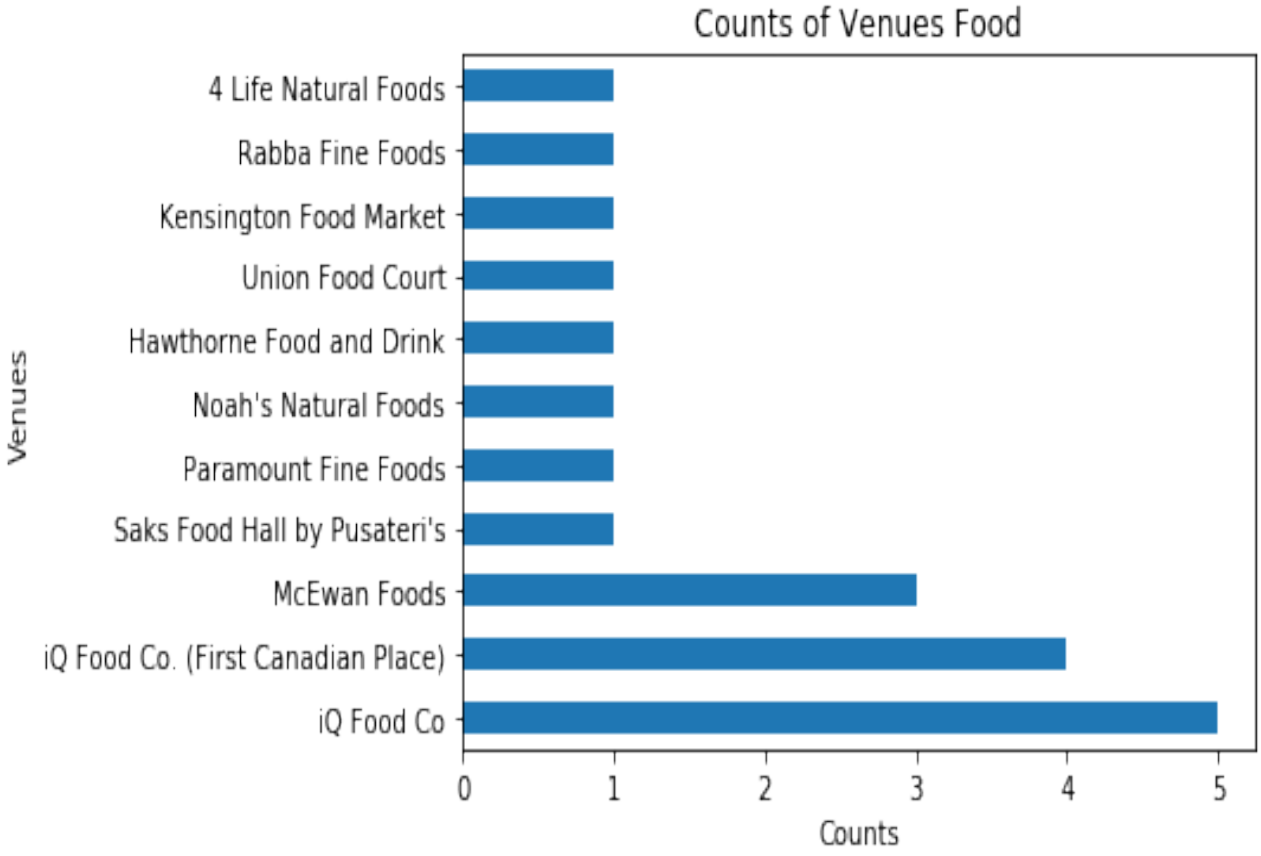

t Disco, ['Nightclub']
Old Toronto
Prince Edward Viaduct
Don Valley Parkway
Danforth Avenue
Pape Avenue
Avenue Road
Bayview Avenue; Bloor Street
Richmond Street; Eastern Avenue
Lake Shore Boulevard
Lake Shore Boulevard; Don Valley Parkway
157
154
Spadina Avenue
153
Exhibition Place
Billy Bishop
Toronto - Hanlan's Point

Popular Food venues : Indirect Concurencies

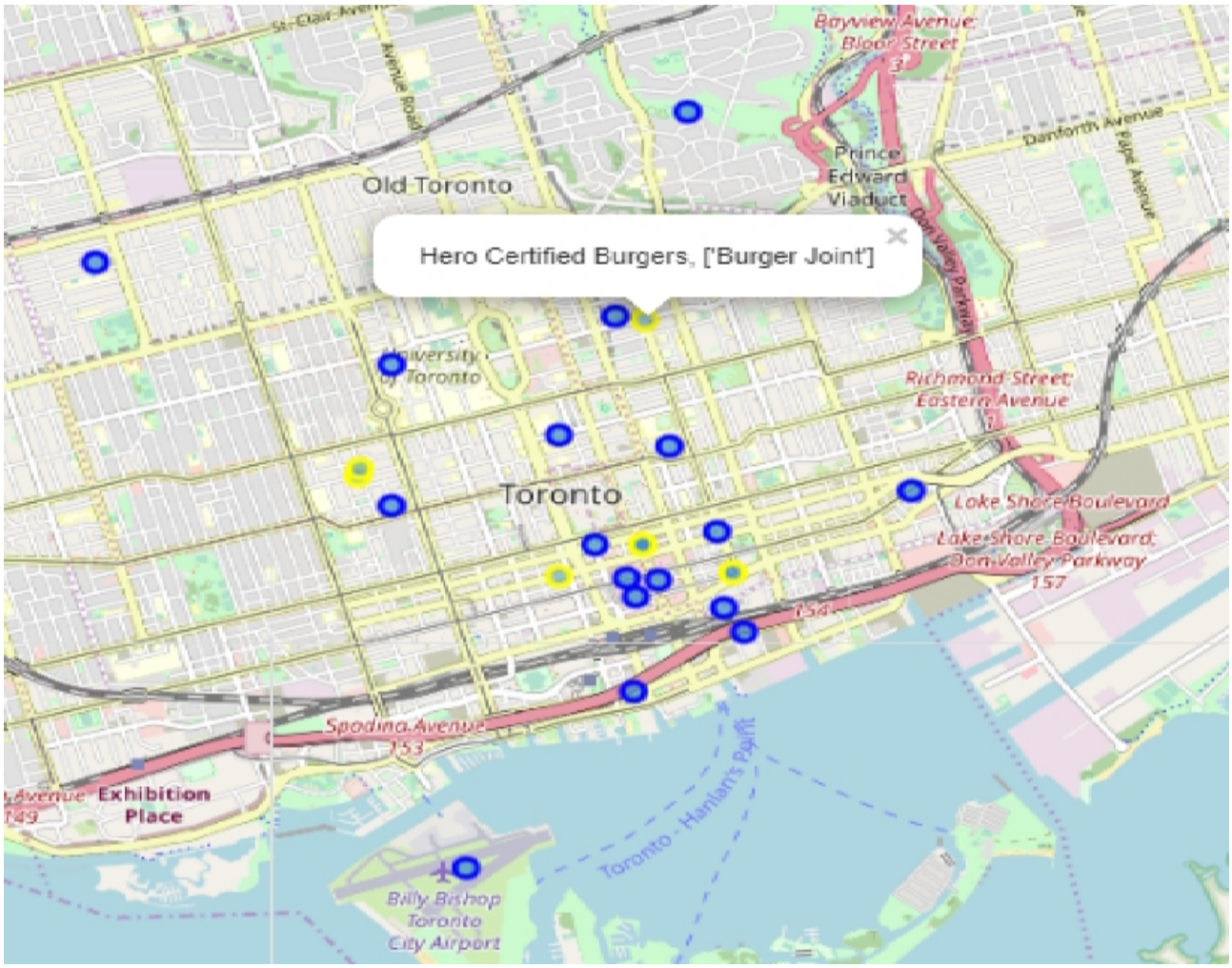


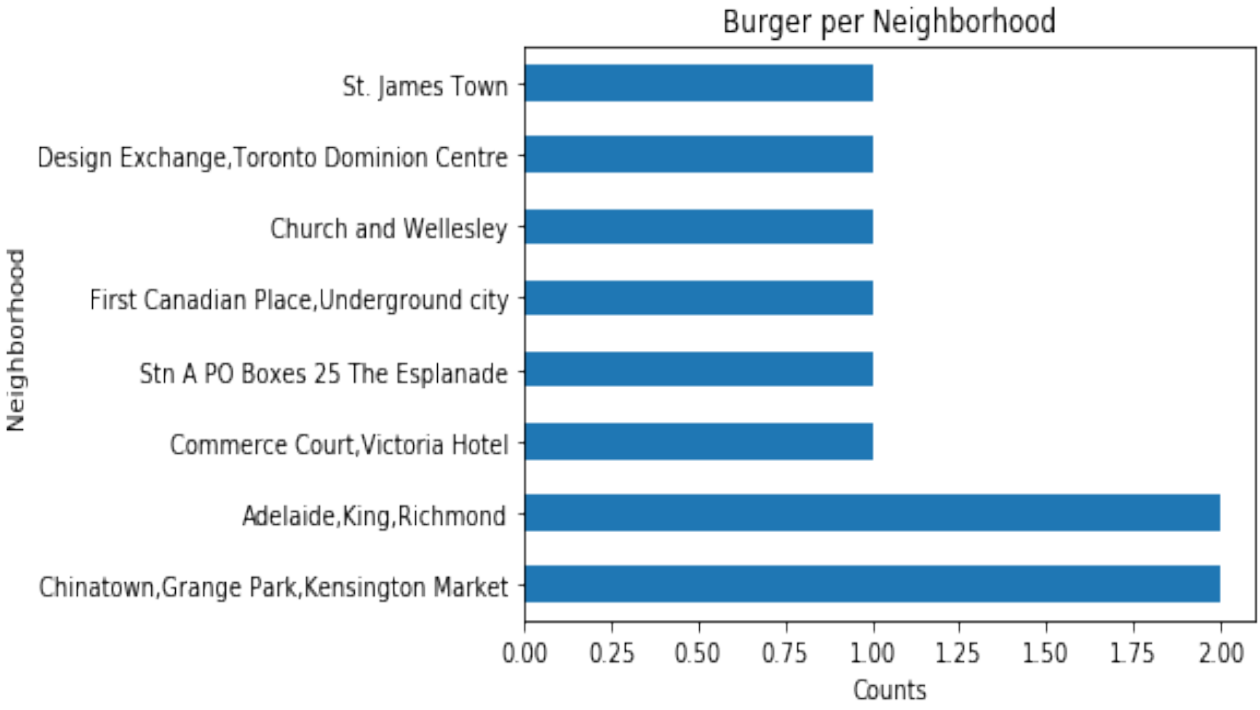


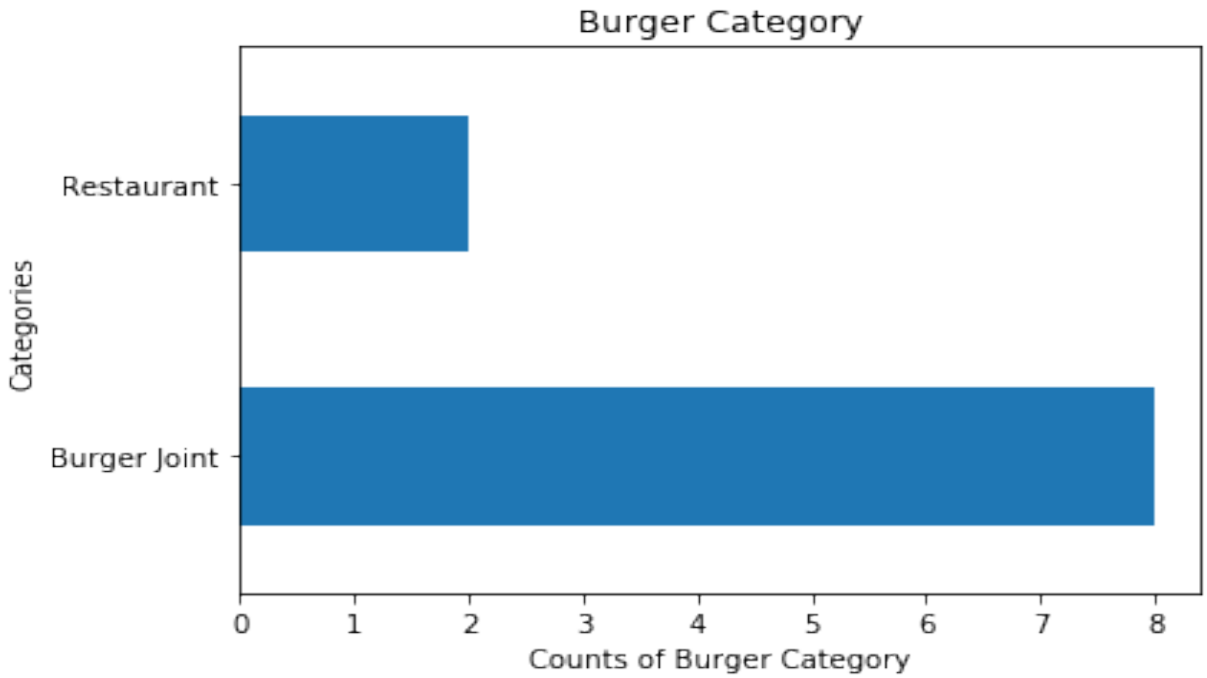

Food Categories Counts
Gastropub
Food Court
Middle Eastern Restaurant
Convenience Store
Farmers Market
Food & Drink Shop
Deli / Bodega
Fast Food Restaurant
Salad Place
Categories
0
1
2
3
4
5
Counts of Food Category

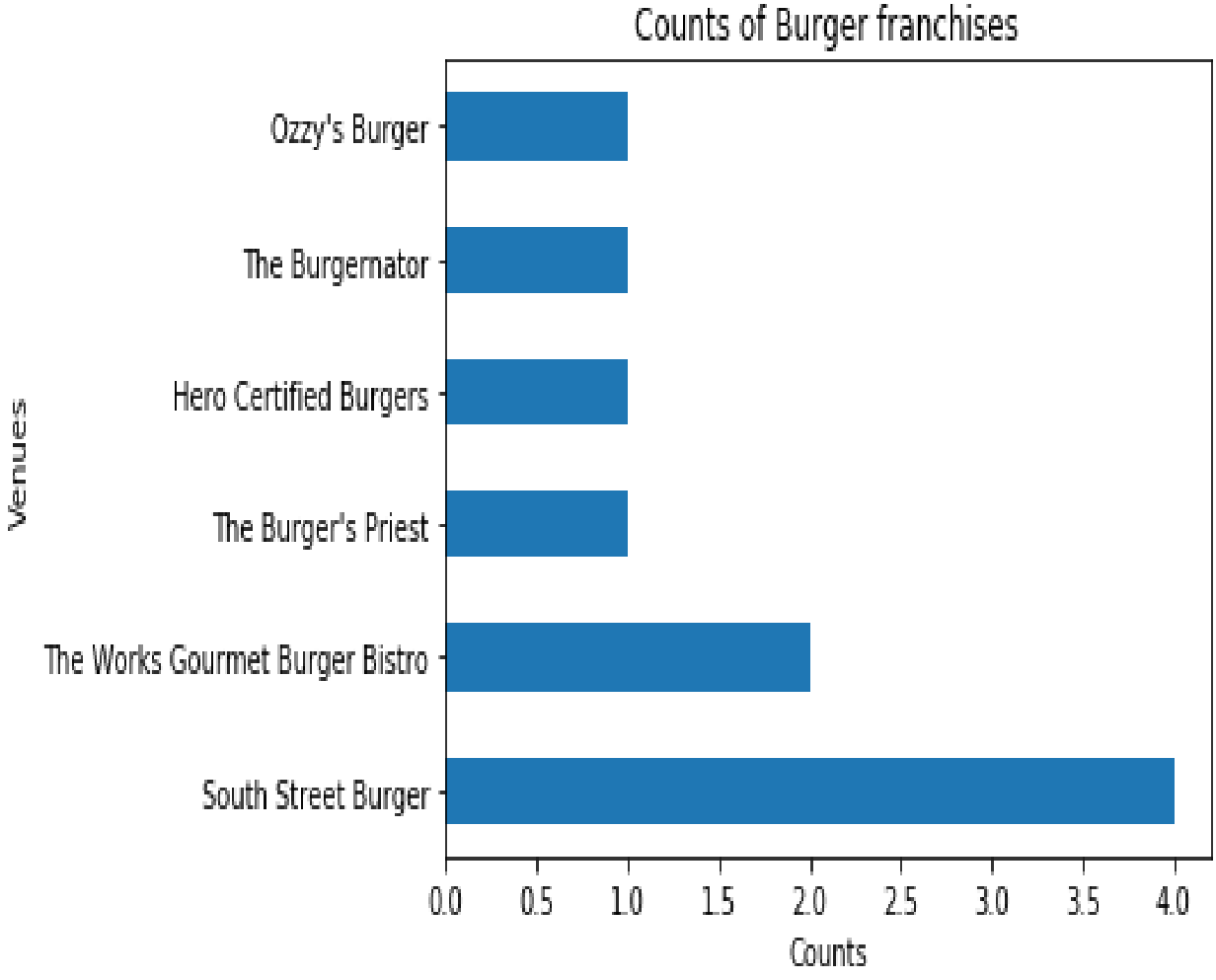

Counts of Venues Food
4 Life Natural Foods
Rabba Fine Foods
Kensington Food Market
Union Food Court
Hawthorne Food and Drink
Noah's Natural Foods
Paramount Fine Foods
Saks Food Hall by Pusateri's
McEwan Foods
iQ Food Co. (First Canadian Place)
iQ Food Co
Venues
0
1
2
3
4
5
Counts

POPULAR Burger Venues

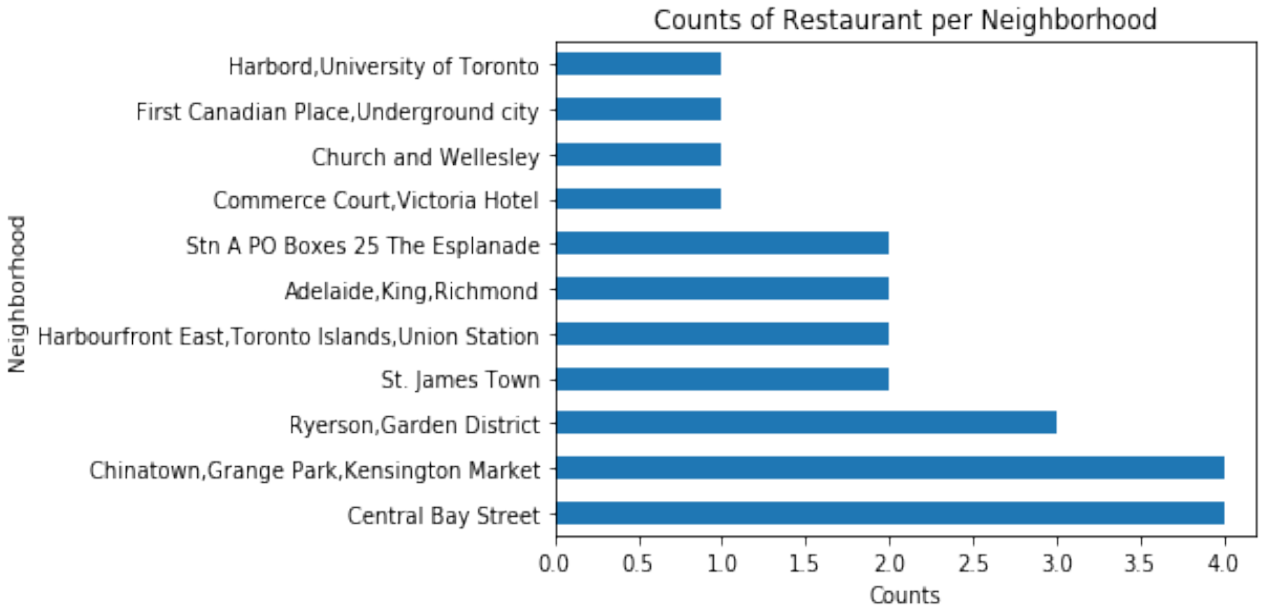


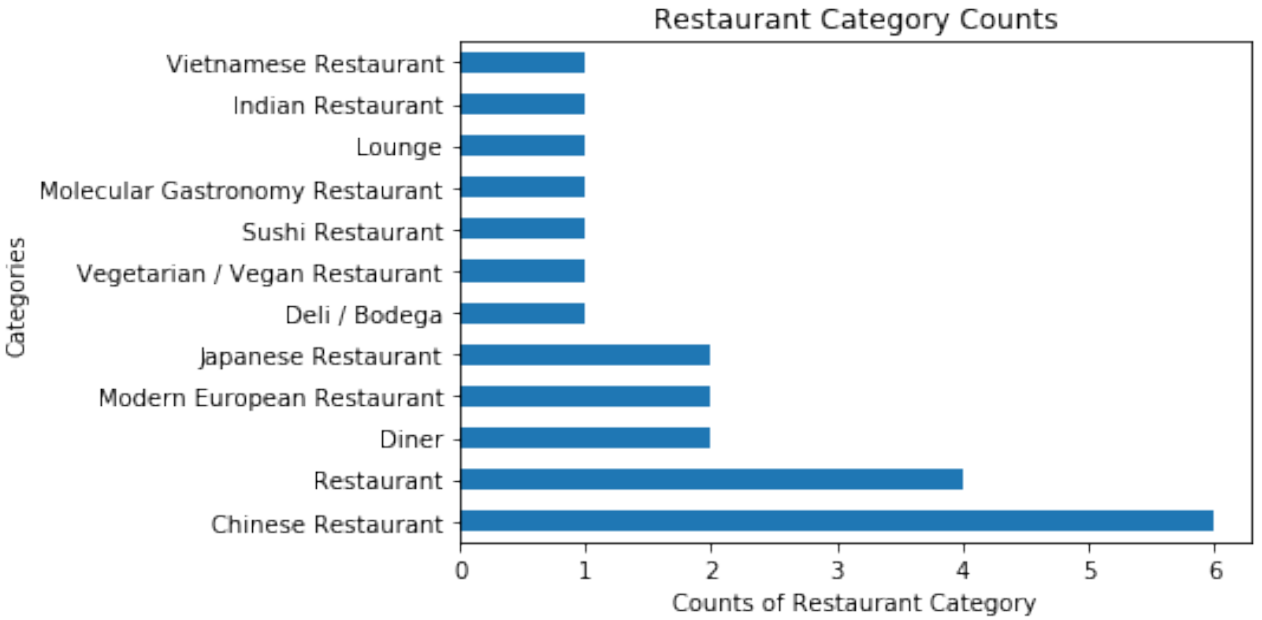



Burger Category
Restaurant
Burger Joint
Categories
0
1
2
3
4
5
6
7
8
Counts of Burger Category

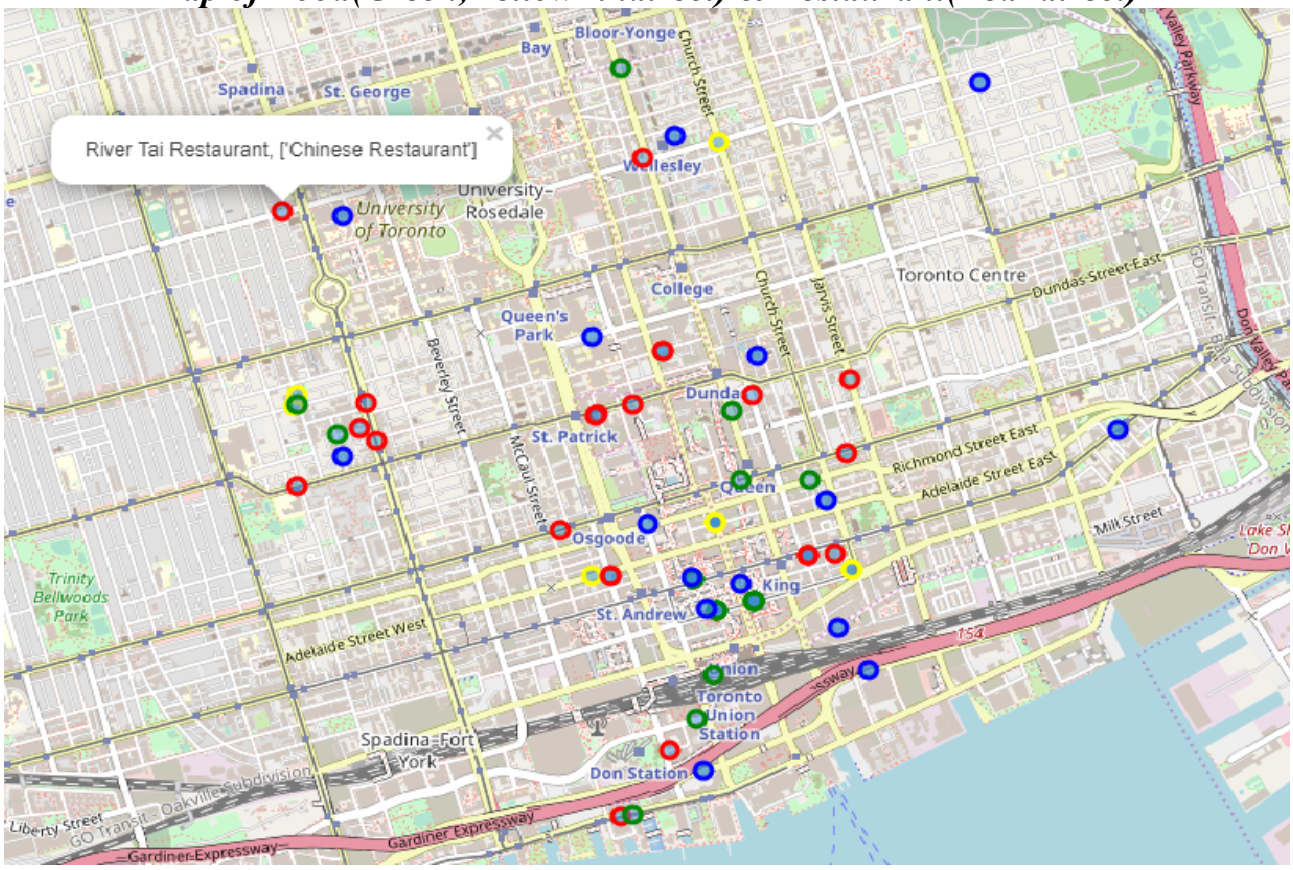

Counts of Burger franchises
Ozzy's Burger
The Burgernator
Hero Certified Burgers
The Burger's Priest
The Works Gourmet Burger Bistro
South Street Burger
Venues
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Counts

Restaurant Populars VENUES : direct rivalry

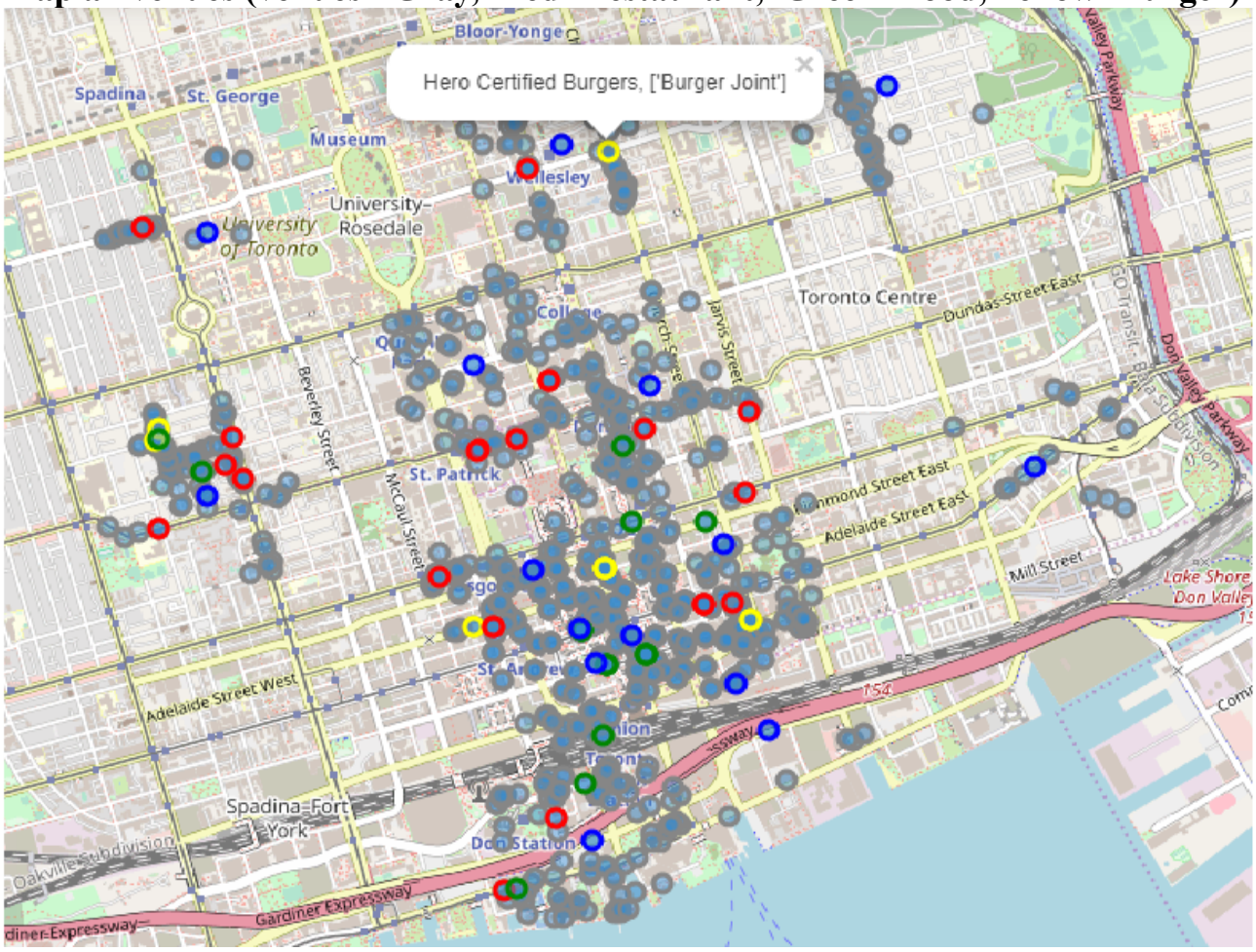




Map of Food(Green, Yellow=indirect) & Restaurant(Red=direct)



Map all venues (venues= Gray, Red=Restaurant, Green=Food, Yellow=Burger)



Machine Learning : Clustering K-Means

I want to Cluster my Borough.
For that, I one hot encoding the table, and group the neighborhood for obtain the mean of frequency for each Categories.

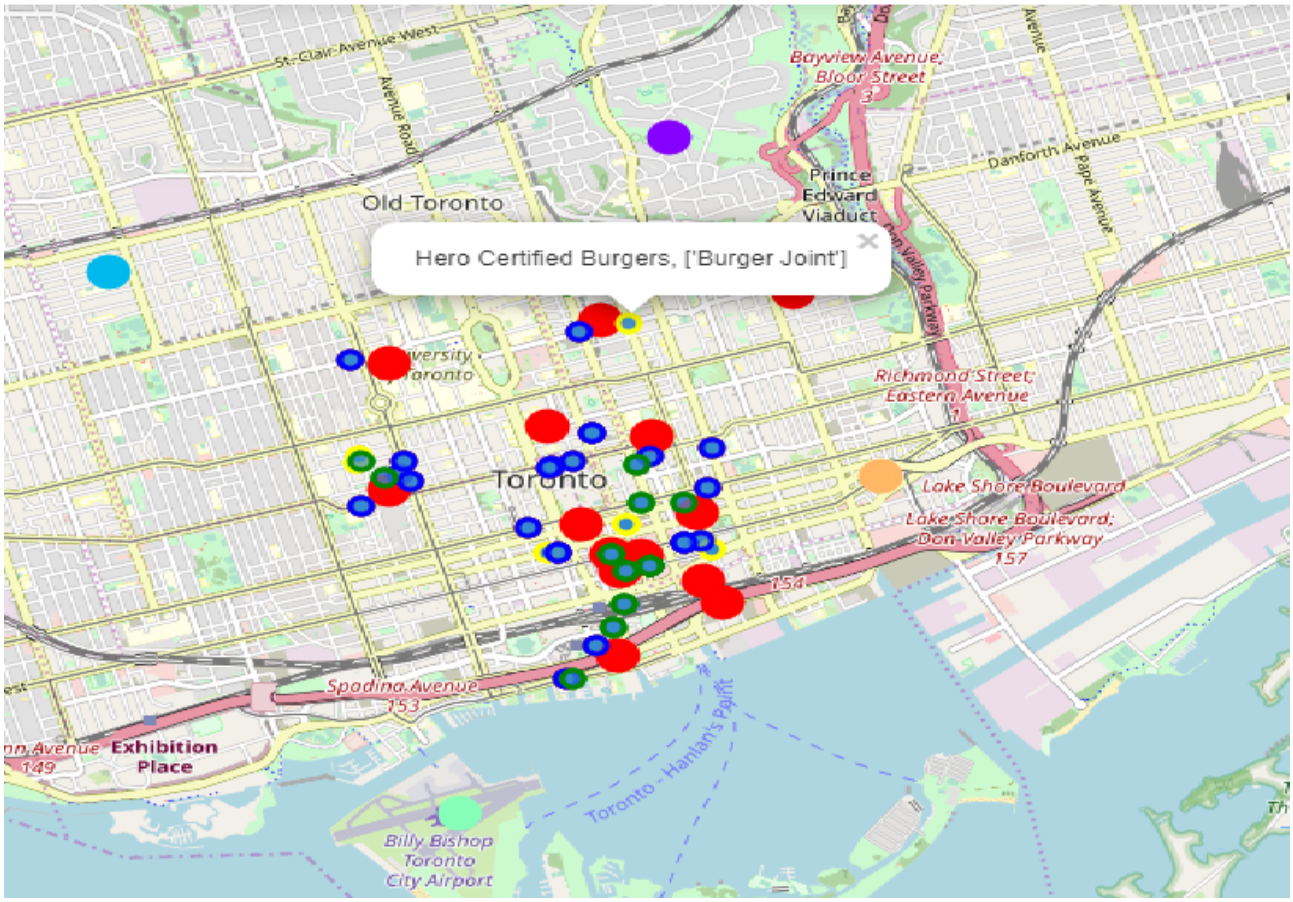
	Neighborhood	Wings Joint	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Art Gallery
0	Adelaide,King,Richmond	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.038462	0.000000
1	Berczy Park	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
2	CN Tower,Bathurst 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
4	Central Bay Street	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000
5	Chinatown,Grange Park,Kensington Market	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000

After that, I can create with a function, this table with order of the most common venues for each Neighborhood.

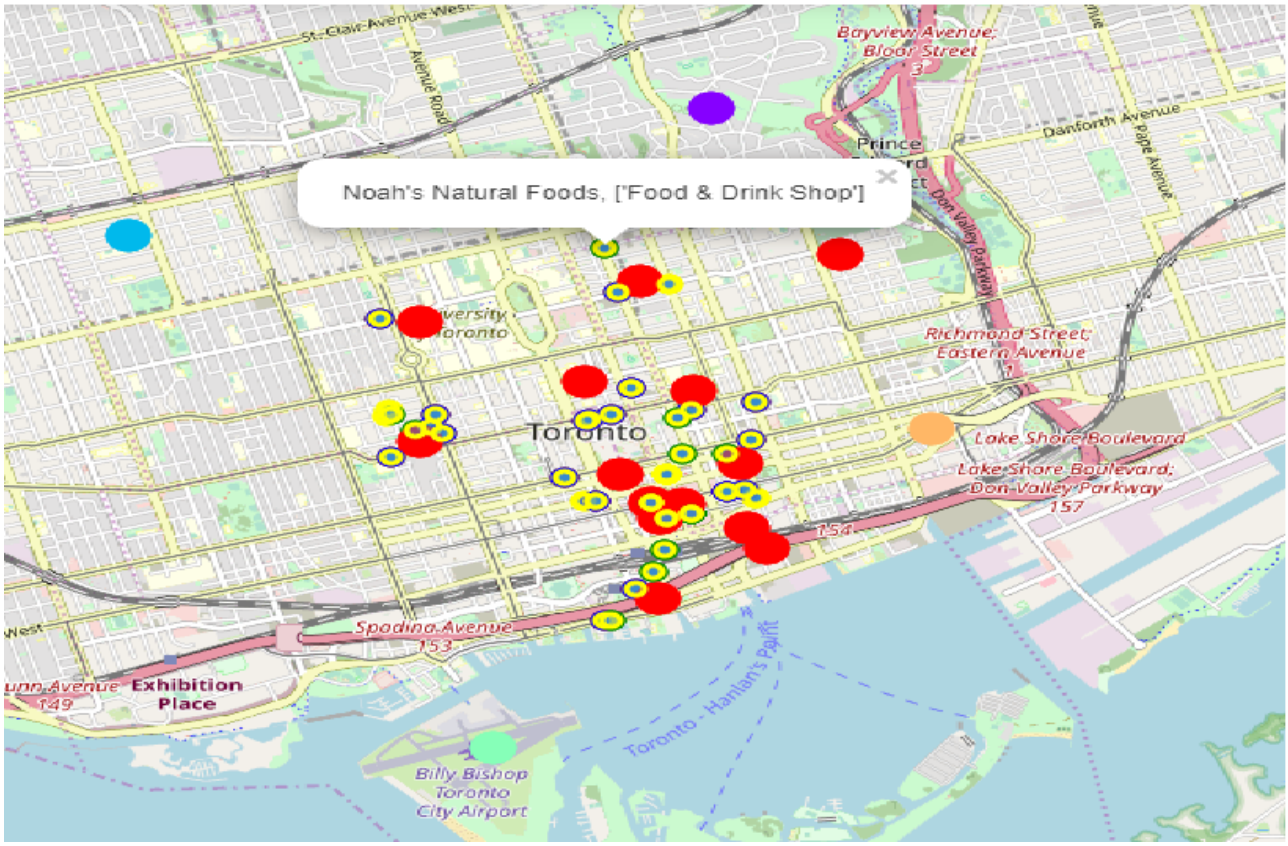
	Neigh	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
0	Adelaide,King,Richmond	Coffee Shop	Japanese Restaurant	Steakhouse	
1	Berczy Park	Cocktail Bar	Liquor Store	Fountain	
2	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Lounge	Airport Terminal	Boutique	Co
3	Cabbagetown,St.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

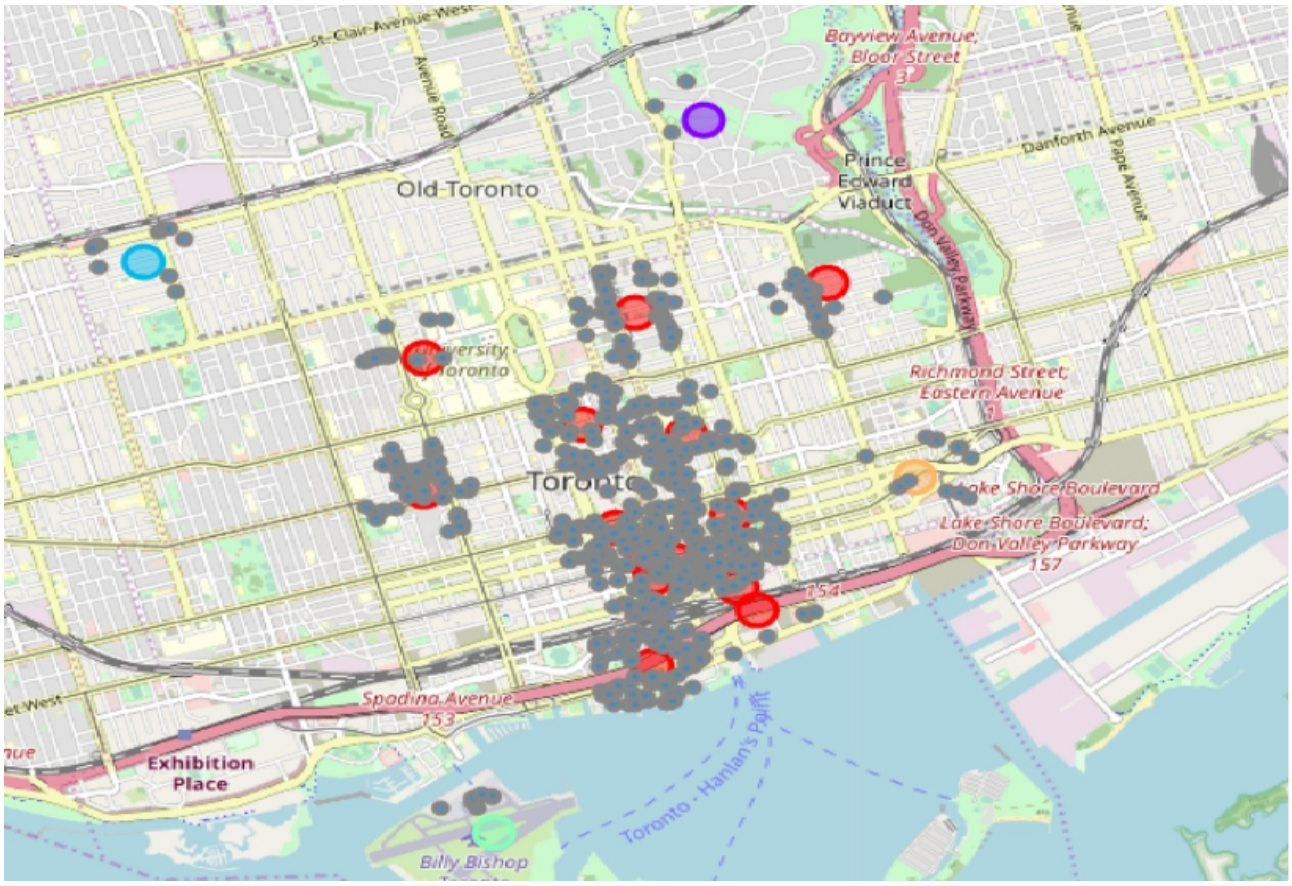
Go Clustering

I used the Kmeans algorithms, for segment each Borough with the Most Common venues in both the same Cluster. I obtain with the Elbow Point k4 is optimal clustering !
But I choose the K5 because the silhouette score are not more decreased with clusters.

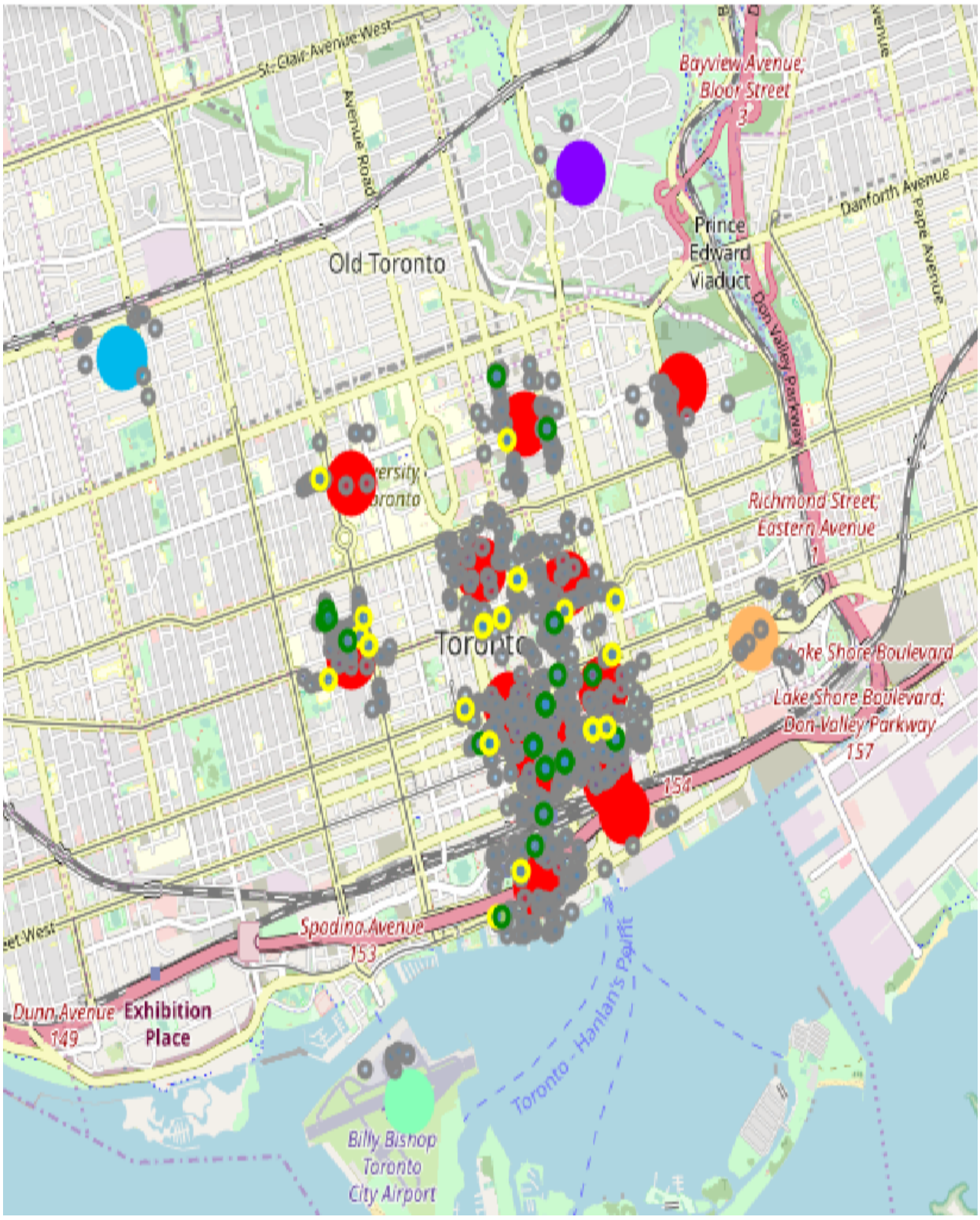


Map Cluster + Eat venues (yellow)





Map : Cluster + Indirect & Direct rivalry + all venues



Comments :

Red cluster are satured with all Eat venues, and all most common venues in gray.

4. Results

If you want to minimize the concurencies :

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

Arguements : you don't have popular restaurant venues and indirect concurencies surroundings area. And all venues are next to cluster 1.

Also, the environnement of this Cluster are more quiet, surroundings Lake, Park, Nature.. And you can touch different customers eventually.

5. 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 of this Borough (demographics, education, cultural...)