Analyzing Public Catering Establishments in Toronto

IBM Capstone Project

1 Introduction

The outbreak of covid-19, occurred the previous year, has radically changed people's usual life style. Since that time people have not mostly been able to go out and visit public places. Plenty of cafes and restaurants have been closed down due to imposed restrictions. Taking into account these facts, on the one hand running the restaurant business is a highly risked decision. On the other hand the pandemic is not a persistent global phenomenon, and in some countries including Canada it can be said that it has been brought under control. So that opportunities for recovering business or running the new one are going to appear. Then there may be other possible variants to open a restaurant or cafe such as providing delivery or takeaway options.

A client wants to open a public catering establishment in Toronto and needs to assess what a format of such a place should be chosen. Is it better to open a small pizzeria or an expensive restaurant with a big number of various dishes? The other question is concerned with a pandemic situation. The client wants to estimate risks for both clients and employees. Moreover, in dependence on the number of COVID-19 cases in neighbourhoods restrictions for public places might be different. Then the client prefers opening their place in a safe neighbourhood in terms of crime rate.

2 Data

2.1 Data sources

According to the objective, the following factors should be investigated:

- Number of existing public catering establishments in the neighbourhoods
- Epidemiological situation in the neighbourhoods
- Crime rates in the neighbourhoods

To resolve all these questions, the following data will be used:

- The dataset with neighbourhood crime rates from the City of Toronto's Open Data Portal is to search for the neighbourhoods with the low crime rate. Types of crimes should be investigated as well.
- The dataset with covid-19 cases in Toronto from the City of Toronto's Open Data Portal is to exclude the neighbourhoods with a high number of new cases.
- Foursquare location data is used to obtain the most common food venues in the neighbourhoods.

- The dataset with the coordinates of Toronto neighbourhoods is for their interactive geospatial visualization.
- The geo data in the json format is to obtain the coordinates of the neighbourhoods for their interactive geospatial visualization on choropleth maps.

2.2 Data preprocessing

Multiple resources with various data are used. It is necessary to investigate these resources to select the appropriate data, remove unnecessary one and resolve issues with missing values if they exist.

2.2.1 Getting data with crime rates

The open source which is the City of Toronto's Open Data Portal is helpful for getting free various datasets. The data in this portal is regularly updated. To compose a data frame with crime rates in Toronto, the dataset 'Neighbourhood Crime Rates' is used.

The dataset contains the data by neighbourhood for various crime types. This data includes the count of crimes and the crime rate per 100,000 population for each of the past 7 years. To estimate risks concerned with crime rates for the new public catering establishment, the more appropriate ones which are robberies and burglaries will be considered. Moreover, to make the data more applicable, the total number of crimes of each type per neighbourhood and the corresponding mean number of the crime rate for all these years will be calculated. The composed data frame with crime rates in Toronto is shown in Figure 1.

	Neighbourhood	Assault_Rate_mean	Burglaries_Rate_mean	Robbery_Rate_mean	Crimes_Total
0	Agincourt North	244.798486	158.419463	94.582293	1067
1	Agincourt South-Malvern West	462.185571	314.380986	102.594551	1565
2	Alderwood	287.495757	209.450686	61.767394	497
3	Annex	833.750386	475.984500	152.495671	3346
4	Banbury-Don Mills	276.324557	234.616229	46.386083	1141

Figure 1. Data frame with crime rates in Toronto

2.2.2 Getting data with COVID-19 cases

To compose a data frame with COVID-19 cases in Toronto, the dataset 'COVID-19 Cases in Toronto' from the City of Toronto's Open Data Portal is used as well.

The dataset contains the data by neighbourhood for all confirmed and probable cases since the first one was reported in January 2020. To estimate risks concerned with rates of COVID-19 cases for opening the new public catering establishment, the ones occurred since March 2021 will be considered. The composed data frame with COVID-19 cases in Toronto is shown in Figure 2.

	Neighbourhood	Covid_Cases_Total
0	Agincourt North	527
1	Agincourt South-Malvern West	560
2	Alderwood	222
3	Annex	310
4	Banbury-Don Mills	361

Figure 2. Data frame with COVID-19 cases in Toronto

2.2.3 Providing the coordinates of the neighbourhoods

Both data frames with COVID-19 cases and crime rates in Toronto are combined to get a single data frame with all the necessary data. Several neighbourhood names were fixed since their names had been slightly different.

Nominatim geocoder from the geopy library is used to get the coordinates for the neighbourhoods. The coordinates of plenty of neighbourhoods presented in the data frame are not defined. So that such neighbourhoods can't be visualized, and the corresponding rows should be dropped. The data frame with the coordinates of Toronto neighbourhoods obtained with Nominatim geocoder is shown in Figure 3.

	Neighbourhood	Covid_Cases_Total	Assault_Rate_mean	Burglaries_Rate_mean	Robbery_Rate_mean	Crimes_Total	Latitude	Longitude
0	Agincourt North	527	244.798486	158.419463	94.582293	1067	43.808038	-79.266439
1	Agincourt South-Malvern West	560	462.185571	314.380986	102.594551	1565	NaN	NaN
2	Alderwood	222	287.495757	209.450686	61.767394	497	43.601717	-79.545232
3	Annex	310	833.750386	475.984500	152.495671	3346	43.670338	-79.407117
4	Banbury-Don Mills	361	276.324557	234.616229	46.386083	1141	NaN	NaN

Figure 3. Data frame with coordinates of Toronto neighbourhoods obtained with Nominatim geocoder

Moreover, these results are changed from time to time which is not convenient for the current work. So that the data obtained with the use of Nominatim geocoder are not reliable. Therefore another source which is [the dataset with the coordinates of Toronto neighbourhoods]() will be used to get the coordinates for more neighbourhoods. The new data frame with the coordinates of Toronto neighbourhoods is shown in Figure 4.

	Neighbourhood	Covid_Cases_Total	Assault_Rate_mean	Burglaries_Rate_mean	Robbery_Rate_mean	Crimes_Total	Longitude	Latitude
0	Agincourt North	527	244.798486	158.419463	94.582293	1067	-79.281616	43.797406
1	Agincourt South-Malvern West	560	462.185571	314.380986	102.594551	1565	-79.289169	43.785187
2	Alderwood	222	287.495757	209.450686	61.767394	497	-79.553204	43.595500
3	Annex	310	833.750386	475.984500	152.495671	3346	-79.412147	43.674431
4	Banbury-Don Mills	361	276.324557	234.616229	46.386083	1141	-79.326505	43.732570

Figure 4. New data frame with coordinates of Toronto neighbourhoods

The coordinates of all neighbourhoods presented in the initial data frame with crime rates and COVID-19 cases are defined. Several neighbourhood names were fixed again since their names had been slightly different. The total number of neighbourhoods in the cleaned data frame is 140.

3 Methodology

3.1 Data visualization

Firstly, the Nominatim geocoder from the geopy library is used to get the latitude and longitude of Toronto. Then The Folium library is used to visualize Toronto neighbourhoods on the interactive map using their earlier obtained coordinates. The map of Toronto with the neighbourhoods designated on it is presented in Figure 5.

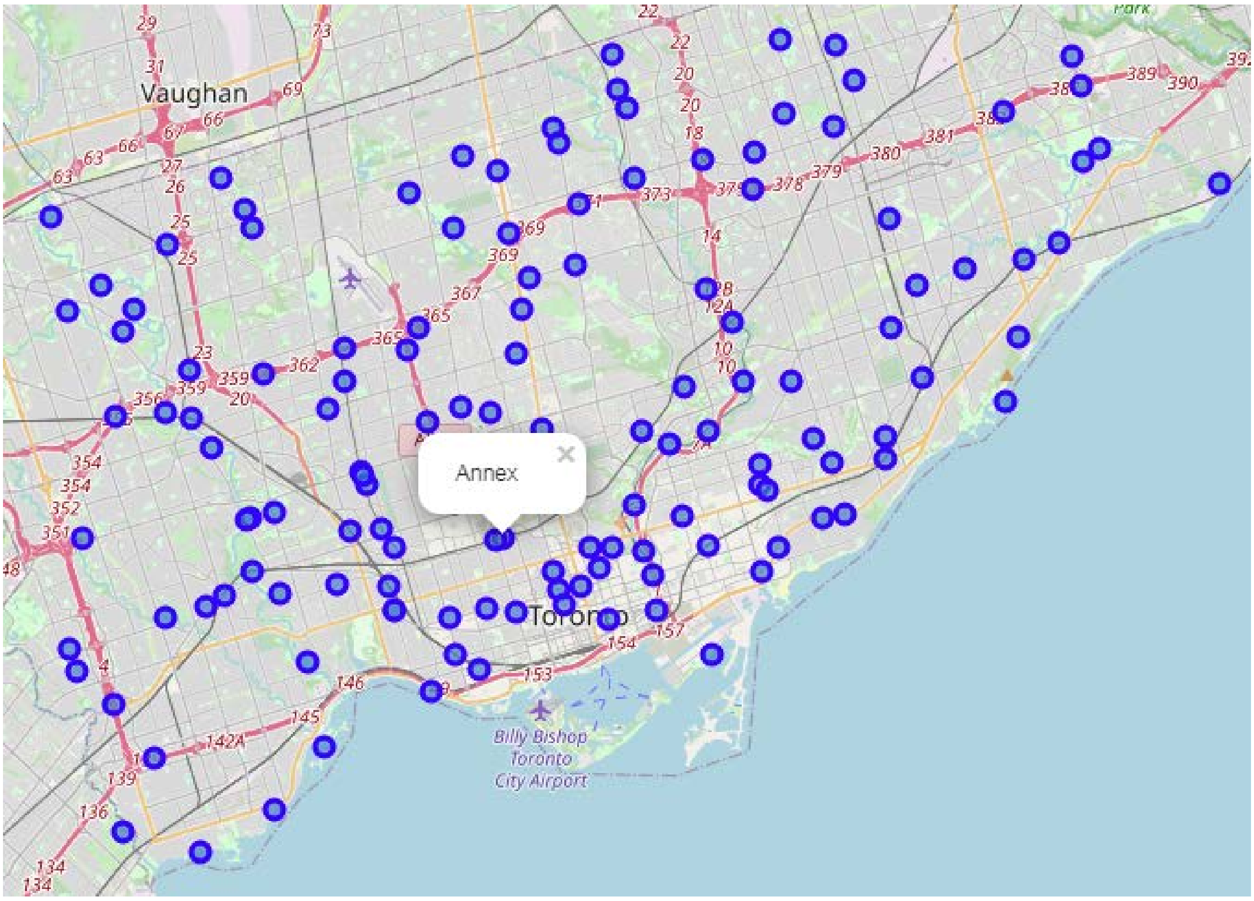


Figure 5. Map of Toronto with designated neighbourhoods

3.2 Exploring Neighbourhoods of Toronto

3.2.1 Getting categories of food venues

First of all, it should be assessed what food venue categories are available in Foursquare. The composed data frame with available food venues is shown in Figure 6.

	level0.name	level1.name	level2.name	level3.name
321	Latin American Restaurant	South American Restaurant	Brazilian Restaurant	Pastelaria
322	Latin American Restaurant	South American Restaurant	Brazilian Restaurant	Southeastern Brazilian Restaurant
323	Latin American Restaurant	South American Restaurant	Brazilian Restaurant	Southern Brazilian Restaurant
324	Latin American Restaurant	South American Restaurant	Brazilian Restaurant	Tapiocaria
325	Latin American Restaurant	South American Restaurant	Peruvian Restaurant	Peruvian Roast Chicken Joint

Figure 6. The data frame with available food venues

According to the results there are 326 unique categories of food venues consisted of no more than 4 levels.

3.2.2 Getting food venues in Toronto

Food venues in Toronto neighbourhoods are obtained with the use of the Foursquare API and within the area restricted by the radius of 600 meters relative to the neighbourhood coordinates. The food venues data in conjunction with the one on the corresponding Toronto neighbourhoods are merged into a common data frame, which is shown in Figure 7.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt North	43.797406	-79.281616	Subway	43.797503	-79.282181	Sandwich Place
1	Agincourt North	43.797406	-79.281616	凱璽	43.801003	-79.283363	Asian Restaurant
2	Agincourt South-Malvern West	43.785187	-79.289169	Burger King	43.784205	-79.292606	Fast Food Restaurant
3	Agincourt South-Malvern West	43.785187	-79.289169	Subway	43.783728	-79.292873	Sandwich Place
4	Agincourt South-Malvern West	43.785187	-79.289169	Congee Queen 皇后名粥	43.783910	-79.292319	Cantonese Restaurant

Figure 7. Data frame with food venues in Toronto

There are 93 unique categories and 1451 food venues among them in 126 Toronto neighbourhoods. There are 14 neighbourhoods without any food venues. They can be considered as a possible place to open a new one. The data frame with such neighbourhoods is presented in the Figure 8.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1451	Alderwood	43.595500	-79.553204	NaN	NaN	NaN	NaN
1452	Centennial Scarborough	43.769728	-79.145626	NaN	NaN	NaN	NaN
1453	Cliffcrest	43.711273	-79.225220	NaN	NaN	NaN	NaN
1454	Humbermede	43.735709	-79.549080	NaN	NaN	NaN	NaN
1455	Long Branch	43.589748	-79.524738	NaN	NaN	NaN	NaN

Figure 8. Data frame with neighbourhoods without food venues in Toronto

A problem with names of food venue categories in Toronto is that they can correspond to any level of the ones available in Foursquare like it is shown in the example in

Figure 9. One venue category is ‘Restaurant’ which corresponds to the one of level 0 in Foursquare. The other category is “Ramen Restaurant” which corresponds to the one of level 2 in Foursquare since it is a subcategory of “Japanese Restaurant” which, in turn, is a subcategory of “Asian Restaurant”.

1050	Willowdale West	43.772978	-79.413791	Sansotei Ramen 三草亭	43.776709	-79.413927	Ramen Restaurant
1051	Rosedale-Moore Park	43.671837	-79.379375	O. Noir	43.669145	-79.382505	Restaurant

Figure 9. Food venue categories of different level

To compare food venue categories existing in Toronto and presented in Foursquare, it is necessary to combine data with all possible food venue categories to make them of the same level.

According to the obtained results roughly 265 food venue categories don't exist in Toronto. However, this number is not precise since names of some food venue categories exist for each level in the new composed data frame, and some food venue categories available in Toronto have a common category such as ‘Restaurant’ which might have been specified more accurately.

Then one food venue category, shown in Figure 10, is presented only in Toronto and should be just dropped as it doesn't give us useful information.

	Venue Category	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Food	Niagara	43.638903	-79.420759	Liberty Village Rotisserie & Grill	43.639941	-79.422095
1	Food	Dufferin Grove	43.652743	-79.431641	The Depanneur	43.652904	-79.429575
2	Food	Woodbine Corridor	43.672073	-79.309942	Beaches Natural Foods	43.669265	-79.303633
3	Food	Oakridge	43.716585	-79.304827	Sunrise Bar & Grill	43.719829	-79.301050
4	Food	Malvern	43.788985	-79.225960	Catering Club	43.792914	-79.222192
5	Food	South Parkdale	43.632999	-79.438836	Urban Herbivore	43.636648	-79.435984
6	Food	Humber Heights-Westmount	43.680013	-79.506078	Pringles	43.675705	-79.507663

Figure 10. Data frame with food venue category existing only in Toronto

3.2.3 Assessing the total number of food venues in each neighbourhood

The more appropriate data to be assessed is the total number of food venues in each Toronto neighbourhood. In Figure 11 the corresponding histogram is presented.

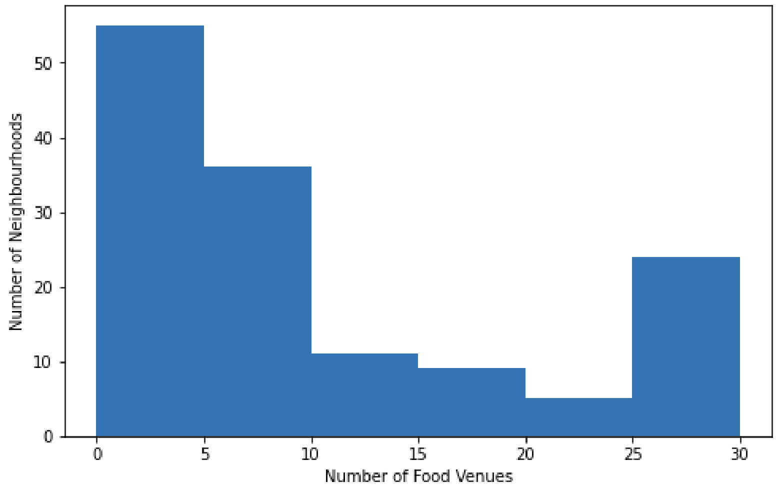


Figure 11. Histogram of number of food venues in Toronto neighbourhoods

The histogram shows that there are 55 neighbourhoods with less than 5 food venues. At the same time 24 neighbourhoods have more than 25 food venues.

It can be observed from the results that there are plenty of neighbourhoods where only a few food venues exist. On the other hand in some neighbourhoods many such places have already opened. The results don't imply that a new food venue should be opened in a neighbourhood with less number of places. There may be many reasons why some neighbourhoods attract owners to open their restaurants and the others don't.

3.2.4 Assessing the total number of venues in each food venue category

Another data to be assessed is the total number of venues in each food venue category existing in Toronto. The corresponding bar chart is shown in Figure 12. Since there are a lot of food venue categories, they were confined to the ones whose count is more than 30.

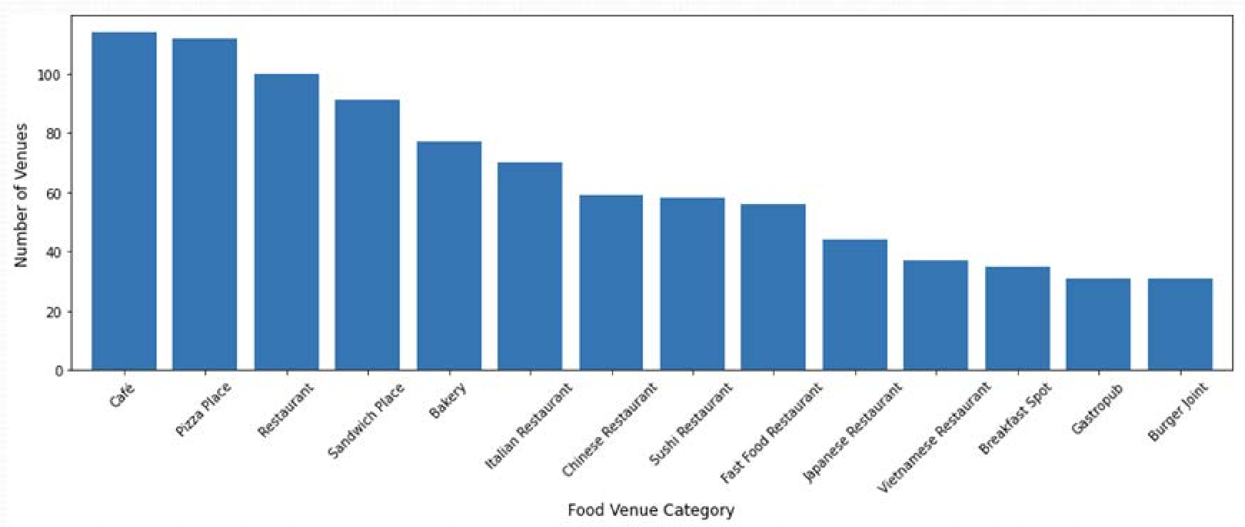


Figure 12. The total number of venues in each food venue category in Toronto

It's clear from the bar chart that most of food venues fall into the following categories: cafes, pizza and sandwich places, bakeries, various restaurants, especially the Italian, Asian and Fast Food ones.

A histogram of the venues in the food venue categories in Toronto is shown in Figure 13.

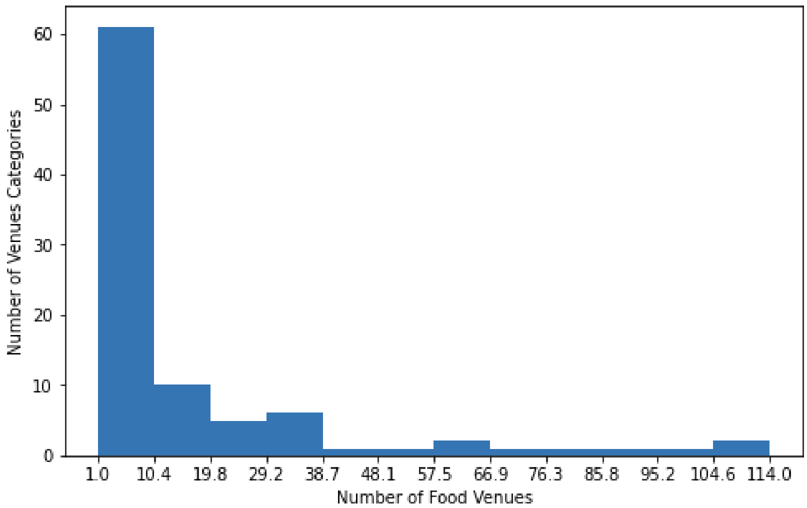


Figure 13. Histogram of venues in food venue categories in Toronto

According to the histogram no more than 10 venues exist for 61 food venue categories. In addition to the previous results it can be observed that among food venue categories, existing in Toronto, only a few venues are presented for most of them.

3.2.5 Getting the top food venues in each neighbourhood

The number of food venues in each Toronto neighbourhood is not so large. Therefore, only 5 top food venues will be considered. Since in some neighbourhoods their number is less than 5, empty items it the data frame will be identified as 'NoVenue'. The data frame with the 5 top food venues in each neighbourhood is shown in Figure 14.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agincourt North	Sandwich Place	Asian Restaurant	NoVenue	NoVenue	NoVenue
1	Agincourt South-Malvern West	Pizza Place	Chinese Restaurant	Restaurant	Vietnamese Restaurant	Deli / Bodega
2	Alderwood	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue
3	Annex	Café	Sandwich Place	Pizza Place	Burger Joint	Diner
4	Banbury-Don Mills	Wings Joint	Restaurant	Food Truck	Deli / Bodega	NoVenue

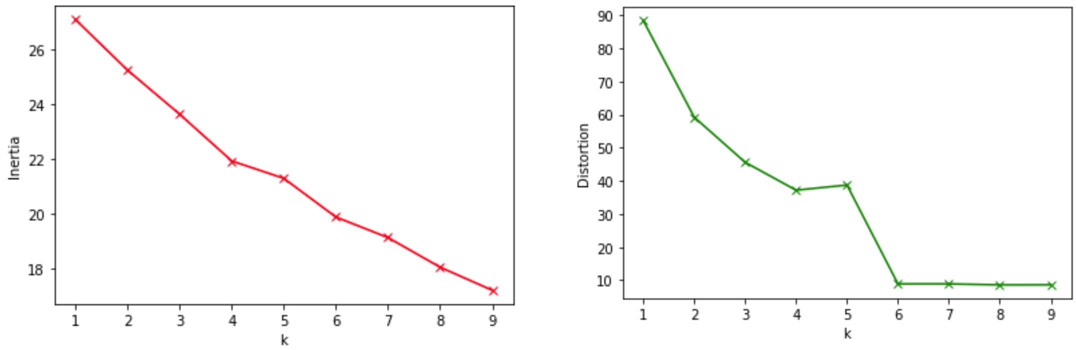
Figure 14. The 5 top food venues in each neighbourhood

3.3 Clustering neighbourhoods

For further analysis of the obtained results with most common venues in Toronto, the neighbourhoods can be divided into non-overlapping subsets of similar objects as in some of them similar top food venues are placed and in the others different food venues are placed. The unsupervised K-Means algorithm for clustering these neighbourhoods is an acceptable and commonly used method.

3.3.1 K Means Clustering

Firstly, neighbourhoods where no food venues exist at all are necessary to be dropped since they don't have a distinguished feature for clustering. After that the appropriate number of clusters should be defined before running K-Means for clustering the neighbourhoods. For this aim the widely used method is the Elbow one which can be utilized with various metrics to compute distances. The most common way shown in Figure 15 (a) doesn't provide an appropriate result but the other one shown in Figure 15 (b) yields a definite k value.



a) b)

Figure 15. Elbow method using Inertia (a) and (b) Canberra Metric (b)

It is seen from the graph in Figure 15 (b) that the optimum k value is 6. Thus K-Means is run to cluster the neighbourhoods into 6 clusters. The map of Toronto with the clustered neighbourhoods designated on it is presented in Figure 16.

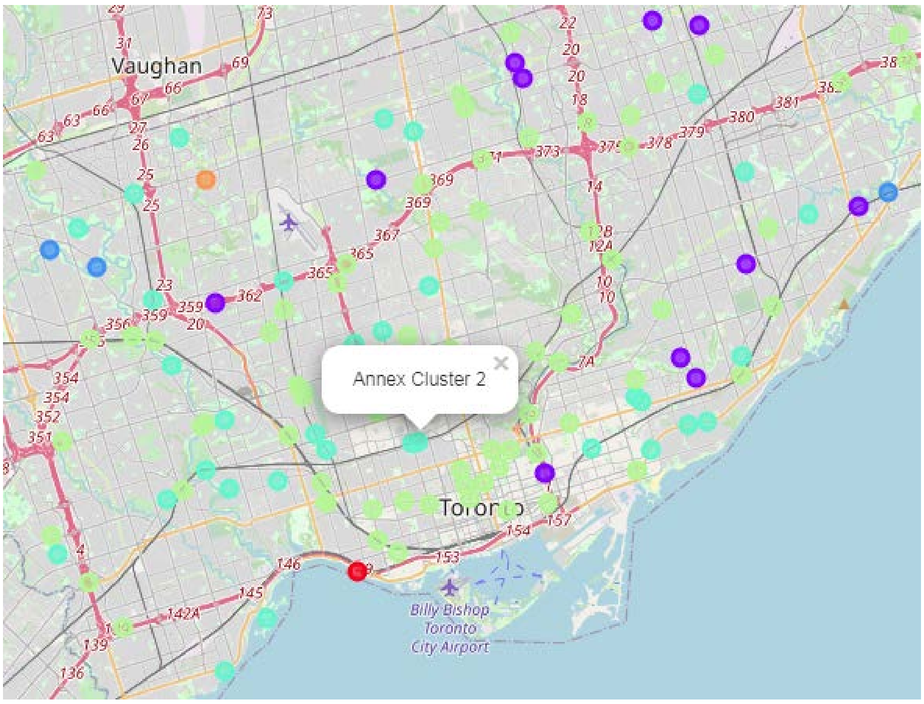


Figure 16. Map of Toronto with designated clustered neighbourhoods

The obtained data frame with cluster numbers specified for neighbourhoods and top food venues in them is presented in Figure 17.

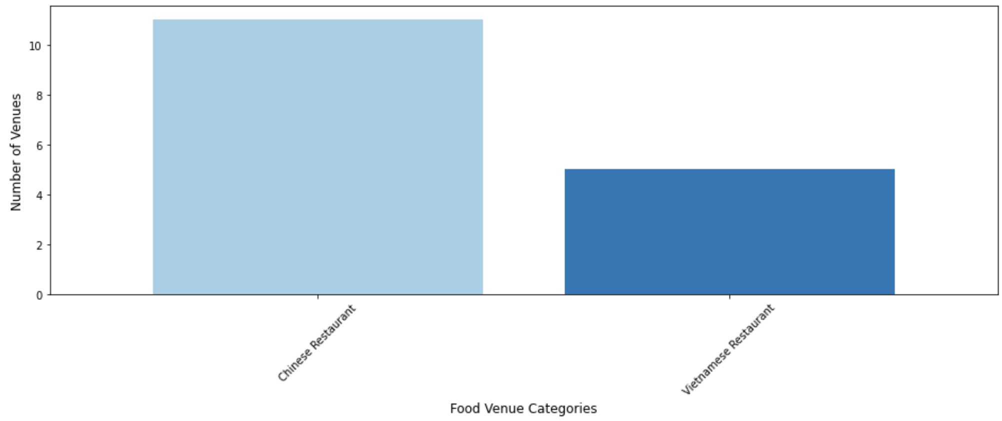
	Cluster_Number	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	3	Agincourt North	Sandwich Place	Asian Restaurant	NoVenue	NoVenue	NoVenue
1	2	Agincourt South-Malvern West	Pizza Place	Chinese Restaurant	Restaurant	Vietnamese Restaurant	Deli / Bodega
2	2	Annex	Café	Sandwich Place	Pizza Place	Burger Joint	Diner
3	3	Banbury-Don Mills	Wings Joint	Restaurant	Food Truck	Deli / Bodega	NoVenue
4	0	Bathurst Manor	Chinese Restaurant	Café	NoVenue	NoVenue	NoVenue

Figure 17. Data frame with clustered neighbourhoods and top food venues

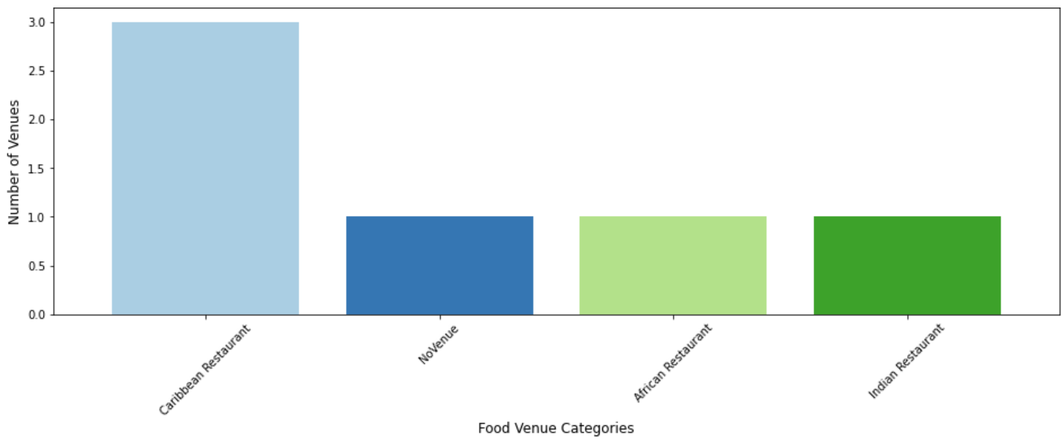
3.3.2 Analyzing clusters

In many neighbourhoods the number of food venues is less than 3. So that the first two most common venues will be assessed to determine an appropriate label for each cluster. Moreover, in dependence on the number of food venue categories in a cluster the minimum number of considered venues can be increased to show the topmost categories.

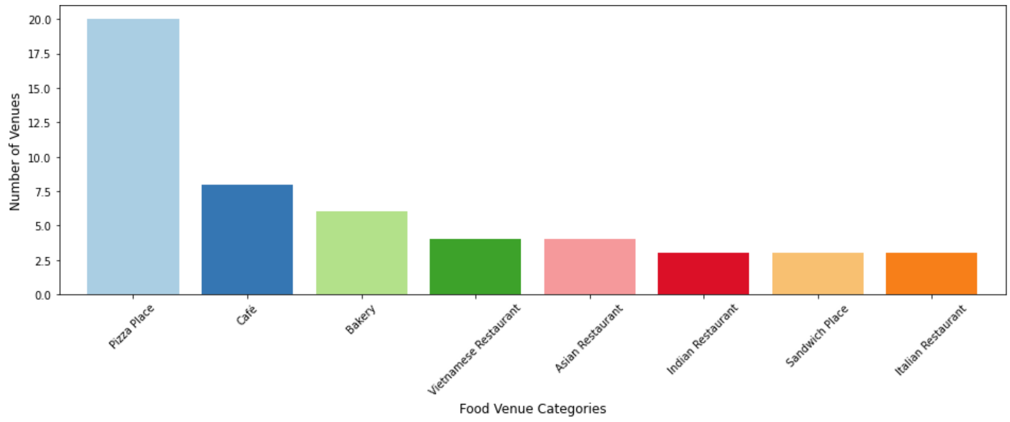
Bar charts with the total number of venues in topmost food venue categories in the clusters are presented in Figure 18.



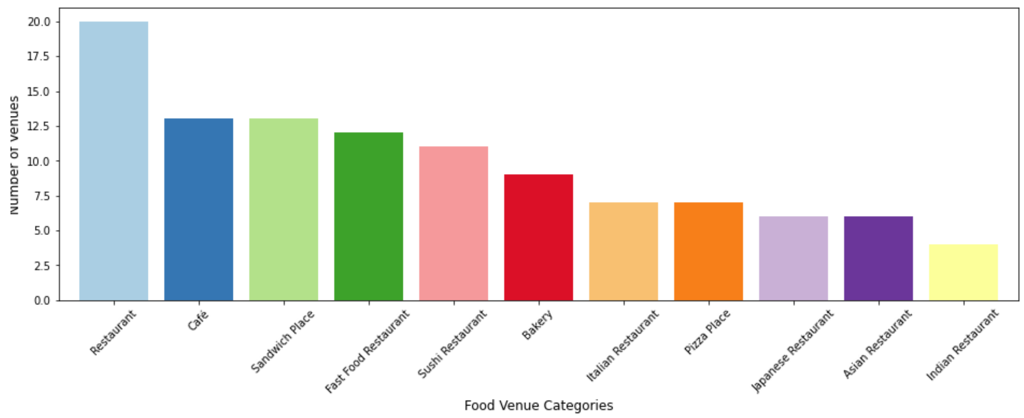
(a)



(b)



(c)



(d)

Figure 18. Total number of venues in topmost food venue categories in clusters: 0 – (a), 1 – (b), 2 – (c), 3 – (d).

The bar chart shows that a distinguishing feature of the zero cluster is the vast majority of Chinese and Vietnamese restaurants. Then a peculiarity of the first cluster is a few Caribbean restaurants while the restaurants of other types are represented by a single place. Next, the presence of a lot of pizza places is a unique characteristic of the second cluster. After that the third cluster is distinguished by various restaurants and fast food places. Moreover, the cuisine for overwhelming majority of restaurants is not specified. Finally, the last two clusters are presented with single venues which are Middle Eastern restaurant and American restaurant correspondingly. Thus on the basis of the analysis the following labels will be provided for clusters:

- Cluster 0 : "Chinese & Vietnamese Restaurants"
- Cluster 1 : "Caribbean Restaurants"
- Cluster 2 : "Pizza places"
- Cluster 3 : "Various Restaurants and Fast Food Places"
- Cluster 4 : "Middle Eastern Restaurant"
- Cluster 5 : "American Restaurant"

3.3.3 Visualization of clustered and non-clustered neighbourhoods

First of all, data frames of Toronto with clustered and non-clustered neighbourhoods with all gathered data for them are created. These data frames are shown in the Figure 19.

	Neighbourhood	Covid_Cases_Total	Assault_Rate_mean	Burglaries_Rate_mean	Robbery_Rate_mean	Crimes_Total	Longitude	Latitude	Cluster_Number	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Label
0	Agincourt North	527	244.798486	158.419463	94.582293	1067	-79.281616	43.797406	3	Sandwich Place	Asian Restaurant	NoVenue	NoVenue	NoVenue	Various Restaurants and Fast Food Places
1	Agincourt South-Malvern West	560	462.185571	314.380906	102.594551	1565	-79.289169	43.785187	2	Pizza Place	Chinese Restaurant	Restaurant	Vietnamese Restaurant	Deli / Bodega	Pizza places
2	Annex	310	833.750386	475.984500	152.495671	3346	-79.412147	43.674431	2	Café	Sandwich Place	Pizza Place	Burger Joint	Diner	Pizza places
3	Banbury-Don Mills	361	276.324557	234.616229	46.386083	1141	-79.326505	43.732570	3	Wings Joint	Restaurant	Food Truck	Deli / Bodega	NoVenue	Various Restaurants and Fast Food Places
4	Bathurst Manor	624	344.008300	184.767071	66.344721	698	-79.430503	43.757578	0	Chinese Restaurant	Café	NoVenue	NoVenue	NoVenue	Chinese & Vietnamese Restaurants

(a)

	Neighbourhood	Covid_Cases_Total	Assault_Rate_mean	Burglaries_Rate_mean	Robbery_Rate_mean	Crimes_Total	Longitude	Latitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alderwood	222	287.495757	209.450686	61.767394	497	-79.553204	43.595500	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue
1	Centennial Scarborough	258	332.824714	124.137040	42.164441	490	-79.145626	43.769728	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue
2	Cliffcrest	395	673.169171	281.115271	146.171497	1297	-79.225220	43.711273	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue
3	Humbermede	841	731.669271	221.528786	149.759330	1261	-79.549080	43.735709	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue
4	Long Branch	209	585.183371	283.050829	136.637676	759	-79.524738	43.589748	NoVenue	NoVenue	NoVenue	NoVenue	NoVenue

(b)

Figure 19. Data frames with clustered and non-clustered neighbourhoods in Toronto

Then The Folium library is used to visualize clustered and non-clustered Toronto neighbourhoods on the interactive map using their earlier obtained coordinates. Such a map is presented in Figure 20. The non-clustered neighbourhoods are designated with grey colour.

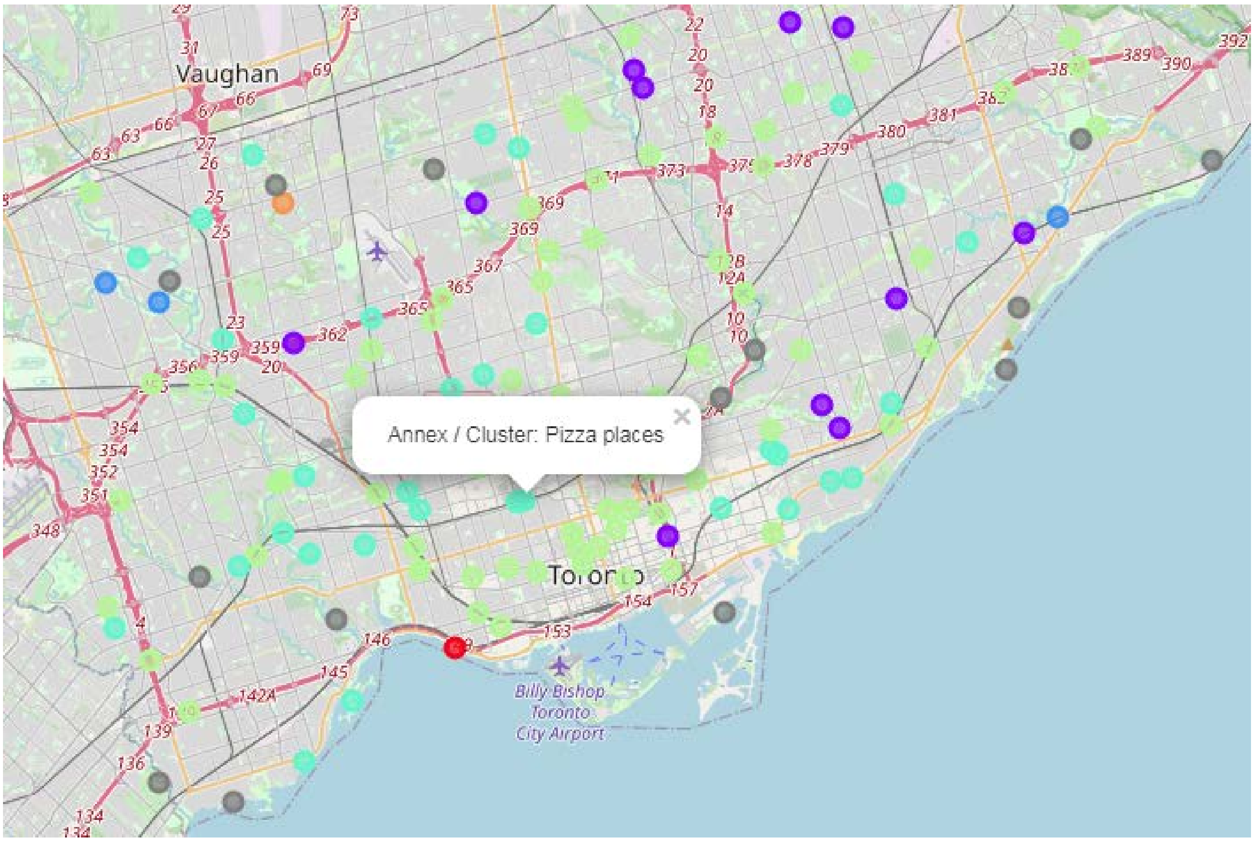


Figure 20. Map of Toronto with clustered and non-clustered Neighbourhoods

4 Results

Visualization of labeled Toronto neighbourhoods on choropleth maps allows estimating both risks concerned with crime rates and the ones due to rates of COVID-19 cases. The map of Toronto with labeled neighbourhoods representing an aggregate summary of COVID-19 cases occurred since March this year is shown in Figure 21.

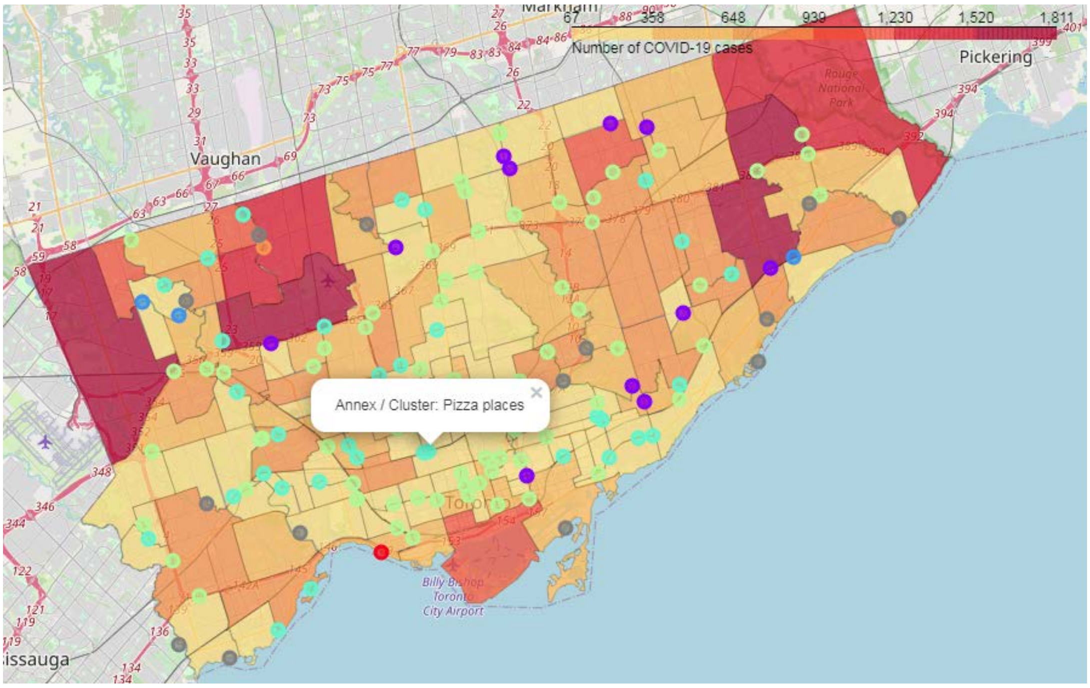


Figure 21. Map of Toronto with labeled neighbourhoods showing COVID-19 rate

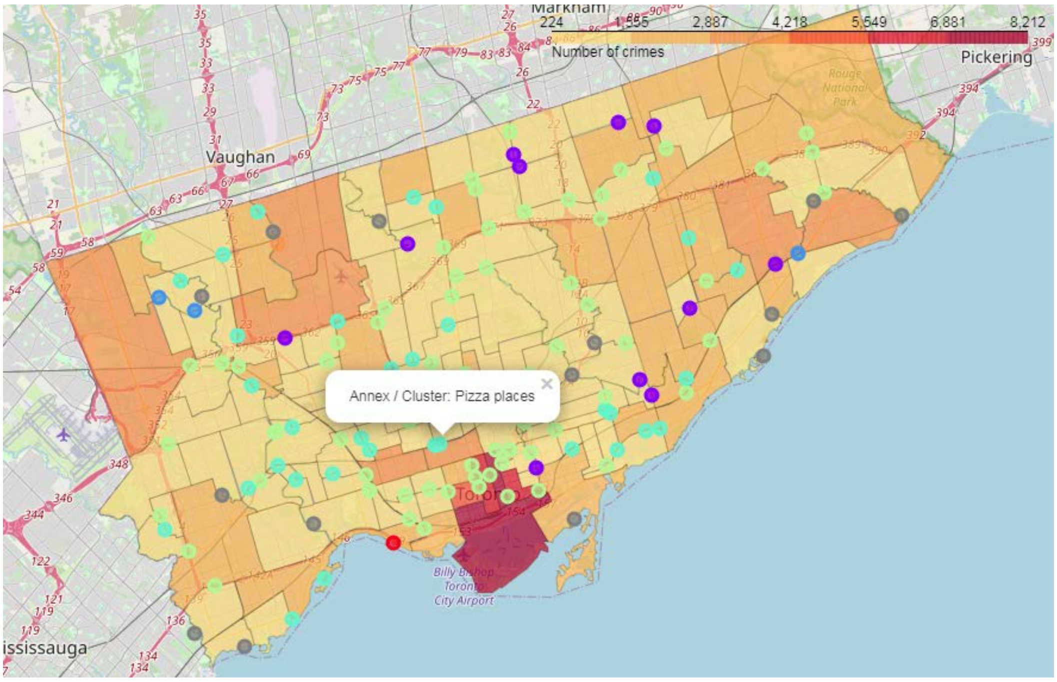


Figure 22. Map of Toronto with labeled neighbourhoods showing crime rate

The non-clustered neighbourhoods are designated with grey colour on both maps. The corresponding map with the crime rate of each type per 100,000 population might have been presented. However, it would not show significantly important results.

5 Discussion

The analysis shows that Toronto can offer a good opportunity to run a catering business. On the one hand a lot of food venues of different categories have already opened in the city. And it may seem that a serious competition is observed and owners of new food venues will be confronted with difficulties. However, having explored carefully the obtained data, a conclusion can be reached that there are many different opportunities for the new owners.

Firstly, there are several neighbourhoods in Toronto where there are no food venues at all. Certainly, the reasons why food venues are absent in them should be clarified. However, it is seen from the map with labeled Toronto neighbourhoods that the ones where no food venues exist are located closer to the outskirts of the city. Then it can be seen that venues of some categories are really in demand as for instance pizza and sandwich places, cafes and several types of restaurants, especially the Chinese, Italian, Sushi and fast food ones. Thus opening a venue of any of these categories will be quite risky in neighbourhoods where they already exist but the other ones can be considered.

It should also be taken into account that in many neighbourhoods there are only a few food venues. Moreover, the first two most common venues were considered in the bar charts with the total number of venues in topmost food venue categories. That means that only a few venues of less popular category may exist. So that the right decision will be to examine neighbourhoods which are not full of food venues. After that it will be useful to analyze what venues may fit in such neighbourhoods.

Finally, the crime rate and the COVID-19 rate were analyzed and shown on the choropleth maps. It is seen that Toronto is a really safe city and only in a few neighbourhoods the number of crimes is larger comparatively with the other ones. The number of COVID-19 cases is quite small for such a big city. Thus according to the obtained results most of neighbourhoods satisfy client's requirements on safety.

6 Conclusion

In the analysis various data on Toronto neighbourhoods were gathered and explored. The cluster analysis was applied to extract significant insights. The obtained results were assessed and visualized on the interactive maps. Observations to determine where a new public catering establishment can be open and what kind of food venue it can be were clarified.