Where to live: Exploring the Coolest Neighbourhoods in Toronto

Ramesh Baskaran

November 30, 2020

1.0 Introduction

Toronto's strength is based on a diverse and stable economy. It is the second fastest growing urban centre in North America and Canada's technology hub. This drives a steady stream of people moving into the city because of job opportunities. Living in Toronto depends on your preferences and lifestyle. Toronto is a mosaic of distinct neighbourhoods, each with its own identities and attracts. These neighbourhoods are continually changing, and boundaries can become blurred and contentious. An analysis will guide some of the critical characteristics of each locality within the city of Toronto and its suburbs.

Seeking a decent place to live can be one of the most daunting tasks when living in or moving to Toronto. From low crime rates to a great education system, there are many important factors to consider for that perfect place that you and your family can call home. Thus, it is worth investing time and effort in planning and researching key factors to help you determine which of the neighbourhoods in Toronto suits you and your family's needs.

A single bachelor condo will set you back between CA$1500-1600 a month in rent to live alone, and you can expect to have to compete with others just to get the rental. The rental market in Toronto is competitive, with demand continuing to outweigh supply. Before committing to a longer-term and more costly settlement, you need information such as the feasibility of leasing an apartment, security concerns, and amenities that will give you some time to find out the Toronto neighbourhood that you think is right for you and your family, if applicable.

The objective of this project is to find the safest and coolest neighbourhoods in Toronto to live in or relocate to. This project will consider crime, apartment rentals, population density, and amenities while comparing the similarities between the neighbourhoods and their local townships. These are crucial considerations to think about when deciding as to where to live. Security and affordability are the key decision-makers for relocation. Do you feel safe in this neighbourhood? Can you afford to live in a neighbourhood like that? And how close is it to mass transit? Where is the best place to live near the beach or the park? This project offers insight and addresses these concerns.

1.1 Target Audience

Who would be more interested in this project? What kind of clients or a group of people will benefit from this research?

- Prospective property buyers and tenants, whether they be business owners or residents who are looking to buy a property at a location.
- Local and global business investors searching for lucrative business opportunities must know who their customers are, what they want, where they live and what they can afford. This review would provide the current and new business owners

to tailor their marketing and sales efforts to specifically reach the segment of communities that will most likely buy their products or services.

- Real estate investment prospects for local and non-resident investors. As Toronto continues to resolve its affordable housing crisis, the city says it is now actively looking to identify partners, sites, and properties that could be sufficient to support the goal of creating tens of thousands of affordable rental homes over the next decade.
- Toronto City Councils are made up of municipalities and councillors responsible for making local decisions on managing townships. The research would consider local facts and conditions, particularly crime rates and rental prices. This can allow service managers to design and deliver various programmes and services to people experiencing safety issues and housing affordability.
- New immigrants to Toronto. Toronto is one of the most multicultural metropolitan areas in the world. Each year tens of thousands of immigrants from all over the world select the city as their new home. This study will provide knowledge and decision-making on which neighbourhood to settle down.

2.0 Data Collection

2.1 Sources of Data on Toronto Neighbourhoods and Venues Categories

a) List of Postal Code from Canada: M

I used the aforementioned wiki link to get all the details about the neighbourhoods in Toronto. This link includes the postal code, borough, and the name of all the neighbourhoods in Toronto. Unfortunately, I noticed that the wiki link could only acquire around 100 postal codes during the Course Lab assignment. As such, I continued to source elsewhere.

Toronto City Government Open Data defined and listed 140 neighbourhood boundaries. Thus, I added another 40 neighbourhood names to the earlier shortage of wiki datasets. Postal codes have been defined using the World Zip Code search engine. I have also managed to get the population density of each neighbourhood from this source.

b) Geospatial data, the .csv format provided during the Course Lab assignment, will be used to obtain the neighbourhood geographical coordinates.

c) The Foursquare API will be explored for location and other information about various venues in Toronto. It will give me ample details about the venues in terms of names, categories and locations (latitude and longitude).

2.2 Sources of Data for Crime

a) The Toronto Police Service Public Safety Data Portal will be accessed to get the crime data. The crime data is split into two components by type of crime, namely, Homicide and Major Crime Indicators(MCI). These datasets span the years 2004 to 2019. I will combine both datasets and choose the year 2019 as the most recent year to compare with other datasets.

b) I will also use GeoJSON data on Toronto to chart the total crime on the choropleth map later.

2.3 Sources of Data for Apartment Rental

a) I will use the dataset from Kaggle. The data was collected for the year 2018 and contained seven columns consisting of Bedroom, Bathroom, Den, Address, Latitude, Longitude and Price.

2.4 How the data will be used to solve the problem

First and foremost, I will try to get the 140 neighbourhoods with postal codes using both datasets from the wiki link and Open Data source. I will add in the population density of each neighbourhood. Then I will merge the crime dataset into the neighbourhood dataset. I can also measure the crime rate per 1000 residents as I have the total crime and population for each neighbourhood. After I have done that, I will use the geospatial data to get each neighbourhood's geographical coordinates. Then explore the Foursquare API to find the various venues according to their categories.

This is where the exciting part begins. I plan to combine neighbourhood data with crime rate, population density and the venue categories instead of just venue and neighbourhood data like we did in the Course Lab. I realize that many of the Capstone projects on GitHub have only clustered neighbourhood groups and merge with venue information. My key contribution to this project is the integration of additional details (i.e., crime rate and population density) into the clustering of neighbourhood groups. In other words, my clustering group will contain information on the neighbourhood, crime rate, population, and the 10 Most Common Venues.

The apartment rental dataset will be used to map the location of the rental places containing information on average prices, the number of bedrooms, whether or not they are overpriced, and the neighbourhood areas. I will use the overall crime count to plot on each neighbourhood boundaries using a choropleth map. Finally, I will combine all these maps into a single consolidated map and classify each cluster with its unique characteristics. I will identify the coolest neighbourhoods in Toronto with comprehensive details on crime rate, population density, 10 Most Common Venues, and average rental rates with the number of bedrooms.

3.0 Methodology

This section represents the project's key component, where the data is compiled and prepared for analysis. All null values, duplicate rows, and outliers will be omitted from the data used in the study. For the sake of brevity and to prevent code repetition, all the data will be preprocessed and ready to be analyzed in the Result section.

The strategy is based on the mapping of the data mentioned in Section 2.0. This visual approach and maps with popup labels allow quick identification of location, price and feature, thus making the selection very simple.

There are four maps to be constructed as follows:

a) Choropleth Heat Map of Total Crime

To show the crime data for Toronto, I will use Folium's choropleth map to show the density of recorded crime in the Toronto neighbourhoods. First, I will count the number of crime incidents in each neighbourhood to demonstrate the density of crime in the Toronto neighbourhood.

b) Circle Marker Map of Apartment Rental

As mentioned, this data was obtained from the Kaggle website. This data will be preprocessed and cleaned. I calculated and included another feature called OverPrice to see if rentals are overpriced based on the median values of each rental listed in the neighbourhoods.

c) Circle Marker Map of Clustered Neighbourhood

To get the 10 most common venue data, I will use the GET function to extract all the venues based in each borough from the Foursquare API and add them into the merged dataframe. I will use the geospatial geographical data for Toronto. Next, I will analyze each neighbourhood in each borough. Then I will group the neighbourhoods by taking the mean frequency of occurrence for venue type and add them into a new dataframe to display the top 10 venues for each neighbourhood.

To compare the venues in the neighbourhoods of Toronto, I will use K-Means clustering. It is the most common technique in unsupervised learning, where data is grouped based on the similarity of the data points. The algorithm can extract inferences from the nature of data objects and then create distinct classes to group them appropriately. Before using K-Means, we need to specify the value of K, which is the number of clusters we want to group our neighbourhoods into. Generally, we are not aware of the number of groups present in our neighbourhoods data, and it becomes challenging to evaluate the optimal value of K. Fortunately, we have several methods to determine the optimal K value. I will use the Elbow curve method to determine the optimal K values before running the cluster analysis.

d) Consolidated Map of the Toronto Neighbourhood

I am combining all the above maps into one solid map. In other words, I plan to superimpose the total crime, apartment rentals, and cluster maps on a single consolidated map. This will show some critical insights into the characteristics of neighbourhoods and, possibly some connections between population, crime rate, apartment rentals, and venue categories.

4.0 Results and Discussion

4.1 Results

4.1.1 Choropleth Heat Map of Total Crime

The Toronto Homicide Data and the Toronto MCI Data were preprocessed and then merged into a standard dataframe . The Toronto Crime Data covers eight types of crime. This dataframe only contains data from the year 2019.

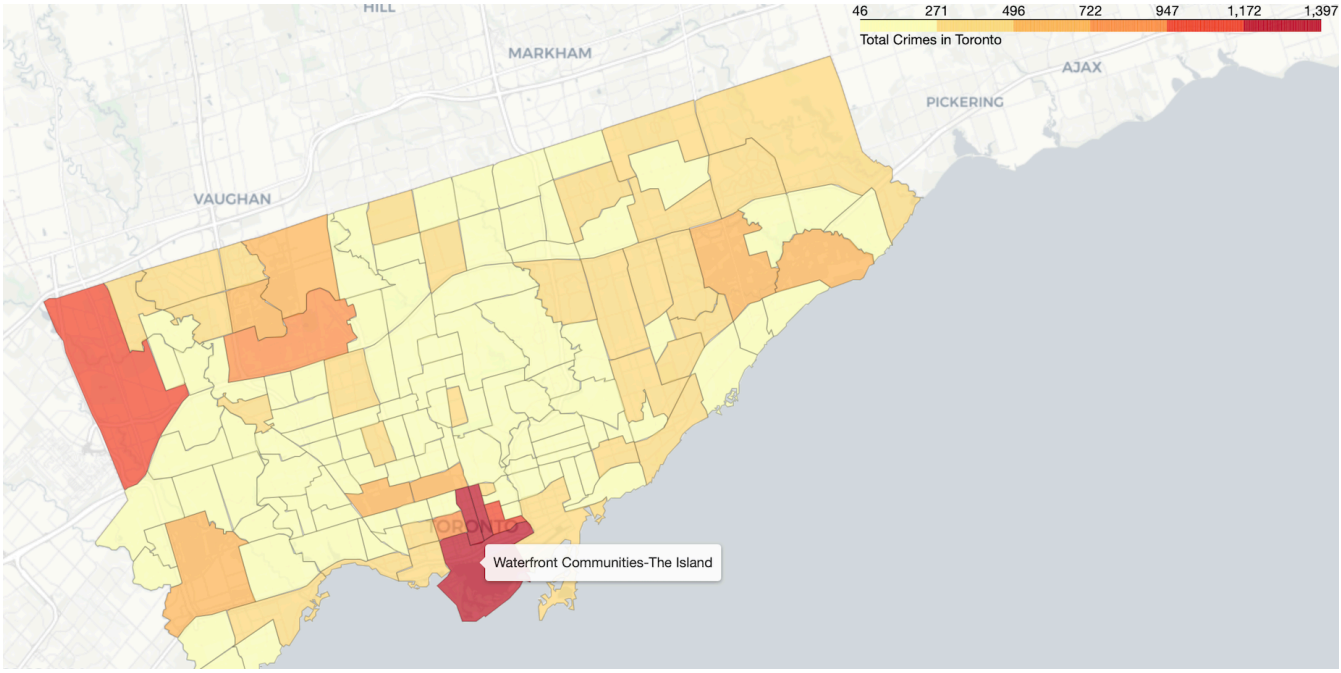


Image 1: Total Crime for 2019 in Toronto

Image 1 shows the frequency count of Total Crime for 2019 in Toronto. The lighter colours suggest a lower crime count, and the darker colours indicate a higher crime count. This map shows which neighbourhoods registered higher crime rates in Toronto in 2019. As can be seen clearly, there is only one neighbourhood called Waterfront Communities -The Island, which has the highest crime density relative to other low-density areas. This map also displays a popup that shows the respective neighbourhood's name corresponding to its crime count.

4.1.2 Circle Marker Map of Apartment Rental

Apartment rental data was obtained from Kaggle with a .csv format. The data included rental prices, addresses, latitude, and longitude. The addresses did not have postal codes, so I reverse geocoded the latitude and longitude values using Nominatim and RateLimiter. The reason is to try to match the postal code with the 140-area neighbourhood postal code. Unfortunately, some of the rentals were not listed in some neighbourhoods. Due to lack of rental listing in all the neighbourhoods, I could not include

or merge the rental data for clustering. There were so many apartments listed in the same building or street, so I decided to average the price based on the number of bedrooms.

I have also calculated if the property is overpriced based on median values. Importantly, median values are less likely to be skewed by very high or very low rental rates. In any event, where a property's rental rate is markedly different from the median value (i.e., overprice), it's a signal that you should question and investigate further the reasons for the difference, which will help you assess whether they are reasonable or not.

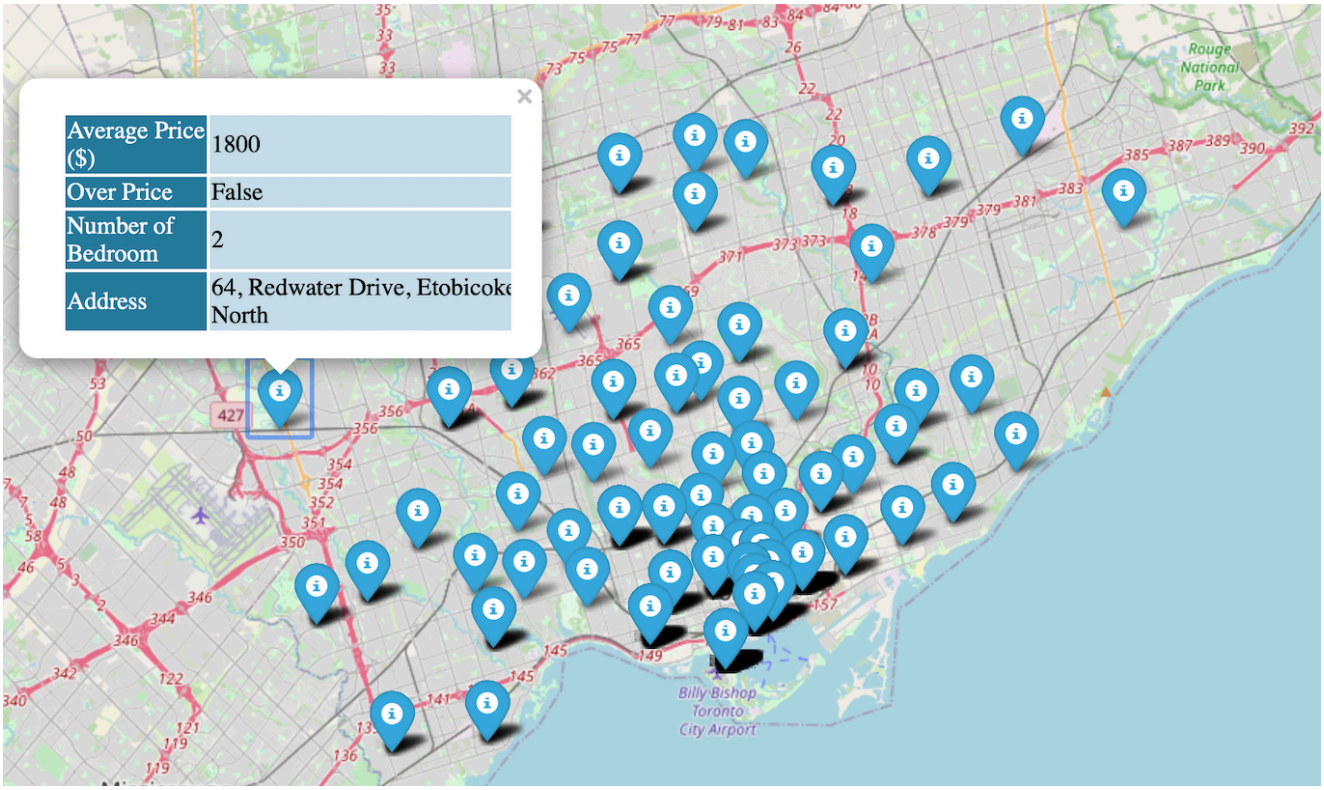
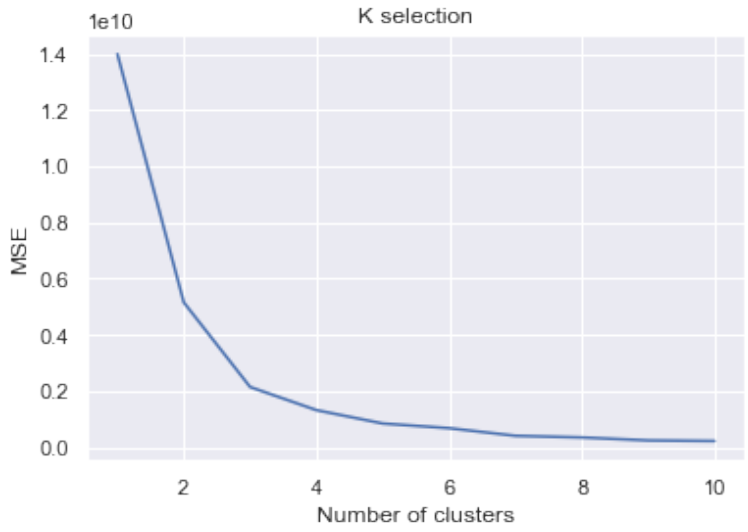


Image 2: Apartment Rental 2018 in Toronto

Image 2 displays the apartment rentals 2018 in Toronto. As you can see, the popup menu shows the average rental price based on the number of bedrooms, the street address and whether or not the property is overpriced.

4.1.3 Circle Marker Map of Clustered Neighbourhood

I used the Elbow method to evaluate the best value of K. The mean squared error (MSE) is plotted by the number of clusters. The elbow in the graph below shows that the optimum value of K is 3.



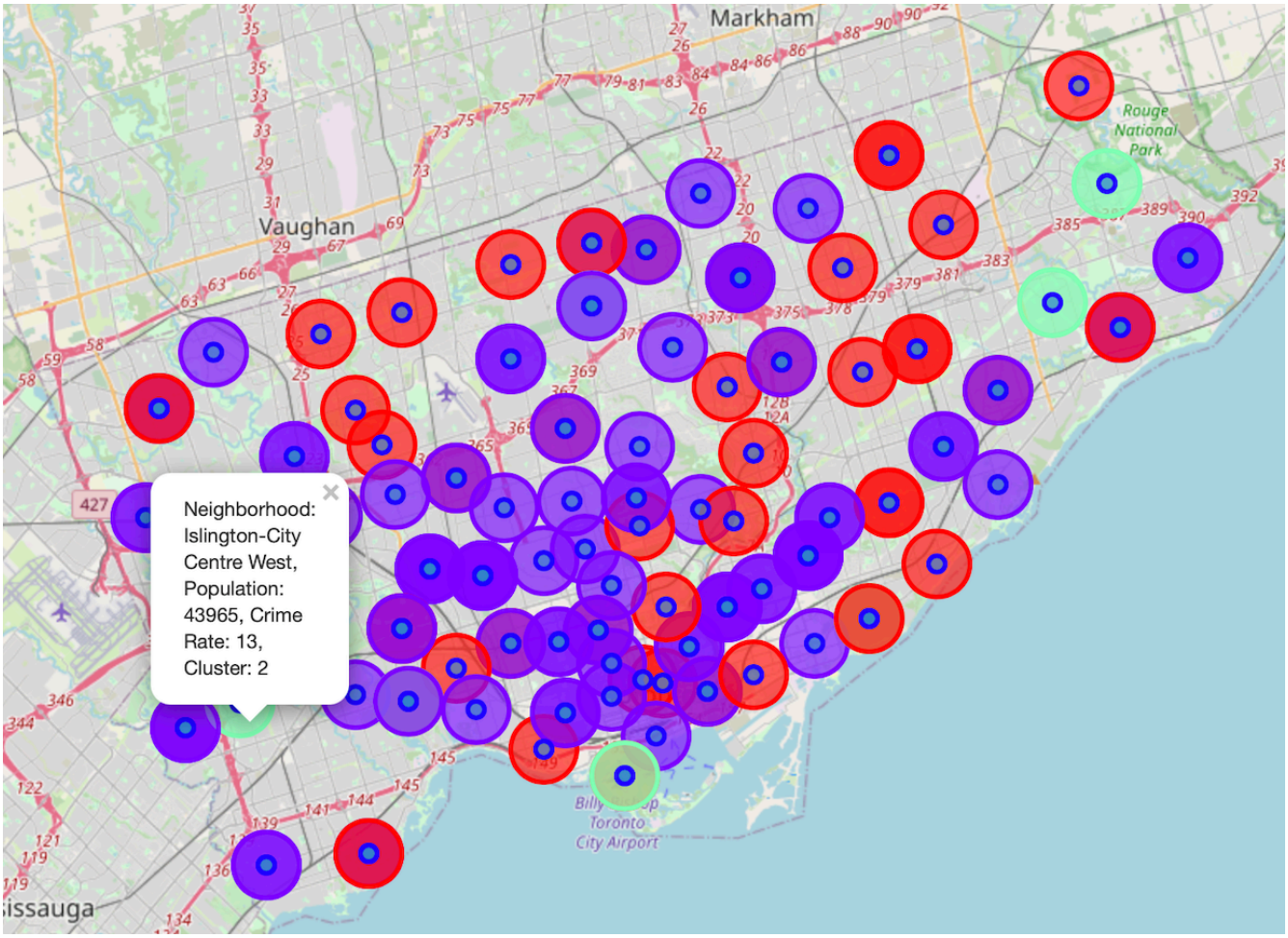


Image 3: Clustered Neighbourhoods in Toronto

Image 3 visualizes the Toronto neighbourhood clusters. In this image, you can see how the different clusters are located in Toronto. There are three clusters in Toronto. The red markers of Cluster 1 are made up of 48 neighbourhoods. The purple markers of Cluster

2 consist of 85 neighbourhoods. The green markers represent Cluster 3,[1] which comprises of 7 neighbourhoods. It is important to note that even though Cluster 3 displays four markers on the map, this means that there are few neighbourhoods in a single marker. The majority of markers belong to Cluster 2. The popup menu displays the neighbourhood's name, the population density based on the Census 2016, the crime rate per 1000 residents, and the cluster number.

4.1.4 Consolidated Map of the Toronto Neighbourhood

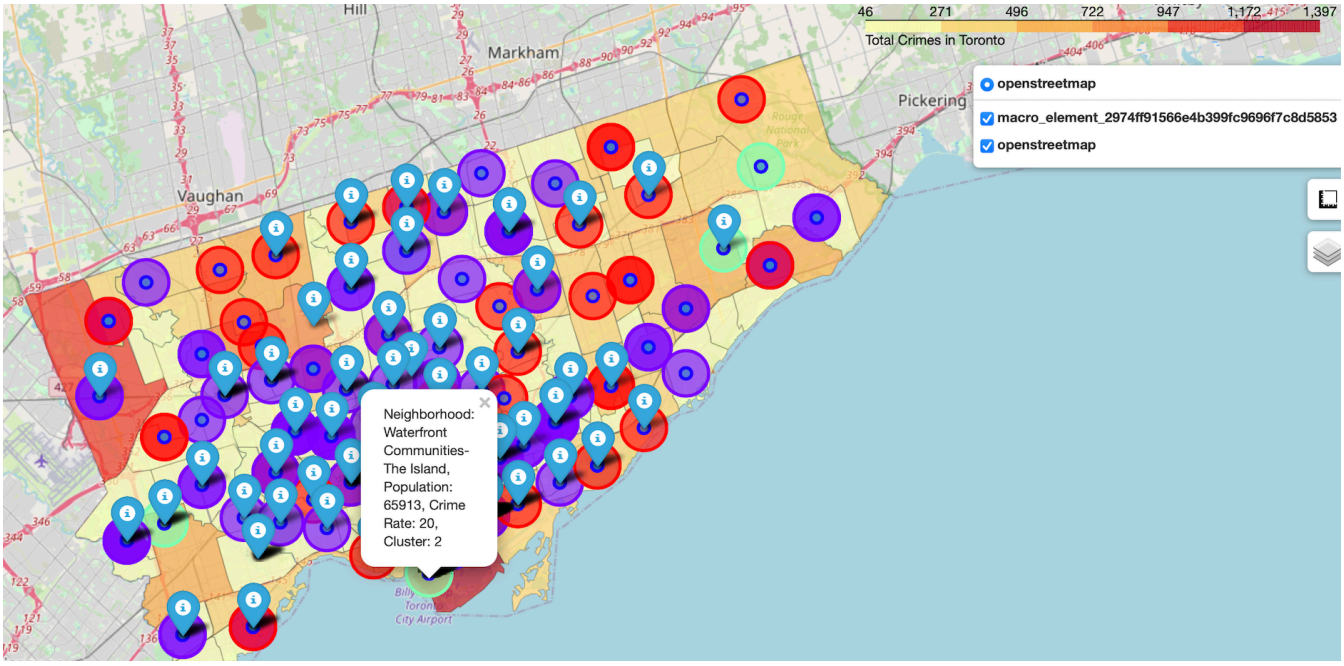


Image 4: Clustered Neighbourhoods with Total Crime and Apartment Rental

In order to visualize the possible relationship between total crime (Image 1), apartment rentals (Image 2) and neighbourhoods with crime rates and population density (Image 3), I have superimposed all these images into a single consolidated map, as shown in Image 4. Here, the Toronto Map shows both the crime and population densities in the neighbourhoods and the rental rates with an overpriced indicator in the city.

[1]Cluster modeling output defines the clusters as 0, 1, and 2. Hereby, the number of clusters are predefined or labelled as 0 = Cluster 1; 1 = Cluster 2; and 2 = Cluster 3. In other words, the green markers in the map indicates Cluster 2 which is predefined and referred to as Cluster 3 in the article.

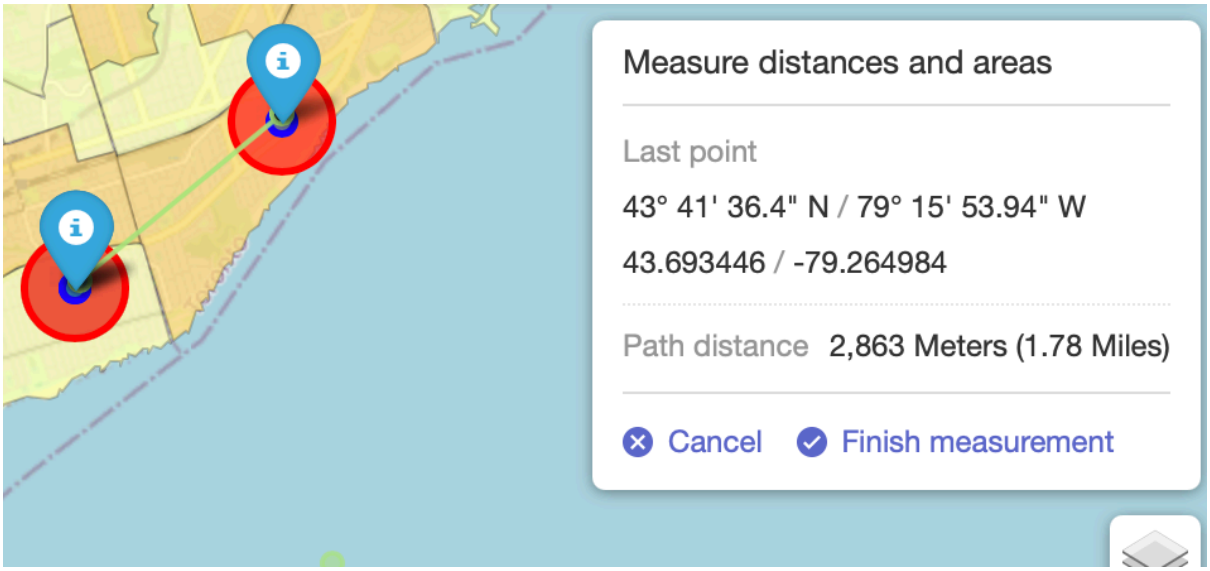


Image 5: Distance Measurement

Another interesting aspect of the map is that I have included a linear distance measurement layout that allows you to measure the distance from one position to another. For example, as seen in Image 5, the linear distance measurement from 174 Scarborough Rd to 45 Rhydwen Ave, Scarborough is approximately 2.9 km. Google Map indicates the direction distance(driving) between these streets is 3.1 km. The distance measuring difference between Google map and Image 5 is around +- 2 km. In a nutshell, you can easily calculate the distance to commute back and forth of the areas you are interested in.

4.2 Discussion

	Cluster 1	Cluster 2	Cluster 3
Neighborhoods	48	85	7
Average Population 2016	26180	13257	49726
Average Crime Rate Per 1000	15	12	10
1st Most Common Venue	Coffee Shop(6)	Park(11)	Print Shop(1)
2nd Most Common Venue	Park(4)	Coffee Shop(13)	Women's Store(2)
3rd Most Common Venue	Park(4)	Sandwich Place(9)	Pub(2)
4th Most Common Venue	Bakery(6)	Women's Store(7)	Eastern European Restaurant(1)
5th Most Common Venue	Women's Store(5)	Women's Store(9)	Dim Sum Restaurant(2)
6th Most Common Venue	Dumpling Restaurant(3)	Dumpling Restaurant(11)	Dumpling Restaurant(1)
7th Most Common Venue	Intersection(4)	Drugstore(11)	Donut Shop(1)
8th Most Common Venue	Donut Shop(5)	Donut Shop(12)	Donut Shop(1)
9th Most Common Venue	Dog Run(4)	Doner Restaurant(11)	Dog Run(1)
10th Most Common Venue	Diner(5)	Dog Run(13)	Dog Run(1)

The number in the parentheses shows the frequencies of the items.

Table 1: Summary of Clusters

The summary of clusters is shown in Table 1. There is a plurality of the neighbourhoods in Cluster 2, followed by Cluster 1, and Cluster 3 has the least number of neighbourhoods. As you can see from the map, Cluster 3 is spread out compared to Clusters 1 and 2, where they are clustered close together. As a result, Clusters 1 and 2 tend to have more or less the same common categories as Parks, Coffee Shops, Women Stores, Dumpling Restaurants, Dog Run, and Sandwich or Bakery shops. Cluster 3 is less frequented by Women Stores, Dim Sum Restaurants, Pubs, and Dog Run. In other words, less social activities in Cluster 3 neighbourhoods. There are distinct variations in lifestyle and neighbourhood characteristics between Cluster 3 and other clusters. Notice that Cluster 3 also has the highest population density. More data is needed to determine if this suburban cluster is more of a quiet countryside living and its demographic segmentation. This will have to be further examined.

Thus, the information from Table 1 may not help to make choices about the coolest neighbourhood to live in. What characteristics of the neighbourhood make it the coolest place to live in based on this project analysis?

Neighborhood	Total Crime	Population 2016	Cluster	1 Bedroom		2 Bedroom	
Guildwood	46	9917	2	1,285	a	1,434	a
Woodbine-Lumsden	54	7865	2	**		**	
Lambton Baby Point	57	43993	3	1,147	b	1,616	c
Markland Wood	67	10554	2	**		**	
Old East York	71	9233	2	1,099	b	1,330	b
Yonge-St.Clair	79	12528	2	1,504	b	2,058	c
Humber Heights-Westmount	81	10948	2	**		1,569	b
Maple Leaf	81	10111	2	1,252	c	1,301	c
Etobicoke West Mall	83	11848	2	1,415	b	1,664	a
Alderwood	85	12054	2	1,193	b	1,357	c
Lawrence Park North	85	14607	2	1,302	b	1,616	b
Bayview Woods-Steeles	90	13154	2	**		**	
Kingsway South	95	9271	2	1,525	d	1,901	d
Pleasant View	95	15818	2	**		**	
Casa Loma	96	10968	2	1,592	a	2,218	c
Centennial Scarborough	99	13362	2	**		**	
Forest Hill South	99	10732	2	1,508	c	1,888	b

The following letter codes are used to indicate the reliability of the estimates: a - Excellent, b- Very good, c - Good, d - Fair (Use with Caution).
** Data suppressed to protect confidentiality or data not statistically reliable.

Table 2: Neighbourhoods with Lowest Crime Rate

Table 2 displays the details needed to make wise decisions about the best neighbourhood to live in. I selected neighbourhoods with a cumulative crime of less than 100 recorded incidents in 2019. These neighbourhoods are then identified into Clusters 1, 2, and 3. I find that most of the low crime rate neighbourhoods are in Cluster 2. Only the Lambton Baby Point neighbourhood is in Cluster 3. None in Cluster 1 qualifies for low crime rates. I have also included the latest average values of apartment rentals from an external source. The data on average rent by bedroom type by neighbourhood was obtained from Canada Mortgage and Housing Corporation based on October 2019 rental rates. I observe that these neighbourhoods' average rental rates are far below the 2018 dataset average rates.

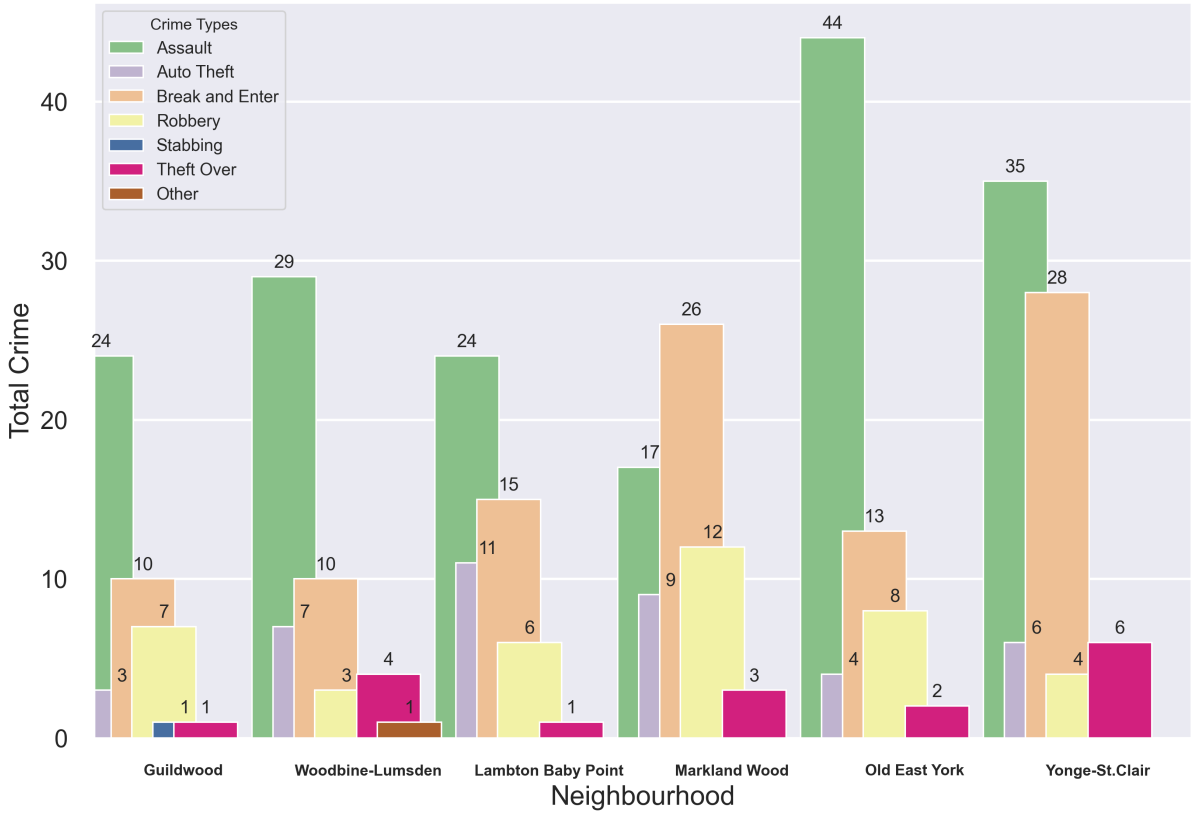


Chart 1: Top Safest Neighbourhoods Crime Types

Chart 1 shows the total crime for 2019 according to crime types for top safest neighbourhoods. Assault registered the highest cases, followed by Break and Enter, Robbery and Auto Theft. These incidences account for about 80% of the crime reported. The study recommends that the Toronto Police Services Board and the Neighbourhood Watch groups be encouraged to continuously increase their surveillance and monitoring functions in Toronto.

Neighbourhood	Walkability	Bikeability	Public Transit	Affordability	Greenspace	Recreation
Guildwood	Good	Excellent	Okay	$$$$	Excellent	Good
Woodbine-Lumsden	Excellent	Good	Excellent	$$$	Good	Excellent
Lambton Baby Point	Excellent	Excellent	Good	$$$$	Excellent	Excellent
Markland Wood	Good	Good	Good	$$$$	Excellent	Excellent
Old East York	Excellent	Good	Excellent	$$$	Good	Excellent
Yonge-St.Clair	Excellent	Excellent	Excellent	$$$$	Good	Excellent

Statistics have been obtained from the Neighbourhood Guide.

Table 3: Top Safest Neighbourhoods in Toronto

Table 3 lists Toronto's top six safest neighbourhoods to live in and their rated amenities. Lambton Baby Point and Yonge-St. Clair are the best neighbourhoods to live unless affordability is an issue. Based on the study, I find that the neighbourhoods stated in Table 3 are the safest and coolest places to live in or move to.

5.0 Conclusion

The project's objective was to identify which neighbourhoods in Toronto are the safest and coolest places to live in. My main contribution to this project is incorporating additional data like crime rate and population density into neighbourhood clustering. Subsequently, merge the cluster neighbourhoods with the 10 most common venues. Besides, I have also included apartment rental data in the analysis. After conducting data preprocessing and data exploratory analysis, the study runs the unsupervised K-Means algorithm to cluster the segmentation of neighbourhoods' similarities into 3 clusters. The number of clusters was decided using the Elbow method for optimal K. Interactive maps were constructed to visualize these analyses to help better decide the best neighbourhood to live in.

Based on the study, Guildwood, Woodbine-Lumsden, Lambton Baby Point, Markland Wood, Old East York, and Yonge-St. Clair are the best neighbourhoods to live in or relocate with the lowest crime rates and well-rated amenities. Among them, Lambton Baby Point and Yonge-St. Clair are the top two coolest neighbourhoods to live in subject to affordability.

Stakeholders will make the final decision on best neighbourhoods based on their preferences and lifestyles, the unique characteristics of locations such as proximity to the park or the beach, noise levels, proximity to public transport, affordability of real estate, prices, the social and economic dynamics of each neighbourhood. This project will somehow aid stakeholders in narrowing their quest for the best residential neighbourhoods.