Applied Data Science Capstone – Week 4 – Capstone Project – The Battle of Neighborhoods (Week 2) – Final Report

Project Title:

The Battle of Neighborhoods (Week 2) – Final Report

Section 1: Introduction and Business Problem

The purpose of this data project is to explore the popular venues or facilities in different neighborhoods in Toronto. Toronto was the most populous metropolitan area in Canada in 2019, with a population or around 6.47 million people (Statista, 2020). Toronto attracts many tourists and new immigrants every year because it is highly economically developed and is one of the most vibrant cities in North America.

Stakeholders. This data project targets these tourists and immigrants who consider moving to Toronto for either recreational visits (travelling and sightseeing) or business purposes (opening and expanding to a new business). Primarily, tourists would like to find popular venues, scenic spots, or superstar cafes and restaurants to visit, while immigrants may find neighborhood(s) with less intense competition in their own interested fields of businesses. For example, an immigrant who runs grocery stores may find a neighborhood with fewer grocery stores of similar kinds more attractive to start their very first business in Canada. All these procedures are costly in terms of time and resources.

Business Problems and Data Project Objectives. This data project is meant to help these groups of stakeholders (tourists and new immigrants) to learn more about different neighborhoods in Toronto. By comparing different neighborhoods in Toronto, these stakeholders would obtain information on which types of venues (e.g. cafes, restaurants, rental car locations, banks, gas stations, gyms, etc.) are the most (or least) popular. With this analysis, stakeholders could decide which neighborhoods have the features of venues that they would like to see. We assume that tourists are more interested in neighborhoods with more popular restaurants and/or tourist spots, while immigrants are more interested in neighborhoods with less intense competition of venues in their preferred fields of businesses.

Toronto is also well known for its higher housing prices. It may affect both tourists and immigrants. Tourists may find hotel or Airbnb accommodation more expensive in certain neighborhoods, while immigrants may find their homes less affordable in certain neighborhoods. This would affect their decisions whether to visit or stay in a neighborhood or not. We will use housing price data in Toronto to give suggestions to our stakeholders, especially those who are sensitive to accommodation/rental expenditures.

Recently, Toronto has been hit hard by the global Covid-19 pandemic. We see soaring Covid-19 active cases and death tolls, unfortunately. We expect that tourists and immigrants may also want to factor in the overall health and safety conditions into their consideration whether to move into a neighborhood or not. We will also use the Covid-19 cases data to provide insights from public health perspectives to potential tourists and immigrants who may want to avoid visiting the hardest hit neighborhoods in Toronto.

Section 2: Data Sources and Descriptions

Data Sources. In this data project, we will use 5 data sources:

- List of Postal Codes of Canada, obtained from Wikipedia. This data is the same as what we used in previous weeks of this course. The URL is: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. It contains the information of Postal Codes, Boroughs and Neighborhoods in Toronto. There are 103 neighborhoods.
- House Price Data for all neighborhoods in Toronto sorted by Postal Codes, obtained from House Price Hub. The URL is: https://housepricehub.com/cities/city/Toronto. It contains the information of Postal Codes, Average House Prices for all Postal Codes in Toronto.
- Covid-19 Cases for all neighborhoods in Toronto sorted by Postal Codes, obtained from Open Data Portal Toronto. The URL is: https://open.toronto.ca/dataset/covid-19-cases-in-toronto/.
- Geospatial Data (Geographical Coordinates of Each Postal Code in Toronto). This data file is taken from the previous weeks of this course. The URL is: http://cocl.us/Geospatial_data. It contains the information of Postal Codes, Latitude and Longitudes for each of these Postal Codes.
- Foursquare Social Location Service Data, obtained from Foursquare Developer account and using API requests. The requests can be made by specifying Client ID, Client Secret, Version, Latitude and Longitude (of Neighborhoods that you want to search for), Radius, and Limit (number of venues returned by Foursquare API).

Data Descriptions. The first 4 data sources have the following variables:

- Source 1: Toronto, columns = {PostalCode, Borough, Neighborhood}
- Source 2: houseprice, columns = {PostalCode, AvgPrice}

- Source 3: covid, columns = {PostalCode, CovidCases}
- Source 4: geocode, columns = {PostalCode, Latitude, Longitude}

We will then clean each of these dataframes and join/merge all these sources into one single dataframe, which we call ‘df’ containing: columns = {PostalCode, Borough, Neighborhood, Latitude, Longitude, AvgPrice, Count}. The column Count refers to the Covid-19 cases in each neighborhood. Note that we focus on confirmed cases only, probable cases are excluded.

Borough Selection. To simplify our analysis, we focus on the neighborhoods in Scarborough, Toronto. The geographical coordinate of Scarborough, Toronto are 43.773077, -79.257774. There are two reasons to focus on Scarborough. First, Scarborough is a popular destination for new immigrants in Canada, making it one of the most diverse and multicultural areas in the Greater Toronto Area. It particularly suits our context of providing location-based information to new immigrants defined in our Introduction and Business Problem section. Second, there are no missing values of Average House Price and Covid Cases data for neighborhoods in Scarborough, so that we can preserve the most comprehensive information of neighborhoods in Scarborough. A previous version of this project focuses on Downtown Toronto, but then Downtown Toronto seems to have more missing values of Average House Price and Covid Cases data for its neighborhoods.

More Data Descriptions on Foursquare API. By making the Foursquare API requests, we can obtain the detailed information of popular venues within a specified radius of a specified neighborhood (with latitude and longitude values). Take our API requests for Malvern, Scarborough as an example. It has 43.806686…, -79.194354… as neighborhood latitude and longitude values. For the top venues, it gives ‘SEPHORA’ (the store name) as the name of venue, ‘300 Borough Drive’ as the location, 43.775016…, -79.258109… as the latitude ang longitude values, ‘217’ as the distance, ‘M1P 4P5’ as the 6-digit postal code, ‘Cosmetics Shops’ as the category of venue, and etc.

There are, of course, much more information from Foursquare, including menus (for places like restaurants, cafes, etc.), photos, and comments, for all these venues. We restrict the radius to be 1000 meters.

Python Libraries and Packages. This data project requires the following dependencies: NumPy (to handle data in a vectorized manner), Pandas (for data analysis), JSON (to handle JSON files), XML (to process XML), Geocoder (to convert an address into latitude and longitude), Requests (library to handle requests), Matplotlib (plotting tools), Scikit-Learn (use k-means clustering), Beautiful Soup (for parsing HTML and XML documents), and Folium (map rendering library).

Section 3: Methodology

Explore Nearby Venues and One Hot Encoding. For each of the neighborhoods in Scarborough, we make use of a function called getNearbyVenues to obtain the nearest venue categories for all neighborhoods. To further analyze each neighborhood, we use one hot encoding to get dummies for each of the unique venue categories for each of the neighborhoods. With this information, we can find the most popular (or least popular if we wish) venue categories. To simplify analysis, we drop venue categories with 0.5% share of all categories. This will speed up our computational analysis.

Finding the Most and Least Common Venues for Each Neighborhood. With the above information obtained from Foursquare API, we can find the most and least common venues for each neighborhood. This is very helpful for both tourists and new immigrants, because tourists might look for popular spots to visits, while new immigrants might look for least popular spots which are good to start a business without much competition. Take Scarborough, Toronto as an example. The Neighborhood of Agincourt has the following 3 most common venues:

- Skating Rink,
- Breakfast Spot,
- Latin American Restaurant,

while it has the following 3 least common venues:

- Accessories Store,
- Indian Restaurant,
- Intersection.

Machine Learning: K-Means Clustering Approach and Segmentation of Neighborhoods. To compare the similarities (or dissimilarities) between different neighborhoods, we will make use of k-Means Clustering Approach to segment and categorize all these neighborhoods into clusters based on the similarities of their venues' characteristics. We set the number of clusters to be five. Although we do not observe neighborhoods to be equally distributed in each of the clusters, we have at least one neighborhood in each cluster, which is nice. The clustering approach and the resulting segmentation of neighborhoods also serve as important information to tourists and new immigrants, as the clustering guides them to choose similar neighborhoods that satisfy their needs (either venues for tourists, or venues to start a business). A complete set of results will be presented in the next section.

Exploratory Analysis: Graphical and Visualization Analysis. After we learn which neighborhoods are similar (and dissimilar) from the segmentation of neighborhoods. We take alternative perspectives to look at different neighborhoods, which are housing prices and number of confirmed Covid-19 cases. They are helpful in capturing (roughly) the accommodation or rental expenditures, and public health and safety level of each of the postal code location for our stakeholders.

Section 4: Results and Discussions

Part (a): Explore Nearby Venues and One Hot Encoding.

One Hot Encoding. The following table shows whether the neighborhoods have a specific kind of venue categories. 1 means yes, and 0 means no. Overall, restaurants seem to be quite popular, but accessories stores are absent in most of these neighborhoods.

	Neighborhood	Accessories Store	American Restaurant	Athletics & Sports	Auto Garage	Bakery	Bank	Bar	Breakfast Spot	Bus Line	Bus Station	Café	Caribbean Restaurant
0	Malvern, Rouge	0	0	0	0	0	0	0	0	0	0	0	0
1	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	0	1	0	0	0	0	0
2	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	1	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
5	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
6	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
7	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
8	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0
9	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0

	Neighborhood	Chinese Restaurant	Clothing Store	Coffee Shop	College Stadium	Construction & Landscaping	Department Store	Electronics Store	Fast Food Restaurant	Fried Chicken Joint	Gas Station	Genera Entertainmen
0	Malvern, Rouge	0	0	0	0	0	0	0	1	0	0	
1	Rouge Hill, Port Union, Highland Creek	0	0	0	0	0	0	0	0	0	0	
2	Rouge Hill, Port Union, Highland Creek	0	0	0	0	1	0	0	0	0	0	
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	1	0	0	0	
5	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	
6	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	
7	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	
8	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	
9	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	

One Hot Encoding and Distribution of Venue Categories of Different Neighborhoods. The following table shows the distributions of venue categories for each of the neighborhoods. We observe large variations across venue categories in different neighborhoods.

	Neighborhood	Accessories Store	American Restaurant	Athletics & Sports	Auto Garage	Bakery	Bank	Bar	Breakfast Spot	Bus Line	Bus Station	Café	Caribbean Restaurant	Chinese Restaurant	Clothing Store	Coffee Shop	College Stadium	Construction & Landscaping	Department Store	Electronics Store	Fast Food Restaurant	Fried Chicken Joint	Gas Station	G Entertai
0	Agincourt	0.0	0.0	0.000	0.0	0.000	0.000000	0.0	0.200	0.0	0.0	0.00	0.000	0.000000	0.2	0.00	0.00	0.0	0.00	0.000	0.000000	0.000000	0.000000	
1	Birch Cliff, Cliffside West	0.0	0.0	0.000	0.0	0.000	0.000000	0.0	0.000	0.0	0.0	0.25	0.000	0.000000	0.0	0.00	0.25	0.0	0.00	0.000	0.000000	0.000000	0.000000	
2	Cedarbrae	0.0	0.0	0.125	0.0	0.125	0.125000	0.0	0.000	0.0	0.0	0.00	0.125	0.000000	0.0	0.00	0.00	0.0	0.00	0.000	0.000000	0.125000	0.125000	
3	Clarks Corners, Tam O'Shanter, Sullivan	0.0	0.0	0.000	0.0	0.000	0.083333	0.0	0.000	0.0	0.0	0.00	0.000	0.083333	0.0	0.00	0.00	0.0	0.00	0.000	0.083333	0.083333	0.083333	
4	Cliffside, Cliffcrest, Scarborough Village West	0.0	0.5	0.000	0.0	0.000	0.000000	0.0	0.000	0.0	0.0	0.00	0.000	0.000000	0.0	0.00	0.00	0.0	0.00	0.000	0.000000	0.000000	0.000000	
5	Dorset Park, Wexford Heights, Scarborough Town...	0.0	0.0	0.000	0.0	0.000	0.000000	0.0	0.000	0.0	0.0	0.00	0.000	0.200000	0.0	0.00	0.00	0.0	0.00	0.000	0.000000	0.000000	0.000000	
6	Golden Mile, Clairlea, Oakridge	0.0	0.0	0.000	0.0	0.200	0.000000	0.0	0.000	0.2	0.1	0.00	0.000	0.000000	0.0	0.00	0.00	0.0	0.00	0.000	0.000000	0.000000	0.000000	
7	Guildwood, Morningside, West Hill	0.0	0.0	0.000	0.0	0.000	0.125000	0.0	0.125	0.0	0.0	0.00	0.000	0.000000	0.0	0.00	0.00	0.0	0.00	0.125	0.000000	0.000000	0.000000	
8	Kennedy Park, Ionview, East Birchmount Park	0.0	0.0	0.000	0.0	0.000	0.000000	0.0	0.000	0.0	0.0	0.00	0.000	0.000000	0.0	0.25	0.00	0.0	0.25	0.000	0.000000	0.000000	0.000000	
9	Malvern, Rouge	0.0	0.0	0.000	0.0	0.000	0.000000	0.0	0.000	0.0	0.0	0.00	0.000	0.000000	0.0	0.00	0.00	0.0	0.00	0.000	1.000000	0.000000	0.000000	

Part (b): Finding the Most and Least Common Venues for Each Neighborhood.

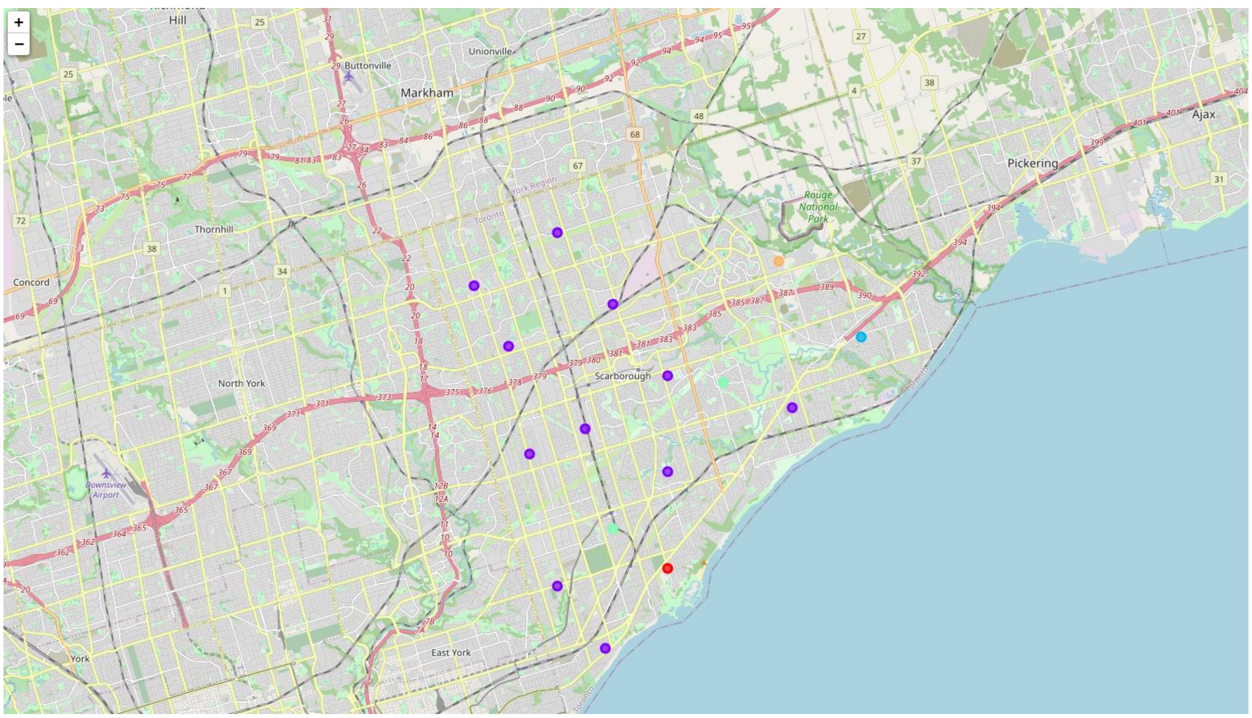
The Most Common Venues. The most common venues of each of these selected neighborhoods look very different. Most venues are entertainment, tourism, and food & beverage related. A few examples are: skating rink, (Hakka/Indian/Fast Food) restaurants, motel and train station. These neighborhoods seem to be great for tourists!

	Neighborhood	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
0	Agincourt	Skating Rink	Breakfast Spot	Latin American Restaurant	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Gas Station
1	Birch Cliff, Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym	Grocery Store	Gas Station	Fried Chicken Joint
2	Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym
3	Clarks Corners, Tam O'Shanter, Sullivan	Pizza Place	Chinese Restaurant	Noodle House	Thai Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Intersection	Bank	Italian Restaurant
4	Cliffside, Cliffcrest, Scarborough Village West	Motel	American Restaurant	Vietnamese Restaurant	Gym	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
5	Dorset Park, Wexford Heights, Scarborough Town...	Indian Restaurant	Vietnamese Restaurant	Pet Store	Chinese Restaurant	Auto Garage	Bakery	American Restaurant	Grocery Store	General Entertainment	Gas Station
6	Golden Mile, Clairlea, Oakridge	Bakery	Bus Line	Ice Cream Shop	Park	Bus Station	Metro Station	Intersection	Soccer Field	Fried Chicken Joint	Coffee Shop
7	Guildwood, Morningside, West Hill	Rental Car Location	Electronics Store	Medical Center	Intersection	Bank	Restaurant	Mexican Restaurant	Breakfast Spot	College Stadium	Construction & Landscaping
8	Kennedy Park, Ionview, East Birchmount Park	Train Station	Hobby Shop	Department Store	Coffee Shop	Hakka Restaurant	Gym	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint
9	Malvern, Rouge	Fast Food Restaurant	Vietnamese Restaurant	Clothing Store	Gym	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Electronics Store	Department Store

The Least Common Venues. The least common venues of these selected neighborhoods are similar. Most venues are related to daily needs of residents, e.g. accessories stores, train station and intersection. For new immigrants, it may be a great idea to open accessories stores in these neighborhoods, if same-industry competitors are major concerns for new businesses. Italian restaurants seem to be a nice alternative option.

	Neighborhood	1st Least Common Venue	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	5th Least Common Venue	6th Least Common Venue	7th Least Common Venue	8th Least Common Venue	9th Least Common Venue	10th Least Common Venue
0	Agincourt	Accessories Store	Indian Restaurant	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Medical Center	Metro Station	Mexican Restaurant	Middle Eastern Restaurant
1	Birch Cliff, Cliffside West	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
2	Cedarbrae	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	Accessories Store	Hobby Shop	Train Station	Indian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station
4	Cliffside, Cliffcrest, Scarborough Village West	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
5	Dorset Park, Wexford Heights, Scarborough Town...	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
6	Golden Mile, Clairlea, Oakridge	Accessories Store	Train Station	Indian Restaurant	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Mexican Restaurant
7	Guildwood, Morningside, West Hill	Accessories Store	Train Station	Indian Restaurant	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Metro Station	Middle Eastern Restaurant
8	Kennedy Park, Ionview, East Birchmount Park	Accessories Store	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant	Middle Eastern Restaurant
9	Malvern, Rouge	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant

Machine Learning: K-Means Clustering Approach and Segmentation of Neighborhoods.



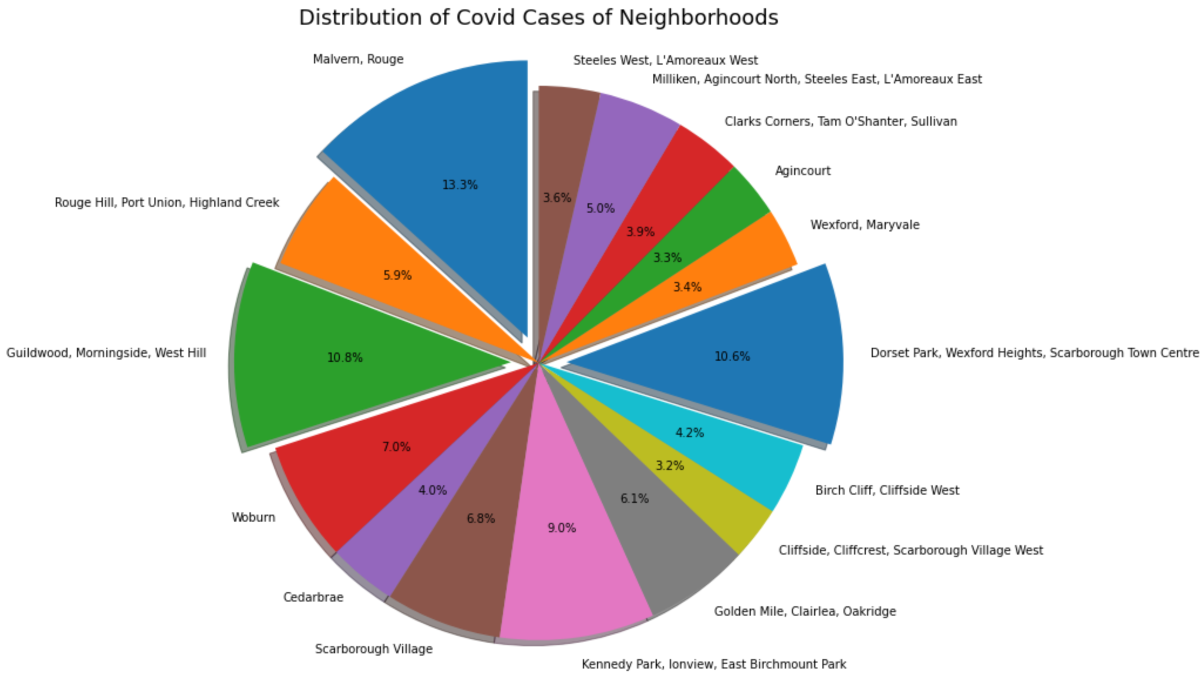
Within Cluster 2 there are more neighborhoods that share similar characteristics. If tourists find one of these neighborhoods interesting, then they may (very likely) find other neighborhoods in the same cluster interesting. Other clusters seem to have only 1 or 2 neighborhoods, that by the k-means clustering algorithm, they are somewhat unique in terms of location-specific characteristics.

ster bels	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	1st Least Common Venue	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	5th Least Common Venue	6th Least Common Venue	7th Least Common Venue	8th Least Common Venue	9th Least Common Venue	10th Least Common Venue
1.0	Rental Car Location	Electronics Store	Medical Center	Intersection	Bank	Restaurant	Mexican Restaurant	Breakfast Spot	College Stadium	Construction & Landscaping	Accessories Store	Train Station	Indian Restaurant	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Metro Station	Middle Eastern Restaurant
1.0	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
1.0	Smoke Shop	Jewelry Store	Playground	Vietnamese Restaurant	Clothing Store	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Accessories Store	Indian Restaurant	Intersection	Italian Restaurant	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
1.0	Bakery	Bus Line	Ice Cream Shop	Park	Bus Station	Metro Station	Intersection	Soccer Field	Fried Chicken Joint	Coffee Shop	Accessories Store	Train Station	Indian Restaurant	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Mexican Restaurant
1.0	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym	Grocery Store	Gas Station	Fried Chicken Joint	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
1.0	Indian Restaurant	Vietnamese Restaurant	Pet Store	Chinese Restaurant	Auto Garage	Bakery	American Restaurant	Grocery Store	General Entertainment	Gas Station	Accessories Store	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station	Mexican Restaurant
1.0	Accessories Store	Auto Garage	Bakery	Sandwich Place	Middle Eastern Restaurant	Smoke Shop	General Entertainment	Grocery Store	Clothing Store	Fried Chicken Joint	Ice Cream Shop	Indian Restaurant	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station
1.0	Skating Rink	Breakfast Spot	Latin American Restaurant	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Gas Station	Accessories Store	Indian Restaurant	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Medical Center	Metro Station	Mexican Restaurant	Middle Eastern Restaurant
1.0	Pizza Place	Chinese Restaurant	Noodle House	Thai Restaurant	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Intersection	Bank	Italian Restaurant	Accessories Store	Hobby Shop	Train Station	Indian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station
1.0	Bakery	Playground	Park	Vietnamese Restaurant	Clothing Store	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Accessories Store	Indian Restaurant	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station
1.0	Fast Food Restaurant	Grocery Store	Chinese Restaurant	Gym	Pharmacy	Pizza Place	Coffee Shop	Breakfast Spot	Bank	Sandwich Place	Accessories Store	Train Station	Intersection	Italian Restaurant	Jewelry Store	Korean BBQ Restaurant	Latin American Restaurant	Lounge	Medical Center	Metro Station

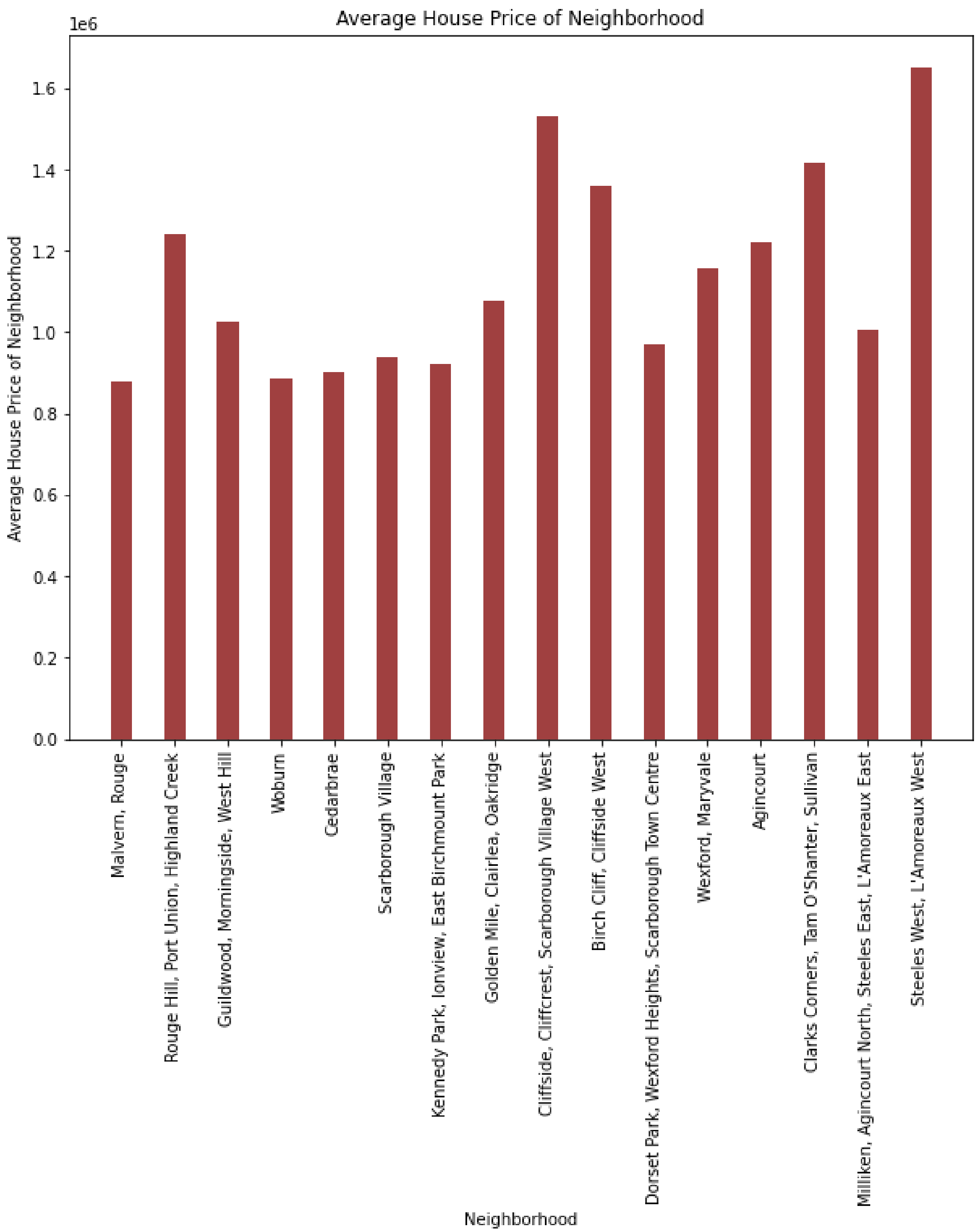
It is also helpful to compile a comprehensive list of neighborhoods with their most and least common venues. To narrow down the list, we focus on the neighborhoods in Cluster 2. Assuming a new immigrant would like to move to one of the neighborhoods in Cluster 2, opening a Mexican or Middle Eastern Restaurant would be a great choice because they are the least common venues in these neighborhoods. To avoid the most intense competition for each of the neighborhoods, a new immigrant can simply look at the 1^{st} most common venue column.

Exploratory Analysis: Graphical and Visualization Analysis.

Distribution of Covid-19 Cases of Neighborhoods. As described in the data section, it makes use of the Covid-19 cases data in Toronto to create a pie plot for all neighborhoods. Clearly, there are large variations in cases across neighborhoods in Toronto. To make it simpler to interpret, the 3 hardest-hit neighborhoods (the 3 separated slices of the pie), are separated out of the original pie. It is easy to observe that these 3 neighborhoods seem to the riskiest choices for tourists and new immigrants to visit, in terms of public health and safety. If public health and safety is the top concern, our stakeholders can choose the pies with smallest case share, namely, Agincourt. Wexford and Maryvale are almost as good as Agincourt.



Bar Plots of Average House Prices of Neighborhoods. House prices are often the concerns of tourists and new immigrants. For tourists, higher house prices usually imply higher accommodation expenditures during their trips; for new immigrants, higher house prices mean higher rental expenditures. With the bar plot below, we learn that Steeles West and L'Amoreaux West are the most expensive neighborhoods, followed by Cliffside, Cliffcrest, and Scarborough Village West. Tourists and new immigrants may be reluctant to move into these places. Malvern, Rouge and Woburn are the most affordable neighborhoods, which are good for budget tourists and price-sensitive immigrants.



Average House Price of Neighborhood
1e6
Average House Price of Neighborhood
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Malvern, Rouge
Rouge Hill, Port Union, Highland Creek
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
Kennedy Park, Ionview, East Birchmount Park
Golden Mile, Clairlea, Oakridge
Cliffside, Cliffcrest, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Wexford Heights, Scarborough Town Centre
Wexford, Maryvale
Agincourt
Clarks Corners, Tam O'Shanter, Sullivan
Milliken, Agincourt North, Steeles East, L'Amoreaux East
Steeles West, L'Amoreaux West
Neighborhood

Section 5: Conclusion

In this capstone project, we use multiple techniques to explore different neighborhoods in Toronto. The techniques include: one hot encoding and finding distribution of venue categories of different neighborhoods, finding the most and least common venues in each of the neighborhoods, k-means clustering approach to segment neighborhoods based on their (dis)similarities of location-based characteristics, and finally, graphical and visualization analysis.

We combine 5 different data sources to obtain a comprehensive dataset to analyze features of different neighborhoods. Combining all data sources allows us to obtain a comprehensive dataset of postal codes, neighborhoods, latitude, longitudes, nearby venues, average house prices, and Covid-19 confirmed cases.

The k-means clustering approach gives a lot of insights for tourists and new immigrants' decisions to choose a suitable neighborhood to explore or open a new business. Take our analysis as an example. If new immigrants want to find a place with less intense competition to start a business, a Mexican or Middle Eastern Restaurant would be a great choice in Cluster 2. Meanwhile, tourists will simply look at the most common venues to see if they fit the purposes of trips.

The house price and Covid-19 cases data give additional perspectives to tourists and immigrants to find the most suitable neighborhoods. Agincourt is seen as the best in public health and safety because it has the lowest Covid-19 case share among all neighborhoods, while Malvern, Rouge and Woburn are the most affordable neighborhoods, which are good for budget tourists and price-sensitive immigrants.