Battle of the Neighborhoods

Opening an indoor playground in Toronto

Figure 1: Indoor Playground. Source: https://www.softplay.com/blog/tips-for-indoor-playground-safety/

Contents

Problem and discussion of the background:

According to IBISWorld, a market research firm, increased disposable income and health awareness have greatly benefited the children's fitness center franchise industry, which includes indoor playgrounds (1). Parents, teachers and caretakers are working hard to combat childhood obesity, resulting in high demand for engaging playgrounds.

Indoor playgrounds are gaining popularity because they offer comfortable and secure play areas, and they give kids a chance to burn energy regardless of weather conditions. Parents and caretakers can relax as kids freely explore fun games and stimulating activities.

(1): https://www.softplay.com/blog/starting-an-indoor-playground/#market-research-and-location

Many start-ups are looking for new potential opportunities to open a new playground business in mature markets such as Toronto. The objective of this report is to use market data to find the optimal placement of a new playground business. A successful recommendation takes into consideration:

- The demographic of the area
- The population income
- The competition location
- And the synergies with other businesses with high traffic areas such as malls which increases the need for having playground areas

Data sources:

To complete this analysis, many sources of information are needed. The link to each source is listed when applicable below. All data was pulled during the execution of this analysis from March 10th to the 20th, 2021.

Source 1: Wikipedia: List of neighbourhoods in Toronto

Link: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto For administrative purposes, the City of Toronto divides the city into 140 neighbourhoods. These divisions are used for internal planning purposes. The boundaries and names often do not conform to the usage of the general population or designated business improvement areas. A number of neighbourhood maps of Toronto do exist, some produced by real estate firms and some by Internet portals. A project to map the neighbourhoods according to the common usage of the residents was done by the Toronto Star newspaper. Based on feedback from Star readers, it has produced the most comprehensive, albeit informal, neighbourhood map.

Each neighbourhood has a unique identifier, “CDN number”

CDN number	City-designated neighbourhood	Former city/borough	Neighbourhoods covered	Map
129	Agincourt North	Scarborough	Agincourt and Brimwood	
128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern	
20	Alderwood	Etobicoke	Alderwood	
95	Annex	Old City of Toronto	The Annex and Seaton Village	
42	Banbury-Don Mills	North York	Don Mills	

Figure 2: Toronto Neighbourhood list from Wikipedia

Source 2: Neighbourhood Demographic and Income

Knowing the customer’s age and income is a good way to start this market analysis. The target customer of indoor playground is children with age range from 2 to 12. One way to find the demographic data for Toronto area is the 2015 Toronto Census open data that is publicly available at https://open.toronto.ca/dataset/neighbourhood-profiles/.

In this file, the word neighbourhood is used to refer to the City of Toronto's 140 social planning neighbourhoods. These neighbourhoods were developed by the City of Toronto to help government and community organizations with their local planning by providing socio-economic data at a meaningful geographic area. Each neighbourhood is also referred to by a neighbourhood number. Each data point in this file is presented for the City's 140 neighbourhoods, as well as for the City of Toronto as a whole. For example, Agincourt North is a neighborhood in Toronto that is assigned a code 129.

_id	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West
1	Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number		129	128
2	Neighbourhood Information	Neighbourhood Information	City of Toronto	TSNS2020 Designation		No Designation	No Designation
3	Population	Population and dwellings	Census Profile 98-316-X20160	Population, 2016	2,731,571	29,113	23,757
4	Population	Population and dwellings	Census Profile 98-316-X20160	Population, 2011	2,615,060	30,279	21,988
5	Population	Population and dwellings	Census Profile 98-316-X20160	Population Change 2011-2016	4.50%	-3.90%	8.00%

Figure 3: Toronto Open Data 2015 Census Data

This census data provides as well for each Toronto neighbourhood:

- "The total number of Children (0-14 years)" identified with **id 10**
- "Total income: Average amount ($)" identified with **id 2273**

Source 3: Locating the competition and the potential synergies with other businesses: Foursquare

Location is crucial to the success of an indoor playground business. Finding the competition and which location do they serve is very important. Also picking a location near a high traffic draw like a mall or shopping center will help bring in more foot traffic (1) The selected source of information for this data is Foursquare. We will focus on finding the distribution of the following venues for each Toronto neighbourhood:

• Playground (CatergoryId = 4bf58dd8d48988d1e7941735)

• Shopping Mall (CatergoryId = 4bf58dd8d48988d1fd941735)

• Shopping Plaza (CatergoryId = 5744ccdfe4b0c0459246b4dc)

Source 4: Geophysical Location data: Geopy library

Geopy is a Python client for several popular geocoding web services. Geopy makes it easy to locate the coordinates of addresses, cities, countries, and landmarks across the globe, including Toronto Ontario, using third-party geocoders and other data sources. The geospatial information will be used to locate all identified venues and neighborhoods using the longitude and latitude coordinates.

Source 5: Toronto neighborhoods Geojson

Visualization is very powerful in conveying the results of this analysis. It is important to have the Toronto-geojson to superimpose the results on the top of Toronto map. The Geojson source is the Toronto Open Data Set https://open.toronto.ca/dataset/neighbourhoods/

Conveniently, the file has the feature “AREA_SHORT_CODE” that is the same as the neighborhood ID found in the census data.


```
{
"type": "FeatureCollection",
"crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
"features": [
{ "type": "Feature", "properties": { "_id": 11481, "AREA_ID": 2480141, "AREA_ATTR_ID": 26005521, "PARENT_AREA_ID": null, "AREA_SHORT_CODE": 96,
.417811656038793, 43.681925277740298 ], [ -79.418121371159799, 43.682710522285703 ], [ -79.4182734938897, 43.683096166233497 ], [ -79.41830144
.412631283351004, 43.6884334123732 ], [ -79.412643206881896, 43.688516872543701 ], [ -79.412669085002705, 43.68866901429419 ], [ -79.41217980
```


Figure 4: Toronto Geopy JSON file screenshot

Methodology

Steps Overview:

Below is an overview of the steps followed in this market analysis to get to the short list of recommended neighborhoods for opening a new playground business:

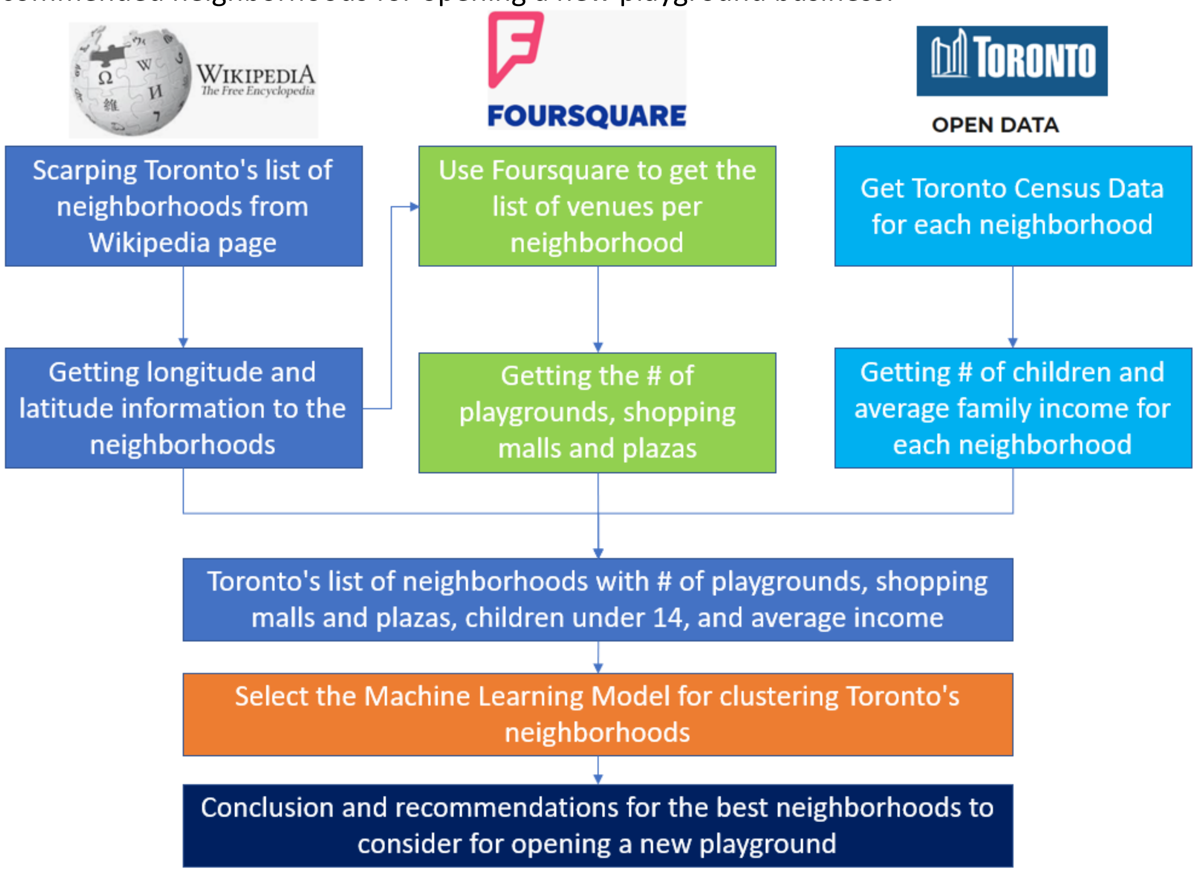

Figure 5: Methodology steps overview

In the next sections, we will go over the pre-processing steps that were executed for each step, and will describe the detailed methodology

Data Pre-Processing:

Toronto's list of neighborhoods with longitude and latitude coordinates

The list of Toronto's neighborhoods is available under Wikipedia page: https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto. To scrape off the data from this page, I used BeautifulSoup library. The 10th table on this page contains the list of neighborhoods with the respective CDN number. Created a dataframe with this data:

Out[4]:

	CDN number	City-designated neighbourhood	Former city/borough	Neighbourhoods covered	Map	Unnamed: 5
0	129	Agincourt North	Scarborough	Agincourt and Brimwood	NaN	NaN
1	128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern	NaN	NaN
2	20	Alderwood	Etobicoke	Alderwood	NaN	NaN
3	95	Annex	Old City of Toronto	The Annex and Seaton Village	NaN	NaN
4	42	Banbury-Don Mills	North York	Don Mills	NaN	NaN
...	...	...	...	...	...	...
135	94	Wychwood	Old City of Toronto	NaN	NaN	NaN
136	100	Yonge and Eglinton	Old City of Toronto	Chaplin Estates	NaN	NaN
137	97	Yonge-St.Clair	Old City of Toronto	NaN	NaN	NaN
138	27	York University Heights	North York	NaN	NaN	NaN
139	31	Yorkdale-Glen Park	North York	Glen Park, Lawrence Heights	NaN	NaN

140 rows × 6 columns

Figure 6: Pandas Dataframe with the scraped list of Toronto's neighborhoods

I need to get the coordinates for each neighborhood in Toronto. This information is required when later we will explore the venues in each neighborhood using Foursquare. To get these coordinates, I used the Geopy library, and search for each neighborhood the associated longitude and latitude. Added this information to the original dataframe. I noticed that some of the neighborhoods did not return any coordinates. There were 30 neighborhoods.

Out[10]:

	CDN number	City-designated neighbourhood	Former city/borough	Neighbourhoods covered	Map	Unnamed: 5	Longitude	Latitude
0	129	Agincourt North	Scarborough	Agincourt and Brimwood	NaN	NaN	-79.266439	43.808038
1	128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern	NaN	NaN	-79.257689	43.781969
2	20	Alderwood	Etobicoke	Alderwood	NaN	NaN	-79.545232	43.601717
3	95	Annex	Old City of Toronto	The Annex and Seaton Village	NaN	NaN	-79.407117	43.670338
4	42	Banbury-Don Mills	North York	Don Mills	NaN	NaN	-79.365270	43.752683
5	34	Bathurst Manor	North York	Bathurst Manor	NaN	NaN	-79.456367	43.763893
6	76	Bay Street Corridor	Old City of Toronto	Bay Street, Financial District	NaN	NaN	-79.390499	43.672135
7	52	Bayview Village	North York	Bayview Village	NaN	NaN	-79.376662	43.769197
8	49	Bayview Woods-Steeles	North York	Bayview Woods	NaN	NaN	-79.382973	43.798127
9	39	Bedford Park-Nortown	North York	Bedford Park, Ledbury Park, and Nortown	NaN	NaN	NaN	NaN
10	112	Beechborough-Greenbrook	York	Keelesdale and Silverthorn	NaN	NaN	NaN	NaN
11	127	Bendale	Scarborough	Bendale	NaN	NaN	-79.255336	43.753520
12	122	Birchcliffe-Cliffside	Scarborough	Birch Cliff and Cliffside	NaN	NaN	NaN	NaN
13	24	Black Creek	North York	Jane and Finch	NaN	NaN	-79.485495	43.695400

Figure 7: List of Toronto's neighborhoods with most Longitude and Latitude information

Further exploration and testing show that, by entering a subset of the neighborhood's name, the coordinates are returned. For example, geolocation did not return a result for the neighborhood "Kensington-Chinatown", by entering just "Kensington" the geolocation is returned.

```
In [12]: # Use geopy library to get the latitude and longitude values for each NaN addresses.
address = 'Kensington, Toronto City, ON'

geolocator = Nominatim(user_agent="on_explorer")
location = geolocator.geocode(address)
latitude1 = location.latitude
longitude1 = location.longitude
print('The geograpical coordinate of Kensington-Chinatown are {}, {}.'.format(latitude1, longitude1))

The geograpical coordinate of Kensington-Chinatown are 43.6552136, -79.4022604.
```

Figure 8: Searching manually for each neighborhood's coordinates

This manual process was successful for all locations. I tracked the results for each of the 30 neighborhood and added manually this information to the CSV file with the original Geopy results. Once done, I imported the file as the new dataframe. Subsequently, I dropped all columns that are not needed:

Out[27]:

	CDN number	City-designated neighbourhood	Longitude	Latitude
0	129	Agincourt North	-79.266439	43.808038
1	128	Agincourt South-Malvern West	-79.278549	43.785353
2	20	Alderwood	-79.545232	43.601717
3	95	Annex	-79.407117	43.670338
4	42	Banbury-Don Mills	-79.365270	43.752683

Figure 9: List of Toronto's neighborhoods with longitude and latitude coordinates

List of neighborhoods with number of playgrounds and malls

To get the number of existing playgrounds, shopping malls, and shopping plazas, I used Foursquare API. I used a radius of 2 kilometre and limited the results to 1000. A radius of 2 kilometres seems reasonable given the usual size of a neighborhood, and as well taking into account the accessibility of the venue for residents in an adjacent neighborhood potentially.

I tried to apply a search for specific venues for the categories that I was interested in as below, but the results were coming with all types of venues. I kept the search criteria open to get all venues:

- Playground (CatergoryId = 4bf58dd8d48988d1e7941735)
- Shopping Mall (CatergoryId = 4bf58dd8d48988d1fd941735)
- Shopping Plaza (CatergoryId = 5744ccdfe4b0c0459246b4dc)

Out[32]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt North	43.808038	-79.266439	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
1	Agincourt North	43.808038	-79.266439	Menchie's	43.808338	-79.268288	Frozen Yogurt Shop
2	Agincourt North	43.808038	-79.266439	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
3	Agincourt North	43.808038	-79.266439	Samosa King - Embassy Restaurant	43.810152	-79.257316	Indian Restaurant
4	Agincourt North	43.808038	-79.266439	Shoppers Drug Mart	43.808894	-79.269854	Pharmacy

Figure 10: List of venues in Toronto

I used one hot spot encoding to transform the "Venue Category" to individual columns

Out[34]:

	Neighborhood	Zoo Exhibit	Accessories Store	Afghan Restaurant	African Restaurant	Airport	American Restaurant	Amphitheater	Antique Shop	Argentinian Restaurant	...	Video Game Store	Vietnamese Restaurant
0	Agincourt North	0	0	1	0	0	0	0	0	0	...	0	5
1	Agincourt South-Malvern West	0	0	0	0	0	0	0	0	0	...	0	2
2	Alderwood	0	0	0	0	0	1	0	0	0	...	1	1
3	Annex	0	0	0	0	0	0	0	0	0	...	0	1
4	Banbury-Don Mills	0	0	0	0	0	0	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
135	Wychwood	0	0	0	0	0	1	0	0	0	...	0	0
136	Yonge and Eglinton	0	0	0	0	0	0	0	0	0	...	0	1
137	Yonge-St.Clair	0	0	0	0	0	1	0	0	0	...	0	0
138	York University Heights	0	0	0	0	0	0	0	0	0	...	0	1
139	Yorkdale-Glen Park	0	1	0	0	0	0	0	1	1	...	0	1

140 rows × 330 columns

Figure 11: List of neighborhoods with the total number of venues per category

I filtered the 3 venues of interest and summed both shopping malls and shopping plazas as one column. This will help in the modeling step. Prior attempts of clustering the neighborhoods with shopping malls and shopping plazas as separate factors did not lead to a convincing grouping.

Out[38]:

	Neighborhood	Playground	Shopping Malls and Plazas
0	Agincourt North	0	2
1	Agincourt South-Malvern West	0	2
2	Alderwood	0	1
3	Annex	1	1
4	Banbury-Don Mills	0	1
...	...	...	...
135	Wychwood	1	0
136	Yonge and Eglinton	0	0
137	Yonge-St.Clair	0	0
138	York University Heights	0	0
139	Yorkdale-Glen Park	1	0

140 rows × 3 columns

Figure 12: List of Toronto's neighborhoods with the number of playgrounds, shopping malls and plazas

Get Toronto Census Data for each neighborhood

In this section I will explore the data from 2015 Toronto census. The data is accessible from open data publicly available from https://open.toronto.ca/dataset/neighbourhood-profiles/. I downloaded the CSV data and imported it as Pandas Dataframe.

Out[39]:

	_id	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood
0	1	Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129.00	128.00	20.00
1	2	Neighbourhood Information	Neighbourhood Information	City of Toronto	TSNS2020 Designation	NaN	No Designation	No Designation	No Designation
2	3	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2016	2731571.00	29113.00	23757.00	12054.00
3	4	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2011	2615060.00	30279.00	21988.00	11904.00
4	5	Population	Population and dwellings	Census Profile 98-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%

5 rows × 146 columns

Figure 13: Toronto's list of neighborhoods with census data

The census data has 2383 indicators for the 140 neighborhoods.

Getting # of children and average family income for each neighborhood

I filtered the 3 indicators of interest, and I transposed the dataframe to have the neighborhoods as rows, and got rid of the any additional columns, and renamed the header:

- ID 1: Neighborhood number
- ID 10: Children (0-14 years)
- ID 2273: Total income: Average amount ($)

Out[44]:

	City-designated neighbourhood1	CDN number	Children_0-14	Total_Income
1	Agincourt North	129.00	3840.00	30414.00
2	Agincourt South-Malvern West	128.00	3075.00	31825.00
3	Alderwood	20.00	1760.00	47709.00
4	Annex	95.00	2360.00	112766.00
5	Banbury-Don Mills	42.00	3605.00	67757.00

Let's examin the data types from this census

In [45]:
```
toronto_Children_Income.dtypes
```

Out[45]:
```
City-designated neighbourhood1    object
CDN number                        object
Children_0-14                     object
Total_Income                      object
dtype: object
```

Figure 14: Census data for Toronto

I had difficult times in the early exploration of the data and the analysis, I received many errors when trying to map the neighborhoods for example with census data. Further investigations have shown that there is a data type issue in the data from the census. By examining the field types, I realized that the numbers were imported as objects, which led me to go back and change the field type in the CSV file, and to remove the separator "," for the thousand. I had to cast a numeric type to these numbers, which took care of the issues I had earlier.

In earlier analysis, I realized that it will be easier of I get a sense of the clusters in category style of number of children and income, so I applied binning approach as following:

```
group_Children = ["Low","Medium","High"]
group_Income = ["Low Income","Lower-Middle Income","Upper-Middle Income","High Income"]
```

Out[48]:

	City-designated neighbourhood1	CDN number	Children_0-14	Total_Income	Children-binned	Income-binned
1	Agincourt North	129.0	3840.0	30414.0	Medium	Low Income
2	Agincourt South-Malvern West	128.0	3075.0	31825.0	Low	Low Income
3	Alderwood	20.0	1760.0	47709.0	Low	Low Income
4	Annex	95.0	2360.0	112766.0	Low	Lower-Middle Income
5	Banbury-Don Mills	42.0	3605.0	67757.0	Medium	Low Income

Figure 15: Toronto's neighborhoods with number of children under 14 and Total Income

Merging all information together

Using the joining technique, I merged all 3 data frames using the “neighborhood number” as the key. See results below:

Out[51]:

	CDN number	City-designated neighbourhood	Longitude	Latitude	Playground	Shopping Malls and Plazas	Children_0-14	Total_Income	Children-binned	Income-binned
0	129	Agincourt North	-79.266439	43.808038	0	2	3840.0	30414.0	Medium	Low Income
1	128	Agincourt South-Malvern West	-79.278549	43.785353	0	2	3075.0	31825.0	Low	Low Income
2	20	Alderwood	-79.545232	43.601717	0	1	1760.0	47709.0	Low	Low Income
3	95	Annex	-79.407117	43.670338	1	1	2360.0	112766.0	Low	Lower-Middle Income
4	42	Banbury-Don Mills	-79.365270	43.752683	0	1	3605.0	67757.0	Medium	Low Income

Figure 16: List of Toronto's neighborhoods with number of playgrounds, shopping malls and plazas, number of children under 14, and Income

Now I can start the modeling

Machine Learning Model:

Model selection

Toronto has 140 neighborhoods. We need a method to cluster these neighborhoods in a homogenous way. We do not have any labels for each neighborhood, which means that we will need an unsupervised learning approach. In the other hand, we have continuous, not categorical data. This suggests using the K-means clustering as the preferred method

Selecting the optimal K

To get our optimum K value, I used the **Elbow Point** Technique. In this technique, many simulations are performed iteratively using different K values, and calculating the sum of the errors for each K value.

Obviously, the higher K, the better is the accuracy. The optimum K is where there is a sharp decrease in improving the error despite increasing K. The elbow in this case was identified at K=5, which means 5 clusters.

The data used for this step is normalized. Normalization is a statistical method that helps mathematical-based algorithms interpret features with different magnitudes and distributions equally. We use StandardScaler() to normalize our dataset

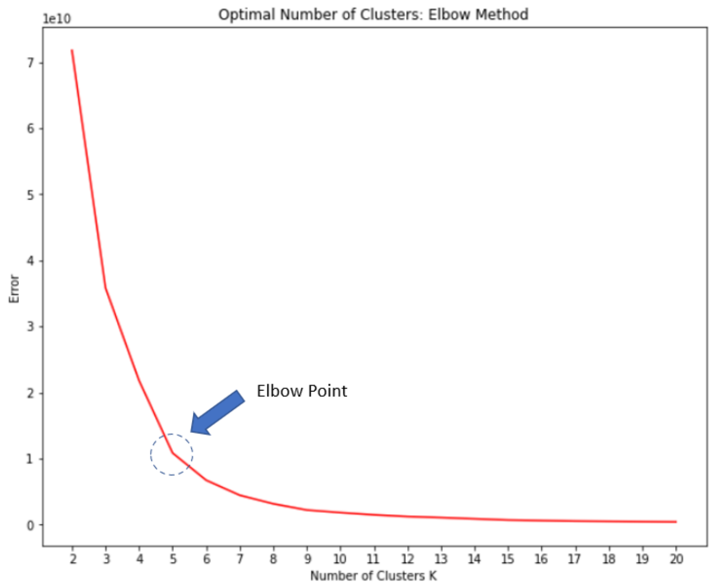

Figure 17: Optimal Number of Clusters: Elbow method

Results

Toronto neighborhood clusters

Toronto's 140 neighborhoods are clustered into 5 clusters based on the following criteria:

- Number of children under the age of 14
- Average income
- Number of playgrounds
- And the number of shopping malls and plazas, which are good synergy businesses with high traffic

By exploring the different clusters, I assigned a short descriptive to each cluster as following:

- **Cluster 0:** Medium to high # children, low competition, high traffic, and low income
- **Cluster 1:** Low to medium # children, low competition, low traffic, and low income
- **Cluster 2:** Low to medium # children, high competition, low to medium traffic and low to middle income
- **Cluster 3:** Low to medium # children, no competition, medium traffic and low to medium income

- **Cluster 4:** Low # children, low competition, low traffic and high income

Cluster #	# Neighborhoods	Average # Children_0-14	Average # Total Income	Average # Playground	Average # Shopping Malls and Plazas
0	17	5595	35804	0.12	1.41
1	45	2479	47185	0.00	0.16
2	29	2405	61358	1.10	0.34
3	44	2552	49290	0.00	0.59
4	5	1894	210937	0.40	0.20

Discussion and recommendation

From the quadrant below, we can see that most of the competition is in cluster 2. Cluster 3 is very similar to cluster 2 in demographic, number of children and in traffic, except that it does not have competition. This cluster is the first recommendation to focus on.

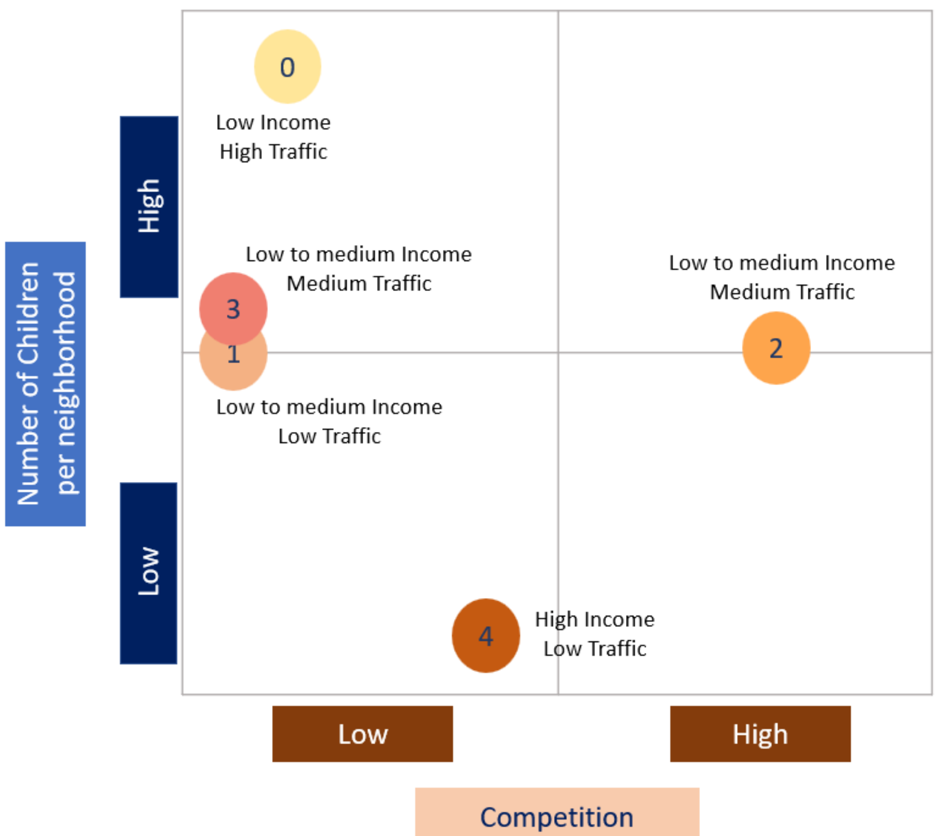

Figure 18: Toronto Neighborhood Cluster Quadrant

The list of the recommended cluster 3 Toronto neighborhoods is below:

- Alderwood
- Banbury-Don Mills
- Bay Street Corridor

- Bayview Village
- Bayview Woods-Steeles
- Bedford Park-Nortown
- Black Creek
- Broadview North
- Brookhaven-Amesbury
- Church-Yonge Corridor
- Clanton Park
- Crescent Town
- East End-Danforth
- Elms-Old Rexdale
- Eringate-Centennial-West Deane
- Etobicoke West Mall
- Henry Farm
- Humber Heights-Westmount
- Humber Summit
- Humbermede
- Kingsview Village-The Westway
- Leaside-Bennington
- Maple Leaf
- Markland Wood
- Mimico
- Moss Park
- New Toronto
- North St. James Town
- O'Connor-Parkview
- Pelmo Park-Humberlea
- Pleasant View
- Rexdale-Kipling
- Rustic
- St. Andrew-Windfields
- Stonegate-Queensway
- The Beaches
- Thistletown-Beaumond Heights
- University
- Victoria Village
- West Humber-Clairville
- Westminster-Branson
- Willowdale West
- Willowridge-Martingrove-Richview
- Woodbine Corridor

I used Folium and Choropleth map type to map the recommended neighborhoods in cluster 3 to open a playground business

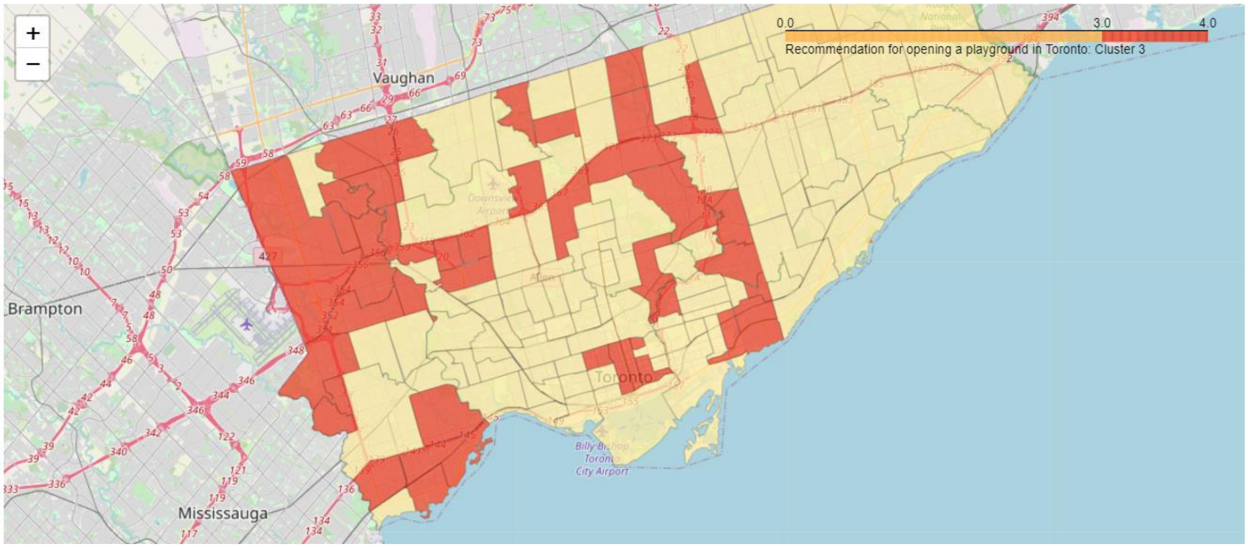

Figure 19: Neighborhoods in dark orange are recommended for opening a playground business

Conclusion

The goal of this project was to recommend the best neighborhoods to open a new playground in Toronto. To build this recommendation, it was important to understand the success factors for this kind of business: The demographics, the income, the competition and the synergies with other businesses such as shopping malls and plazas. Clustering the neighborhoods into 5 optimum clusters using the K-means led to the formation of similar neighborhoods for these factors.

I used the Beautifulsoup web scrapping library to get the neighborhoods from Wikipedia, and I used Foursquare to get the venues information for playgrounds, shopping malls and shopping plazas. And I used Toronto Open Data to get the 2015 census results for the number of children under 14, and the average income per neighborhood. The census data might have changed over the past 6 years, the assumption is that the data is still valid.

To visualize the recommended neighborhoods, I used Folium libraries along with the geojson file for Toronto neighborhoods sourced from Toronto Open Data.

As a future exploration, more factors can be added for a better decision making. For example, the real-estate cost for a location, as well as the availability of adequate locations with a high ceiling as a pre-requisite.

Also, I would be interested to investigate advanced techniques in the clustering such as Feature engineering which is the process of using domain knowledge to choose which data metrics to input as features into a machine learning algorithm. Feature engineering plays a key role in *K*-means clustering; using meaningful features that capture the variability of the data is essential for the algorithm to find all of the naturally-occurring groups.