INTRODUCTION

Introduction

Toronto is a financial capital of Canada, its downtown core has large concentration of high-rise office buildings for Canadian and International companies. For several decades Toronto downtown experienced rapid growth in residential condominiums, driven by professionals working in the financial service industry and proximity of their offices.

2020 global pandemic had significant impact on the operations of financial companies. Offices were shut down, and accent was made on the remote work from home. Toronto downtown core became a ghost town, with people looking for more suitable Toronto neighborhoods to relocate.

Business Problem

This research will focus on the people currently living in crowded Toronto downtown core, and considering relocating to other Toronto neighborhoods, which will better satisfy their interests and social needs. We will explore Toronto neighborhoods and build a data model, which will help people to find an optimal place to live depending on their specific preferences. Special consideration will be made for the 2020 COVID pandemic, in selecting new neighborhoods.

DATA

In order to achieve our goal, we need detailed information about Toronto's segmentation by neighborhoods and its amenities.

This research uses two main data sources. The first one is Toronto neighborhoods listing from Wikipedia. This data set is important to establish main framework, define neighborhoods locations and names.

The second data set is a listing of Toronto amenities and facilities from Foursquare API. This set of data provides important insights into each neighborhood and will allow compare neighborhoods to each other. We will focus on amenities which allow social distancing, considering specific of current COVID pandemic and focus of this research.

Data cleaning and wrangling

First, we will clean Toronto neighborhoods data set. The set contains postal codes "Not assigned" to specific boroughs, these lines have to be dropped, as they don't contain any relevant data for our analysis. Next, we check that there is no "not assigned" neighborhoods, duplicate lines or missing values. Once it is done, we add geolocation longitude and latitude to each neighborhood.

Table 1. Example of neighborhood data.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

As a result, we have 103 unique neighborhoods and 10 boroughs. In order to make data easier to understand and visualize, we will cluster neighborhoods in several buckets and create a folium map of Toronto neighborhoods. We can see a heavy concentration of neighborhoods around downtown core, the place we want to relocate from.

Table 2. Toronto neighborhood clusters.
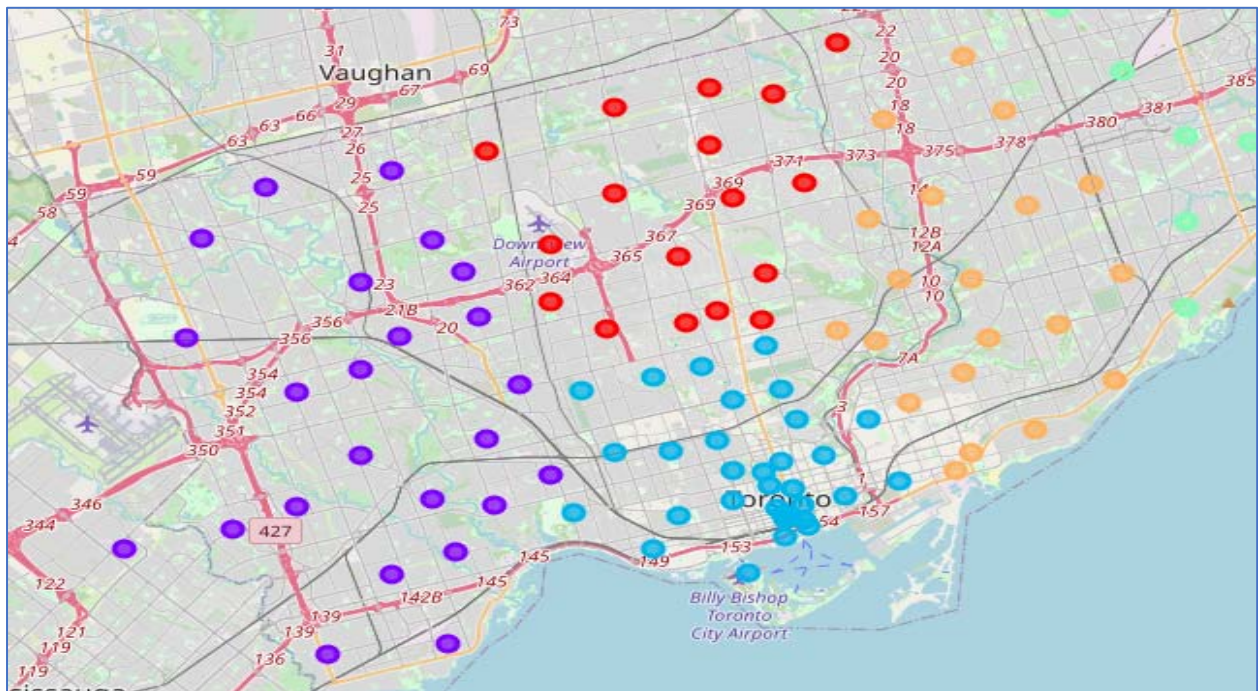
The next data set contains Toronto facilities and venues from Foursquare API. Considering that it is COVID pandemic, a lot of amenities are on lockdown. For this reason, we will disregard restaurants, and entertainment facilities, which are not essential services and have issues with social distancing rules. We will focus mainly on proximity of grocery chains, outdoor facilities, parks, public transit, and medical centers.

Original venue data from Foursquare was extracted do data frame and mapped based on the geolocation to existing neighborhoods from the previous step.

Table 3. Neighborhoods and venues mapped with geolocation.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	TTC stop #8380	43.752672	-79.326351	Bus Stop
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	Parkwoods	43.753259	-79.329656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
4	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena

Next, we drop the venues related to restaurants and entertainment facilities, as it is not relevant for our research, and focus on venues with outdoor activities, parks, dog playgrounds, grocery stores and medical offices. As a result, we have a solid data frame for further data analysis and modeling.

It is curious to check if there are any data correlations between venues. As expected, we don't find any significant correlations between venues. The largest correlation of 0.55 is only between Curling Ice and Athletic venues.

Table 4. Venues correlation.

	Athletics & Sports	Bakery	Baseball Field	Beach	Bus Line	Bus Station	Bus Stop	Curling Ice	Department Store
Athletics & Sports	1.000000	0.036051	-0.008427	-0.025282	-0.051177	-0.036735	0.254850	0.555525	-0.051752
Bakery	0.036051	1.000000	-0.061966	0.030629	0.097458	0.164543	-0.057182	-0.041716	-0.069869
Baseball Field	-0.008427	-0.061966	1.000000	-0.016380	-0.033159	-0.021867	-0.022453	-0.016380	-0.031688
Beach	-0.025282	0.030629	-0.016380	1.000000	-0.021767	-0.015624	-0.014739	-0.010753	-0.022011
Bus Line	-0.051177	0.097458	-0.033159	-0.021767	1.000000	0.190643	-0.029836	-0.021767	-0.044557
Bus Station	-0.036735	0.164543	-0.021867	-0.015624	0.190643	1.000000	-0.021417	-0.015624	0.686015
Bus Stop	0.254850	-0.057182	-0.022453	-0.014739	-0.029836	-0.021417	1.000000	0.489074	-0.030172
Curling Ice	0.555525	-0.041716	-0.016380	-0.010753	-0.021767	-0.015624	0.489074	1.000000	-0.022011
Department Store	-0.051752	-0.069869	-0.031688	-0.022011	-0.044557	0.686015	-0.030172	-0.022011	1.000000
Discount Store	0.009177	-0.063749	-0.022162	-0.030229	-0.061193	0.418774	-0.041437	-0.030229	0.384560
Dog Run	-0.033884	-0.055911	-0.021954	-0.014411	-0.029173	-0.020940	-0.019754	-0.014411	-0.029501

We will also check descriptive statistics of our data.

Table 5. Descriptive statistic.

	Athletics & Sports	Bakery	Baseball Field	Beach	Bus Line	Bus Station	Bus Stop	Curling Ice	D
count	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	94.000000	
mean	0.006218	0.019547	0.009733	0.000180	0.010106	0.002996	0.004179	0.001520	
std	0.025642	0.048848	0.061946	0.001748	0.048405	0.019988	0.029561	0.014735	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
75%	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
max	0.142857	0.250000	0.500000	0.016949	0.250000	0.166667	0.250000	0.142857	

From this point we can start analyzing and tailoring data for specific stakeholders needs.

For example, we can create a list of all neighborhoods, and calculate mean frequency. Then we can list all neighborhoods with top 5 venues for each of them. This can be helpful for a person who already has a neighborhood in mind, or just browsing through the list. Data can easily point top venues for each neighborhood.

Table 5. Most abundant activity for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Agincourt	Skating Rink	Train Station	Discount Store	Grocery Store	Golf Course
1	Alderwood, Long Branch	Skating Rink	Train Station	Discount Store	Grocery Store	Golf Course
2	Bathurst Manor, Wilson Heights, Downsview North	Supermarket	Pharmacy	Train Station	Discount Store	Grocery Store
3	Bayview Village	Train Station	Dog Run	Harbor / Marina	Grocery Store	Golf Course
4	Bedford Park, Lawrence Manor East	Pharmacy	Grocery Store	Discount Store	Golf Course	Garden

Table 6. Most frequent activity for each neighborhood.

```
    ----Bathurst Manor, Wilson Heights, Downsview North----
                    venue  freq
0             Pharmacy    0.04
1          Supermarket    0.04
2        Grocery Store    0.04
3                 Park    0.04
4    Athletics & Sports   0.00


    ----Bayview Village----
                    venue  freq
0    Athletics & Sports   0.0
1             Pharmacy    0.0
2                 Lake    0.0
3    Light Rail Station   0.0
4               Market    0.0
```

Based on the current data, we can check what are the most frequent facilities for each of the Toronto neighborhoods. Further we can rank those venues or focus on the specific ones.

If a person looking for a neighborhood with park, we can rank top 10 neighborhoods with the highest park frequency. We can recommend Willowdale and Newtonbrook.

Table 7. Neighborhoods with highest park frequency.

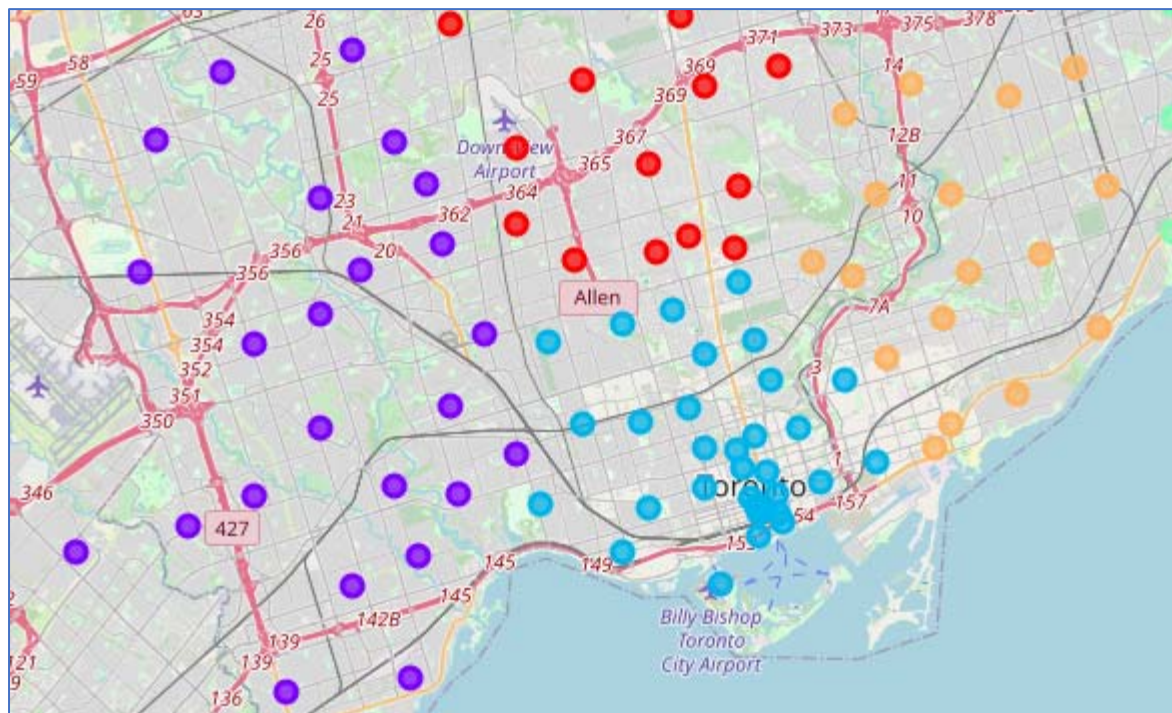|	Neighborhood	Park
88	Willowdale, Newtonbrook	1.000000
10	Caledonia-Fairbanks	0.500000
65	Rosedale	0.500000
93	York Mills West	0.500000
50	Milliken, Agincourt North, Steeles East, L'Amo...	0.333333
81	The Kingsway, Montgomery Road, Old Mill North	0.333333
26	East Toronto, Broadview North (Old East York)	0.250000
30	Forest Hill North & West, Forest Hill Road Park	0.250000
44	Kingsview Village, St. Phillips, Martin Grove ...	0.250000
46	Lawrence Park	0.250000

Someone who has a dog might be interested in neighborhood with pet playground. Below is the list of top 5 such neighborhoods.

Table 8. Neighborhoods with highest dog playground frequency.

	Neighborhood	Dog Run
37	Hillcrest Village	0.200000
59	Parkdale, Roncesvalles	0.071429
15	Church and Wellesley	0.012500
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000

RESULTS

As a result of this project, we build a model of Toronto neighborhoods with COVID friendly venues. The model is based on the Toronto geolocation data, and Foursquare venues. We clean up data and grouped it in several cluster. With majority of neighborhoods concentrating around busy downtown core.

The model allows find suitable neighborhood based on specific preferences, such as park hospitals, dog playgrounds, grocery stores and proximity to public transit.

For example, the top neighborhoods with outdoor parks are Willowdale and Newtonbrook.

	Neighborhood	Park
88	Willowdale, Newtonbrook	1.000000
10	Caledonia-Fairbanks	0.500000
65	Rosedale	0.500000
93	York Mills West	0.500000

DISCUSSION

I observed several things during data analysis stage. First there is no positive / negative correlations between current data set, which is a predictable outcome, as venues are build or naturally situated around the city. But it was interesting to see if correlations are existing.

Another thing is concentration of 103 neighborhoods. We can clearly see that neighborhoods are heavily concentrated around Toronto's downtown core, and less spread toward to city borders.

CONCLUSION

Residents of Toronto Downton neighborhoods started looking for better places to live, as global COVID pandemic impacted busy financial center.

In order to help with new neighborhood search, we developed a model based on Toronto geolocation and Foursquare venues. The model contains venues which are highly attractable during pandemic, such as parks, dog playgrounds, outdoor venues, and medical centers.

Models allows to factor specific individual needs of each person seeking to relocate to a new neighborhood and help them find optimal living solution.

For example, a dog owner currently living in Toronto downtown would be interested in Hillcrest Village neighborhood with the highest mean playground rating among 103 Toronto neighborhoods.

	Neighborhood	Dog Run
37	Hillcrest Village	0.200000
59	Parkdale, Roncesvalles	0.071429
15	Church and Wellesley	0.012500
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000