# Capstone project – the battle of neighborhoods

### **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.

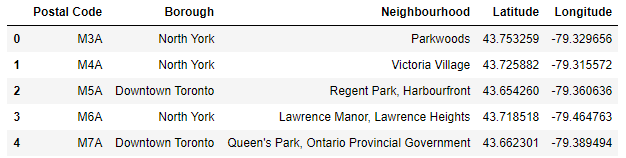
This research uses two main data sources. The first one is Toronto neighborhoods listing from Wikipedia. This data set is important to establish main framework, define neighborhoods locations and names.

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.

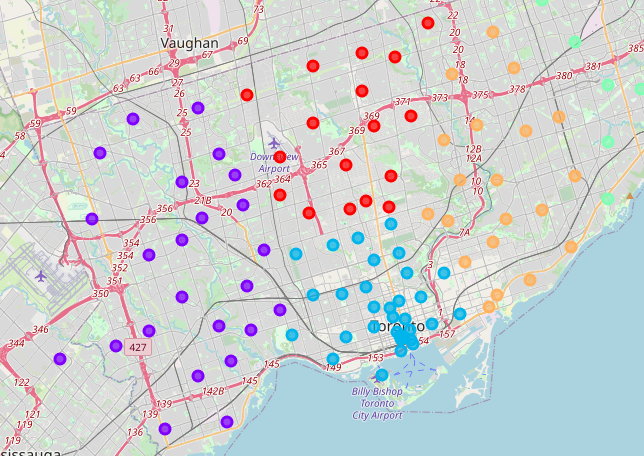
### **methodology**

**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.

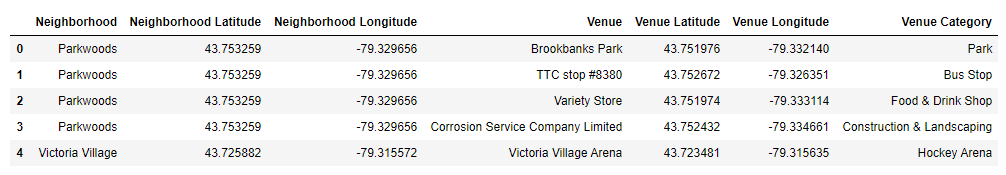
Table 1. Example of neighborhood data.

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.

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.

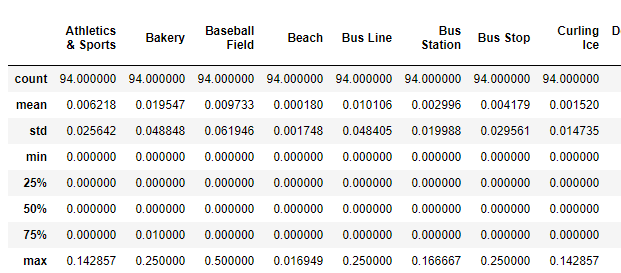
It is curious to check if there are any data correlations between venues. As expected, we don’t find any significant correlations between venues. The largest correlation of 0.55 is only between Curling Ice and Athletic venues.

Table 4. Venues correlation.



We will also check descriptive statistics of our data.

Table 5. Descriptive statistic.



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.

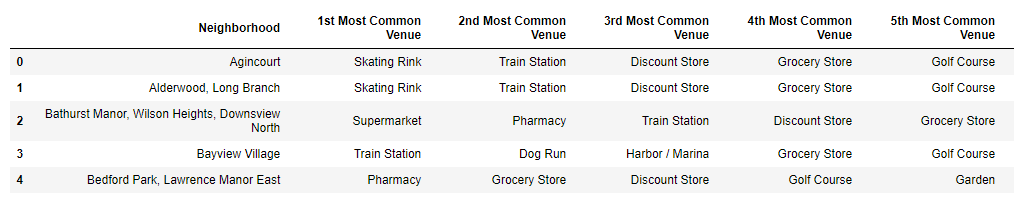
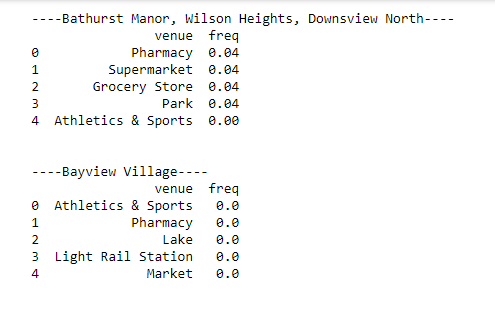
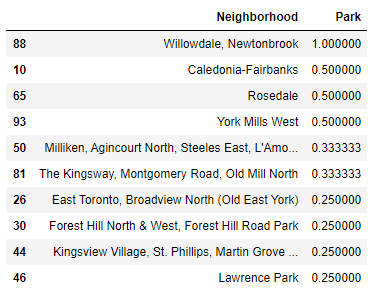
Table 5. Most abundant activity for each neighborhood.

Table 6. Most frequent activity for each neighborhood.

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.

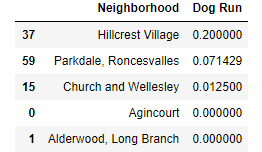
If a person looking for a neighborhood with park, we can rank top 10 neighborhoods with the highest park frequency. We can recommend Willowdale and Newtonbrook.

Table 7. Neighborhoods with highest park frequency.



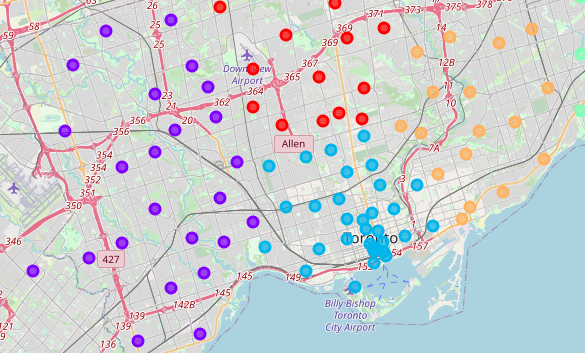
Someone who has a dog might be interested in neighborhood with pet playground. Below is the list of top 5 such neighborhoods.

Table 8. Neighborhoods with highest dog playground frequency.



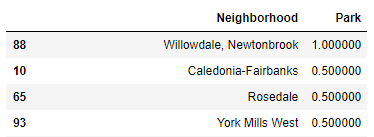
### **results**

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.

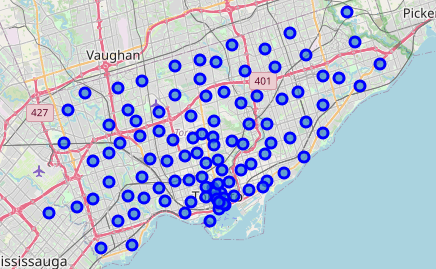
For example, the top neighborhoods with outdoor parks are Willowdale and Newtonbrook.



### **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.

