Finding the best place for Yoga Studio in Toronto

Capstone Project Report

Design By
MEHRNOOSH NOBAKHT

Introduction

Business Problem

Yoga studios are essential for everyone. In this digital, social, media-energized virtual world, it is evident to me that people still crave human connection. Yoga studio is a place for gathering. A place where hearts can connect. A place where interaction and relationships develop in personal face-to-face relationships. Yoga helps people of all ages physically, mentally and emotionally and it can give them special vitality to perform daily activities. Given the competition of clubs to attract customers, the project is looking for a suitable location in Toronto to establish a yoga studio.

Objective

The objective of this project is to find the best neighborhood in Toronto to open a Yoga Studio using Foursquare location data.

Problem Description

We will look for **locations that are not already crowded with Yoga studios**. We are also particularly interested in **areas with no nearby Yoga studio**. If the first two conditions are met, we would also prefer locations **as close to the city center as possible**.

Based on this criterion, we will use our data science skills to generate a few promising neighborhoods. The benefits of each area will then be clearly expressed so that stakeholders can choose the best possible final location.

Target Audience

- Business people who want to invest or open a startup Yoga studio.
- The individual who wishes to run their own yoga studio as a side business.
- Marketing firms that want to choose the best location to offer their services and products.

Research Question

I will attempt to answer these questions:

- “Where should I open a yoga studio in Toronto?”
- "Which factors are important for choosing the best location to start up a Yoga Studio?"

Data Description

Based on definition of our problem, factors that will influence our decision are:

- number of existing yoga studio in the neighborhood
- number of and distance to yoga studios in the neighborhood, if any
- List of Boroughs and neighborhoods of Toronto with their geodata (latitude and longitude)
- distance of neighborhood from city center

To do this project we need the following data:

1. Toronto City data that contains Borough, Neighborhoods along with there latitudes and longitudes

Data Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Geographical Location data using Geocoder Package

Data Source: https://cocl.us/Geospatial_data

3. Yoga studio Data using Foursquare API

Data Source: https://foursquare.com/developers/apps

All data related to locations of Yoga Studio will be obtained via the FourSquare API utilized via the Request library in Python.

Foursquare is a US tech company from New York focusing on location data.

We can get the name, latitude and longitude for each venue.

Data Preprocessing

We use data preprocessing to clean the dataset.

We do the following preprocessing operations:

- Drop rows having null value and value assigned as "Not assigned"
- Process the cells that have an assigned borough
- Get the latitude and the longitude coordinates of each Postal code
- Merging the geographical data frame with neighborhood data frame according to the Postal Code

Our Neighborhood Candidate Data is shown in the below table.

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

After cleaning data, we should add the Latitude and Longitudes of each location.

We build a data frame of the postal code of each neighborhood along with the borough and neighborhood name.

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

We merge the geographical dataframe with neighborhood dataframe according to the Postal Code

	PostalCode	Borough	Neighborhood	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	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

we find out how many neighborhoods are in each borough

```
Borough
Central Toronto            9
Downtown Toronto          17
Downtown Toronto Stn A     1
East Toronto               4
East Toronto Business      1
East York                  4
East York/East Toronto     1
Etobicoke                 11
Etobicoke Northwest        1
Mississauga                1
North York                24
Queen's Park               1
Scarborough               17
West Toronto               6
York                       5
```

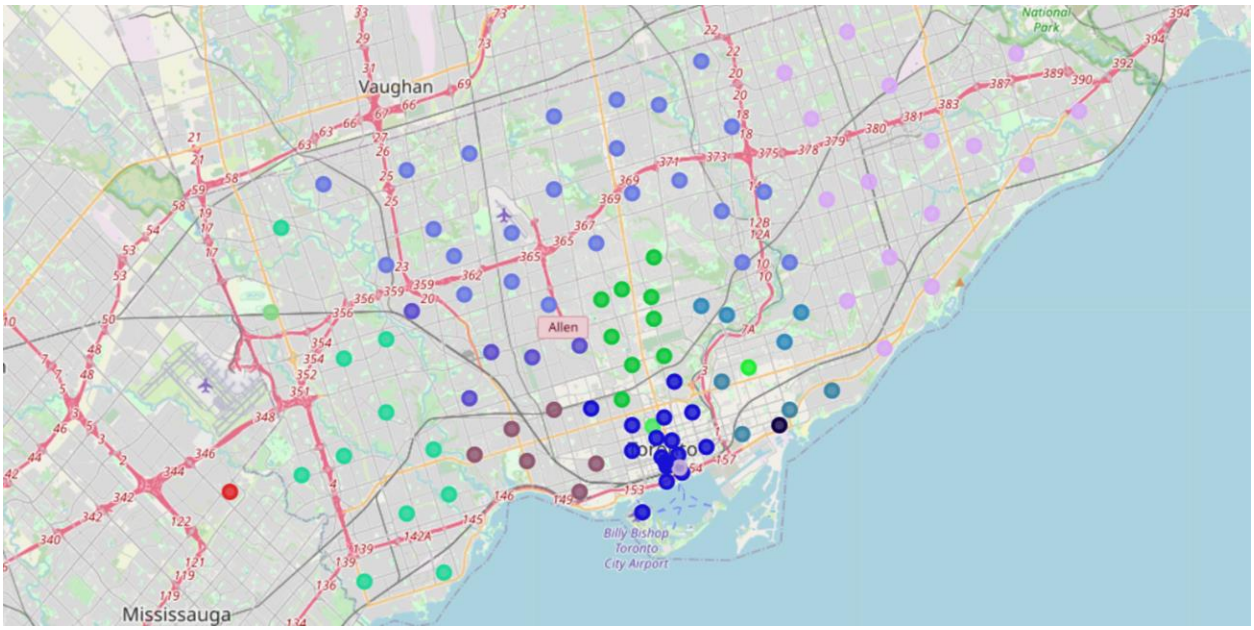
Visualizing all the boroughs in Toronto

Create a list and store all unique borough names

```
['North York',
 'Downtown Toronto',
 "Queen's Park",
 'Etobicoke',
 'Scarborough',
 'East York',
 'York',
 'East Toronto',
 'West Toronto',
 'East York/East Toronto',
 'Central Toronto',
 'Mississauga',
 'Downtown Toronto Stn A',
 'Etobicoke Northwest',
 'East Toronto Business']
```

The geographical coordinates of Toronto are

43.704607733980595, -79.3971529116505



Methodology

In this project, we will focus our efforts on identifying areas of Toronto with a low number of yoga studios. We will limit our analysis to an area ~5km around the city center. In the first step, we gathered the necessary information: the location of every Yoga studio within 5 kilometers of the Toronto center, as determined by Foursquare data.

The second step in our analysis will be the calculation and exploration of 'yoga studio density' across different areas of Toronto - we will use heatmaps to identify a few promising areas close to the center with a low number of yoga studios and focus our attention on those areas.

In the third and final step, we will concentrate on the most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders:

We will take into consideration locations without Yoga studios within a radius of 500 meters. We will present a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones/neighborhoods/ addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Data Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of yoga studio in every area candidate.

The number of venues per neighborhood

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	5	5	5	5	5	5
Alderwood, Long Branch	7	7	7	7	7	7
Bathurst Manor, Wilson Heights, Downsview North	21	21	21	21	21	21
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	23	23	23	23	23	23
Berczy Park	58	58	58	58	58	58
Birch Cliff, Cliffside West	4	4	4	4	4	4
Brockton, Parkdale Village, Exhibition Place	22	22	22	22	22	22
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	17	17	17	17	17	17
Caledonia-Fairbanks	4	4	4	4	4	4

There are 272 unique venue categories.

We analyze each neighborhood
The column "Venue Category" has a categorical value. As a result, we must convert it to numerical values using one hot encoding.
Then group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

	Neighborhoods	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.043478	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Here we only require the "Neighborhoods" and "Yoga Studio" columns for the clustering. So, we'll group these two columns.

	Neighborhoods	Yoga Studio
0	Agincourt	0.0
1	Alderwood, Long Branch	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0
3	Bayview Village	0.0
4	Bedford Park, Lawrence Manor East	0.0

We will use k-means algorithm for clustering (k=4).

Then we create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

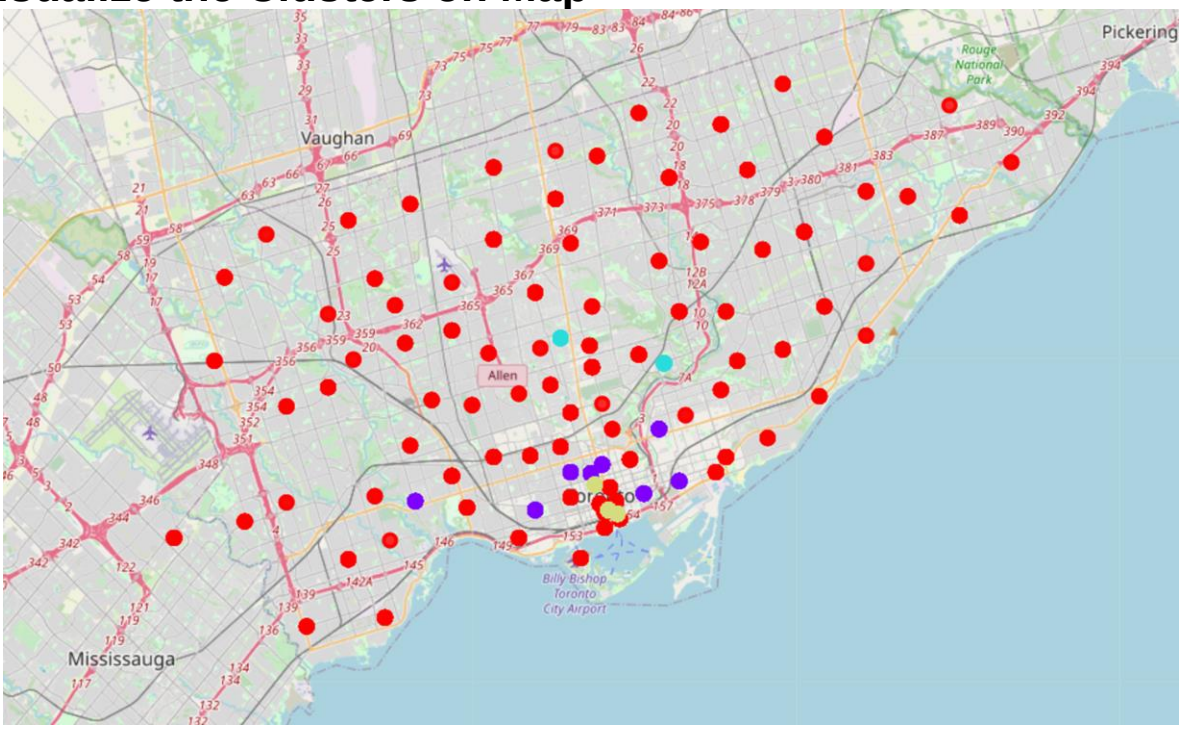
	Neighborhood	Yoga Studio	Cluster Labels
0	Agincourt	0.0	0
1	Alderwood, Long Branch	0.0	0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0
3	Bayview Village	0.0	0
4	Bedford Park, Lawrence Manor East	0.0	0

We merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood

	Neighborhood	Yoga Studio	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
31	Enclave of M5E	0.010417	3	43.646435	-79.374846	43.648147	-79.378752
31	Enclave of M5E	0.010417	3	43.646435	-79.374846	43.649615	-79.371747
31	Enclave of M5E	0.010417	3	43.646435	-79.374846	43.650400	-79.376700
31	Enclave of M5E	0.010417	3	43.646435	-79.374846	43.646791	-79.378769
17	Commerce Court, Victoria Hotel	0.010000	3	43.648198	-79.379817	43.651836	-79.378107

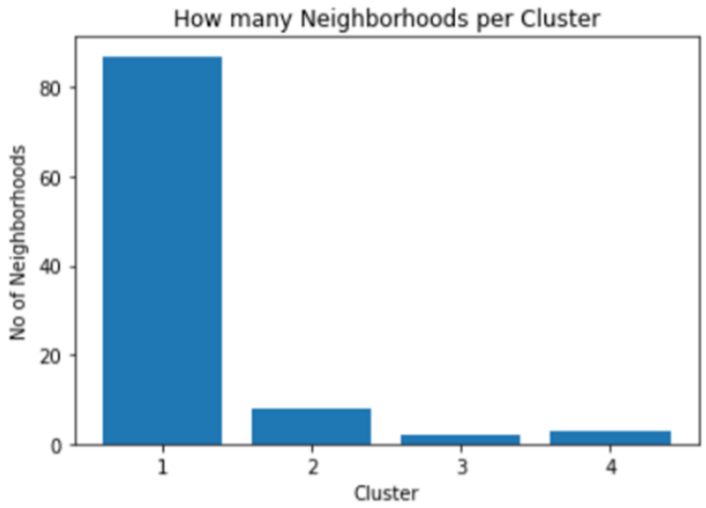
We see that there are a total of 14 locations with yoga studio in Toronto

Visualize the Clusters on Map



Number of Neighborhoods in each Cluster

	Neighborhood	Yoga Studio	Cluster Labels
0	Agincourt	0.0	0
1	Alderwood, Long Branch	0.0	0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0
3	Bayview Village	0.0	0
4	Bedford Park, Lawrence Manor East	0.0	0



Number of neighborhoods in each cluster

```
0    87
1     8
3     3
2     2
```

Results

Cluster 1

	Borough	Neighborhood	Yoga Studio	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
0	North York	Parkwoods	0.0	43.753259	-79.329656	43.752432	-79.334661
1	North York	Parkwoods	0.0	43.753259	-79.329656	43.751974	-79.333114
2	North York	Parkwoods	0.0	43.753259	-79.329656	43.754387	-79.333021
3	North York	Parkwoods	0.0	43.753259	-79.329656	43.751976	-79.332140
4	North York	Victoria Village	0.0	43.725882	-79.315572	43.726086	-79.313620

Cluster 2

	Borough	Neighborhood	Yoga Studio	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
0	Downtown Toronto	Regent Park, Harbourfront	0.023256	43.65426	-79.360636	43.653447	-79.362017
1	Downtown Toronto	Regent Park, Harbourfront	0.023256	43.65426	-79.360636	43.653559	-79.361809
2	Downtown Toronto	Regent Park, Harbourfront	0.023256	43.65426	-79.360636	43.653249	-79.358008
3	Downtown Toronto	Regent Park, Harbourfront	0.023256	43.65426	-79.360636	43.656265	-79.357119
4	Downtown Toronto	Regent Park, Harbourfront	0.023256	43.65426	-79.360636	43.656369	-79.356980

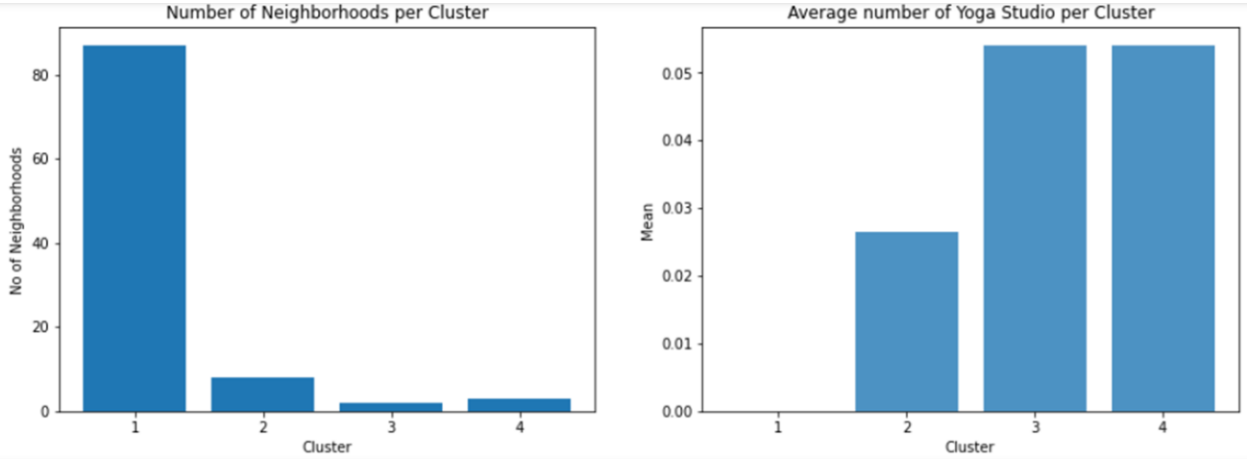
Cluster 3

	Borough	Neighborhood	Yoga Studio	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
0	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705810	-79.347044
1	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.704058	-79.348094
2	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.704596	-79.349670
3	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705751	-79.352054
4	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705511	-79.347064

Cluster 4

	Borough	Neighborhood	Yoga Studio	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
0	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705810	-79.347044
1	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.704058	-79.348094
2	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.704596	-79.349670
3	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705751	-79.352054
4	East York	Thorncliffe Park	0.05	43.705369	-79.349372	43.705511	-79.347064

Number of neighborhoods per cluster vs Average number of yoga studio in each Cluster



Discussion

The East York neighborhood of Toronto (cluster-3) has the highest average number of yoga studios.

Downtown Toronto has the second-highest concentration of yoga studios.

Looking at nearby venues, the best location for a new yoga studio is Parkwood, North York (cluster-1) because there are many Neighborhoods in that area but only a few yoga studios, removing any competition.

The second-best neighborhoods with a lot of potential are in Cluster-1, such as Victoria Village in North York. In this area, there is a good opportunity to open a new yoga studio.

Conclusion

This concludes the best findings for this project and advises the entrepreneur to open a yoga studio in these areas where there is little to no competition.

The k-Means algorithm is used for clustering the neighbors in city during this project.

Here we take a yoga studio as an example. We can do the same process to find the best place or neighborhood to open a start-up company.