

Applied Data Science Capstone Project

Introduction

A Registered Yoga Teacher in Toronto wants to open her Yoga & Meditation Studio to teach authentic Indian Hatha & Ashtanga Yoga. She expects the mental health issues to escalate as a consequence of Covid-19 and wants to help people to take care of their mental & physical health.

Objective

Finding the right location for such a studio is paramount to its success. The objective of this capstone project is to find the most suitable location for an entrepreneur to open a new Yoga Studio positioned as Authentic Indian Yoga.

Business Problem

Here is the business question I aim to solve with this project:

Where do I (an entrepreneur) open my Yoga Studio in Toronto after taking into consideration:

- Where are the majority of Yoga centres located? Identify the Yoga hub in the city, if it exists?
- Are there other Yoga Studios in the above area positioned as authentic Indian?

Target Audience

An entrepreneur who wants to open a yoga studio in Toronto.

This analysis will help them

- Understand the neighborhoods where yoga studios are concentrated in Toronto to get a sense of where yoga enthusiasts live/work?
- Or whether they are equally distributed across the city with a few major clusters?

This will help them make a data-driven decision regarding which neighborhoods might be ideal for opening her studio.

Data Requirements

To solve this problem, here are the data requirements:

1. List of neighborhoods in Toronto, Canada
2. Latitude and Longitude of these neighborhoods
3. Location of existing yoga studios

Data Sources/Analysis

- Scraping Wikipedia to find Toronto neighborhoods
- Using Geocoder package (or the CSV file with the co-ordinates used in Week 3) to get the latitude & longitude details of these neighborhoods
- Using Foursquare API to get venue data related to these neighborhoods
- Using K-Means Clustering Analysis to identify if multiple clusters of yoga studios exist or if there is a single cluster of neighborhoods where majority of yoga studios are located
- If there is a single major cluster, using data from Foursquare & Google search to understand if there are yoga studios positioned as authentic Indian ones to figure out the competitive landscape for such a boutique studio.

Methodology

1. To get the list of neighborhoods in Toronto, I would be scraping the following Wikipedia page:

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

This is accomplished by utilizing pandas HTML table scraping method as it is convenient to pull tabular data directly from a web page into the data frame.

Result: List of Neighborhood names & postal codes

2. To get the coordinates of these neighborhoods, I tried using Geocoder Package but it did not seem to work properly.

Then, I used the CSV file (list of Toronto pin-codes & their latitudes & longitudes data) provided in the course to match the coordinates of Toronto neighborhoods.

3. Now, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates if the data makes sense.
4. Next, I used the Foursquare API to pull the list of top 100 venues within a 500 meters radius of these pin codes.

To enable this, I created a Foursquare developer account to obtain an account ID and API key. Using these, I was able to pull the names, categories, latitude, and longitude details of the venues. I checked the unique categories of these venues.

Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. I prepared the data for the clustering analysis and specifically looked for yoga studios as a category.

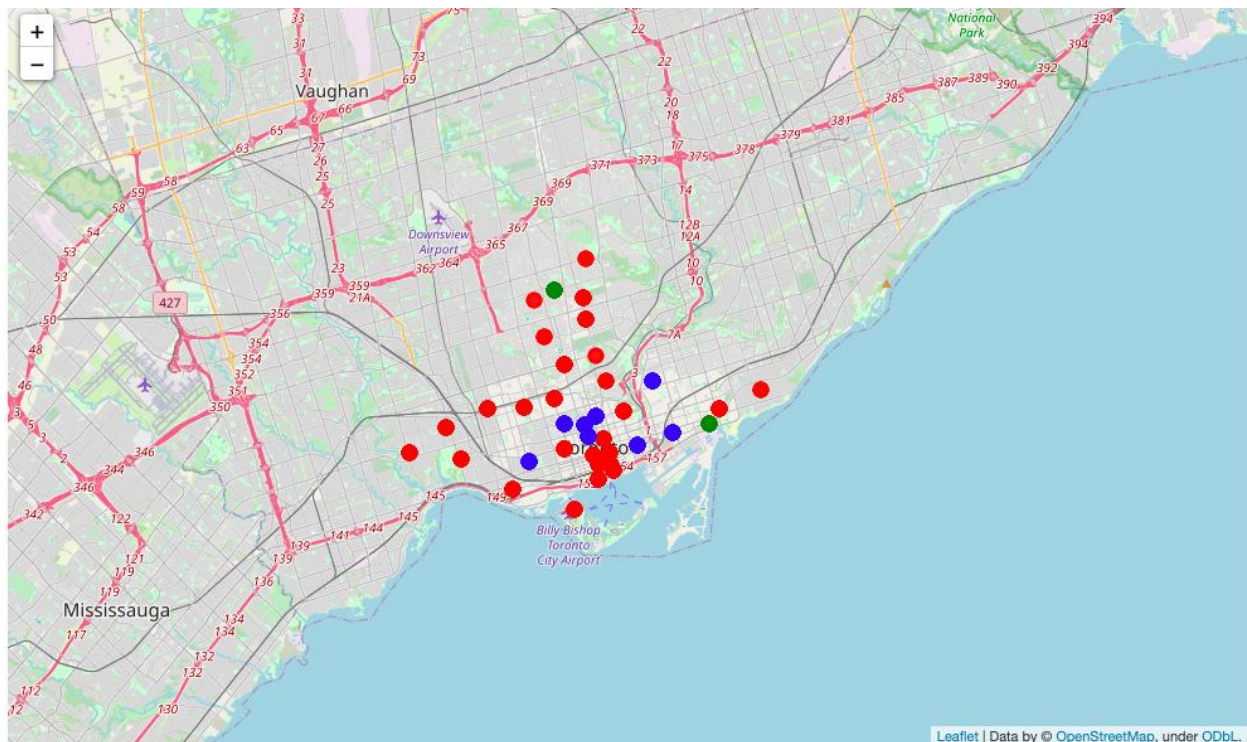
5. Lastly, I performed the clustering method by using k-means clustering.

K-means clustering algorithm identifies 'k' number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. The idea is to produce clusters with maximum intra-cluster and minimum inter-cluster distances.

It is a simple & popular unsupervised machine learning algorithm. Since, the objective of this analysis is to identify clusters of yoga studios, I have clustered the neighborhoods in Toronto into 3 clusters based on the frequency of presence of "Yoga Studios" in these neighborhoods.

Result

Cluster Map



The results from k-means clustering show that there is 1 primary cluster of neighborhoods (close to Downtown, Toronto) with the highest concentration of Yoga Studios.

Here is the screenshot of the cluster:

```
#Cluster 1
to_merged.loc[(to_merged['Cluster Labels'] == 1) & (to_merged['Venue Category'] == 'Yoga Studio')]
```

	Neighborhood	Yoga Studio	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Central Bay Street	0.014706	1	43.657952	-79.387383	The Yoga Sanctuary	43.661499	-79.383636	Yoga Studio
36	The Danforth West, Riverdale	0.023810	1	43.679557	-79.352188	Moksha Yoga Danforth	43.677622	-79.352116	Yoga Studio
23	Queen's Park, Ontario Provincial Government	0.030303	1	43.662301	-79.389494	The Yoga Sanctuary	43.661499	-79.383636	Yoga Studio
38	University of Toronto, Harbord	0.029412	1	43.662696	-79.400049	Sivananda Yoga Centre	43.662754	-79.402951	Yoga Studio
24	Regent Park, Harbourfront	0.021277	1	43.654260	-79.360636	The Yoga Lounge	43.655515	-79.364955	Yoga Studio
6	Church and Wellesley	0.027397	1	43.665860	-79.383160	The Yoga Sanctuary	43.661499	-79.383636	Yoga Studio
6	Church and Wellesley	0.027397	1	43.665860	-79.383160	Bikram Yoga Yonge	43.668205	-79.385780	Yoga Studio
19	Little Portugal, Trinity	0.024390	1	43.647927	-79.419750	YogaSpace	43.647607	-79.420133	Yoga Studio
32	Studio District	0.025000	1	43.659526	-79.340923	Spirit Loft Yoga	43.663548	-79.341333	Yoga Studio

Inference & Recommendations

This indicates most of the yoga enthusiasts are living/working in and around Cluster 1 & this might be a good location to start looking for available locations.

The only downside is heightened competition.

However, a google search on all the above venues listed in the screenshots suggests that there is only 1 which is an authentic Indian Yoga Centre (Sivananda Yoga Centre) in this vicinity.

Hence, there is an opportunity to position the yoga studio as an authentic Indian one. The fact that the entrepreneur for whom I was doing this analysis (my wife 😊) is highly trained (900+ Hours of Training) & certified by multiple Yoga institutes in India would lend credibility to the claim and attract potential customers.

The success would eventually depend on the quality of the teachings but it's important to be present in an area where the customers are active. Hence, my recommendation would be to start looking for a suitable location in this cluster only.