Choosing an ideal location to open a new CrossFit box

Suzie Amey – 22 February 2019

Table of Contents

1. Introduction

1.1 Background

CrossFit is a class of fitness philosophy and lifestyle founded in the year 2000 by Greg Glassman and Lauren Jenai. It has grown in popularity over the last 20 years and has become an international competitive sport. It is described as 'Functional Fitness' and incorporates Olympic weightlifting, high-intensity interval training, gymnastics skills, powerlifting, calisthenics, among other disciplines, and is designed to bring a dramatic uptake in fitness and strength with a feeling of comradery as it is practiced in groups with the leading of a qualified coach.

1.2 Problem

A client is looking to open a new CrossFit training gym in the city of Toronto, Ontario, and has asked for us to come up with an ideal location based on statistics and data from the area. A CrossFit training gym is known as a Box, and in order to call themselves a CrossFit Box, each gym must pay a yearly membership fee to the CrossFit brand which makes them an affiliate. As such, Crossfit memberships for individuals are known to be quite pricey, hence, the ideal location would likely be in an area where the household income is on the higher end of the scale. Also, it would be good to find a location which is not too close to any other CrossFit Box, as the residents of that area may already be members somewhere else. It would also be a good idea to consider whether the neighbourhood is one where health and fitness is a priority, so one where there are a number of sports and fitness shops and health food stores would be a good fit.

1.3 Interest/Audience

The audience in this case is the client who is looking to open a new Box. Their obvious reason for interest is to open the Box in an area where they will bring in enough clients to make their gym a success.

2. Data

2.1 Data Sources

As mentioned, the average income of a neighbourhood is going to be a factor in deciding where to open a new CrossFit Box; the following website has information on just such information:

https://www03.cmhc-schl.gc.ca/hmip-pimh/en/TableMapChart/TableMatchingCriteria?GeographyType=MetropolitanMajorArea&GeographyId=2270&CategoryLevel1=Population%2C%20Households%20and%20Housing%20Stock&CategoryLevel2=Household%20Income&ColumnField=HouseholdIncomeRange&RowField=Neighbourhood&SearchTags%5B0%5D.Key=Households&SearchTags%5B0%5D.Value=Number&SearchTags%5B1%5D.Key=Statistics&SearchTags%5B1%5D.Value=AverageAndMedian

Another factor which may be beneficial to consider is Population Density. Opening a Box in an area with low population density means there would be fewer potential clients to attract. The link below has information on population density for Toronto Neighbourhoods:

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

I will also use Foursquare location data to find out which neighbourhoods already have a CrossFit Box, and also, I will look for areas which show an interest in personal health and fitness; for example, these areas could have health food and sports equipment stores and a number of other outlets for physical activities.

3. Methodology

3.1 Reading in the Data

Firstly, I read the income data from the given website into a Pandas Data Frame in my Jupyter notebook, kept only the columns which were relevant to my research, and renamed these columns to make the data frame easier to understand.

	Neighbourhood	Income
0	Agincourt/Malvern	68108
1	Ajax/Pickering	98363
2	Alderwood	83249
3	Aurora	106761
4	Banbury-Don Mills/York Mills	90620

Next, I read in data to a new data frame to connect postal codes and Latitude/Longitude values to each neighbourhood, and got rid of any 'Not assigned' values. I then merged this data frame with the first one, and sorted the new data frame in descending order of income for each neighbourhood.

	Neighbourhood	Latitude	Longitude	Income
27	Rosedale	43.679563	-79.377529	114625
1	Rouge	43.806686	-79.194353	101482
7	The Beaches	43.676357	-79.293031	100365
25	Moore Park	43.689574	-79.383160	97762
4	Guildwood	43.763573	-79.188711	87538
28	Alderwood	43.602414	-79.543484	83249
16	Riverdale	43.679557	-79.352188	77819
19	Cliffcrest	43.716316	-79.239476	75291
23	Forest Hill North	43.696948	-79.411307	70920
9	Bathurst Manor	43.754328	-79.442259	69634

From this information, I was able to select the top 5 neighbourhoods to research further for suitability, these were:

- **Rosedale**
- **Rouge**
- **The Beaches**
- **Moore Park**
- **Guildwood**

3.2 Creating a Map with Folium

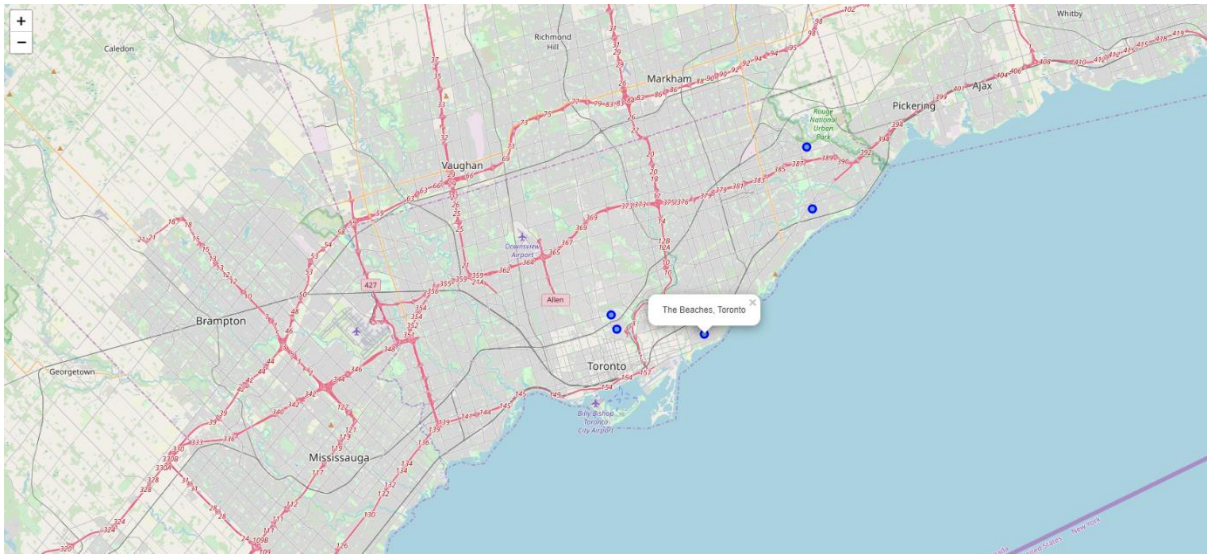
I used Geolocator to find the Latitude and Longitude values for the city of Toronto, and then used these values with Folium to Create a visual map of Toronto with markers showing the five selected locations, found previously.

Next, I created a map of Toronto, using Folium, with markers for the locations of the top 5 Neighbourhoods

```
In [92]: address = 'Toronto, CA'

         geolocator = Nominatim(user_agent="toronto_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Toronto are 43.653963, -79.387207.
```

```
In [93]: # create map of Toronto using latitude and longitude values
         map_toronto = folium.Map(location=[latitude, longitude], zoom_start=10)

         # add markers to map
         for lat, lng, neighbourhood in zip(top_five['Latitude'], top_five['Longitude'], top_five['Neighbourhood']):
             label = '{}, {}'.format(neighbourhood, 'Toronto')
             label = folium.Popup(label, parse_html=True)
             folium.CircleMarker(
                 [lat, lng],
                 radius=5,
                 popup=label,
                 color='blue',
                 fill=True,
                 fill_color='#3186cc',
                 fill_opacity=0.7,
                 parse_html=False).add_to(map_toronto)

         map_toronto
```



Map of Toronto, C.A.

3.3 Foursquare

Using Foursquare, I created a search query for 'Crossfit' and applied it to each neighbourhood individually.

```
In [95]: address1 = 'Rosedale, Toronto'

         geolocator = Nominatim(user_agent="foursquare_agent")
         location = geolocator.geocode(address1)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of Rosedale are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Rosedale are 43.6783556, -79.3807457.
```

```
In [96]: search_query = 'CrossFit'
         radius = 500
```

```
In [97]: url2 = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius)
         url2
```

```
Out[97]: 'https://api.foursquare.com/v2/venues/search?client_id=QELSOPIU4YTD3H0NS03HBFGWEHUEREWU51ACI10SV2BCGFIA&client_secret=RP02DMBFW3WDUSAAYHOSVZSKEYFHUHWEK1JJWZ5FPS0RQQB4&ll=43.6783556,-79.3807457&v=201
         dius=500'
```

```
In [98]: results = requests.get(url2).json()
         results
```

```
Out[98]: {'meta': {'code': 200, 'requestId': '5e5129b721678501c4c7cec'},
          'response': {'venues': [{'id': '4b7f01d9f964a520e00f30e3',
             'name': 'uptown fitness',
             'location': {'lat': 43.672672,
              'lng': -79.379627,
              'labeledLatLngs': [{'label': 'display',
```

From running this query, I found that, out of the five original neighbourhoods, only two of them didn't already have a CrossFit Box, they were:

- **Rouge**
- **Guildwood**

I then ran a further query to see whether either of these two neighbourhoods showed any interest in sports and recreation. I got no results back from Rouge, and two venues returned from Guildwood.

From this, I would recommend that Guildwood be chosen as an appropriate location to open a new CrossFit Box.

4. Results

From my data analysis, my initial findings for a suitable location, based on income were:

Neighbourhood	Mean Annual Income ($)	Latitude	Longitude
Rosedale	114625	43.6796	-79.3775
Rouge	101482	43.8067	-79.1944
The Beaches	100365	43.6764	-79.293
Moore Park	97762	43.6896	-79.3832
Guildwood	87538	43.7636	-79.1887

After using Foursquare to find whether any of these neighbourhoods already had a CrossFit Box open in the area, I was able to narrow my results down to:

Neighbourhood	CrossFit Box
Rosedale	Yes
Rouge	No
The Beaches	Yes
Moore Park	Yes
Guildwood	No

Rouge and Guildwood were then examined further and, the final location chosen to recommend was **Guildwood, Toronto.**

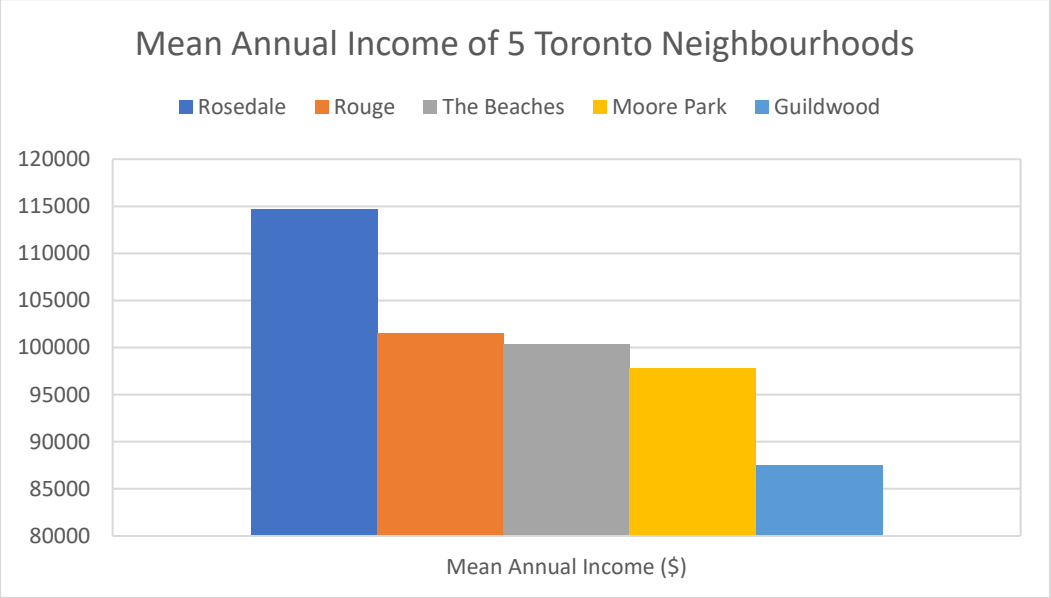
5. Conclusion

After analysing the data, the first set of results gave me a list of 5 possible locations which I could recommend for my client to open a new CrossFit Box, based on average income in the area – as discussed in the introduction, CrossFit membership tends to be quite pricey, so this factor was very important to consider. I wanted to narrow the list down as much as possible, and thought that any neighbourhood already having a CrossFit Box would likely not be suitable, as the residents there would then have to choose between more than one place to become a member, reducing the number of potential customers for my client. Using Foursquare, I found two neighbourhoods which did not already have a CrossFit Box registered in the area, so these were my final shortlisted locations to recommend. I wanted to make the decision as easy as possible for my client at the same time as giving them the best recommendation, so I thought of any other factors which could turn a potential area into a successful place to open their business. As CrossFit members are likely

to be interested in sports and personal health/fitness, I ran a query to look for evidence of interest in these factors in the two neighbourhoods. Guildwood was the one which returned results showing an interest in sports, so my final recommendation to my client would be to look at opening their new CrossFit Box in this neighbourhood.

6. Discussion

Although I was able to narrow down, first five neighbourhoods, then two, and then one final neighbourhood to recommend to my client using the given tools of data analysis in Pandas and then Foursquare location data, I believe that this exercise could've gone into much more detailed analysis of the many neighbourhoods of Toronto to find the perfect location for opening a new CrossFit Box. For example, even within the 5 neighbourhoods that I first extracted from the data, there was quite a lot of variation in the selection factor, i.e. mean household income, as shown in the graph below.



Other factors which could've been considered are:

- Population density
- Average age of population
- Type of neighbourhood, i.e. Rural or Urban
- Proximity to public transport routes
- Availability of property
- Cost of property to rent or buy
- Space to build new properties or develop existing ones

And the list could go on.

For this assignment, however, the idea was to show that we could

1. Choose a problem/topic for investigation
2. Use data science methods to solve analyse data towards solving the problem
3. Investigate a geographical location using location data, such as Foursquare
4. Provide evidence of our methodology and results
5. Discuss these results and present them to an audience

As such, I believe that the criteria have been met as required.