The Battle of the Neighborhoods

Christopher González Delgado

Introduction

Physical exercise and the desire to have a t, muscular and attractive body are a trend in our society. Therefore, it would not be unreasonable for a person to see an ideal business opportunity in setting up a gym. And although the idea is good, there is an unknown question that we must solve: where should we place the gym so that it is a successful site? Through this project you will be able to observe the methodology to locate in a good area the business of the 21st century: the tness centers.

The objective of this project is to nd the ideal areas to open a gym in Toronto taking into account some characteristics of the area. The audience that could show the most interest consists of investors looking for a safe bet or entrepreneurs who want to start their business.

Data description and extraction

Firstly, we have a database (obtained from wikipedia) that oers us the postal code, the boroughs and the neighborhoods, and secondly, another database that consists of the coordinates of these places. On the other hand, the Python package "geopy" has been used to obtain the coordinates of Toronto on the map, while the Foursquare platform has been used to obtain information about the venues in Toronto.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village

Methodology - Understanding the problem

Obviously the main problem is: where should we place the gym? But to complete this question
we must also add certain conditions: where should we place the gym if we take into
account the other shops and establishments in each neighborhood?

Methodology - Preparation of the data | Processing

The data we have obtained, is dirty i.e. it has some unassigned variables
whose value we cannot intuit or calculate. Therefore, rst of all we are going to delete all the data whose "Borough" has the value "Not assigned". Once this is done, we can see that there are several postal codes that appear more than once and with dierent neighborhoods, so we 1
will proceed to merge these neighborhoods in the same row for each postal code. Finally, we will focus only on those boroughs that explicitly have the word "Toronto" in their name.

```
In [2]: df = pd.read_html('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M')
        df = df[0]
        dff = df.drop( df[(df['Borough']=='Not assigned')].index)
        dff = dff.reset_index(drop=True) #Let's reset the index

In [8]: dff.columns=(['Postal Code', 'Borough','Neighborhood'])
        dff.head()
        df3= dff.groupby(['Postal Code', 'Borough'])['Neighborhood'].apply(list).apply(lambda x: ", ".join(x)).to_frame()
        df3 = df3.reset_index() #Let's add the index to each row
```

Methodology - Working with the data

The procedure that has been carried out has been as follows: First of all we have obtained the list of venues in toronto.

In [13]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):
    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)
        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&lim
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)
        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']
        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])
    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [14]:
```
LIMIT=50
toronto_venues = getNearbyVenues(names=DFF['Neighborhood'],
                                 latitudes=DFF['Latitude'],
                                 longitudes=DFF['Longitude']
                                )
```

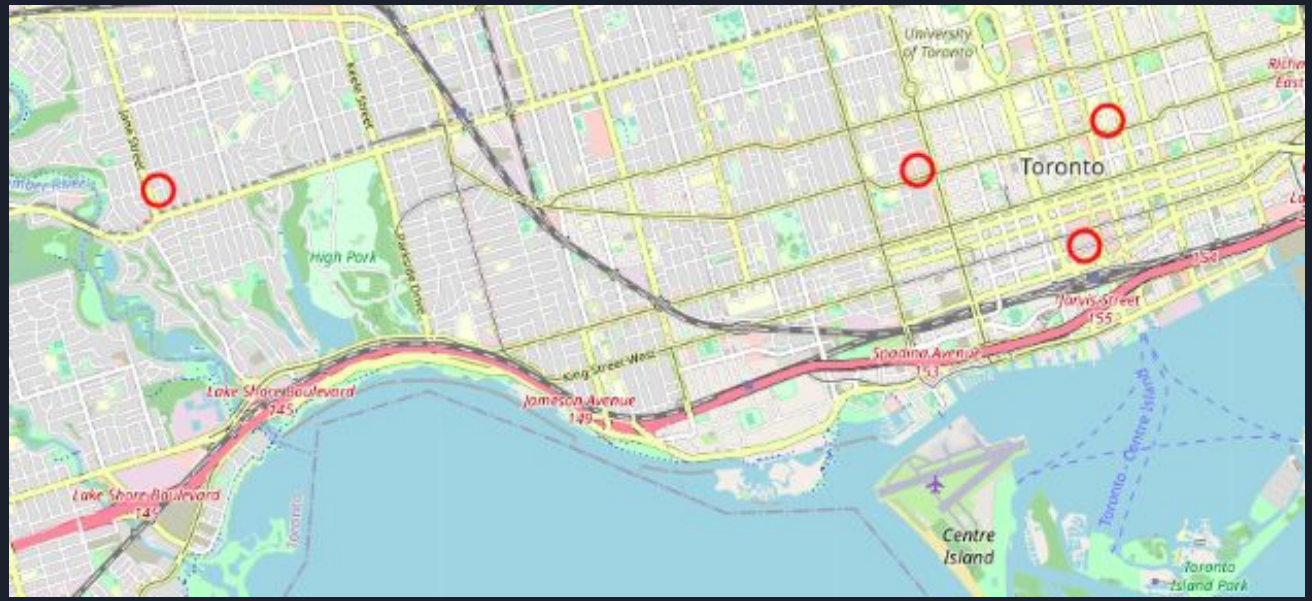
Methodology - Working with the data

From these data, it has been possible to search for the neighborhoods that do not have any (commonly) gym or tness center in their surroundings (we are interested in the fact that they do not have gyms, so ours will draw the attention of people who live near that area). On the other hand, it has been taken into account that it is an area with many cafes and hotels. The rst is evident: coee is a good sports supplement and it is common to drink coee only before training, in addition to being common to take some food or protein shake after nishing workouts. And the second reason is because when people go on vacation they are likely to want to maintain moderate physical activity: if we can locate a gym near a hotel we can have a fairly constant ow of customers.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Garden District, Ryerson	Coffee Shop	Café	Clothing Store	Cosmetics Shop	Ramen Restaurant	Tea Room	Theater	Bookstore	Fast Food Restaurant	Japanese Restaurant
1	Toronto Dominion Centre, Design Exchange	Coffee Shop	Café	Seafood Restaurant	Hotel	Japanese Restaurant	Restaurant	Beer Bar	Bakery	Gym / Fitness Center	Deli / Bodega
2	Kensington Market, Chinatown, Grange Park	Café	Mexican Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Burger Joint	Vietnamese Restaurant	Bar	Pizza Place	Dessert Shop	Record Shop
3	Runnymede, Swansea	Café	Coffee Shop	Pizza Place	Pub	Sushi Restaurant	Italian Restaurant	Yoga Studio	Bookstore	Bar	Smoothie Shop

Methodology - Working with the data

Finally, a map has been constructed using Folium to visualize the areas where it would be ideal to set up a gym



Discussion and conclusions

Through this project, it has been possible to demonstrate that using powerful data science tools and using a simple algorithm, really powerful results can be obtained. Specically, we have seen that there are (at least) 4 areas in which it would be ideal to open a gym according
to the conditions, which are: cafes and hotels nearby.
Much deeper analysis could be done using other variables that play a vital role in all of this. For
example, including the average income of the neighborhoods or the average age. If we include
the rent we can focus on a more auent public that is willing to pay higher amounts for the service; Taking into account the age of potential clients, we can look for a young audience as they are interested in obtaining a muscular and aesthetic body.

END