Introduction/Business Problem

Our stakeholder is a company, which plans to build a new food restaurant, espically for fitness affine people, who only want to eat healthy and vital food.

Now they know that they will build their restaurant in Toronto, but they are uncertain which location should be the best.

So now it is our job to analyze the different neighborhoods to find the perfect place for the new food restaurant, so that it is in range for fitness affine people. Therefore we will need to analyze every neigborhood and check their actual venues. So we can think about which people will life near that location and so, which people may want to eat something in the new restaurant.

Data
We will use three different data source to fullfill the assignment.

First of all we use the wikipedia page:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
to get the postal codes, neighborhoods and boroughs from the city of toronto.

As a second datasource we use the .csv data from the last assignment, so that we can get the geo-coordinates of each neighborhood (latitude and longitude)

The last datasource is the foursquare api, which we will use the get informations about the neighboorhoods. E.g. we can check if there are any yoga studios or other fitness center in the neighborhood

Methodology

First of all we took the data from the wikipedia webpage, so that we got information about the neighborhoods and boroughs. After this dataframe was created, we took longitude and latitude information and merged these two tables, on the same Postal Code.

	PostalCode	Borough	Neighborhood
0	M4B	East York	Woodbine Gardens, Parkview Hill
1	M6P	West Toronto	High Park, The Junction South
2	M6L	North York	Maple Leaf Park, North Park, Upwood Park
3	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...
4	M9P	Etobicoke	Westmount

	Postal Code	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

	PostalCode	Borough	Neighborhood	Postal Code	Latitude	Longitude
0	M4B	East York	Woodbine Gardens, Parkview Hill	M4B	43.706397	-79.309937
1	M6P	West Toronto	High Park, The Junction South	M6P	43.661608	-79.464763
2	M6L	North York	Maple Leaf Park, North Park, Upwood Park	M6L	43.713756	-79.490074
3	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	M4V	43.686412	-79.400049
4	M9P	Etobicoke	Westmount	M9P	43.696319	-79.532242

After archiving this table, we used the foursquare API to get venue information for each neighborhood. This results in a venues table, which looks like this:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Jawny Bakers	43.705783	-79.312913	Gastropub
1	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Toronto Climbing Academy	43.709136	-79.314790	Rock Climbing Spot
2	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	East York Gymnastics	43.710654	-79.309279	Gym / Fitness Center
3	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Shoppers Drug Mart	43.705892	-79.312410	Pharmacy
4	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	TD Canada Trust	43.705740	-79.312270	Bank
5	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	Pizza Pizza	43.705159	-79.313130	Pizza Place

Now we wanted to categorize this data. Therefore we took a look at the Venue Category. We splitted the category in two types: Restaurants and Sport.
Therefore we counted every sport place as a 1, a Fast Food place as -1 and a Restaurant as 1 point.
So we got this table.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	RestaurantNumber	SportNumber
2	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	East York Gymnastics	43.710654	-79.309279	Gym / Fitness Center	0.0	1
12	Woodbine Gardens, Parkview Hill	43.706397	-79.309937	TKTO - Toronto Knife Throwing Organization	43.709966	-79.313411	Athletics & Sports	0.0	1
14	High Park, The Junction South	43.661608	-79.464763	nodo	43.665303	-79.465621	Italian Restaurant	1.0	0
18	High Park, The Junction South	43.661608	-79.464763	Playa Cabana Cantina	43.665315	-79.465548	Mexican Restaurant	1.0	0
28	High Park, The Junction South	43.661608	-79.464763	Silk	43.665291	-79.466238	Thai Restaurant	1.0	0

Then we used the score points and summed up every Number for each Neighborhood.
Neighborhoods with Nan values, got removed from the dataframe.

	PostalCode	Borough	Neighborhood	Postal Code	Latitude	Longitude	SportNumber	RestaurantNumber
0	M4B	East York	Woodbine Gardens, Parkview Hill	M4B	43.706397	-79.309937	2.0	0.0
1	M6P	West Toronto	High Park, The Junction South	M6P	43.661608	-79.464763	0.0	4.0
3	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	M4V	43.686412	-79.400049	1.0	3.0
4	M9P	Etobicoke	Westmount	M9P	43.696319	-79.532242	0.0	2.0

In the last step we clustered the neighborhoods in 5 clusterts, which resulted in the following Map of Toronto:
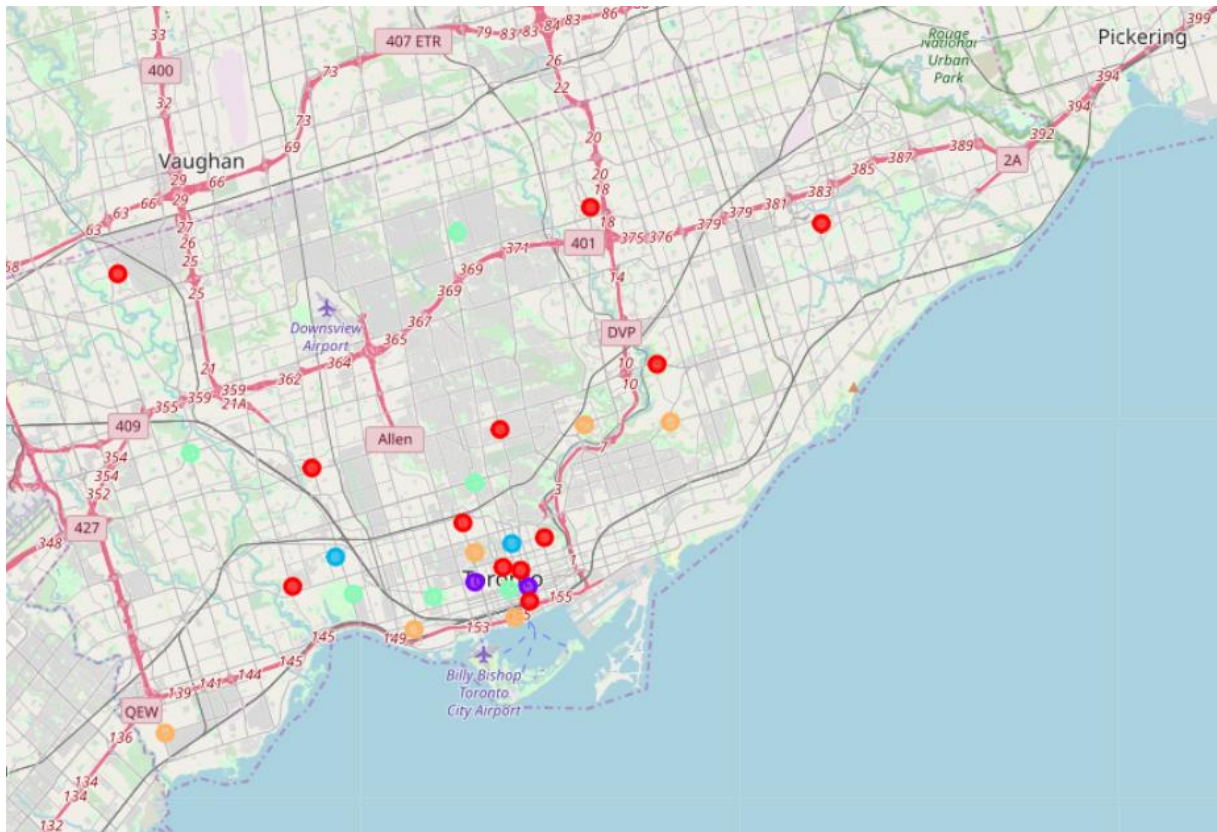
You can see, that those are not all neighborhoods of Toronto, because we dropped those, who had not have a score of SportPlaces + RestaurantPlaces > 0.

Results

In cluster 1-3 are no sport places. So they are not interesting for us. Cluster 4 contains one item with a sport studio, but also has a relativ high quote of restaurants When taking a look at cluster five:

	PostalCode	Latitude	Longitude	SportNumber	RestaurantNumber
0	M4B	43.706397	-79.309937	2.0	0.0
7	M8W	43.602414	-79.543484	1.0	0.0
9	M4H	43.705369	-79.349372	1.0	1.0
11	M6K	43.636847	-79.428191	1.0	1.0
42	M5J	43.640816	-79.381752	1.0	0.0
94	M5S	43.662696	-79.400049	1.0	0.0

You can see, that there are only those neighboorhoods, which have >= 1 sport places and not so much restaurants.

So it would be the best plan to build the new restaurant in this cluster.

Discussion

According to data, which we got. We were able to say, where the best place for the new restaurant will be. But we also got some information of the sports and restaurant affinity of the neighborhoods in the city of Toronto.
One problem in this report is, that we had to define manually, which venue category is Fast Food, Sport or a Restaurant. Therefore we used string comparison with keywords. But if we forgot some keyword, we also missed the value of that row, so our data could be a little bin inconsistently.

Conclusion

In this report we showed, that the best place to build a restaurant is in the neighborhood with the PostalCode M4B in Toronto. Because there are some sport studios and actually no restaurants.