

A wide-angle photograph of the Toronto skyline across Lake Ontario. The CN Tower is the central focus, surrounded by various skyscrapers and buildings. The water in the foreground is dark blue with some whitecaps, and a few sailboats are visible. The sky is a vibrant blue with wispy white clouds.

OPENING A LOW SALAD BAR IN TORONTO

USING THE FOURSQUARE API

2012 George Socka

THE GOAL

Finding the best location
for a low calorie salad bar
in Toronto



THE AUDIENCE



Data driven entrepreneurs looking to establish a healthy food restaurant and has identified their preferred clientele as fitness enthusiasts are the target audience for this project.



HOW WILL WE GET THERE

a. General Approach

Collect the Toronto city data from the previous week's notebook.

Using Foursquare API, we will get all venues for each neighbourhood then narrow down to neighbourhoods in the borough with the highest gyms

The feature to be extracted

The “star” feature to be extracted is the number of Gyms in the each in each neighbourhood, the frequency data will enable to discover the neighbourhood with the highest. This will be done by one-hot encoding the venue data gotten directly from foursquare

The option of franchising

clustering based on the singular feature that is gyms, the goal is find similar neighbourhood that can serve as upcoming location should the business attempt to grow

Data Visualization and some statistical analysis.

Analysing using Clustering (Specially K-Means):

Find the best value of K

Visualize the neighbourhoods with a number of Gyms Restaurants. Compare the Neighbourhoods to Find the Best Place for Starting up a Restaurant. Inference From these Results and related Conclusions

WIKI + GEOSPATIAL DATA

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.81	-79.19
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78	-79.16
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76	-79.19
3	M1G	Scarborough	Woburn	43.77	-79.22
4	M1H	Scarborough	Cedarbrae	43.77	-79.24

Data downloaded from the Cousera site and scraped from table from the wiki page ['https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) were combined into one table

GETTING VENUES

Nearby venues were derived in JSON format then converted into a Pandas data frame and the important columns filtered out

```
venue.location.address venue.location.crossStreet venue.location.lat venue.location.lng venue.location.labeledLatLngs \
```

```
1530 Albion Rd Albion Mall 43.74 -79.58 [{"label": "display", "lat": 43.741685, "lng": ...
```

```
80-1530 Albion Rd at Kipling Ave. (Albion Centre) 43.74 -79.58 [{"label": "display", "lat": 43.74120870478487...
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.81	-79.19	Wendy's	43.81	-79.20	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.78	-79.16	Chris Effects Painting	43.78	-79.16	Construction & Landscaping
2	Rouge Hill, Port Union, Highland Creek	43.78	-79.16	Royal Canadian Legion	43.78	-79.16	Bar
3	Guildwood, Morningside, West Hill	43.76	-79.19	RBC Royal Bank	43.77	-79.19	Bank
4	Guildwood, Morningside, West Hill	43.76	-79.19	G & G Electronics	43.77	-79.19	Electronics Store

WHATS OUR STAR FEATURE

	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum
0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

There were 274 possible venue categories
However only one feature was clearly
stated in the objective as the target
audience i.e., the density of GYMS, thus
that feature was filtered out of the dataset
and sorted in descending order of
magnitude to provide the best location,
this was determined to be the Willowdale,
Newton brook in the North York
borough

WHO HAS THE MOST GYMS

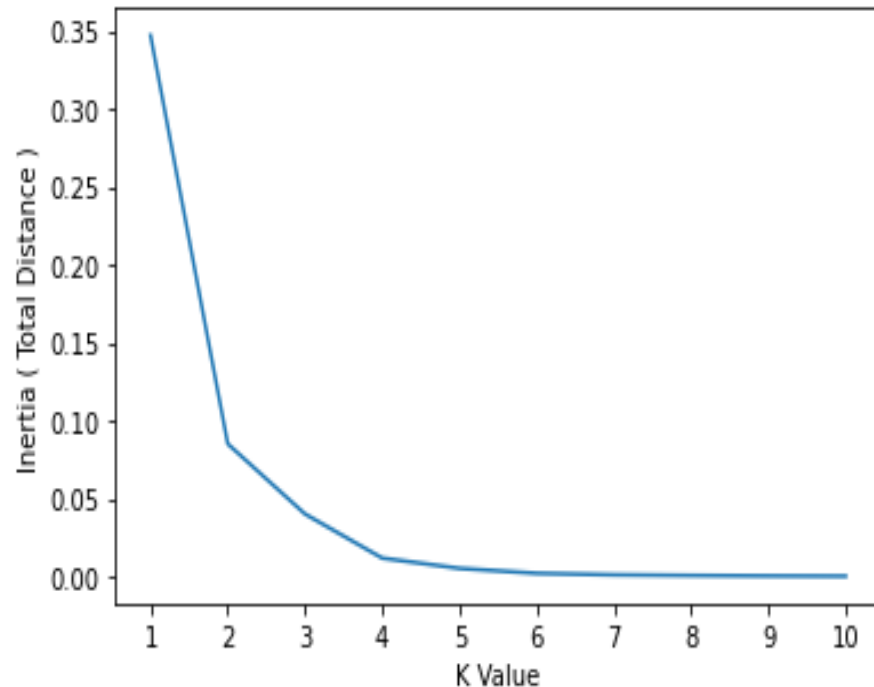
	Neighborhood	Gym
95	Willowdale, Newtonbrook	0.50
21	Don Mills North	0.25
1	Alderwood, Long Branch	0.14
22	Don Mills South	0.10
58	New Toronto, Mimico South, Humber Bay Shores	0.08



WILLOW DALE NEWTON BROOK HAS
THE MOST GYMS

LETS FIND MORE LIKE IT?

```
done sorting  
end inner loop  
Iteration 1, inertia 0.00016526477255519856  
center shift 0.000000e+00 within tolerance 3.506871e-07
```



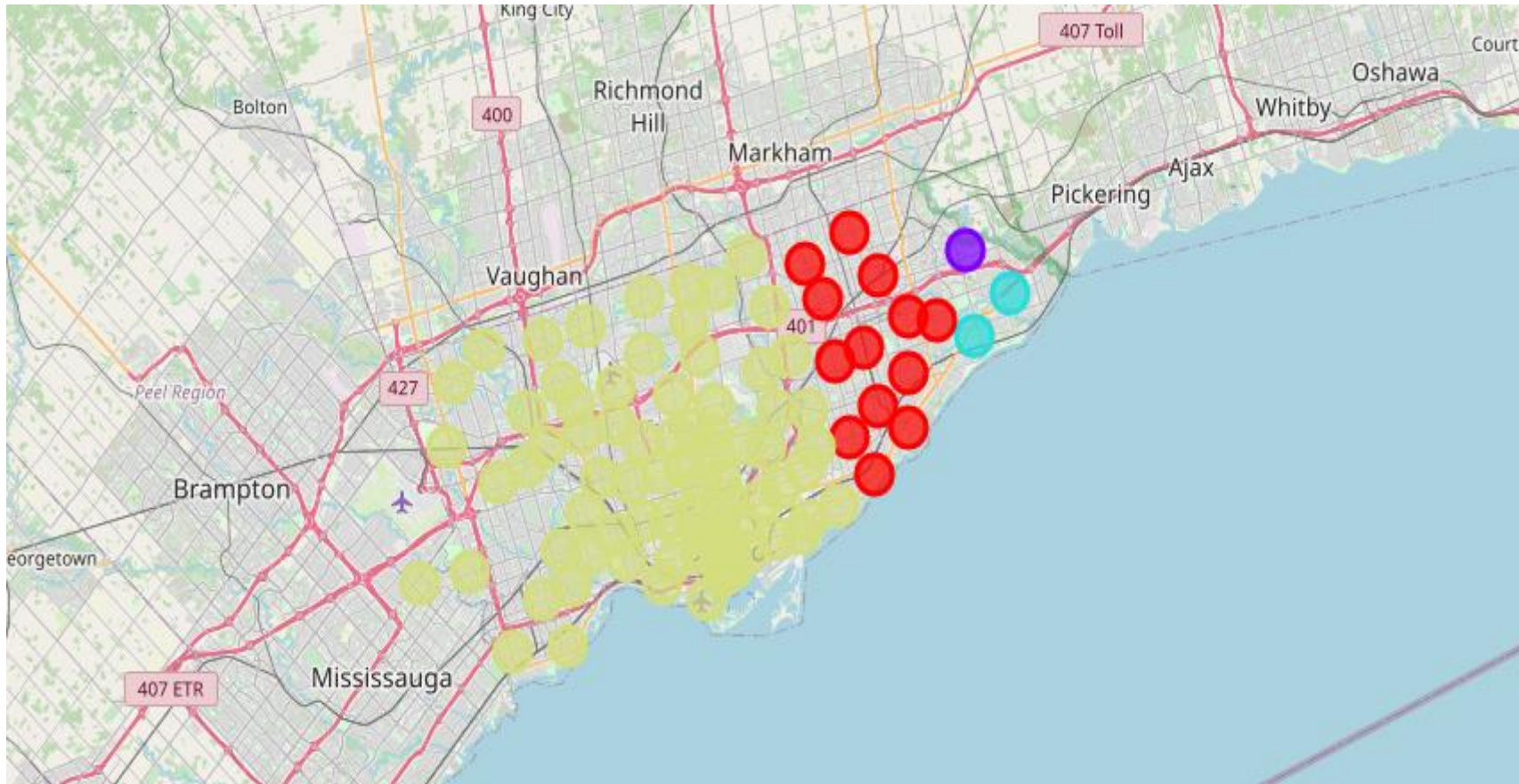
In order to determine similar neighbourhoods in the event of company franchising desired by the executives of the target audience, the neighbourhoods were clustered with gym density as the only criteria

a. K-means Clustering

i. Applying standard algorithms and their problems

I applied the Kmeans Clustering algorithm using Create a function that calculates Inertia for n times by sweeping through 1 to n to find the optimal cluster number. The optimal cluster number was determined to be 4 and that was utilize to create clusters,.

THE RESULTS



there were 83
neighbourhoods that
were highly suited for
expansion

THE CONCLUSION

The client is best advised to open their first venue in Willowdale, Newton brook, most of Toronto's neighbourhood appear to have a high density of gyms and the client is advised to open venues in order of gym frequencies in cluster 4

OPPORTUNITY FOR IMPROVEMENT

- a. Increasing the number of features utilized would have that would have improved the precision of our clustering, this could be made possible by
 - i. The inclusion of gym adjacent venues such as yoga studios and sports centres
 - ii. The study of similar businesses such as other low calories salad bars and juice places hoping to selected cluster based on the inverse relationship between all forms of similar and competing business
 - iii. This would result in the ideal location being one with a high density of fitness related facilities (gyms, yoga studios, sports centres) and a low density of healthy food options.
- a. Further exploratory analysis to discover relationships between venue categories, preferably to use this to determine features to be selected in combination with the Gyms
- a. The use of multiple clustering algorithms Such as mean shifting, hierarchical clustering method among others. This would have allowed for accuracy comparisons between models

THANK YOU

