



Finding the optimal spot for
our authentic Italian
Restaurant

The Battle of
Neighbourhoods
Sebastiaan Vrij

Table of Contents

Introduction	2
Business Problem.....	2
Data	2
Methodology	3
Results.....	4
Recommendations	5
Appendex 1 List overview Area's	6

Introduction

In this is a project for IBM Data Science Professional Certificate. In this project, I am creating a hypothetical scenario for a concept that there may not be enough Italian Restaurants in Toronto Area. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Italian restaurant in Toronto, Canada.

Since there are lots of restaurants in Toronto we will try to detect ****locations that are not already crowded with restaurants****. We are also particularly interested in ****areas with no or almost Italian restaurants in vicinity****.

We will use our data science powers to generate a few most promising neighbourhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

- We want to map the area with information from Wikipedia + Geospatial_Coordinates.csv
- We want to map the number of restaurants in Toronto, this data we will get from Foursquare API

Methodology

First, I need to get the list of neighbourhoods in Toronto, Canada. This is possible by extracting the list of neighbourhoods from Wikipedia:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M I did the web scraping by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a webpage into the data frame.

However, it is only a list of neighbourhood's names and postal codes. I need to get their coordinates to utilize Foursquare to pull the list of venues near these neighbourhoods. To get the coordinates, I used Geocoder Package (Geospatial_Coordinates.csv).

Match the coordinates of Toronto neighbourhoods. After gathering these coordinates, I visualize the map of Toronto using Folium package to verify whether these are correct coordinates. Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I am pulled the names, categories, latitude, and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues.

Then, I analyse each neighbourhood by grouping the rows by neighbourhoods and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for "Italian restaurants". 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. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighbourhoods in Toronto into 3 clusters based on their frequency of occurrence for "Italian food". Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

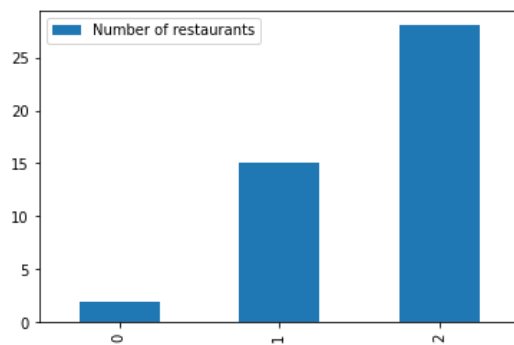
Results

Clusters

With the K-mean we determined 3 clusters to see how many restaurants are in each cluster.

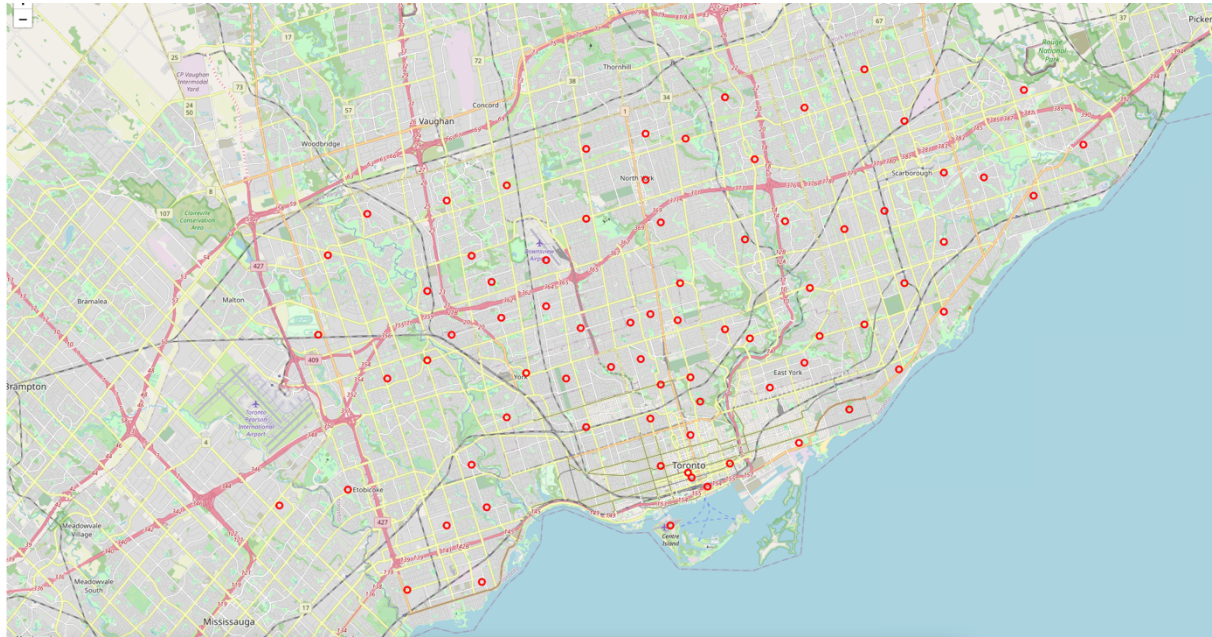
Below you can find the results of the number of restaurants in each cluster.

Segment	Number of Italian Restaurants
0	2
1	15
2	28



We can see that segment 0 has the least number of restaurants. Therefore, we want to be in this area.

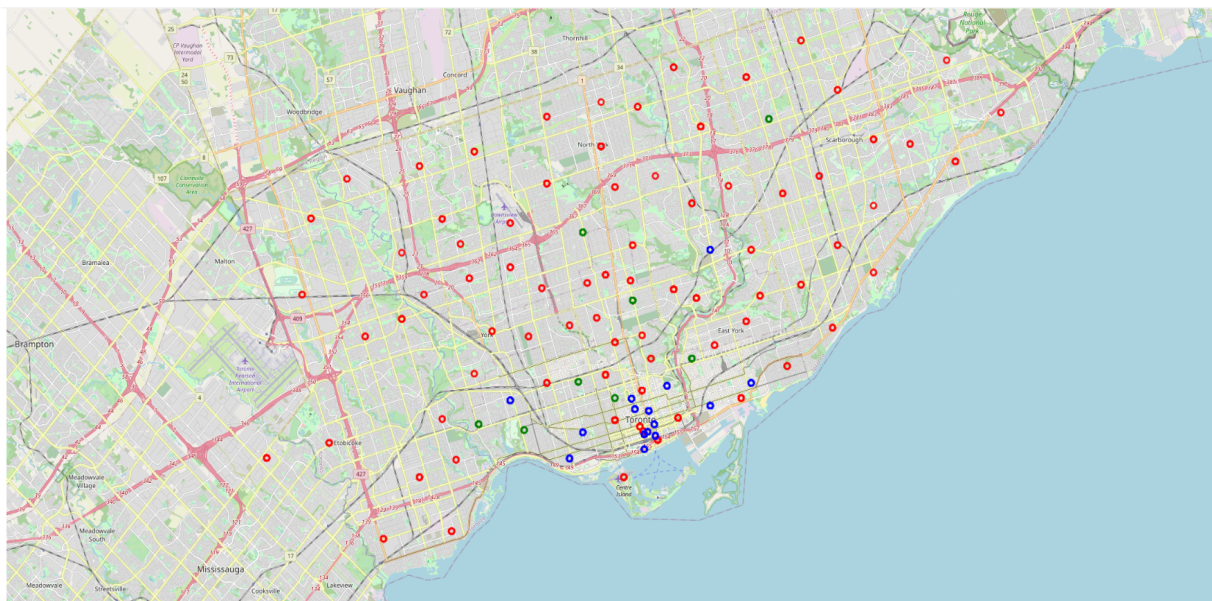
Below you can find how area 0 looks like



Recommendations

Most of the Italian restaurants are in cluster 1 lowest in Cluster 0. Looking at nearby venues it seems cluster 1 might be a good location as there are not a lot of Italian restaurants in these areas. We therefore recommend the entrepreneur to open an authentic Italian restaurant in these locations. If we look to the total map of all the area's

Area 0 = Red
 Area 1 = Blue
 Area 2 = Green



We might want to explore the areas close to the blue and green areas first because there are likely to be more downtown.

Appendix 1 List overview Area's

Agincourt

Berczy Park

CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South
Niagara, Island airport
Birch Cliff, Cliffside West

Bathurst Manor, Wilson Heights, Downsview North

Alderwood, Long Branch

Downsview Central

Downsview East

Davisville North

First Canadian Place, Underground city

Bayview Village

Cliffside, Cliffcrest, Scarborough Village West

Clairville, Humberwood, Woodbine Downs, West Humber, Kipling Heights, Rexdale, El
ms, Tandridge, Old Rexdale

Church and Wellesley

Humber Summit

Del Ray, Mount Dennis, Keelsdale and Silverthorn

Runnymede, The Junction North

The Kingsway, Montgomery Road, Old Mill North

Humberlea, Emery

Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens

Hillcrest Village

Dorset Park, Wexford Heights, Scarborough Town Centre

Don Mills North

Cedarbrae

Caledonia-Fairbanks

Fairview, Henry Farm, Oriole

Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood

Enclave of M4L

Downsview West

Dufferin, Dovercourt Village

York Mills West

Enclave of L4W

Guildwood, Morningside, West Hill

Golden Mile, Clairlea, Oakridge

Glencairn

Forest Hill North & West

Downsview Northwest

Scarborough Village

Richmond, Adelaide, King

Leaside

Lawrence Park

Lawrence Manor, Lawrence Heights

Moore Park, Summerhill East

Milliken, Agincourt North, Steeles East, L'Amoreaux East

Malvern, Rouge

Regent Park, Harbourfront

Parkwoods

Parkview Hill, Woodbine Gardens

Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The Queensway East, Royal York South East, Kingsway Park South East

Northwood Park, York University

New Toronto, Mimico South, Humber Bay Shores

North Park, Maple Leaf Park, Upwood Park

North Toronto West

Woodbine Heights

Roselawn

Thorncliffe Park

The Danforth East

Victoria Village

Willowdale South

Wexford, Maryvale

Weston

Westmount

The Beaches

Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park

Steeles West, L'Amoreaux West

South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumont Heights
, Thistletown, Albion Gardens

The Annex, North Midtown, Yorkville

Kensington Market, Chinatown, Grange Park

Kennedy Park, Ionview, East Birchmount Park

Woburn

Willowdale, Newtonbrook

Willowdale West

Rouge Hill, Port Union, Highland Creek

Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West

Rosedale

Humewood-Cedarvale