

# The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

## Project Report

Xiaoyu

# Table of Contents

<b>INTRODUCTION: BUSINESS PROBLEM .....</b>	<b>3</b>
<b>DATA.....</b>	<b>3</b>
<b>METHODOLOGY.....</b>	<b>4</b>
<b>RESULT .....</b>	<b>8</b>
<b>DISCUSSION.....</b>	<b>8</b>
<b>CONCLUSION .....</b>	<b>9</b>

## Introduction: Business Problem

As increasing numbers of consumers want to dine out or take prepared food home, the number of food-service operations has skyrocketed today. But there's still room in the market for your food-service business. Choosing location is vitally important to start a restaurant.

This project is trying to find an optimal location for a restaurant. In particular, this project will be targeted to stakeholders interested in opening a **Chinese 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 Chinese restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods 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

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of Chinese restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using geopy library
- number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- coordinate of Toronto center will be obtained using geopy library

## Methodology

Firstly, we find the latitude and longitude of the center of Toronto City using the geopy library.

The geographical coordinate of center of Toronto are 43.6534817, -79.3839347.

Secondly, we explore the neighborhood in the city of Toronto, Canada. For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore the neighborhoods in Toronto. We scrape the Wikipedia 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) and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format. We clean up the neighborhood data by removing all the unknown field and group by the 'Postal Code'. There are total 103 neighborhoods in our dataset.

Thirdly, we get the geographical data that containing the latitude and longitude from [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data). We join two pandas dataframes and create the table below:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

103 rows x 5 columns

Table 1 Geospatial Data of Neighborhoods in Toronto, Canada

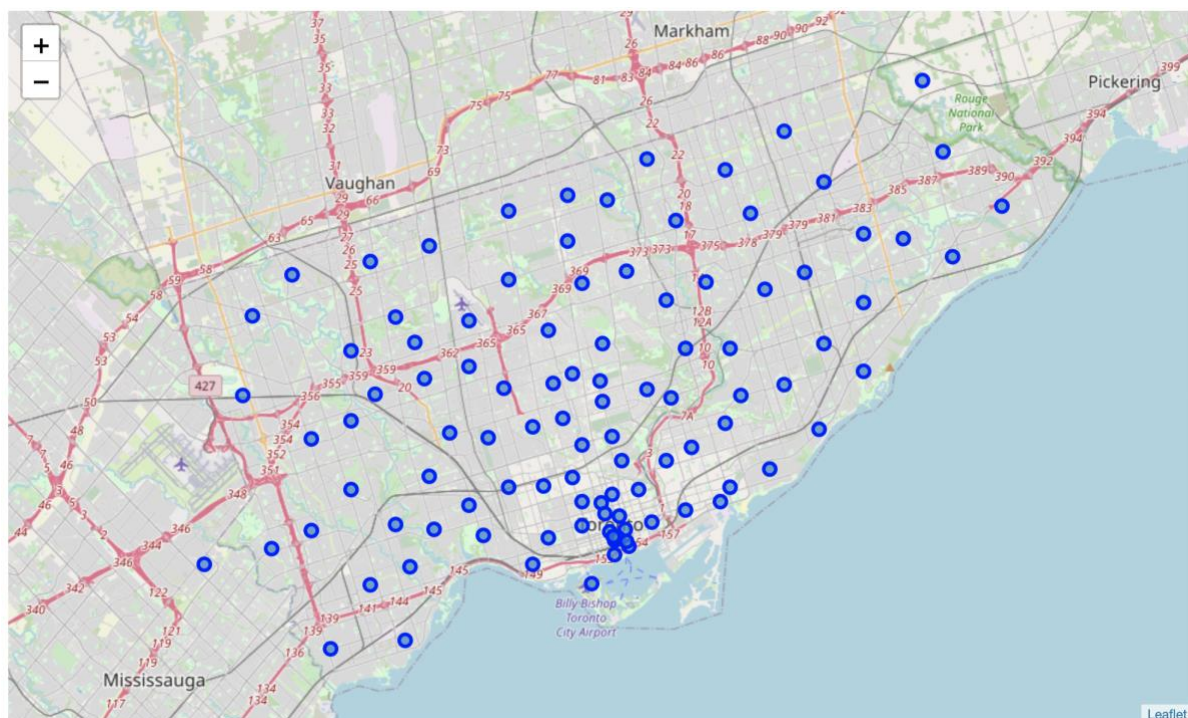


Figure 1 Neighbourhoods in Toronto, Canada

Fourthly, we calculate the distance from the neighbourhoods to the center using geopy library and append to our dataframe. (The distance is in km.)

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Distances_from_Centers
0	M3A	North York	Parkwoods	43.753259	-79.329656	6.400372
1	M4A	North York	Victoria Village	43.725882	-79.315572	7.778012
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2.601491
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	9.122914
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.646674
...	...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	13.734695
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	0.268979
100	M7Y	East Toronto	Business reply mail Processing Centre	43.662744	-79.321558	6.967320
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	12.797783
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	15.312405

103 rows × 6 columns

Table 2 Geospatial Data (including distances from center) of Neighborhoods in Toronto, Canada

Fifthly, we will use Foursquare API to get info on restaurants in each neighborhood. We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Chinese restaurant' category, as we need info on Chinese restaurants in the neighborhood. Then we calculate the total number of restaurants and Chinese restaurants in the neighborhood, and we append two new columns to our dataframe.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Distances_from_Centers	Chinese Restaurant	Total
0	M3A	North York	Parkwoods	43.753259	-79.329656	6.400372	0.0	0.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	7.778012	0.0	2.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2.601491	0.0	5.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	9.122914	0.0	1.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.646674	0.0	6.0
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	16.562078	NaN	NaN
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	21.405203	0.0	1.0
7	M3B	North York	Don Mills	43.745906	-79.352188	4.023882	1.0	10.0
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	8.334155	0.0	1.0
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0.563126	1.0	24.0
10	M6B	North York	Glencairn	43.709577	-79.445073	6.922737	0.0	3.0
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	19.069812	NaN	NaN
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	25.096198	0.0	0.0
13	M3C	North York	Don Mills	43.725900	-79.340923	5.029201	1.0	10.0
14	M4C	East York	Woodbine Heights	43.695344	-79.318389	7.369382	0.0	0.0
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0.951829	0.0	22.0
16	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	5.010316	0.0	0.0
17	M9C	Etobicoke	Eringate, Bloordeale Gardens, Old Burnhamthorpe...	43.643515	-79.577201	21.580345	0.0	0.0
18	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	21.917138	0.0	1.0
19	M4E	East Toronto	The Beaches	43.676357	-79.293031	10.160860	0.0	0.0

Table 3 Geospatial Data (including distance from center, total number of Chinese restaurants and restaurants) of Neighborhoods in Toronto, Canada



There are total 514 restaurants in the city of Toronto, of which 15 are Chinese restaurants. The percentage of Chinese restaurants is 2.92%.

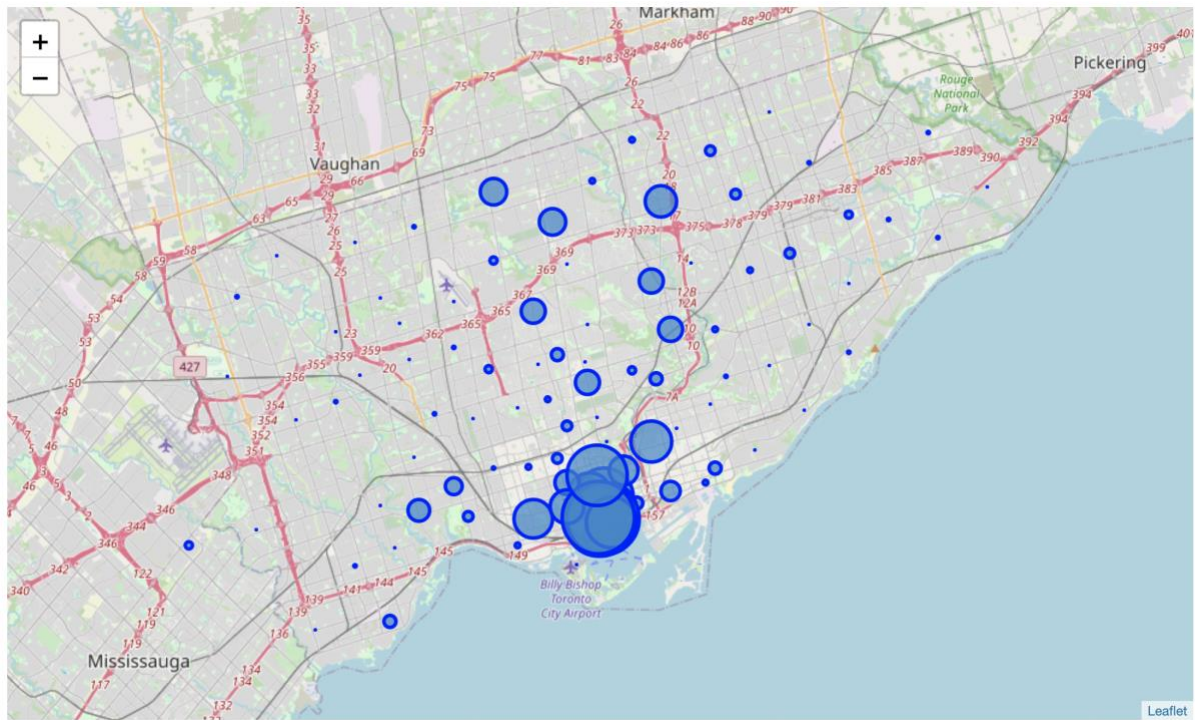


Figure 2 Number of restaurants in the neighborhood of Toronto, Canada

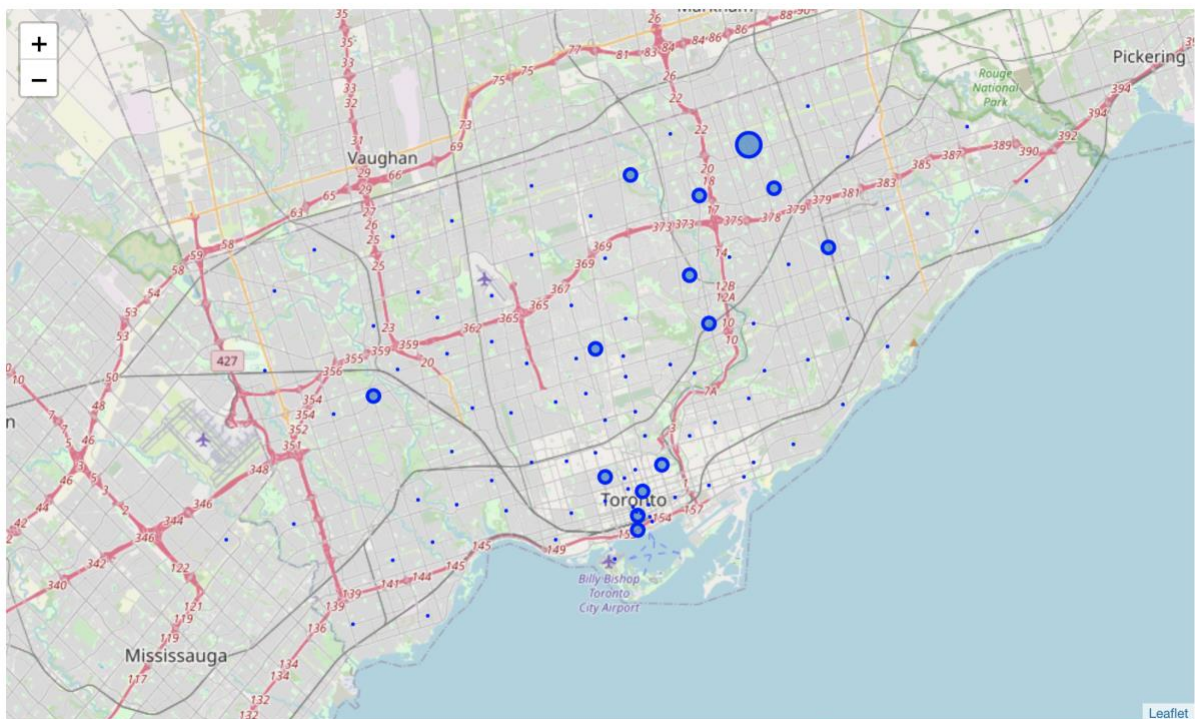


Figure 3 Number of Chinese restaurants in the neighborhood of Toronto, Canada

Finally, we sort our dataframe by distance from center of Toronto and remove the neighborhood that contains Chinese restaurants. We pick up 10 neighborhood candidates for consideration.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Distances_from_Centers	Chinese Restaurant	Total
30	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	0.092617	0.0	24.0
97	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	0.211978	0.0	30.0
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	0.268979	0.0	25.0
24	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0.395811	0.0	18.0
48	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	0.472455	0.0	30.0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.646674	0.0	6.0
83	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	0.747643	0.0	0.0
91	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	0.894210	0.0	0.0
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0.951829	0.0	22.0
92	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846	1.025117	0.0	21.0

Table 4 Neighborhood of Toronto DataFrame

## Result

From the dataframe, We can see that **Moore Park, Summerhill East** is nearest to the center of Toronto with least restaurants. **Rosedale** can also be taken into consideration since it is also not far away from center. **Queen's Park, Ontario Provincial Government** is also a good choice since it's near to the center and with relatively less restaurants in the neighborhood.

## Discussion

As we can see from the dataframe, there are less restaurants opening in Moore Park, Summerhill East, Rosedale and Queen's Park, Ontario Provincial Government area. There may be some reasons for people not opening restaurants in these area. Further investigation in these area may be needed to find the optimal solution. The suggested solution is only based on data analysis that may not take actual conditions into consideration.



## Conclusion

In this project, we have proposed some optimal solutions to open a Chinese restaurant in Toronto, Canada based on geospatial data analysis and data visualization. We choose the candidate neighborhood in Toronto based on distance from the center of Toronto, the number of restaurants in the neighborhood and the total number of Chinese restaurants in the neighborhood. The result can be useful for stakeholders interested in opening a Chinese restaurant in Toronto, Canada.