Choose the Best Restaurant Location in a Metropolitan Area

Zixiang Zhou

1. Introduction

Choosing the best restaurant location of a given type can be difficult in a metropolitan area where there are hundreds of neighborhoods. There are many metrics to determine whether a location is good or bad. However, in this exercise, the metric used to determine the optimal location is the number of similar restaurants nearby to minimize the competition as much as possible. Taking advantage of Foursquare API, we are able to determine the number of venues by searching for keywords in an API call such as "noodles", "sushi", "Italian" and etc. by setting a coordinate as the center of a circle as the search area with a configurable radius value. In a metropolitan area, there are many boroughs and under each borough, there are multiple neighborhoods. Due to the time and the number of API Calls constraints, the granularity applied to this analysis is at neighborhood level. Therefore, the question to answer in this exercise is which neighborhood is the best to open a new [Restaurant Style*] restaurant. (*Note: Restaurant style can be changed to explore more options.)

Borough	Neighborhood	Latitude	Longitude	Num_of_Noodle_Shop
Port Hope	Port Hope	43.951575	-78.293970	0
Port Colborne	Port Colborne	42.886239	-79.251390	0
Grimsby	Grimsby	43.193209	-79.560692	0
Barrie	North,East	44.389311	-79.690174	0
Barrie	South,West	44.389311	-79.690174	0
Keswick	Keswick	44.239617	-79.468656	0
Midland	Midland	44.750147	-79.885712	0
Whitby	Central	43.879820	-78.942175	0
Welland	West	42.992218	-79.248419	0

In this exercise, we will look into all neighborhoods outside Downtown Toronto but in the Greater Toronto Area (GTA) to find candidate location(s) to open a new noodle shop.

2. Data acquisition and cleaning

The 1st dataset to be used for this analysis shall contain information about boroughs and neighborhoods under each borough in GTA outside downtown Toronto. To easily and automatically get the data, web scrapping can be used to take available public information from the Wikipedia page and store the return as a JSON file and then transform it to a Panda dataframe for the further data processing. Wikipedia Page to scrap the RAW data from: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_L.

RAW Data looks like:

```
[                     0  \
 0             L1APort Hope
 1       L1BBowmanville(East)
 2       L1CBowmanville(West)
 3   L1ECourtice(Bowmanville)
 4         L1GOshawa(Central)
 5       L1HOshawa(Southeast)
 6       L1JOshawa(Southwest)
 7            L1KOshawa(East)
 8           L1LOshawa(North)
 9           L1MWhitby(North)
 10      L1NWhitby(Southeast)
 11      L1PWhitby(Southwest)
 12        L1RWhitby(Central)
 13        L1SAjax(Southwest)
 14        L1TAjax(Northwest)
 15   L1VPickering(Southwest)
 16       L1WPickering(South)
 17     L1XPickering(Central)
 18       L1YPickering(North)
 19             L1ZAjax(East)
```

Figure 1- RAW Data from Web

The process of cleaning the RAW data is described in the Jupyter notebook.

After cleaning, the dataset looks like the following:

	PostalCode	Borough	Neighborhood
0	L1A	Port Hope	Port Hope
1	L1B	Bowmanville	East
2	L1C	Bowmanville	West
3	L1E	Courtice	Bowmanville
4	L1G	Oshawa	Central
5	L1H	Oshawa	Southeast

Figure 2- Cleaned Neighborhood Dataset

The 2nd dataset to append to the 1st one is the geographical location in terms of coordinates (latitude, longitude) for each neighborhood. To get coordinates efficiently, one can utilize the geocoder to convert “Neighborhood, Province” string to a tuple of latitude and longitude via making geocoder API calls.

After cleaning, the dataset looks like the following:

```
neighborhoods.head(5)
```

	Borough	Neighborhood	Latitude	Longitude
0	Port Hope	Port Hope	43.951575	-78.293970
1	Bowmanville	East	43.912300	-78.689167
2	Bowmanville	West	43.912300	-78.689167
3	Courtice	Bowmanville	43.904861	-78.788314
4	Oshawa	Central	43.897556	-78.863532

Figure 3 - Cleaned Data with Coordinates

What is essential to the analysis in this exercise is finding the number of venues around each coordinate presented in the 2nd dataset. Foursquare API offers developers a way to efficiently making calls to search for venues around a geographical location with a configurable search range.

The URL used to send the JSON request is as follows:

url='https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, ***latitude****,* ***longitude****, ACCESS_TOKEN, VERSION,* ***search_query****,* ***radius****, LIMIT)*

results = requests.get(url).json()

search_query = “Noodle”

radius = 15000

Due to the limited number of regular Foursquare call can be made today, sometimes the following error shows up:

```
<ipython-input-58-332b82af6324>:25: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize inste
ad
  dataframe = json_normalize(venues)

---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-58-332b82af6324> in <module>
     20
     21     # assign relevant part of JSON to venues, if an error reported here, it means daily quota exceeded!
---> 22     venues = results['response']['venues']
     23
     24     # tranform venues into a dataframe

KeyError: 'venues'
```

Figure 4 - Daily Quota Exceeded

In order to alleviate the impact to our analysis due to this limitation, we can sort the entries in ascending order in terms of the number of nearby noodle shops (within 15km):

I only picked top 3 as follows:

```
In [63]: neigborhoods_noodle_shops_sorted_top3 = neigborhoods_noodle_shops_sorted.head(3)

In [64]: neigborhoods_noodle_shops_sorted_top3
```

Out[64]:

	Borough	Neighborhood	Latitude	Longitude	Num_of_Noodle_Shop
0	Port Hope	Port Hope	43.951575	-78.293970	0
35	Port Colborne	Port Colborne	42.886239	-79.251390	0
36	Grimsby	Grimsby	43.193209	-79.560692	0

Figure 5- Top 3 Candidates

3. Data Analysis

From the previous section, we know Port Hope, Port Colborne and Grimsby are areas where there is no exiting noodle shop.

A. Port Hop, Population: 16,753 and Population Density: 993.6/km^2
B. Port Colborne, Population: 18,306 and Population Density: 150.1/km^2
C. Grimsby, Population: 27,314 and Population Density: 396.3/km^2

The best candidate to open a new noodle shop is Port Hop, Ontario.

4. Visualization

Use the Folium library to visualize these three locations on the map:

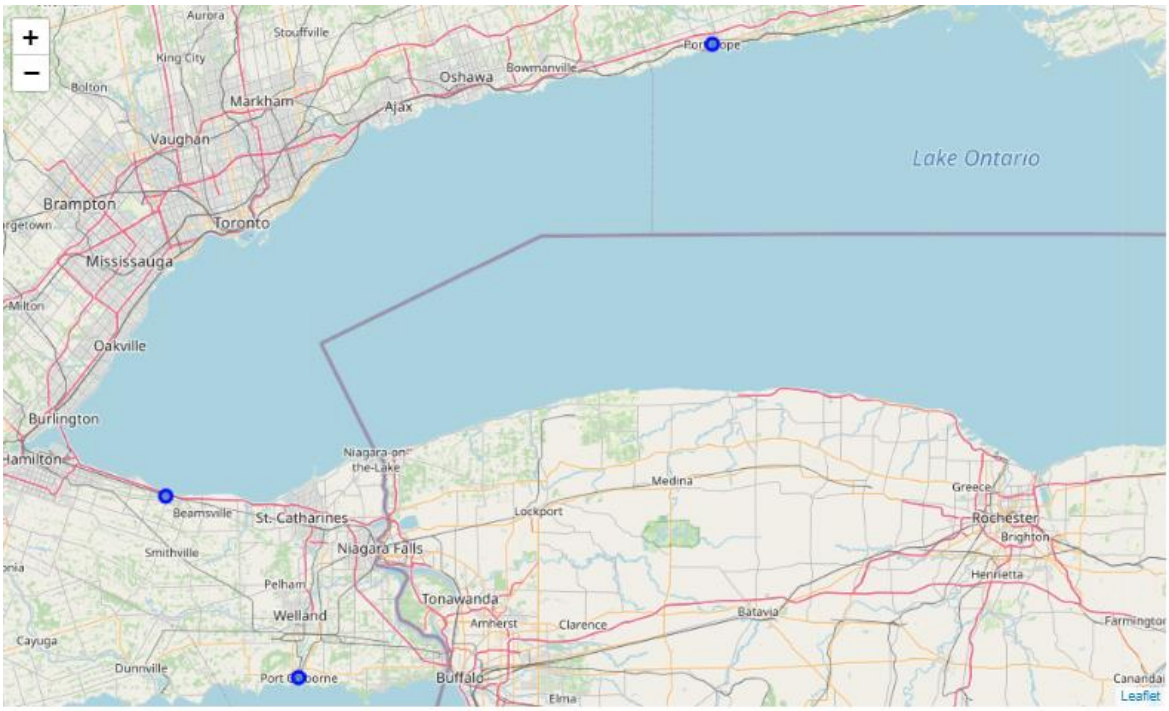

Figure 6- Visualization on Locations

5. Recommendation and Further Analysis

Actually, there are many other neighbors that do not have existing noodle shops or just have a few. The further analysis can take advantage of web scrapping to get population, population density and even ethnic group ratio data for each neighborhood to make a more insightful decision.

Borough	Neighborhood	Latitude	Longitude	Num_of_Noodle_Shop
Port Hope	Port Hope	43.951575	-78.293970	0
Port Colborne	Port Colborne	42.886239	-79.251390	0
Grimsby	Grimsby	43.193209	-79.560692	0
Barrie	North,East	44.389311	-79.690174	0
Barrie	South,West	44.389311	-79.690174	0
Keswick	Keswick	44.239617	-79.468656	0
Midland	Midland	44.750147	-79.885712	0
Whitby	Central	43.879820	-78.942175	0
Welland	West	42.992218	-79.248419	0