Restaurant business recommendation in Toronto

1 Introduction

The idea of this project is to provide insights to investors who are interested in opening a new restaurant or purchase an existing restaurant in one of Toronto's suburbs. Traditionally, the investors would usually determine where to open or acquire a food business based on their own experience and judgement. However, with the support of this project, people can determine the best suburbs to set up the restaurant and potentially the best location and food types served in the restaurant. The result of the project can be used to inform the decision-making process of the investment, which would be data-driven.

2 Data Source

The data source of this project can be downloaded from the server of Foursquare and Wikipedia. First of all, we will download Toronto's neighbourhood and borough lists from Wikipedia. Then, the neighbourhood's latitude and longitude can be determined by using Python Geocoder.

Table 2.1. Toronto neighbourhood data

	Postal Code	Neighbourhood	Borough	Latitude	Longitude
0	M1B	Malvern, Rouge	Scarborough	43.8066863	-79.194353
1	M1C	Rouge Hill, Port Union, Highland Creek	Scarborough	43.7845351	-79.160497
2	M1E	Guildwood, Morningside, West Hill	Scarborough	43.7635726	-79.188712
3	M1G	Woburn	Scarborough	43.7709921	-79.216917
4	M1H	Cedarbrae	Scarborough	43.773136	-79.239476
5	M1J	Scarborough Village	Scarborough	43.7447342	-79.239476
6	M1K	Kennedy Park, Ionview, East Birchmount Park	Scarborough	43.7279292	-79.262029
7	M1L	Golden Mile, Clairlea, Oakridge	Scarborough	43.7111117	-79.284577
8	M1M	Cliffside, Cliffcrest, Scarborough Village West	Scarborough	43.716316	-79.239476
9	M1N	Birch Cliff, Cliffside West	Scarborough	43.692657	-79.264848
...					

For each neighbourhood, the venues which specifically corresponds to food category can be downloaded. Most importantly, the type of the food provided in each restaurant (or say venue) can be collected from this data, which would enable us to determine which food type is most popular in this suburb.

In addition, once we decide the suburb to set up the restaurant, we can also use clustering technique to further separate the suburb to different blocks, which would guide us to determine the location of the restaurant in the suburb. The selection of the block can be supported by the local renting cost, etc.

3 Methodology

1. First of all, the coordinates of existing suburbs can be collected based on their post codes and names. The package used to determine the coordinates is Python Geocoder. The detail of the coordinates is shown in Table 2.1.
2. Then, the "explore" API provided by Foursquare is used to search food venues in every suburb. In this search process, we need to define the search radius for the venue. We have considered that the land size of Toronto is about 630 km^2 and its longest distance from Eastern to Western border is about 48 km, while the distance between the Southern and Northern border is 21 km. Since we have considered about 100 suburbs here, the search radius needs to be able to principally cover most of the Toronto region. In this case, we have set the radius to 2km, which would lead to some overlapping of different search regions, and therefore same venues may be allocated to different suburbs.
3. After deriving the venue list, there are some duplicates of each venue in the list, which needs to be removed. We will calculate the Euclidean distance between the venue and its corresponding suburbs. The profile of the venue with its closest suburb will be kept.
4. Subsequently, the unique venue list of every suburb can be derived. We can firstly explore the relationship between the number of venue and the number of venue categories in each suburb.
5. Finally, the top venue types in the selected suburbs can be further analysed. For instance, we will analyse 10 suburbs with the highest number of food venues, assuming that the investor wants to set up a restaurant in these suburbs. Then, the most common category of restaurants in these suburbs can be displayed, while these suburbs can be clustered in order to allow the investor can target some similar regions with a given restaurant type in mind.

4 Results and discussion

The relationship between the number of venues and the number of venue categories is shown in Figure 1. We notice that the diversity of the venue categories increases with the number of venues.

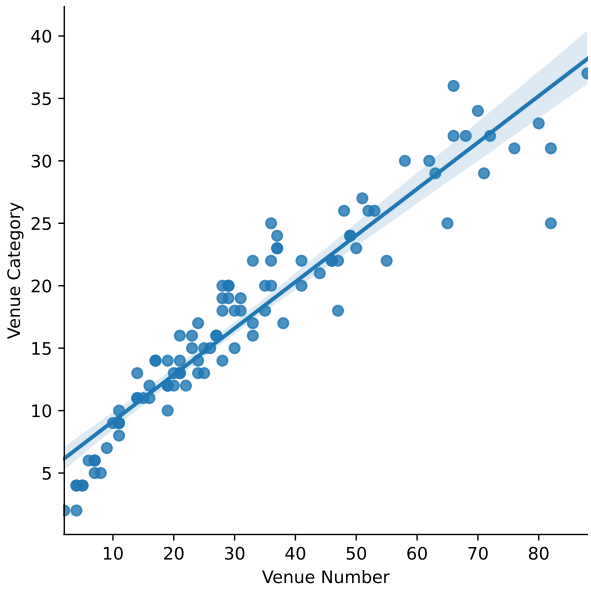

Figure 1. Regression plot of venue number and venue category in all suburbs

Then, 10 suburbs with the highest number of venues have been selected here to further explore the potential business opportunity. The profiles of these suburbs are shown in Table 4.1.

Table 4.1. Top ten suburbs with the highest number of venues in Toronto

Postal Code	Neighbourhood	Borough	Latitude	Longitude
M1S	Agincourt	Scarborough	43.7942003	-79.262029
M1V	Milliken, Agincourt North, Steeles East, L'Amoreaux East	Scarborough	43.8152522	-79.284577
M2M	Willowdale, Newtonbrook	North York	43.789053	-79.408493
M3B	Don Mills	North York	43.7459058	-79.352188
M3C	Don Mills	North York	43.7258997	-79.340923
M4E	The Beaches	East Toronto	43.6763574	-79.293031
M4M	Studio District	East Toronto	43.6595255	-79.340923
M6C	Humewood-Cedarvale	York	43.6937813	-79.428191
M7R	Canada Post Gateway Processing Centre	Mississauga	43.6369656	-79.615819
M8W	Alderwood, Long Branch	Etobicoke	43.6024137	-79.543484
M8Z	Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West	Etobicoke	43.6288408	-79.520999

First of all, we will convert the venue number of each category as a frequency to reflect the suburb's food culture. Then, by using K means method, we can cluster these suburbs into different groups. The number of clusters has been set to three here. Based on the result, we can see that:

- Category 1: In Scaborough region, the Chinese restaurants are the most common ones.
- Category 2: In Willowdale and Newtonbrook region, Korean and Middle Eastern restaurants are the most common ones.
- Category 3: In other seven regions, pizza place sounds like a more common business.

Table 4.2, features of restaurants in top 10 suburbs

Neighbourhood	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Agincourt	Scarborough	2	Chinese Restaurant	Restaurant	Bakery	Pizza Place
Milliken, …	Scarborough	2	Chinese Restaurant	BBQ Joint	Japanese Restaurant	Korean Restaurant
Willowdale, Newtonbrook	North York	1	Korean Restaurant	Middle Eastern Restaurant	Restaurant	Cafe
Don Mills	North York	0	Japanese Restaurant	Pizza Place	Restaurant	Cafe
Don Mills	North York	0	Japanese Restaurant	Pizza Place	Restaurant	Cafe
The Beaches	East Toronto	0	Pizza Place	Sandwich Place	Bakery	Breakfast Spot
Studio District	East Toronto	0	Bakery	Vietnamese Restaurant	Cafe	Diner
Humewood-Cedarvale	York	0	Pizza Place	Restaurant	Italian Restaurant	Sushi Restaurant
Canada Post Gateway Processing Centre	Mississauga	0	Restaurant	Middle Eastern Restaurant	Sandwich Place	Chinese Restaurant
Alderwood, Long Branch	Etobicoke	0	Pizza Place	Restaurant	Fast Food Restaurant	Sandwich Place

Mimico NW, …	Etobicoke	0	Restaurant	Sandwich Place	Fast Food Restaurant	Sushi Restaurant

5 Conclusion

The program developed in this project can be used to help investors to understand the feature of food venue business in different regions of Toronto. By using the venue data provided by Foursquare, we can find the top suburbs where food venues are popular. The program can also be used to explore the categories of existing food venue in a given suburb, which would guide the business decisions in terms of whether the investor should target a over-crowded but popular market, or should set up a restaurant venue which has never existed before in this suburb.