FINAL PROJECT

Toronto Japanese restaurant Location Selection

Author: Yushan T.

INTRODUCTION

Many decisions in today's society are driven by geographic data, so we have more opportunities to understand how location intelligence can enhance the analysis of businesses, governments, utilities, and other organizations or audiences. In the current complex business environment, location selection is particularly important for catering companies because the location is the direct entrance to offline traffic. The purpose of this project is to use machine learning and data analysis to find the most suitable location to open a Japanese restaurant. This analysis can help new restaurant chains in Japan to choose a location. Data analysis results can be directly used to quickly assess and effectively control the risk of location selection.

BUSSINESS PROBLEM

The number of Japanese restaurants has grown at a rate of 10% in the Toronto area. MISUYA is a very famous Japanese restaurant in Japan. Investors want to set up different restaurant chains in the Toronto area. The solution I implemented is to use the vector maps of these cities and load more detailed geographic data, and use the geographic information system (GIS) to achieve batch processing and quantitative analysis. Help to help customers MISUYA find and determine the best location in the city with commercial geographic data.

DATA SOURCE

List of postal codes of Canada: M

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Postal Code	Borough	Neighborhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson

Get all the information about Toronto's existing communities. This page contains the postcode, the names of the boroughs and all communities in Toronto.

METHODOLOGY

In order to help this company, choose the most favorable location, we took a deep approach to extracting layers, and conducted in-depth analysis from the locations of various areas in Toronto.

Get all the information about Toronto's existing communities. This page contains the postcode, the names of the boroughs and all communities in Toronto.

[3]:
```python
import lxml
```

[4]:
```python
df = pd.read_html('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M')[0]
```

[5]:
```python
df.head()
```

[5]:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

[6]:
```python
df_dropna = df[df.Borough != "Not assigned"].reset_index(drop=True)
df_dropna.head()
```

[6]:

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront

In order to get the location and other information about various places in Toronto, I am using Foursquare's Explore API. By using Foursquare's Explore API, we obtained detailed information about Toronto's existing venues and collected their names, categories, and locations (longitude and latitude).

```python
venues_df.head()
```

(1620, 9)

[30]:

	PostalCode	Borough	Neighborhood	BoroughLatitude	BoroughLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	M4E	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	M4E	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop

I use Folium to draw interactive leaflet maps using coordinate data. View various regions and we need to verify that these coordinates are correct.

[22]:

Next, I used the Foursquare API to extract a list of the top 100 locations within a radius of 500 meters.

[29]:
```
radius = 500
LIMIT = 100

venues = []
```

[31]:
```
venues_df.groupby(["PostalCode", "Borough", "Neighborhood"]).count()
```

[31]:

PostalCode	Borough	Neighborhood	BoroughLatitude	BoroughLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
M4E	East Toronto	The Beaches	4	4	4	4	4	4
M4K	East Toronto	The Danforth West, Riverdale	43	43	43	43	43	43
M4L	East Toronto	India Bazaar, The Beaches West	19	19	19	19	19	19
M4M	East Toronto	Studio District	40	40	40	40	40	40
M4N	Central Toronto	Lawrence Park	3	3	3	3	3	3
M4P	Central Toronto	Davisville North	8	8	8	8	8	8

With this data, I can also check how many unique categories I can get from these places.

```
[33]: venues_df['VenueCategory'].unique()[:50]
```

```
[33]: array(['Trail', 'Health Food Store', 'Pub', 'Neighborhood',
       'Cosmetics Shop', 'Greek Restaurant', 'Italian Restaurant',
       'Ice Cream Shop', 'Yoga Studio', 'Brewery',
       'Fruit & Vegetable Store', 'Dessert Shop', 'Restaurant',
       'Pizza Place', 'Juice Bar', 'Bookstore', 'Bubble Tea Shop',
       'Furniture / Home Store', 'Grocery Store', 'Spa', 'Coffee Shop',
       'Bakery', 'Caribbean Restaurant', 'Café', 'Indian Restaurant',
       'Japanese Restaurant', 'Lounge', 'Frozen Yogurt Shop',
       'American Restaurant', 'Liquor Store', 'Gym', 'Fish & Chips Shop',
       'Fast Food Restaurant', 'Sushi Restaurant', 'Park', 'Pet Store',
       'Steakhouse', 'Burrito Place', 'Movie Theater', 'Sandwich Place',
       'Fish Market', 'Gay Bar', 'Seafood Restaurant', 'Cheese Shop',
       'Middle Eastern Restaurant', 'Comfort Food Restaurant',
       'Stationery Store', 'Wine Bar', 'Thai Restaurant',
       'Coworking Space'], dtype=object)
```

I check if the japanese restaurant exist in the venue category or not.

```
[34]: "Japanese Restaurant" in venues_df['VenueCategory'].unique()
```

```
[34]: True
```

Then, I analyzed each neighbor by grouping rows by neighbors and averaging the frequency of occurrence for each category of places. This is to prepare the cluster for later.

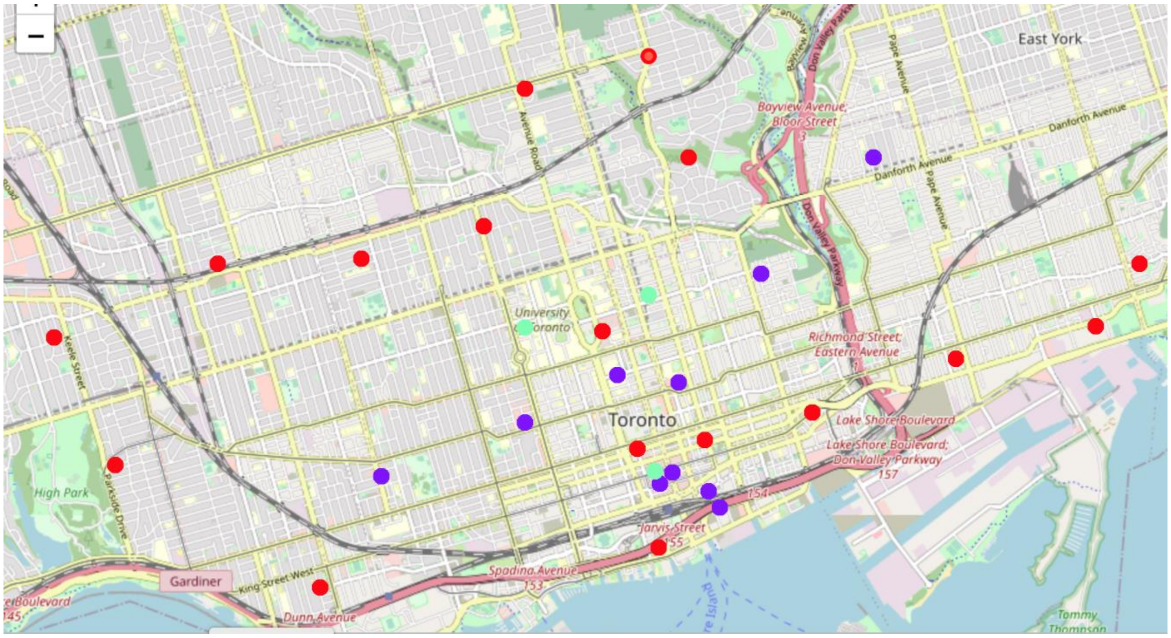
	PostalCode	Borough	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Mu
0	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	0	0	0	0	
1	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	0	0	0	0	
2	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	0	0	0	0	
3	M4E	East Toronto	The Beaches	0	0	0	0	0	0	0	0	0	0	0	
4	M4K	East Toronto	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	0	0	

	PostalCode	Borough	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop
0	M4E	East Toronto	The Beaches	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	M4K	East Toronto	The Danforth West, Riverdale	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023256	0.000000
2	M4L	East Toronto	India Bazaar, The Beaches West	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	M4M	East Toronto	Studio District	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.050000	0.000000

I perform the clustering method by using k-means clustering. The idea of the K-Means algorithm is very simple. For a given sample set, the sample set is divided into K clusters according to the distance between the samples. Make the points in the clusters as close as possible, and make the distance between the clusters as large as possible. It is relatively simple to implement, and the clustering effect is also good, so it is widely used.

I divided the communities in Toronto into 3 categories based on the frequency of Japanese restaurants. Based on the results I will be able to recommend the ideal location to open the restaurant.

The result of k-means clustering shows that we can divide Toronto Street into 3 clusters according to how many Japanese restaurants are in each block:



- *Category 0: A community with a large number of Japanese restaurants.*
- *Category 1: There are not many Japanese restaurants in the community.*
- *Category 2: Community with few Japanese restaurants*

The results are visualized in the figure above, cluster 0 is red, cluster 1 is purple, and cluster 2 is green.

CONCLUSION

Most Japanese restaurants are located in Type 0 and Type 1 areas near Richmond, Adelaide, King, st. James town and Toronto dominion centre, and the lowest in Type 2 areas in the Church and Wellesley area. In addition, there is also a good opportunity to open near UofT. Looking at the nearby venues, it seems that Group 2 may be a good location because there are not many Japans in these areas. Therefore, the project recommends MISUYA to open chain restaurant restaurants in these locations.