

Battle of the Neighbourhoods

1. Introduction

1.1 Business Problem

Rasa is a well-known restaurant brand from Malaysia and has recently expanded its business to Toronto, Canada. What sets Rasa apart from every other ethnic restaurant is that the dishes served at Rasa are as colourful as Malaysia's diverse culture. At Rasa, customers can not only enjoy Malay cuisines, but Chinese and Indian as well. Due to the huge success of its first international branch in Toronto, the owners of Rasa have decided to expand their business in New York, a city known for its diverse population.

1.2 Target Audience

The next step for the owners now is to identify a suitable location in one of New York City's neighbourhood to open up the restaurant. Besides being useful to the owners, this project can also assist other business owners looking to expand their businesses anywhere around New York city.

2. Data

2.1 New York Data

The first step in this project is to identify a suitable dataset that contains relevant information of the neighbourhoods in New York city. Luckily for us, this data is readily available online at https://geo.nyu.edu/catalog/nyu_2451_34572.

Using this dataset, we extract features that will be used in the analysis, which are Borough, Neighbourhood, Latitude, and Longitude. The dataset consists of 5 boroughs and 306 neighbourhoods, as visualized in Figure 1 below. The first five entries of this dataframe will look like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Table 1

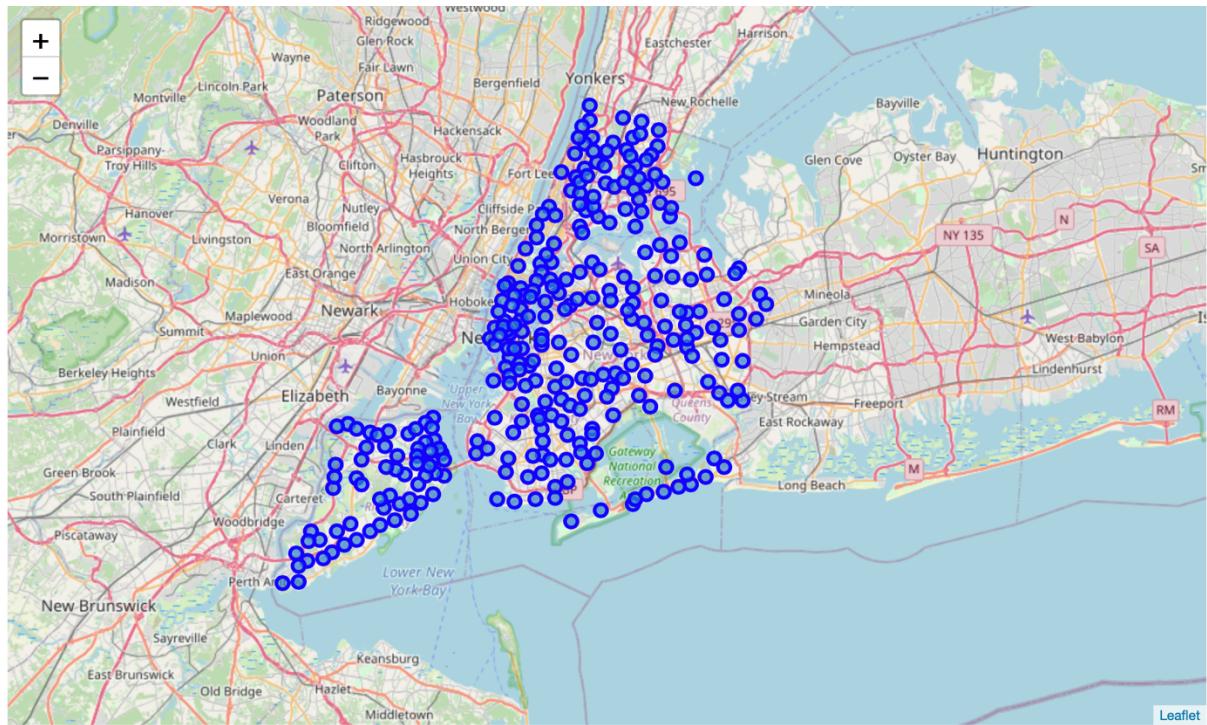


Figure 1

However, notice how there are too many neighbourhoods to analyse. To overcome this problem, we focus on only one borough instead, which is Manhattan, since that is where most of New York City's main attractions (Broadway, Central Park, and Empire State Building) are located at. By doing this, we managed to narrow down the initial dataset into 40 neighbourhoods, which is shown in Figure 2.

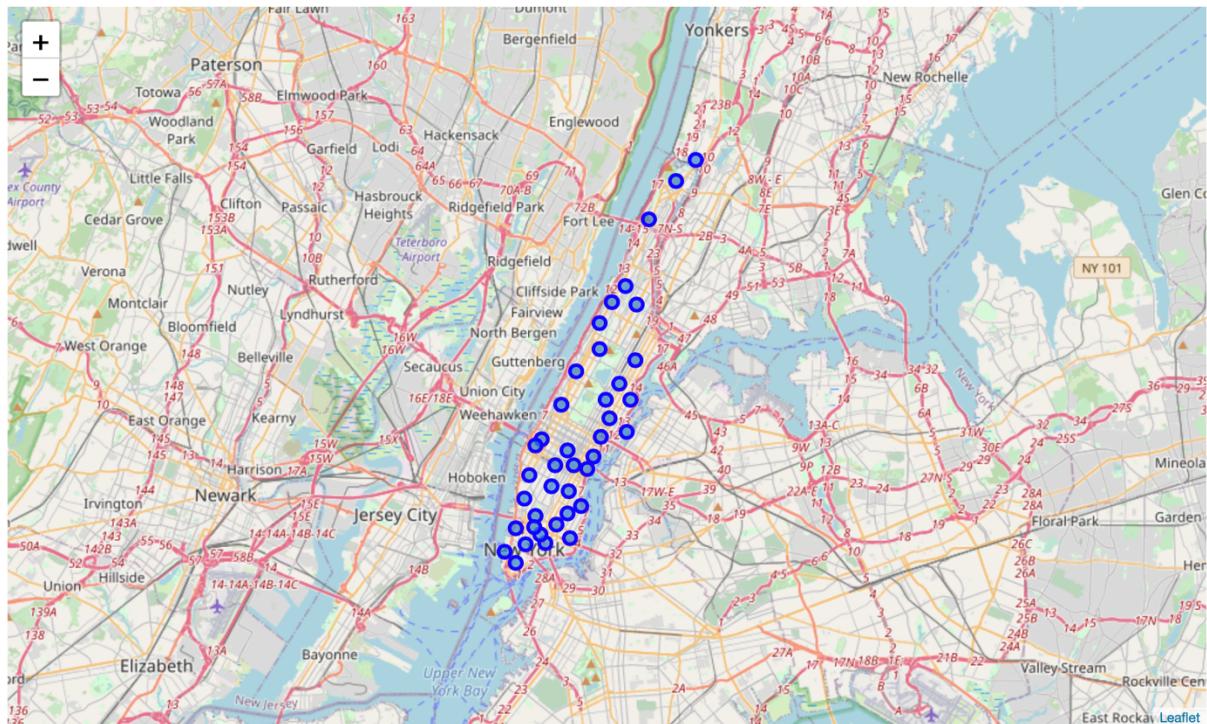


Figure 2

2.1 Foursquare API

Once we are satisfied with our final dataframe, we proceed to explore venues available in each neighbourhood. This is done to get a general assumption of what kind of venues are available nearby. To do this, we extract information from Foursquare and identify the top 100 most common venues within a radius of 500 meters in each neighbourhood. This step returned a total of 3140 venues, with the first five listed below in Table 2.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Table 2

3. Methodology

3.1 Analysing Each Neighbourhood

Since the name of venues are of no value to us, we decide to look at the venue category instead. However, looking at the venue categories individually in each neighbourhood will not provide any sort of analysis to us. Hence, we take the mean of the frequency of occurrence of each category and group rows by neighbourhood. The first 5 rows and 12 columns are shown in Table 3 below.

Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepas Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	Battery Park City	0.000000	0.00	0.00	0.000000	0.015873	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Carnegie Hill	0.000000	0.00	0.00	0.000000	0.011364	0.000000	0.000000	0.000000	0.011364	0.000000	0.011364
2	Central Harlem	0.000000	0.00	0.00	0.065217	0.043478	0.000000	0.000000	0.000000	0.000000	0.021739	0.000000
3	Chelsea	0.000000	0.00	0.00	0.000000	0.040000	0.000000	0.000000	0.000000	0.000000	0.070000	0.000000
4	Chinatown	0.000000	0.00	0.00	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 3

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
0	Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Plaza	Gourmet Shop	Burger Joint	Food Court	Shopping Mall
1	Carnegie Hill	Coffee Shop	Wine Shop	Café	Italian Restaurant	Yoga Studio	Gym	Gym / Fitness Center	Bookstore	Japanese Restaurant	Cocktail Bar
2	Central Harlem	African Restaurant	Chinese Restaurant	Gym / Fitness Center	American Restaurant	Fried Chicken Joint	Bar	French Restaurant	Seafood Restaurant	Bookstore	Boutique
3	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	American Restaurant	Café	Italian Restaurant	Bakery	French Restaurant	Market	Cocktail Bar
4	Chinatown	Chinese Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Ice Cream Shop	Spa	Bar	Cocktail Bar	Hotpot Restaurant	Optical Shop

Table 4

However, there are a total of 328 unique venue categories to look at. To further assist our analysis, we narrow it down to only the 10 most common venue categories in each neighbourhood. This gives us a dataframe that looks like Table 4.

3.2 Clustering the Neighbourhoods

Now that we have obtained the top 10 most common venue in each neighbourhood, we will use K-Means clustering to group the neighbourhoods into 5 different clusters. We append the cluster labels, latitude, and longitude for each neighbourhood into Table 4, and obtain Table 5 below.

Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	
0	Manhattan	Marble Hill	40.876551	-73.910660	4	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Diner	Miscellaneous Shop	Pizza Place	Steakhouse
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Ice Cream Shop	Spa	Bar	Cocktail Bar
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Café	Bakery	Mobile Phone Shop	Coffee Shop	New American Restaurant	Tapas Restaurant	Latin American Restaurant	Park
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant	Park	Frozen Yogurt Shop	Pizza Place
4	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Pizza Place	Coffee Shop	Café	Deli / Bodega	Mexican Restaurant	Yoga Studio	Sushi Restaurant	Bakery

Table 5

4. Results

A visualization of the 5 different clusters is shown in Figure 3 below.

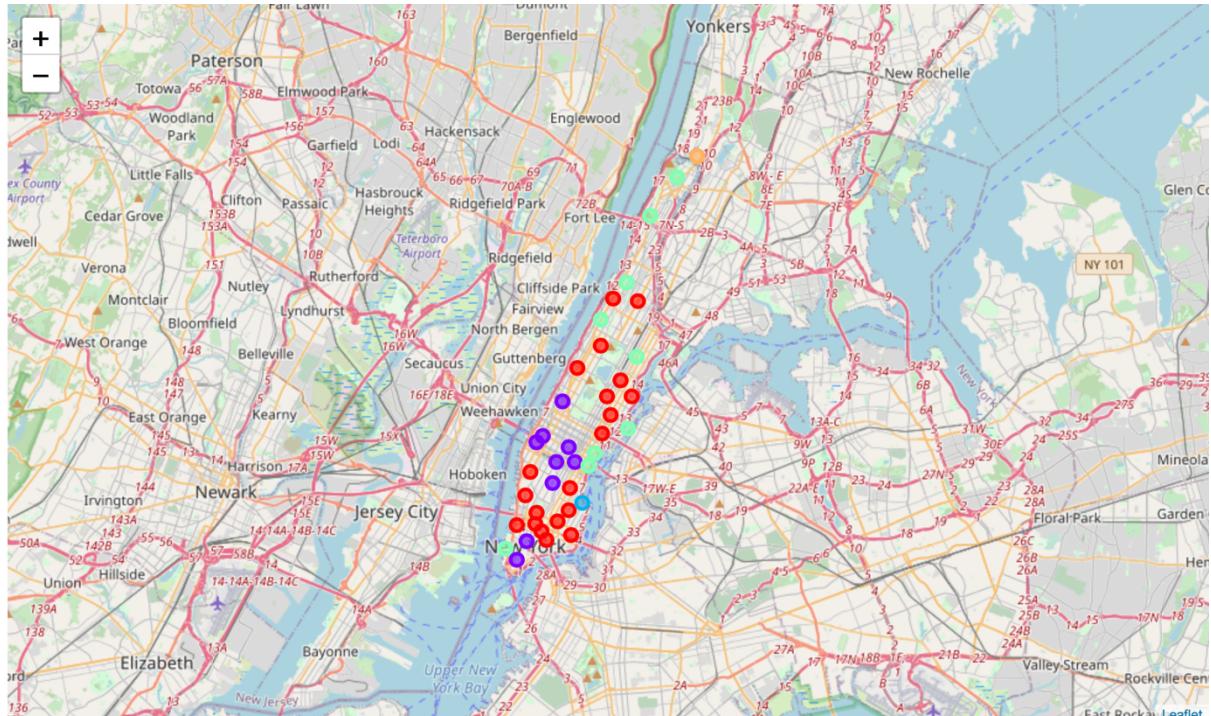


Figure 3

4.1 Cluster 1 (Red)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Chinatown	Chinese Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Ice Cream Shop	Spa	Bar	Cocktail Bar	Hotpot Restaurant	Optical Shop
5	Manhattanville	Coffee Shop	Deli / Bodega	Mexican Restaurant	Italian Restaurant	Seafood Restaurant	Indian Restaurant	Café	Liquor Store	Lounge	Sushi Restaurant
6	Central Harlem	African Restaurant	Chinese Restaurant	Gym / Fitness Center	American Restaurant	Fried Chicken Joint	Bar	French Restaurant	Seafood Restaurant	Bookstore	Boutique
8	Upper East Side	Italian Restaurant	Coffee Shop	Bakery	Juice Bar	Gym / Fitness Center	Yoga Studio	Wine Shop	Spa	French Restaurant	Sushi Restaurant
9	Yorkville	Italian Restaurant	Gym	Coffee Shop	Bar	Deli / Bodega	Sushi Restaurant	Pizza Place	Diner	Japanese Restaurant	Wine Shop
10	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Cocktail Bar	Café	Gym	Gym / Fitness Center	Burger Joint	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Indian Restaurant	Wine Bar	Bakery	Coffee Shop	Ice Cream Shop	Seafood Restaurant	Sports Bar	Pub
17	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	American Restaurant	Café	Italian Restaurant	Bakery	French Restaurant	Market	Cocktail Bar
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	Café	Indian Restaurant	Chinese Restaurant	Ice Cream Shop	Coffee Shop	Clothing Store	Sandwich Place	Pilates Studio
19	East Village	Bar	Cocktail Bar	Pizza Place	Mexican Restaurant	Coffee Shop	Korean Restaurant	Wine Bar	Dessert Shop	Ice Cream Shop	Filipino Restaurant
20	Lower East Side	Café	Coffee Shop	Chinese Restaurant	Cocktail Bar	Bakery	Park	Ramen Restaurant	Art Gallery	Sandwich Place	French Restaurant
21	Tribeca	Italian Restaurant	Park	American Restaurant	Wine Bar	Spa	Greek Restaurant	Coffee Shop	Café	Skate Park	Burger Joint
22	Little Italy	Bakery	Chinese Restaurant	Spa	Bubble Tea Shop	Italian Restaurant	Café	Sandwich Place	Thai Restaurant	Ice Cream Shop	Pizza Place
23	Soho	Italian Restaurant	Clothing Store	Boutique	Coffee Shop	Sandwich Place	Mediterranean Restaurant	French Restaurant	Bakery	Ice Cream Shop	Café
24	West Village	Italian Restaurant	American Restaurant	New American Restaurant	Wine Bar	Cocktail Bar	Park	Coffee Shop	Bakery	Jazz Club	Pizza Place
25	Manhattan Valley	Coffee Shop	Bar	Mexican Restaurant	Yoga Studio	Pizza Place	Bubble Tea Shop	Café	Peruvian Restaurant	Park	Arts & Crafts Store
27	Gramercy	Bar	Coffee Shop	Italian Restaurant	Bagel Shop	Playground	Mexican Restaurant	Cocktail Bar	American Restaurant	Ice Cream Shop	Pizza Place
30	Carnegie Hill	Coffee Shop	Wine Shop	Café	Italian Restaurant	Yoga Studio	Gym	Gym / Fitness Center	Bookstore	Japanese Restaurant	Cocktail Bar
31	Noho	Italian Restaurant	Pizza Place	Coffee Shop	Grocery Store	Mexican Restaurant	Hotel	Cocktail Bar	Japanese Restaurant	Wine Bar	American Restaurant
34	Sutton Place	Italian Restaurant	Gym	Park	Coffee Shop	Gym / Fitness Center	Furniture / Home Store	Bagel Shop	Lingerie Store	Latin American Restaurant	Ice Cream Shop

4.2 Cluster 2 (Purple)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Lincoln Square	Plaza	Café	Concert Hall	Performing Arts Venue	Italian Restaurant	Theater	Gym / Fitness Center	Gym	American Restaurant	Indie Movie Theater
14	Clinton	Theater	Gym / Fitness Center	Italian Restaurant	Wine Shop	Gym	American Restaurant	Coffee Shop	Sandwich Place	Cocktail Bar	Hotel
15	Midtown	Coffee Shop	Hotel	Theater	Bakery	Clothing Store	Sandwich Place	Pizza Place	Cuban Restaurant	Tailor Shop	Indian Restaurant
16	Murray Hill	Sandwich Place	Coffee Shop	Japanese Restaurant	Hotel	American Restaurant	Gym / Fitness Center	Pizza Place	Cuban Restaurant	Juice Bar	Mediterranean Restaurant
29	Financial District	Coffee Shop	American Restaurant	Bar	Pizza Place	Italian Restaurant	Cocktail Bar	Hotel	Juice Bar	Steakhouse	Event Space
32	Civic Center	Coffee Shop	Gym / Fitness Center	French Restaurant	Park	Hotel	Cocktail Bar	Yoga Studio	Spa	American Restaurant	Café
33	Midtown South	Korean Restaurant	Hotel	Dessert Shop	Japanese Restaurant	Café	Burger Joint	Gym / Fitness Center	Cocktail Bar	Coffee Shop	Scenic Lookout
38	Flatiron	Gym / Fitness Center	Mediterranean Restaurant	Café	Gym	New American Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Park	Wine Shop	Toy / Game Store
39	Hudson Yards	Hotel	Gym / Fitness Center	American Restaurant	Italian Restaurant	Coffee Shop	Park	Gym	Bar	Dog Run	Café

4.3 Cluster 3 (Blue)

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
37 Stuyvesant Town	Boat or Ferry	Park	Playground	Harbor / Marina	Fountain	Farmers Market	Skating Rink	Bistro	Gas Station	Baseball Field

4.4 Cluster 4 (Teal)

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2 Washington Heights	Café	Bakery	Mobile Phone Shop	Coffee Shop	New American Restaurant	Tapas Restaurant	Latin American Restaurant	Park	Italian Restaurant	Supermarket
3 Inwood	Mexican Restaurant	Café	Restaurant	Lounge	Chinese Restaurant	Park	Frozen Yogurt Shop	Pizza Place	Spanish Restaurant	Bakery
4 Hamilton Heights	Pizza Place	Coffee Shop	Café	Deli / Bodega	Mexican Restaurant	Yoga Studio	Sushi Restaurant	Bakery	Caribbean Restaurant	Chinese Restaurant
7 East Harlem	Mexican Restaurant	Thai Restaurant	Bakery	Spa	Latin American Restaurant	Deli / Bodega	Sandwich Place	Taco Place	Cocktail Bar	Café
11 Roosevelt Island	Park	Farmers Market	Dog Run	Outdoors & Recreation	Supermarket	Coffee Shop	Sandwich Place	Food & Drink Shop	Greek Restaurant	Gym
26 Morningside Heights	Bookstore	American Restaurant	Park	Coffee Shop	Deli / Bodega	Burger Joint	Sandwich Place	Indian Restaurant	New American Restaurant	Supermarket
28 Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Plaza	Gourmet Shop	Burger Joint	Food Court	Shopping Mall
35 Turtle Bay	Italian Restaurant	Coffee Shop	Sushi Restaurant	Café	Park	Deli / Bodega	French Restaurant	Seafood Restaurant	Pharmacy	Karaoke Bar
36 Tudor City	Café	Park	Deli / Bodega	Sushi Restaurant	Mexican Restaurant	Garden	Seafood Restaurant	Coffee Shop	Thai Restaurant	Diner

4.5 Cluster 5 (Orange)

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Marble Hill	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Diner	Miscellaneous Shop	Pizza Place	Steakhouse	Supplement Shop	Seafood Restaurant

5. Discussion

Out of the initial 40 neighbourhoods, the K-Means clustering method has grouped 20 neighbourhoods into Cluster 1, 9 into Cluster 2, 1 in Cluster 3, 9 in Cluster 4, and 1 in Cluster 5.

Based on the clustering done by the machine learning algorithm, we identified two different clusters that may be suitable to open up Rasa's new restaurant. Looking at both Cluster 1 and Cluster 2, there are a lot of ethnic restaurants in the area which may imply the presence of a diverse population. However, Cluster 1 would be a second option compared to Cluster 2. This is because in Cluster 1, we can see there are also more Asian restaurants in the area compared to Cluster 2. Choosing Cluster 2 may reduce the risk of competition among other Asian restaurants, since Rasa's main dishes are targeted towards those who prefer Asian cuisines (Malay, Chinese, and Indian).

Cluster 3 is not suitable since it is more of a recreational area due to its nearby location to the East River. While Cluster 4 is a more suitable area for food businesses focusing on light meals, such as coffee shops and bakeries. This may also be due to a number of parks available in the area, New Yorkers would enjoy walks in the park while enjoying coffee and snacks instead of

heavy meals. Cluster 5 itself is located too far from the city centre and may not be suitable for any kind of business to expand there.

6. Conclusion

Using K-Means clustering on location data for New York and Foursquare API, we were able to narrow down 40 different neighbourhoods into 5 different clusters. Based on the analysis done in this study, we have decided that the best place for Rasa to expand their business is in Cluster 2. While Cluster 1 is could also be a potential location, Cluster 2 reduces competition risk. However, once Rasa has established itself in New York City with its upcoming branch, it could be a good idea for Rasa to expand in Cluster 1 as well.