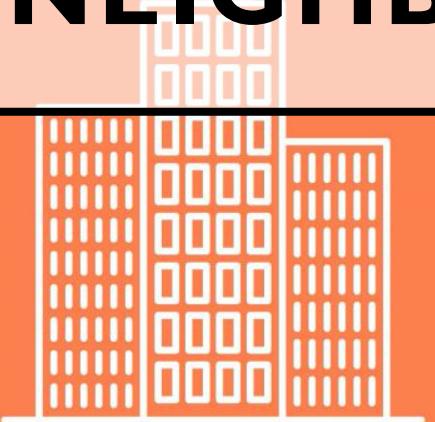
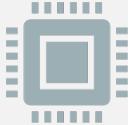




# THE BATTLE OF THE NEIGHBORHOODS



# INTRODUCTION



Customer: Rasa is a well-known restaurant brand from Malaysia and has recently expanded its business to Toronto, Canada.



Objective: 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.



Business Problem: 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.



Target Audience: Besides being useful to the owners, this project can also assist other business owners looking to expand their businesses anywhere around New York city.

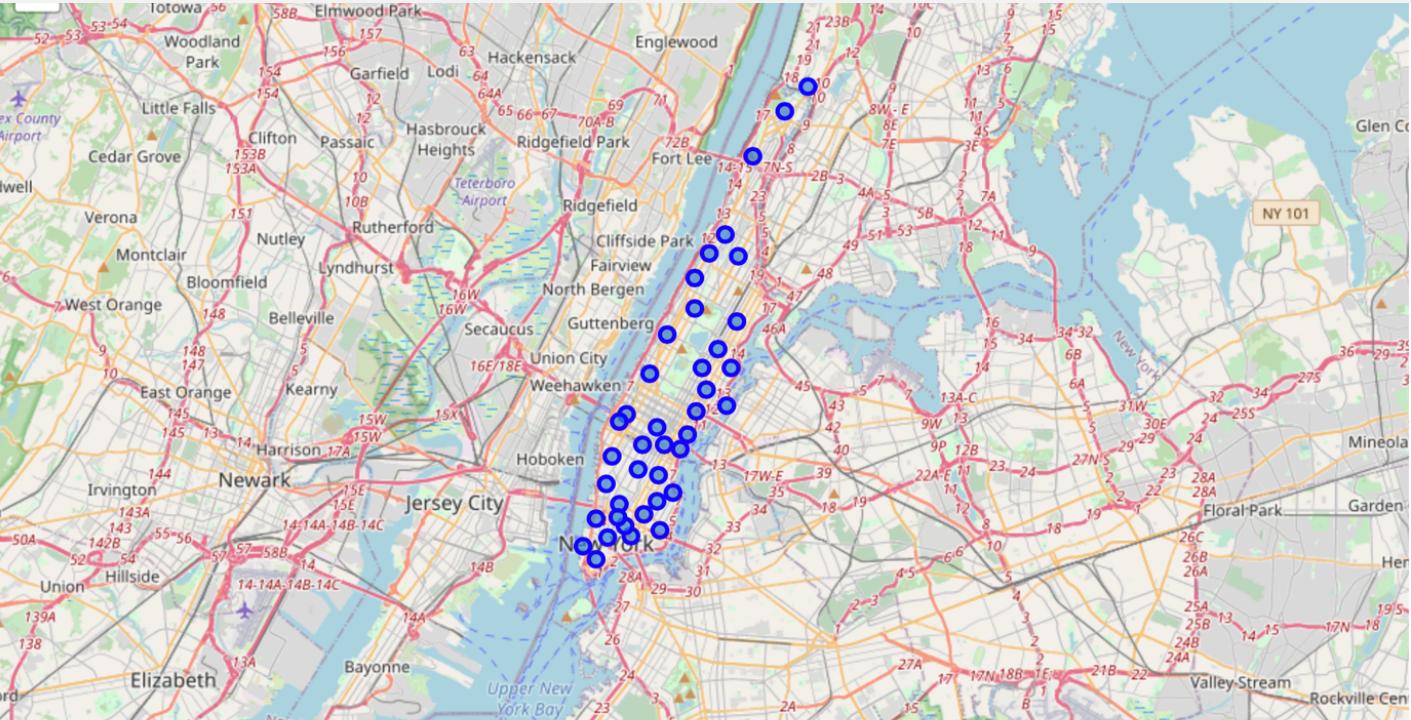
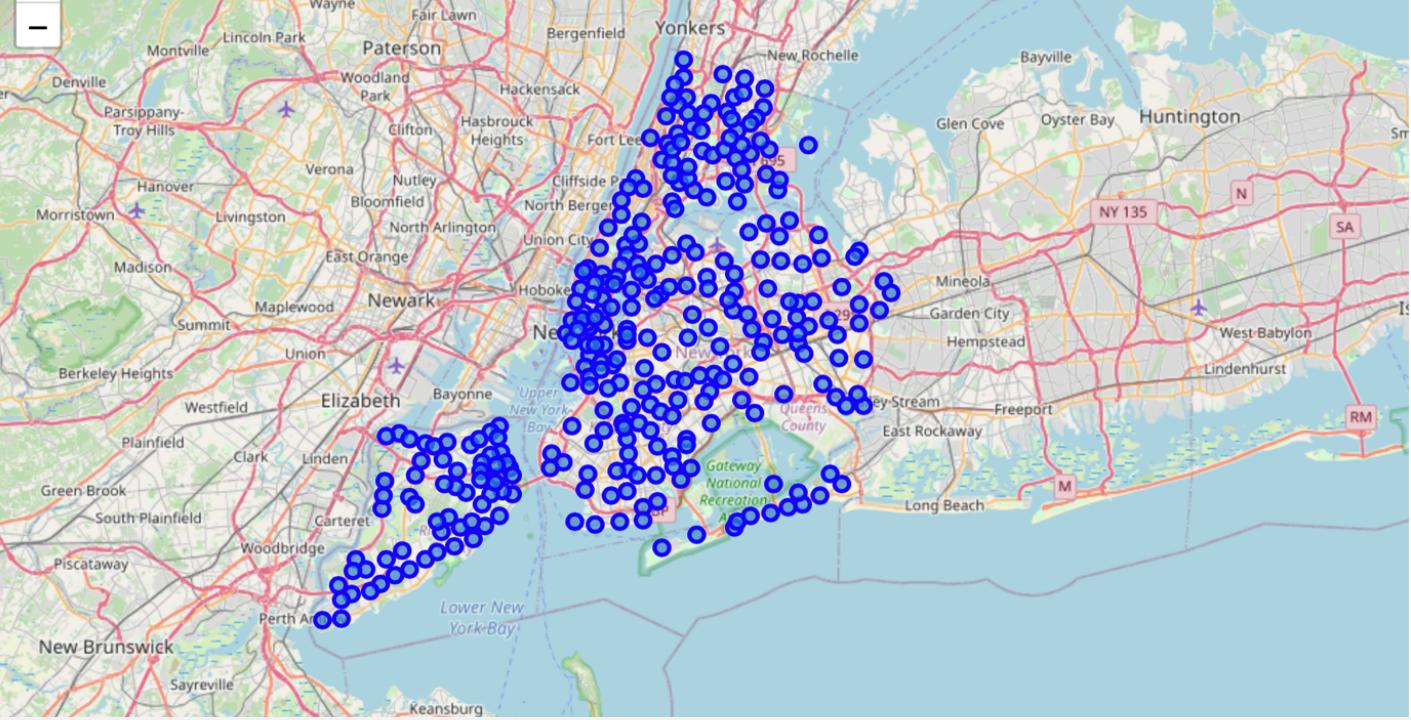
# DATA

- New York location data:  
[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572).
- Contains 5 boroughs and 306 neighbourhoods
- Using this dataset, we extract features that will be used in the analysis, which are Borough, Neighbourhood, Latitude, and Longitude.

	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

# DATA

- 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 .



# DATA

- Once we are satisfied with our final dataframe, we proceed to explore venues available in each neighbourhood.
- 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.

	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

# METHODOLOGY

- 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.
- 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.

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

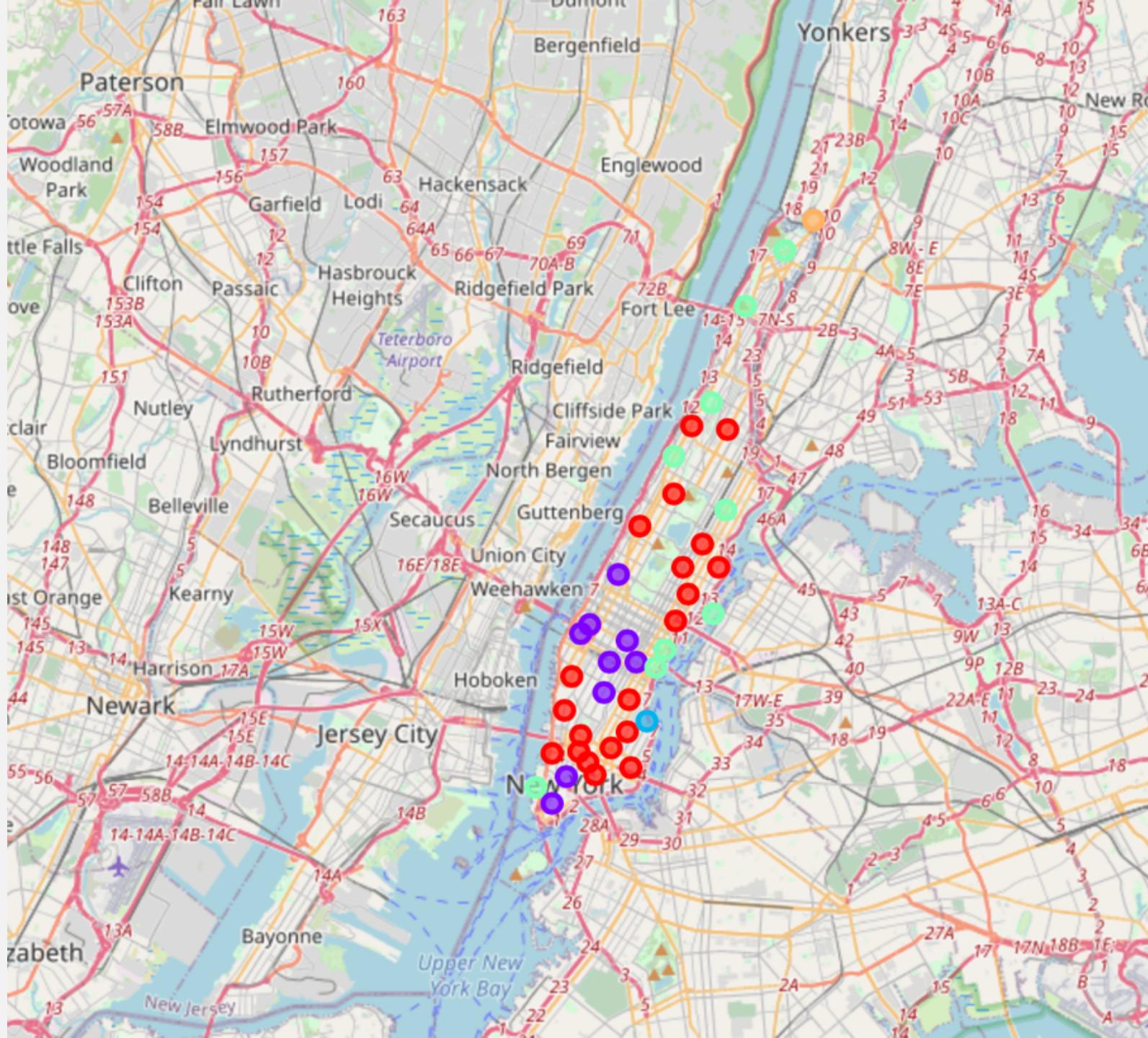
# METHODOLOGY

- 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.

	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

## RESULTS

- Cluster 1 (Red):
  - 20 Neighbourhoods
- Cluster 2 (Purple):
  - 9 Neighbourhoods
- Cluster 3 (Blue):
  - 1 Neighbourhood
- Cluster 4 (Teal):
  - 9 Neighbourhoods
- Cluster 5 (Orange)
  - 1 Neighbourhood



## DISCUSSION

- 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).

- 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 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.

## CONCLUSION