

IBM Data Science Capstone Project

The Restaurant Battle of Neighborhoods in New York City

Introduction

New York City is one of the most visited cities, full of vibrant international food scene, and ranked as the top cities for food [1]. This can be good and bad news for tourists as finding the right place to stay, close to restaurants with desired cuisines, is undoubtedly a huge challenge. Great food or restaurant is always one of the biggest factors in determining strategic location to stay for tourists.

This project is aimed to provide recommendations to tourists in New York City with neighborhood clusters, categorised by top cuisines in the vicinity, which facilitates them in choosing the best place, either for long or short stay. The chosen scenario for this project is to recommend neighborhoods in Bronx, New York, to tourists who are looking for an area with specific cuisine such as Italian, Spanish etc.

Description of data

The following data is required to conduct the project analysis:

1. List of neighborhoods in Bronx, New York, including geo-coordinates:
 - source: https://cocl.us/new_york_dataset (stored in json format)
 - relevant information is extracted using python library json
2. Top venues in each neighborhood, focusing of restaurant categories:
 - data is extracted from Foursquare through an API using python library request

Methodology

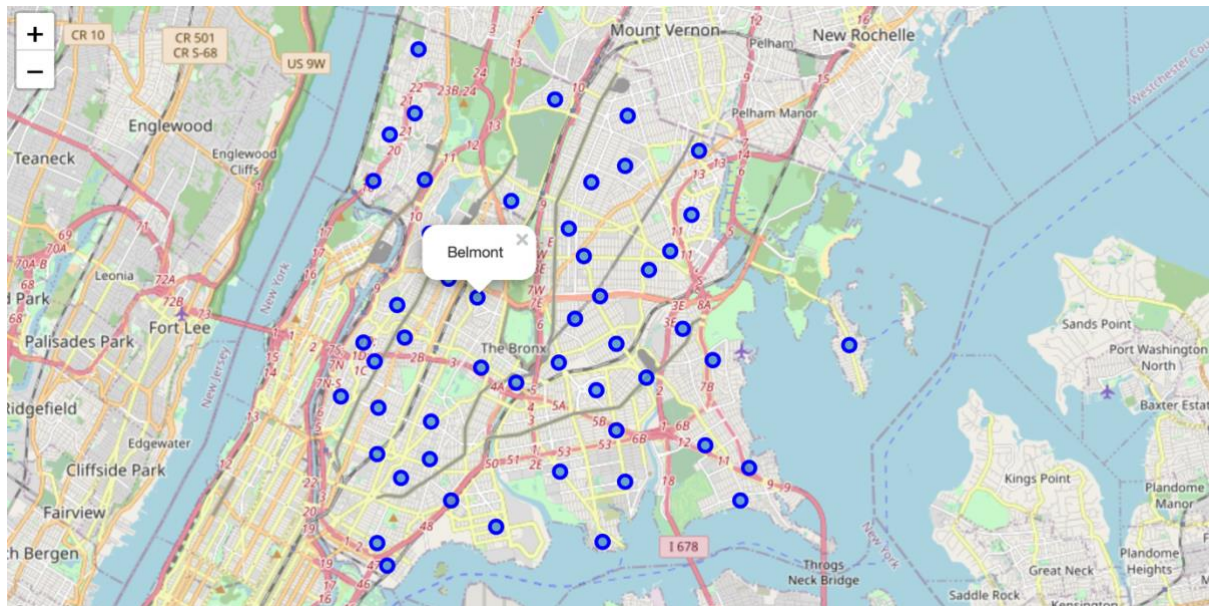
1. Data preparation

A list of neighborhoods, including geo-coordinates, is extracted from json file using python library request, and transformed into pandas dataframe for better management of data exploration and manipulation.

	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

Top rows of neighborhoods with geo-coordinates

Next, a visual representation of each neighborhood is created using python library folium, with initial location of New York City, where its geospatial data is obtained from python library Nominatim.



Map of neighborhoods in Bronx, New York

2. Data exploration and manipulation

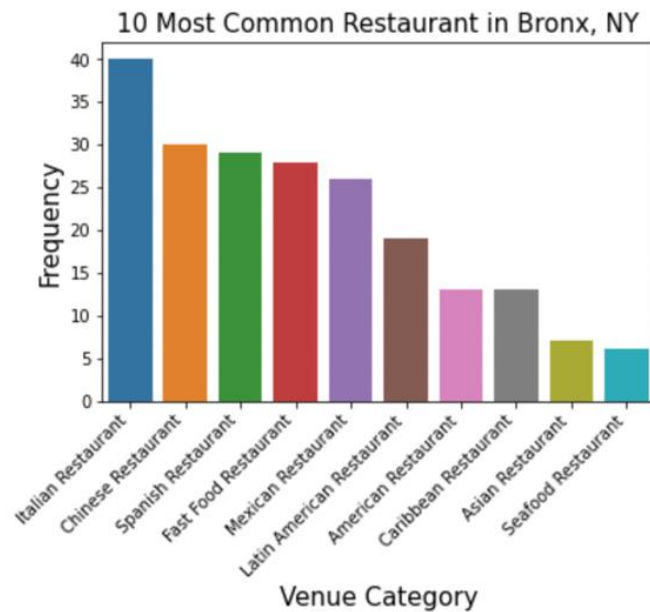
Each neighborhood is further explored with venues' names, categories and geo-coordinates tagged in Foursquare. Through Foursquare API, data is received in json format with limit of 100 venues for each neighborhood. The venues are collected based on a radius of 500 metres from respective neighborhood geo-coordinates.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Top rows of venues with geo-coordinates for each neighborhood

Foursquare results provide a list of 1,179 venues registered all over Bronx, where 239 venues are restaurant with 28 unique categories, such as Italian, Spanish etc. For better clustering representation, restaurant venues without any specific cuisine (category: Restaurant) are removed in the dataset.

Python library seaborn and matplotlib are used to plot a bar chart of 10 most common restaurant in the whole borough. It shows that Italian cuisine is the most popular in Bronx, followed by Chinese and Spanish restaurants. This observation may infer that northern part of New York City has a relatively high number of people from these countries, and their food have been appreciated by both local and tourist staying in this area.



As part of data requirement to use machine learning technique, where only numerical data can be processed in the algorithm, dataset with restaurant venues is transformed into dummy variables using one-hot encoding (0/1). The processed dataset can be further analysed by grouping the neighborhoods to show the mean frequency of occurrence in each restaurant category.

	Neighborhood	African Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Eastern European Restaurant	Fast Food Restaurant	French Restaurant	Rest
1	Co-op City	0	0	0	0	0	0	0	0	0	1	0	
2	Eastchester	0	0	0	0	1	0	0	0	0	0	0	
3	Eastchester	0	0	0	0	0	0	0	0	0	0	0	
4	Eastchester	0	0	0	0	1	0	0	0	0	0	0	
5	Eastchester	0	0	0	0	0	1	0	0	0	0	0	

Top rows of neighborhoods with dummy variables of restaurant category

	Neighborhood	African Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Eastern European Restaurant	Fast Food Restaurant	French Restaurant	Rest
0	Allerton	0.0	0.000000	0.0	0.0	0.0	0.500	0.0	0.0	0.000000	0.500000	0.0	
1	Baychester	0.0	0.000000	0.0	0.0	0.0	0.000	0.0	0.0	0.000000	0.333333	0.0	
2	Bedford Park	0.0	0.000000	0.0	0.0	0.0	0.375	0.0	0.0	0.000000	0.000000	0.0	
3	Belmont	0.0	0.035714	0.0	0.0	0.0	0.000	0.0	0.0	0.035714	0.035714	0.0	
4	Bronxdale	0.0	0.000000	0.0	0.0	0.0	0.200	0.0	0.0	0.200000	0.000000	0.0	

Top rows of neighborhoods with mean frequency of each restaurant category

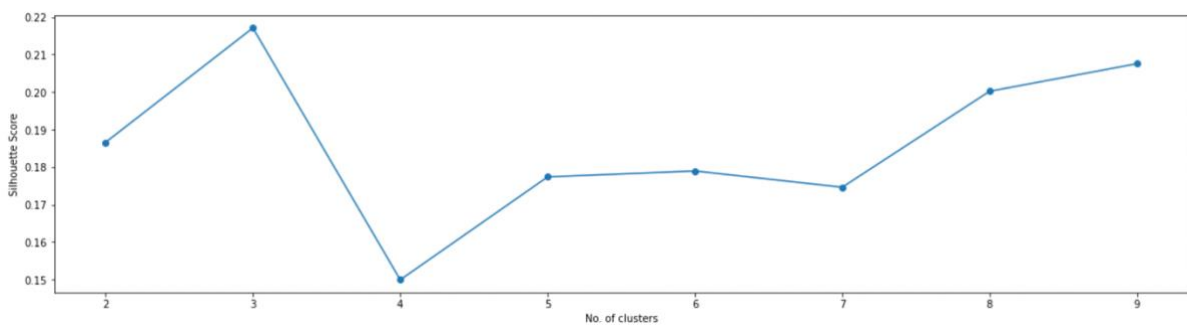
This dataset is further manipulated into dataframe of neighborhoods with the most common restaurant venue, as shown below:

	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	Allerton	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Cuban Restaurant
1	Baychester	Spanish Restaurant	Mexican Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant
2	Bedford Park	Chinese Restaurant	Mexican Restaurant	Italian Restaurant	Spanish Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant
3	Belmont	Italian Restaurant	Spanish Restaurant	Mexican Restaurant	American Restaurant	Eastern European Restaurant	Fast Food Restaurant	Japanese Restaurant	Greek Restaurant	Arepa Restaurant	Asian Restaurant
4	Bronxdale	Italian Restaurant	Spanish Restaurant	Chinese Restaurant	Eastern European Restaurant	Mexican Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant

Top rows of neighborhoods with the 10 most common restaurant venue

3. Data clustering

To perform data clustering, the most popular unsupervised machine learning technique, which is K-means, will be used. This method requires pre-defined number of clusters, and it is not practical and efficient to select based on above visual representation or trial and error approach. Thus, silhouette scores are plotted within a range of clusters, as shown below:



Line plot of silhouette score against no. of clusters

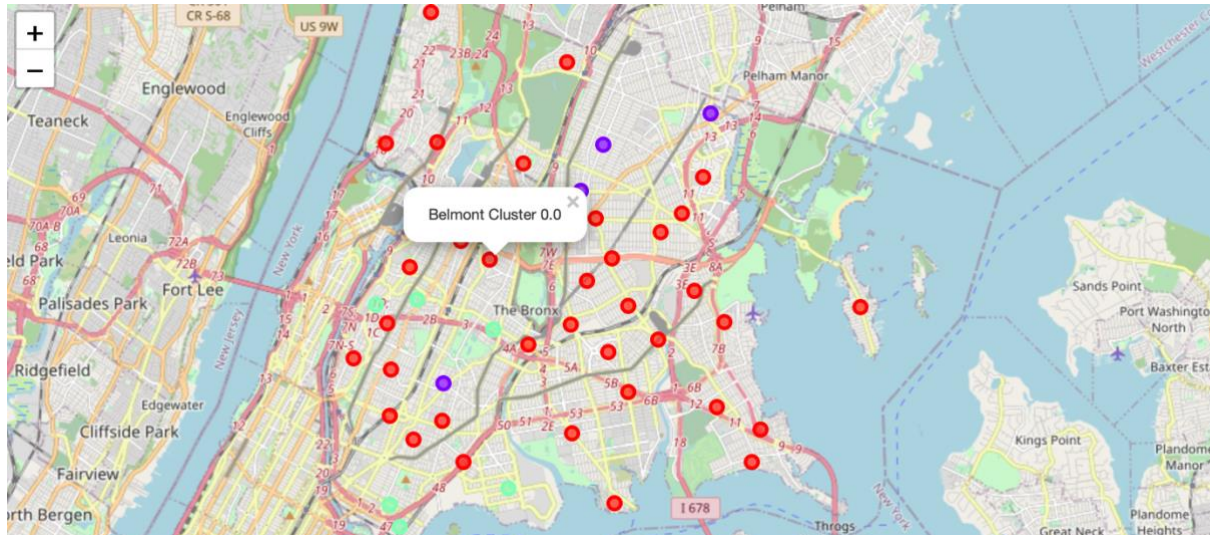
The optimal number of clusters is determined as 3 for the neighborhood dataset, and K-means algorithm provides cluster labels to each neighborhood.

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	9th Most Common Venue	10th Most Common Venue
Co-op City	40.874294	-73.829939	0.0	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
Eastchester	40.887556	-73.827806	1.0	Caribbean Restaurant	Chinese Restaurant	Seafood Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Comfort Food Restaurant
Kingsbridge	40.881687	-73.902818	0.0	Mexican Restaurant	Latin American Restaurant	Spanish Restaurant	Fast Food Restaurant	Caribbean Restaurant	Seafood Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	French Restaurant
Woodlawn	40.898273	-73.867315	0.0	Italian Restaurant	American Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
Norwood	40.877224	-73.879391	0.0	American Restaurant	Spanish Restaurant	Caribbean Restaurant	Chinese Restaurant	Mexican Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	Arepa Restaurant	Asian Restaurant

Top rows of neighborhoods with labelled clusters

Results

A visual representation of the clusters is created on the map using folium. The list of neighborhoods with labelled cluster is separated into dataframe of respective cluster and a bar chart of top 3 most common restaurant venue is plotted for each cluster. This is to ensure better data representation when providing recommendation to tourists with specific favourite cuisine.



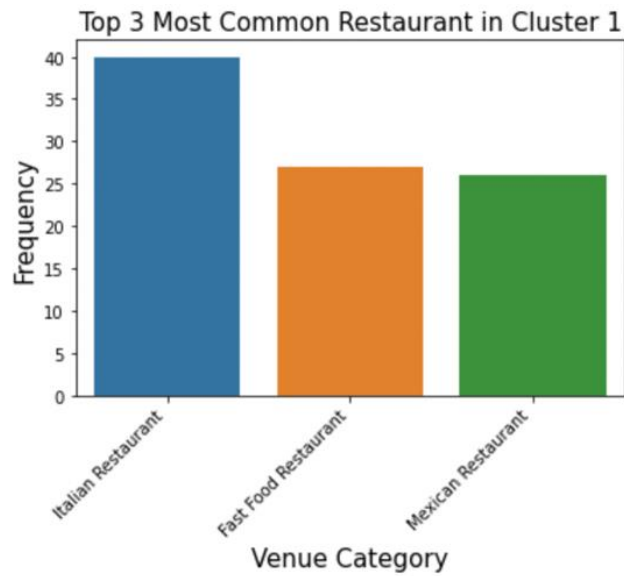
Map of 3 clusters of neighborhoods in Bronx, NY

Cluster 1

The first cluster is the biggest cluster and scattered all over Bronx, which indicates Italian, Fast Food and Mexican restaurants are well appreciated by a wide range of neighborhoods.

	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	Co-op City	Fast Food Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
5	Kingsbridge	Mexican Restaurant	Latin American Restaurant	Spanish Restaurant	Fast Food Restaurant	Caribbean Restaurant	Seafood Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	French Restaurant
6	Woodlawn	Italian Restaurant	American Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
7	Norwood	American Restaurant	Spanish Restaurant	Caribbean Restaurant	Chinese Restaurant	Mexican Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	Arepa Restaurant	Asian Restaurant
9	Baychester	Spanish Restaurant	Mexican Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant
10	Pelham Parkway	Italian Restaurant	Sushi Restaurant	Chinese Restaurant	Mexican Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant
11	City Island	Seafood Restaurant	Italian Restaurant	American Restaurant	Spanish Restaurant	French Restaurant	Indian Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant
12	Bedford Park	Chinese Restaurant	Mexican Restaurant	Italian Restaurant	Spanish Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant
13	University Heights	African Restaurant	American Restaurant	Chinese Restaurant	Fast Food Restaurant	Latin American Restaurant	Indian Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant
15	Fordham	Spanish Restaurant	Fast Food Restaurant	Chinese Restaurant	Latin American Restaurant	African Restaurant	Greek Restaurant	American Restaurant	Caribbean Restaurant	Mexican Restaurant	Arepa Restaurant

Top rows of neighborhoods in Cluster 1

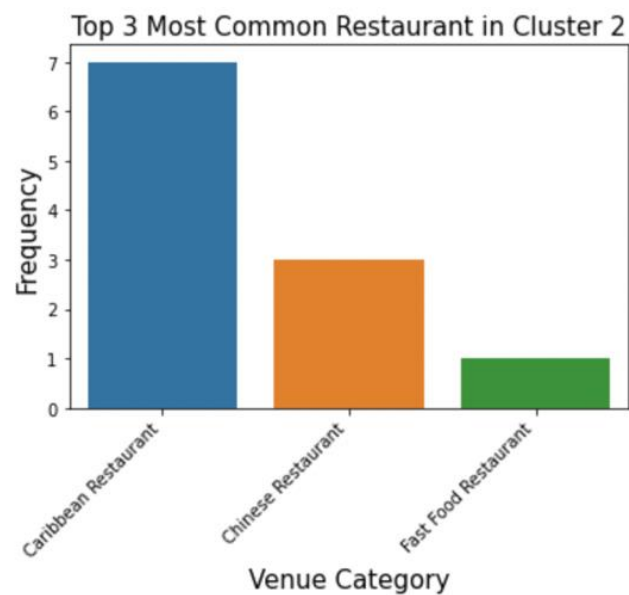


Cluster 2

The second cluster is much smaller than the first cluster, and mostly located in the northern part of Bronx, with Caribbean and Chinese cuisines as the most popular in the area.

	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	Eastchester	Caribbean Restaurant	Chinese Restaurant	Seafood Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Comfort Food Restaurant
8	Williamsbridge	Caribbean Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Eastern European Restaurant
40	Olinville	Caribbean Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Eastern European Restaurant
45	Claremont Village	Chinese Restaurant	Caribbean Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Comfort Food Restaurant	Cuban Restaurant	Eastern European Restaurant

List of neighborhoods in Cluster 2

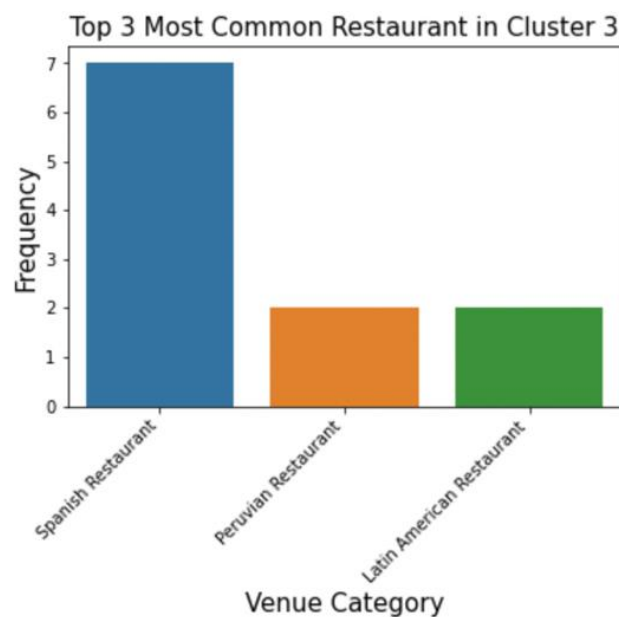


Cluster 3

The third cluster is slightly larger than the second cluster, and generally located in the opposite area ie. southern part of Bronx, with Spanish restaurants being the people's favourite who stay in the neighborhoods.

	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
14	Morris Heights	Spanish Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
16	East Tremont	Spanish Restaurant	Puerto Rican Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant
20	Mott Haven	Spanish Restaurant	Peruvian Restaurant	Latin American Restaurant	Vietnamese Restaurant	Greek Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant
21	Port Morris	Spanish Restaurant	Peruvian Restaurant	Latin American Restaurant	Vietnamese Restaurant	Greek Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant
23	Hunts Point	Spanish Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant
48	Mount Hope	Spanish Restaurant	Chinese Restaurant	Vietnamese Restaurant	Indian Restaurant	American Restaurant	Arepa Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Cuban Restaurant

List of neighborhoods in Cluster 3



Discussion

Based on the clusters shown above, tourists whose favourite is Italian and Mexican cuisines, neighborhoods in the first cluster can be recommended for their staycation. Similar way for second and third clusters. The neighborhoods in those areas can be suggested to tourists who love Caribbean, Spanish or Chinese food, depending on their preference.

Conclusion

K-means clustering method is used to identify a set of locations with common characteristics, based on optimal number of clusters. It provides a set of recommended neighborhoods to tourists, suited with their preferred cuisines in the vicinity, which facilitates their decision making in choosing the right location to stay. The model can also be implemented on other boroughs in New York City, as well as any city in the world, provided there is sufficient dataset from Foursquare for clustering modeling.

References

[1] <https://www.travelandleisure.com/food-drink/worlds-best-cities-for-food>

The Jupyter notebook of the whole analysis can be found on the following GitHub link:

<https://github.com/acapamin/IBM-Capstone-Project/blob/main/Capstone%20Project%20Final.ipynb>