Analysing restaurants in Brooklyn, New York

Aditya Deepak Nayar

Introduction:

Background

New York in one of the most diverse cities in the world with people from several countries residing there .One of the most visible effects of this can be seen in the variety of eateries and restaurants that are found there with restaurants serving almost all types of cuisines. But all of them are not evenly distributed across the city and usually restaurants serving a particular cuisine might be more popular in one neighbourhood while other cuisines may be popular in some other neighbourhoods. These are affected by a lot of factors and some of which are evident whereas others are not. So a chef or a restaurant chain looking to open a restaurant in a particular neighbourhood will have to consider the popularity of the dishes or cuisine it is planning to serve in that given neighbourhood and it would also be beneficial to know the areas where they are popular.

Problem

Grouping different neighbourhoods in Brooklyn having similar food patterns of eateries so as to make decisions on which neighbourhood to open the restaurant in so as to get maximum customers.

Interest

This problem will be considered by several large fast food chains, burger joints and so on to decide the area in which they should start their branch to get maximum customers. They can use the data to decide on the kind of dishes they should focus

more on in different outlets and also the kind of competition they will have in different areas and the offers that should be given and so on. This data can also be used by chefs who are specialised in a particular cuisine to decide the area where they would like to open their restaurant and the taste of people in different neighbourhoods.

Data

Data Sources

The details about the latitude and longitude of the neighbourhoods in New York were available in https://geo.nyu.edu/catalog/nyu_2451_34572 which I had downloaded from an IBM sever where it was present in json format which was cleaned and stored into a data frame. The details about the restaurants were obtained by using the Foursquare API.

Data Cleaning

Initially the json file containing the latitude and longitude details were downloaded and then they were converted into a data frame with neighbourhood, borough, latitude and longitude as the features. As this data contained details of all boroughs in New York I filtered it so as to contain neighbourhoods in Brooklyn alone.

Then requests were made using the Foursquare API which returned the details of all the venues in a radius of 500 meters around the neighbourhood's latitude and longitude. The request also returned the category of the given venue. These details were then put into a data frame.

To filter out the other venues that were not restaurants or eateries, by checking the venue category one can find out that all eateries had either a restaurant, joint, bistro or place in their venue type. Shops like ice cream shops and candy shops were not included. Then these new columns having one of the above four words were put into another list and then their values were appended into the data frame through which all the eateries were stored in a single data frame.

Feature Selection

For the analysis all the venues that were eateries were chosen. Along with this the neighbourhood name was also chosen to enable clustering later on was also selected. The neighbourhood coordinates are also stored to map the clusters at a later stage. There were no redundancies in the results of the Foursquare API.

Methodology

Libraries

The libraries used were pandas, numpy ,sci-kit- learn ,geopy , matplotlib , folium and json.

Exploratory data analysis

The data frame containing all the required features was obtained as shown in the data section. A count of the restaurants of each neighbourhood was obtained and also the frequency of the venues in each venue category was calculated. This was then converted into one-hot encoding as shown.

	Neighborhood	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bar	Beer Bar	Bistro	Brazilian Restaurant	Burger Joint	Burrito Place	Cajun / Creole Restaurant	Cantonese Restaurant
0	Bay Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Bay Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bay Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Bay Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Bay Ridge	0	0	0	0	0	0	0	0	0	0	0	0	0	0

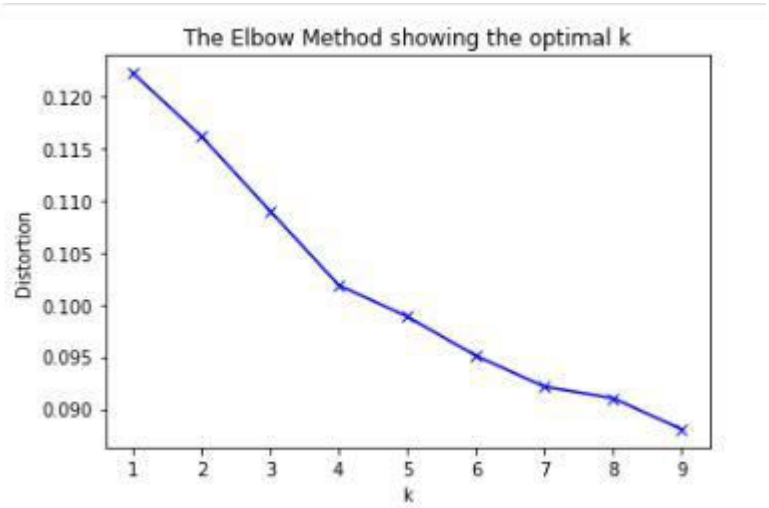
Then the data was grouped so that all the values in the table will be normalised and so that the values can be used for clustering (as clustering algorithms don't take string values) and after this was merged with the venue category names to give a list of popular venues in decreasing order.

	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	9th Most Common Venue	10tl Mos Commo Venu
0	Brooklyn	Bay Ridge	40.625801	-74.030621	3	Italian Restaurant	Pizza Place	Greek Restaurant	American Restaurant	Bar	Hookah Bar	Sandwich Place	Seafood Restaurant	Chinese Restaurant	Middl Easter Restauran
1	Brooklyn	Bensonhurst	40.611009	-73.995180	1	Chinese Restaurant	Pizza Place	Italian Restaurant	Sushi Restaurant	Hotpot Restaurant	Dumpling Restaurant	Shabu-Shabu Restaurant	Russian Restaurant	American Restaurant	Asia Restauran
2	Brooklyn	Sunset Park	40.645103	-74.010316	3	Latin American Restaurant	Pizza Place	Mexican Restaurant	Italian Restaurant	Sandwich Place	Hotel Bar	Hotpot Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Foo Restauran

Modelling

To model the above data an un-supervised algorithm was required as we had to group similar restaurants together .Here the k – means clustering was used to group the data.

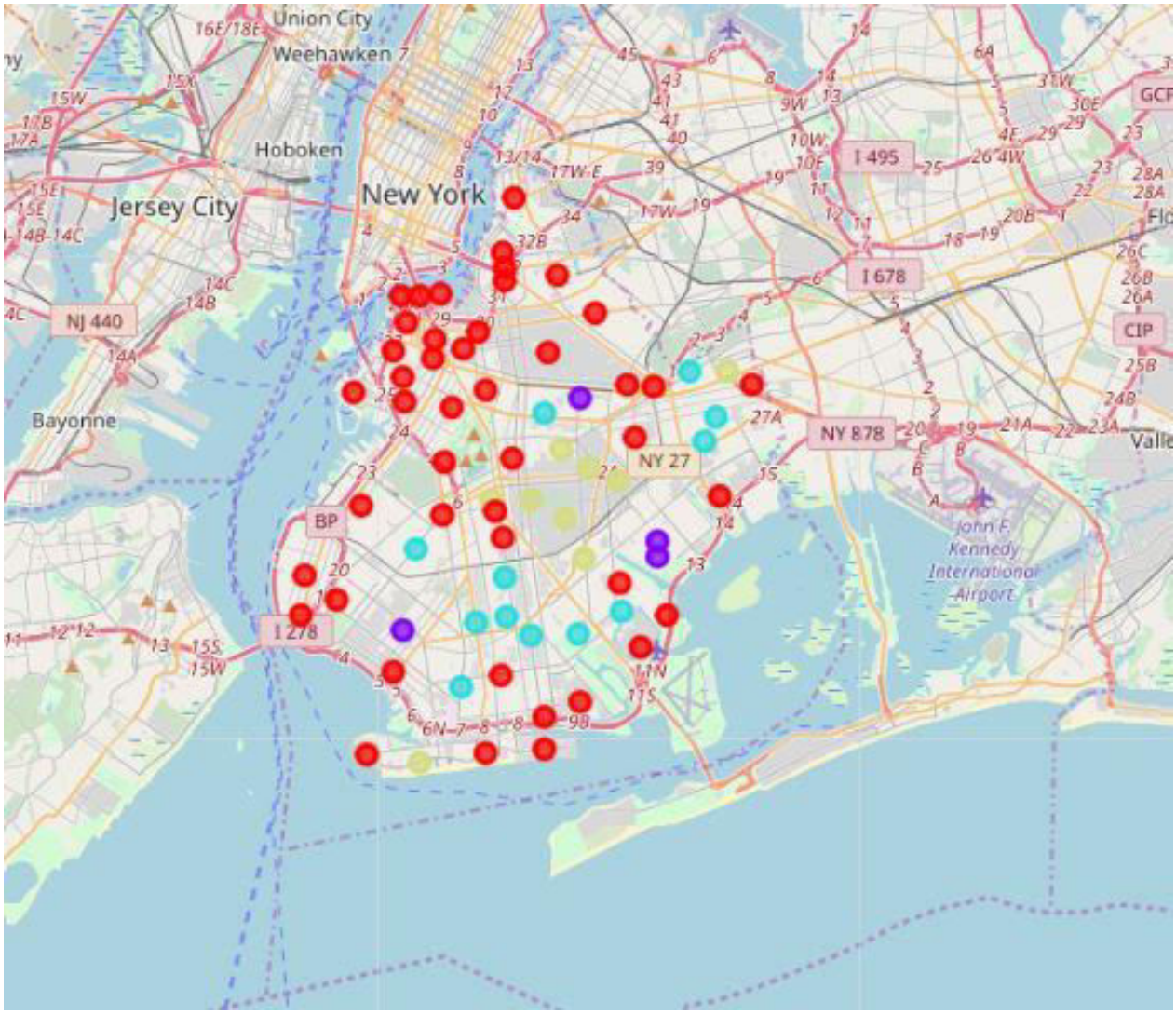
For getting the optimum number of clusters we used the elbow technique to find the value of k i.e. the number of clusters the data is to be grouped into. The graph obtained is shown.



Here the optimum value of k was found to be 4

Results

After using the k – means clustering algorithm in the above data, 4 sets of clusters were obtained with each cluster containing neighbourhoods with similar sets of popular restaurants.



Clustered neighbourhoods map

Then a bar graph was plotted showing the category of the most popular restaurants in each cluster.

	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
8	Flatbush	Mexican Restaurant	Caribbean Restaurant	Italian Restaurant	Juice Bar	Sandwich Place	Pizza Place	Chinese Restaurant	Middle Eastern Restaurant	German Restaurant	Fast Food Restaurant
10	East Flatbush	Chinese Restaurant	Caribbean Restaurant	Wings Joint	Greek Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Fried Chicken Joint	Gay Bar
25	Cypress Hills	Fried Chicken Joint	Fast Food Restaurant	Chinese Restaurant	Latin American Restaurant	Salon / Barbershop	Pizza Place	South American Restaurant	Spanish Restaurant	Gay Bar	Ethiopian Restaurant

Cluster1

	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
1	Bensonhurst	Chinese Restaurant	Pizza Place	Italian Restaurant	Sushi Restaurant	Hotpot Restaurant	Dumpling Restaurant	Shabu-Shabu Restaurant	Russian Restaurant	American Restaurant
4	Gravesend	Pizza Place	Italian Restaurant	Eastern European Restaurant	Bar	Chinese Restaurant	Hookah Bar	German Restaurant	Falafel Restaurant	Fast Food Restaurant
7	Manhattan Terrace	Pizza Place	Japanese Restaurant	Steakhouse	German Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant

Cluster2

	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
28	Canarsie	Asian Restaurant	Chinese Restaurant	Caribbean Restaurant	Wings Joint	Halal Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Fried Chicken Joint	Gay Bar
59	Paerdegat Basin	Asian Restaurant	Chinese Restaurant	Wings Joint	Halal Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	Fried Chicken Joint	Gay Bar

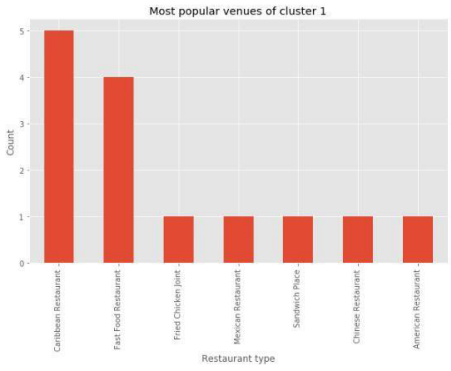
Cluster3

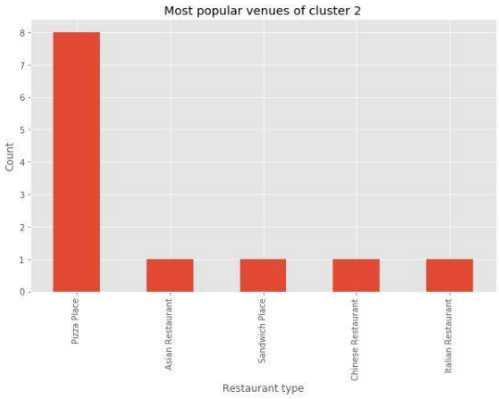
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bay Ridge	Italian Restaurant	Pizza Place	Greek Restaurant	American Restaurant	Bar	Hookah Bar	Sandwich Place	Seafood Restaurant	Chinese Restaurant	Middle Eastern Restaurant
2	Sunset Park	Latin American Restaurant	Pizza Place	Mexican Restaurant	Italian Restaurant	Sandwich Place	Hotel Bar	Hotpot Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
3	Greenpoint	Bar	Pizza Place	Cocktail Bar	Mexican Restaurant	Polish Restaurant	Salon / Barbershop	Sandwich Place	French Restaurant	Italian Restaurant	New American Restaurant
5	Brighton Beach	Eastern European Restaurant	Restaurant	Sushi Restaurant	Fast Food Restaurant	Russian Restaurant	Korean Restaurant	Mediterranean Restaurant	Taco Place	Fried Chicken Joint	Indian Restaurant

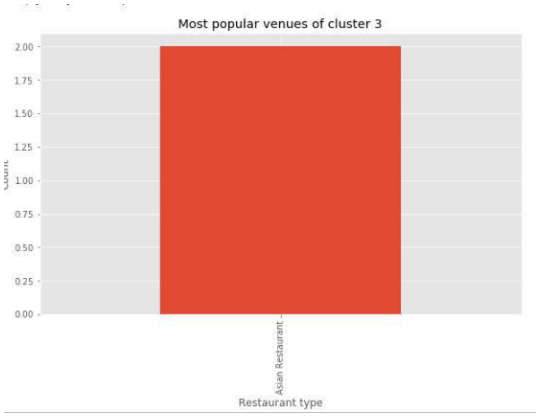
Cluster4

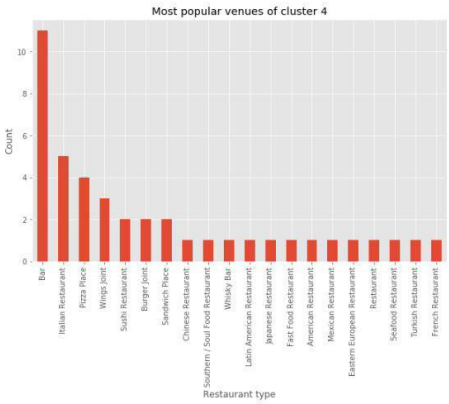
Through this the following results can be made

In the neighbourhoods in cluster 1 Caribbean restaurants were the first choice for most of the residents of the area which was followed by fast food restaurants and fried chicken joints. So in neighbourhoods such as Flat Brush or Cypress hills in Brooklyn Caribbean restaurants and fast food joints are more popular and receive more customers. Similarly in other clusters pizza places, Asian restaurants and bars topped the list. The graph of the most popular venues in each cluster are shown.









Discussions

In the above clusters cluster1 and cluster4 have the most number of neighbourhoods and cover almost all neighbourhoods in Brooklyn. So as bars and Caribbean restaurants are popular in these cluster we can say that they are one of the most common categories of restaurants in Brooklyn.

Similarly by looking at the above data one can even find the places where their restaurants have a better chance of success. For example if a company like KFC wants to open a new outlet in a neighbourhood in Brooklyn where they want to get a steady number of customers they have a better chance of success at a neighbourhood in cluster1 or cluster4 where fried chicken joints and wings joints are popular .

They can also choose a neighbourhood where they are less in number so as to put something new in the market.

Conclusions

In this project I analysed different categories of restaurants in Brooklyn, New York to find the popular cuisines in different neighbourhoods. This was done by analysing the popularity (likes) received by each restaurant in the Foursquare website. The data obtained will help chefs and food chains decide on the correct place to open their restaurant based on their style of food.

Moreover this can also be used by customers who are looking to try out popular cuisines in that particular neighbourhood. It can be used by food apps to give better recommendations to the user about different venues in the neighbourhood.

To get better clusters and more details information like restaurant reviews and user ratings can be taken from the Foursquare API.