

Capstone Project - The Battle of Neighborhoods

Introduction:

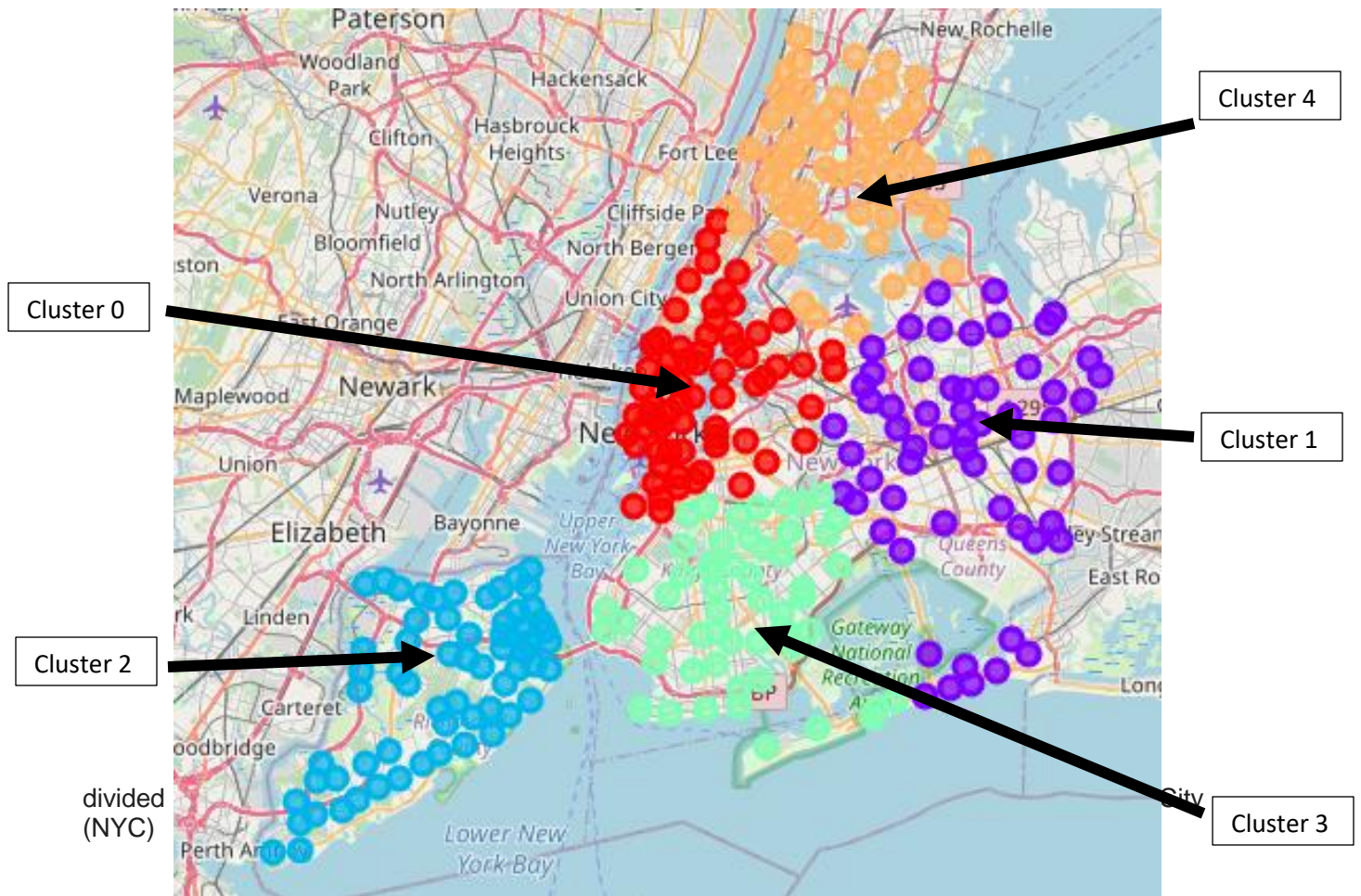
A new gastronomist from the Latin America area has decided to expand his exploratory cuisine to parts of the United States! He has chosen New York City to begin the journey as it has a large diversity of people, some of the best restaurants, and is a land of opportunity in the world of professional cooks. As he is unfamiliar with the area, he needs help deciding where to put his restaurant for the best chance at success. To optimize his chances of success a place will be chosen where most restaurants are located as it is an indicator that it is popular to go out for food in these areas and one where Mexican cuisine is more common as it may be an indicator that people will enjoy his cuisine in those areas.

Data:

The data that will be collected during this study will be the location data of neighborhoods in New York City to first divide it into clusters so each neighborhood can be analyzed separately. Once it is divided it will also be imperative to see what restaurants are in NYC as a general understanding will help us further understand what is happening. These two pieces of information help perform a top-down analysis allowing you to see the bigger picture before diving into a specific piece of NYC. Once this information is collected and analyzed we will develop an idea of what restaurants are most popular in the developed clusters using one-hot encoding techniques to determine which cluster will provide the best opportunity for a Latin-American based restaurant.

Methodology:

In this section, I will describe the data analysis conducted to achieve the results. To obtain a general idea of what NY looked like I used data from a lab in the course to obtain the geographical data of New York City (NYC) then used the geopy library to obtain the coordinates of NYC. Once all the data was downloaded, I decided to cluster NYC into 5 clusters using k-means clustering. I chose $k=5$ because there are 5 neighborhoods and because I did not know how the neighborhoods were originally divided, I was curious to see if the results would vary from the original neighborhood borders. The results would show that it would only vary slightly which can be seen in the following map of NYC.

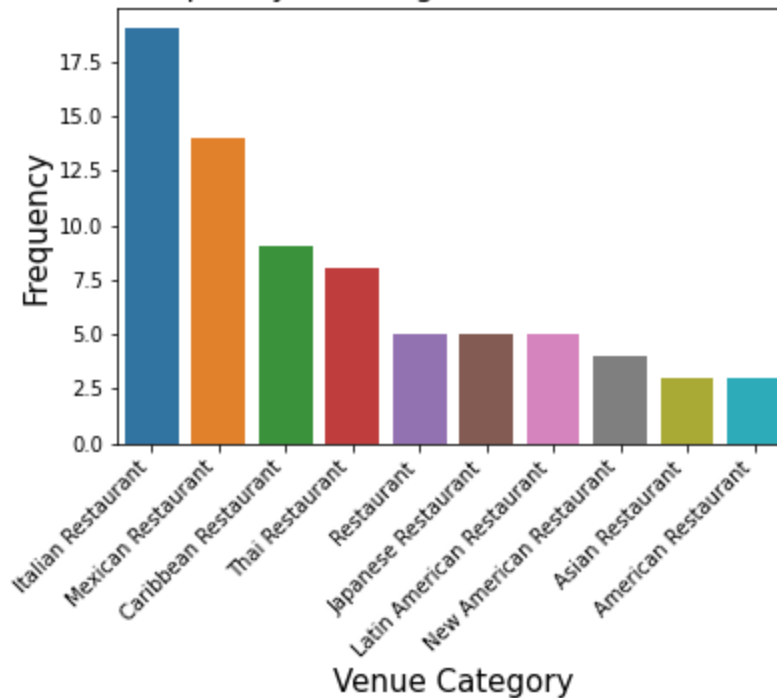


After finding the clusters on the map, I decided find the center of each cluster developed as a relative search point as everything could then be traced back to the center of the city. The average latitudes and longitudes can be seen below:

	Cluster Labels	Latitude	Longitude
0	0	40.734036	-73.966889
1	1	40.701191	-73.797676
2	2	40.588851	-74.137414
3	3	40.626466	-73.941361
4	4	40.843482	-73.876488

Before I went further into the analysis, I wanted to see how popular each cuisine type was in all of NYC so I conducted a search using the foursquare API to see which restaurants occurred most frequently.

10 Most Frequently Occurring Venues in 5 clusters of NY City



Looking at the above chart it can be noted that Mexican Restaurants are second most popular, right behind Italian cuisine which would suggest that it has a strong presence in the area and NYC would be a good area to open a restaurant. If there were no presence this would make the area more at risk as the cuisine could be considered too new for the area and could be intimidating.

	Cluster #	Number of Restaurants
1	Cluster 0	12
2	Cluster 1	27
3	Cluster 2	20
4	Cluster 3	26
5	Cluster 4	18

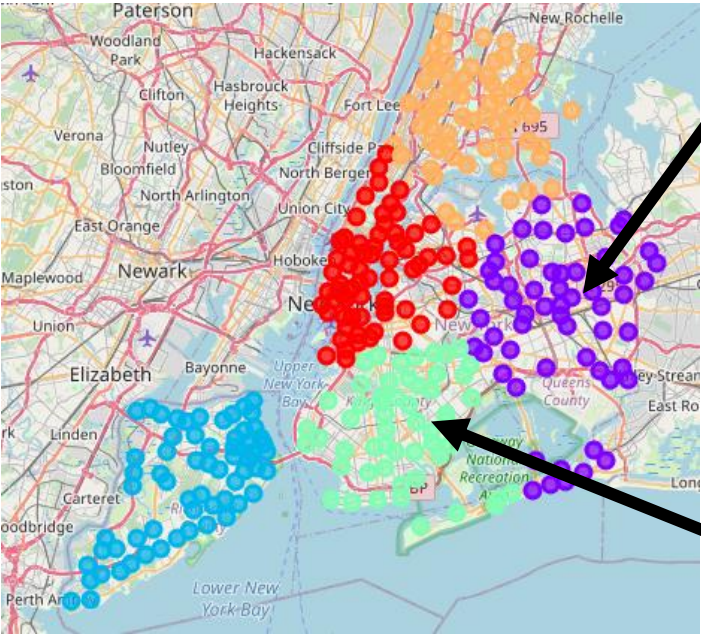
It can be noted from the above graph that the most popular food destinations are in Cluster 1 and Cluster 3 which would suggest that these clusters are popular food destination which may make these clusters ideal to open a restaurant as they may contain more people open to the suggestion of a Latin-American cuisine. Taking a further look at what is popular in each cluster we will further look at the frequencies of

each restaurant in each cluster using one hot encoding to determine where there may be the most competition and what areas he is most likely to get business.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cluster 0	Thai Restaurant	Italian Restaurant	Scandinavian Restaurant	Japanese Restaurant	Mediterranean Restaurant
1	Cluster 1	Caribbean Restaurant	Mexican Restaurant	Thai Restaurant	Greek Restaurant	Vegetarian / Vegan Restaurant
2	Cluster 2	Italian Restaurant	Restaurant	Japanese Restaurant	Indian Restaurant	Latin American Restaurant
3	Cluster 3	Caribbean Restaurant	Mexican Restaurant	American Restaurant	Thai Restaurant	Restaurant
4	Cluster 4	Italian Restaurant	Mexican Restaurant	Latin American Restaurant	Cuban Restaurant	Fast Food Restaurant

After finding the most popular restaurants in each cluster and where most of the food destinations are it is time to perform a final analysis.

Results:



Cluster	1
1st Most Common Venue	Caribbean Restaurant
2nd Most Common Venue	Mexican Restaurant
3rd Most Common Venue	Thai Restaurant
4th Most Common Venue	Greek Restaurant
5th Most Common Venue	Vegetarian / Vegan Restaurant

Cluster	3
1st Most Common Venue	Caribbean Restaurant
2nd Most Common Venue	Mexican Restaurant
3rd Most Common Venue	American Restaurant
4th Most Common Venue	Thai Restaurant
5th Most Common Venue	Restaurant

Discussion:

This analysis looked at which cluster had the most restaurants and assumed this would be a common food area. It also assumed that because other restaurants were there it would fit in with them. It did not look at revenue or traffic of each of these restaurants which are big factors in determining where you would put a restaurant and what tier restaurant he would be having. If it is assumed that this would be an average restaurant then either of these locations may work.

Conclusion:

Cluster 1 and 3 seem to be the most popular for restaurants supporting they are common areas for food which would make for a good restaurant area. It also showed that Mexican cuisine was not one of the most popular which suggests that the placement of one would not congest the area with Mexican food but support it as it is still a common venue. After looking at those factors it would be deemed that either of these locations would be suitable for the placement of a Latin-American restaurant.

Code via Github:

https://github.com/GellyMan20/Coursera_Capstone