2020

Food Store Location Hunt

RAVI TEJA PRABHALA VENKATA

1. Introduction

1.1 Background:

New York City has been teeming with a plethora of cosmopolitan restaurants. The city hosts a large number of people from different nations and ethnicities with various neighborhoods across the three major boroughs. Despite being arguably the busiest city in the world, it still has room for more and more businesses. We want to explore this prospect and provide insights about a feasible business venture in the city.

1.2 Problem:

The project explores the possibility of setting up a new food store business in the vicinity of New York City boroughs. We are exploring the 3 major boroughs – Manhattan, Brooklyn and Queens as the prospective location for the store. The food store is a specialty food store which offers food items specific to ethnicities and customs specific to Mediterranean, Greek, Middle Eastern all the way to Japan including almost every Asian country.

1.3 Interest

Our interest is only in finding out the hotspots in these three boroughs that have more Asian/Mediterranean/Greek restaurants. These hotspots are a key to our business as it becomes relatively easy and convenient to set up such a food store closer to these hotspots. Since patrons who visit these restaurants might like those cuisines, our food store offers a lot more for them to take home.

2. Data Acquisition

2.1 Data Sources:

Our data sources are OpenCageData APIs for geo location information and Foursquare APIs for New York City data. From Foursquare APIs, we specifically extract restaurant data and use it to deduce our analysis.

From OpenCageData, we get the following data:

City	Latitude	Longitude
Manhattan	40.789624	-73.959894
Brooklyn	40.650104	-73.949582
Queens	40.749824	-73.797634

From Foursquare APIs, we get data sample such as following:

Name	Latitude	Longitude	Category Name	Category Short Name
Da Capo	40.787679	-73.953899	Café	Café
Earl's Beer & Cheese	40.787331	-73.951725	American Restaurant	American
Marinara Pizza Upper East	40.782538	-73.953359	Pizza Place	Pizza
Dig Inn	40.780332	-73.954728	American Restaurant	American
Levain Bakery	40.777354	-73.955284	Bakery	Bakery

2.2 Data Cleansing:

When we extract data from Foursquare APIs, we can get every location's information in all three boroughs. We will cleanse this data and only extract "Food" category data. From this category, we get all restaurants located in a defined radius from the coordinates of each borough.

Once we get this data, we filter it to extract restaurants specific to only our interests i.e.
Greek, Mediterranean, Middle East, Asian, Indian, etc. We also include Seafood and Vegetarian/Vegan restaurant types as well since the store offers food specific to these categories as well.

List of restaurant types sample looks like the following table:

Category Name	Category Short Name
Japanese Restaurant	Japanese
Sushi Restaurant	Sushi
Donut Shop	Donuts
American Restaurant	American
Salad Place	Salad

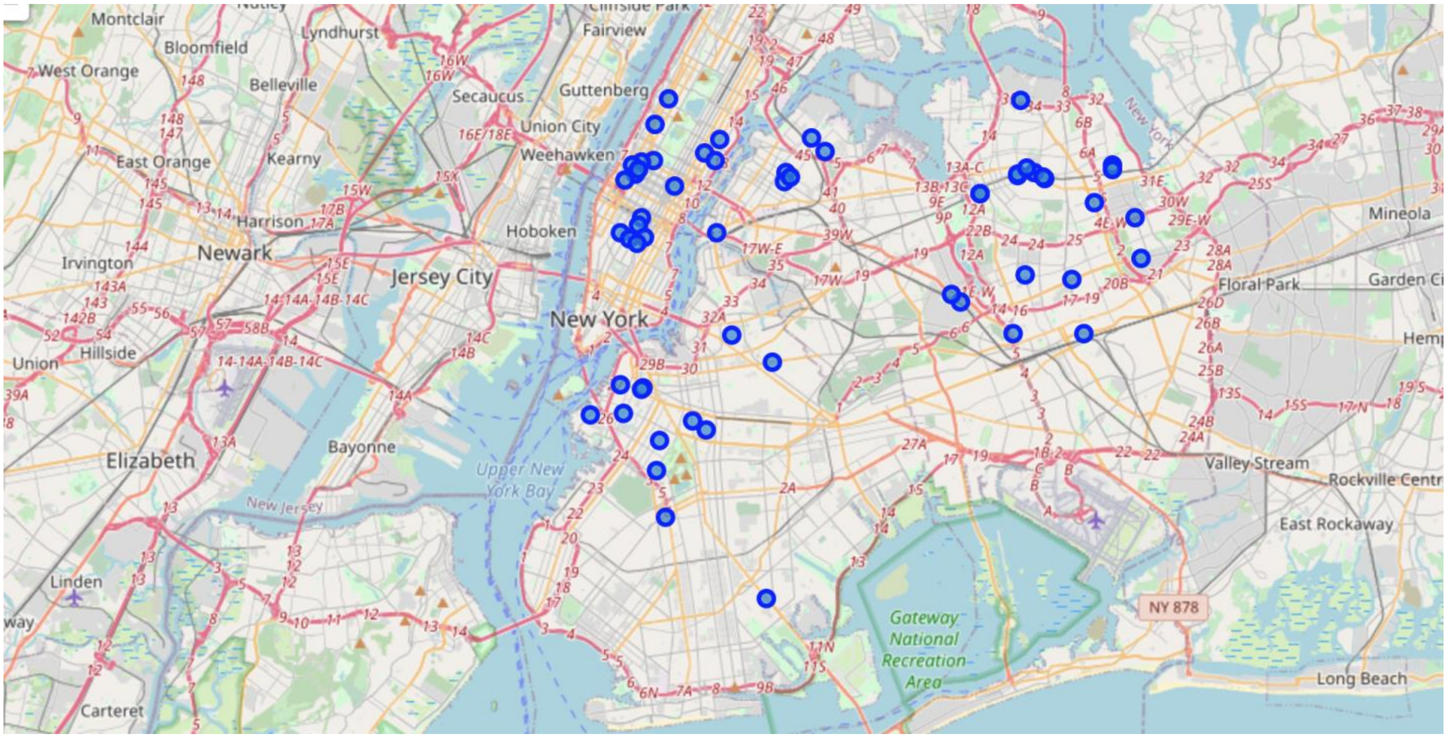
3. Exploratory Data Analysis

3.1 Data Analysis:

Since we only have data about restaurants, their cuisines and their location with no demographic data available, we need to leverage the existing data to make better decisions about the prospective business idea. Most of these locations are probably workplaces with relatively lesser domestic resident neighborhoods. Despite that fact, the store can still have the very same patrons if they can afford to spend some time with very little traveling distance.

We filter the data to match the restaurant types with the following list:

Category Type
Chinese
Falafel
Greek
Japanese
Korean
Lebanese
Mediterranean
Moroccan
Seafood
Sushi
Szechuan
Thai
Udon
Vegetarian / Vegan



As depicted in the map using blue circles, these restaurants are scattered across these three boroughs. It would be easier if we can use demographic data to achieve more accurate results but since we are counting on patrons who visit directly from their work locations, we only need the location of these types of restaurants.

Our major focus is on the location of these restaurants and the types of restaurants. We leverage this data and apply our machine learning to conclude best possible locations for our store business.

We have what we need, and we know what needs to be done, so there is no need for any further data analysis and hence no charts or graphs depicting the data relationships.

4. K Means clustering

4.1 Choice of machine learning algorithm:

We have a variety of machine learning algorithms that solve a wide range of problems. In our current scenario, we need to figure out the hotspots, or as we call them, clusters. Our goal is to figure out the clusters of these types of restaurants in New York City especially in the above mentioned three boroughs. We want to create dense clusters so that we can derive the centroids of these clusters that can be the potential location for the store.

Our data sources are OpenCageData APIs for geo location information and Foursquare APIs for New York City data. From Foursquare APIs, we specifically extract restaurant data and use it to deduce our analysis.

4.2 Implementation:

We use the geo location of the restaurants that are of our interest, filtered specifically that match the cuisine types we outlined. We perform K Means clustering on this data set and find out the optimal number of clusters i.e. k.

After clustering the data, we get the following dataset (only showing a few records):

Name	Latitude	Longitude	Category Name	Category	Cluster
The Mermaid Inn	40.788744	-73.974243	Seafood Restaurant	Seafood	2
PuTawn Local Thai Kitchen	40.774599	-73.951042	Thai Restaurant	Thai	0
Sala Thai	40.780124	-73.980475	Thai Restaurant	Thai	2
Up Thai	40.769898	-73.957598	Thai Restaurant	Thai	0
Tanoshi Sushi	40.767747	-73.953203	Sushi Restaurant	Sushi	0

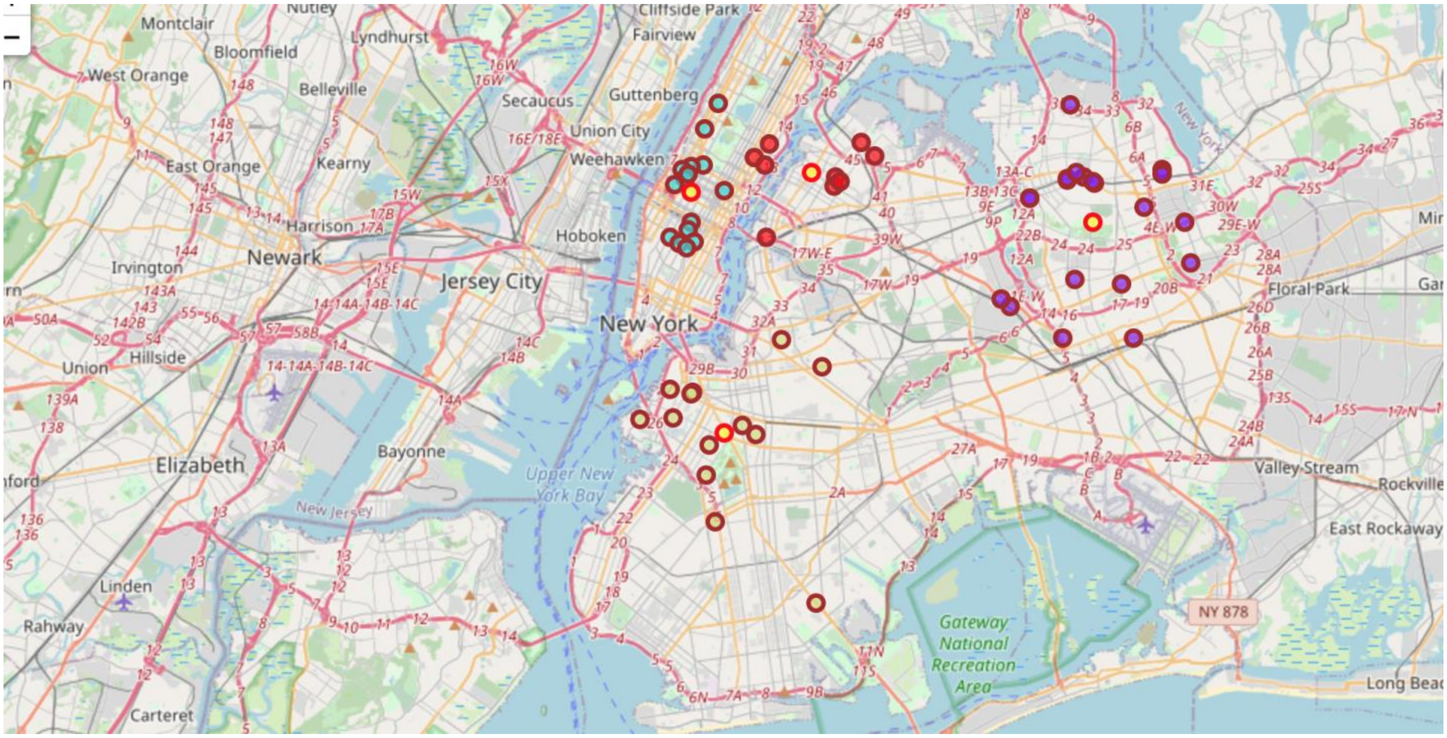
Then we depict the clustered data in a map to visualize the clusters.

Each colored circle represents an element in a cluster and each cluster has a specific color. In the above picture, we can see that we have 4 different color-filled circles representing 4 different clusters.

Here's how each cluster performed. Cluster 1 (Queens – purple) has the best density followed by Cluster 2 (Manhattan – cyan) and Cluster 3 (Brooklyn – yellow).

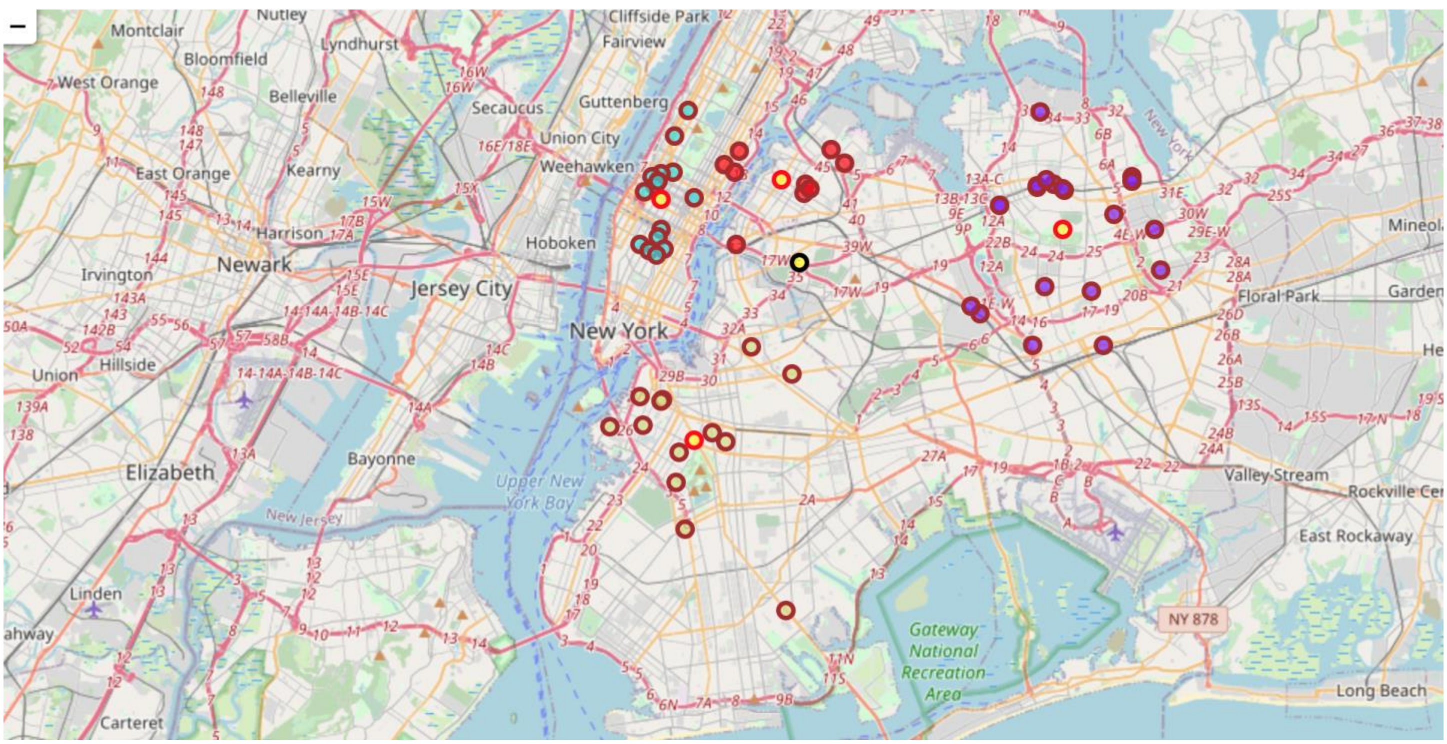
Cluster	Number of Restaurants
1	20
2	15
3	13
0	9

Let's visualize the clusters along with their centroids.



In the above map, for each cluster, a centroid is depicted by a circle outlined by red and filled with yellow. Each of these centroids can serve as a potential nearby location for our store.

Let us also calculate and depict the centroid of these centroids if we want to establish only one major store that’s close to all these clusters.



The circle outlined by black and filled with yellow color in the center is our centroid of centroids.

5. Results

Based on the above analysis, we have two feasible solutions that can be proposed.

5.1 Multiple store outlets

Driven by the data as depicted in the above maps, should the business decide to open the store in multiple locations, given enough resources and access, the stores can be established closer to each of the centroids of individual clusters in the order of their ranking. As in this case, it's Queens, Manhattan, and Brooklyn as top 3 from our results. Furthermore, we can also cluster restaurants based on type instead of proximity if the business plans to set up stores per ethnicity. Also, demographic information such as census would be an important factor in delving the insights further.

5.2 Single Store:

If the business decides to set up only one outlet and wants close proximity to all these restaurants, the centroid of centroids would serve as the optimal location. The business should do some research around that location for an affordable location in and around the premises.

6. Conclusion

The above insights provide choices of location for the food store outlet. The business must consider other factors such as leasing costs, infrastructure planning and marketing, etc. and pick the location that's practical and fruitful for growth.