

Finding the Location of NYC's Newest Bagel Shop

Melissa DiMaio

April 15, 2021

1. Introduction

1.1 Background

Any proud New Yorker can tell you that the city's classic bagel shops are like none other in the world. Let's say we'd like to contribute to this backbone of the city by opening a brand-new bagel shop somewhere in New York City. There are several things we'd want to consider before opening our shop. Firstly, we'd like the shop to be close to where we live. For this project, I chose Central Queens. Secondly, we would want to find a neighborhood which isn't too expensive. Thirdly, there would have to be enough demand for our bagels so we will have to check how many bagel shops each neighborhood already has. Using these steps, we will find the optimal neighborhood to set up our shop.

1.2 Data Used

The following data was used in this project:

- For longitude and latitude values of New York City neighborhoods, I used a JSON file which can be accessed [here](#). This file was provided by the Applied Data Science Capstone staff at coursera.org
- To define Central Queens, I used the Queens neighborhood divisions on [Wikipedia](#)
- To find average rent prices by neighborhood in Queens, I used data from [RENTCafe](#)
- Finally, to search for nearby venues, I used [Foursquare API](#)

2. Methodology

2.1 Data Cleansing

Very minimal data cleansing needed to be done. Firstly, latitude and longitude values needed to be scraped from the JSON value and put into a dataframe. Secondly, prices in the average rents table, which were presented in a CSV format, needed to be converted from object type into float type.

2.2 Neighborhood Selection

Our first concern was that we wanted our bagel shop to be close to our home in Central Queens. Our dataset of longitude and latitude values includes every neighborhood in New York City, so we'll have to narrow it down just a bit. Using the definition of Central Queens from Wikipedia and the latitude and longitude values in the dataset, we limit our dataset to only neighborhoods close to home.

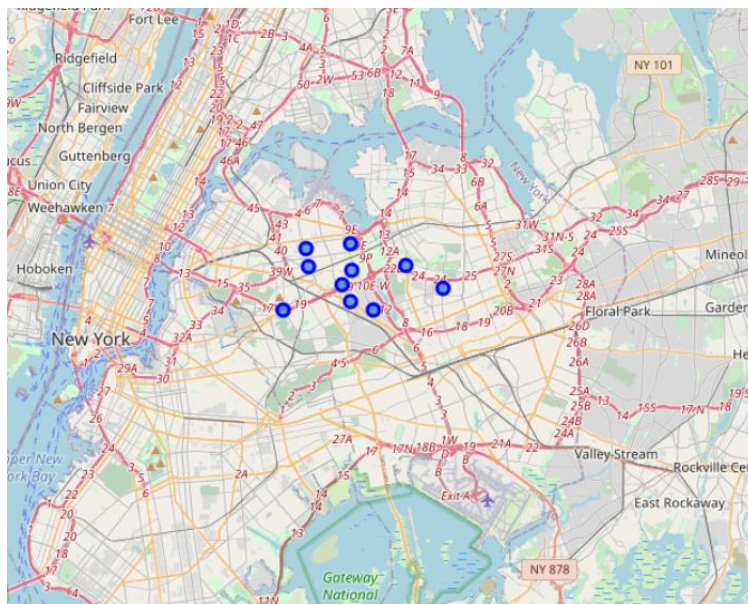


Fig 1: Map of Neighborhoods in Central Queens

Supposing we are someone of modest means, we would secondly like to look at rent prices in each neighborhood before making any decisions. Here, we must drop some neighborhoods which don't have average rent values associated with them in RentCafe's dataset. This is okay, because many neighborhood subdivisions in Queens overlap with each other. As we can see from the map below, we'll still be getting a good general survey.

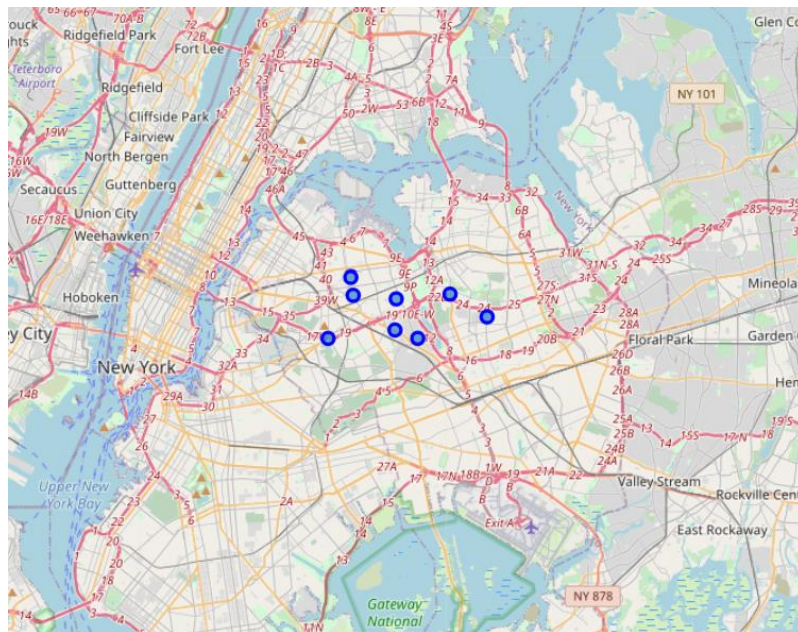
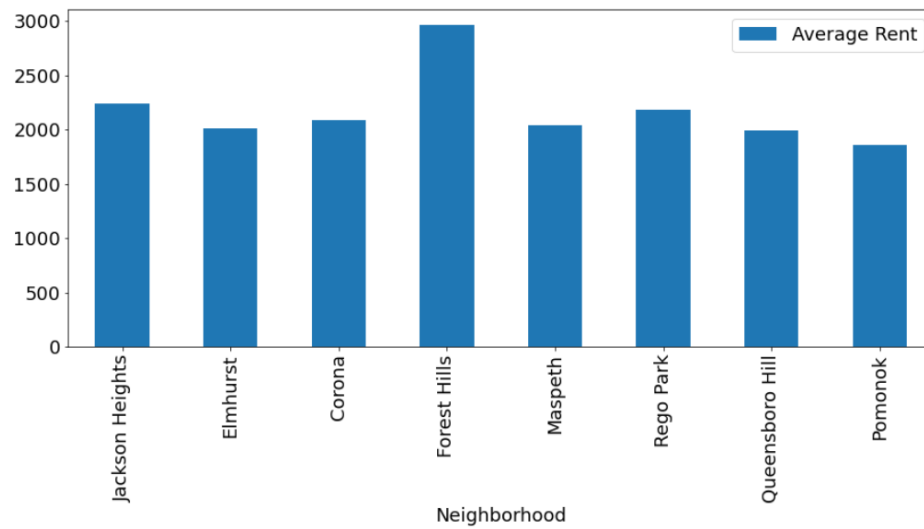


Fig 2: Map of Neighborhoods with Associated Average Rent Values

Looking at a graph of rent prices in Central Queens neighborhoods, we can see that Forest Hills is the most expensive, and maybe a bit above our price range. We will choose to eliminate Forest Hills as an option for our bagel shop.

Fig 3: Graph of Average Rent by Neighborhood

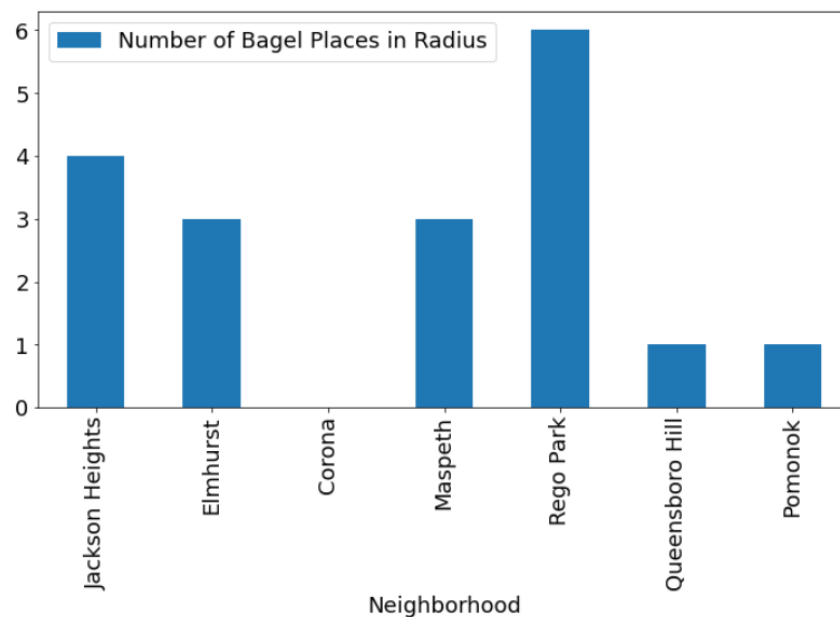


Now that we have the neighborhoods we'd like to examine, we can move on to using the Foursquare API.

2.3 Using the Foursquare API

Next, we use the Foursquare API to search for businesses with the word “bagel” in their name within a 500-foot radius of each neighborhood’s latitude and longitude values. We put these businesses into a dataframe and record the count for each neighborhood. Below are our results.

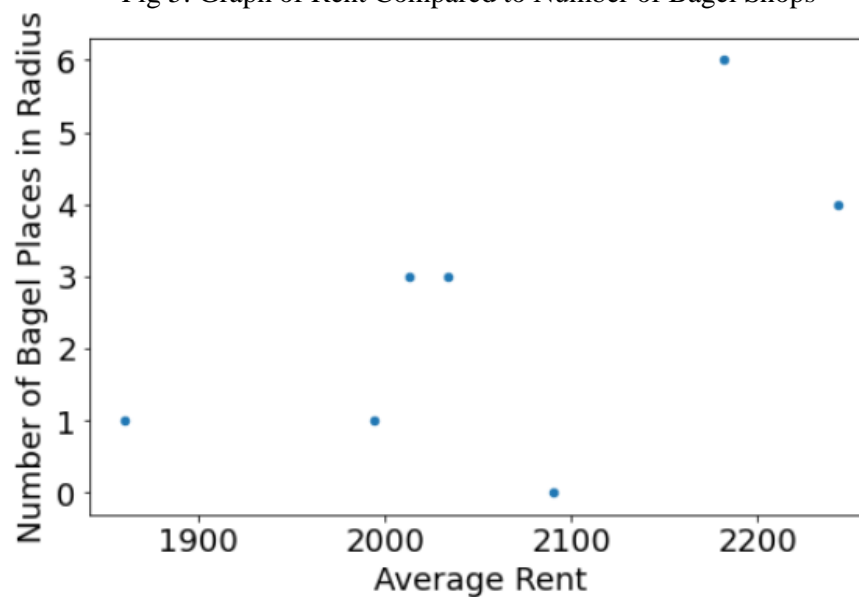
Fig 4: Graph of number of bagel places in 500 foot radius by neighborhood



2.4 Exploratory Analysis

The graph below compares average rent to the number of bagel shops. The two features appear positively correlated, that is, the higher the rent, the more bagel shops, so our answer won't be so easy.

Fig 5: Graph of Rent Compared to Number of Bagel Shops



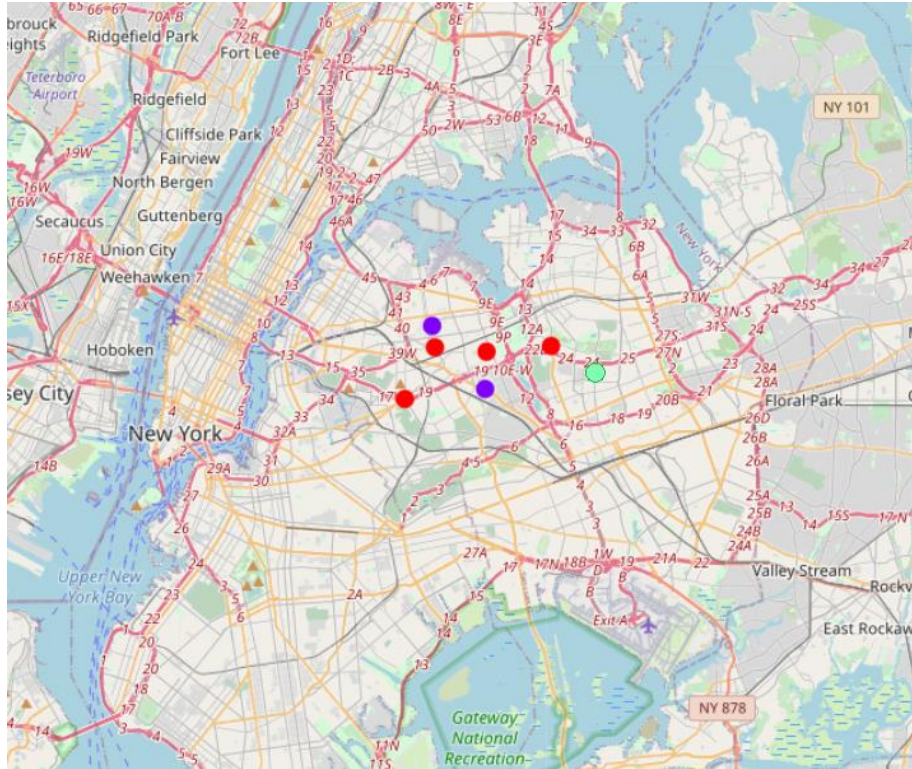
2.5 Clustering

Lastly, we use the k-means clustering algorithm to cluster neighborhoods into three clusters, using average rent price and number of bagel shops as independent variables. This technique was used because there is no one “right” answer to this problem and as such it is helpful to see the neighborhoods split into categories to choose from.

3. Results

Our k means clustering divides our candidate neighborhoods into 3 clusters based on rent and number of bagel shops. Below is a dataframe of the clusters along with the other features as well as the clusters displayed on a map.

Borough	Neighborhood	Latitude	Longitude	Average Rent	Number of Bagel Places in Radius	Cluster Labels
Queens	Jackson Heights	40.751981	-73.882821	2243.0	4	0
Queens	Elmhurst	40.744049	-73.881656	2013.0	3	2
Queens	Corona	40.742382	-73.856825	2090.0	0	2
Queens	Maspeth	40.725427	-73.896217	2034.0	3	2
Queens	Rego Park	40.728974	-73.857827	2182.0	6	0
Queens	Queensboro Hill	40.744572	-73.825809	1994.0	1	2
Queens	Pomonok	40.734936	-73.804861	1860.0	1	1



4. Discussion

How do we interpret these clusters? The purple dots, which correspond to the neighborhoods of Jackson Heights and Rego Park, are places where the number of bagel shops is the highest, but rent is also high. We could open our bagel shop here if we wanted a location where we know there is high demand and if we had enough money to open it. The red dots, which correspond to Elmhurst, Corona, Queensboro Hill, and Maspeth, are places where there are less bagel shops, but the rent is a bit lower. Finally, the green dot, which corresponds to Pomonok, is an area where there are also few bagel shops, but which has the lowest rent overall. Which cluster we choose is up to how much we want to pay.

5. Conclusion

In conclusion, there is no one best place to open a bagel shop in New York City. A business owner's choice will depend on several different factors, including proximity to home, expenses, and the number of bagel shops already in the area. Our model gives us several categories to choose from in this regard. In the future, we could improve this model could be more accurate if we tested it on more neighborhoods. Additionally, we could include other features such as taxes and other financial data, reviews of nearby bagel shops, or the history of each neighborhood to develop a more complex model.