Best Neighborhood in Brooklyn to open a Pizzeria

Keegan Pieri

12/23/2020

1. Introduction

1.1 Background

New York City, Brooklyn in particular, is known for having some of the best pizza in the world. With this reputation there can feel like an oversaturation of Pizza options in certain neighborhoods. In fact, as we will see later, there are 123 Pizza Places listed as venues in Brooklyn. Even with this large selection of Pizzerias, certain neighborhoods may have a high demand for Pizza yet few to no options nearby. Understanding which neighborhoods these are would be quite advantageous for someone looking to open up a Pizzeria.

1.2 Problem

Figure out where the supply of Pizzeria's is not meeting the demand.

1.3 Interest

This problem must be solved to make an educated decision on where to open a Pizzeria in Brooklyn

2. Data Description

2.1 Data Sources

The initial data source I used contained the coordinates of each of the 306 neighborhoods across all 5 boroughs in New York City. This dataset can be found here: https://geo.nyu.edu/catalog/nyu_2451_34572

The second data source used was from the Foursquare API, and was used to gather information about the venues in Brooklyn.

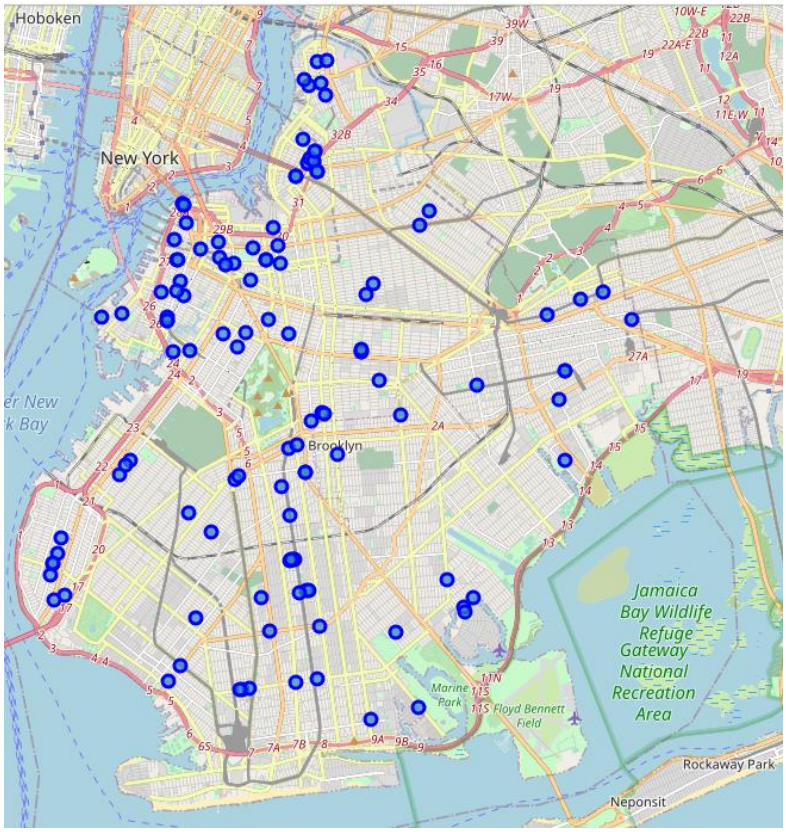
2.2 Data Retrieval & Cleaning

I began by adding column names and creating a data frame from the NYC Neighborhood dataset. I then filtered for only Neighborhoods in the borough of Brooklyn, as shown by the sample below:

:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

Next I pulled the top 10 most visited/reviewed venues for each Brooklyn Neighborhood from the Foursquare data. This data was merged with the existing data frame. Examining this data, I found there are 123 venues in Brooklyn listed as "Pizza Places". Below is a map to help visualize how these venues are spread out across the borough:



3. Methodology

3.1 Data Analysis

I began the data analysis by normalizing the venue data by taking the mean of frequency of occurrence of each category. This allowed my to then sort the top 10 venues in each Neighborhood by their frequency, 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
Bath Beach	Chinese Restaurant	Bubble Tea Shop	Pizza Place	Cantonese Restaurant	Gas Station	Donut Shop	Fast Food Restaurant	Italian Restaurant	Pharmacy	Peruvian Restaurant
Bay Ridge	Italian Restaurant	Spa	Pizza Place	Bar	American Restaurant	Greek Restaurant	Bagel Shop	Pharmacy	Sandwich Place	Hookah Bar
Bedford Stuyvesant	Coffee Shop	Bar	Café	Deli / Bodega	Pizza Place	Gift Shop	Bagel Shop	Juice Bar	Gourmet Shop	Cocktail Bar
Bensonhurst	Sushi Restaurant	Italian Restaurant	Chinese Restaurant	Ice Cream Shop	Bakery	Donut Shop	Cha Chaan Teng	Noodle House	Sporting Goods Shop	Coffee Shop
Bergen Beach	Harbor / Marina	Baseball Field	Playground	Park	Athletics & Sports	Event Service	Event Space	Factory	Falafel Restaurant	Farm

3.2 Machine Learning

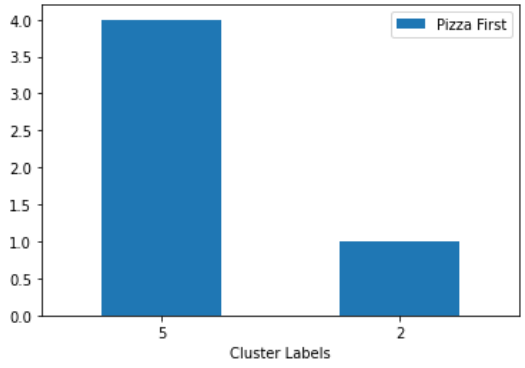
To determine which type of neighborhood has a high demand for Pizzerias I grouped the neighborhoods into like clusters; allowing me to determine which cluster had the highest preference for Pizzerias. This was done using K-Means clustering, with a K value of 7. This sorted each neighborhood into one of 5 clusters and added a column to identify the cluster, as shown below:

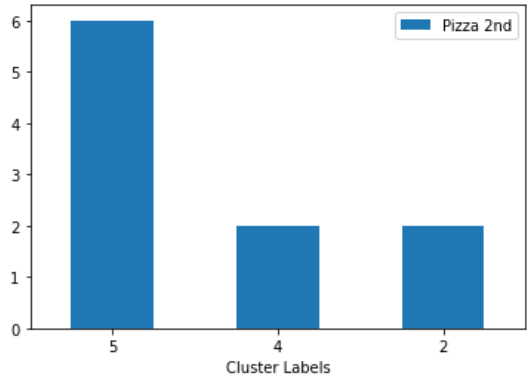
Borough	Neighborhood	Latitude	Longitude	Cluster Labels
Brooklyn	Bay Ridge	40.625801	-74.030621	5
Brooklyn	Bensonhurst	40.611009	-73.995180	5
Brooklyn	Sunset Park	40.645103	-74.010316	4
Brooklyn	Greenpoint	40.730201	-73.954241	5
Brooklyn	Gravesend	40.595260	-73.973471	5

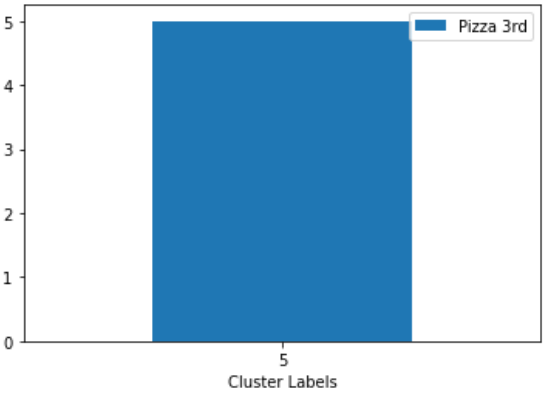
4. Results

4.1 Determine High Demand Pizza Cluster

To determine which of the 7 clusters represented neighborhoods which have a high demand for Pizza, I totaled the number of 1st, 2nd, and 3rd Most Common venues that were “Pizza Place” for each cluster. Clearly, Cluster #5 has the highest demand:



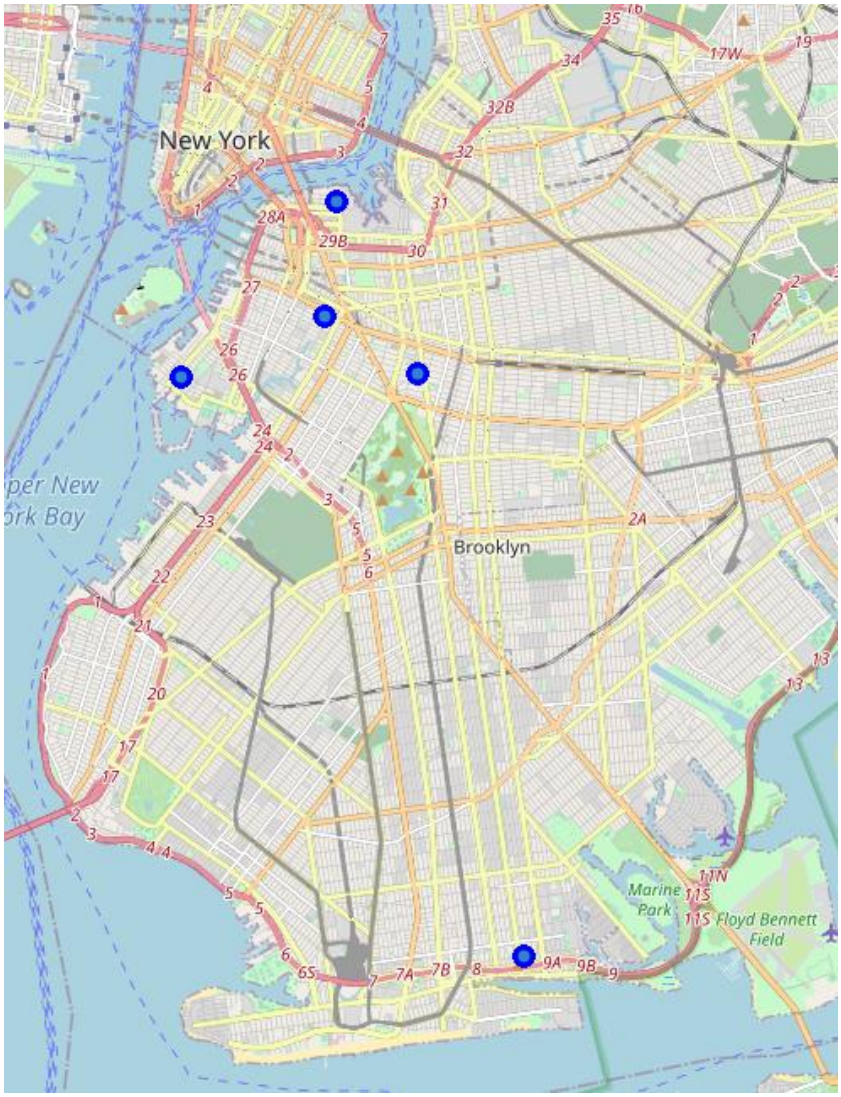




4.2 Identify Low Supply Pizza Neighborhoods

Since we know that Cluster 5 has the highest demand for Pizza, the next step was to look at the neighborhoods to determine where supply might be low. To do this I removed all Neighborhoods who had a Pizza Place as a Top ten most common venue. Additionally I removed neighborhoods which had potentially similar venues (Italian Restaurants & Deli/Bodegas). The result was the following 5 neighborhoods: Sheepshead Bay, Prospect Heights, Red Hook, Boerum Hill, & Vinegar Hill.

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
Sheepshead Bay	Dessert Shop	Turkish Restaurant	Boat or Ferry	Yoga Studio	Grocery Store	Creperie	Hotel	Diner	Restaurant	Outlet Store
Prospect Heights	Bar	Mexican Restaurant	Wine Shop	Bakery	Coffee Shop	Cocktail Bar	Diner	Ice Cream Shop	Southern / Soul Food Restaurant	Beer Bar
Red Hook	Seafood Restaurant	Art Gallery	Park	Bar	Bagel Shop	Wine Shop	Flower Shop	Ice Cream Shop	Café	American Restaurant
Boerum Hill	Coffee Shop	Dance Studio	Bar	Arts & Crafts Store	Sandwich Place	Furniture / Home Store	French Restaurant	Bakery	Spa	Chinese Restaurant
Vinegar Hill	Food Truck	Art Gallery	Coffee Shop	Café	Sandwich Place	Entertainment Service	Bakery	Scenic Lookout	Latin American Restaurant	Bike Rental / Bike Share



5. Discussion

This analysis indicates that there are 5 neighborhoods without Pizzerias in Brooklyn that are part of the cluster, as determined by KMeans, where demand for Pizza tends to be high. Given this result it seems these neighborhoods would be promising locations to open a Pizzeria.

However, additional analysis could be conducted to help determine the suitability of Brooklyn neighborhoods for a new Pizzeria. From a quantitative perspective, different machine learning methods could be used to determine like clusters. Additionally, new factors such as income, population, and ethnicity could be introduced to improve the accuracy of the clustering and determine the appropriate location for a new Pizzeria.

Additionally it is important to consider qualitative factors in addition to the quantitative analysis. For instance, the history and success of Pizzerias in these 6 neighborhoods should be examined. Locals in each neighborhood should also be consulted to opine on whether they feel a new Pizzeria would be popular in their neighborhood.

6. Conclusion

This analysis indicates that an individual looking to open a Pizzeria is likely to have success in the following neighborhoods: Sheepshead Bay, Prospect Heights, Red Hook, Boerum Hill, and Vinegar Hill. An individual should use the beforementioned additional quantitative and qualitative assessments to determine which of these 6 neighborhoods is most suitable for them to open a new Pizzeria in Brooklyn!