

THE BATTLE OF NEIGHBORHOODS

An Analysis of Neighborhoods and Venues of Brooklyn by Using Data Science Methodologies

1. Introduction

The center of Brooklyn history has started in 1863. Brooklyn is the most populous borough in New York City. It has about 2.6 million residents. Brooklyn has about 70 neighborhoods. Also, it has a multicultural structure and has 200 different languages spoken. Its population is 40 percent foreign born [1]. This diversity is reflected to Brooklyn's neighborhoods, economy and ideas. In addition, due to its history, there are major museums, libraries and art centers.

Starting a specific business in a new place is a real challenge for an entrepreneur. Venue types around the place would be a good point to start business. If there are many categorical venues with the starting business around this place, there will be competitive cases and this will compel the entrepreneur in many different aspects. In addition, population of a borough will be another issue to take care. If population of the borough is higher, then the potential of the growth of this business will be higher too. Considering these two cases together, it can be possible to find the best borough to start a new business.

Because of its population and diverse culture starting a food business in Brooklyn could be a good idea. In this project a baker wants to amplify his bakery chains in a different neighborhood Brooklyn, who has already chains in Manhattan in New York. He should choose the best place to have high income and grow his business. Therefore, in this project this issue is aimed to be solved with the best solution by using some data science methodologies.

Methodologies and solutions used in this projects could be suggestive for entrepreneurs who want to start a new business in a borough and do not have any idea about its venues, population and common places.

2. Data

Population of each borough is gathered by using web scraping method from a website [2]. It has been searched in web and only the population of main thirty six neighborhoods in Brooklyn is found. In addition, by using Geocoder Python package [3], latitude and longitude data of each neighborhood is retrieved. Also, to get venues and venue categories of each neighborhood Forsquare API [4] is

used. Moreover, to create a map of neighborhoods by using latitude and longitude values, Folium package [5] is used.

3. Methodology

3.1 Exploratory data analysis

3.1.1 Exploring neighborhoods in Brooklyn

Firstly, boroughs and neighborhoods in each borough with their latitude and longitude values in New York [6] are demonstrated via table and pinned map by using Folium Map. Then, Brooklyn is filtered from New York boroughs.

	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
5	Brooklyn	Brighton Beach	40.576825	-73.965094
6	Brooklyn	Sheepshead Bay	40.586890	-73.943186
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
8	Brooklyn	Flatbush	40.636326	-73.958401
9	Brooklyn	Crown Heights	40.670829	-73.943291
10	Brooklyn	East Flatbush	40.641718	-73.936103
11	Brooklyn	Kensington	40.642382	-73.980421
12	Brooklyn	Windsor Terrace	40.656946	-73.980073

1. Latitudes and Longitudes of Brooklyn Neighborhoods



2. Brooklyn Neighborhoods

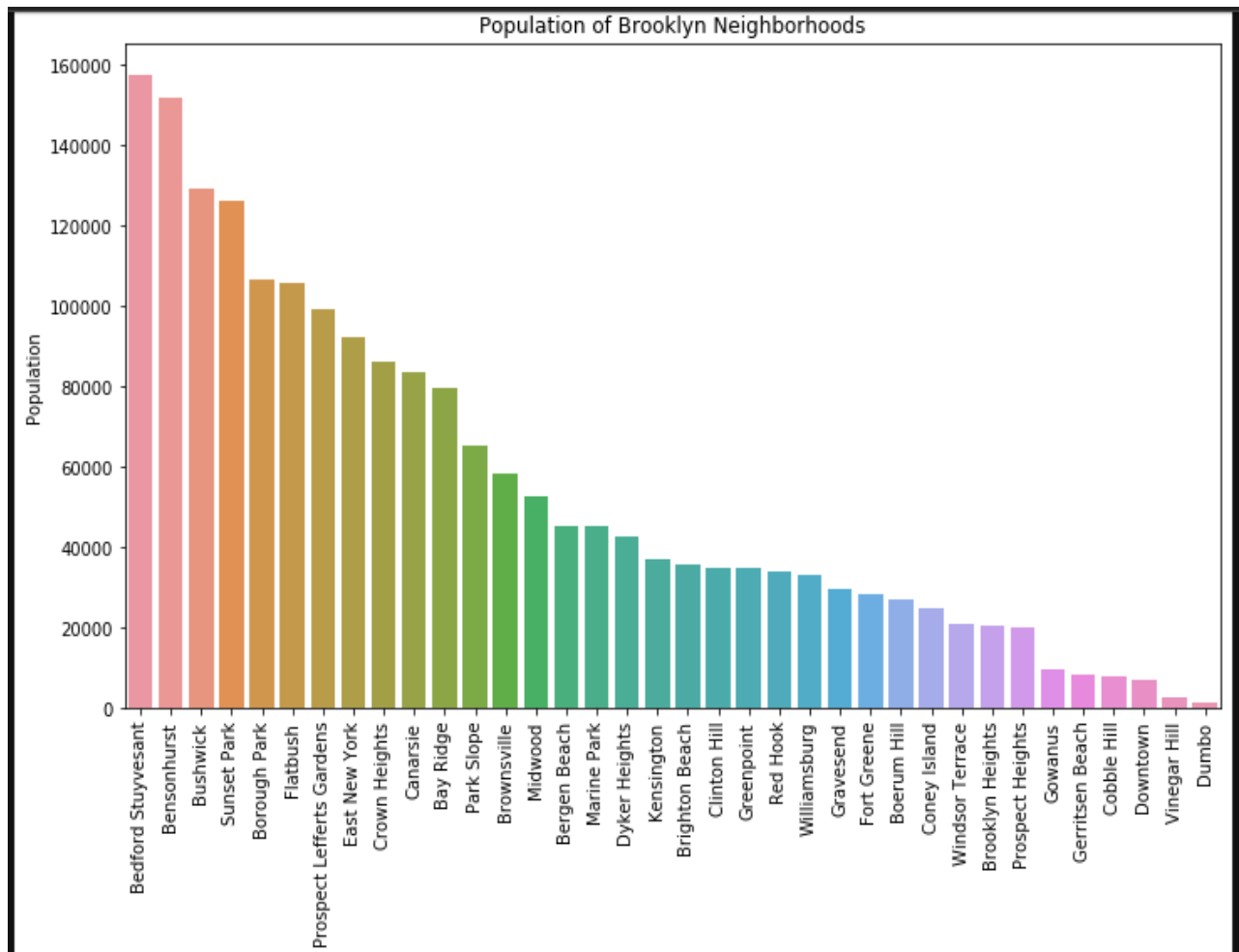
3.1.2 Explore population of neighborhoods in Brooklyn

The population of each neighborhood in Brooklyn is scrapped from the website. It is cleaned and merged with latitude and longitude values of Brooklyn neighborhoods. Then the values are sorted by population as descending to examine the populous neighborhoods in first place.

	Borough	Neighborhood	Latitude	Longitude	Population
14	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785	157530
1	Brooklyn	Bensonhurst	40.611009	-73.995180	151705
13	Brooklyn	Bushwick	40.698116	-73.925258	129239
2	Brooklyn	Sunset Park	40.645103	-74.010316	126000
24	Brooklyn	Borough Park	40.633131	-73.990498	106357
6	Brooklyn	Flatbush	40.636326	-73.958401	105804
31	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	99287
21	Brooklyn	East New York	40.669926	-73.880699	91958
7	Brooklyn	Crown Heights	40.670829	-73.943291	85886
22	Brooklyn	Canarsie	40.635564	-73.902093	83693

3. Population of each neighborhood of Brooklyn

The graph of population of neighborhoods in Brooklyn is demonstrated in figure 4.



4. Population of neighborhoods of Brooklyn

3.1.3 Explore venues in neighborhoods of Brooklyn

By using Foursquare API venues, venue latitudes, venue longitudes and venue longitudes are explored.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Georgian Dream Cafe and Bakery	40.625586	-74.030196	Caucasian Restaurant
5	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
6	Bay Ridge	40.625801	-74.030621	Ho' Brah Taco Joint	40.622960	-74.031371	Taco Place
7	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore
8	Bay Ridge	40.625801	-74.030621	Mimi Nails	40.622571	-74.031477	Spa
9	Bay Ridge	40.625801	-74.030621	A.L.C. Italian Grocery	40.623051	-74.031224	Grocery Store

5. Venues in neighborhoods of Brooklyn

3.2 One Hot Encoding to analyze each neighborhood

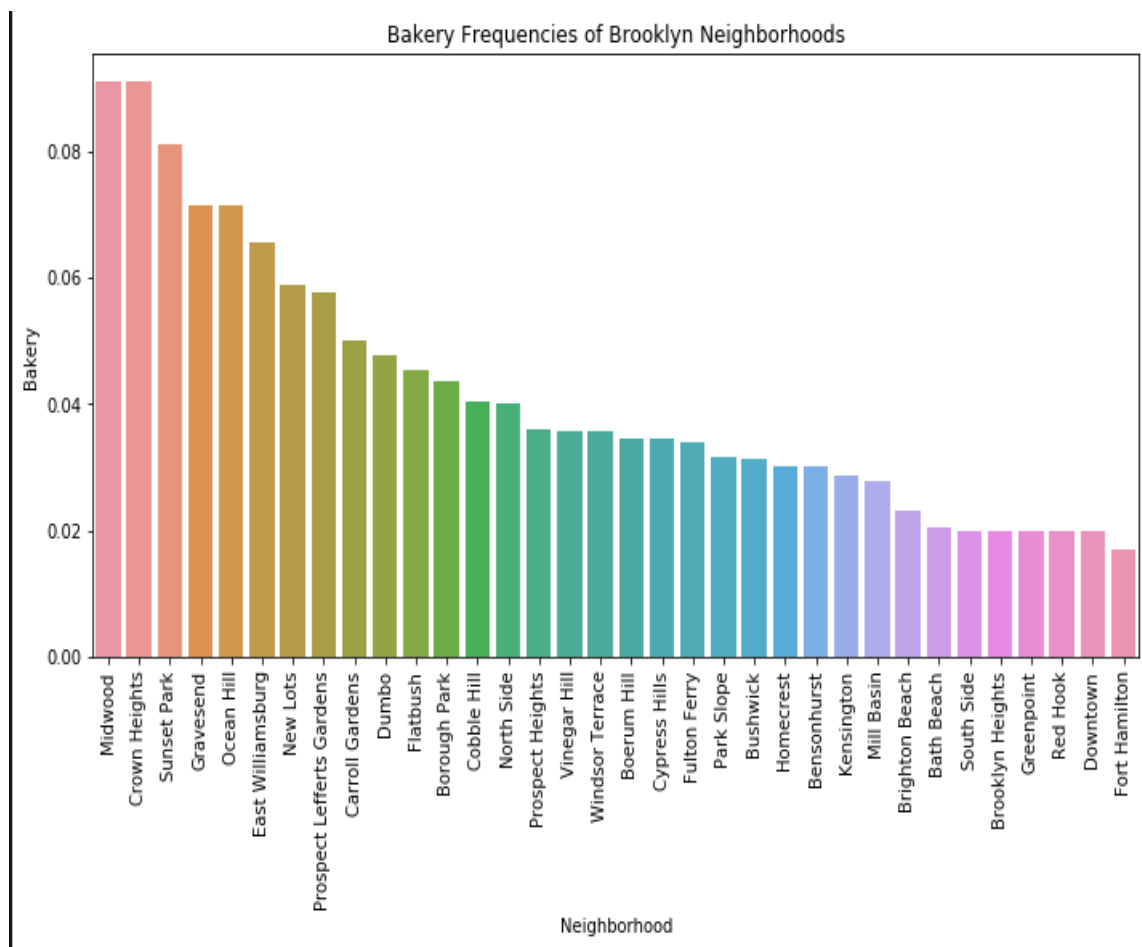
One Hot Encoding is used to analyze each neighborhood. Then, rows in One Hot Encoding table are grouped by neighborhood and by taking the mean of the frequency of occurrence of each venue category. By this analysis, frequency of each venue category clearly be examined.

	Neighborhood	Accessories Store	Adult Boutique	Airport Terminal	American Restaurant	Antique Shop	Arepa Restaurant
0	Bath Beach	0.000000	0.00	0.000000	0.000000	0.000000	0.00
1	Bay Ridge	0.000000	0.00	0.000000	0.037500	0.000000	0.00
2	Bedford Stuyvesant	0.000000	0.00	0.000000	0.000000	0.000000	0.00
3	Bensonhurst	0.000000	0.00	0.000000	0.000000	0.000000	0.00
4	Bergen Beach	0.000000	0.00	0.000000	0.000000	0.000000	0.00
5	Boerum Hill	0.000000	0.00	0.000000	0.000000	0.011494	0.00
6	Borough Park	0.000000	0.00	0.000000	0.043478	0.000000	0.00
7	Brighton Beach	0.000000	0.00	0.000000	0.000000	0.000000	0.00
8	Broadway Junction	0.000000	0.00	0.000000	0.000000	0.000000	0.00
9	Brooklyn Heights	0.000000	0.00	0.000000	0.010000	0.000000	0.00
10	Brownsville	0.000000	0.00	0.058824	0.000000	0.000000	0.00
11	Bushwick	0.000000	0.00	0.000000	0.015625	0.000000	0.00
12	Canarsie	0.000000	0.00	0.000000	0.000000	0.000000	0.00

6. Venue frequencies in each neighborhood

3.3 Filter bakery venues for each neighborhood

After analyzing venue category frequencies, bakery frequencies for each neighborhood is determined. Then, table is sorted by frequency of bakeries for each neighborhood in descending order to examine which neighborhood has the most bakery frequency and which has the least bakery frequency. Neighborhoods which have zero frequency of bakery venue are not included in the dataframe. Another dataframe is created to determine neighborhoods have no bakery venue.



7. Bakery frequency of each neighborhood

3.4 Analyze best places to start bakery business

Neighborhoods that do not have bakery frequency and have the most population would be best choices to start a new bakery. Therefore, populous neighborhoods table and neighborhoods do not have bakery venues table are merged to clearly examine the best places.

	Neighborhood	Bakery
60	Sea Gate	0.0
61	Sheepshead Bay	0.0
52	Paerdegat Basin	0.0
56	Prospect Park South	0.0
63	Starrett City	0.0
59	Rugby	0.0
66	Weeksville	0.0
58	Remsen Village	0.0
67	Williamsburg	0.0
47	Mill Island	0.0
51	Ocean Parkway	0.0

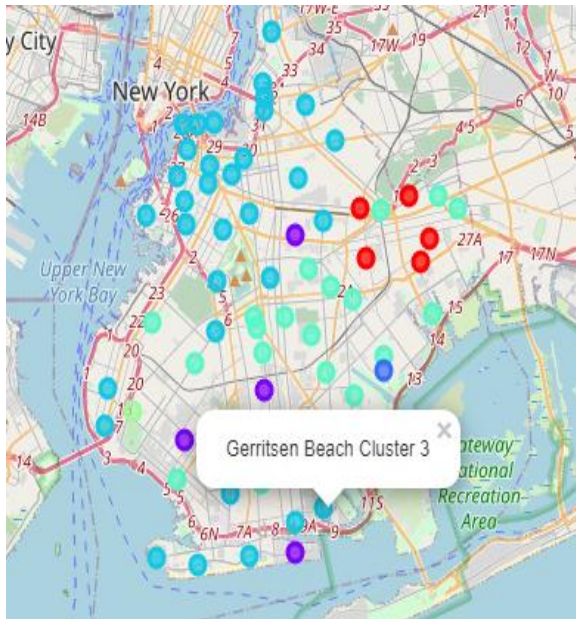
Neighborhood	Population	Bakery Frequency	Borough	Latitude	Longitude
Bedford Stuyvesant	157530	0.0	Brooklyn	40.687232	-73.941785
Borough Park	106357	0.0	Brooklyn	40.633131	-73.990498
Flatbush	105804	0.0	Brooklyn	40.636326	-73.958401
East New York	91958	0.0	Brooklyn	40.669926	-73.880699
Canarsie	83693	0.0	Brooklyn	40.635564	-73.902093
Bay Ridge	79371	0.0	Brooklyn	40.625801	-74.030621
Brownsville	58300	0.0	Brooklyn	40.663950	-73.910235
Bergen Beach	45231	0.0	Brooklyn	40.615150	-73.898556
Marine Park	45231	0.0	Brooklyn	40.609748	-73.931344
Dyker Heights	42419	0.0	Brooklyn	40.619219	-74.019314
Clinton Hill	34791	0.0	Brooklyn	40.693229	-73.967843
Williamsburg	32926	0.0	Brooklyn	40.707144	-73.958115
Fort Greene	28335	0.0	Brooklyn	40.688527	-73.972906
Coney Island	24711	0.0	Brooklyn	40.574293	-73.988683
Gowanus	9450	0.0	Brooklyn	40.673931	-73.994441
Gerritsen Beach	8353	0.0	Brooklyn	40.590848	-73.930102

8. Neighborhoods have no bakery

9. Best places to open bakery

3.5 Cluster neighborhoods by their venue category frequencies (K-Clustering)

K-means is used to cluster the neighborhood into 8 clusters. Clusters are created by using venue category frequencies for each neighborhood. As a result, neighborhoods have similar properties in venue category are determined. On figure 10 each cluster is demonstrated by different colors in the map by using Folium package. Moreover, neighborhoods in Cluster 1 table is demonstrated on figure 11. When each cluster is examined, neighborhoods in Cluster 1 in red color, Cluster 3 in blue color, Cluster 6 in light green color, Cluster 7 in yellow color, and Cluster 8 in orange color are most appropriate places to start a new bakery. Because in this clusters bakery is not in the 10th most common venue category. So, opening new bakery in these neighborhoods would make easier to grow this business because if there is less bakery there would be less rivalry to make more sale.



10 Cluster map

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
14	Brownsville	Fried Chicken Joint	Chinese Restaurant	Playground
26	East New York	Deli / Bodega	Fried Chicken Joint	Plaza
43	Ocean Hill	Deli / Bodega	Fried Chicken Joint	Bus Stop
58	New Lots	Grocery Store	Pizza Place	Fried Chicken Joint
67	Highland Park	Deli / Bodega	Latin American Restaurant	Fried Chicken Joint

11 Cluster 1

4. Result and Discussion

Some data analysis are made to find best neighborhoods in Brooklyn to start a new bakery business that has chains in Manhattan. To determine best place two cases are considered, which are;

- population of the neighborhood and
- bakery frequency in that neighborhood.

If the neighborhood has more population, possibility of more sale in bakery would increase. In addition, if the bakery frequency of the neighborhood is less, the competition ratio between other bakeries will decrease. As a result, this would increase possibility of sale ratio and make owner determine adjustable prices for the products.

The population of each neighborhood is scrapped and sorted by population to find most populous neighborhoods. After examining venue categories in neighborhoods, bakery frequencies are found in each neighborhood. Then neighborhoods that have zero bakery frequency are determined. Afterwards, neighborhoods that have zero bakery and have more population are combined and 16 neighborhoods are found in below.

Neighborhood	Population
Bedford Stuyvesant	157530
Borough Park	106357
Flatbush	105804
East New York	91958
Canarsie	83693
Bay Ridge	79371
Brownsville	58300
Bergen Beach	45231
Marine Park	45231
Dyker Heights	42419
Clinton Hill	34791
Williamsburg	32926
Fort Greene	28335
Coney Island	24711
Gowanus	9450
Gerritsen Beach	8353

These neighborhoods could be the best places to start new bakery business.

In addition to expand the number of best place choices to open a new bakery, K-Means Clustering is applied to data. After finding venue category frequencies in each neighborhood in Brooklyn, similar neighborhoods are grouped by their ten most common venue categories. Therefore, the clusters which have the least bakery frequencies could be best places to start new bakery business. Consequently, five clusters are found as promising places. They are listed below.

Cluster 1	Cluster 3	Cluster 6	Cluster 7	Cluster 8
Brownsville	Paerdegat Basin	Dyker Heights	Bergen Beach	Mill Island
East New York				
Ocean Hill				
New Lots				
Highland Park				

When two results are examined, the neighborhoods which are;

- Ocean Hill
- New Lots
- Highland Park
- Paerdegat Basin
- Mill Island

can be added to best places list. However, population data could not be found for these neighborhoods. In the end, 21 neighborhoods could be the most appropriate neighborhoods to start a new bakery for the owner. Expanded list is in below.

Neighborhood	Population
Bedford Stuyvesant	157530
Borough Park	106357
Flatbush	105804
East New York	91958
Canarsie	83693
Bay Ridge	79371
Brownsville	58300
Bergen Beach	45231
Marine Park	45231
Dyker Heights	42419
Clinton Hill	34791
Williamsburg	32926
Fort Greene	28335
Coney Island	24711
Gowanus	9450
Gerritsen Beach	8353
Ocean Hill	-
New Lots	-
Highland Park	-
Paerdegat Basin	-
Mill Island	-

In this project, best neighborhoods to start a bakery are determined. However, it is not implied that these are actually optimal places. The information about neighborhoods is provided to bakers who want to start a new bakery in a neighborhood that the owner does not have enough information. While examining the population of neighborhoods in Brooklyn, only the 36 main crowded neighborhoods are analyzed. Because the population of all neighborhoods could not be reached. For the future projects, population of all neighborhoods could be evaluated for starting a new bakery. Moreover, best neighborhoods can be suggested for another starting business fields and also after determining city center, potential zones for starting a business can be suggested with respect to their distances to the center.

5. Conclusion

Aim of this project was to suggest optimal neighborhoods to start a new bakery in Brooklyn. To achieve this goal, two criteria are taken care of which are population and the frequency of bakeries in that neighborhood. If a neighborhood that is the most populous and has the least number of bakery could be an optimal place to start this business. After gathering population data of 36 main neighborhoods, the venue types are found for each neighborhood and bakery frequencies are discovered. Then by merging two data, 16 optimal places are revealed. In addition, to have more options of optimal neighborhoods, K-Means Clustering is used. By using this analyze, 10 most common venues are found for each neighborhood and they are clustered. The clusters that does not have bakery venues in 10 most common venues are chosen as optimal neighborhoods. And five more neighborhoods are added to the optimal neighborhoods list. As a result 21 neighborhoods are revealed as optimal places to start a new bakery business.

Bibliography

[1] *About Brooklyn*: URL: <https://www.brooklyn-usa.org/about-brooklyn/>

[2] *Population of Brooklyn Neighborhoods*. URL: <https://www.worldatlas.com/articles/brooklyn-neighborhoods-by-population.html>

[3] *Geocoder*. URL: <https://geocoder.readthedocs.io/index.html>

[4] *Foursquare*: URL: <https://foursquare.com/>

[5] *Folium Map*: URL: <https://pypi.org/project/folium/>

[6] *New York Boroughs*: URL: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json