

## Capstone - The Battle of Neighborhoods (Week 2)

### REPORT

#### INTRODUCTION

This is a capstone project for IBM Data Science Professional Certificate. In this project, we will process a hypothetical scenario for opening a new bookstore in Brooklyn/New York City. Bookstores always have an important role for people, especially for educated people and students in large cities like NYC. There are many bookstores in NYC and serious considerations are required to open a new bookstore than it seems. Finding the best location to open a bookstore is one of the most important decisions for investors, and we will complete this project to help stakeholders find the most suitable location.

#### BUSINESS PROBLEM

The objective of this project is to find the most suitable location for the investor to open a new bookstore in Brooklyn/NYC. By using data analysis and tools, and machine learning algorithms such as clustering, this project aims to provide solutions to answer the question; ***“If an entrepreneur wants to open a new bookstore, what would be the best location for it in Brooklyn?”***

#### TARGET AUDIENCE

The investor who wants to open a new bookstore in Brooklyn/NYC.

#### DATA

To find the best solution, we will need the following data;

- List of neighborhoods in New York City.
- Latitude and longitude coordinates of these neighborhoods.
- Venue data related to bookstores. This will help us find the neighborhoods which are more suitable to open a bookstore.

#### EXTRACTING THE DATA

- Scrapping data of NYC neighborhoods from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- Getting latitude and longitude data of NYC neighborhoods via Geocoder library
- Using Foursquare API to get venue data related to NYC neighborhoods

## METHODOLOGY

This project applies various data analysis techniques like;

- Get the data from open sources
- Working with Foursquare API
- Data cleaning
- Data wrangling
- Machine Learning (K-means Clustering)
- Map Visualization

We use the data of NYC neighborhoods from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

This source has 5 boroughs and 306 neighborhoods of New York City as well as the latitude and longitude data of each neighborhood.

```
neighborhoods.head()
```

```
[6]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Let's see how many neighborhoods have each borough

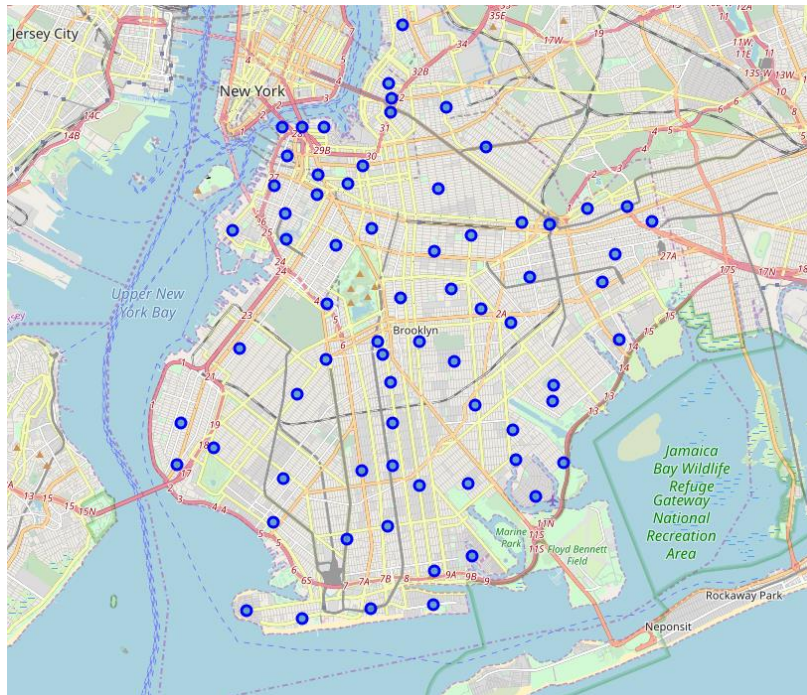
```
[7]: neighborhoods['Borough'].value_counts()
```

```
[7]: Queens      81  
     Brooklyn   70  
     Staten Island 63  
     Bronx      52  
     Manhattan   40  
     Name: Borough, dtype: int64
```

In this project, we focus on the Brooklyn area, and we can see the Brooklyn area has 70 neighborhoods.

We get the geographical coordinates of these neighborhoods by using Geocoder Python package. This map performs a simple validation of correctness of our data.

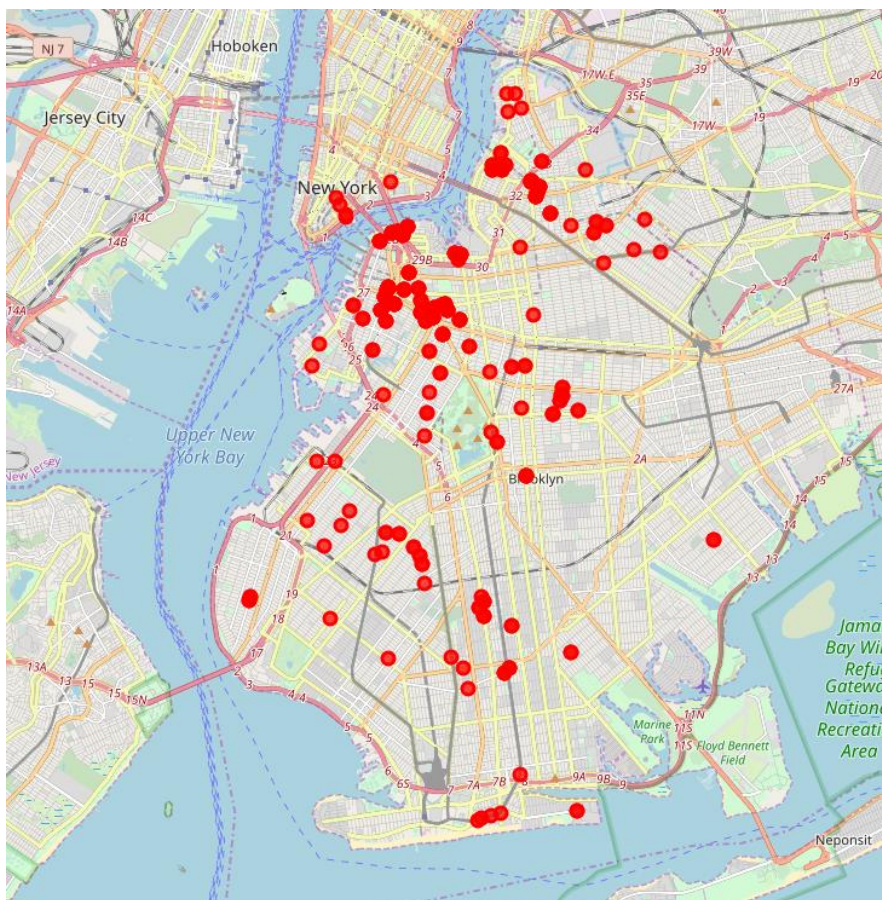
In this map, we can see the neighborhoods in Brooklyn;



We use Foursquare API to the venues (bookstores) of these neighborhoods in Brooklyn.

```
[21]: neighborhoods = neighborhoods[neighborhoods['Borough'] == 'Brooklyn'].reset_index(drop=True)
newyork_venues_bookstore = getNearbyVenues(names=neighborhoods['Neighborhood'], latitudes=neighborhoods['Latitude'], longitudes=neighborhoods['Longitude'], radius=1000, categoryIds='4bf58dd8d848988d114951735')
newyork_venues_bookstore.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore
1	Bay Ridge	40.625801	-74.030621	Book Mark Shop	40.623848	-74.031059	Bookstore
2	Bensonhurst	40.611009	-73.995180	Excilbur Comics	40.611408	-73.991451	Bookstore
3	Sunset Park	40.645103	-74.010316	Koch Comics Warehouse	40.653740	-74.011697	Bookstore
4	Sunset Park	40.645103	-74.010316	IC Christian Bookstore	40.640991	-74.014412	Bookstore





We use this feature to group the neighborhoods into K-means clustering algorithm. And also, the Folium library to visualize the neighborhoods in Brooklyn and its emerging clusters.

```
[30]: num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Brooklyn_grouped['Neighborhood']

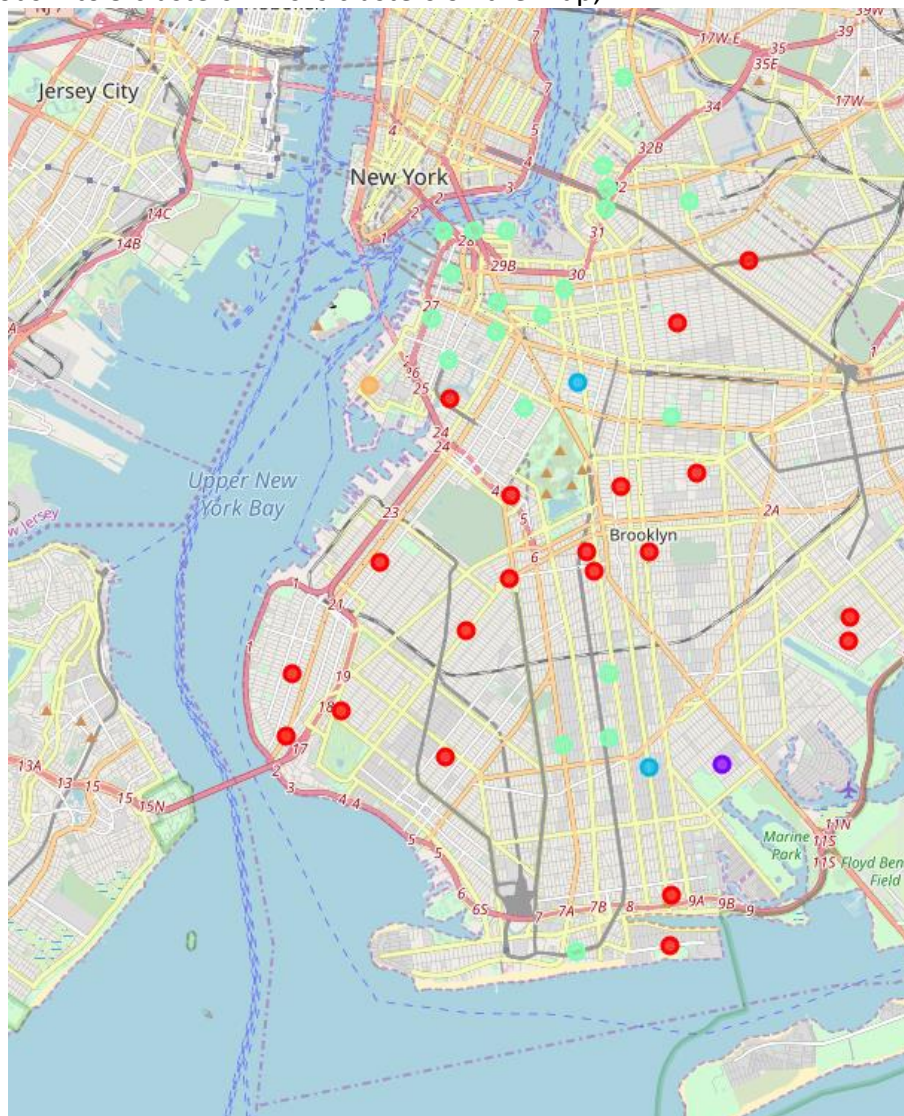
for ind in np.arange(Brooklyn_grouped.shape[0]):
    neighborhood_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Brooklyn_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

	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
0	Bay Ridge	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
1	Bedford Stuyvesant	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
2	Bensonhurst	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
3	Boerum Hill	Bookstore	Gift Shop	Newsstand	Cosmetics Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library
4	Borough Park	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store

## RESULTS

The results from K-means clustering show that we can categorize Brooklyn neighborhoods into 5 clusters. This is clusters on the map;



## Cluster 0

[37]:	Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 0, Brooklyn_merged.columns[[1] + list(range(5, Brooklyn_merged.shape[1]))]]										
[37]:	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
0	Bay Ridge	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
1	Bensonhurst	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
2	Sunset Park	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
6	Sheepshead Bay	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
11	Kensington	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
12	Windsor Terrace	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
16	Bushwick	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
17	Bedford Stuyvesant	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
22	Gowanus	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
28	Canarsie	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
31	Manhattan Beach	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
34	Borough Park	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
35	Dyker Heights	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
42	Prospect Lefferts Gardens	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
47	Prospect Park South	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
53	Fort Hamilton	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
54	Ditmas Park	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
55	Wingate	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
59	Paerdegat Basin	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
69	Erasmus	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store

## Cluster 1

[38]:	Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 1, Brooklyn_merged.columns[[1] + list(range(5, Brooklyn_merged.shape[1]))]]										
[38]:	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
37	Marine Park	Used Bookstore	Video Store	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store	Cosmetics Shop

## Cluster 2

[39]:	Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 2, Brooklyn_merged.columns[[1] + list(range(5, Brooklyn_merged.shape[1]))]]										
[39]:	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
13	Prospect Heights	Bookstore	Used Bookstore	Coffee Shop	Video Store	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop
68	Madison	Bookstore	Used Bookstore	Video Store	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store

## Cluster 3

[40]:	Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 3, Brooklyn_merged.columns[[1] + list(range(5, Brooklyn_merged.shape[1]))]]										
[40]:	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
3	Greenpoint	Bookstore	Restaurant	Video Store	Used Bookstore	Toy / Game Store	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
5	Brighton Beach	Bookstore	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store
7	Manhattan Terrace	Bookstore	Miscellaneous Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Library	Gift Shop	Furniture / Home Store
9	Crown Heights	Bookstore	Coffee Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop
15	Williamsburg	Bookstore	Furniture / Home Store	Comic Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library
18	Brooklyn Heights	Bookstore	Newsstand	Gift Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library	Furniture / Home Store
19	Cobble Hill	Bookstore	Gift Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Furniture / Home Store
20	Carroll Gardens	Bookstore	Gift Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Furniture / Home Store
23	Fort Greene	Bookstore	Cosmetics Shop	Coffee Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library
24	Park Slope	Bookstore	Used Bookstore	Gift Shop	Video Store	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Furniture / Home Store
38	Clinton Hill	Bookstore	Coffee Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop
40	Downtown	Bookstore	Newsstand	Gift Shop	Cosmetics Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library
41	Boerum Hill	Bookstore	Gift Shop	Newsstand	Cosmetics Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library
46	Midwood	Bookstore	Miscellaneous Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Library	Gift Shop	Furniture / Home Store
49	East Williamsburg	Bookstore	Library	Furniture / Home Store	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Gift Shop
50	North Side	Bookstore	Toy / Game Store	Furniture / Home Store	Comic Shop	Video Store	Used Bookstore	Restaurant	Newsstand	Miscellaneous Shop	Library
51	South Side	Bookstore	Toy / Game Store	Furniture / Home Store	Comic Shop	Video Store	Used Bookstore	Restaurant	Newsstand	Miscellaneous Shop	Library
52	Ocean Parkway	Bookstore	Miscellaneous Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Library	Gift Shop	Furniture / Home Store
61	Fulton Ferry	Bookstore	Comic Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop
62	Vinegar Hill	Bookstore	Newsstand	Coffee Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library	Gift Shop
65	Dumbo	Bookstore	Newsstand	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Miscellaneous Shop	Library	Gift Shop	Furniture / Home Store

## Cluster 4

[41]:	Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 4, Brooklyn_merged.columns[[1] + list(range(5, Brooklyn_merged.shape[1]))]]										
[41]:	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
21	Red Hook	Bookstore	Bike Shop	Video Store	Used Bookstore	Toy / Game Store	Restaurant	Newsstand	Miscellaneous Shop	Library	Gift Shop

### **RECOMMENDATIONS**

In this project we only consider the factor of frequency of occurrences of venues to categorize neighborhoods. And this analysis is performed on limited data. It can be augmented by many other factors (e.g. other statistical measures like variance, various percentiles etc. of this number, colleges, schools, public transportation etc.) We are also relying on the data from Foursquare for our analysis.

Under these conditions, from our cluster analysis, it looks that Cluster 1 and Cluster 4 areas are the best choices to open a new bookstore. We would recommend **Red Hook** and **Marine Park** neighborhoods as a potential place for the new bookstore.