Clustering Safest Neighborhoods

ADSC – Final Project Report

By: Harsh Vardhan Jaiswal

20th June 2020

Capstone Project – Battle of Neighborhoods (Week 1)

1) Introduction

1.1 Background

According to the data from U.S. Census Bureau, an average person in U.S. moves more than 11 times in his/her lifetime. This shows that people are not really satisfied with the place where they decide to live. Hence it is really important for people to choose their neighborhoods carefully after doing a lot research, so that they are satisfied in where they decide to move in. When we are looking for living in a new place the first concern that most of the people have is the safety. The neighborhood that is not safe and has a high crime rate is not really fit for living and has large scale implications on people's lives.

Therefore keeping this principle in mind I have prepared this project which will help people to make this important decision and will help them to save a lot of time and money.

1.2 Problem

The problem is to determine the safest neighborhoods in New York for living. This will be determined by using the dataset that is freely available at the official website of New York City. The link of the data is here (NYC crime data). As this is a very large dataset containing records for more than 1 year we will be taking the data for the first quarter of the year 2020 i.e. from Jan 2020 – Mar 2020. Along with this dataset I will also be using the New York geo JSON file which was provided in one of the Labs of this course (New York geo json). This file contains all Neighborhoods, Boroughs and their geographical coordinates.

After processing the Crime data I will be merging it with the neighborhood dataframe created from the above geojson file. And then finally I will be working on this combined dataset.

The aim of this project is to first find the safest neighborhoods and then by using Foursquare location dataset determine the 10 most common venues in each neighborhood and then finally dividing it into clusters using k-means clustering algorithm

1.3 Interested Audience

This model will be useful to the people who are looking to re-locate or settle in New York City. It can also help the real estate agents or companies to find the best neighborhoods for their clients to live in.

2) Data Acquisition, Cleaning and Processing

2.1 Data Acquisition

a) New York Crime Data

The crime data for the New York City is freely available on the official site of New York City. As the original dataset has data for many previous years, we will be taking the most recent data i.e. the first quarter of 2020 (Jan 2020 to Mar 2020). There are a lot of columns in this dataset hence the columns that we will be using are highlighted with Yellow color. The data set will be containing the following columns:

Column Name	Column Description
CMPLNT_NUM	Randomly generated persistent ID for each complaint
ADDR_PCT_CD	The precinct in which the incident occurred
BORO	The name of the borough in which the incident occurred
CMPLNT_FR_DT	Exact date of occurrence for the reported event (or starting date of occurrence, if CMPLNT_TO_DT exists)
CMPLNT_FR_TM	Exact time of occurrence for the reported event (or starting time of occurrence, if CMPLNT_TO_TM exists)
CMPLNT_TO_DT	Ending date of occurrence for the reported event, if exact time of occurrence is unknown
CMPLNT_TO_TM	Ending time of occurrence for the reported event, if exact time of occurrence is unknown
CRM_ATPT_CPTD_CD	Indicator of whether crime was successfully completed or attempted, but failed or was interrupted prematurely
HADEVELOPT	Name of NYCHA housing development of occurrence, if applicable
HOUSING_PSA	Development Level Code
JURISDICTION_CODE	Jurisdiction responsible for incident. Either internal, like Police(0), Transit(1), and Housing(2); or external(3), like Correction, Port Authority, etc.
JURIS_DESC	Description of the jurisdiction code
KY_CD	Three digit offense classification code
LAW_CAT_CD	Level of offense: felony, misdemeanor, violation
LOC_OF_OCCUR_DESC	Specific location of occurrence in or around the premises; inside, opposite of, front of, rear of
OFNS_DESC	Description of offense corresponding with key code
PARKS_NM	Name of NYC park, playground or greenspace of occurrence, if applicable (state parks are not included)
PATROL_BORO	The name of the patrol borough in which the incident occurred
PD_CD	Three digit internal classification code (more granular than Key Code)
PD_DESC	Description of internal classification corresponding with PD code (more granular than Offense Description)

PREM_TYP_DESC	Specific description of premises; grocery store, residence, street, etc.
RPT_DT	Date event was reported to police
STATION_NAME	Transit station name
SUSP_AGE_GROUP	Suspect's Age Group
SUSP_RACE	Suspect's Race Description
SUSP_SEX	Suspect's Sex Description
TRANSIT_DISTRICT	Transit district in which the offense occurred.
VIC_AGE_GROUP	Victim's Age Group
VIC_RACE	Victim's Race Description
VIC_SEX	Victim's Sex Description
X_COORD_CD	X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Y_COORD_CD	Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Latitude	Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

b) Neighborhood Dataset

The neighborhood dataset for is prepared using the New York geo json file . This dataset contains the list of all the neighborhoods as per boroughs and their coordinates.

Column Name	Column Description
Borough	Name of the Borough
Neighborhood	Name of the Neighborhood
Latitude	Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

2.2 Data Cleaning and Processing

Before we can merge the Crime dataset with the Neighborhood data we need to do some data cleaning and processing by following these steps:

1) Remove all the data that is not in the first quarter of the year 2020 or has missing information.
2) As you can see that the above column list for Crime data does not have a Neighborhood column. Hence we need to add a Neighborhood column before we can merge the 2 datasets.

In order to do so we need to use GeoPy API and pass the Longitude and Latitude values in it to find the Neighborhood for each row. And then merge it with Crime dataset using Location as the key.

[22]:

	Locations	Neighborhoods
0	(40.65699087900003, -73.87657444799999)	East New York
1	(40.67458330800008, -73.93022154099998)	Eastern Parkway
2	(40.817877907000025, -73.91695668199996)	Melrose
3	(40.75201086000004, -73.93587196099996)	Sunnyside
4	(40.81477097700008, -73.92511075099996)	Mott Haven

Figure 1: Neighborhood dataset generated from GeoPy API

3) Drop the irrelevant columns from the dataset.
4) Prepare a pivot table based on the type of crime. In U.S. the crime is divided into 3 levels i.e. Felony, Misdemeanor and Violation

[29]:

	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION
0	Alphabet City	MANHATTAN	25.0	48.0	27.0
1	Annadale	STATEN ISLAND	3.0	12.0	2.0
2	Arlington	STATEN ISLAND	21.0	63.0	20.0
3	Arrochar	STATEN ISLAND	2.0	9.0	2.0
4	Arverne View	QUEENS	37.0	101.0	38.0

Figure 2: Crime data after cleaning and processing

The second data i.e. Neighborhood Data prepared from the geo json file of the New York City. In order to use for analysis we need to convert it into a data frame.

[8]:

	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

Figure 3: Neighborhood data after processing

After this we merge the above Crime dataset with the Neighborhood data we get the following the dataset:

[31]:

	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude
0	ANNADALE	STATEN ISLAND	8	20	8	36	40.538114	-74.178549
1	ARLINGTON	STATEN ISLAND	21	63	20	104	40.635325	-74.165104
2	ARROCHAR	STATEN ISLAND	2	9	2	13	40.596313	-74.067124
3	ARVERNE	QUEENS	39	104	38	181	40.589144	-73.791992
4	ASTORIA	QUEENS	55	133	41	229	40.768509	-73.915654

Figure 4: Final dataset after merging Crime and Neighborhood data

Using the above dataset we can find the safest neighborhood in New York City which is as follows:

	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude
69	EDGEWATER PARK	BRONX	0	1	0	1	40.821986	-73.813885
190	SOUTH SIDE	BROOKLYN	0	1	0	1	40.710861	-73.958001
37	CHARLESTON	STATEN ISLAND	2	1	0	3	40.530531	-74.232158
39	CHELSEA	STATEN ISLAND	0	1	3	4	40.594726	-74.189560
109	HUGUENOT	STATEN ISLAND	0	3	1	4	40.531912	-74.191741

Figure 5: Safest neighborhoods

The above data will be used to generate the Venues for each neighborhood using the Foursquare API.

3) Methodology

3.1 Exploratory Data Analysis

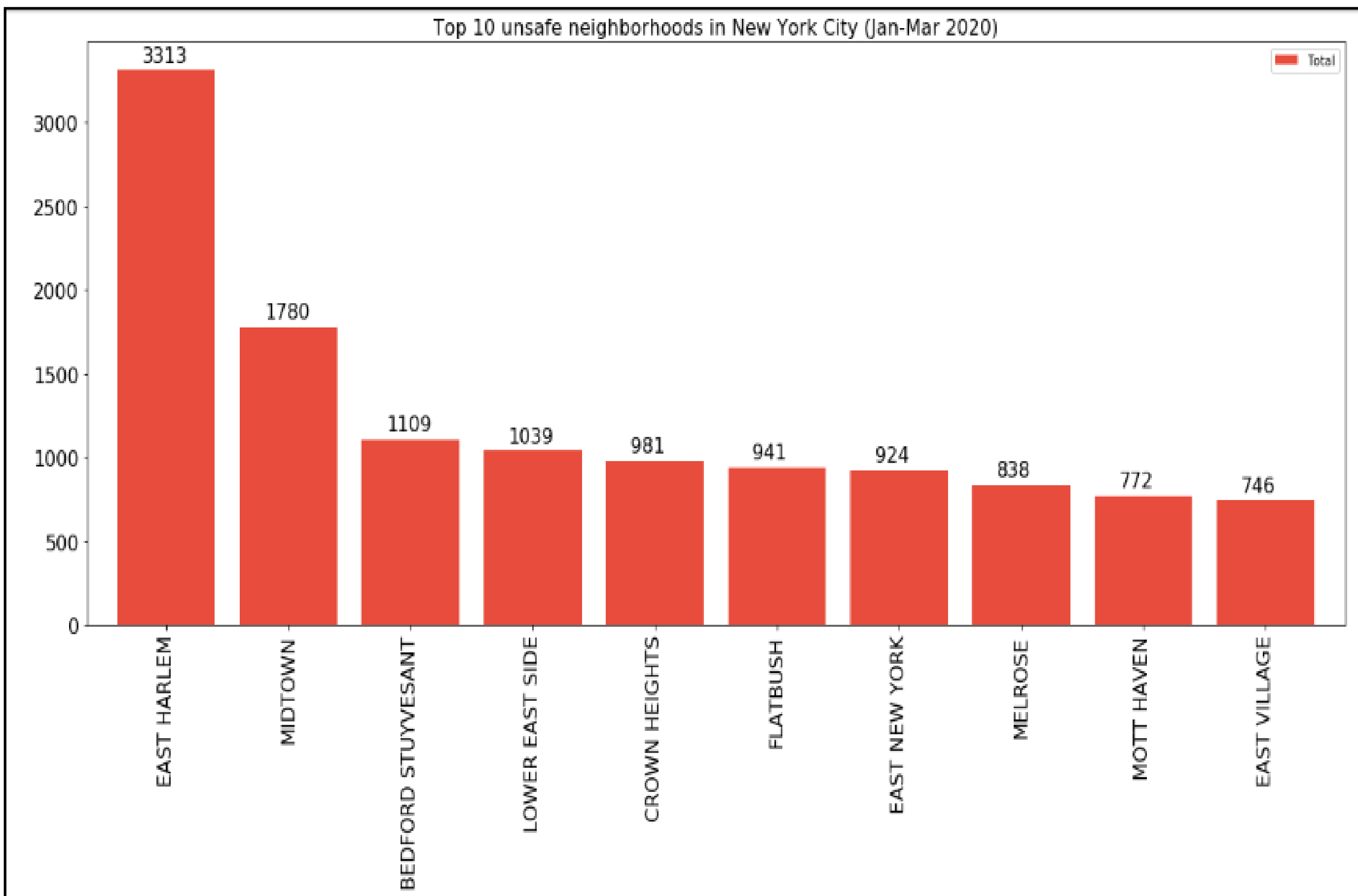
3.1.1. Statistical summary of crimes

In order to use prepare the descriptive statistics for the crime data I have used the describe function in Python. This function gives us the mean, standard deviation, minimum, 25th Percentile, Median (50th Percentile), 75th Percentile and maximum values associated with no. of Felonies, Misdemeanor, Violations and Total no. of crimes reported in the first quarter of 2020 in the city of New York.

	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude
count	231.000000	231.000000	231.000000	231.000000	231.000000	231.000000
mean	66.857143	114.844156	35.748918	217.450216	40.704096	-73.952877
std	98.386865	171.142813	52.060226	315.964536	0.097375	0.121677
min	0.000000	1.000000	0.000000	1.000000	40.505334	-74.246569
25%	8.000000	16.500000	6.500000	33.500000	40.622638	-74.010491
50%	33.000000	67.000000	20.000000	130.000000	40.705179	-73.944182
75%	85.500000	155.500000	44.000000	286.500000	40.769667	-73.863736
max	901.000000	1855.000000	557.000000	3313.000000	40.898273	-73.715481

3.1.2 Neighborhoods with the highest crime rates

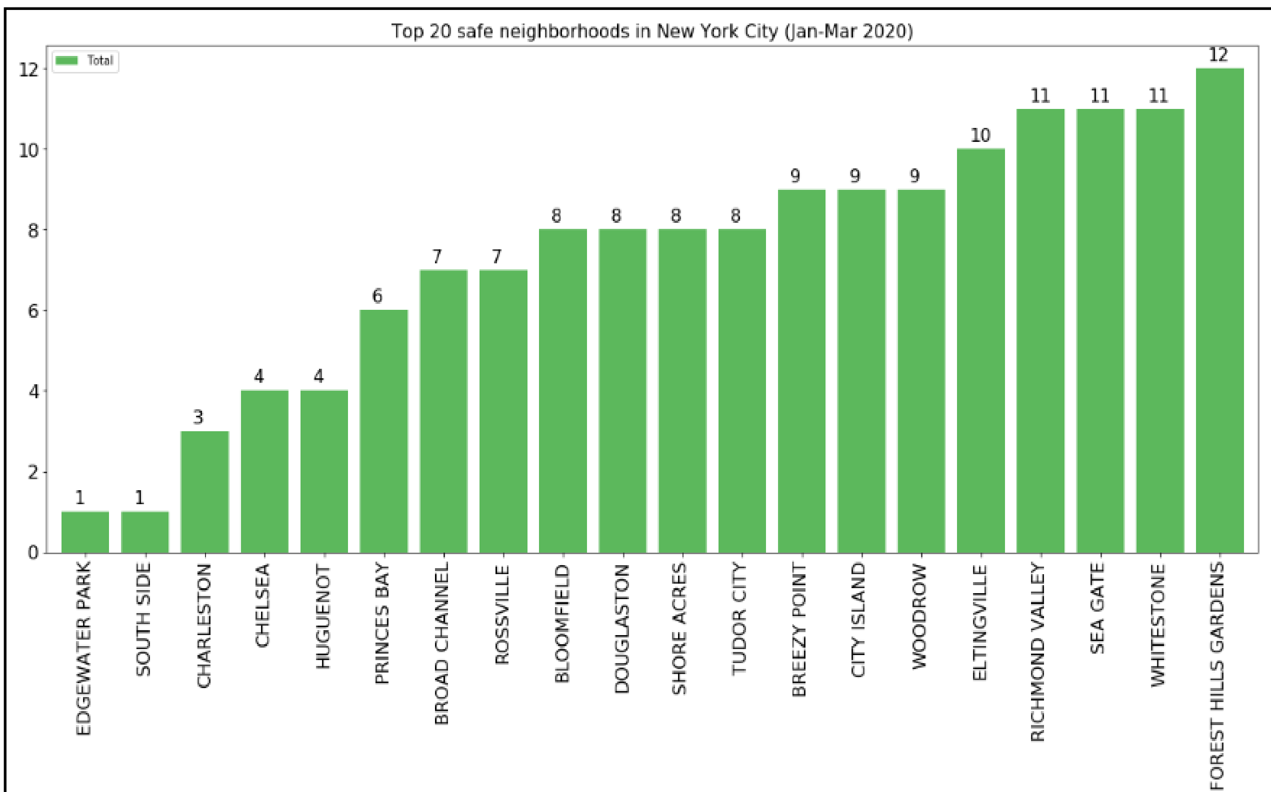
For the analysis 10 neighborhoods with highest crime rate in New York City are taken as per the first quarter of 2020. Following are the bar-chart and table to provide more information in detail.



	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude
0	EAST HARLEM	MANHATTAN	901	1855	557	3313	40.792249	-73.944182
1	MIDTOWN	MANHATTAN	641	951	188	1780	40.754691	-73.981669
2	BEDFORD STUYVESANT	BROOKLYN	308	589	212	1109	40.687232	-73.941785
3	LOWER EAST SIDE	MANHATTAN	303	558	178	1039	40.717807	-73.980890
4	CROWN HEIGHTS	BROOKLYN	324	479	178	981	40.670829	-73.943291
5	FLATBUSH	BROOKLYN	332	453	156	941	40.636326	-73.958401
6	EAST NEW YORK	BROOKLYN	409	354	161	924	40.669926	-73.880699
7	MELROSE	BRONX	234	449	155	838	40.819754	-73.909422
8	MOTT HAVEN	BRONX	185	462	125	772	40.806239	-73.916100
9	EAST VILLAGE	MANHATTAN	201	432	113	746	40.727847	-73.982226

3.1.3 Neighborhoods with the lowest crime rates

For the analysis 20 neighborhoods with lowest crime rate in New York City are taken as per the first quarter of 2020. Following are the bar-chart and table to provide more information in detail.



	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude
0	EDGEWATER PARK	BRONX	0	1	0	1	40.821986	-73.813885
1	SOUTH SIDE	BROOKLYN	0	1	0	1	40.710861	-73.958001
2	CHARLESTON	STATEN ISLAND	2	1	0	3	40.530531	-74.232158
3	CHELSEA	STATEN ISLAND	0	1	3	4	40.594726	-74.189560
4	HUGUENOT	STATEN ISLAND	0	3	1	4	40.531912	-74.191741
5	PRINCES BAY	STATEN ISLAND	2	3	1	6	40.526264	-74.201526
6	BROAD CHANNEL	QUEENS	4	3	0	7	40.603027	-73.820055
7	ROSSVILLE	STATEN ISLAND	1	3	3	7	40.549404	-74.215729
8	BLOOMFIELD	STATEN ISLAND	0	4	4	8	40.605779	-74.187256
9	DOUGLASTON	QUEENS	3	4	1	8	40.766846	-73.742498
10	SHORE ACRES	STATEN ISLAND	0	5	3	8	40.609719	-74.066678
11	TUDOR CITY	MANHATTAN	3	5	0	8	40.746917	-73.971219
12	BREEZY POINT	QUEENS	1	6	2	9	40.557401	-73.925512
13	CITY ISLAND	BRONX	2	2	5	9	40.847247	-73.786488
14	WOODROW	STATEN ISLAND	4	3	2	9	40.541968	-74.205246
15	ELTINGVILLE	STATEN ISLAND	0	4	6	10	40.542231	-74.164331
16	RICHMOND VALLEY	STATEN ISLAND	1	7	3	11	40.519541	-74.229571
17	SEA GATE	BROOKLYN	4	5	2	11	40.576375	-74.007873
18	WHITESTONE	QUEENS	3	8	0	11	40.781291	-73.814202
19	FOREST HILLS GARDENS	QUEENS	7	5	0	12	40.714611	-73.841022

3.2 Modeling

After finding the 20 safest neighborhoods with low crime rates, a separate data frame with only those 20 neighborhoods was created for further analysis. Once that was done, the goal was find all the venues nearby each neighborhood using the Four-Square API this was done using the longitude and latitude of each neighborhood in the above data frame. The data frame would look like the table mentioned below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	EDGEWATER PARK	40.821986	-73.813885	Muscle Maker Grill	40.819391	-73.817298	American Restaurant
1	EDGEWATER PARK	40.821986	-73.813885	Tommy's Pizzeria	40.819573	-73.817482	Pizza Place
2	EDGEWATER PARK	40.821986	-73.813885	The Miles Coffee Bar	40.819462	-73.817352	Coffee Shop
3	EDGEWATER PARK	40.821986	-73.813885	Tosca Café	40.819204	-73.817467	Bar
4	EDGEWATER PARK	40.821986	-73.813885	The Wicked Wolf	40.819688	-73.817359	Pub

As you can see that the above data frame is in length format. Hence we need to do hot-encoding based on the venue category provided above for all the 20 Neighborhoods. Based on the analysis of the above data frame we found out that there are 144 unique venue categories in the data frame. After the data frame is prepared it looks somewhat like this.

	Neighborhood	American Restaurant	Antique Shop	Arepa Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Beach	Beer Garden	Big Box Store	Bike Shop	Boat or Ferry	Bookstore	Boxing Gym	Breakfast Spot
0	EDGEWATER PARK	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	EDGEWATER PARK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	EDGEWATER PARK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	EDGEWATER PARK	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	EDGEWATER PARK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

After this step we create a new dataframe from the above to figure out 10 common venues across all the 20 neighborhoods. This is done by grouping the rows by neighborhood and taking mean of the frequency of each venue category.

	Neighborhood	American Restaurant	Antique Shop	Arepa Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Beach	Beer Garden	Big Box Store	Bike Shop	Boat or Ferry	Bookst
0	BLOOMFIELD	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.00	0.000000	C
1	BREEZY POINT	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.5	0.00	0.000000	0.00	0.000000	C
2	BROAD CHANNEL	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.00	0.000000	C
3	CHARLESTON	0.035714	0.00	0.00	0.000000	0.035714	0.000000	0.00	0.000000	0.035714	0.000000	0.000000	0.000000	0.0	0.00	0.071429	0.00	0.000000	C
4	CHELSEA	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.000000	0.00	0.000000	C

Now we will be creating a dataframe with 10 most common venues for each neighborhood. After preparing the dataframe it looks like as show 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
0	BLOOMFIELD	Theme Park	Recreation Center	Burger Joint	Bus Stop	Yoga Studio	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Event Space
1	BREEZY POINT	Beach	Trail	Monument / Landmark	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food
2	BROAD CHANNEL	Deli / Bodega	Pizza Place	Sporting Goods Shop	Dive Bar	Other Nightlife	Bus Station	Event Space	Food & Drink Shop	Food	Fast Food Restaurant
3	CHARLESTON	Big Box Store	Coffee Shop	Cosmetics Shop	Japanese Restaurant	Furniture / Home Store	Pet Store	Music Venue	Kids Store	Hardware Store	Gym / Fitness Center
4	CHELSEA	Steakhouse	Spanish Restaurant	Park	Sandwich Place	Dry Cleaner	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Event Space

Then after this we merge the above dataset with the dateframe that contains all the information about the 20 safest neighborhoods.

In order to find similarities in the neighborhoods k-means clustering algorithm will be used. In this analysis the neighborhoods will be divided into 3 clusters, this will be done on the basis of type of venues/amenities around the neighborhood.

4) Results

After running k-means clustering algorithm, the neighborhoods are divided into 3 clusters based on the venues/amenities around them.

Cluster 1

The first cluster is smallest cluster with only 1 neighborhood and has venues like Theme Park, Recreation center, Fast food places and Yoga Studio. This cluster is missing out on basic amenities like supermarket or grocery/departmental stores.

	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude	Cluster Labels	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
8	BLOOMFIELD	STATEN ISLAND	0	4	4	8	40.605779	-74.187256	0	Theme Park	Recreation Center	Burger Joint	Bus Stop	Yoga Studio	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Event Space

Cluster 2

The second cluster is biggest cluster with only 17 out of 20 neighborhoods it has venues like Restaurants, Bars/Pubs, Grocery/Departmental stores and other essential amenities.

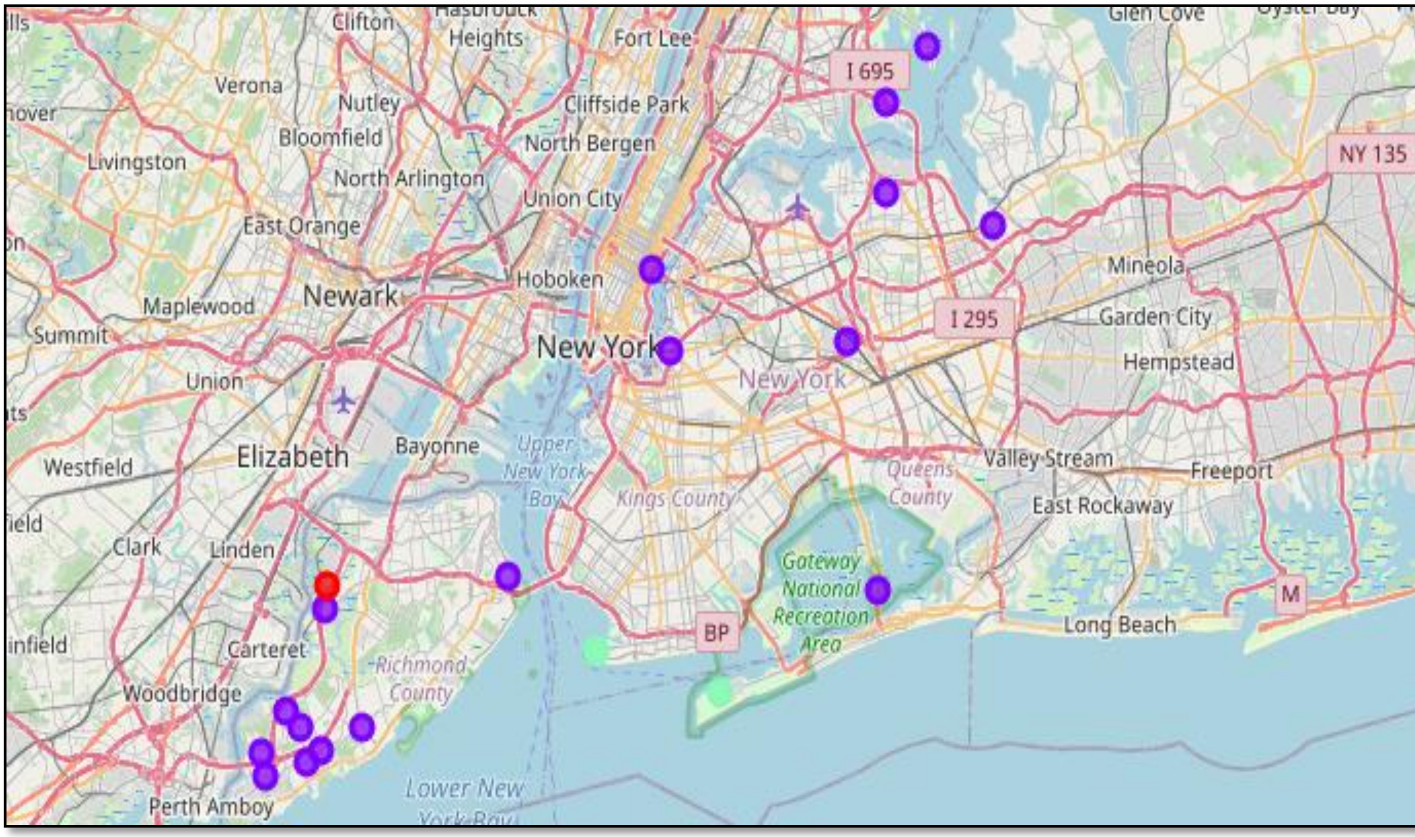
	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	EDGEWATER PARK	BRONX	0	1	0	1	40.821986	-73.813885	1	Italian Restaurant	Pizza Place	Donut Shop	American Restaurant	Liquor Store	Park	Pub	Coffee Shop	Chinese Restaurant	Sports Bar
1	SOUTH SIDE	BROOKLYN	0	1	0	1	40.710861	-73.958001	1	Bar	Coffee Shop	American Restaurant	Pizza Place	Wine Bar	Breakfast Spot	Yoga Studio	Japanese Restaurant	Sushi Restaurant	Chinese Restaurant
2	CHARLESTON	STATEN ISLAND	2	1	0	3	40.530531	-74.232158	1	Big Box Store	Coffee Shop	Cosmetics Shop	Japanese Restaurant	Furniture / Home Store	Pet Store	Music Venue	Kids Store	Hardware Store	Gym / Fitness Center
3	CHELSEA	STATEN ISLAND	0	1	3	4	40.594726	-74.189560	1	Steakhouse	Spanish Restaurant	Park	Sandwich Place	Dry Cleaner	Food Truck	Food & Drink Shop	Food	Fast Food Restaurant	Event Space
4	HUGUENOT	STATEN ISLAND	0	3	1	4	40.531912	-74.191741	1	Italian Restaurant	Asian Restaurant	Donut Shop	Sandwich Place	Moving Target	Bank	Deli / Bodega	Train Station	Ice Cream Shop	Heliport
5	PRINCES BAY	STATEN ISLAND	2	3	1	6	40.526264	-74.201526	1	Pizza Place	Bank	Pharmacy	Pet Store	Construction & Landscaping	Bagel Shop	Sushi Restaurant	Chinese Restaurant	Dry Cleaner	Food & Drink Shop
6	BROAD CHANNEL	QUEENS	4	3	0	7	40.603027	-73.820055	1	Deli / Bodega	Pizza Place	Sporting Goods Shop	Dive Bar	Other Nightlife	Bus Station	Event Space	Food & Drink Shop	Food	Fast Food Restaurant
7	ROSSVILLE	STATEN ISLAND	1	3	3	7	40.549404	-74.215729	1	Pizza Place	Bagel Shop	American Restaurant	Chinese Restaurant	Liquor Store	Grocery Store	Moving Target	Dry Cleaner	Donut Shop	Deli / Bodega
9	DOUGLASTON	QUEENS	3	4	1	8	40.766846	-73.742498	1	Deli / Bodega	Chinese Restaurant	Bank	Ice Cream Shop	Lounge	Diner	Convenience Store	Donut Shop	Fast Food Restaurant	Shanghai Restaurant
10	SHORE ACRES	STATEN ISLAND	0	5	3	8	40.609719	-74.066678	1	Italian Restaurant	Bus Stop	Intersection	Deli / Bodega	Bar	Supermarket	Gastropub	Furniture / Home Store	Music Store	Food
11	TUDOR CITY	MANHATTAN	3	5	0	8	40.746917	-73.971219	1	Park	Café	Mexican Restaurant	Deli / Bodega	Garden	Seafood Restaurant	Wine Shop	Gym	Greek Restaurant	Dog Run
13	CITY ISLAND	BRONX	2	2	5	9	40.847247	-73.786488	1	Thrift / Vintage Store	Seafood Restaurant	Grocery Store	Boat or Ferry	Diner	Smoke Shop	Pizza Place	Pharmacy	Park	Deli / Bodega
14	WOODROW	STATEN ISLAND	4	3	2	9	40.541968	-74.205246	1	Pharmacy	Grocery Store	Donut Shop	Cosmetics Shop	Coffee Shop	Chinese Restaurant	Miscellaneous Shop	Mexican Restaurant	Diner	Martial Arts Dojo

Cluster 3

The third cluster contains 2 neighborhoods it has venues Beach, Restaurants, Departmental stores, Yoga studios, Spa and Sports club and other essential amenities.

	Neighborhood	Borough	FELONY	MISDEMEANOR	VIOLATION	Total	Latitude	Longitude	Cluster Labels	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
12	BREEZY POINT	QUEENS	1	6	2	9	40.557401	-73.925512	2	Beach	Trail	Monument / Landmark	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Food Truck	Food & Drink Shop	Food
17	SEA GATE	BROOKLYN	4	5	2	11	40.576375	-74.007873	2	American Restaurant	Beach	Bus Station	Spa	Sports Club	History Museum	French Restaurant	Department Store	Dessert Shop	Diner

Visualizing neighborhoods divided in 3 clusters. Clusters are color coded like Cluster 1 is Red, Cluster 2 is Purple and Cluster 3 is Sea Green.



5) Discussion

The idea behind the project was to help people to find the safest neighborhood to live with basic amenities which can make people's life easier. For example after the above analysis we can say that Neighborhoods that are present in Cluster 2 and 3 will be ideal for stay as per the various requirements. Like if a person wants to have basic amenities in his/her vicinity then they can easily choose neighborhood from Cluster 2 as per their requirement and if person prefers to have a beach nearby his/her house along with amenities then they can choose from Cluster 3. Ultimately it depends on people's choices and requirements which neighborhood to choose.

6) Conclusion

Choosing a place to live can be hectic task that takes up significant amount of time, money and effort. But going through technical or data oriented approach can save people significant amount of their resources and effort. This model has turned out to be very helpful shortlisting places and then further narrowing it down based on the requirement. Such models aren't helpful at the individual level but also at a large scale for example real estate agents or companies can find ideal places for their clients using this approach.

7) Future directions

Since in this problem, we are trying to find a suitable neighborhood for a client. We can also take another aspect of a person's requirement i.e. Budget whether a person buying a property or looking for a rented one. We should be able to accommodate that requirement into this model as well. This is something on which one can work on improving the model further.