The Battle of Neighborhoods

==

Crime and Venues of Brooklyn

Vojtech Hlava

2021

This paper was made as part of the Applied Data Science Capstone Course
(Part of Coursera IBM Data Science Professional Certificate specialization)

Introduction

In this version of the battle of neighborhoods I will focus on New York. But to keep it differently the analysis will be directed on Brooklyn. It is not the only difference. In the week 2 lab session of the IBM Data Science Capstone Project course our task was to cluster Manhattan neighborhoods based on Foursquare data. Now is time to do it a bit differently. Let's find out how would the result look like if we use also NYPD data?

The story behind this idea is as follows. When someone is looking for a new appartement and doesn't have much information about a new neighborhood you may be interested in many criteria. You can ask yourself questions like "How do people spend their free time in the neighborhood?", "Is this neighborhood save for my family?" or "Are there quality schools for my children nearby?" and so on. For the first two mentioned questions the following report is going to give an answer to such person.

Lots of people are searching for a new apartment because they want to find cheaper rent or perhaps, they want to live closer to their workplace. Nevertheless, they may not be familiar with neighborhoods where they are looking for an apartment. So, it would be helpful for them if they have a tool that tells them this neighborhood is like that one you already know.

This project will focus on comparing neighborhoods based on venues good for spending free time and their safeness.

Data acquisition and cleaning

For this project I will be using and combining multiple data sources.

Data sources

First data source that is necessary for a neighborhood comparison is a geographical dataset. We use neighborhood dataset from week 2 in the same way as it was used in the lab session. Dataset contains centroids of all neighborhoods in New York. But for the purpose of this analysis I filter values for items where name of the borough is equal to Brooklyn. This step filters data by borough and it results in the data frame containing 70 neighborhoods.

	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

Figure 1 First 5 items of the neighborhood dataset

Main part of this paper is built upon analysis of NYPD data. There are a lot of datasets on a website https://opendata.cityofnewyork.us/ and all of them are free to download and analyze. Even though they can be downloaded in various formats and preprocessed manually. I use API link provided by the webservice itself. In such a case I will work with json file. I use two datasets that should provide me a quality data and result of this analysis should by comparable with other forms of neighborhood comparisons. I selected a NYPD Arrest dataset and NYPD Complaint dataset. Combination of these two datasets might be sufficient for covering the whole spectrum of criminal activity. Up to date version of the datasets should provide a good comparison to Foursquare data I use for a control analysis.

NYPD Arrest Data (Year To Date) contains a breakdown of every arrest effected in NYC by the NYPD during the current year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning. Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. In addition, information related to suspect demographics is also included. This data can be used by the public to explore the nature of police enforcement activity.

Column Name	**Column Description**
ARREST_KEY	Randomly generated persistent ID for each arrest
ARREST_DATE	Exact date of arrest for the reported event
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)
KY_CD	Three digit internal classification code (more general category than PD code)
OFNS_DESC	Description of internal classification corresponding with KY code (more general category than PD description)
LAW_CODE	Law code charges corresponding to the NYS Penal Law, VTL and other various local laws
LAW_CAT_CD	Level of offense: felony, misdemeanor, violation
ARREST_BORO	Borough of arrest. B(Bronx), S(Staten Island), K(Brooklyn), M(Manhattan), Q(Queens)
ARREST_PRECINCT	Precinct where the arrest occurred
JURISDICTION_CODE	Jurisdiction responsible for arrest. Jurisdiction codes 0(Patrol), 1(Transit) and 2(Housing) represent NYPD whilst codes 3 and more represent non NYPD jurisdictions
AGE_GROUP	Perpetrator’s age within a category
PERP_SEX	Perpetrator’s sex description
PERP_RACE	Perpetrator’s race description
X_COORD_CD	Midblock X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Y_COORD_CD	Midblock Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Latitude	Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

Table 1 NYPD arrest dataset parameters and their explanation

NYPD Complaint Data Current (Year To Date) includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) for all complete quarters so far this year (2019).

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)

Table 2 NYPD complain dataset parameters and their explanation

The last data source is Foursquare API. I download json file containing venue information through API. My setting of an API call returns 100 venues in the radius of 500 meters for each neighborhood.


```
{'meta': {'code': 200, 'requestId': '602032c27b777166bed7a05d'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'},
    {'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Bay Ridge',
  'headerFullLocation': 'Bay Ridge, Brooklyn',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 81,
  'suggestedBounds': {'ne': {'lat': 40.63030106951066,
    'lng': -74.02470273356597},
   'sw': {'lat': 40.62130106051065, 'lng': -74.03653865351028}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4b895827f964a5206c2d32e3',
       'name': 'Pilo Arts Day Spa and Salon',
       'location': {'address': '8412 3rd Ave',
        'lat': 40.62474788273414,
        'lng': -74.03059056940135,
        'labeledLatLngs': [{'label': 'display',
          'lat': 40.62474788273414,
          'lng': -74.03059056940135},
         {'label': 'entrance', 'lat': 40.624726, 'lng': -74.030697}],
        'distance': 117,
        'postalCode': '11209',
        'cc': 'US',
        'city': 'Brooklyn',
        'state': 'NY',
        'country': 'United States',
        'formattedAddress': ['8412 3rd Ave',
         'Brooklyn, NY 11209',
         'United States']},
       'categories': [{'id': '4bf58dd8d48988d1ed941735',
         'name': 'Spa',
         'pluralName': 'Spas',
         'shortName': 'Spa',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shop
          'suffix': '.png'},
         'primary': True}],
       'photos': {'count': 0, 'groups': []}},
      'referralId': 'e-0-4b895827f964a5206c2d32e3-0'},
     {'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4ad09cf7f964a520bed820e3',
       'name': 'Bagel Boy',
       'location': {'address': '8002 3rd Ave',
        'crossStreet': '80th St',
```


Figure 2 information contained in Foursquare JSON file

All three datasets contain GPS location and category attribute for each item. It means I can easily assign item from datasets to neighborhoods.

Data cleaning

Both NYPD datasets contain attributes where is information about category of the criminal offence and GPS coordinates of the event. Because both datasets are formatted differently, I made two separate data frames and merge them afterwards.

From the NYPD Arrest dataset, I select OFNS_DESC, Latitude and Longitude attributes and create pandas data frame in the way that attribute OFNS_DESC is named "Offense" instead.

	Offense	Latitude	Longitude
0	FELONY ASSAULT	40.808798	-73.916184
1	ROBBERY	40.845956	-73.937813
2	FELONY ASSAULT	40.682398	-73.840079
3	FELONY ASSAULT	40.689336	-73.800409
4	FORGERY	40.634708	-74.124241

Figure 3 Arrest data frame

From the NYPD Complain dataset, I select OFNS_DESC, Latitude and Longitude attributes and create pandas data frame in the way that all columns has the same name as in the first data frame.

	Offense	Latitude	Longitude
0	MURDER & NON-NEGL. MANSLAUGHTER	40.625769	-73.991417
1	MURDER & NON-NEGL. MANSLAUGHTER	40.674583	-73.930222
2	RAPE	40.823101	-73.869690
3	MURDER & NON-NEGL. MANSLAUGHTER	40.887451	-73.847608
4	MURDER & NON-NEGL. MANSLAUGHTER	40.800222	-73.930848

Figure 4 Complaint data frame

The trickiest part of data cleaning is to assign each offense to its proper neighborhood. I choose to assign each item to its neighborhood purely based on the value direct distance between offense coordinates and the coordinates of a neighborhood. The closest neighborhood is assigned to each offense.

	Offense	Latitude	Longitude	Neighborhood
0	FELONY ASSAULT	40.808798	-73.916184	Greenpoint
1	ROBBERY	40.845956	-73.937813	Greenpoint
2	FELONY ASSAULT	40.682398	-73.840079	City Line
3	FELONY ASSAULT	40.689336	-73.800409	City Line
4	FORGERY	40.634708	-74.124241	Bay Ridge

Figure 5 Arrest data frame with neighborhood column

	Offense	Latitude	Longitude	Neighborhood
0	MURDER & NON-NEGL. MANSLAUGHTER	40.625769	-73.991417	Borough Park
1	MURDER & NON-NEGL. MANSLAUGHTER	40.674583	-73.930222	Weeksville
2	RAPE	40.823101	-73.869690	Greenpoint
3	MURDER & NON-NEGL. MANSLAUGHTER	40.887451	-73.847608	Greenpoint
4	MURDER & NON-NEGL. MANSLAUGHTER	40.800222	-73.930848	Greenpoint

Figure 6 Complain data frame with neighborhood column

Now is time to merge both data frames and count offense occurrences and grouped them by neighborhoods. This process ends with following data frame where counted values are normalized.

	Neighborhood	ADMINISTRATIVE CODE	AGRICULTURE & MRKTS LAW-UNCLASSIFIED	ANTICIPATORY OFFENSES	ARSON	ASSAULT 3 & RELATED OFFENSES	BURGLAR'S TOOLS	BURGLARY	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL TRESPASS	...	PETIT LARCENY
0	Bath Beach	0.0	0.0	0.0	0.0	0.428571	0.0	0.000000	0.142857	0.000000	...	0.000000
1	Bay Ridge	0.0	0.0	0.0	0.0	0.166667	0.0	0.000000	0.166667	0.027778	...	0.194444
2	Bedford Stuyvesant	0.0	0.0	0.0	0.0	0.068966	0.0	0.034483	0.172414	0.000000	...	0.068966
3	Bensonhurst	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.125000	0.000000	...	0.125000
4	Bergen Beach	0.0	0.0	0.0	0.0	0.363636	0.0	0.000000	0.090909	0.000000	...	0.000000

Figure 7 Offense data frame containing normalized counts of each offense category

Because I want to compare clusters made with help of NYPD data with Foursquare data clusters. I have to prepare them to the same format of the data frame. Whole process is a bit easier with Foursquare data due to the fact Foursquare API returns list of venues closest to each neighborhood. Thanks to this I can easily grouped the list of venues by neighborhood the same way as before. This process results in the following data frame.

	Neighborhood	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	...	Video Game Store	Video Store	Vietnamese Restaurant
0	Bath Beach	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.020000	0.02	0.000000
1	Bay Ridge	0.0	0.0	0.036585	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.012195	0.00	0.012195
2	Bedford Stuyvesant	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
3	Bensonhurst	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
4	Bergen Beach	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000

Figure 8 Foursquare data frame containing normalized counts of each venue category

Data analysis

To find out in what neighborhood does an incident occurs we assign an incident to the closest neighborhood center based on coordinates of an incident and coordinates of a neighborhood.

Because we pick attributes to our data frame that are the same in both NYPD datasets we can easily merge both data frame into one. We are interested in number of occurrences of each incident category grouped by assigned neighborhood. It results into a data frame that contains the 5 most common incidents in each neighborhood by NYPD.

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
0	Bath Beach	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	MISCELLANEOUS PENAL LAW	CRIMINAL MISCHIEF & RELATED OF	FELONY ASSAULT
1	Bay Ridge	PETIT LARCENY	CRIMINAL MISCHIEF & RELATED OF	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	MISCELLANEOUS PENAL LAW
2	Bedford Stuyvesant	CRIMINAL MISCHIEF & RELATED OF	MISCELLANEOUS PENAL LAW	ROBBERY	FELONY ASSAULT	GRAND LARCENY
3	Bensonhurst	FELONY ASSAULT	HARRASSMENT 2	OFF. AGNST PUB ORD SENSBLTY &	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY
4	Bergen Beach	ASSAULT 3 & RELATED OFFENSES	GRAND LARCENY	SEX CRIMES	MURDER & NON-NEGL. MANSLAUGHTE	MURDER & NON-NEGL. MANSLAUGHTER

Figure 9 Head of crime data frame with 5 most common offenses

We create a data frame that contains 5 most common venue categories for each neighborhood in below table.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bath Beach	Chinese Restaurant	Pharmacy	Donut Shop	Bubble Tea Shop	Cantonese Restaurant
1	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Greek Restaurant	Chinese Restaurant
2	Bedford Stuyvesant	Coffee Shop	Café	Pizza Place	Deli / Bodega	Bar
3	Bensonhurst	Chinese Restaurant	Ice Cream Shop	Sushi Restaurant	Donut Shop	Italian Restaurant
4	Bergen Beach	Harbor / Marina	Park	Athletics & Sports	Baseball Field	Playground

Figure 10 Head of venue data frame with 5 most common venues

Clustering

As an input for following clustering we use the data frame that contains normalized incidence of arrests and complains recorded by NYPD in Brooklyn neighborhoods.

	Neighborhood	ADMINISTRATIVE CODE	AGRICULTURE & MRKTS LAW-UNCLASSIFIED	ANTICIPATORY OFFENSES	ARSON	ASSAULT 3 & RELATED OFFENSES	BURGLAR'S TOOLS	BURGLARY	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL TRESPASS	...	PETIT LARCENY	POSSESSIO OF STOLE PROPERT
0	Bath Beach	0.0	0.0	0.0	0.0	0.428571	0.0	0.000000	0.142857	0.000000	...	0.000000	0
1	Bay Ridge	0.0	0.0	0.0	0.0	0.166667	0.0	0.000000	0.166667	0.027778	...	0.194444	0
2	Bedford Stuyvesant	0.0	0.0	0.0	0.0	0.068966	0.0	0.034483	0.172414	0.000000	...	0.068966	0
3	Bensonhurst	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.125000	0.000000	...	0.125000	0
4	Bergen Beach	0.0	0.0	0.0	0.0	0.363636	0.0	0.000000	0.090909	0.000000	...	0.000000	0
...	...	...	...	...	...	...	...	...	...	...	...	...	
65	Vinegar Hill	0.0	0.0	0.0	0.0	0.133333	0.0	0.000000	0.133333	0.000000	...	0.333333	0
66	Weeksville	0.0	0.0	0.0	0.0	0.111111	0.0	0.000000	0.111111	0.000000	...	0.000000	0
67	Williamsburg	0.0	0.0	0.0	0.0	0.000000	0.0	0.200000	0.000000	0.000000	...	0.200000	0
68	Windsor Terrace	0.0	0.0	0.0	0.0	0.166667	0.0	0.166667	0.166667	0.000000	...	0.000000	0
69	Wingate	0.0	0.0	0.0	0.0	0.187500	0.0	0.062500	0.250000	0.000000	...	0.125000	0

70 rows × 45 columns

Figure 11 Crime data frame ready for clustering

Similar data frame was acquired from Foursquare data. As shown below there is a data frame that contains incidences of venues in all Brooklyn neighborhoods.

	Neighborhood	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	...	Video Game Store	Video Store	Vietnamese Restaurant
0	Bath Beach	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.020000	0.02	0.000000
1	Bay Ridge	0.0	0.0	0.036585	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.012195	0.00	0.012195
2	Bedford Stuyvesant	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
3	Bensonhurst	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
4	Bergen Beach	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
65	Vinegar Hill	0.0	0.0	0.033333	0.000000	0.0	0.0	0.066667	0.000000	0.0	...	0.000000	0.00	0.000000
66	Weeksville	0.0	0.0	0.071429	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000
67	Williamsburg	0.0	0.0	0.000000	0.000000	0.0	0.0	0.029412	0.000000	0.0	...	0.000000	0.00	0.000000
68	Windsor Terrace	0.0	0.0	0.034483	0.034483	0.0	0.0	0.000000	0.034483	0.0	...	0.000000	0.00	0.000000
69	Wingate	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.000000	0.00	0.000000

70 rows × 291 columns

Figure 12 Venue data frame ready for clustering

To determine the best number of clusters I use an elbow method. The results are show on the following pictures.

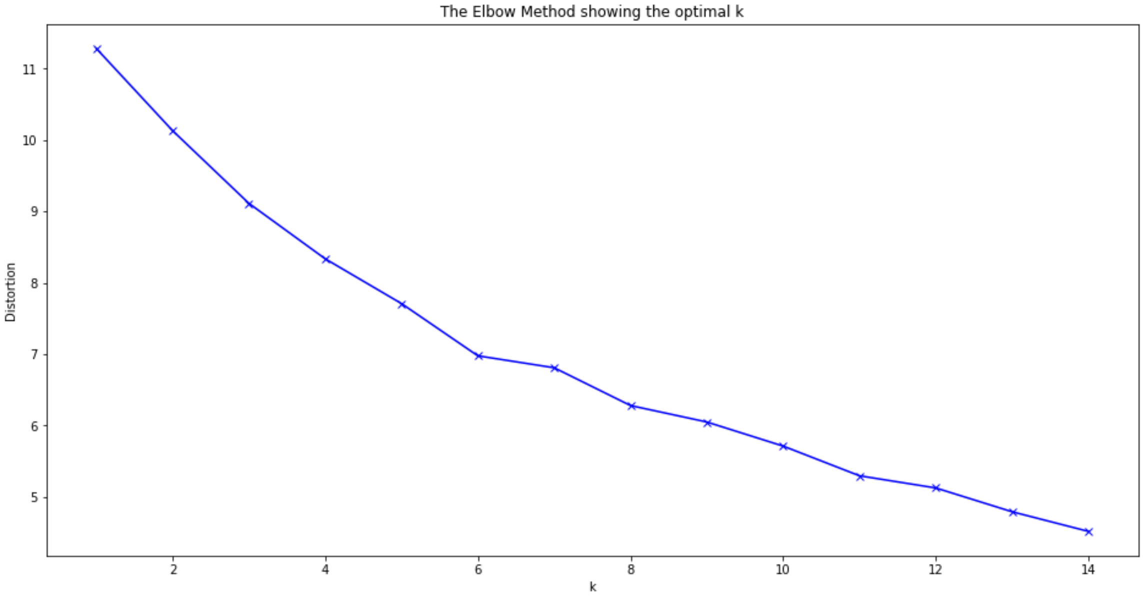

Figure 13 Result of an elbow method for crime dataset

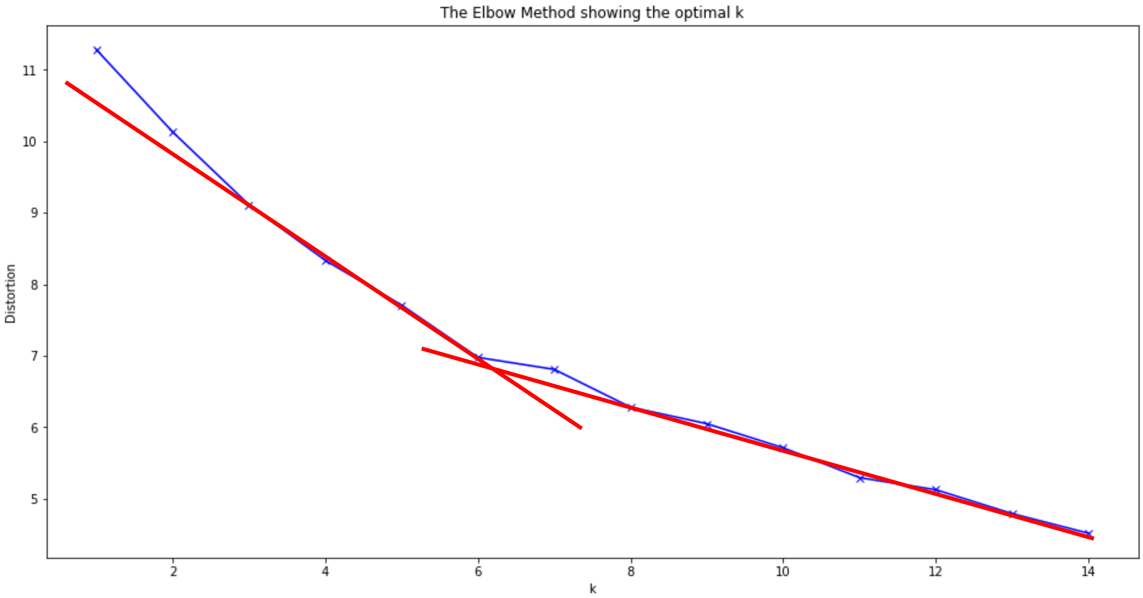

Figure 14 Result of an elbow method for crime dataset (optimal number of clusters)

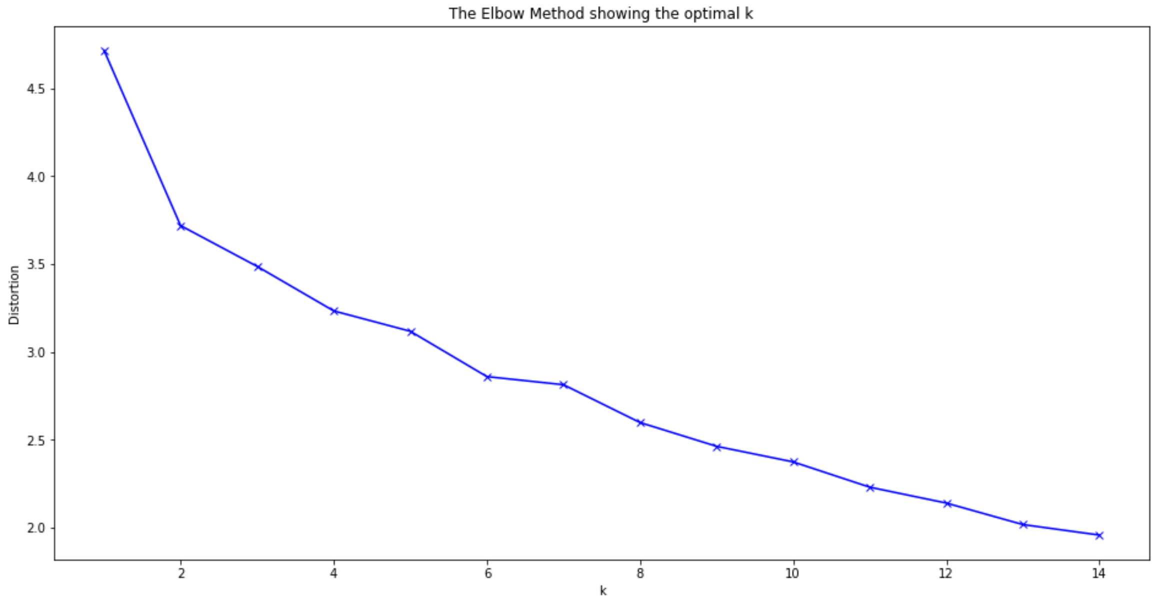

Figure 15 Result of an elbow method for venue dataset

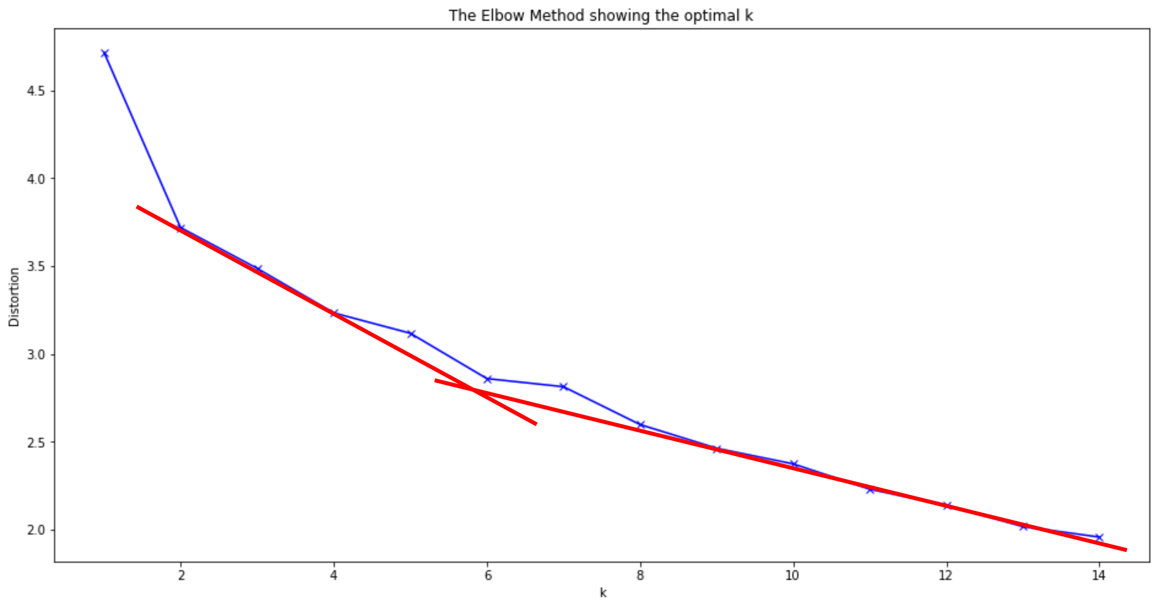

Figure 16 Result of an elbow method for venue dataset t(optimal number of clusters)

Both results are quite inconclusive but I determine that for the crime data frame I continue with 6 clusters and with 6 clusters for venue the data frame as well.

Result

The work isn't ending by adding a cluster number to the dataframe. It is allways a good practise to present results in some graphical way.

Because dataframe contains goegraphical coordiantes it is convenient to show the result on a map.

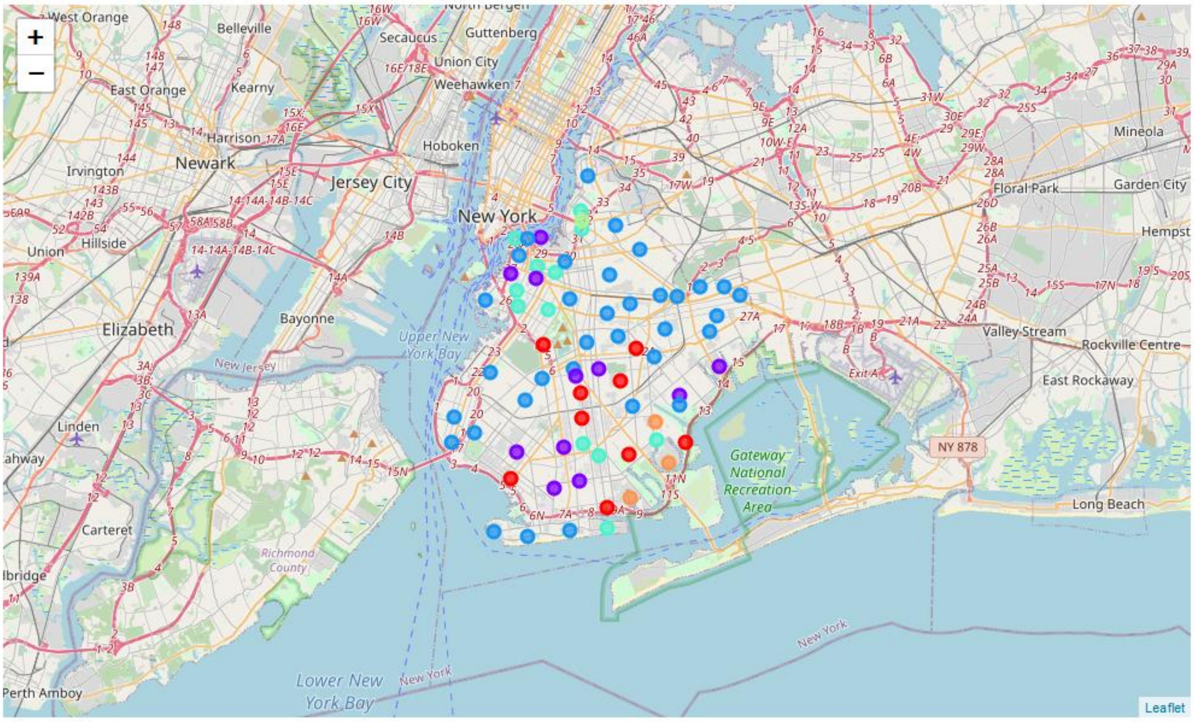

Figure 17 Result of crime data clustering (colors represent different clusters)

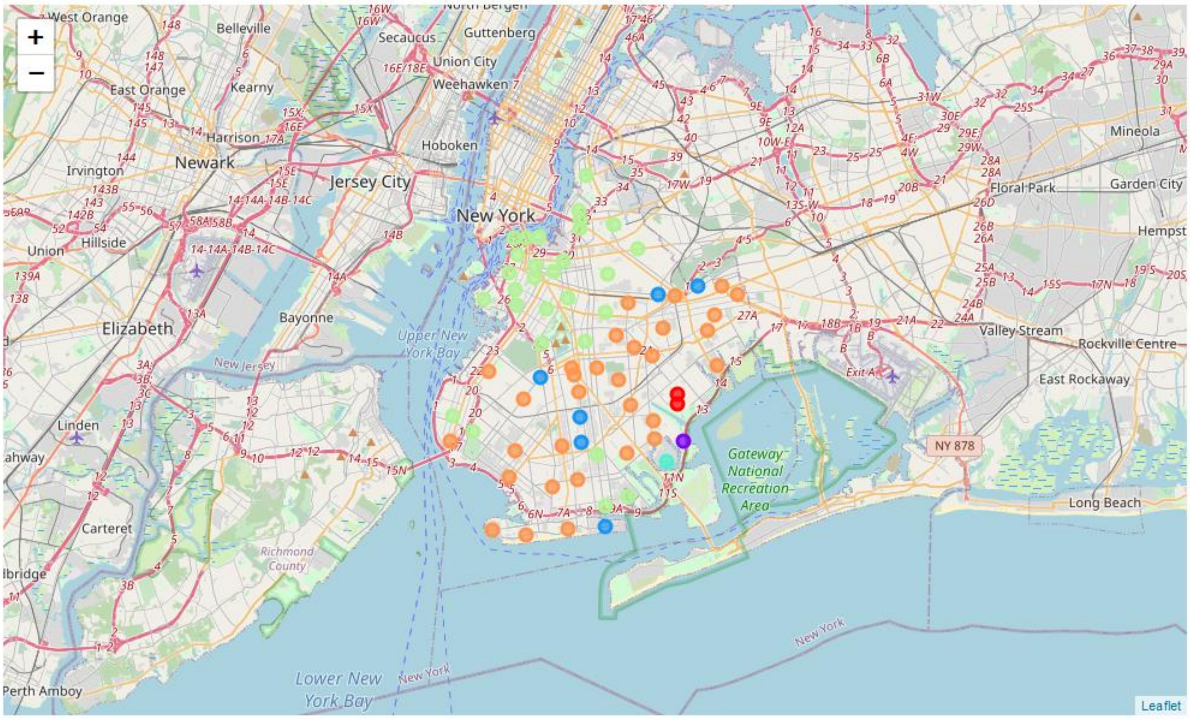

Figure 18 Result of venue data clustering (colors represent different clusters)

In the following tables there are lists of neighborhoods with five most common offense category and venue category, respectively.

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
6	Sheepshead Bay	HARRASSMENT 2	ASSAULT 3 & RELATED OFFENSES	BURGLARY	VEHICLE AND TRAFFIC LAWS	FELONY ASSAULT
8	Flatbush	ASSAULT 3 & RELATED OFFENSES	GRAND LARCENY	BURGLARY	CRIMINAL TRESPASS	VEHICLE AND TRAFFIC LAWS
10	East Flatbush	HARRASSMENT 2	GRAND LARCENY OF MOTOR VEHICLE	BURGLARY	ASSAULT 3 & RELATED OFFENSES	UNAUTHORIZED USE OF A VEHICLE
12	Windsor Terrace	DANGEROUS WEAPONS	HARRASSMENT 2	ASSAULT 3 & RELATED OFFENSES	BURGLARY	CRIMINAL MISCHIEF & RELATED OF
33	Bath Beach	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	MISCELLANEOUS PENAL LAW	CRIMINAL MISCHIEF & RELATED OF	FELONY ASSAULT
37	Marine Park	ASSAULT 3 & RELATED OFFENSES	OFF. AGNST PUB ORD SENSBLTY &	OTHER TRAFFIC INFRACTION	VEHICLE AND TRAFFIC LAWS	FELONY ASSAULT
45	Bergen Beach	ASSAULT 3 & RELATED OFFENSES	GRAND LARCENY	SEX CRIMES	MURDER & NON-NEGL. MANSLAUGHTE	MURDER & NON-NEGL. MANSLAUGHTER
46	Midwood	ASSAULT 3 & RELATED OFFENSES	BURGLARY	VEHICLE AND TRAFFIC LAWS	FELONY ASSAULT	HOMICIDE-NEGLIGENT,UNCLASSIFIE
56	Rugby	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	FELONY ASSAULT	ROBBERY	MURDER & NON-NEGL. MANSLAUGHTER

Figure 19 1st cluster of crime data frame

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
1	Bensonhurst	FELONY ASSAULT	HARRASSMENT 2	OFF. AGNST PUB ORD SENSBLTY &	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY
4	Gravesend	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY	FELONY ASSAULT	VEHICLE AND TRAFFIC LAWS	DANGEROUS WEAPONS
19	Cobble Hill	FELONY ASSAULT	MISCELLANEOUS PENAL LAW	ARSON	PETIT LARCENY	VEHICLE AND TRAFFIC LAWS
27	Starrett City	PETIT LARCENY	MURDER & NON-NEGL. MANSLAUGHTER	ASSAULT 3 & RELATED OFFENSES	FELONY ASSAULT	HARRASSMENT 2
28	Canarsie	HARRASSMENT 2	MURDER & NON-NEGL. MANSLAUGHTER	PETIT LARCENY	FELONY ASSAULT	VEHICLE AND TRAFFIC LAWS
41	Boerum Hill	HARRASSMENT 2	PETIT LARCENY	MISCELLANEOUS PENAL LAW	MURDER & NON-NEGL. MANSLAUGHTER	DANGEROUS WEAPONS
52	Ocean Parkway	PETIT LARCENY	HARRASSMENT 2	DANGEROUS WEAPONS	ROBBERY	OFF. AGNST PUB ORD SENSBLTY &
54	Ditmas Park	PETIT LARCENY	DANGEROUS WEAPONS	HARRASSMENT 2	GRAND LARCENY OF MOTOR VEHICLE	ASSAULT 3 & RELATED OFFENSES
62	Vinegar Hill	PETIT LARCENY	ASSAULT 3 & RELATED OFFENSES	MURDER & NON-NEGL. MANSLAUGHTER	CRIMINAL MISCHIEF & RELATED OF	HARRASSMENT 2
66	Homecrest	PETIT LARCENY	DANGEROUS DRUGS	FELONY ASSAULT	DANGEROUS WEAPONS	HARRASSMENT 2
69	Erasmus	FELONY ASSAULT	MISCELLANEOUS PENAL LAW	ASSAULT 3 & RELATED OFFENSES	OFF. AGNST PUB ORD SENSBLTY &	PETIT LARCENY

Figure 20 2nd cluster of crime data frame

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
0	Bay Ridge	PETIT LARCENY	CRIMINAL MISCHIEF & RELATED OF	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	MISCELLANEOUS PENAL LAW
2	Sunset Park	PETIT LARCENY	ASSAULT 3 & RELATED OFFENSES	VEHICLE AND TRAFFIC LAWS	OFF. AGNST PUB ORD SENSBLTY &	CRIMINAL MISCHIEF & RELATED OF
3	Greenpoint	ASSAULT 3 & RELATED OFFENSES	PETIT LARCENY	FELONY ASSAULT	CRIMINAL MISCHIEF & RELATED OF	HARRASSMENT 2
5	Brighton Beach	DANGEROUS DRUGS	FELONY ASSAULT	GRAND LARCENY OF MOTOR VEHICLE	ROBBERY	ASSAULT 3 & RELATED OFFENSES
9	Crown Heights	ASSAULT 3 & RELATED OFFENSES	BURGLARY	SEX CRIMES	CRIMINAL MISCHIEF & RELATED OF	DANGEROUS WEAPONS
11	Kensington	ROBBERY	FELONY ASSAULT	ASSAULT 3 & RELATED OFFENSES	GRAND LARCENY	PETIT LARCENY
13	Prospect Heights	HARRASSMENT 2	ROBBERY	PETIT LARCENY	CRIMINAL MISCHIEF & RELATED OF	VEHICLE AND TRAFFIC LAWS
14	Brownsville	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	OFF. AGNST PUB ORD SENSBLTY &	MISCELLANEOUS PENAL LAW	MURDER & NON-NEGL. MANSLAUGHTER
16	Bushwick	CRIMINAL MISCHIEF & RELATED OF	FELONY ASSAULT	ROBBERY	PETIT LARCENY	ASSAULT 3 & RELATED OFFENSES
17	Bedford Stuyvesant	CRIMINAL MISCHIEF & RELATED OF	MISCELLANEOUS PENAL LAW	ROBBERY	FELONY ASSAULT	GRAND LARCENY
18	Brooklyn Heights	FELONY ASSAULT	GRAND LARCENY	CRIMINAL MISCHIEF & RELATED OF	CRIMINAL TRESPASS	FOR OTHER AUTHORITIES
21	Red Hook	GRAND LARCENY OF MOTOR VEHICLE	ROBBERY	MURDER & NON-NEGL. MANSLAUGHTER	VEHICLE AND TRAFFIC LAWS	DANGEROUS WEAPONS
25	Cypress Hills	PETIT LARCENY	ASSAULT 3 & RELATED OFFENSES	GRAND LARCENY	FELONY ASSAULT	HARRASSMENT 2
26	East New York	ROBBERY	FELONY ASSAULT	GRAND LARCENY	ASSAULT 3 & RELATED OFFENSES	FORGERY
29	Flatlands	PETIT LARCENY	VEHICLE AND TRAFFIC LAWS	GRAND LARCENY	MURDER & NON-NEGL. MANSLAUGHTE	ASSAULT 3 & RELATED OFFENSES
32	Coney Island	DANGEROUS WEAPONS	PETIT LARCENY	OFF. AGNST PUB ORD SENSBLTY &	ROBBERY	BURGLARY
34	Borough Park	SEX CRIMES	OFF. AGNST PUB ORD SENSBLTY &	INTOXICATED & IMPAIRED DRIVING	HARRASSMENT 2	GRAND LARCENY
35	Dyker Heights	DANGEROUS DRUGS	VEHICLE AND TRAFFIC LAWS	MISCELLANEOUS PENAL LAW	ROBBERY	MURDER & NON-NEGL. MANSLAUGHTER
38	Clinton Hill	MISCELLANEOUS PENAL LAW	OFFENSES AGAINST THE PERSON	GRAND LARCENY	ASSAULT 3 & RELATED OFFENSES	MURDER & NON-NEGL. MANSLAUGHTER
39	Sea Gate	ASSAULT 3 & RELATED OFFENSES	PETIT LARCENY	HARRASSMENT 2	GRAND LARCENY	MURDER & NON-NEGL. MANSLAUGHTER
42	Prospect Lefferts Gardens	OTHER OFFENSES RELATED TO THEF	BURGLARY	MISCELLANEOUS PENAL LAW	MURDER & NON-NEGL. MANSLAUGHTER	OFF. AGNST PUB ORD SENSBLTY &
43	Ocean Hill	ASSAULT 3 & RELATED OFFENSES	HARRASSMENT 2	ROBBERY	GRAND LARCENY	CRIMINAL MISCHIEF & RELATED OF
44	City Line	ASSAULT 3 & RELATED OFFENSES	PETIT LARCENY	FELONY ASSAULT	CRIMINAL MISCHIEF & RELATED OF	HARRASSMENT 2
47	Prospect Park South	VEHICLE AND TRAFFIC LAWS	GRAND LARCENY	ROBBERY	ASSAULT 3 & RELATED OFFENSES	OFFENSES AGAINST PUBLIC ADMINI
49	East Williamsburg	HARRASSMENT 2	GRAND LARCENY	MISCELLANEOUS PENAL LAW	DANGEROUS WEAPONS	ASSAULT 3 & RELATED OFFENSES
53	Fort Hamilton	VEHICLE AND TRAFFIC LAWS	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY	HARRASSMENT 2	MISCELLANEOUS PENAL LAW
55	Wingate	CRIMINAL MISCHIEF & RELATED OF	ASSAULT 3 & RELATED OFFENSES	FELONY ASSAULT	PETIT LARCENY	THEFT-FRAUD
57	Remsen Village	MISCELLANEOUS PENAL LAW	ASSAULT 3 & RELATED OFFENSES	VEHICLE AND TRAFFIC LAWS	MURDER & NON-NEGL. MANSLAUGHTER	PETIT LARCENY
58	New Lots	UNAUTHORIZED USE OF A VEHICLE	GRAND LARCENY OF MOTOR VEHICLE	DANGEROUS WEAPONS	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY
59	Paerdegat Basin	UNAUTHORIZED USE OF A VEHICLE	FELONY ASSAULT	MISCELLANEOUS PENAL LAW	CRIMINAL MISCHIEF & RELATED OF	OFFENSES AGAINST THE PERSON
63	Weeksville	MURDER & NON-NEGL. MANSLAUGHTER	MISCELLANEOUS PENAL LAW	ASSAULT 3 & RELATED OFFENSES	CRIMINAL MISCHIEF & RELATED OF	HARRASSMENT 2
64	Broadway Junction	DANGEROUS DRUGS	FELONY ASSAULT	ASSAULT 3 & RELATED OFFENSES	CRIMINAL MISCHIEF & RELATED OF	PETIT LARCENY
65	Dumbo	GRAND LARCENY OF MOTOR VEHICLE	ASSAULT 3 & RELATED OFFENSES	OFF. AGNST PUB ORD SENSBLTY &	DANGEROUS DRUGS	BURGLARY
67	Highland Park	HARRASSMENT 2	OFF. AGNST PUB ORD SENSBLTY &	SEX CRIMES	ROBBERY	RAPE

Figure 21 3rd cluster of crime data frame

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
7	Manhattan Terrace	PETIT LARCENY	GRAND LARCENY	VEHICLE AND TRAFFIC LAWS	DANGEROUS WEAPONS	HARRASSMENT 2
15	Williamsburg	HARRASSMENT 2	MURDER & NON-NEGL. MANSLAUGHTER	BURGLARY	PETIT LARCENY	DANGEROUS WEAPONS
20	Carroll Gardens	GRAND LARCENY	PETIT LARCENY	MURDER & NON-NEGL. MANSLAUGHTER	BURGLARY	VEHICLE AND TRAFFIC LAWS
22	Gowanus	DANGEROUS DRUGS	GRAND LARCENY	BURGLARY	PETIT LARCENY	DANGEROUS WEAPONS
23	Fort Greene	MISCELLANEOUS PENAL LAW	MURDER & NON-NEGL. MANSLAUGHTER	GRAND LARCENY	DANGEROUS DRUGS	VEHICLE AND TRAFFIC LAWS
24	Park Slope	MISCELLANEOUS PENAL LAW	GRAND LARCENY	MURDER & NON-NEGL. MANSLAUGHTER	PETIT LARCENY	VEHICLE AND TRAFFIC LAWS
31	Manhattan Beach	GRAND LARCENY	OFFENSES AGAINST THE PERSON	VEHICLE AND TRAFFIC LAWS	INTOXICATED & IMPAIRED DRIVING	HARRASSMENT 2
40	Downtown	PETIT LARCENY	GRAND LARCENY	BURGLARY	CRIMINAL MISCHIEF & RELATED OF	MISCELLANEOUS PENAL LAW
50	North Side	BURGLARY	HARRASSMENT 2	PETIT LARCENY	GRAND LARCENY	FELONY ASSAULT
60	Mill Basin	HARRASSMENT 2	GRAND LARCENY	PETIT LARCENY	INTOXICATED & IMPAIRED DRIVING	VEHICLE AND TRAFFIC LAWS
61	Fulton Ferry	GRAND LARCENY	BURGLARY	PETIT LARCENY	ASSAULT 3 & RELATED OFFENSES	FOR OTHER AUTHORITIES
68	Madison	BURGLARY	PETIT LARCENY	VEHICLE AND TRAFFIC LAWS	DANGEROUS WEAPONS	HARRASSMENT 2

Figure 22 4th cluster of crime data frame

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
51	South Side	HARRASSMENT 2	VEHICLE AND TRAFFIC LAWS	UNAUTHORIZED USE OF A VEHICLE	HOMICIDE-NEGLIGENT,UNCLASSIFIE	GRAND LARCENY OF MOTOR VEHICLE

Figure 23 5th cluster of crime data frame

	Neighborhood	1st Most Common Offense	2nd Most Common Offense	3rd Most Common Offense	4th Most Common Offense	5th Most Common Offense
30	Mill Island	FELONY ASSAULT	VEHICLE AND TRAFFIC LAWS	UNAUTHORIZED USE OF A VEHICLE	HOMICIDE-NEGLIGENT,UNCLASSIFIE	HARRASSMENT 2
36	Gerritsen Beach	VEHICLE AND TRAFFIC LAWS	FELONY ASSAULT	ASSAULT 3 & RELATED OFFENSES	HOMICIDE-NEGLIGENT,UNCLASSIFIE	HARRASSMENT 2
48	Georgetown	FELONY ASSAULT	ASSAULT 3 & RELATED OFFENSES	VEHICLE AND TRAFFIC LAWS	HOMICIDE-NEGLIGENT,UNCLASSIFIE	HARRASSMENT 2

Figure 24 6th cluster of crime data frame

In the following tables there are lists of neighborhoods with five most common venues category and venue category, respectively.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
28	Canarsie	Asian Restaurant	Food	Grocery Store	Caribbean Restaurant	Home Service
59	Paerdegat Basin	Child Care Service	Food	Home Service	Asian Restaurant	Gym

Figure 25 1st cluster of venue data frame

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
45	Bergen Beach	Harbor / Marina	Park	Athletics & Sports	Baseball Field	Playground

Figure 26 2nd cluster of venue data frame

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Manhattan Terrace	Pizza Place	Ice Cream Shop	Cosmetics Shop	Donut Shop	Liquor Store
11	Kensington	Thai Restaurant	Grocery Store	Pizza Place	Ice Cream Shop	Park
31	Manhattan Beach	Bus Stop	Ice Cream Shop	Playground	Sandwich Place	Harbor / Marina
43	Ocean Hill	Deli / Bodega	Grocery Store	Southern / Soul Food Restaurant	Supermarket	Fried Chicken Joint
46	Midwood	Pizza Place	Convenience Store	Ice Cream Shop	Candy Store	Bakery
67	Highland Park	Liquor Store	Spanish Restaurant	Pizza Place	Grocery Store	Latin American Restaurant

Figure 27 3rd cluster of venue data frame

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
30	Mill Island	Pool	Yoga Studio	Fish Market	Farm	Farmers Market

Figure 28 4th cluster of venue data frame

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Greek Restaurant	Chinese Restaurant
3	Greenpoint	Pizza Place	Coffee Shop	Bar	Cocktail Bar	Grocery Store
6	Sheepshead Bay	Dessert Shop	Turkish Restaurant	Yoga Studio	Karaoke Bar	Hotel
9	Crown Heights	Pizza Place	Café	Museum	Bagel Shop	Burger Joint
12	Windsor Terrace	Deli / Bodega	Plaza	Grocery Store	Park	Diner
13	Prospect Heights	Bar	Mexican Restaurant	Cocktail Bar	Wine Shop	Coffee Shop
15	Williamsburg	Pizza Place	Coffee Shop	Bagel Shop	Yoga Studio	Event Space
16	Bushwick	Bar	Coffee Shop	Mexican Restaurant	Deli / Bodega	Discount Store
17	Bedford Stuyvesant	Coffee Shop	Café	Pizza Place	Deli / Bodega	Bar
18	Brooklyn Heights	Deli / Bodega	Yoga Studio	Park	Coffee Shop	Gym
19	Cobble Hill	Playground	Pizza Place	Cocktail Bar	Coffee Shop	Deli / Bodega
20	Carroll Gardens	Italian Restaurant	Coffee Shop	Pizza Place	Bakery	Cocktail Bar
21	Red Hook	Seafood Restaurant	Art Gallery	Bar	Park	Café
22	Gowanus	Italian Restaurant	Bar	Furniture / Home Store	Gym / Fitness Center	Wine Shop
23	Fort Greene	Flower Shop	Wine Shop	Italian Restaurant	Pizza Place	Cocktail Bar
24	Park Slope	Coffee Shop	Bagel Shop	Burger Joint	Pizza Place	Mexican Restaurant
35	Dyker Heights	Burger Joint	Bagel Shop	Park	Golf Course	Grocery Store
36	Gerritsen Beach	Ice Cream Shop	Bar	Harbor / Marina	Park	Restaurant
38	Clinton Hill	Italian Restaurant	Pizza Place	Mexican Restaurant	Wine Shop	Thai Restaurant
40	Downtown	Burger Joint	Coffee Shop	Bar	Sandwich Place	Chinese Restaurant
41	Boerum Hill	Coffee Shop	Dance Studio	Bar	Bakery	French Restaurant
42	Prospect Lefferts Gardens	Café	Pizza Place	Deli / Bodega	Caribbean Restaurant	Bakery
49	East Williamsburg	Deli / Bodega	Bar	Bakery	Concert Hall	Cocktail Bar
50	North Side	Coffee Shop	Yoga Studio	Pizza Place	American Restaurant	Bakery
51	South Side	Pizza Place	Bar	Coffee Shop	American Restaurant	Wine Bar
61	Fulton Ferry	Park	Ice Cream Shop	Roof Deck	Scenic Lookout	Coffee Shop
62	Vinegar Hill	Food Truck	Bike Rental / Bike Share	Coffee Shop	Café	Art Gallery
65	Dumbo	Park	Bakery	Scenic Lookout	Dog Run	Pizza Place
68	Madison	Bagel Shop	Deli / Bodega	Dessert Shop	Italian Restaurant	Spa

Figure 29 5th cluster of venue data frame

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Bensonhurst	Chinese Restaurant	Ice Cream Shop	Sushi Restaurant	Donut Shop	Italian Restaurant
2	Sunset Park	Bank	Bakery	Mexican Restaurant	Latin American Restaurant	Pizza Place
4	Gravesend	Pizza Place	Italian Restaurant	Lounge	Chinese Restaurant	Bus Station
5	Brighton Beach	Restaurant	Russian Restaurant	Eastern European Restaurant	Gourmet Shop	Mobile Phone Shop
8	Flatbush	Deli / Bodega	Mexican Restaurant	Coffee Shop	Caribbean Restaurant	Bank
10	East Flatbush	Chinese Restaurant	Print Shop	Supermarket	Caribbean Restaurant	Park
14	Brownsville	Fried Chicken Joint	Moving Target	Restaurant	Plaza	Convenience Store
25	Cypress Hills	Fried Chicken Joint	Spanish Restaurant	Donut Shop	Fast Food Restaurant	Pizza Place
26	East New York	Deli / Bodega	Spanish Restaurant	Fast Food Restaurant	Food Truck	Metro Station
27	Starrett City	Bus Station	Convenience Store	Pharmacy	American Restaurant	Caribbean Restaurant
29	Flatlands	Pharmacy	Fast Food Restaurant	Caribbean Restaurant	Fried Chicken Joint	Nightclub
32	Coney Island	Baseball Stadium	Theme Park Ride / Attraction	Beach	Farmers Market	Caribbean Restaurant
33	Bath Beach	Chinese Restaurant	Pharmacy	Donut Shop	Bubble Tea Shop	Cantonese Restaurant
34	Borough Park	Bank	Pizza Place	Fast Food Restaurant	Deli / Bodega	Pharmacy
37	Marine Park	Chinese Restaurant	Deli / Bodega	Liquor Store	Gym	Pizza Place
39	Sea Gate	Sports Club	Spa	Home Service	Bus Station	Beach
44	City Line	Donut Shop	Pharmacy	Cosmetics Shop	Grocery Store	Shoe Store
47	Prospect Park South	Caribbean Restaurant	Pizza Place	Mobile Phone Shop	Grocery Store	Fast Food Restaurant
48	Georgetown	Bank	Breakfast Spot	Pharmacy	Donut Shop	Pizza Place
52	Ocean Parkway	Liquor Store	Dessert Shop	Steakhouse	Paper / Office Supplies Store	Gym
53	Fort Hamilton	Deli / Bodega	Italian Restaurant	Sandwich Place	Chinese Restaurant	Diner
54	Ditmas Park	Chinese Restaurant	Caribbean Restaurant	Kids Store	Department Store	Pizza Place
55	Wingate	Fried Chicken Joint	Gym / Fitness Center	Flower Shop	Fast Food Restaurant	Liquor Store
56	Rugby	Grocery Store	Caribbean Restaurant	Bank	Salon / Barbershop	Pharmacy
57	Remsen Village	Caribbean Restaurant	Fast Food Restaurant	Sandwich Place	Supermarket	Chinese Restaurant
58	New Lots	Fried Chicken Joint	Pizza Place	Chinese Restaurant	Grocery Store	Bank
60	Mill Basin	Chinese Restaurant	Pizza Place	Japanese Restaurant	Bagel Shop	Bank
63	Weeksville	Deli / Bodega	Juice Bar	Gas Station	Chinese Restaurant	Grocery Store
64	Broadway Junction	Gas Station	Donut Shop	Fried Chicken Joint	Bus Station	Bus Stop
66	Homecrest	Bank	Donut Shop	Chinese Restaurant	Pizza Place	Sushi Restaurant
69	Erasmus	Caribbean Restaurant	Pharmacy	Music Venue	School	Chinese Restaurant

Figure 30 6th cluster of venue data frame

Conclusion

I showed two ways to cluster Brooklyn neighborhoods. But that there are multiple ways to cluster the same neighborhoods into different clusters. Each of the way would give us unique perspective on the Brooklyn borough.

Firstly, I used NYPD data for clustering. There is a lot of variability in data, so it is hard to choose the best number of clusters.

Second option served the purpose of “control group”. The procedure is the same as was shown in the week 2 lab session. Foursquare data shows a lot of homogeneity when not filtered.

The results of those two clustering couldn’t be more different or at least behavior of the clustering method couldn’t be different. This is caused by input data which contains whole range of felony and venues, respectively. Such behavior results in incomparable results between both clustering approaches.