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