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