Moving to London from Saint-Petersburg.

Gennadiy Garshin, May 2020.

Table of Contents

1. Introduction.

1.1. Problem declaration.

The family faced the question about moving to London. Having visited city several times it was decided to choose the southern part of the capital. In this research I will try to find several options for possible neighborhoods to move to. The starting point of a comfortable stay will be the current place of living in Saint-Petersburg. The analysis will be made on the basis of open sources such as foursquare, wikipedia and uk government.

1.2. Problem discussion.

To search for the right neighborhoods, I'll use the clustering segmentation. Attributes for segmentation will be common places around current location. To do this, we need to prepare two datasets:
1) with the current place of living and surrounding common venues - bars, restaurants, fitness clubs, supermarkets, etc.
2) London's main neighborhoods and surrounding public areas are within walking distance.

The list of areas in London is here. Data about nearby venues will be received from Foursquare using API. Clustering will be done by k-Means method.

After that we will look at the obtained neighborhoods on the map and try to assess their prospects by adding information about their distance from the center.

In the end of the analysis section, I will make a few comparative graphs with rental prices for the resulting neighborhoods and their crime situation based on the open data sets from uk.gov site site.

2. Data preparation.

Any analysis is impossible without data. In this reserch, we will use both type of data - historical data showing the dynamics of changes in any factors, and static data, showing the real state of things. Let's prepare the data and import it into Pandas data frames to use in our analysis.

2.1. Dataset 1. Neighborhood data.

The table with London areas is in the Wikipedia pages. The link is here.
Scrap the data using library BeautifulSoup and record it into Pandas DataFrame.
Field list: **Location, London borough, Post town, Postcode district, Dial code, OS grid ref**.

Information we need:
Location - Specific neighborhood name.
London borough - borough of the City of London.
Postcode district - first three letters from postal codes of this neighborhoods.

After transformation we have such dataframe.

	Neighborhood	Borough	Postcode	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	SE2	51.49245	0.12127
11	Anerley	Bromley	SE20	51.41009	-0.05683
21	Balham	Wandsworth	SW12	51.44822	-0.14839
22	Bankside	Southwark	SE1	51.49996	-0.09568
27	Barnes	Richmond upon Thames	SW13	51.47457	-0.24212

2.2. Dataset 2. Nearby Venues.

Using the foursquare service, we can get data about nearby places at the specified coordinates.
Field list: **Neighborhood, Neighborhood Latitude and Longitude, Venue, Venue Latitude and Longitude, Venue Category**.

Information we need:
Neighborhood - Specific neighborhood name.
Neighborhood Latitude and Longitude - geographical coordinates.
Venue Category - category of venue for classification.

After transformation we have such dataframe with neighborhood and their nearby venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbey Wood	51.49245	0.12127	Lesnes Abbey	51.489526	0.125839	Historic Site
1	Abbey Wood	51.49245	0.12127	Lidl	51.496152	0.118417	Supermarket
2	Abbey Wood	51.49245	0.12127	Sainsbury's	51.492826	0.120524	Supermarket
3	Abbey Wood	51.49245	0.12127	Abbey Wood Railway Station (ABW)	51.490825	0.123432	Train Station
4	Abbey Wood	51.49245	0.12127	Co-op Food	51.487650	0.113490	Grocery Store

This dataframe contains information about the neighborhoods, its coordinates, and places within it.

2.3. Dataset 3. London Crime data.

Data about London crimes for last 24 months for each borough. This dataframe is consist of historycal data about London crimes in each month. Using it we can see how the number of crimes has changed over time in each borough. Dataset was copied from uk.gov. site

Field list: **MajorText, MinorText,LookUp_BorougName, 201805,201806…**

Information we need:
MajorText -name of crime type.
LookUp_BorougName - borough name of the City of London.
Columns with specific month 201805..202004 - number of incidents

	MajorText	MinorText	LookUp_BoroughName	201805	201806	201807	201808	201809	201810	201811	201812	201901	201902	201903	201904
0	Arson and Criminal Damage	Arson	Barking and Dagenham	4	12	6	5	3	8	5	1	5	2	5	5
1	Arson and Criminal Damage	Criminal Damage	Barking and Dagenham	126	123	127	101	107	132	105	88	97	127	138	130
2	Burglary	Burglary - Business and Community	Barking and Dagenham	24	33	30	18	33	32	39	33	45	24	29	27
3	Burglary	Burglary - Residential	Barking and Dagenham	93	77	94	84	99	94	106	164	114	107	99	96
4	Drug Offences	Drug Trafficking	Barking and Dagenham	8	6	8	7	10	7	7	4	5	2	6	5

2.4. Dataset 4. London average rents by borough.

Data about London neighborhood rent prices.
This Dataarame includes data about the types of apartments and their median prices in the context of dates. Dataset was copied from uk.gov. site.

Field list: **Year, Quarter,code,Area,Category,Count of rents, Average, Lower quartile, Median, Upper quartile.**

Information we need:
Year
Quarter
Area - borough name of the City of London.
Category - type of apartments
Median - the median price for category

	Year	Quarter	Code	Area	Category	Count of rents	Average	Lower quartile	Median	Upper quartile
0	2011	Q2	E09000001	City of London	Room	-	-	-	-	-
1	2011	Q2	E09000002	Barking and Dagenham	Room	92	336	282	347	390
2	2011	Q2	E09000003	Barnet	Room	945	450	399	433	500
3	2011	Q2	E09000004	Bexley	Room	119	390	347	390	433
4	2011	Q2	E09000005	Brent	Room	344	469	390	457	550

3. Analysis.

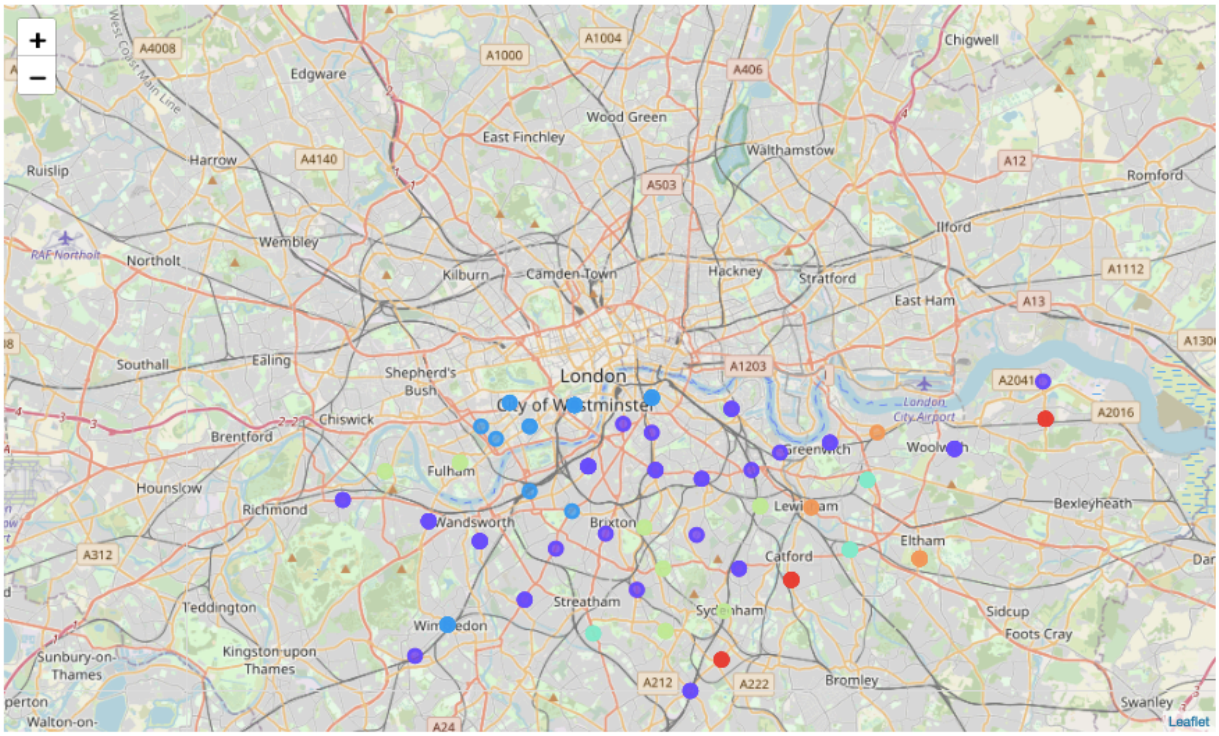
In this section of the analysis, we will do the main work on selecting a suitable place of living. Based on clustering segmentation, we will identify several clusters. To do this, we will use k-means clustering. After that, we will several analysis on the results obtained to narrow the resultset. We will conduct a comparative analysis of the characteristics of the borough - prices for renting apartments and the criminogenic situation. It will help us to choose our final goal in the future.

3.1. Preparation dataset for clustering.

Afer all preparation including one-hot encoding, finding the top-ten venues category we can implement clustering.

3.3. Clustering result.

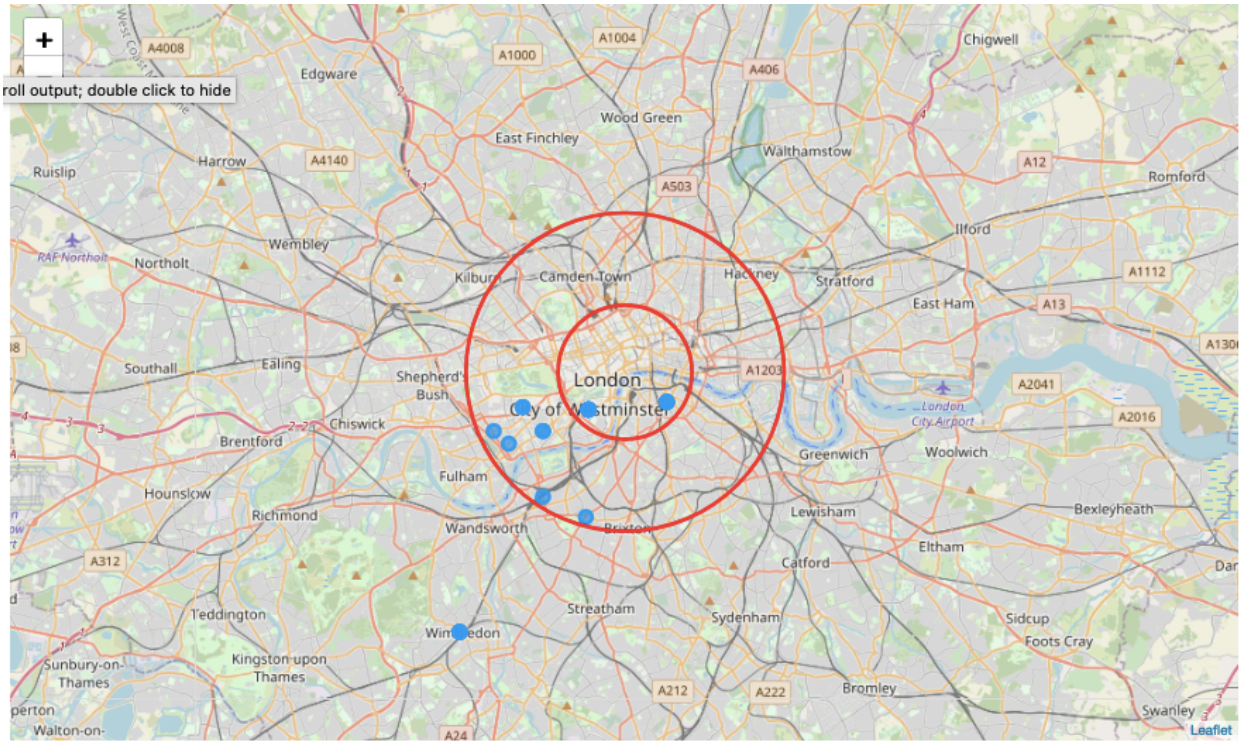
After clustering segmentation we can show it on the map.



3.4. Cluster on the map.

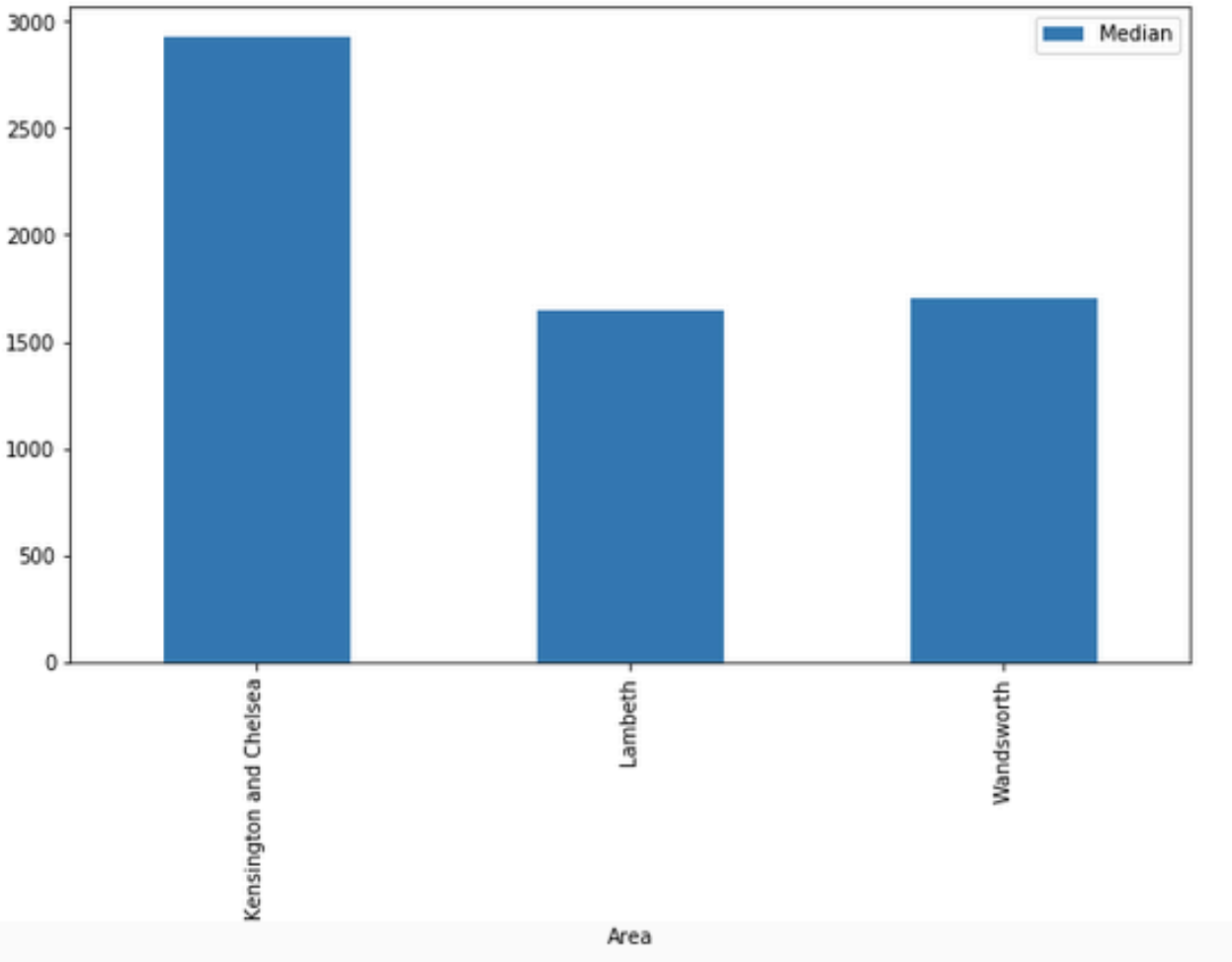
Now we can explore cluster on the map.
Let's add two circles on the map that illustrate the minimum and maximum distance from the city center. These distances were chosen for personal reasons of comfortable distance from the center.

As we can see on the map below the neighborhoods and their boroughs satisfying chosen distances are:
1) Neighborhood: Earls Court. Borough: Kensington and Chelsea.
2) Neighborhood: South Kensington. Borough: Kensington and Chelsea.
3) Neighborhood: Chelsea. Borough: Kensington and Chelsea.
4) Neighborhood: Battersea. Borough: Wandsworth.
5) Neighborhood: Clapham. Borough: Lambeth.

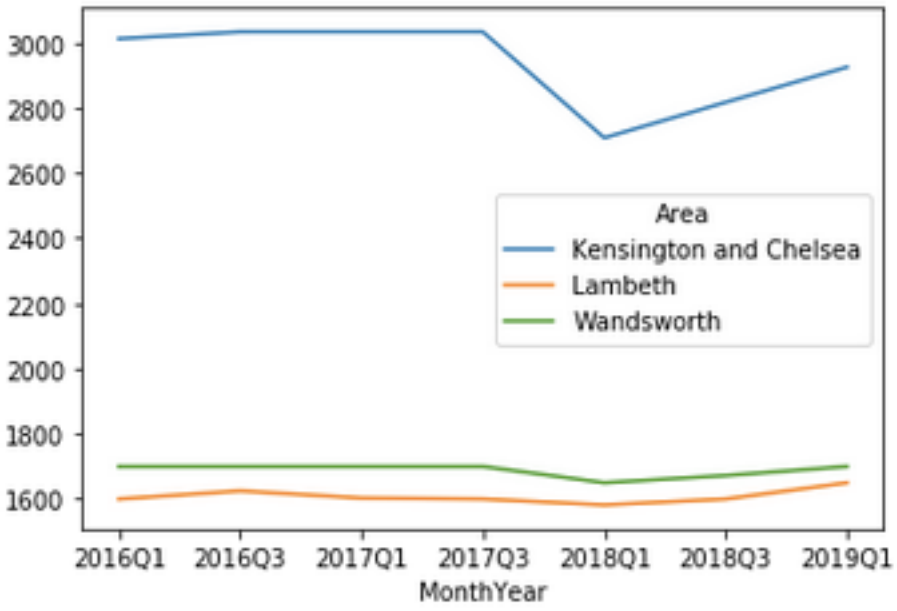


3.5. The apartment rental prices data in the London boroughs.

Now let's evaluate prices for neighborhoods.



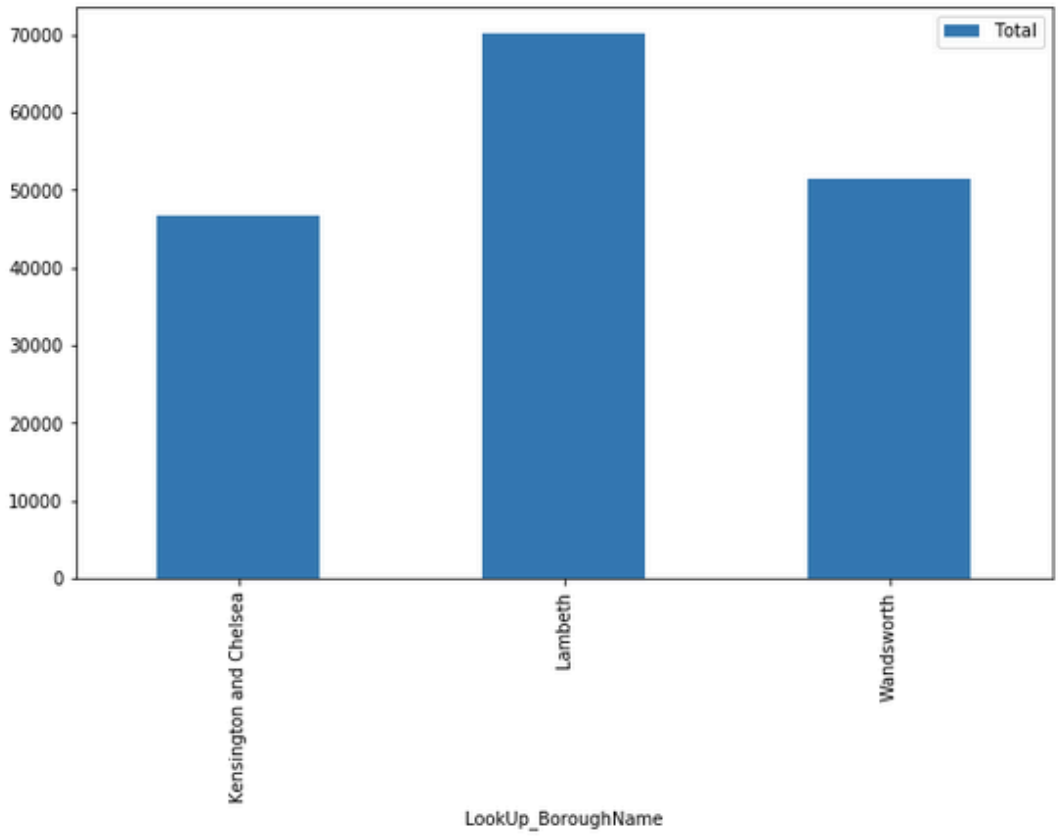
And let's look at the dynamics of price changes over the past few years.



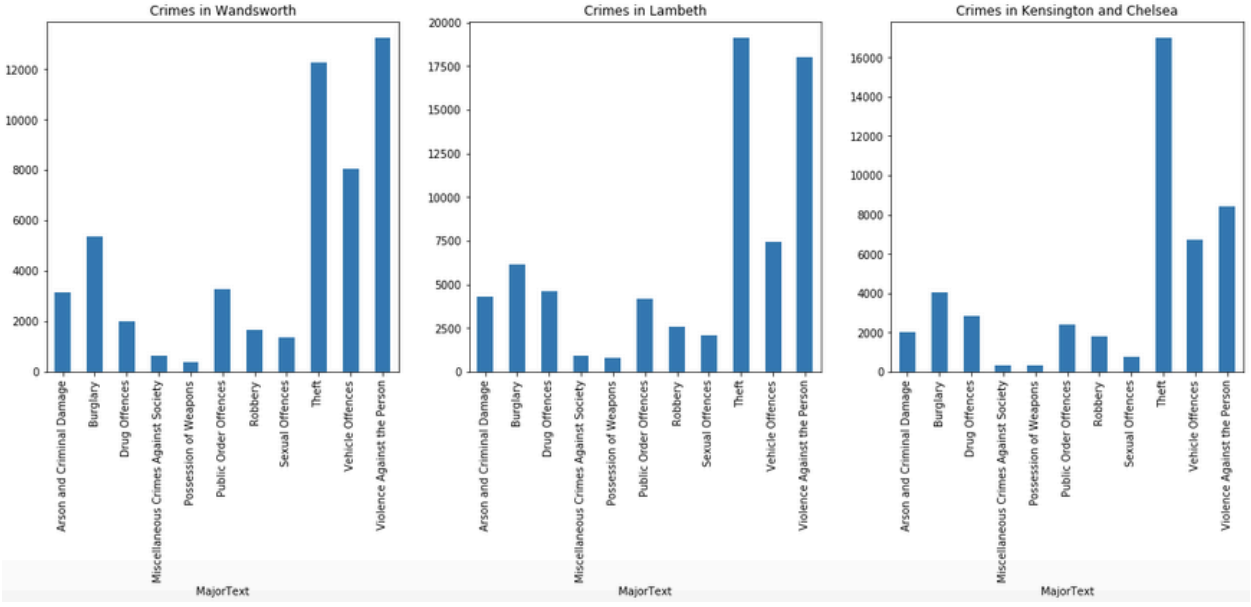
As we can see on charts, Kensington and Chelsea is extremely expensive borough in rent.
Wandsworth and Lambeth are cheaper.
Over the past four years the price dynamics have roughly coincided for all three boroughs.

3.6. Crime data for London boroughs.

Total crimes by chosen boroughs.



Crime types by boroughs.



4. Results.

In this research we tried to determine the most comfortable areas for moving to London from St. Petersburg.
Based on cluster segmentation, we have selected several neighborhoods that satisfy our conditions - the proximity of venues, so that the family does not feel a shortage of anything from the previous area in the new place of living.
Using the minimum and maximum distance from the city center, we have reduced our resultset to three borough - Kensington and Chelsea, Lambeth, Wandsworth.
After conducting a comparative analysis of rental apartment prices, we conclude that the neighborhoods of the borough of Kensington and Chelsea have the highest prices. The dynamics of price changes shows that prices in the medium term are stable and it is a insignificant factor when choosing an apartment.
We conducted a comparative analysis of the criminal situation in the districts. As you can see from the charts - most crimes happened in Lambeth.

Based on the received data, you can build a list of priorities in this way:
1) Wandsworth, since this borough has the least crimes and is not the most expensive.
2) Kensington and Chelsea, despite the fact that this is the most expensive borough, crimes are still committed less often in it.
3) Lambeth, an unfavorable borough, but also the prices for apartments are the lowest.

5. Discussion

In large modern cities it is difficult to find borough that differs greatly from each other by the criterion of the presence or absence of any venue categories. Without having visited the city once, it is impossible to choose a place remotely where you would like to live comfortably. Perhaps, for such complex tasks it would be appropriate to add priority to each category, so as to prioritize not by the fact of the availability of any venues, but by the importance for the family. Moreover, distance to work will be an important criteria. Since this characteristic was not set initially, I added a distance from the city center. In large cities the time to get to work takes up a decent part of the day.
Perhaps, the comparative analyses of boroughs which has been carried out is not enough to fully understand the situation. You can add to the consideration such characteristics as nationality and religion of people in each borough.

6. Conclusion

In this report we made some work to identify similar neighborhoods in two different cities. Cluster segmentation was selected as the definition method. As a result we received a list of boroughs that meet the conditions. Then we analyzed the received boroughs and looked at them from different angles. Based on the data obtained, we have prepared a list of priority boroughs.