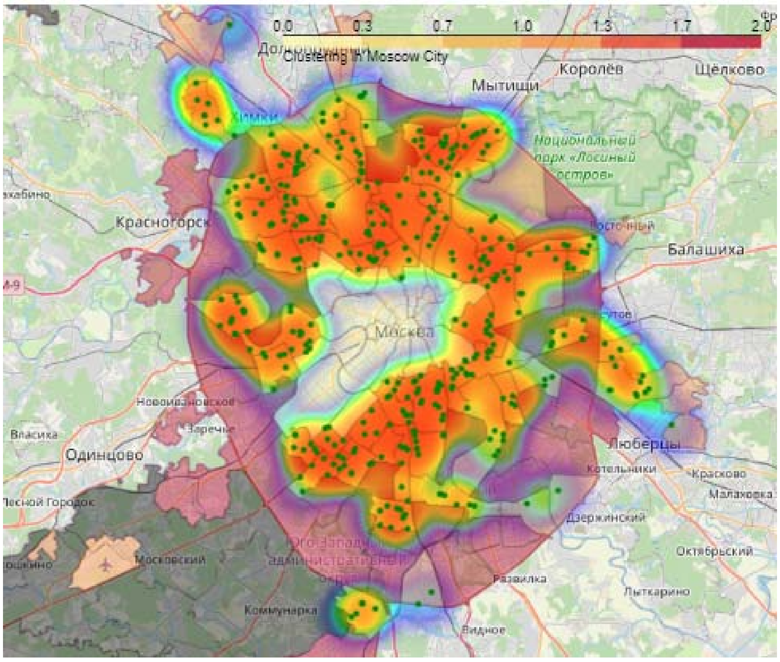
Venues Data Analysis of Moscow City



IBM Professional Certificate in Data Science – Capstone Project (Coursera)

.

By Alex Prkh

Contents

1. Introduction

1.1 Background

Moscow, one of the largest metropolises in the world with a population of more than 15 million people, covers an area around 2561.5 km² with an average density of inheritance of 4924.96 people / km² .

Moscow is divided into 12 districts (125 boroughs, 2 urban boroughs, 19 settlement boroughs).

Moscow has a very uneven population density from 30429 people / km² for one borough, to 560 people / km² for the last borough .

The average cost of real estate varies from 68,768 rubles / m² for the "Кленовское" borough to 438,568 rubles / m² for the "Арбат" borough .

1.2 Problem statement

Owners of social facilities are expected to prefer boroughs with a high population density. Investors will prefer areas with low housing costs and low competitiveness. In this case, we assume that the price for housing corresponds to the rent payment of facilities in one borough. So one can make an assumption about spending to open a new facility.

On the part of residents, the preference is expected for a boroughs with a low cost of housing and good accessibility of social places.

In my research, I will try to determine the optimal places for the location of Auto Workshops in Moscow boroughs, taking into account the number of people, the cost of real estate and the density of other facilities.

The key criteria for selecting suitable locations will be:

- High population of the borough
- Low cost of real estate in the borough
- The absence in the immediate vicinity of other facilities

The main stakeholders of my research will be investors for local Auto Workshops.

2. The Data Set

The data sources I use are publicly available sources but are not easily found online.

Based on the problem and the established selection criteria, to conduct the research, I will need the following information.

Data requirements:

1. main dataset with the list of Moscow Borough, containing the following attributes:
   - name of the each Moscow Borough
   - type of the each Moscow Borough
   - name of the each Moscow District in which Borough is belong to
   - area of the each Moscow Borough in square kilometers
   - the population of the each Moscow Borough
   - housing area of the each Moscow Borough in square meters
   - average housing price of the each Moscow Borough
2. geographical coordinates of the each Moscow Borough
3. shape of the each Moscow Borough in GEOJSON format
4. list of venues placed in the each Moscow Borough with their geographical coordinates and categories

Data for Moscow Boroughs datasets were downloaded from multiple HTTP page combined into one pandas dataframe.

- List of Moscow District and they Boroughs were downloaded from the page Moscow Boroughs
- Information about area of the each Moscow Borough in square kilometers, their population and housing area in square meters were downloaded from the page Moscow Boroughs Population Density
- Information about housing price of the each Moscow Borough were downloaded from the page Moscow Boroughs Housing Price

Geographical coordinates of the each Moscow Borough were queried through Nominatim service. As the Nominatim service are quite unstable it was quite a challenge to request coordinate in several iterations

From the link, I will merge the information into a single file and used this merged file for analysis. The reason for separating the task is because of the long run time for the API data merge to happen.

The result Moscow Boroughs dataset

The prepared and cleared Moscow Boroughs dataset has such view.

The picture below shows a small part of the Moscow Boroughs dataset

Borough_Name	District_Name	Borough_Type	ATO_Borough_Cc	TMO_District_C	Borough_Area	orough_Populatic	Populatic	rough_Housing_Ai	using_Are	Latitude	Longitude	orough_Housing_Pric
Академический	ЮЗАО	Муниципальный округ	45293554	45397000	5.83	109387	18762	2467.00	22.70	55.69	37.58	199999.00
Алексеевский	СВАО	Муниципальный округ	45280552	45349000	5.29	80534	15223	1607.90	20.50	55.81	37.65	199474.00
Алтуфьевский	СВАО	Муниципальный округ	45280554	45350000	3.25	57596	17721	839.30	15.50	55.88	37.58	138021.00
Арбат	ЦАО	Муниципальный округ	45286552	45374000	2.11	36125	17120	731.00	26.00	55.75	37.59	438568.00
Аэропорт	САО	Муниципальный округ	45277553	45333000	4.58	79486	17355	1939.70	25.90	55.80	37.53	234544.00
Бабушкинский	СВАО	Муниципальный округ	45280556	45351000	5.07	88537	17462	1586.30	18.50	55.87	37.66	164324.00
Басманный	ЦАО	Муниципальный округ	45286555	45375000	8.37	110694	13225	1991.80	18.40	55.78	37.69	302021.00
Беговой	САО	Муниципальный округ	45277556	45334000	5.56	42781	7694	791.10	18.80	55.78	37.57	261402.00
Бескудниковский	САО	Муниципальный округ	45277559	45335000	3.30	79603	24122	1391.70	18.40	55.86	37.56	158398.00
Бибирево	СВАО	Муниципальный округ	45280558	45352000	6.45	160163	24831	2521.80	15.80	55.88	37.60	140533.00
Бирюлёво Восточное	ЮАО	Муниципальный округ	45296553	45911000	14.77	155863	10552	2122.20	14.70	55.59	37.66	124645.00
Бирюлёво Западное	ЮАО	Муниципальный округ	45296555	45912000	8.51	88672	10419	1183.20	13.20	55.59	37.64	109421.00
Богородское	ВАО	Муниципальный округ	45263552	45301000	10.24	109324	10676	1744.10	16.90	55.82	37.71	178577.00
Братеево	ЮАО	Муниципальный округ	45296557	45913000	7.63	110021	14419	1585.40	15.50	55.64	37.76	136300.00
Бутырский	СВАО	Муниципальный округ	45280561	45353000	5.04	71458	14178	1236.20	18.30	55.81	37.59	182641.00
Вешняки	ВАО	Муниципальный округ	45263555	45302000	10.72	122285	11407	1976.80	16.20	55.73	37.82	147352.00
Внуково	ЗАО	Муниципальный округ	45268552	45317000	17.42	25471	1462	416.60	17.80	55.61	37.30	113399.00
Войковский	САО	Муниципальный округ	45277565	45336000	6.61	70729	10700	1531.00	23.10	55.82	37.49	207242.00
Восточное Дегунино	САО	Муниципальный округ	45277568	45337000	3.77	98923	26239	1592.50	16.70	55.88	37.56	146300.00

To determine **venues** the service **Foursquare API** was used.

The API of **Foursquare** service have the restriction of 100 **venues**, which it can return in one request.

To obtain list of all **venues** I used the following approach:

- present Moscow area in the form of a regular grid of circles of quite small diameter, no more than 100 **venues** in each circle
- perform exploration using **Foursquare API** with quite bigger radius than circle of a grid to make sure it overlaps/full coverage to don't miss any venues
- cleaning list of venues from duplicates.

This approach and some of the Python code was taken from the work presented here. https://cocl.us/coursera_capstone_notebook

The prepared and cleared Venue dataset has such view.

The picture below shows a small part of it.

```
print('Take a look at the dataframe data types')
print(Moscow_venues_df.dtypes)
```

```
Take a look at the dataframe
                                  Cell_id  Cell_Latitude  Cell_Longitude  \
0   55.49560209571474,37.5786154020309 2      55.495602       37.578615
1    55.50758514958972,37.5417462724848 5      55.507585       37.541746
2    55.50758514958972,37.5417462724848 5      55.507585       37.541746
3  55.50247175433097 6,37.56806302526971 6      55.502472       37.568063
4    55.50076610141684,37.5768334014280 5      55.500766       37.576833

                   Venue_Id              Venue_Name  \
0  4c2325d013c00f47638e88de      Рынок «Удобный»
1  501abe19e4b07bd245dabf68  Пруд "Утиная гавань"
2  58b6a74a109dfe2494c95358          Imperia BMW
3  578e94bc498e584562d31cad        Центр Плова 24
4  5519adb1498e70931fb8eb51               Сушиwok

                              Venue_All_Categories  Venue_Latitude  \
0   [('Hardware Store', '4bf58dd8d48988d112951735')]       55.498413
1             [('Lake', '4bf58dd8d48988d161941735')]       55.509217
2    [('Auto Workshop', '56aa371be4b08b9a8d5734d3')]       55.509046
3  [('Fast Food Restaurant', '4bf58dd8d48988d16e9...       55.503582
4  [('Sushi Restaurant', '4bf58dd8d48988d1d294173...       55.503365

   Venue_Longitude                                  Venue_Location  \
0        37.577748        Симферопольское ш., 17 (Обводная дор.)
1        37.541756                                          Россия
2        37.546187                  Староникольская 84а, Щербинка
3        37.572914                      Симферопольское шоссе, 5Д
4        37.575996  Захарьинские дворики, д. 1, корп. 2, 117148

   Venue_Distance  Borough_Name   Venue_Category_Name  \
0             317  Южное Бутово        Hardware Store
1             181  Южное Бутово                  Lake
2             323  Южное Бутово         Auto Workshop
3             329  Южное Бутово  Fast Food Restaurant
4             294  Южное Бутово      Sushi Restaurant

          Venue_Category_Id
0  4bf58dd8d48988d112951735
1  4bf58dd8d48988d161941735
2  56aa371be4b08b9a8d5734d3
3  4bf58dd8d48988d16e941735
4  4bf58dd8d48988d1d2941735
(20864, 13)
Take a look at the dataframe data types
Cell_id                  object
Cell_Latitude           float64
Cell_Longitude          float64
Venue_Id                 object
Venue_Name               object
Venue_All_Categories     object
Venue_Latitude          float64
Venue_Longitude         float64
Venue_Location           object
Venue_Distance            int64
Borough_Name             object
Venue_Category_Name      object
Venue_Category_Id        object
dtype: object
```

In [62]:

```
# Count duplicates venues
print('Unique Venues {} of {}'.format(Moscow_venues_df['Venue_Id'].nunique(), Mos
```

Cell_id	Venue_Id	Borough_Name	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Category_Name
55.7020821…	511629f5e4b051a081439bf5	Очаково-Матвеевское	"Aminevskoe hotel" restaurant	55.703032	37.454590	Hotel
55.8350558…	5023841de4b0e6fe1a411c7d	Ростокино	"Cosmos 2" Hotel	55.836780	37.665548	Hotel
55.8277624…	505f30d2e4b0d9a2f19a319d	Покровское-Стрешнево	"Karaoke&Bar G-Voice"	55.827876	37.409241	Karaoke Bar
55.6864545…	4efb158da17cdc15b40b98fc	Очаково-Матвеевское	"MOON"	55.686766	37.414477	Furniture / Home Store
55.7213688…	5905a5870123587260ffe1d5	Южнопортовый	"Mime" Film Company (Мим Кинокомпания)	55.722946	37.679820	Film Studio
55.7488985…	5083dcc4e4b0ba1a3249d19f	Вешняки	"Red House" Клуб-Сауна	55.746088	37.838734	Sauna / Steam Room
55.7454108…	50eadc9de4b02662c430d51c	Новокосино	"Александр"	55.744217	37.877648	Department Store
55.7366957…	4eb12a04b63434fc86fa3310	Дорогомилово	"Аргумент - кафе"	55.738145	37.532077	Restaurant
55.7143244…	53a02544498e62c556da1f3f	Хамовники	"Банкет Холл" Лужники	55.715131	37.547142	Russian Restaurant
55.8692166…	5299878d11d2d1319ecea89f	Северное Тушино	"Бегемотики"	55.870727	37.440701	Kids Store
55.7623045…	50162ce6e4b01bcdb30b45e0	Крылатское	"Беговая дорожка" в Крылатском	55.762294	37.416648	Athletics & Sports
55.6249294…	4d877bec99b78cfaf7f5f91f	Орехово-Борисово Севе…	"Борисовский" билиардная	55.624427	37.709809	Bar
55.7949991…	503ccbf9e4b0708fceeb8ad1	Строгино	"Веселуха"	55.795756	37.405038	Dance Studio
55.8866119…	50420be2e4b0b5223de4c8a5	Дмитровский	"Волчий лес" / "Wolf Wood"	55.885273	37.528364	Café
55.6367977…	4f2c1f33e4b0ecad92a8352c	Коньково	"Гермес"	55.639274	37.544578	Convenience Store
55.6645507…	4f6a1b18e4b0ed0504f11293	Марьино	"Городская аптека"	55.662385	37.773821	Pharmacy
55.8777268…	50fbfea6e4b09f8ff7c27c93	Куркино	"Золотые Дуги"	55.880515	37.396922	American Restaurant
55.7902398…	4d43cae40349224b7365f34e	Восточное Измайлово	"Измайловский СДС" Филиал ГУП "Мосзеленх…	55.793075	37.823913	Flower Shop
55.7110205…	56b5e6ed498e16a72e900561	Даниловский	"Комус"	55.709422	37.657847	Paper / Office Supplies Store
55.8952978…	5558da32498ed73c64236d90	Лианозово	"Лавочки"	55.896766	37.580660	Park
55.8951981…	4ead5cf729c2a9bb97952c9e	Дмитровский	"Левый Берег" торговый центр	55.895344	37.503386	Shopping Mall
55.6521319…	4ea54de79adff6343ad6ff45	Тропарёво-Никулино	"Леди & Бродяга"	55.651273	37.470040	Pet Store
55.6833684…	51f7c3b0498e305d9ef6b5b2	Некрасовка	"Магнит"	55.683751	37.928274	Supermarket
55.8798507…	541c4831498e76f1b432ffee	Ярославский	"Магнит"	55.878228	37.729744	Supermarket
55.6628188…	51bea6bf498ea7d17efe1403	Люблино	"Мекона" Сервис	55.661802	37.807258	Auto Workshop

3. Methodology

The libraries and packages used in the Jupyter notebook are listed below:

i) Pandas – For storing and manipulating structured data. Pandas functionality is built on NumPy
ii) Numpy – For multi-dimensional array and matrix data structures.
iii) Geopandas – For storing spatial data coordinates and shape files
iv) Scikit learn – For Machine learning tasks
v) Plotly Visualization Package – For all visualizations (including maps and graphs)
vi) Requests - to send HTTP requests easily
vii) eopy – To retrieve location coordinates

The main steps for this project can be summarized in the flowchart below:

The key criteria for my research are:

- high population of the boroughs
- low cost of real estate in the boroughs area
- the absence in the immediate vicinity of the other fitness facilities

So I need to perform at least two tasks during analysis:

- first is to find boroughs with highest population and smallest housing price
- second is to provide a tool or methodology for determining vicinity of other same facilities in the boroughs

For the first task I try to use some approaches and methods of machine learning. And found out, what of the approaches suits my tasks best. I will use:

- exploratory data analysis, including descriptive statistical analysis, categorical variables analysis and correlation analysis
- segmentation with K-Means clustering

For the second task I decided to use visualization approach to mapping fitness facilities on to the interactive choropleth map and heatmap.

3.1. Exploratory Data Analysis

We have following key features in Moscow Boroughs dataset:

- District - name of the Moscow District in which Borough is belong to
- Area - area of the Moscow Borough in square kilometers
- Population_Density - population density of the Moscow Borough
- Housing_Area - housing area of the Moscow Borough in square meters

Let's analyze features and key criteria using:

- descriptive statistical analysis
- categorical variables analysis
- correlation analysis

3.1.1. Descriptive statistical analysis

The picture below shows basic statistics for all features.

As we can see, Moscow Boroughs has a very uneven population from 12 194 people to 253 943 people.

The average cost of real estate varies from 109 421 rubles/m² to 438 568 rubles/m².

	Area	Population_Density	Housing_Area	Population	Housing_Price
count	120.000000	120.000000	120.000000	120.000000	120.000000
mean	8.706417	13426.608333	1775.684167	99847.608333	190037.316667
std	4.927028	5956.551611	815.978445	44024.992123	66182.885601
min	2.110000	559.000000	69.900000	12194.000000	109421.000000
25%	5.395000	9745.750000	1244.450000	71821.750000	147339.000000
50%	7.680000	13266.000000	1709.450000	93892.000000	168172.500000
75%	10.282500	17151.000000	2206.600000	126545.750000	210978.000000
max	27.570000	30428.000000	4523.000000	253943.000000	438568.000000

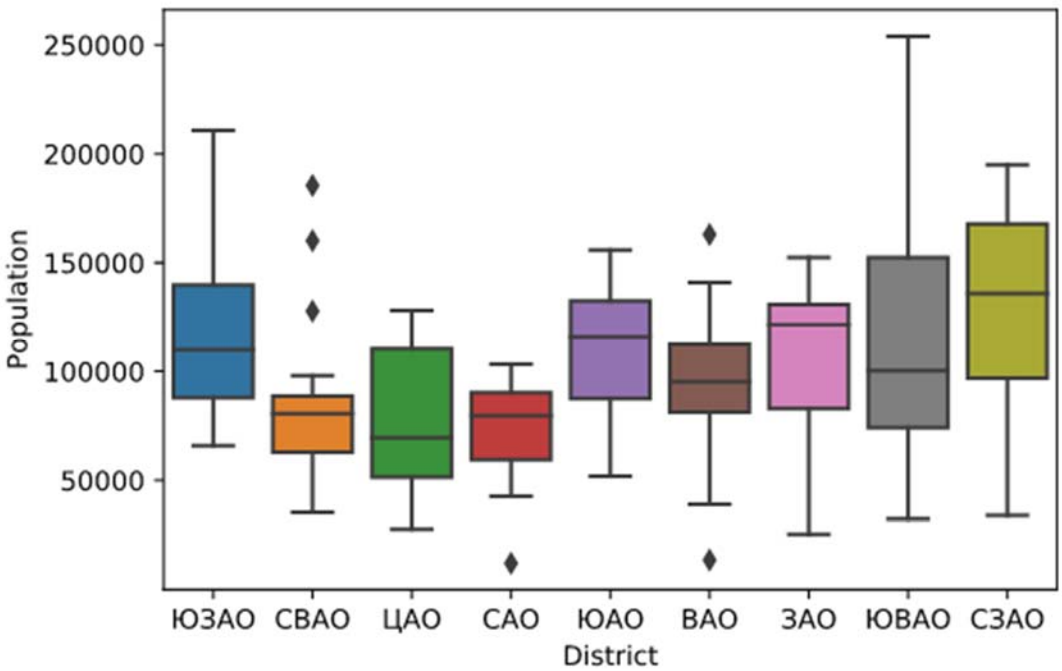
3.1.2. Categorical variables analysis

I have one categorical variable - name of the Moscow District in which Borough is belong to.

Let's analyze relationship between categorical feature 'District' and key criteria using boxplots visualization.

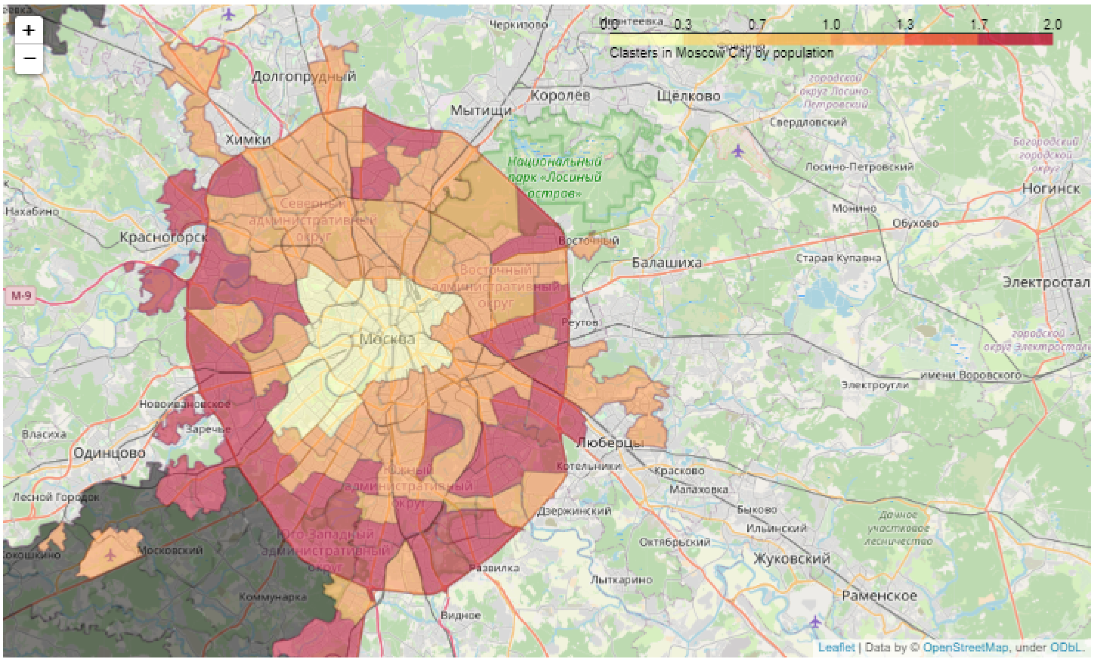
The picture below shows relationship between 'District' and 'Population'.

We can see that the distributions of Population between Boroughs in the different Districts have aт overlap, but we can estimate, that the most populated Boroughs are placed in 'ЮЗАО', 'ЮАО', 'СЗАО' and 'ЗАО' Districts. It is shown on a picture below.



```
    line_opacity=0.2,
    legend_name='Clasters in Moscow City by population')
Moscow_map
```

Out[53]:

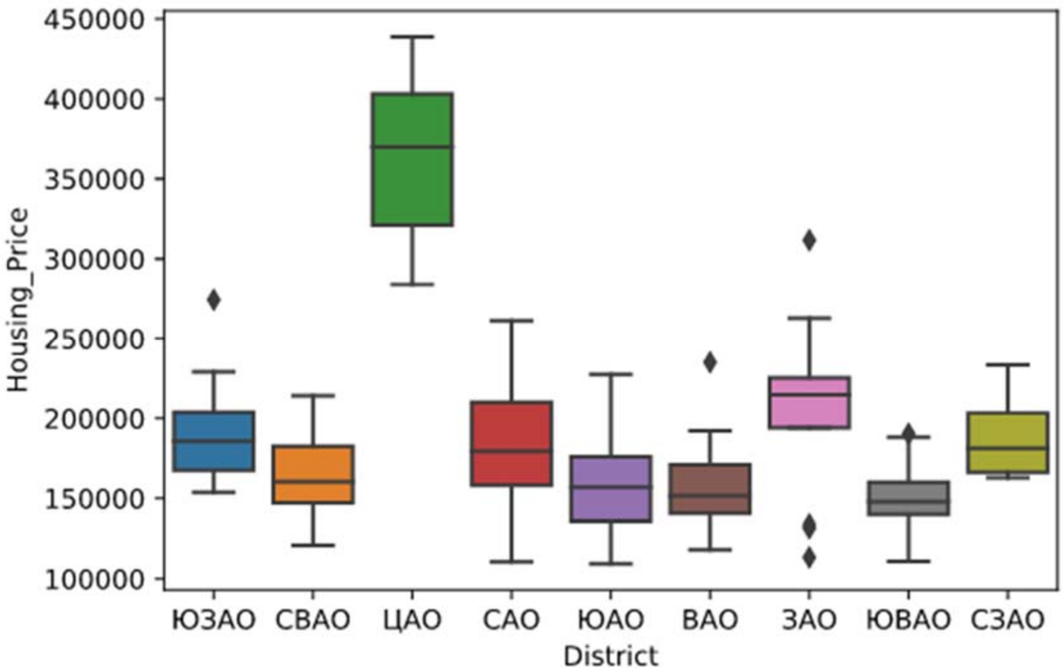


In [54]:
```
Moscow_Borough_df = pd.read_csv("Moscow_Borough_df.csv")
```

The next picture shows relationship between 'District' and 'Housing Price'.

We can see that the distributions of Housing Price between Boroughs in the different Districts are distinct enough.

As the result of boxplots visualization, categorical feature 'District' would be a good potential predictor only of Housing Price.



3.1.3. Correlation analysis

Correlation between 'Area', 'Population_Density' and 'Population' is statistically significant, although the linear relationship isn't extremely strong.

Correlation between 'Housing_Are' and 'Population' is statistically hughly significant, and the linear relationship is extremely strong.

Correlation between 'Area', 'Population_Density', 'Housing_Area' and 'Housing_Price' is not statistically significant, although the linear relationship isn't strong.

Correlation between 'Area' to 'Population_Density' is statistically hughly significant, and the linear relationship is extremely strong.

So we can exclude 'Population_Density' from our considerations.

	Area	Population_Density	Housing_Area	Population	Housing_Price
Area	1.000000	-0.585991	0.344188	0.380587	-0.154996
Population_Density	-0.585991	1.000000	0.289456	0.338621	-0.101348
Housing_Area	0.344188	0.289456	1.000000	0.887856	-0.016971
Population	0.380587	0.338621	0.887856	1.000000	-0.195774
Housing_Price	-0.154996	-0.101348	-0.016971	-0.195774	1.000000

3.2. Clustering

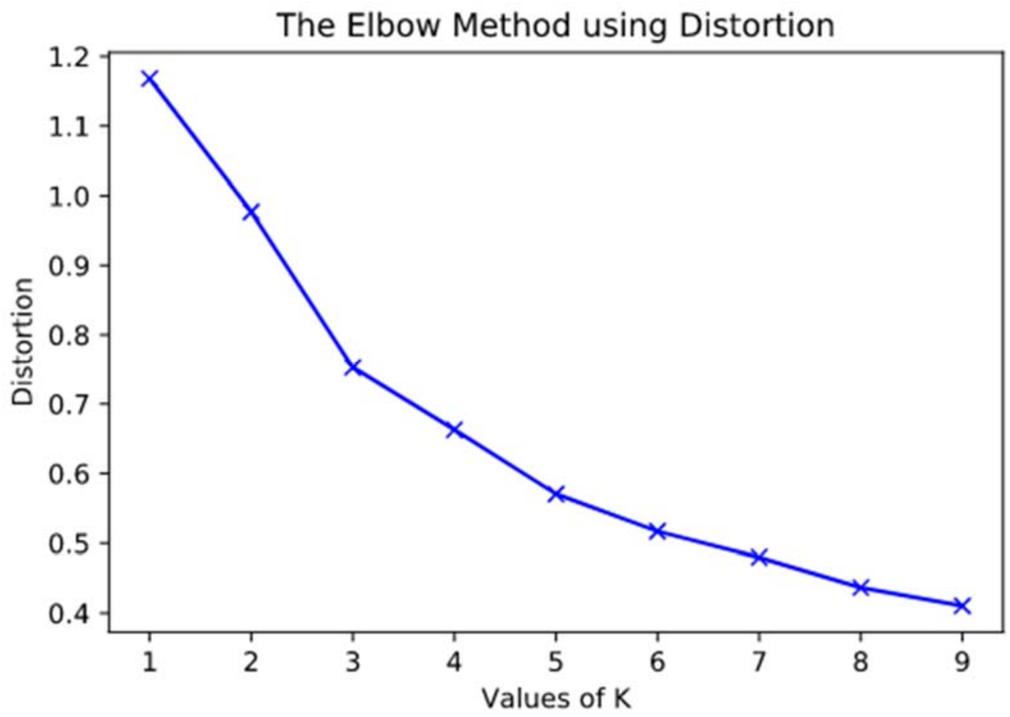
In my research, I decided to try segmentation with K-Means clustering to detect Boroughs that have highest population and smallest housing price.

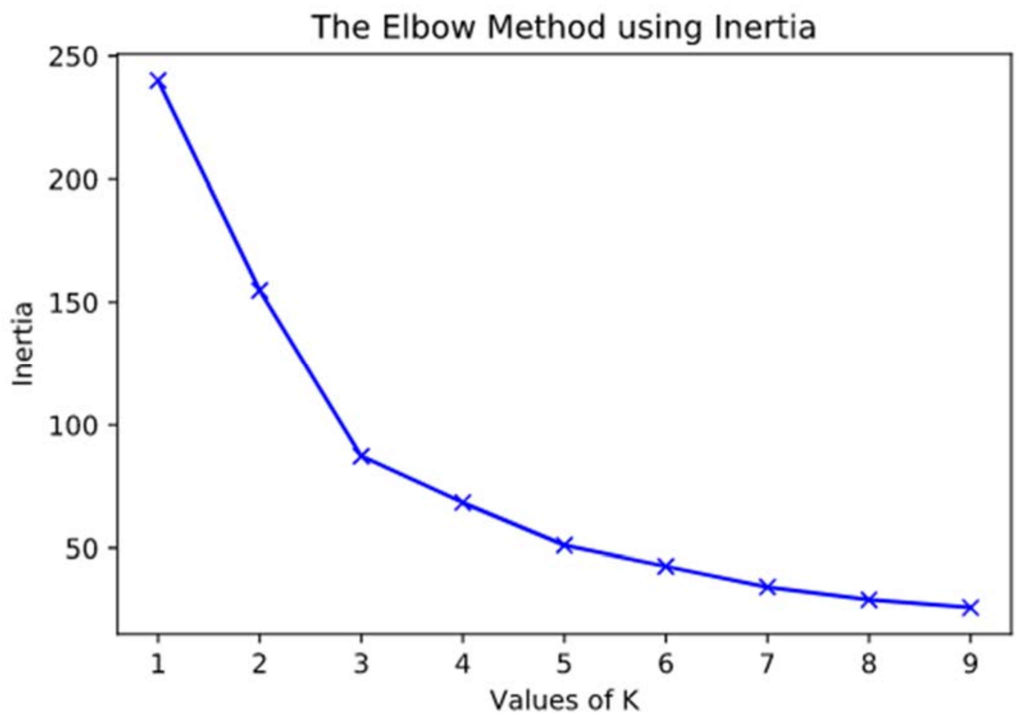
3.2.1. K-Means Clustering with elbow method

To determine right number of clusters, I used elbow method. According elbow method I implemented K-Means clustering from 1 to 10 centroids and calculate distortion and inertia for each variant.

The next pictures show elbow method using Distortion and Inertia. We can see that there are elbows at 3 and 5 centroid.

I decided to use 3 centroid in my research.





3.2.2. Analyze K-Means clusters

To analyze K-Means clusters I calculated some additional statistics:

- count boroughs in the cluster
- sum population in the cluster
- sum area of the cluster
- mean population in the boroughs in the cluster
- mean housing price in the boroughs in the cluster

- % population in the cluster to all Moscow City population
- % area of the cluster to all Moscow City area
- population density in the cluster

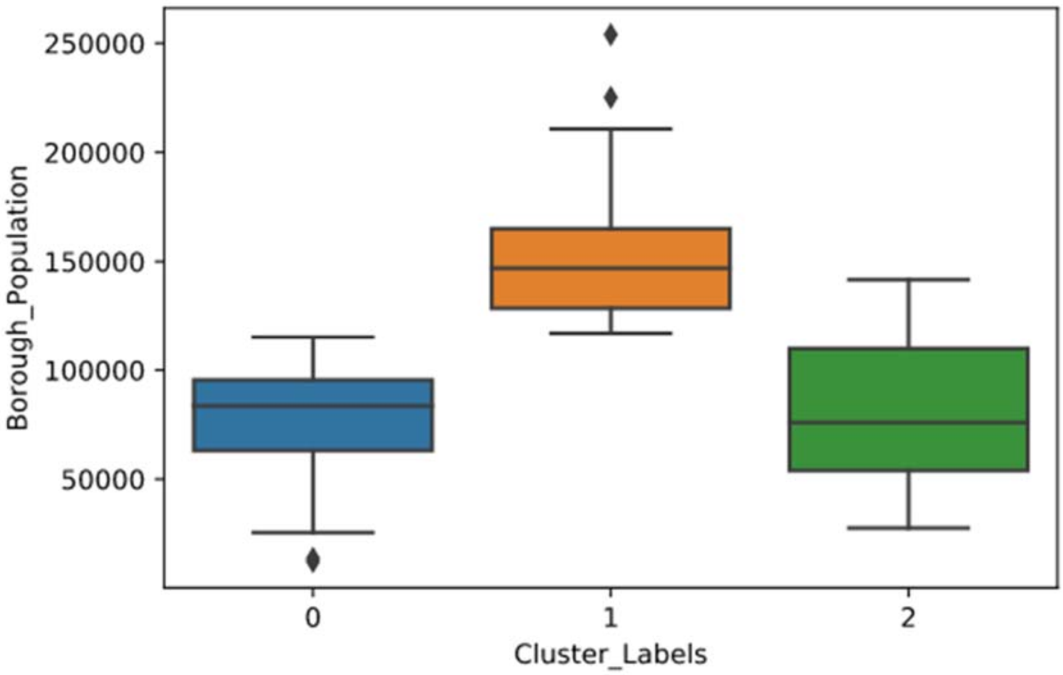
The next pictures show these statistics

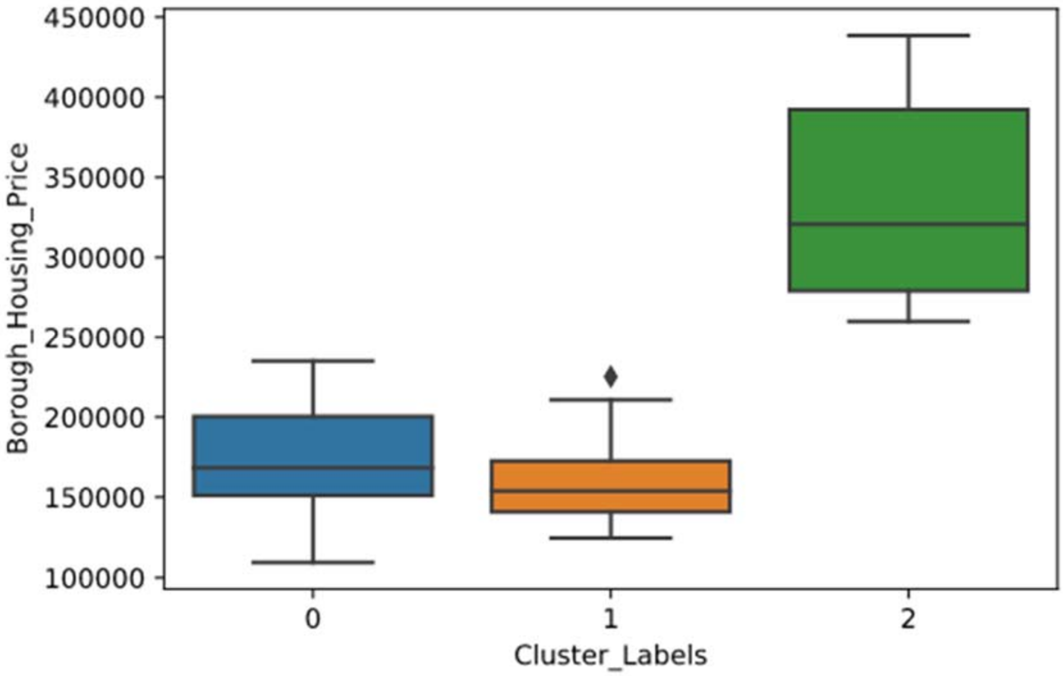
	Cluster_Labels	Population_Mean	Housing_Price_Mean	Population_Sum	Population_%	Borough_Count	Area_Sum	Area_%	Population_Density
0	0	78538.901408	173695.070423	5576262	46.539773	71	539.87	51.673574	10328.897698
1	1	153187.235294	160741.323529	5208366	43.469294	34	391.25	37.448434	13312.117572
2	2	79805.666667	333794.866667	1197085	9.990934	15	113.65	10.877992	10533.084030

As we can see, there are 3 clusters:

- "0" Cluster - characterized by low mean population (78538 people per Borough), relatively high mean housing price (173695 rubles/m²) and low population density (10328 people/km²)
- "1" Cluster - characterized by highest mean population (153187 people per Borough), smallest mean housing price (160741 rubles/m²) and highest population density (13312 people/km²)
- "2" Cluster - characterized by low mean population (79805 people per Borough), highest mean housing price (333794 rubles/m²) and low population density (10533 people/km²)

The next pictures show these clusters using boxplots visualization.





Very good result of the KMean clustering.

"1" Cluster perfectly fits my research criteria:

- boroughs from this cluster have highest mean population and smallest mean housing price
- in 34 boroughs about 43% of the Moscow population occupied only 37% of the Moscow City area, that mean the highest population density

3.2.3. Visualize clusters on choropleth map

The next picture shows all clusters on choropleth map.

As we can see Boroughs in our target "1" Cluster mostly placed in the periphery of the Moscow City.

But not all of the periphery Boroughs are well populated so not meet our criteria.

4. Results and Discussion

The result of my research consisted of:

- List of the optimal Boroughs for the location of facilities centers, according to the main criterias
  - high population of the borough
  - low cost of real estate in the borough
- List of the other competitive facilities in the each Borough from the optimal list
- Interactive choropleth map and heatmap with other competitive facilities in the each Borough

The result dataset for 10th most popular facilities is shown below.

```
for ind in np.arange(M_grouped.shape[0]):
    neighbourhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(M_grouped.iloc[ind, :], num_top_venues)

neighbourhoods_venues_sorted.head(10)
```

Out[93]:

	Borough_Name	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	Академический	Pharmacy	Coffee Shop	Park	Auto Workshop	Bakery	Health Food Store	Wine Shop	Shoe Store	Dance Studio	Supermarket
1	Алексеевский	Auto Workshop	Park	Supermarket	Pizza Place	Hotel	Food & Drink Shop	Coffee Shop	Pet Store	Convenience Store	Pharmacy
2	Алтуфьевский	Supermarket	Auto Workshop	Light Rail Station	Bus Station	Health Food Store	Pizza Place	Eastern European Restaurant	Shoe Store	Pedestrian Plaza	Park
3	Арбат	Coffee Shop	Bakery	Hostel	Hotel	Museum	Concert Hall	Plaza	Gym / Fitness Center	Caucasian Restaurant	Bar
4	Аэропорт	Coffee Shop	Café	Cosmetics Shop	Pharmacy	Park	Wine Shop	Bakery	Salon / Barbershop	Food & Drink Shop	Italian Restaurant
5	Бабушкинский	Park	Pharmacy	Gym	Supermarket	Bus Stop	Gym / Fitness Center	Baby Store	Food & Drink Shop	Fast Food Restaurant	Café
6	Басманный	Coffee Shop	Café	Caucasian Restaurant	Dance Studio	Bar	Bookstore	Gym / Fitness Center	Art Gallery	Beer Bar	Clothing Store
7	Беговой	Coffee Shop	Dance Studio	Gym / Fitness Center	Café	Restaurant	Bar	Hotel	Nightclub	Sandwich Place	Pizza Place
8	Бескудниковский	Bus Stop	Bus Line	Pizza Place	Supermarket	Bookstore	Pharmacy	Gym	Japanese Restaurant	Shop & Service	Eastern European Restaurant
9	Бибирево	Supermarket	Park	Bus Stop	Pharmacy	Gym	Sushi Restaurant	Health Food Store	Gym / Fitness Center	Soccer Field	Fast Food Restaurant

4.1. Dataset of the optimal Boroughs

Result dataset contains columns:

- **Borough_Name** - name of the Moscow Borough
- **District_Name** - name of the Moscow District in which Borough is belong to
- **Borough_Type** - type of the Moscow Borough
- **Borough_Area** - area of the Moscow Borough in square kilometers
- **Borough_Population** - population of the Moscow Borough
- **Borough_Population_Density** - population density of the Moscow Borough
- **Borough_Housing_Area** - housing area of the Moscow Borough in thousands of square meters
- **Borough_Housing_Price** - average housing price of the Moscow Borough

The picture below shows a part of this dataset.

In [32]:
```
Moscow_Borough_df.head()
```

Out[32]:

	Borough_Name	District_Name	Borough_Type	OKATO_Borough_Code	OKTMO_District_Code	Borough_Area	Borough_Population	Borough_Population_Der
0	Академический	ЮЗАО	Муниципальный округ	45293554	45397000	5.83	109387	1
1	Алексеевский	СВАО	Муниципальный округ	45280552	45349000	5.29	80534	1
2	Алтуфьевский	СВАО	Муниципальный округ	45280554	45350000	3.25	57596	1
3	Арбат	ЦАО	Муниципальный округ	45286552	45374000	2.11	36125	1
4	Аэропорт	САО	Муниципальный округ	45277553	45333000	4.58	79486	1

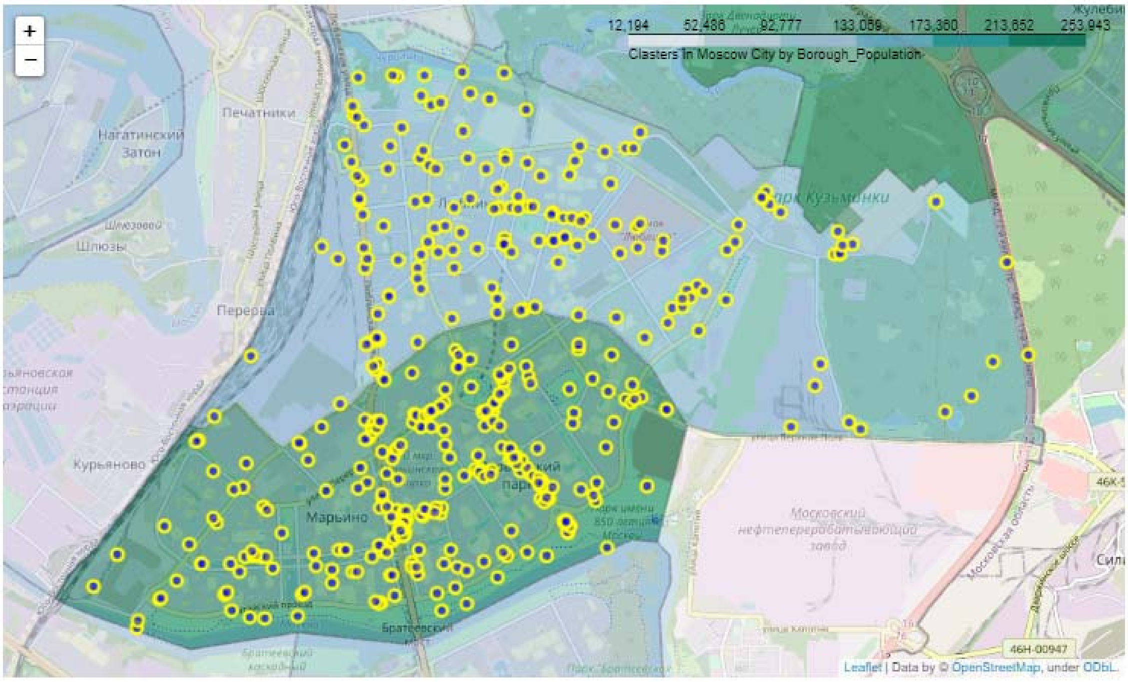
4.2. Dataset of the competitive facilities

There are 422 venues of "Auto Workshop of all 20864 venues in Moscow City. There are 419 venues of all Auto Workshop in 1 Cluster.

Result dataset contains columns:

- **Borough_Name** - name of the Moscow Borough
- **Venue_Name** - name of the fitness facilities
- **Venue_Category_Name** - category of the fitness facilities
- **Venue_Location** - address of the fitness facilities
- **Venue_Latitude** - latitude of the fitness facilities
- **Venue_Longitude** - longitude of the fitness facilities

ut[76]:



n [83]:

```
Moscow_venues_df.groupby('Borough_Name').count()
```

ut[83]:

Borough_Name	index	Cell_id	Cell_Latitude	Cell_Longitude	Venue_Id	Venue_Name	Venue_All_Categories	Venue_Latitude	Venue_Longitude	Venue_Locatio
Академический	171	171	171	171	171	171	171	171	171	17
Алексеевский	156	156	156	156	156	156	156	156	156	15
Алтуфьевский	34	34	34	34	34	34	34	34	34	3
Арбат	203	203	203	203	203	203	203	203	203	20
Аэропорт	162	162	162	162	162	162	162	162	162	16
...	...	...	...	...	...	...				.
Южное Тушино	160	160	160	160	160	160	160	160	160	16
Южнопортовый	177	177	177	177	177	177	177	177	177	17
Якиманка	321	321	321	321	321	321	321	321	321	32
Ярославский	121	121	121	121	121	121	121	121	121	12
Ясенево	292	292	292	292	292	292	292	292	292	29

Активация Windows
Чтобы активировать Windows, перейд
"Параметры".

You can see, that most populated (green) sector is totally occupied by venues, but you can open your own very close nearby in the next district (blue).

4.3. Choropleth map and heatmap of competitive fitness facilities

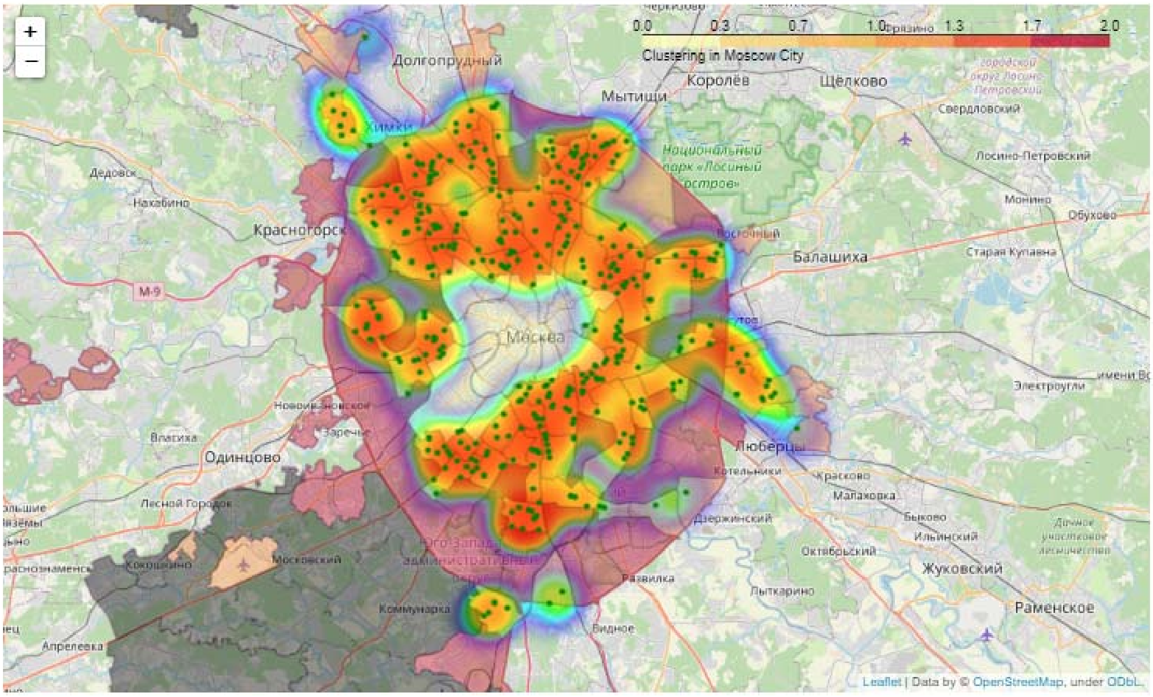
The interactive choropleth map and heatmap of competitive facilities is shown below.

```
HeatMap(heat_data).add_to(Moscow_map)

# Add facilities as markers to Moscow map
for Venue_name, lat, lng in zip(Moscow_rec_venues_df['Venue_Name'], Moscow_rec_venues_df['Venue_Latitude'], Moscow_rec_venues_df
    folium.CircleMarker(
        [lat, lng],
        radius=1, # define how big you want the circle markers to be
        color='green',
        fill=True,
        popup=folium.Popup('{}'.format(Venue_name), parse_html=True),
        fill_color='blue',
        fill_opacity=0.5
    ).add_to(Moscow_map)

# display map
Moscow_map
```

Out[173]:



5. Conclusion

In the course of my research I gathered a lot of information about Moscow Boroughs, such as:

- area of the each Moscow Borough in square kilometers
- the population of the each Moscow Borough
- housing area of the each Moscow Borough in square meters
- average housing price of the each Moscow Borough
- geographical coordinates of the each Moscow Borough
- shape of the each Moscow Borough in GEOJSON format
- list of venues placed in the each Moscow Borough with their geographical coordinates and categories

I have used segmentation with K-Means clustering to detect Boroughs that have highest population and smallest housing price. When I tested the elbow method, I set the optimum k value to 3, but there are another elbow at 5 centroid. Additional analysis can be done with 5 clusters, that can present slightly another set of optimal Boroughs for the facility location.

This visual analysis of the competitive facilities shows that although there is a great number of such facilities (more than 250), there are pockets of low density in our list of the optimal Boroughs set.

6. Appendices

[1] GitHub Repository

[2] Jupyter Notebook Viewer

[3] Moscow Boroughs

[4] Moscow Boroughs Housing Price

[5] Moscow Boroughs GEOJSON