

Venues Data Analysis of Moscow City

Determine the optimal places for the location of fitness centers in Moscow

- ▶ Owners of cafes, fitness centers and other social facilities are expected to prefer boroughs with a high population density. Investors will prefer areas with low housing costs and low competitiveness.
- ▶ On the part of residents, the preference is expected for a boroughs with a low cost of housing and good accessibility of social places.

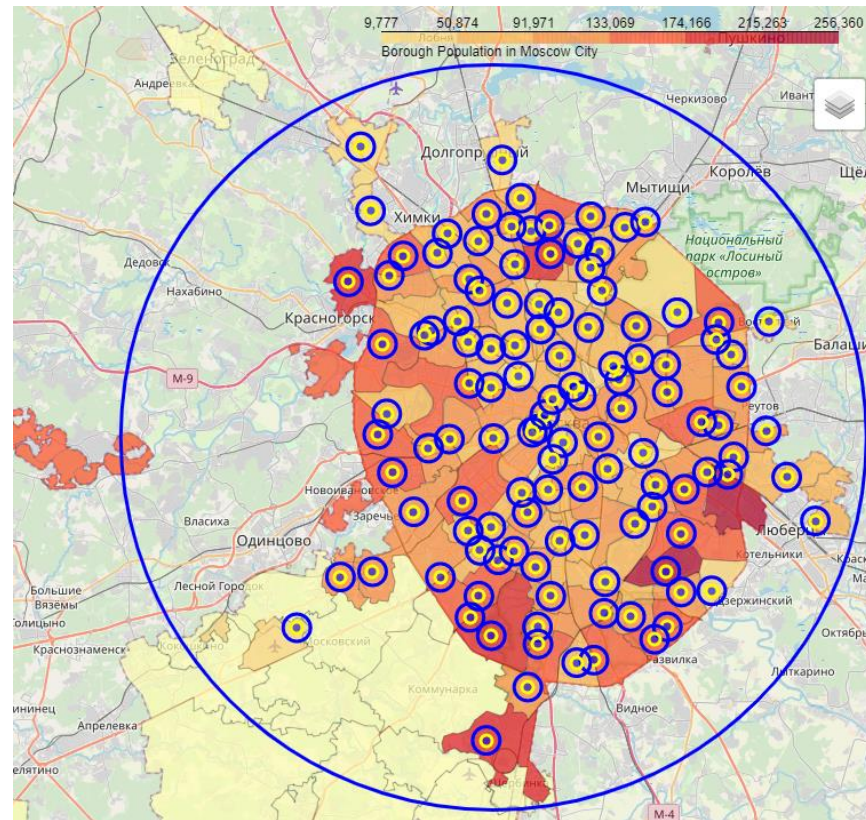
Data acquisition and cleaning

- ▶ Data for Moscow Boroughs dataset 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](#)

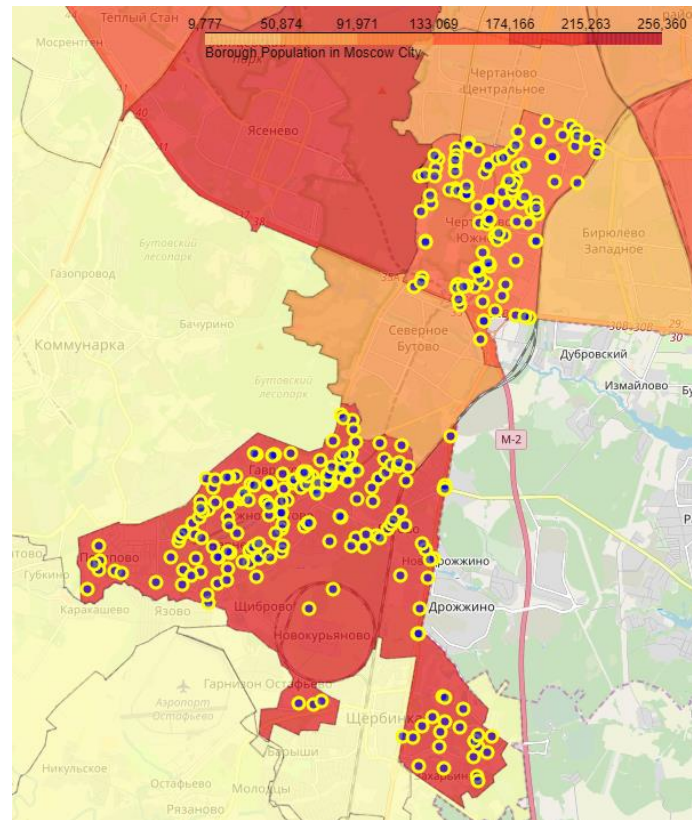
Data acquisition and cleaning

- ▶ To determine **venues** the service **Forsquare API** was used.
- ▶ The API of **Forsquare** 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 **Forsquare 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.

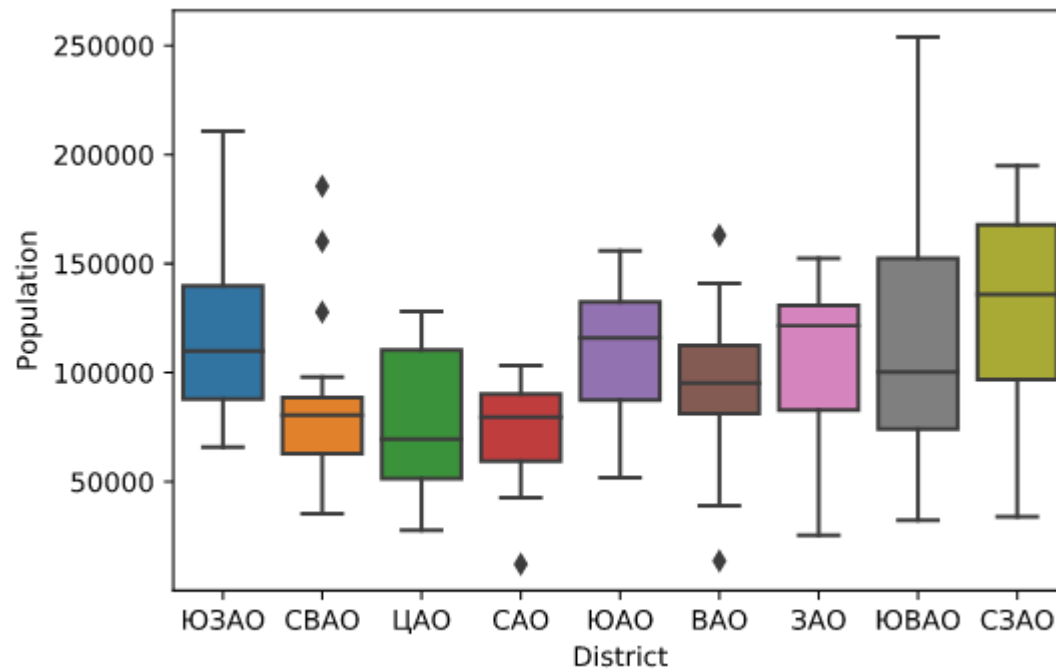
Boroughs population in Moscow City map



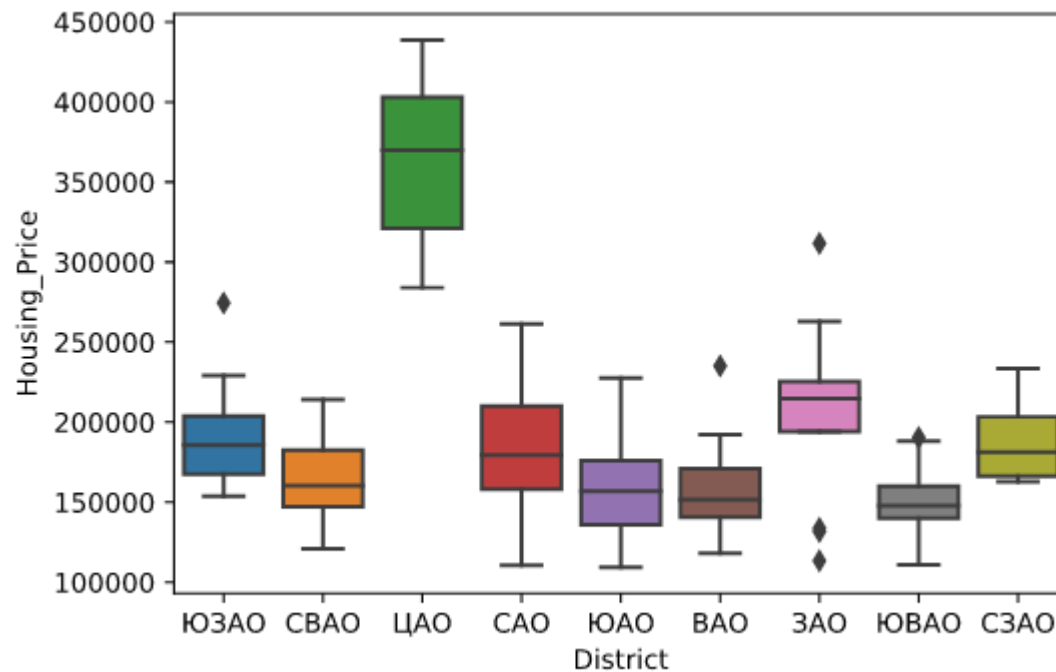
Some Moscow Boroughs and their venues



Distributions of Population between Boroughs in the different Districts



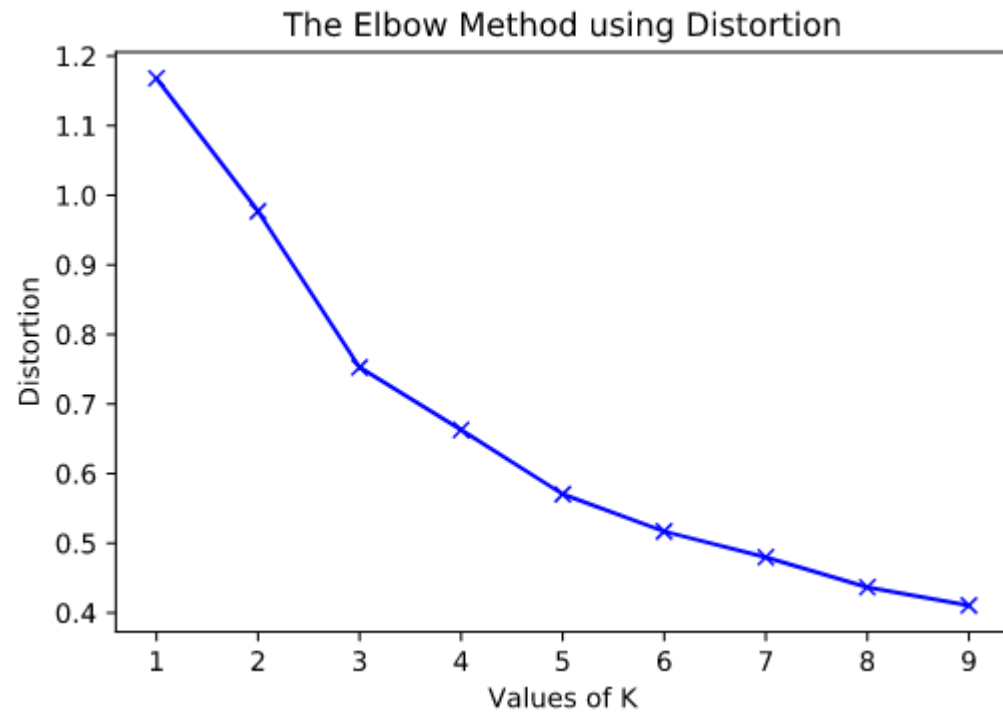
Distributions of House pricing between Boroughs in the different Districts



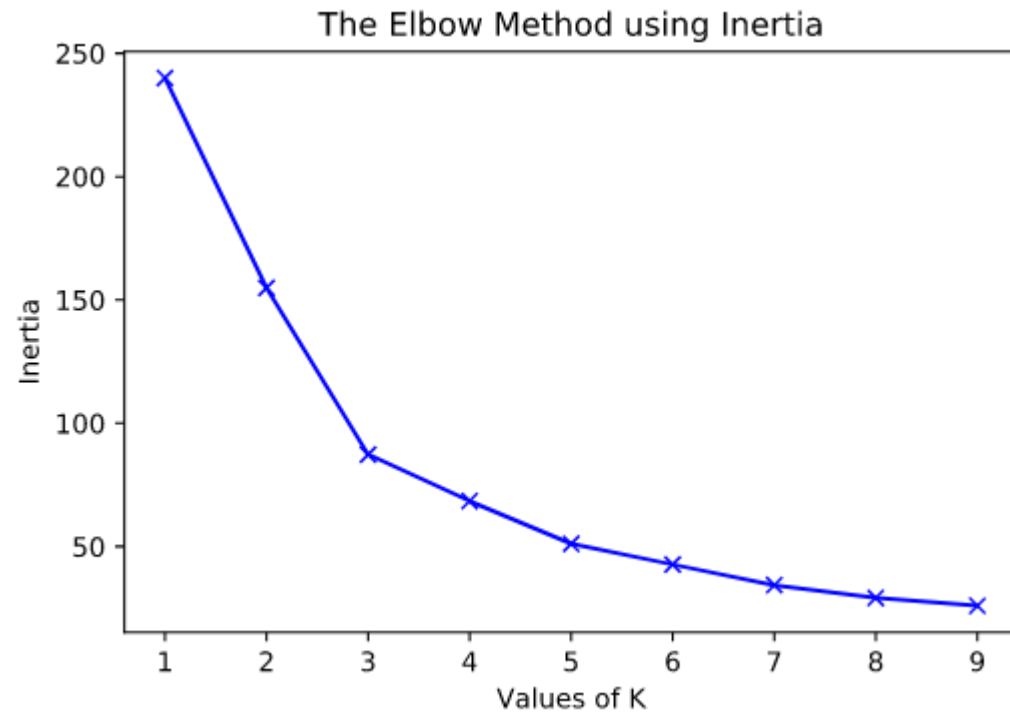
Correlation matrix



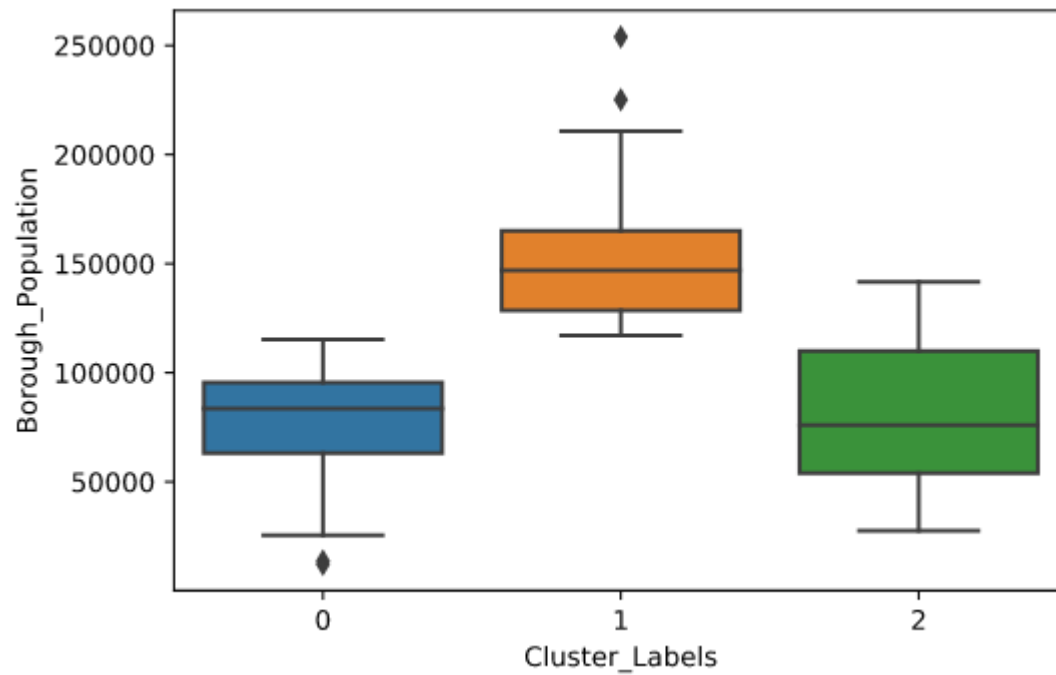
K-Means Clustering with elbow method



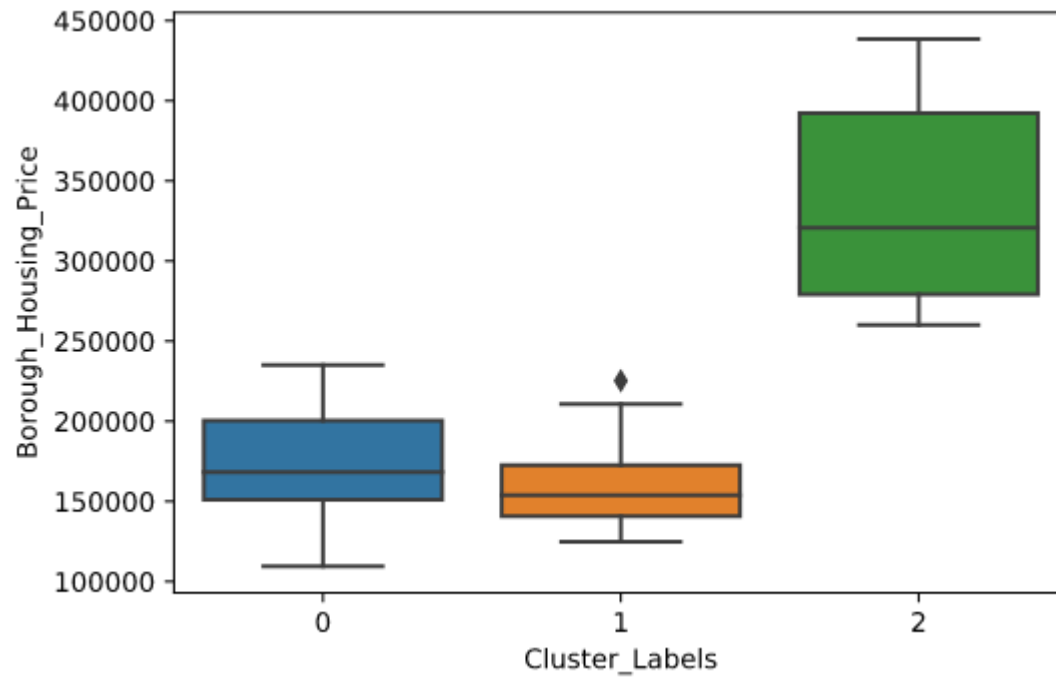
K-Means Clustering with elbow method



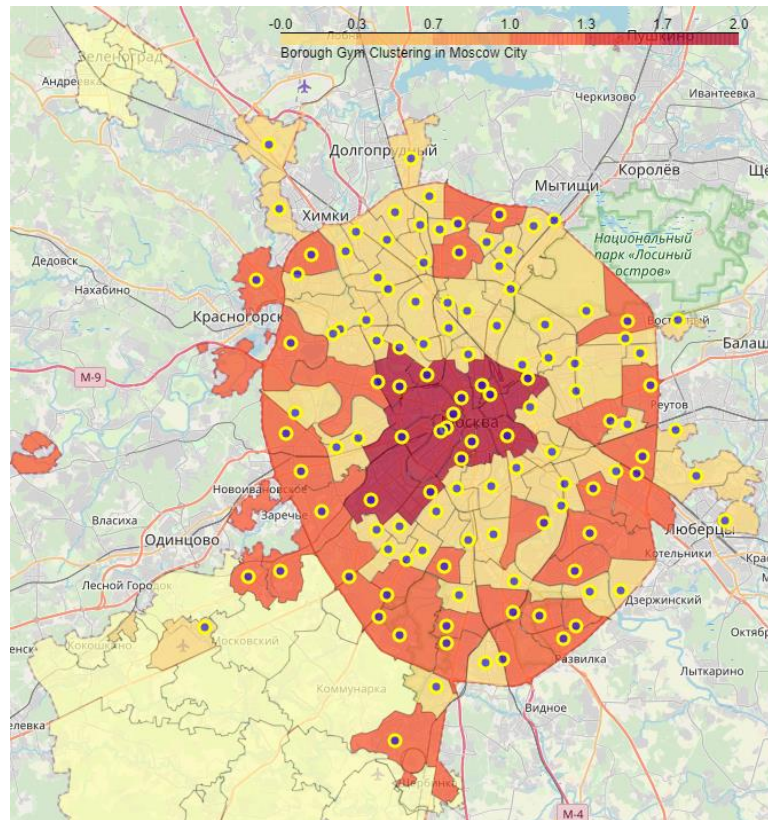
Clusters using boxplots visualization



Clusters using boxplots visualization



Vizualize clusters on choropleth map



Result:

Dataset of the optimal Boroughs

- List of the optimal Boroughs for the location of fitness centers, according to the main criterias
 - high population of the borough
 - low cost of real estate in the borough

	Borough_Name	District_Name	Borough_Type	Borough_Area	Borough_Population	Borough_Population_Density	Borough_Housing_Area	Borough_Housing_Price
0	Бибирево	СВАО	Муниципальный округ	6.45	160163	24831	2521.8	140533.0
1	Бирюлёво Восточное	ЮАО	Муниципальный округ	14.77	155863	10552	2122.2	124645.0
2	Вешняки	ВАО	Муниципальный округ	10.72	122285	11407	1976.8	147352.0
3	Выхино-Жулебино	ЮВАО	Муниципальный округ	14.97	225154	15040	3704.7	140357.0
4	Гольяново	ВАО	Муниципальный округ	14.99	163009	10874	3024.1	144398.0
5	Зюзино	ЮЗАО	Муниципальный округ	5.45	126815	23268	2132.3	180447.0
6	Зябликово	ЮАО	Муниципальный округ	4.38	133278	30428	1879.2	140687.0
7	Ивановское	ВАО	Муниципальный округ	10.19	128267	12587	2075.0	141112.0
8	Коньково	ЮЗАО	Муниципальный округ	7.18	156389	21781	3098.8	190654.0
9	Кузьминки	ЮВАО	Муниципальный округ	8.15	145984	17912	2312.7	157459.0