Capstone Project - The Battle of the Neighborhoods

"Brotherhood of the Neighborhoods"

Applied Data Science Capstone by IBM/Coursera

Konstantin Skrochinsky

June, 2020

Table of contents

Introduction

The purpose of this project is the districts of Moscow, Russia. The choice is not accidental - I was born here, studied, live and work. And most importantly - I love this city.

For those who are not familiar with the history of the city, there is a very short excursion:
The city was founded in 1147, Yuri Dolgoruky, and has a large number of cultural heritage sites (hereinafter referred to as OKN) on its territory. The whole city is divided into districts, which in turn are divided into districts and settlements (125 districts and 21 settlements). As you already understood, the main experimental subjects of this work are the districts of Moscow and historical monuments on their territory. The ultimate goal in the analysis of districts and objects of cultural heritage is to search for patterns and identify related districts according to one or another of their properties. By adding or adjusting various data sets, we can assess the level of development of the districts themselves in different directions: the availability of the necessary infrastructure (shops, public services for the population), cultural and development facilities (libraries, theaters), recreation areas (parks, recreation areas with the family) etc.

Cultural Heritage Sites in Moscow

On the territory of the city of Moscow there are more than 8000 objects under state protection. The largest number of Objects is concentrated in the Central Administrative District. On the territory of the city of Moscow there are 3 objects of cultural heritage included in the UNESCO World Cultural and Natural Heritage List:

- Kremlin Ensemble and Red Square
- Church of the Ascension at Kolomenskoye
- The ensemble of the Novodevichy Convent

Analysis Objectives

This project will go through several stages of work on the data of the city districts regarding cultural heritage objects ... and not only:

1. Classification of OKN according to their properties
2. Classification of city districts according to their cultural heritage objects
3. Obtaining infrastructure data within walking distance and creating a new classification of districts, taking into account the infrastructure that surrounds the monuments
4. Engage the Foursquare API and work with it
5. Let us analyze the ratio of the number of objects of cultural heritage and its relation to the average cost of an apartment
6. Data visualization on a map of Moscow and just graphics

The information obtained may be of interest to people interested in the history of the city, companies involved in real estate transactions in the capital, municipal authorities and everyone who just likes the numbers.
The following data and information sources are involved in the project:

1. Dataset - the basis of the project

"Objects of the cultural heritage of the city of Moscow" contains information about objects under state protection. Fields for data selection (short descriptive format):

- Object Name
- The name of the historical ensemble - when several objects enter the same category
- District - district of location of OKN
- District - district where the OKN is located
- Object category
- Security status
- Type of geometry on the map
- Geo-coordinates

2. Foursquare Service Data

To obtain the categories of infrastructure that surrounds our facilities, we need data from the service Foursquare.com
We will be frank, if it were not for the condition that Foursquare service must be present in solving the problem, the choice would be on a completely different data source. Foursquare service has a very weak information flow and its heyday ended 3-4 years ago - this is true for Moscow (this is my personal opinion and no more). Nevertheless, as a test option for the project, it suits us. In the next variation of the work, we consider a more relevant source.

3. Data on the average cost of apartments in districts

There are a lot of options on the Internet. My choice fell on the source from the site www.rlt24.com due to the presence of complete information on the districts and convenient collection of information. Finding and changing the source is quite simple.

4. Euro to ruble exchange rate

Auxiliary source for the first value of the cost per square meter of real estate to the value in euros at the date of the information on the cost of the apartment.

5. A source of information on geolocation and geoforms of Moscow regions - for visualization on a map.

Great source of information from gislab https://gis-lab.info.

Preparation and data cleaning

For further work, we need to prepare the data that we will use at various stages of the analysis. First of all, we will take data on objects of cultural heritage and their semantics from the “Portal of Open Data of the City of Moscow”, which contains a large number of datasets in various areas of the capital’s life.
Data cleaning is accompanied by checks for technical problems in the data process, sometimes there are failures on the server and we additionally store data in a local database.

We previously downloaded the table containing the coordinates of the districts and their semantics and also placed them locally on the computer - this is justified when receiving data having constant values that do not change in time (thanks to the site ***https://gis-lab.info/qa/moscow-atd.html***).

Some of the data received from the dataset for Moscow OKN contains zero values - these are **ObjectNameOnDoc** and **EnsembleName** columns. We delete these fields.
The **EnsembleNameOnDoc** field contains information on whether the object belongs to the ensemble, i.e. included in the joint group of cultural heritage sites. We convert this field into a logical format containing information about such an affiliation and add this field to our model for segmentation.

```
Columns with NaN:
                         Number                    0
                         global_id                 0
                         AISID                     0
                         ObjectNameOnDoc           0
                         ObjectName             8316
                         EnsembleNameOnDoc      2973
                         EnsembleName           8316
                         AdmArea                   0
                         District                  0
                         SecurityStatus            0
                         Location                  0
                         Category                  0
                         ObjectType                0
                         geoData.coordinates       0
                         geoData.type              0
```

The next data stream is the average cost per square meter of housing in the areas of Moscow, which we will use for paragraph 5 of our goals. Such data is fairly easy to find on the Internet, but we still need to clean and adapt it.
Add the field "Cost in Euros" to the table (*PricrEuro* - the euro exchange rate, at the time of receiving the data, we take it from one of the numerous sites with the exchange rate using the parsing method). The obtained data on the name of the districts had to be cleaned a little so that they had the same spelling in all the tables used. Basically, cleaning the names consists of replacing lowercase letters in spelling areas and replacing the letter "e" with "ё" (Russion symbol).

Information on the average cost of 1 square meter of housing on January 6, 2020

Sorry for the Russian words in the data set - there was only such an option :(

0	NAME	PriceRub	PriceEuro
1.	Хамовники	755473	10888.916114
2.	Якиманка	681050	9816.229461
3.	Арбат	672148	9687.921591
4.	Пресненский	556598	8022.456039
5.	Тверской	541558	7805.678870
6.	Замоскворечье	510406	7356.673393
7.	Мещанский	445415	6419.933698
8.	Раменки	434873	6267.987893
9.	Басманный	429737	6193.960796
10.	Дорогомилово	387803	5589.550303

Our analysis is accompanied by cartographic visualization with the support of the **folium** package, which has excellent tools for working with geodata and an excellent presentation style.
The OKN database contains information about the location of objects in the city in the geojson format. All objects in the database are divided into three types according to their geometric characteristics: point - a small object (for example, sculpture, grave); a polygon is an object having sufficient dimensions so that it can already be drawn completely on the map (for example, a building) and a multipolygon is an object, as a rule, relating to one window, but consisting of several polygons (a composite object). To simplify the data model, we will take data on the geolocation of cultural objects only through their center:

- Objects **Point** - already contains these coordinates.
- Objects **Polygon** - we calculate the center point through the prescribed centroid function (unfortunately, we failed to import the *geopandas* library, which has a built-in tool for such work)
- Objects **Multipolygon** - we take only the first object in its composition and then its center point

Methodology and Data Analysis

This section is dedicated to one of the popular K-mean method clustering methods

Let me remind you that the main goal of this work is to segment Moscow regions using the influence of cultural heritage objects on them. The analysis uses the rather popular K-mean method of clustering. This method has both pros and cons. From the pros, we highlight:

- Ease of implementation
- Algorithm performance

In the minuses add:

- Lack of guarantee to achieve a global minimum, i.e. the search is reduced to a local minimum (easier - not always segmentation will be successful in terms of logic)
- The need for preliminary selection of the number of segments The method can be called "Cheap and cheerful", and who said that this is bad? :)

The step-by-step algorithm will look like this:

- Clustering all windows
- Aggregation and data overlay on the districts of Moscow with the subsequent segmentation of districts
- Connection of Foursquare service for receiving “points of attraction / interests” (POI) surrounding OKNs within walking distance

- Data connection with the average cost of apartments by district
- Merging all received data into one dataset and final segmentation of Moscow districts

In the meantime, the next merger of the two tables. By the way, here we draw your attention to the fact that not all areas of Moscow have OKN: **23** of **146** districts simply do not have them. Alas, apparently our ancestors were not interested in these locations.

```
Number of districts without OKN: 23
Total districts:         146
0                   Киевский
13                    Силино
14                 Кокошкино
24               Тёплый Стан
29             Ломоносовский
37                  Солнцево
56       Проспект Вернадского
64                  Щербинка
68           Бескудниковский
74          Восточное Дегунино
78                Алтуфьевский
80       Чертаново Центральное
94                   Нагорный
100        Северное Медведково
101         Нагатино-Садовники
116               Метрогородок
122    Орехово-Борисово Южное
123                   Марьино
129                  Братеево
130                 Зябликово
131                 Рязанский
140        Восточное Измайлово
144                Некрасовка
```

We collect the final table with data that contains information about the districts themselves, the price per square meter of housing and the geometric center of the district for visualization on the map, plus several more indicators:

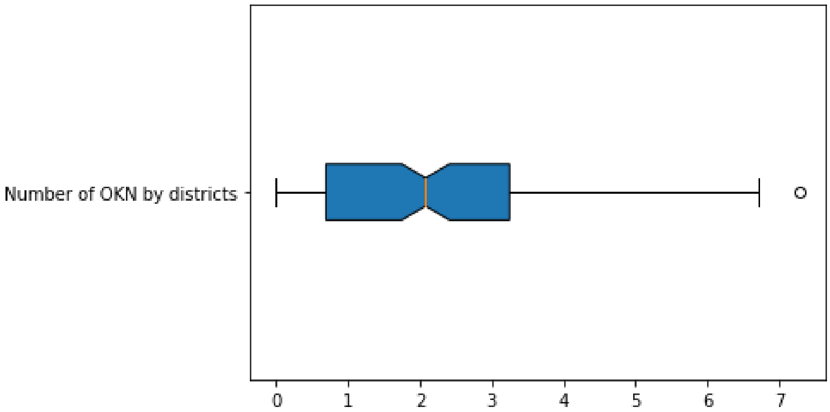
- ***LvL_Price*** - he gradation of the price of the region from 0 to 145. The less, the more expensive. Hit parade of capital prices
- ***Qnt_CHS*** -is the number of OKNs in the district
- ***Differ*** - the rating of districts in terms of price level and the number of OKNs in the district, the higher, the greater the difference in the rating indicators. Hit parade of attractiveness of the area, if you were only interested in the presence of ancient monuments in their territory
- ***Avg.SaleOnChs*** - the ratio of square meter of housing to the number of windows. The smaller, the "more profitable"

Let's go over one of the district ratings - in the category **Differ**. The leaders are the settlements of New Moscow (the expansion of the city due to land lying in the southwest took place in 2012) - this victory was achieved due to cheap apartments and a relatively high level of OKN in their territory. Looking ahead, we say that 75% of the districts of Moscow have this figure of 25 OKN. The district **Lefortovo** deserves a special prize - the largest number of OKNs is **374** with relatively "inexpensive" apartments. The result - 25th place with a cost of sq.m at 3219 euros. District **Khamovniki** - for him there is a special prize - for uniqueness:

- the most expensive area - 10888 euros per sq.m (Ostozhenka and Prestizhenka streets make their contribution)
- the largest number of OKN - 1448 pcs. (one of the oldest districts of Moscow)
- Indicator **Differ** - 0 (zero) the most balanced area for this indicator
- The lowest rate **Avg.SaleOnChs** - only 7.51 euros for one OKN in sq.m (no matter how strange it sounds)

The lower ranks of the rating are occupied by areas with high housing prices and a small number of OKNs in their territory.

1. In no case should the concept of the quality of living in a particular area be interfered with by our analysis - it covers only a small fraction of the data and applies only to cultural monuments (someone may just hate junk and then the rating will have other parameters for values)



2. The value inf (infinity - infinity) in the "Avg.SaleOnChs" indicator indicates that you don't have access to any monument for any money. "No pens - no candy"

The number of cultural heritage sites in each district

Only 13 districts have more than 100 OKN in their territory. Logically, the historical city center has the largest number of objects. And yes, the Khamovniki district with 1,448 objects is allocated against this background among all the districts. And as you can see from the table with housing prices - this is the most expensive area of Moscow. By the way, more than 50% of the OKN in the Khamovniki area is brought by the "Novodevichy Convent Ensemble" and the Novodevichy Cemetery burial site.

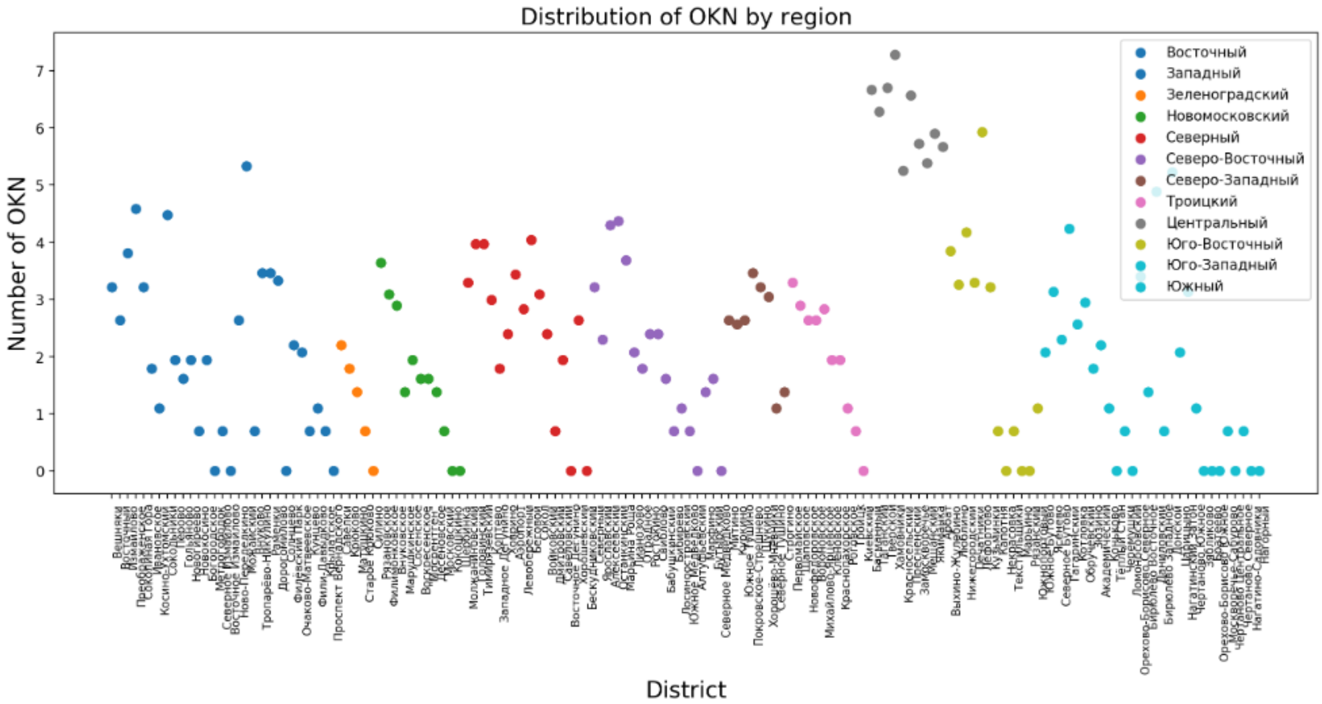
A small statistical calculation of the number of monuments in the districts. If not interested - you can run past.

From top to bottom:

1. ***count*** - 146 number of districts / settlements
2. ***mean*** - 56.95 average number of OKN per district
3. ***std*** - 173.88 standard deviations (google, this is statistics).
   *It can be said that the indicator for the number of OKN in the regions is highly unbalanced: the ratio of indicators ** mean **and** std ** differ by more than three times*
4. ***min*** - 0 OKN the smallest indicator
5. ***25%*** - 1 OKN indicator for 25% of districts
6. ***50%*** - 7 OKN in 50% of the districts
7. ***75%*** - ~ 25 OKN in 75% of the districts
8. ***max*** - 1448 maximum number of OKN (***KHAMOVNIKI Forever***)

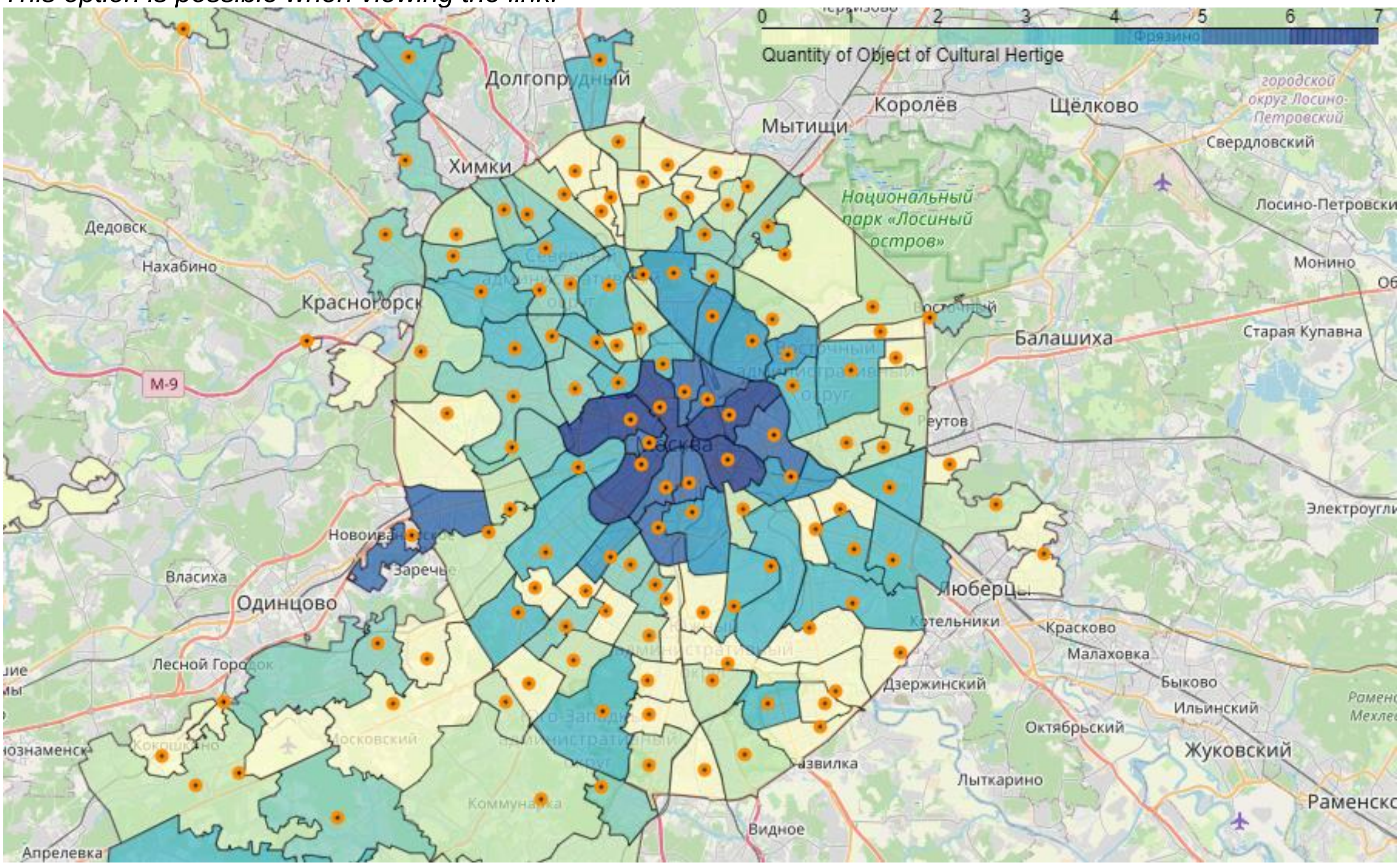
```
count     146.000000
mean       56.958904
std       173.885332
min         0.000000
25%         1.000000
50%         7.000000
75%        24.750000
max      1448.000000
```

Schedule of districts and districts (difficult to read, I know). A bunch of gray dots to the right and up is a galaxy of "elite" districts of the Central District. Four "stars" trying to reach the same heights are the green "star" to the right of the galaxies - "Lefortovo" from the South-East of the okrug, the "Mozhayskaya" star from the West and turquoise "stars" - "Donskaya" and "Danilovskaya" from the Southern District.



Cartography time! Let's build a Choropleth map (well, such a name) of Moscow with a color scheme by the number of monuments in the districts. The only problem, as we have seen, is a very large outburst in the number of OKN in the central regions. We will fight it through data normalization - the logarithmic transformation will give a much better result, the more we have already added such a column to our table.

When markers are activated, information is provided about the area and the number of OKNs on its territory. This option is possible when viewing the link:

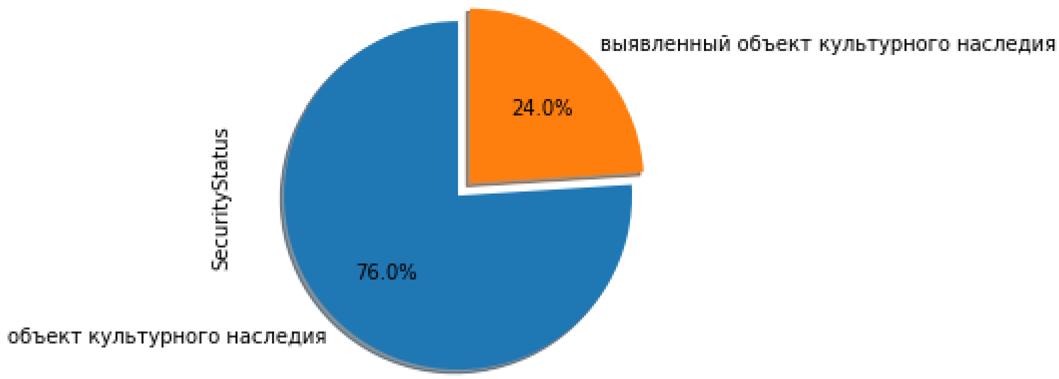


The superiority of the historical center over other districts is clearly visible. The South-Eastern District holds the second place at the expense of the Lefortovo district - with the date of foundation in 1699, it is rightfully one of the old districts of Moscow and has a decent number of monuments allowing it to compete with the central regions.
A small table on the number of OKN by city districts.

Segmentation of OKN and districts by the k-means method

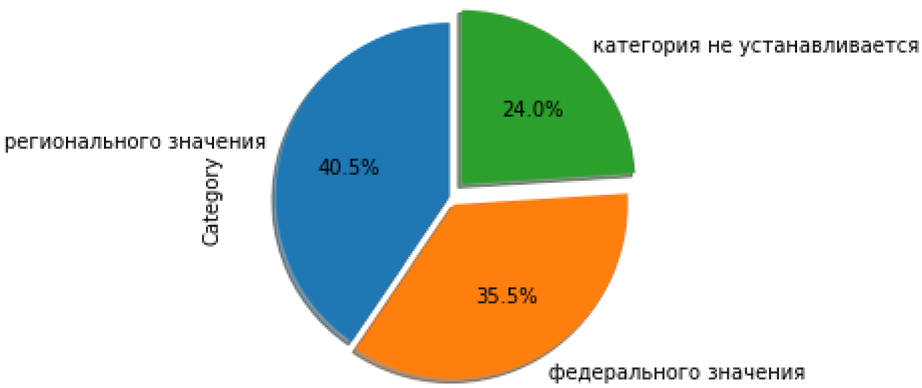
The time has come to see what is interesting in the dataset with objects of cultural heritage and which of these data we will take for analysis. In the [data](# data) section, we have already deleted some of the empty fields and determined the fate of cultural ensembles. What else do we have?

SecurityStatus - the status of the object according to its protected level. “Identified object of cultural heritage”(orange) is a relatively new window that will change its status when it is entered into the registry.




```
объект культурного наследия              6321
выявленный объект культурного наследия   1995
```

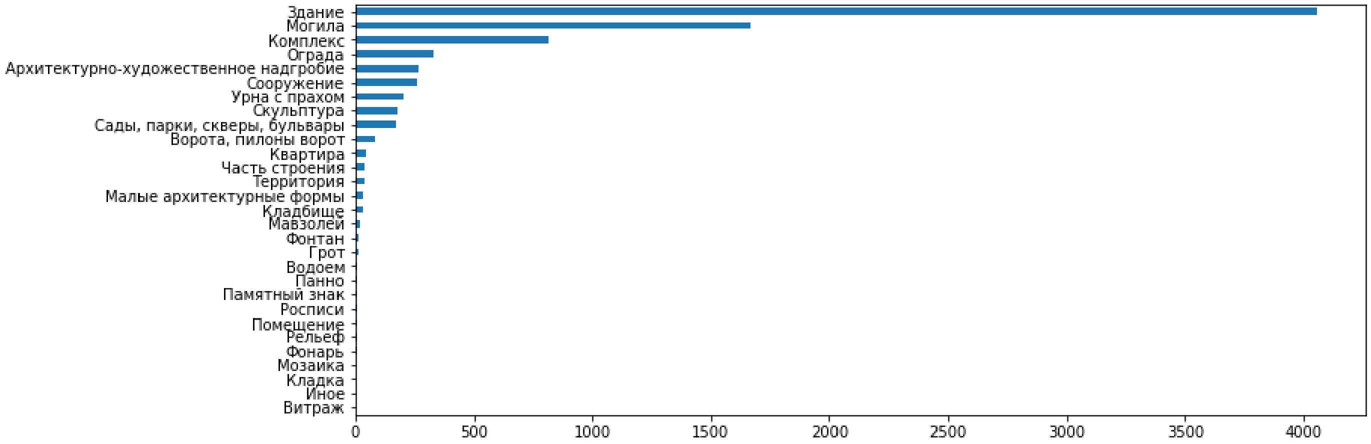
Category - the status of the object relative to its value for history (not sure if it was correctly expressed). The highest level is “Federal Significance”(orange), the medium level is “Regional”(blue) and the lowest is “category is not established”(green) in our case.




```
регионального значения         3367
федерального значения          2954
категория не устанавливается   1995
```

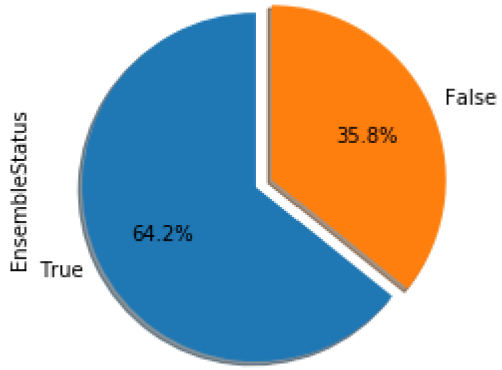
ObjectType - it’s quite simple here - it is a classification according to the type of objects.

And note that a huge number of objects are associated with burial: Grave, Tombstone, Urn with ashes, Cemetery, Mausoleum. It is possible that data of this type could be given an additional status that clearly expresses their affiliation. But in this paper we will not do this, we will leave all the additional improvements for the next version.



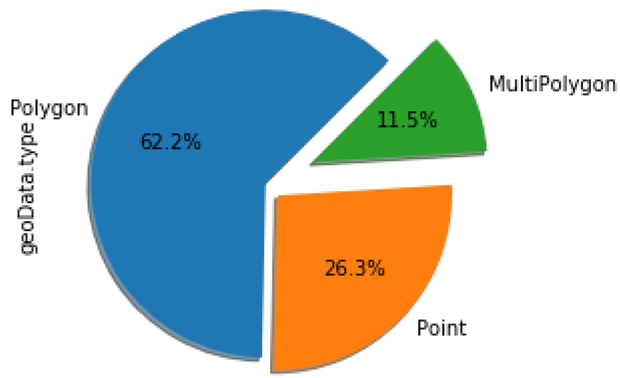

```
Здание                                  4055
Могила                                  1666
Комплекс                                 812
Ограда                                   329
Архитектурно-художественное надгробие    269
Сооружение                               260
Урна с прахом                            204
Скульптура                               181
Сады, парки, скверы, бульвары            171
Ворота, пилоны ворот                      82
```

EnsembleStatus - We have already talked about this indicators, the status of the entry of the object into the ensemble.




```
True     5343
False    2973
```

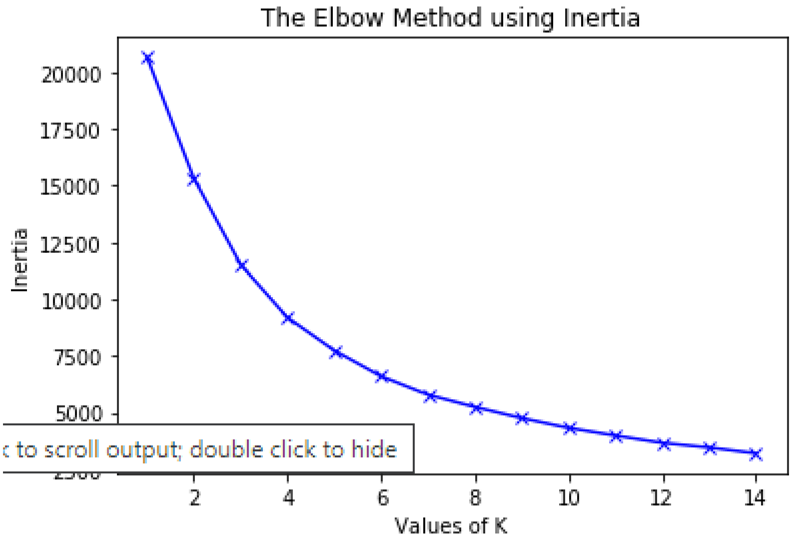
geoData.type - and this was already discussed, the type of geometric configuration of the monument when projected on a map.




```
Polygon         5170
Point           2188
MultiPolygon     958
```

I suppose criticism and I'll say right away - yes, the collection on the semantics of the monuments could be better, at least the year the object was created would look just fine in the set. But we work with what we have and understand that we have a foundation for improvement - now we are building a basis for analysis.

We turn to the segmentation of cultural objects and first of all we determine what number of clusters we need for our data model. The elbow method, which is often used for such tasks, will help us with this. One of the inconveniences in this method is the determination of the number of clusters through visual analysis of the graph. Our choice in quantity will be determined at the point where the line is broken (if very simple).



As you can see, the solution in this case is not the simplest, but by brute force (I chose between 3.4 and 5 clusters), we stop at the number of segments equal to four.

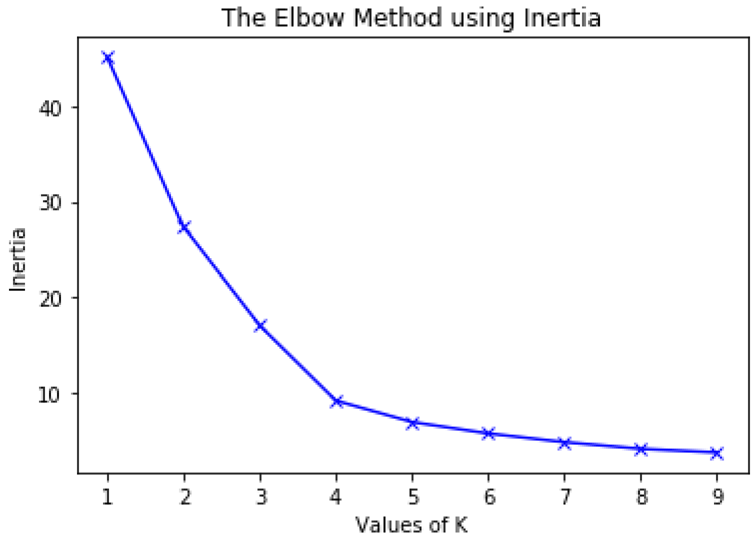
We start the algorithm to process our data ...and look at our resulting clusters

An analysis of the segmentation of our objects is as follows:

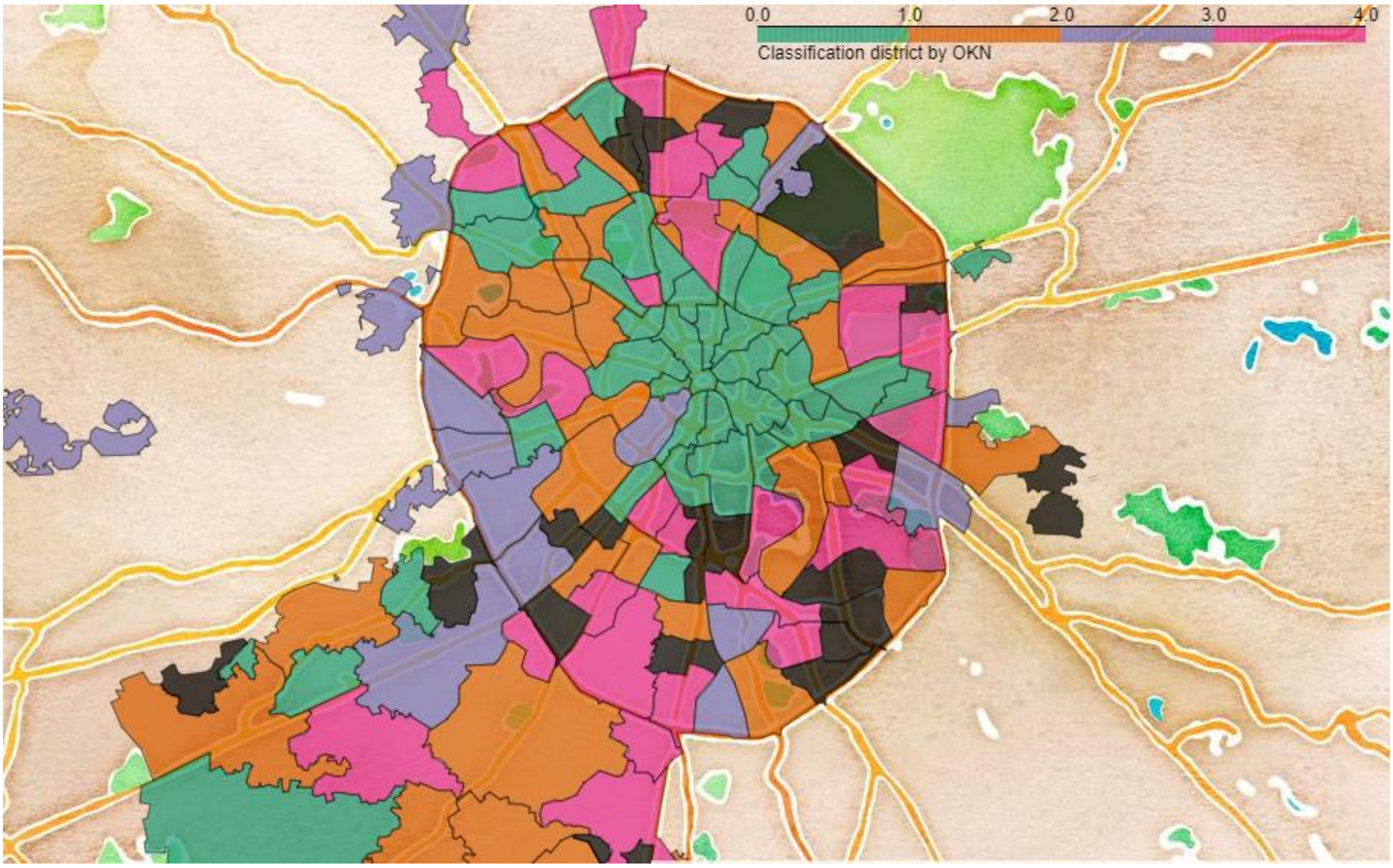
- 0 Cluster - OKN only of regional importance, large objects: ***Large regionals***
- 1 Cluster - the OKN category is not installed, the mix in all other parameters: ***Status-free***
- 2 Cluster - only "point" OKN of federal and regional significance. Burial only: ***Burial***
- 3 Cluster - large OKNs of federal significance: ***Large federals***

We proceed to the merging of the data of our regions with segmented Windows and we get a unique combination (we can say OKNscopic imprint) of the region. But unlike people, it will not be unique (except for areas where the number of historical values is very decent)

Again, we use the "elbow" method to obtain the number of clusters. Now the choice is quite simple - this is again 4 clusters ... a coincidence?



Let us take a look at the work done to segment our regions regarding the types of antiquities located on their territory. Despite the fact that we have 4 clusters, in fact we will also have a fifth - these are areas that do not fall into our classification due to the lack of OKN in their territory (shown in black).



From observations:

- The center is kept together
- Khamovniki feel uncomfortable in the galaxy of "elite" and rushes to the outlying areas ... interesting
- New Moscow is not worse than the center, at least in terms of the OKN ratio

Let's see what's in the segments:

	0	1	2	3
Cluster Labels				
0	0.253601	0.451706	0.080170	0.214523
1	0.874915	0.061197	0.015772	0.048116
2	0.105691	0.034834	0.797566	0.061909
3	0.136137	0.036459	0.032500	0.794904

Districts

- **Mix** or "green" - 0 Cluster - Prevalence of ***Statusless*** monuments, plus equally ***Large federals*** and ***Large regionals***. Pronounced center. New Moscow is versatile :))
- **Big Region** or "orange" - 1 Cluster - ***Large regionals*** rule here. A large group of areas outside the Garden Ring (historic road circle of the city)
- **Burial** or "lilac" - 2 Cluster - areas with large graves. Identified aspirations of Khamovnikov
- **Big Federal** or "pink" - 3 Cluster - large and there ***Large feds***. Another group of districts outside ...

Add data with Foursquare

The time has come for the once-popular Foursquare service to give us some more data to increase the diversity of our clustered areas with information on points of interest (POIs) that are within walking distance of cultural heritage sites.

Point one: we prescribe all personal keys for connecting to the service API, set the step-by-step availability at 500 m, prescribe the maximum number of POI categories that we will take for analysis (30 points).

Next, we write the main function that collects the necessary data from the flow of information that is issued upon request.
How the Foursquare data acquisition service works - we give them the coordinates of our OKNs, they return the POI category, its coordinates and name to us (in fact, there is a lot more data). The quantity depends on the presence of interesting places within walking distance. By the way, it's possible that cultural monuments may fall into our search :)

Due to the limitation of 2500 requests per hour, we have to split our request into four and then glue them together.

And here is our fresh little dataset: almost 120 thousand lines and 421 one unique category of POI! Ufff, more than 50 million values) Of course, we saved it in advance and now use it for our purposes.

```
(119763, 7)
There are 421 uniques categories.
```

	global_id	global_id Latitude	global_id Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	2949468	55.758196	37.568968	Боэми	55.756670	37.570165	Eastern European Restaurant
1	2949468	55.758196	37.568968	Своя школа	55.758648	37.570456	Dance Studio
2	2949468	55.758196	37.568968	Волконский	55.756251	37.571128	Bakery
3	2949468	55.758196	37.568968	The Toy	55.756198	37.566602	Restaurant
4	2949468	55.758196	37.568968	Правда кофе	55.756579	37.565897	Coffee Shop

Housing price analytics in Moscow

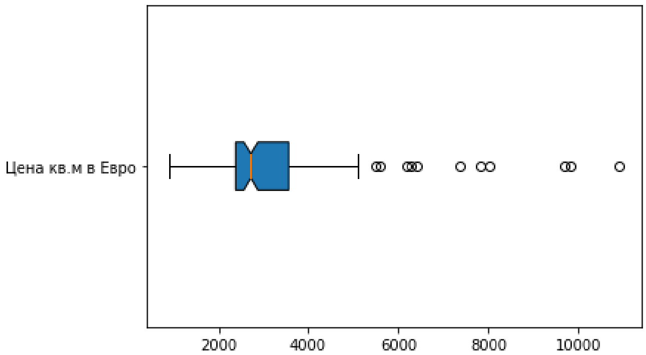
Let's step back a little from history and look at the present - how and where to live in Moscow.

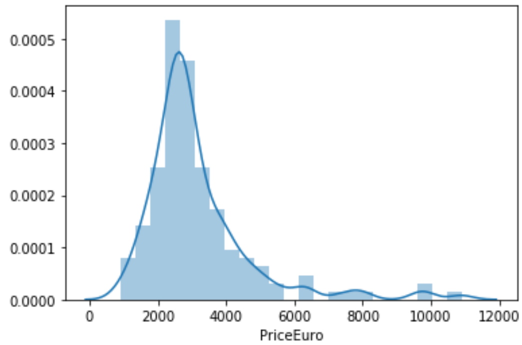
What is interesting about price offers for housing?

	PriceRub	PriceEuro
count	146.000000	146.000000
mean	219362.143836	3161.748974
std	110825.860537	1597.374756
min	62867.000000	906.125685
25%	165587.250000	2386.671231
50%	188667.000000	2719.328337
75%	247219.750000	3563.271116
max	755473.000000	10888.916114

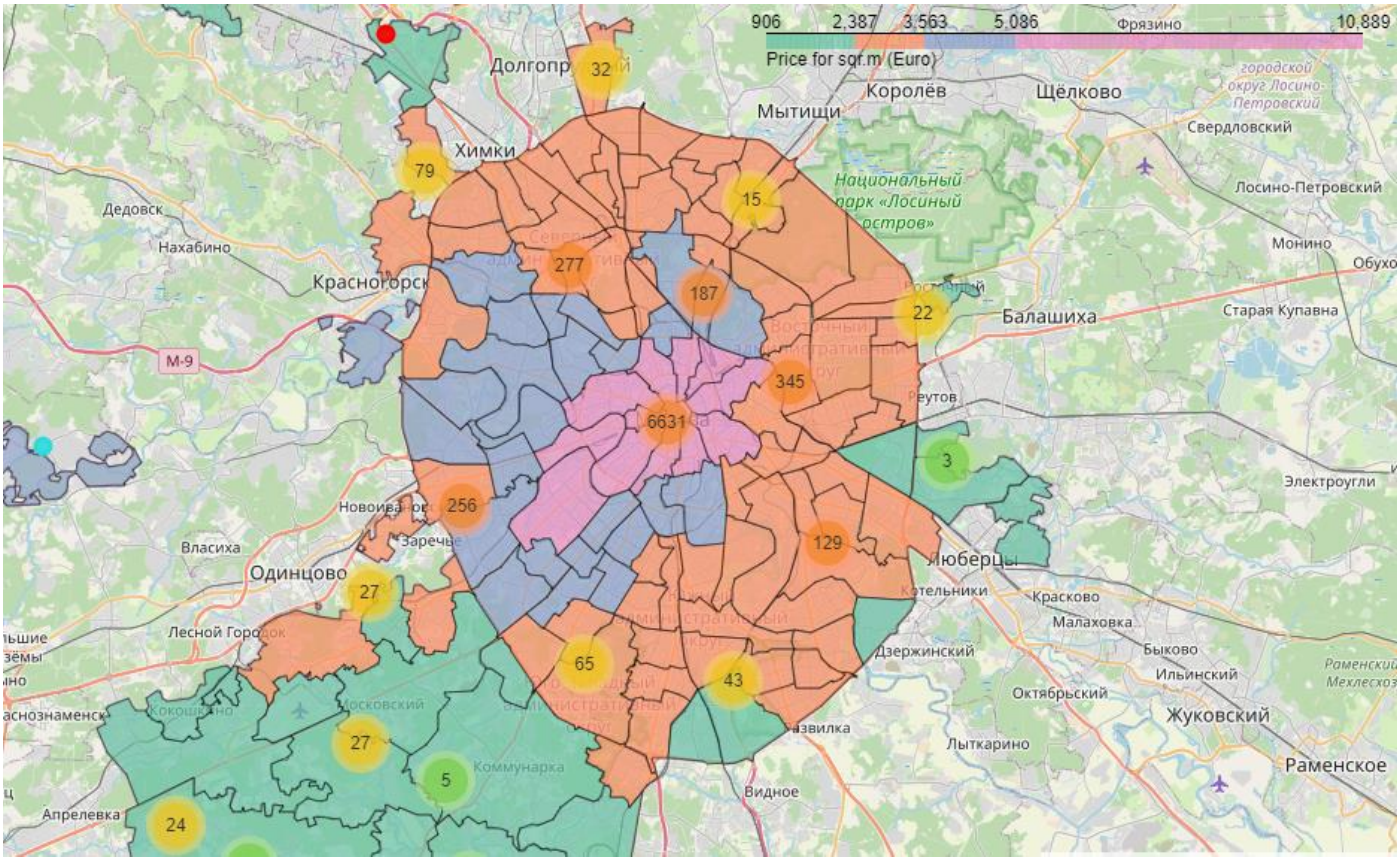
We are already familiar with the type of such statistics, a price imbalance is present, but not as large as in the case of the number of OKNs. We can even draw a graph without resorting to normalization. From the interesting:

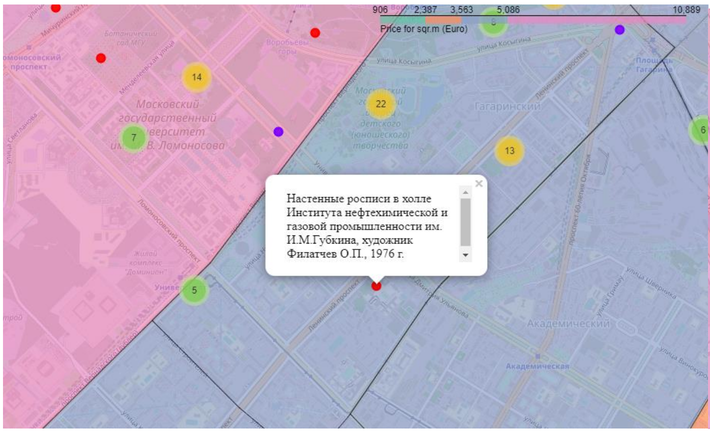
- The average price in Moscow is 3161 Euro per sq.m.
- If the cost of housing is more than 5100 per 1 sq. M - the price is considered an “emission” ie goes beyond statistics and you should have an understanding of what you pay for





A small cartographic visualization on the cost of apartments in Moscow and cultural monuments located on its territory:



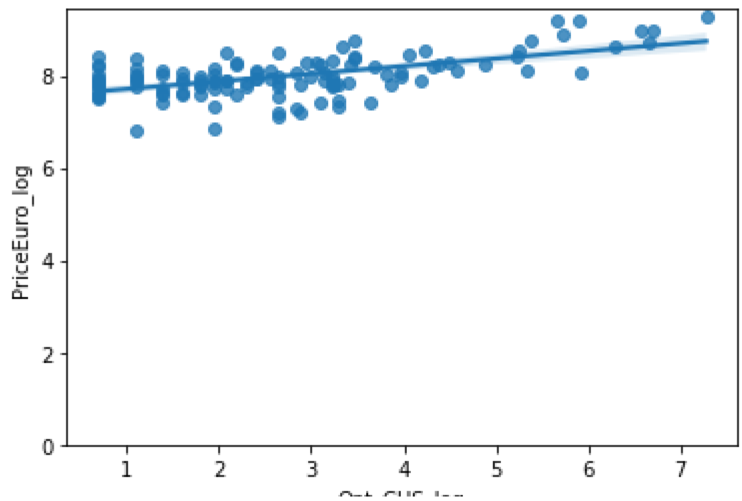


This visualization clearly shows the "wind rose" of the cost of housing in Moscow. He has his own character:

- The flow goes from NE to SW with a slight deviation to the west
- New Moscow - the first price level - "green" (low-cost housing)
- The second-level price offer is an “orange” color. These are the districts near the MKAD (C, NE and SE) and MKAD * and the Third Transport Ring * (B and SE) (city road rings)
- "Blue" areas the third level of prices - the outskirts of NW, SW and West. Central areas of the south and north
- "Pink" areas - the fourth, the most expensive areas in the city - these are just "emissions", the initial price level starts from 5086 euros per sq.m
- MKAD and the Third Ring Road - city road rings

Let's check our hypothesis that the level of the number of cultural objects influences, the way, and indirectly the level of housing prices.
Using logic, we understand that the number of monuments and the cost of apartments in the center are already precisely interconnected, but we will look at this using the correlation method.

```
(0.5876156714644017, 9.015164331686818e-13)
```



	Qnt_CHS_log	PriceEuro_log	Qnt_CHS	PriceEuro
Qnt_CHS_log	1.000000	0.587616	0.710403	0.658065
PriceEuro_log	0.587616	1.000000	0.581863	0.939840
Qnt_CHS	0.710403	0.581863	1.000000	0.729205
PriceEuro	0.658065	0.939840	0.729205	1.000000

As you can see, there is a dependence, but it is not very strong - this can be seen by the small angle of the line. And caused more valuable housing in the Central expensive areas.

Results

We are building the final table with the participation of all the data that we managed to collect.

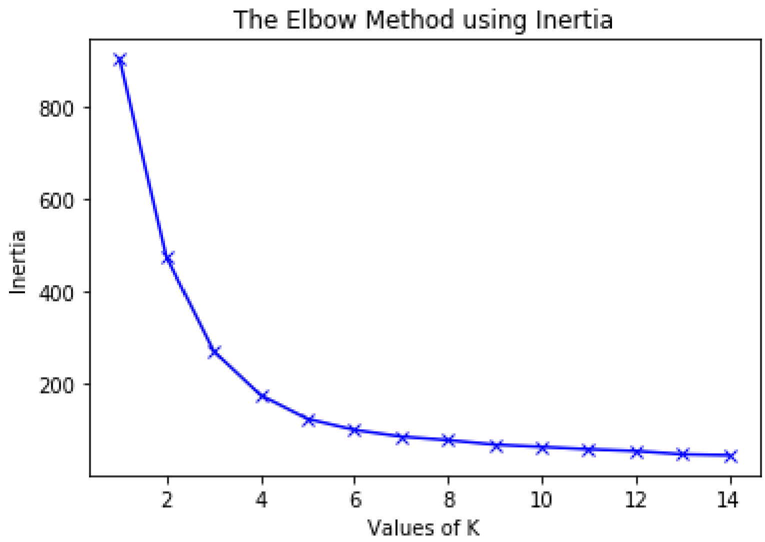
At the stage of receiving data from Foursquire, we “lost” another 29 districts - their OKN had no points of attraction. It's a pity!

The final data on Foursquare, the classification of districts and data on cultural objects are the last two in a normalized form.

	Qnt_CHS_log	Accessories Store	Adult Boutique	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Ent
0	3.637586	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
1	3.295837	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
2	3.295837	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.333333	0.0	0.0	0.0	
3	2.890372	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
4	3.091042	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	

5 rows × 424 columns

Again four clusters - magic :)



River	0.25
Well	0.25
Scenic Lookout	0.25
Park	0.25
Furniture / Home Store	0
Fruit & Vegetable Store	0
Fast Food Restaurant	0
Film Studio	0
Financial or Legal Service	0
Fish & Chips Shop	0

Example of district infrastructure

And here is what we got at the output by district by segment:

PriceEuro	2748.430953
lat	55.721604
lon	37.535446
Qnt_CHS	13.627907
PriceEuro_log	7.873999
Cluster Labels	2.697674
Qnt_CHS_log	2.208451
Park	0.089703
Lake	0.057648
Historic Site	0.055835
Café	0.033234
Bus Stop	0.029239
Auto Workshop	0.023453

CLUSTER 0

PriceEuro	5374.212669

```
Qnt_CHS              298.666667
lat                   55.764161
lon                   37.606928
PriceEuro_log          8.503359
Qnt_CHS_log            5.154189
Cluster Labels         0.166667
Coffee Shop            0.039939
Café                   0.039367
Bus Stop               0.030878
Park                   0.028857
History Museum         0.024266
Historic Site          0.023931
```

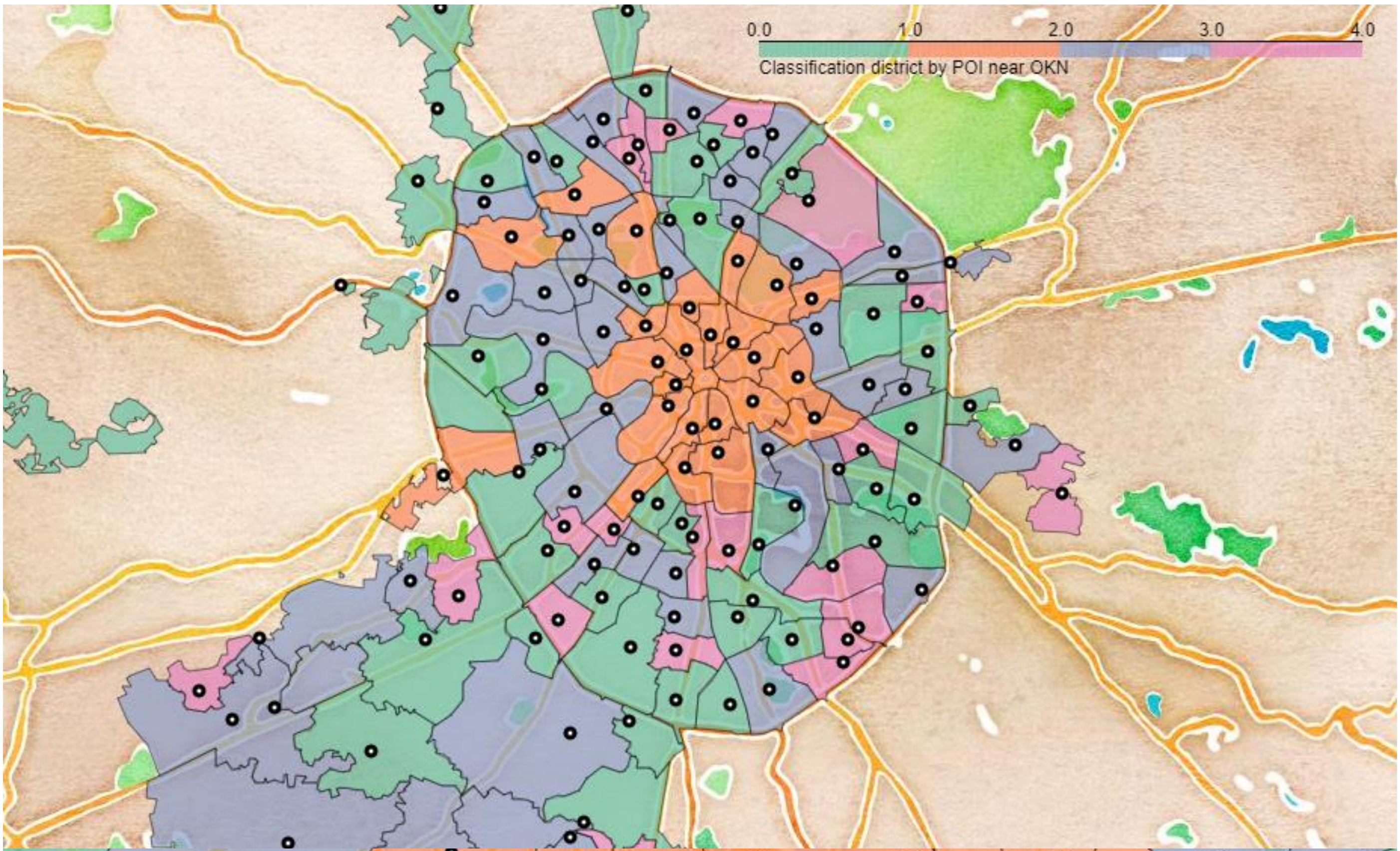
CLUSTER 1

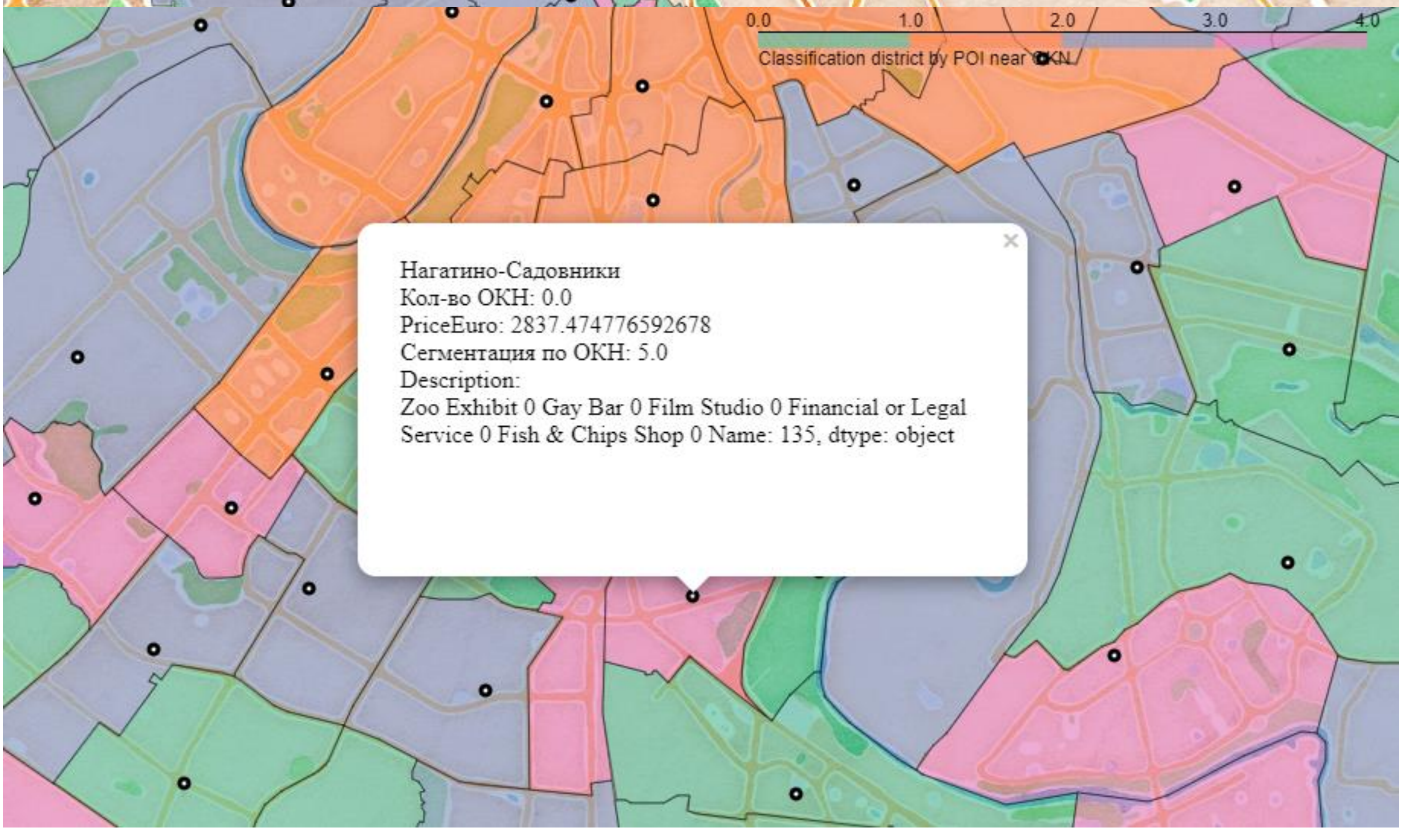
```
PriceEuro              2767.458459
lat                      55.708532
lon                      37.517561
Qnt_CHS                  10.035714
PriceEuro_log             7.859128
Qnt_CHS_log               2.002701
Cluster Labels            0.714286
Park                      0.038932
Café                      0.022150
Bus Stop                  0.022088
Supermarket               0.021995
Food & Drink Shop         0.016572
Cosmetics Shop            0.014759
```

CLUSTER 2

```
PriceEuro              2585.827892
lat                      55.703999
lon                      37.567995
PriceEuro_log             7.826725
Cluster Labels            5.000000
```

CLUSTER 3





Discussion

So the final fraternity of the districts looks exactly like this:

The center was reunited in the first cluster (orange) - a high price tag for housing, a large number of OKN, a variety of infrastructure surrounding ancient monuments.

The zero, green cluster is very similar to the second, purple, both in price and in the average number of OKNs in their territory. A characteristic difference of the zero cluster is a pronounced share in such POIs as a park, lake and historical sites, almost 20% of the infrastructure.

The third cluster is the areas without OKN and related infrastructure.

Are we satisfied with the result? And yes and no, probably the expectations would be in the discovery of the type of anomalous distribution of sentiments. The center would mix with the outskirts, but no. Everything is logical. An excellent continuation of the project - datasets with cultural themes - museums, theaters, libraries, etc. I think this will be the next step. And the battle of the neighbors will continue.

Conclusion

In this project, we examined one of the many cities in the world with the help of antiquities and the cost of housing. Any city, region or whole country can be subjected to such an analysis. Changing the incoming datasets both in quantity and in content, we can expand the input information and look at the object under study from different points of view. The segmentation method we have examined is not the only one, modern methods for studying data have more modern and powerful algorithms, for example, the t-SNE (t-distributed stochastic neighbor embedding) method or DBSCAN (Density-Based Spatial Clustering of Applications with Noise), which can produce completely different results other than what we got. POI data sources are also not limited by Foursquare capabilities - there is a world leader Google, there are local powerful data providers from Yandex or 2GIS in Russia, and in total all this means that the project is only at the beginning of the way.