**The restoration of the playground or the construction of a shopping center**

Lanevskiy Timur

May 11, 2020

**1 Introduction**

**1.1 Background**

In Shchelkovo, Moscow region, Russia, at 55.916471°N, 37.994031°E, a children's playground was closed [1], [2] and trees were cut down in preparation for the construction of a shopping center. The situation caused a public outcry and construction was stopped for a final decision.

**1.2 Problem**

The question being addressed in this analysis is: should a playground be restored at this location, or should the construction of a shopping center continue?

**1.3 Interest**

The stakeholder is the city administration, which manages the sustainable development of the city and decides on further construction. The city administration tries to avoid social discontent in society in the face of irritated mothers with children. On the other hand, it seeks to develop businesses that lead to the prosperity of the city, create new jobs and increase tax revenues to the budget, including increasing the welfare of townspeople’s and the prestige of the city administration.

**2 Data acquisition and cleaning**

**2.1 Data sources**

The following information is used to solve the problem:

* all-Russian statistics on the level of education, which shows the number of children in Russia aged 2 months to 7 years, and the number of kindergartens [3];
* all-Russian statistics on the birth rate, which shows the total number of Russians [4];
* the number of townspeople’s in the study, for example, Shchelkovo [5];
* normative document on the average number of children in kindergarten [6];
* Python library for determining the geographical coordinates of the city under study [7];
* Foursquare to search for research objects and determine geographical coordinates, for example, kindergartens in the city of Shchelkovo [8];
* Yandex for in-depth search of research objects and determining geographical coordinates, for example, kindergartens in the city of Shchelkovo [9];
* Geojson.io for plotting on a map the boundaries of the city in the format GeoJSON [10].

**2.2 Data cleaning**

Foursquare search queries are reinforced by Yandex, which is more confident in finding the objects under study in Russia.

Research objects located outside the city limits were removed from the data frame.

Below is the result of a search query, represented as 10 lines of a data frame, for example, kindergartens in Shchelkovo.



Responses to search queries in the "Name" and "Address" columns are accepted in Cyrillic and are not displayed on folium maps. They are saved in the data frame for reference.

Responses to search queries in the "latitude" and "longitude" columns are displayed on folium maps.

The data frame stores data from a single class, such as a kindergarten. In the economic sense, the term kindergarten is understood as a municipal (budget-funded) institution where parents can leave their child under the supervision of a tutor from 7 am to 7 PM while working.

The search result represents a broader class, where there are both municipal kindergartens with a population of 150 to 350 children, and private kindergartens with a smaller number. In addition, there are other categories of education related to children, such as "children's clubs", among the objects studied. This may lead to a significant difference from the normative document [6].

In an economic sense, a playground is a free outdoor playground with benches for parents.

In contrast, the search result returns a smaller number of mostly paid playgrounds that better fit the definition of "amusement park". The ratio of the number of children to the number of playgrounds is not normalized, but it turns out to be excessively high.

A shopping center is a multi-store, several-story building of an individual project.

**3 Methodologies**

The methodology for data analysis consists of performing sequential steps.

**3.1 Analytical processing of city data frames**

Analytical processing of data frames of search query results for all cities under study. The resulting date frame is shown below.

The "Name" column shows the cities under study in the vicinity of Shchelkovo. It is assumed that their development is similar due to geographical proximity and orientation to the socio-economic policy of the center - Moscow.



Data from neighboring cities will help you visually represent the objects under study and perhaps train the model at the next stage to determine the class of the object under study and answer the question: should we restore the playground on this site, or should we continue building a shopping center?

The information in the column "Children" about the number of children living in the city is obtained by calculation based on the average Russian birth rate of children per 1000 people multiplied by the number of townspeople. The averaging assumption makes an error in estimating the number of children per kindergarten and the number of children per playground.

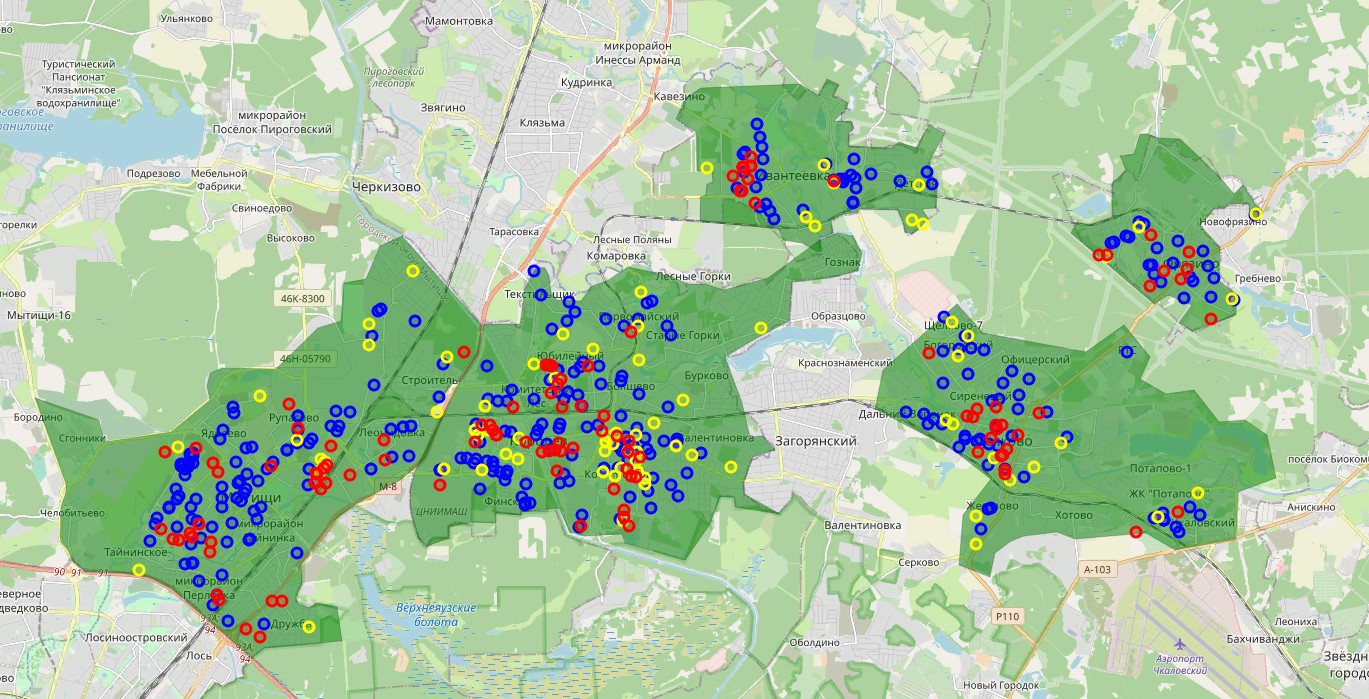
The assumptions made for data in the "Kindergarten", "Playground", and "Shopping center" columns are described in section 2.2 data cleaning.

The data in the "Children per kindergarten", "Children per playground", and "Townspeople per shopping center" columns is calculated by dividing the number of children or townspeople who also include children by the number of objects under study.

The data frame shows that the number of children per playground is too large and is related to the small number of playgrounds found in the search results. Therefore, kindergartens were added to the database for the cities studied to better reflect urban dynamics.

**3.2 Visualization of found objects on the city map**

For each city under study, maps were obtained with the location of separate kindergartens, playgrounds, shopping centers, and all the marked objects. We also received a map with all the cities and all the objects on the map, shown below, where kindergartens are highlighted in blue, shopping centers in red, playgrounds in yellow, and the city borders are highlighted in green.



Visualization will help you see the distribution trend of the studied objects, for example, their concentration or vice versa emptiness.

In several cases, a discrepancy between the address and location of the object was found, including the location on the roadway, inside a residential building, or on the river. We recommend that you check the coordinates of the location of the object under study in reverse order at the address found.

**3.3 Data analysis using machine learning**

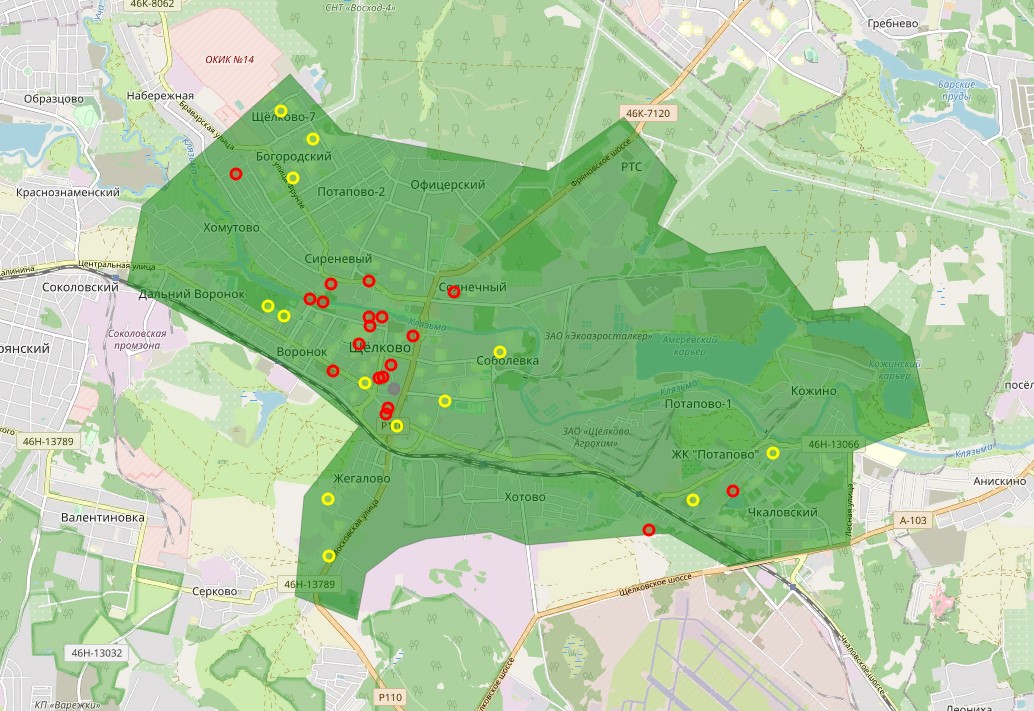
Based on the data obtained for the cities of Korolyov, Mytishchi, Fryazino and Ivanteevka, a test data frame was prepared to train the model and apply it to the data of the city of Shchelkovo to answer the question.

The task relates to supervised learning and consists in classifying the object under study using the K-Nearest Neighbors method.

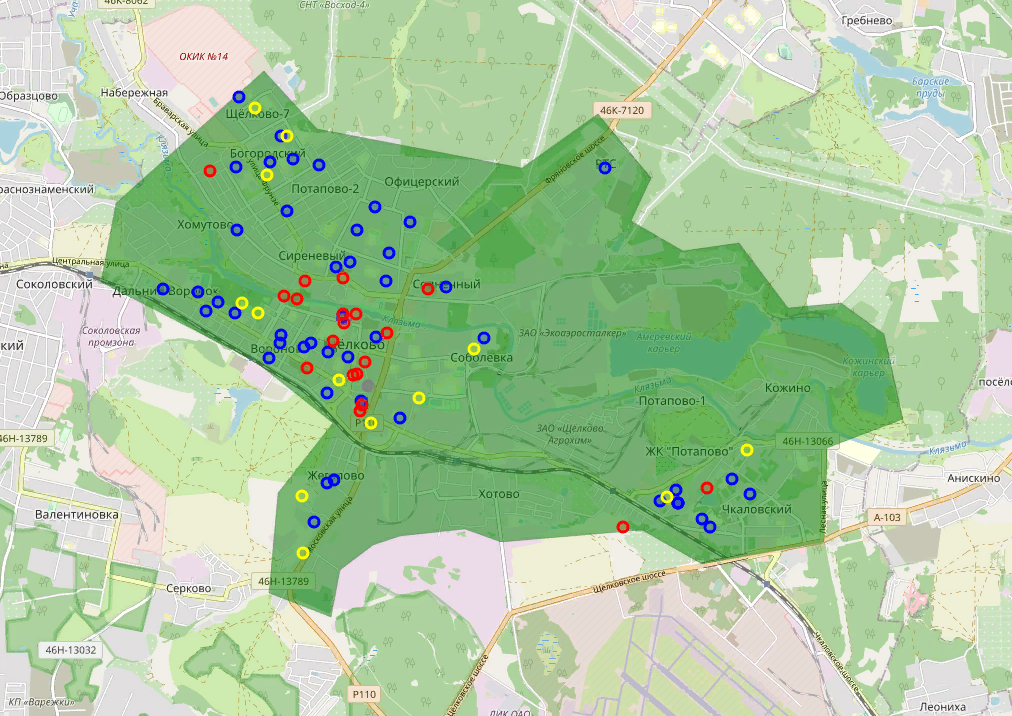
Additional calculations are performed using Decision Tree, Support Vector Machine, and Logistic Regression.

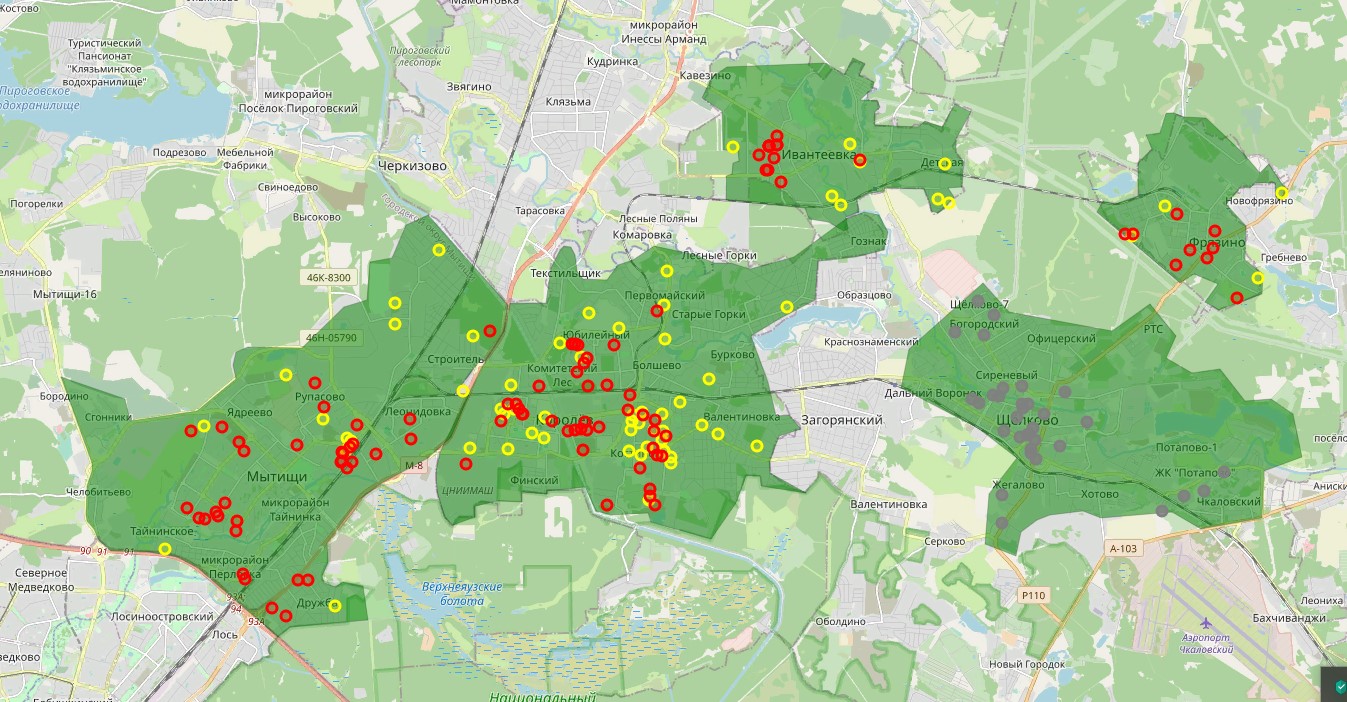
Several calculated cases are shown below.

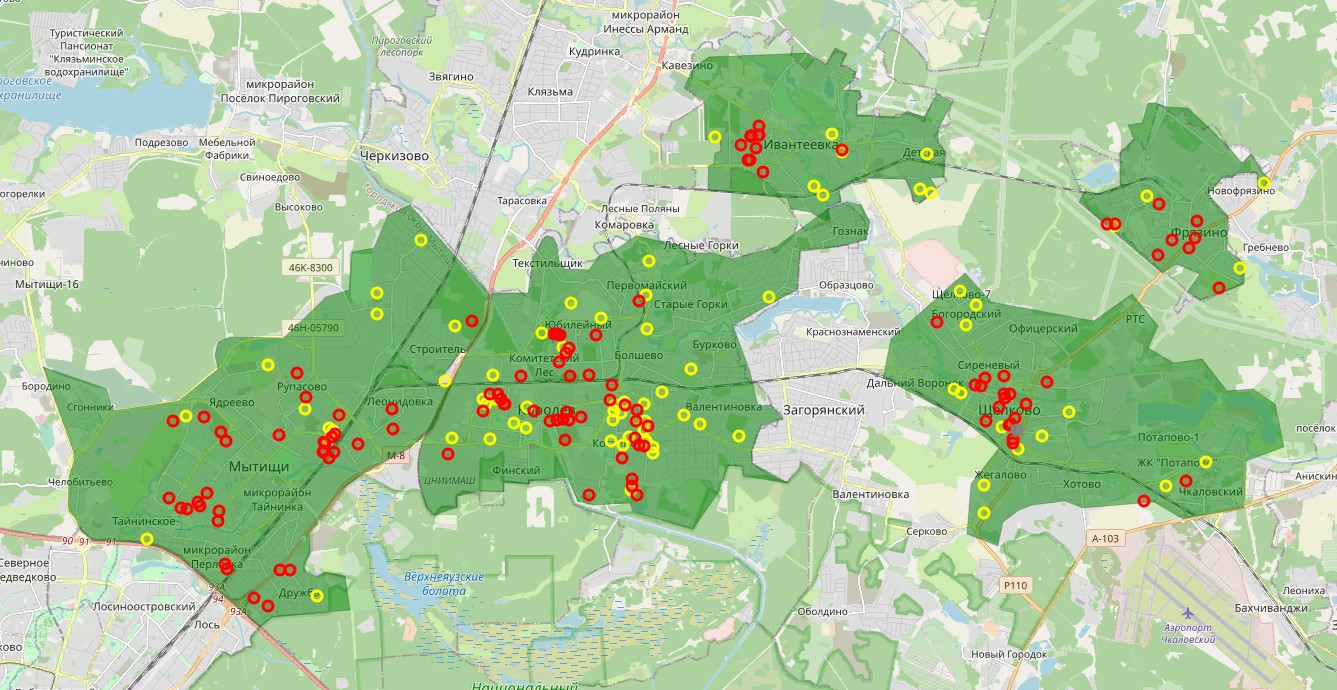
1. Training data consists of data about playgrounds and shopping centers in Shchelkovo. The test data includes the research object marked with a gray marker.



1. Training data consists of data on kindergartens, playgrounds and shopping centers in Shchelkovo. The test data includes the research object marked with a gray marker.



1. Training data consists of data on playgrounds and shopping centers in the cities of Korolyov, Mytishchi, Fryazino, and Ivanteevka. Test data includes playgrounds and shopping centers in Shchelkovo, marked with a gray marker.
2. Training data consists of data on playgrounds and shopping centers in all cities. The test data includes the research object marked with a gray marker.



**4 Results**

The calculation results for various calculation cases are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Jaccard** | **F1-score** | **Class** |
| KNN | 1.00 | 1.00 | shopping center |
| Decision Tree | 0.00 | 0.00 | playground |
| SVM | 1.00 | 1.00 | shopping center |
| Logistic Regression | 1.00 | 1.00 | shopping center |

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Jaccard** | **F1-score** | **Class** |
| KNN | 1.00 | 1.00 | shopping center |
| Decision Tree | 0.00 | 0.00 | kindergarten |
| SVM | 0.00 | 0.00 | kindergarten |
| Logistic Regression | 0.00 | 0.00 | kindergarten |

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Jaccard** | **F1-score** | **Class** |
| KNN | 0.52 | 0.52 | - |
| Decision Tree | 0.42 | 0.42 | - |
| SVM | 0.45 | 0.46 | - |
| Logistic Regression | 0.61 | 0.50 | - |

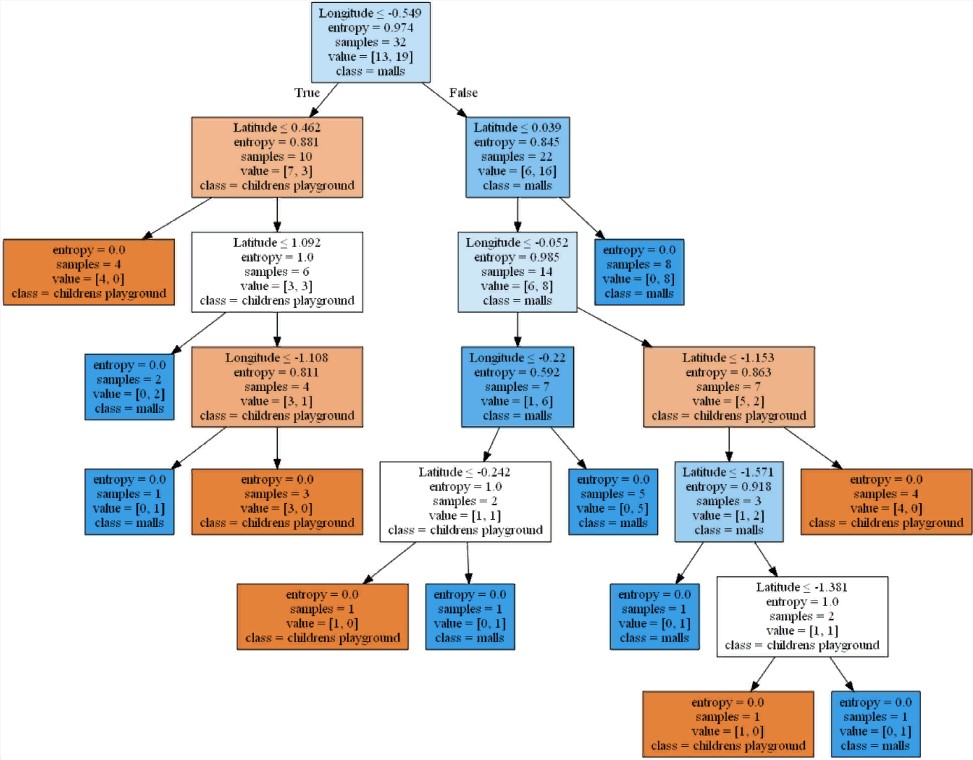
|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Jaccard** | **F1-score** | **Class** |
| KNN | 0.00 | 0.00 | playground |
| Decision Tree | 0.00 | 0.00 | playground |
| SVM | 1.00 | 1.00 | shopping center |
| Logistic Regression | 1.00 | 1.00 | shopping center |

**5 Discussions**

Based on the analytical processing of the frame data, it can be seen that the number of kindergartens, playgrounds and shopping centers in Shchelkovo is not inferior to the average value for neighboring cities. Korolyov has a more developed urban infrastructure. The answer to the question about further action is not clear.

A visual study of the location of objects on the territory of cities shows that kindergartens are located evenly, there are few playgrounds, they are mostly scattered, and shopping centers are concentrated in the regions or rarely located separately on the outskirts of the city. An urban area can be divided by a large object, such as an industrial development.

Upon closer examination, the Shchelkovo (a) basic KNN model and most auxiliary models recommend the construction of a shopping center. It is not clear why Decision Tree does not follow this recommendation. The Decision Tree is shown below.

****

When adding data about kindergartens to the data frame Shchelkovo (b) auxiliary models recommend the construction of kindergartens. It is not possible to consider models "out of the box" you need to know what is "located under the hood".

In the calculated case (c), models based on data from neighboring cities have a 50% probability of determining the class of objects under study. In other words, using a classification method based on data from neighboring cities is not very applicable within the boundaries of the city under study.

In the calculated case (d) with the maximum amount of data, the basic KNN model assumes the restoration of a playground, in contrast to the previous recommendation (a) for the construction of a shopping center. It is likely that an increase in the volume of data from other cities "noise" the model.

**6 Conclusions**

As a result of the research, it is recommended to continue construction of the shopping center at 55.916471°N, 37.994031°E.

The recommendation is based on the collected and processed search data on the number and location of kindergartens, playgrounds, and shopping centers in the cities of Shchelkovo, Korolyov, Mytishchi, Fryazino, and Ivanteevka.

By the number of children per kindergarten and playground and the number of townspeople per shopping center Shchelkovo is not inferior to the neighboring cities.

Playgrounds found in search queries are scattered and are mostly paid "amusement parks". Shopping centers are located mostly compactly.

Calculation models based on supervised learning classify the object under study as a shopping center.

We recommend that you check the coordinates of the location of the object under study in reverse order at the address found.

**7 References**

1. <https://ru.foursquare.com/explore?mode=url&near=Щёлково%2C%20Россия&nearGeoId=72057594038423280&q=детская%20площадка>
2. <https://yandex.ru/maps/10765/shelkovo/search/55.916471°N%2037.994031°E/?ll=37.993265%2C55.916741&sll=37.994031%2C55.916471&source=wizgeo&utm_medium=maps-desktop&utm_source=serp&z=18>
3. <https://rosinfostat.ru/uroven-obrazovaniya/>
4. <https://rosinfostat.ru/rozhdaemost>
5. <https://ru.wikipedia.org/wiki/Щёлково>
6. <http://mtsk.mos.ru/Handlers/Files.ashx/Download?ID=11293>
7. <https://geopy.readthedocs.io/en/stable/>
8. <https://ru.foursquare.com/explore?mode=url&near=Щёлково%2C%20Россия&nearGeoId=72057594038423280&q=детский%20сад>
9. <https://yandex.ru/maps/10765/shelkovo/search/детский%20сад/?clid=2270456&ll=38.005283%2C55.924296&sctx=ZAAAAAgBEAAaKAoSCT%2F%2FPXjt%2FkJAER1aZDvf9UtAEhIJS2z9%2F3%2BNzz8RFe5RP6nTtD8iBQABAgQFKAowADiPptOJ5NiwkoQBQI1USAFVzczMPlgAagJydXAAnQHNzEw9oAEAqAEAvQG0mY5xwgGAAZ%2BLi8AElamf640CucDJvgSl35a%2FBvS9qPauA6yk548E%2F6z73wPJgLL0A%2BbZveIDgevDjwSIva3sA%2Bj9r7kE6Y2C08gB8Ij6kQTF1%2BD6A9PHo48Gouu0nASrpJmdBIDowt0D143inATY6ci2BInygOYE0K%2FerwTym7T1BeCPxscE&sll=38.005283%2C55.924296&sspn=0.141580%2C0.046722&z=13.8hgjl>

1. <http://geojson.io/>