

Capstone Project - The Battle of the Neighborhoods

Sidenko Stanislav Aleksandrovich

20.01.2021

Table of contents:

1. Introduction
 - 1.1 Background
 - 1.2 Problem
 - 1.3 Interest
2. Data
 - 2.1 Data sources
 - 2.2 Data acquisition and cleaning
 - 2.3 Feature selection
3. Methodology
4. Exploratory Analysis
5. Results and Discussion
6. Conclusion

Introduction

1.1. Background

New entertainment venues, such as nightclubs, are constantly opening up in the world. This is a great place to relax and have fun. In them you can not only relax, but also make interesting acquaintances. Such establishments bring a good profit. There are always great opportunities in the implementation of any idea for the design and selection of themes for a nightclub.

1.2. Problem

There is a lot of competition in the market. With all the benefits, many establishments are closed, making one of the main mistakes at the start-choosing the wrong location. In this project, we will choose the optimal location of a nightclub in Moscow.

1.3. Interest

This project will be interesting for entrepreneurs who have decided to open a nightclub in Moscow. It will also be of interest to landlords, as you can find customers for a nightclub and, reasonably, set the price and build a long-term relationship with the tenant.

2. Data

2.1 Data source

We will determine the factors that influence the successful choice of a location for a nightclub:

- Number of Nightclubs located right nearby
- Distance to the center
- Rating of clubs in the neighborhood

Data sources for the project:

1. To determine the coordinates, use the geoby library;
2. All night clubs, their locations and ratings are obtained using the Foursquare API;
3. Using the Pandas library, we convert the json information into a tabular form;
4. Distances by coordinates will be determined using the haversine library.

The search for a place is limited to the third transport ring, which is approximately 7 km in radius.

2.2. Data acquisition and cleaning

To begin with, I determined the coordinates of the center of Moscow, for which Red Square was taken.

Next, it was necessary to get all the night clubs within a radius of 7 km from Red Square. To do this, use the Foursquare API. There was also a problem with limiting the number of night clubs received per request from the Foursquare API.

To solve this problem, you need to divide the territory in question into several zones in order to get all the night clubs located on it. I put 4 points equidistant from the center and will search by radius, intersecting each other. Immediately there is a second problem of calculating the distance by latitude and longitude.

Since the distances are small, we apply a simplification: for latitude-1 degree = 111.1 km, 1 degree of longitude is approximately 64 km. This allowed us to get rid of long formulas for calculating the distance by coordinates.

To overlap the zones we received, we take the search radius 500 meters more (from 7 km of the zone we selected). On the Foursquare API site, we find the nightclub category to search only for the nightclubs we need. From the Foursquare API, we get information in Json format. For convenience, we use the Pandas library to convert the received data to a tabular format. Also, since the sample zones overlap, we remove duplicate clubs and combine the resulting information into one dataset.

2.3. Feature selection

The result is a table with 97 nightclubs and 19 characteristics.

The first step is to select from the information received only those characteristics that will help in the analysis. We will also change the names of the characteristics, making them more understandable.

Discard all unnecessary data, leaving the main thing. The resulting night clubs will be displayed on the map (Fig. 1).

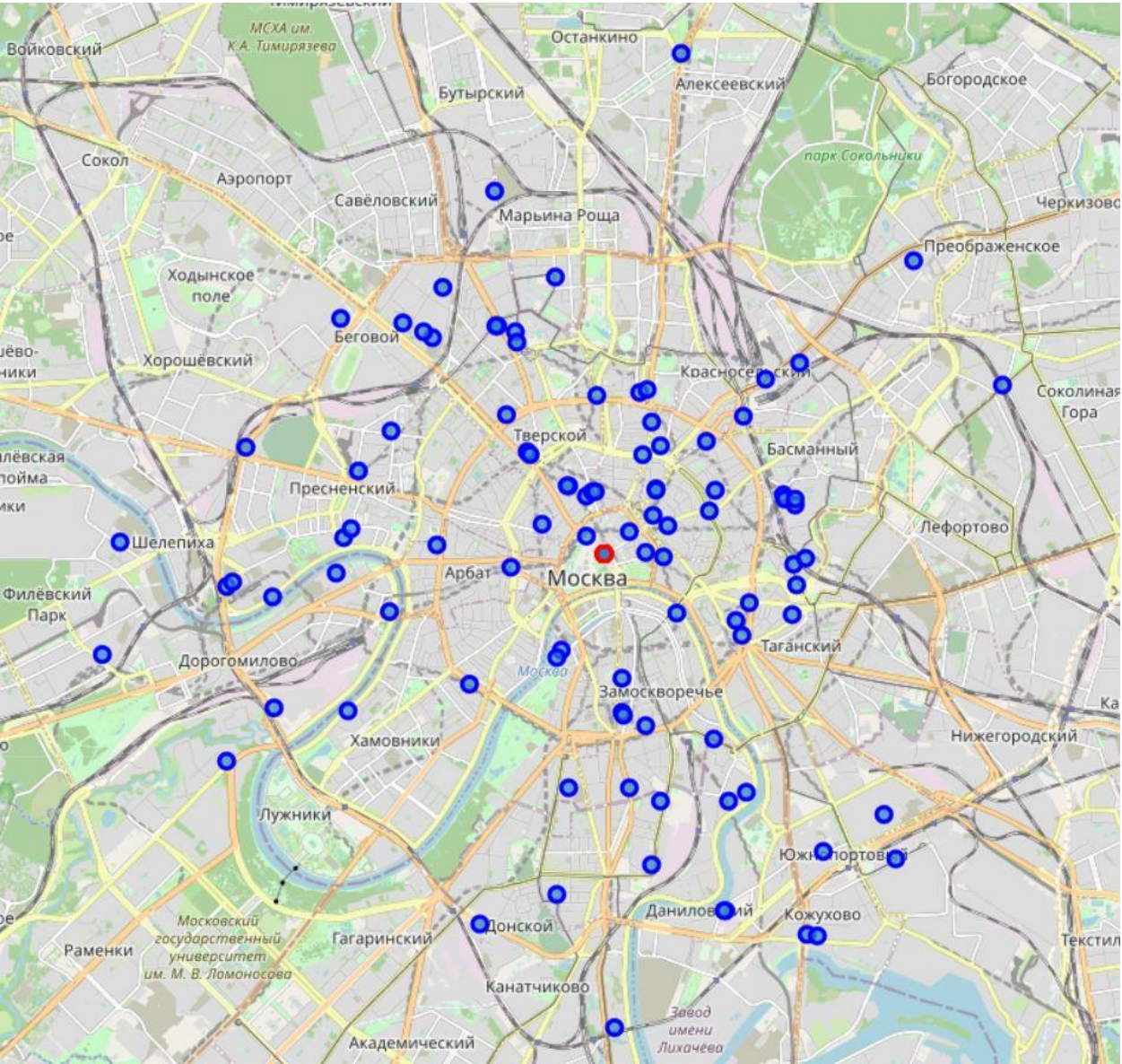


Fig. 1 All night clubs in Moscow.

You will need the distance data from each club to the center of Moscow (Red Square), for this we will use the haversine library. The resulting distances are added to the table.

3. Methodology

In this project, I focused my efforts on finding and defining the territory for the best opening of a nightclub.

First, I received information about the location of all existing nightclubs within the third transport ring in Moscow. I also determined the distance of each night club to the center of Moscow, for which we took the Red Square and got its coordinates at the beginning. For a more complete picture of the night clubs, club ratings were obtained using the Foursquare API.

Secondly, I will create clusters based on the data obtained, for a more accurate determination, different properties will be used, as well as two clustering tools: Kmeans and DBSCAN.

Thirdly, I will test the hypothesis that the closer the distance to the center of Moscow, the higher the rating of nightclubs.

4. Exploratory Analysis

To begin with, let's take the most common clustering method Kmeans, it is quite fast and will allow you to choose the most suitable hyperparameters for it.

Since we can assume the special case that each nightclub will belong to a separate cluster, we will indicate this assumption when searching.

Data on the location of night clubs on the map of Moscow are specified in coordinates, for the best clustering, we will translate all coordinates into radians using the Numpy library.

To select the optimal parameters for the Kmeans clustering method, we use GridSearchCV from the Sklearn library.

Applying the best found parameters, we obtained 75 clusters (Fig.2)

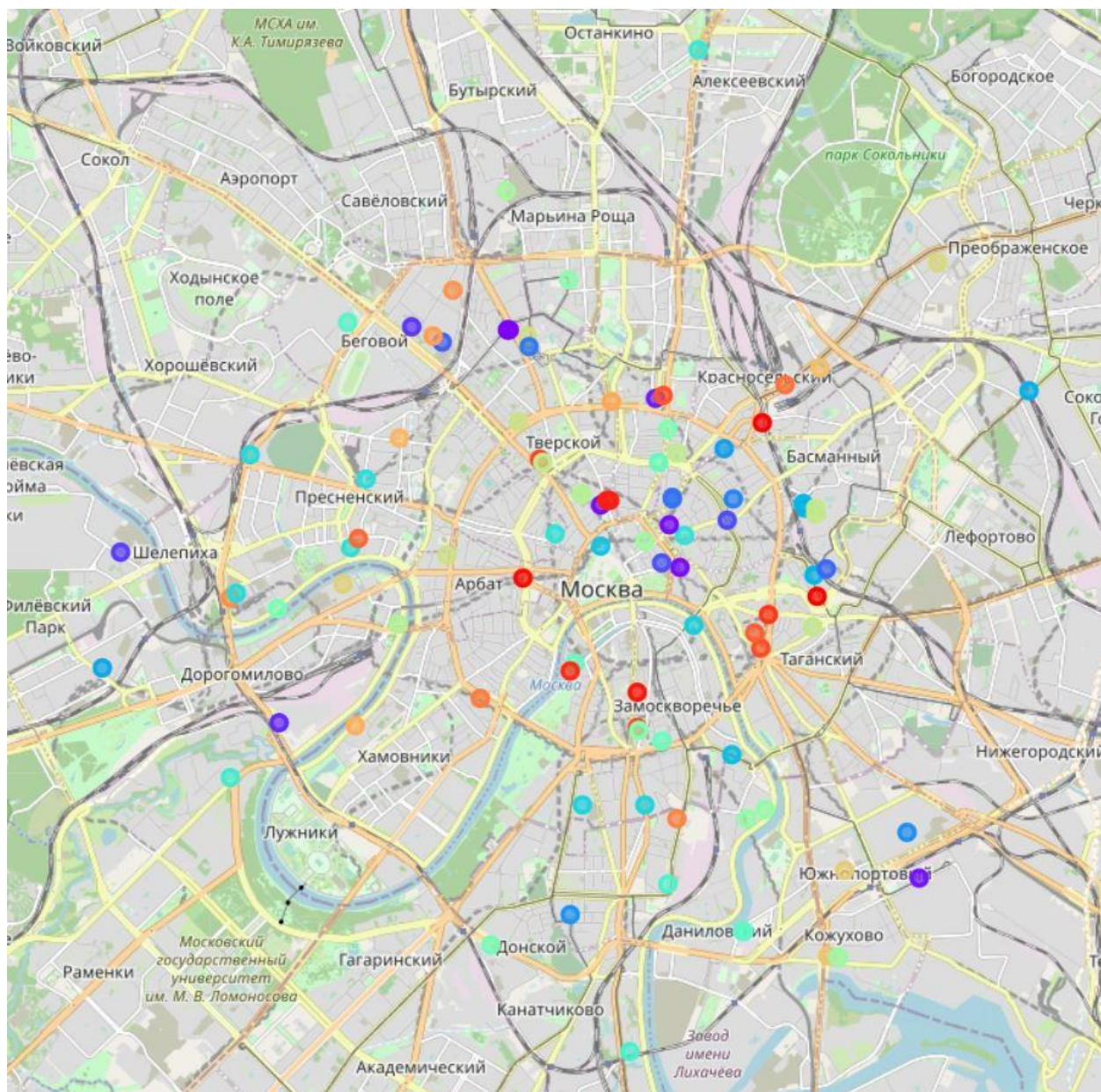


Fig. 2 Division into clusters by the Kmeans method.

When studying the resulting partition, unfortunately, no useful information was revealed. For this reason, the DBSCAN method was re-clustered (Figure 3).

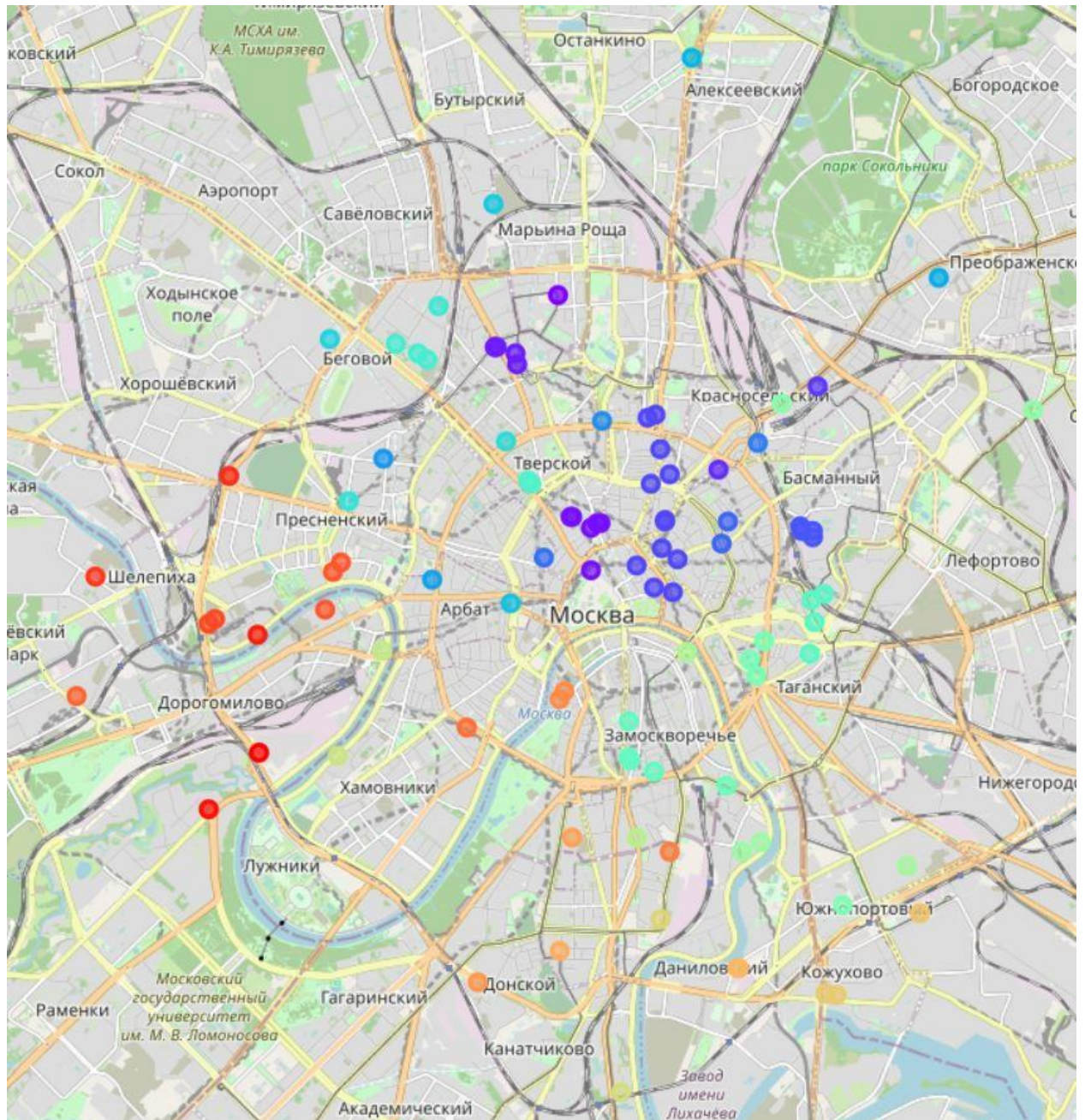


Figure 3 Clustering by the DBSCAN method.

The result is a distinct group of nightclubs suitable for further analysis. The obtained 55 groups of night clubs were included in the table.

As you can see from the resulting map, most nightclubs are separated by quite large distances. Only small groups that are closer to the city center are clearly visible.

For more detailed information about these clusters, a grouping was carried out and all clusters of nightclubs with a number equal to or higher than 4 were selected.

These groups were plotted on a separate map (Fig.4)

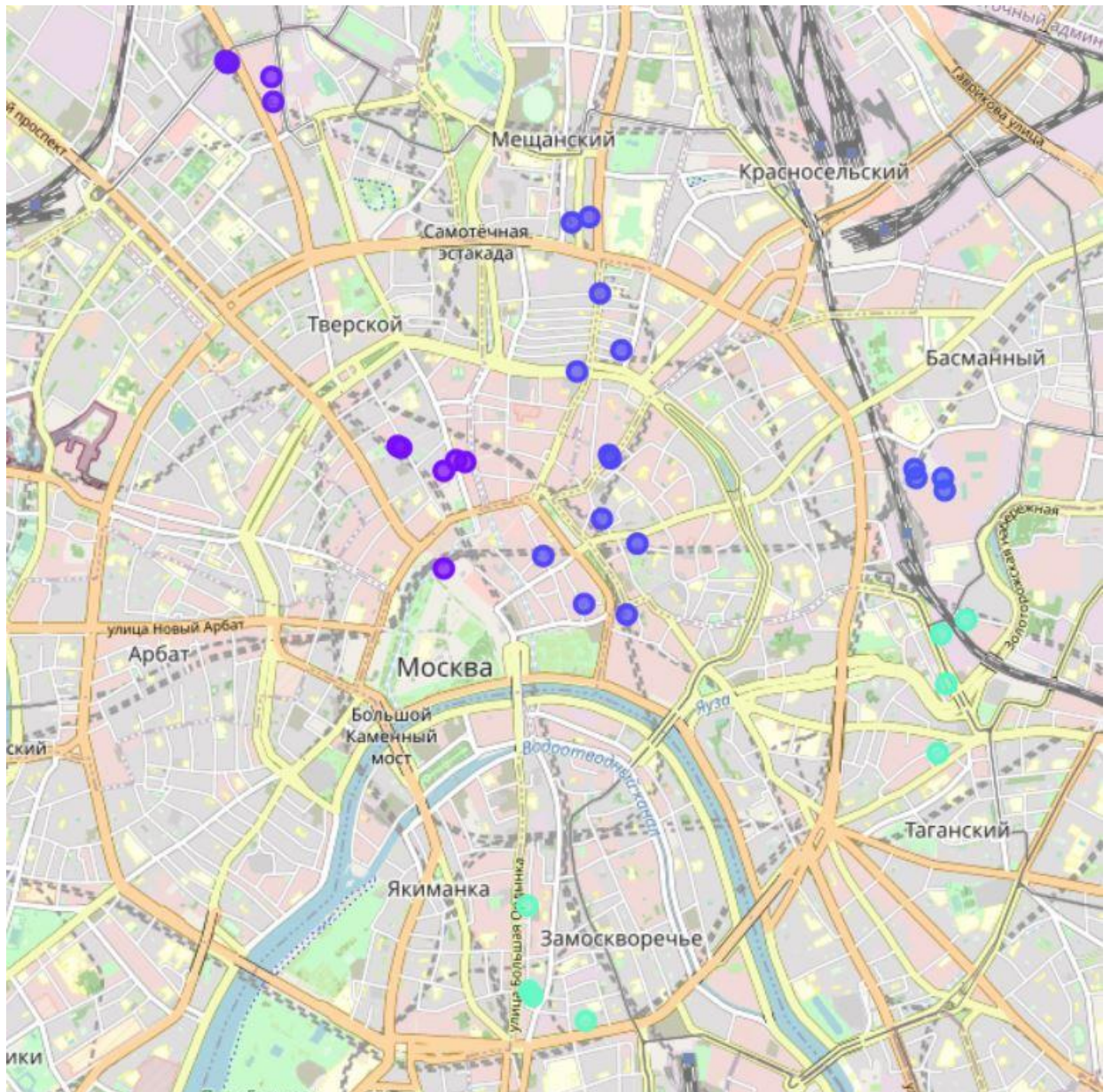


Fig. 4 Distinct groups from 4 clubs.

On the resulting map, the main cluster of nightclubs, when viewed in detail, can be linked to:

- the largest number of clubs open within the garden ring;
- all clubs are located near theaters, museums, and business parks.

Next, we will test the hypothesis about the relationship between distance from the city center and the rating of the club.

To do this, select all the clubs that have a rating higher than 7.0. These clubs will be displayed on the map (Fig. 5).

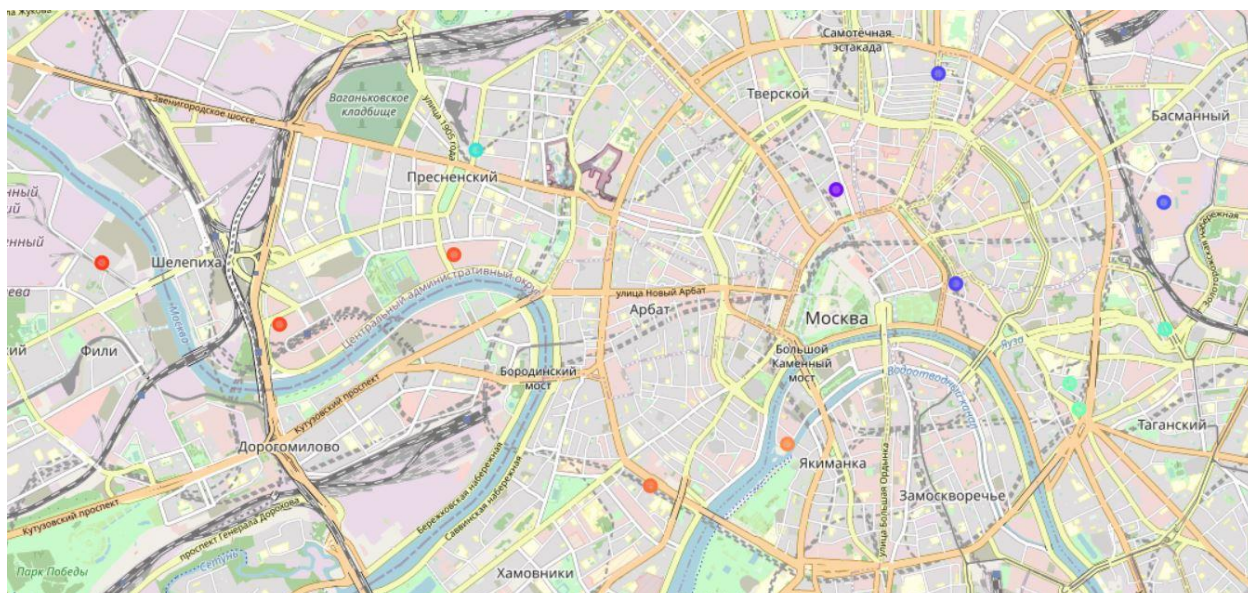


Fig. 5 Clubs with a rating of 7.0 and higher.

Let's calculate the ratio of the total number of clubs to clubs with a rating above 7.0. It turned out that only 14.5% of all clubs have a high rating. In our data, there are a large number of clubs with a rating of 0, these clubs can be considered new. The number of these clubs to all is about 54.5%.

We will display the ratio of the distance to the center and the rating of the clubs on the dot diagram (Fig.6)

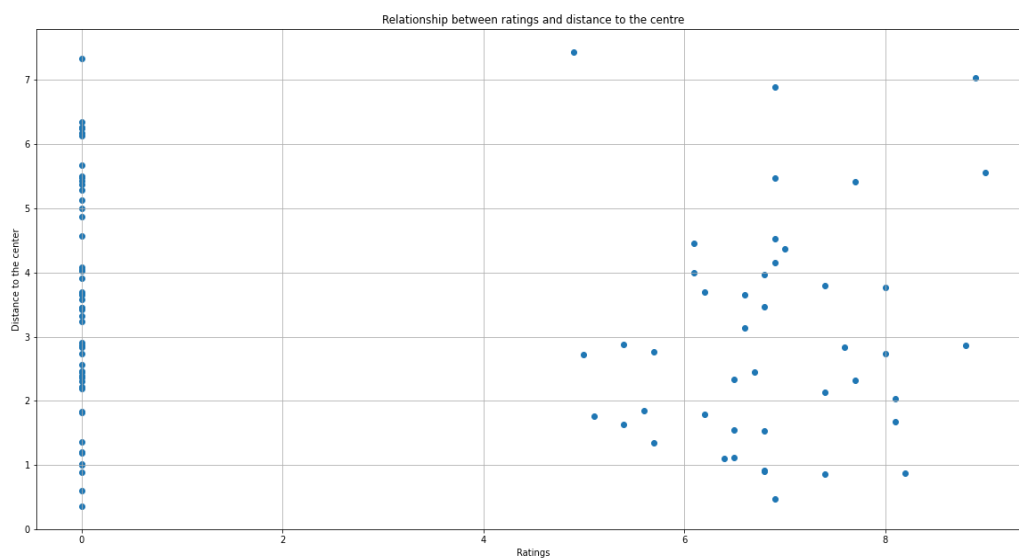


Fig. 6 The ratio of the rating of the club to the distance from the center.

After reviewing the map and the dot chart, we can conclude that the hypothesis about the ratio of the rating to the distance from the center was not confirmed.

5. Result and discussion

In this study, we looked at the night clubs in Moscow to identify the best areas for opening new night clubs.

As can be seen from the study, approximately 14.5% of clubs are competitive. These are the clubs that maintain their rating above 7.0. There is no clear distribution of institutions with a high rating, they are scattered throughout the study area, which gives room for a competitive environment and this gives ample opportunities for opening new ones, which is just observed in the form of 54.5%. A very large number of new nightclubs makes it possible to promote in this area, using professional service and customer orientation, this is confirmed by a very small percentage of clubs with a high rating.

The study clearly shows a trend in the location of clubs close to tourist areas and theaters/museums, as well as business parks are very popular.

In connection with all the identified trends and the density of nightclubs in Moscow, three main recommended zones can be identified. They are located close to tourist areas, as well as there are no clusters of nightclubs (green zones in Figure

In addition, 2 zones are identified, marked in red, in which it is not recommended to open establishments. Despite the fact that they meet the conditions, in these areas there are large concentrations of religious and state institutions, in addition to tourist ones.

A very large spread of night clubs around the city, gives the opportunity to a larger selection with minimal competition.

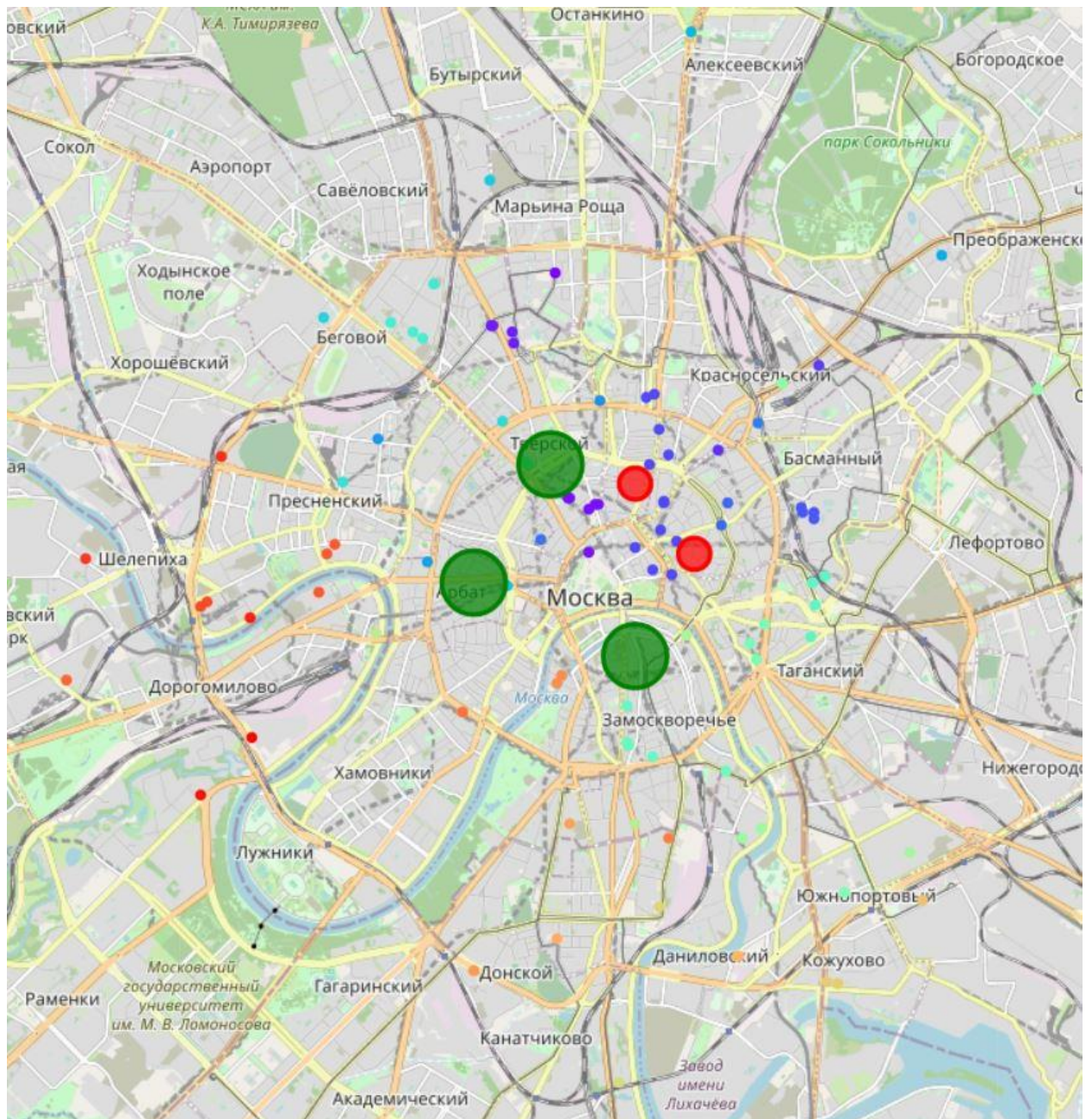


Fig 7. Best and worst areas for opening night clubs.

6. Conclusion

The goal of this project was to identify favorable areas for the opening of nightclubs. For entrepreneurs, this is a good help to their business plan, and for landlords, it is an opportunity to interest tenants. Several maps were created to display the results. The theory about the connection between the distance to the center and the rating of clubs was refuted. The very large number of new nightclubs and their disparity across the map may be the cause of the recent epidemic and an opportunity for new clubs to open when the measures associated with it are relaxed.