Capstone Project - The Battle of the Neighborhoods

Sidenko Stanislav Aleksandrovich

20.01.2021

Table of contents:

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).

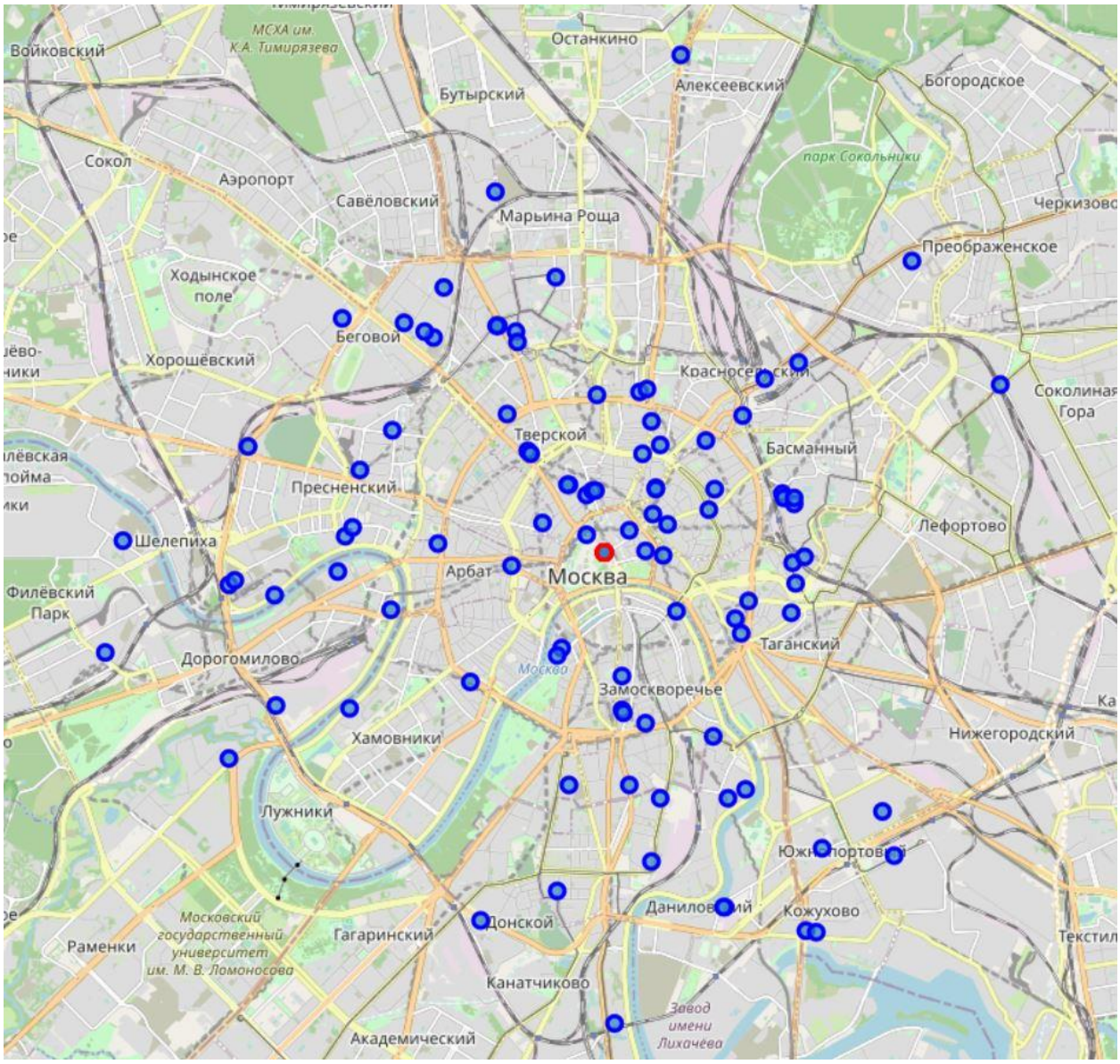

Fig. 1 All night clubs in Moscow.

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.