Capstone Project - The Battle of Neighborhoods

Sergey Bogdanov
27 May 2021

1. Introduction

Moscow city have aproximate 2500 square km area and nearly 13 million population. For people who want to start new business, in our example a coffee shop, it is quite hard to find a "good place". "Good place" is a broad term, it is including:

- flow of people
- location of other coffee shops around
- etc

For business which mostly work with people, finding a place with big flow of people and many other positive factors is a half a success.

2. Data

Based on definition of our problem, factors that will influence our decision are:

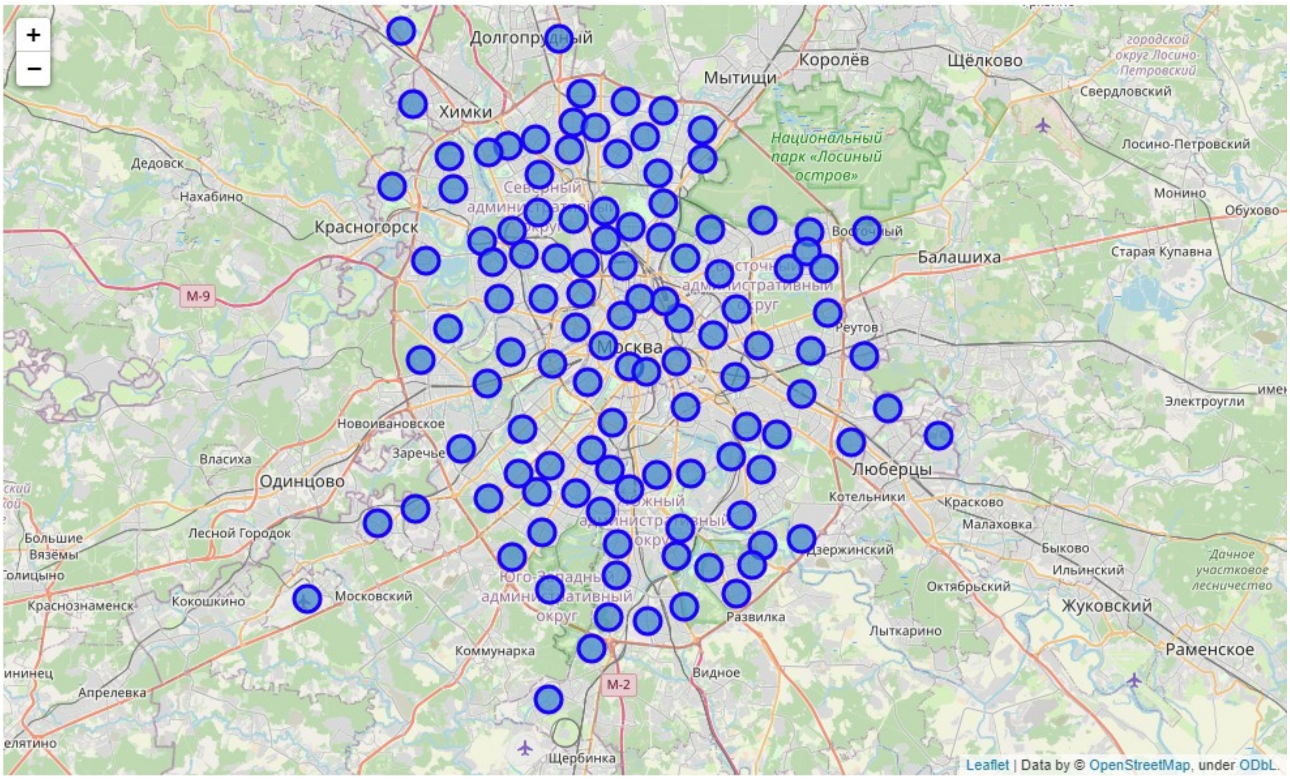
- number of existing food services neighborhood
- number of and distance to coffee shop in the neighborhood, if any
- distance of neighborhood from most popular district center

We decided to use regularly spaced grid of locations, centered around most popular center, to define our neighborhoods.
Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of food services and their type and location in every neighborhood will be obtained using **Foursquare API**

2.1 Finding most popular district

First of all we need to find most popular district for that we scrap wikipedia page for all districts and put them on map for better visualization. We decided to remove far away districts like city of Zelenograd, they are not our target but they are part of Moscow infrastructure.



As we can see there are a lot of possible popular start points so we fetch data from Foursquare API in every district center with radius of 500 meters and then sort it by count of venues.

District	Venue Count
Arbat District	100+
Tverskoy District	86
Zamoskvorechye District	85
Presnensky District	72
Chertanovo Tsentralnoye District	55

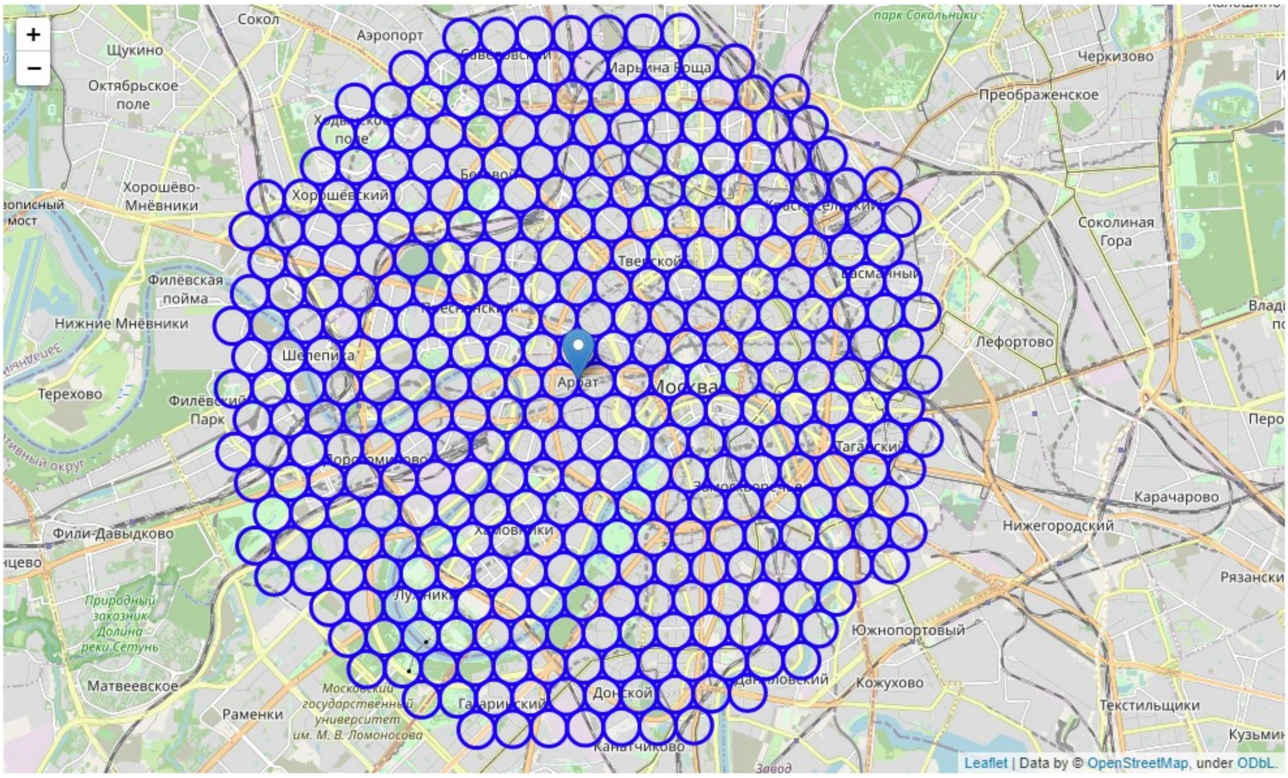
The most popular district is **Arbat District**.

2.1 Neighborhood Candidates

As second step we created grid of cells which covered area of interest which is aprox. 12x12 killometers centered around Arbat District center. The grid of area candidates, equaly spaced, centered around city center and within ~6km from Arbat District center. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Let's create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**.



OK, we now have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors is exactly the same) and within ~6km from Arbat District center.

Let's now use Google Maps API to get approximate addresses of those locations.

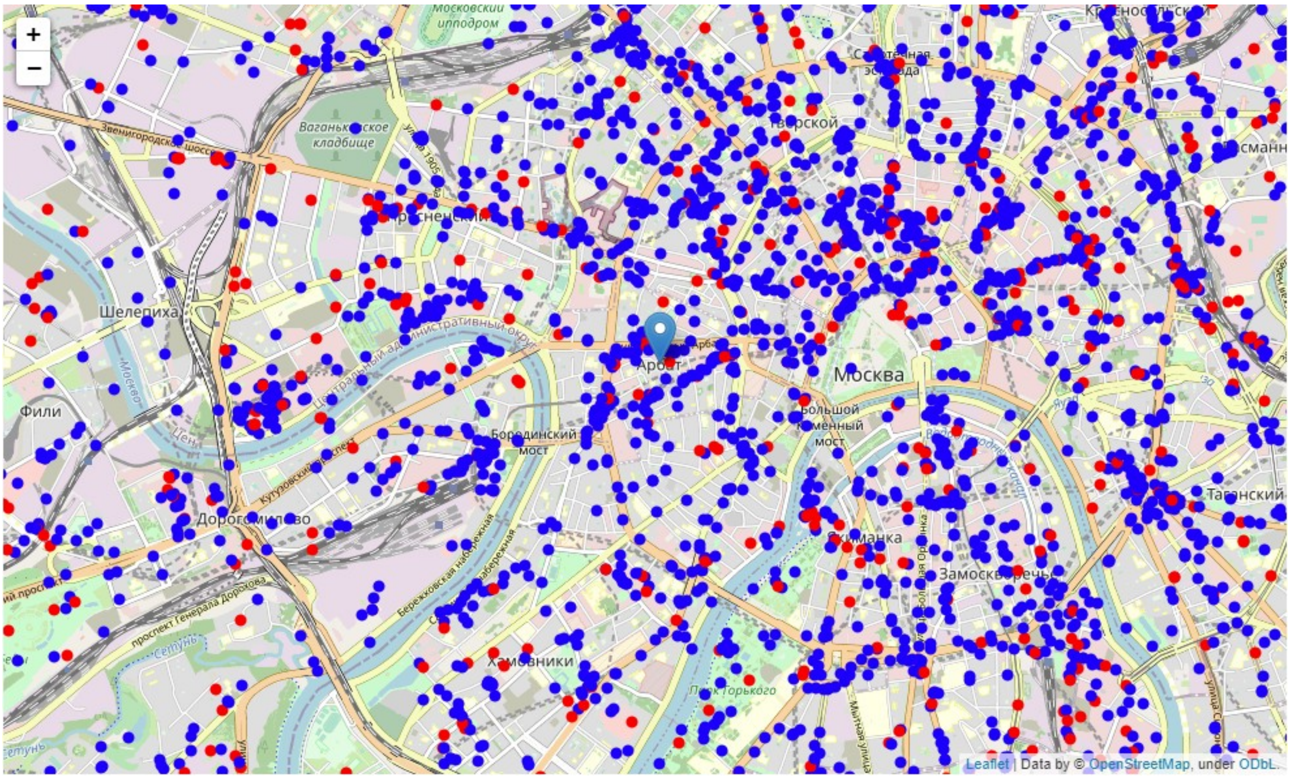
	Address	Latitude	Longitude	X	Y	Distance from center
0	Leninskiy Prospekt, 52, Moskva, 119333	55.699525	37.563091	409692.511369	6.173574e+06	5992.495307
1	Ulitsa Bardina, 6, Moskva, 119334	55.699636	37.572634	410292.511369	6.173574e+06	5840.376700
2	Vavilova Ulitsa, 13 строение 74, Moskva, Mosco...	55.699747	37.582178	410892.511369	6.173574e+06	5747.173218
3	ул. Вавилова, 9А строение 8, Moskva, 117312	55.699856	37.591721	411492.511369	6.173574e+06	5715.767665
4	5-Y Donskoy Proyezd, , корп. 15, Moskva, 119334	55.699965	37.601265	412092.511369	6.173574e+06	5747.173218
5	Zagorodnoye Shosse, 2A, Moskva, 117152	55.700074	37.610809	412692.511369	6.173574e+06	5840.376700

2.3 Filtering

Now that we have our location candidates, let's use Foursquare API to get info on coffee shops in each neighborhood.

We're interested in venues in 'food' category, but only those that are deal with coffee. So we will include only specific foursquare categories in our search.

Let's visialize all collected food services on the map and then color coffee shops in different color.



So now we have all the food in area within few kilometers from Arbat center, and we know which ones are coffee shops.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new coffee shop.

3. Methodology

In this project we will direct our efforts on detecting areas of Moscow that have low coffee shop density, but which are quite popular. We will limit our analysis to area ~6km aroud most popular district center.

In first step we found which district is most popular in our city of choise.

Second step in our analysis we fetch all food venues and cafe venues in 6km from most popular district center.

Third step - we calculate and explore 'cafe density' across nearby areas. We will use heatmaps to identify a few promising areas close to center with low number of food services in general and focus our attention on those areas.

Forth step - we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 300 meters, and we want locations without cafe restaurants in radius of 400 meters. We will winnerent map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of food services in every area candidate:

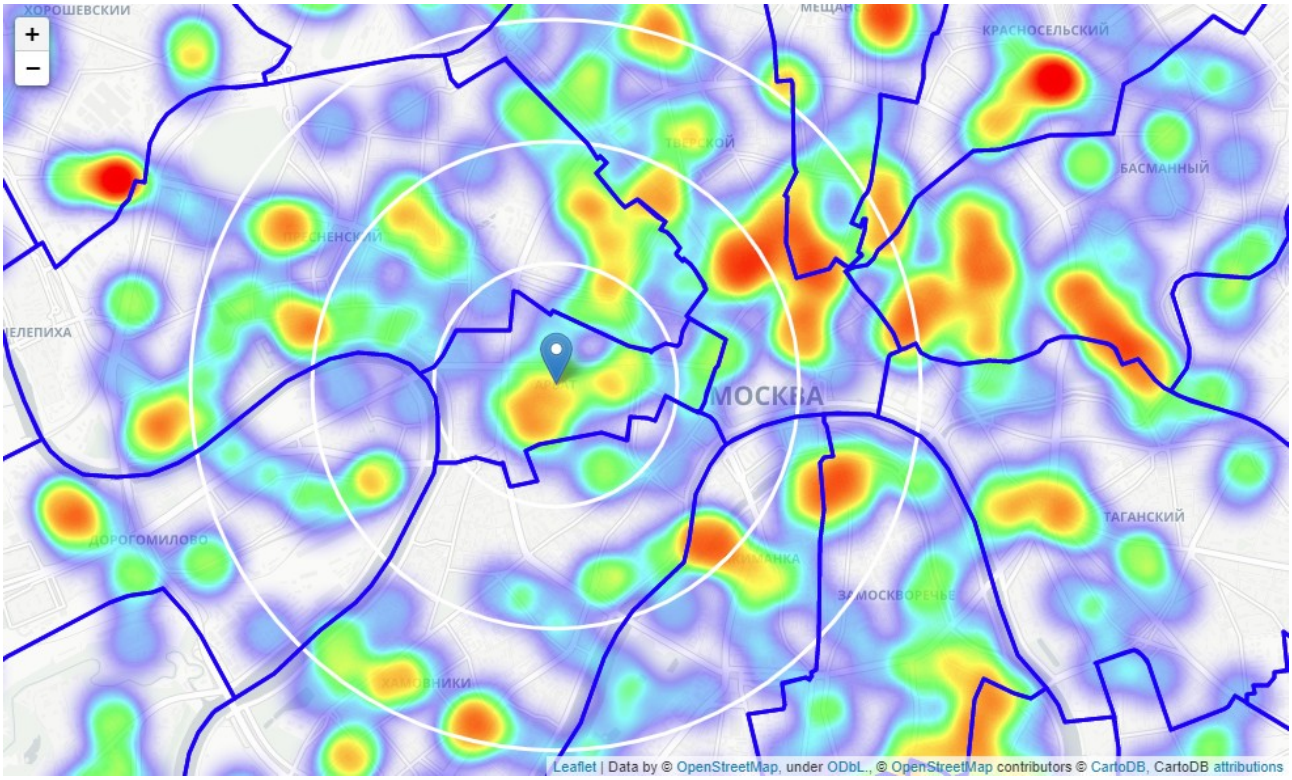
	Address	Latitude	Longitude	X	Y	Distance from center	Foods in area
0	Leninskiy Prospekt, 52, Moskva, 119333	55.699525	37.563091	409692.511369	6.173574e+06	5992.495307	4
1	Ulitsa Bardina, 6, Moskva, 119334	55.699636	37.572634	410292.511369	6.173574e+06	5840.376700	1
2	Vavilova Ulitsa, 13 строение 74, Moskva, Mosco...	55.699747	37.582178	410892.511369	6.173574e+06	5747.173218	5
3	ул. Вавилова, 9А строение 8, Moskva, 117312	55.699856	37.591721	411492.511369	6.173574e+06	5715.767665	0
4	5-Y Donskoy Proyezd, , корп. 15, Moskva, 119334	55.699965	37.601265	412092.511369	6.173574e+06	5747.173218	1
5	Zagorodnoye Shosse, 2A, Moskva, 117152	55.700074	37.610809	412692.511369	6.173574e+06	5840.376700	3

After that let's calculate the distance to nearest coffee shop from every area candidate center (not only those within 300m - we want distance to closest one, regardless of how distant it is).

	Address	Latitude	Longitude	X	Y	Distance from center	Foods in area	Distance to Coffee Shop
0	Leninskiy Prospekt, 52, Moskva, 119333	55.699525	37.563091	409692.511369	6.173574e+06	5992.495307	4	250.695899
1	Ulitsa Bardina, 6, Moskva, 119334	55.699636	37.572634	410292.511369	6.173574e+06	5840.376700	1	375.977637

	Address	Latitude	Longitude	X	Y	Distance from center	Foods in area	Distance to Coffee Shop
2	Vavilova Ulitsa, 13 строение 74, Moskva, Mosco...	55.699747	37.582178	410892.511369	6.173574e+06	5747.173218	5	98.947041
3	ул. Вавилова, 9А строение 8, Moskva, 117312	55.699856	37.591721	411492.511369	6.173574e+06	5715.767665	0	694.152134
4	5-Y Donskoy Proyezd, , корп. 15, Moskva, 119334	55.699965	37.601265	412092.511369	6.173574e+06	5747.173218	1	302.563561
5	Zagorodnoye Shosse, 2A, Moskva, 117152	55.700074	37.610809	412692.511369	6.173574e+06	5840.376700	3	481.427380

OK, so on average cafe can be found within ~275m from every area center candidate.
Let's crate a map showing heatmap / density of coffee shops. Also, let's show borders of Moscow boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Arbat Disctrict center.



This map is pretty scarse. It is because coffee shops have low % of all food services in Moscow.

4.1 Arbat District

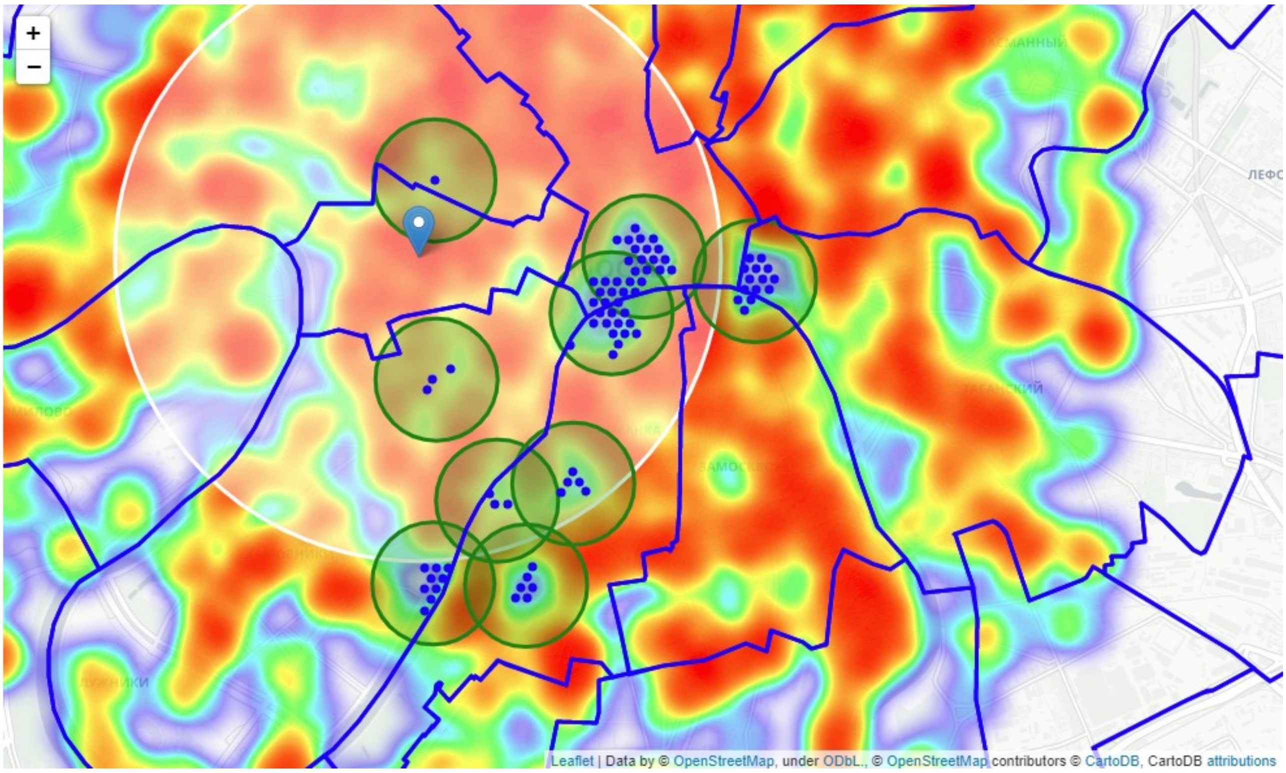
OK. Now let's calculate two most important things for each location candidate: **number of food services** (we'll use radius of **300 meters**) and **distance to coffee shops**.

	Latitude	Longitude	X	Y	Foods nearby	Distance to coffee shop
0	55.724260	37.591639	411542.511369	6.176290e+06	5	76.283398
1	55.725029	37.590815	411492.511369	6.176376e+06	3	162.675027
2	55.725047	37.592407	411592.511369	6.176376e+06	7	108.052055
3	55.725066	37.593999	411692.511369	6.176376e+06	10	129.951255
4	55.725084	37.595590	411792.511369	6.176376e+06	11	71.407647
5	55.725102	37.597182	411892.511369	6.176376e+06	12	35.726406

OK. Let us now **filter** those locations: we're interested only in **locations with no more than five food services in radius of 300 meters**, and **no cafe in radius of 150 meters**.

```
Locations with no more than five foods nearby: 148
Locations with no coffee shops within 150m: 200
Locations with both conditions met: 82
```

Looking good. What we have now is a clear indication of zones with low number of food services in vicinity, and *no* coffee shops at all nearby.
Let us now **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.
We going to use K-means clustering with variable amount of clusters. From several tries we see that good representation of clusters is when number of them are equal 9.
Heatmap with clusters:

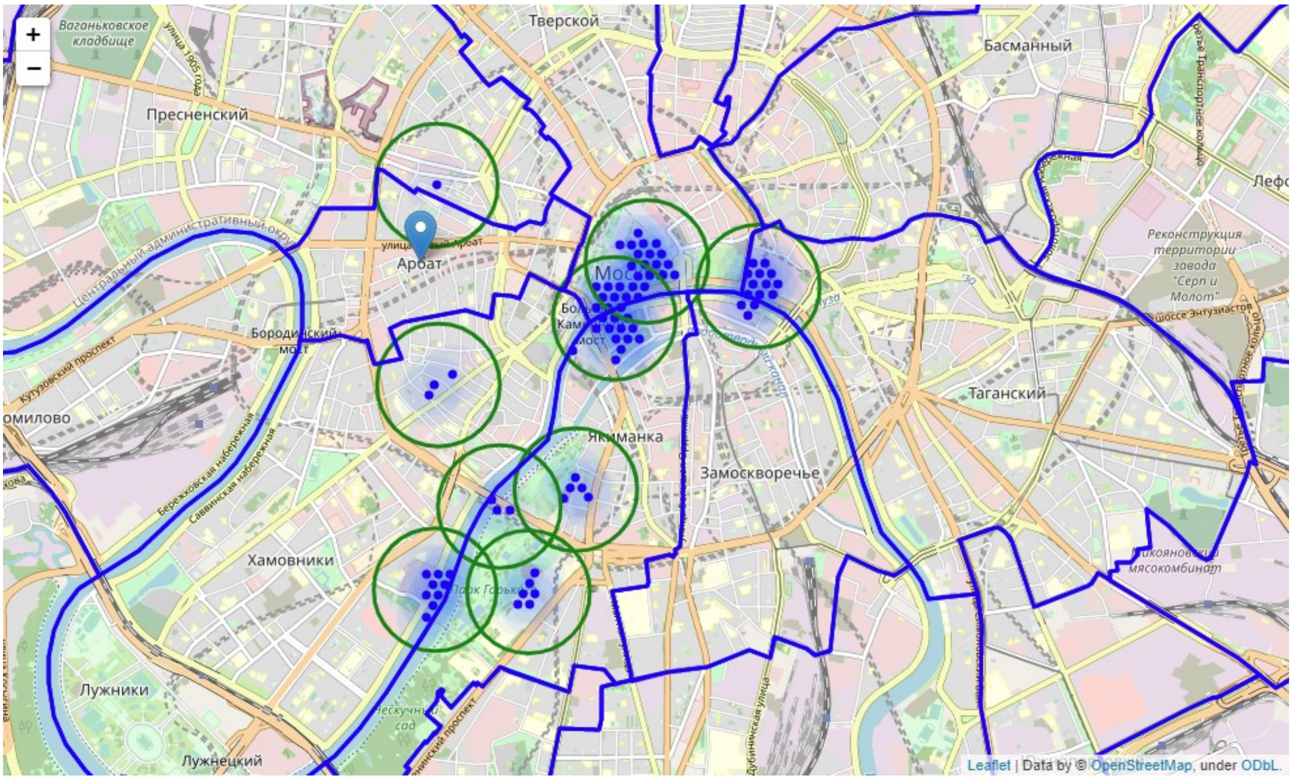


Not bad - our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.
Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.
Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:

We can observe that we clearly have 4 different locations to explore:

- Bolshoy Moskvoretsky Bridge (near Kremlin)
- Prechistenskiy Pereulok
- Park Gor'kogo
- Stolovyy Pereulok

Clean view on our clusters:



Finaly, let's **reverse geocode those candidate area centers to get the addresses** which can be presented to stakeholders.

```
==============================================================
Addresses of centers of areas recommended for further analysis
==============================================================

Kitaygorodskiy Passage, Moskva, 109074             => 2.8km from Arbat District
Center
Leninskiy Prospekt, 6, строение 7, Moskva, 119049  => 2.9km from Arbat District
Center
Sofiyskaya Embankment, 12, Moskva, 119072          => 1.7km from Arbat District
Center
Spasskaya Ulitsa, Moskva, 103132                   => 1.9km from Arbat District
Center
Krymsky Val, 9 строение 45, Moskva, 119049         => 2.1km from Arbat District
Center
Frunzenskaya Naberezhnaya, 22 строение 1, Moskva, 119146 => 2.7km from Arbat
District Center
Stolovyy Pereulok, 11 строение 2, Moskva, 121069   => 0.7km from Arbat District
Center
Chistyy Pereulok, 5A, Moskva, 119034               => 1.0km from Arbat District
Center
Yakimanskiy Pereulok, д. 6, Moskva, 119049         => 2.3km from Arbat District
Center
```

This concludes our analysis. We have created 9 addresses representing centers of zones containing locations with low number of food services and no coffee shops nearby. Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.

5. Results and Discussion

Our analysis shows that although there is a great number of food services in Moscow. Highest concentration of venues is in Arbat District, so we picked it as our 'new city center' and start research there.
We directed our attention to this more narrow area of interest we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than five food services in radius of 300m and those with an coffee shop closer than 150m were removed.
Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.
Result of all this is 9 zones containing largest number of potential new coffee shop locations based on number of and distance to existing venues - both food services in general and coffee shops particularly. This, of course, does not imply that those zones are actually optimal locations for a coffee shop! Purpose of this analysis was to only provide info on areas close to picked center but not crowded with existing food services - it is entirely possible that there is a very good reason for small number of food services in any of those areas, reasons which would make them unsuitable for a new coffee shop regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6. Conclusion

Purpose of this project was to identify Moscow areas close to most popular district center with low number of food services (particularly coffee shops) in order to aid stakeholders in narrowing down the search for optimal location for a new coffee shop. By calculating food service density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby food services. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.
Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.