New restaurant in Moscow.

March, 2019

I. Introduction

Moscow is a highly developing city. One of tendencies for last 2-3 years is a rapid development and constriction of areas on the outskirts of the city, especially Northwest Administrative District.

And new business centers and residential areas means that the district needs more social venues, including restaurants and cafés.

In this project I'll evaluate all existing venues within Northwest Administrative District and try to predict the best area for opening the new restaurant. As it's not the city center and the district is still developing the competition should be lower as well as the rental price for next few years when the expenses of the new business are the highest.

II. Data description.

I used Foursquare to get the list of venues in Moscow.

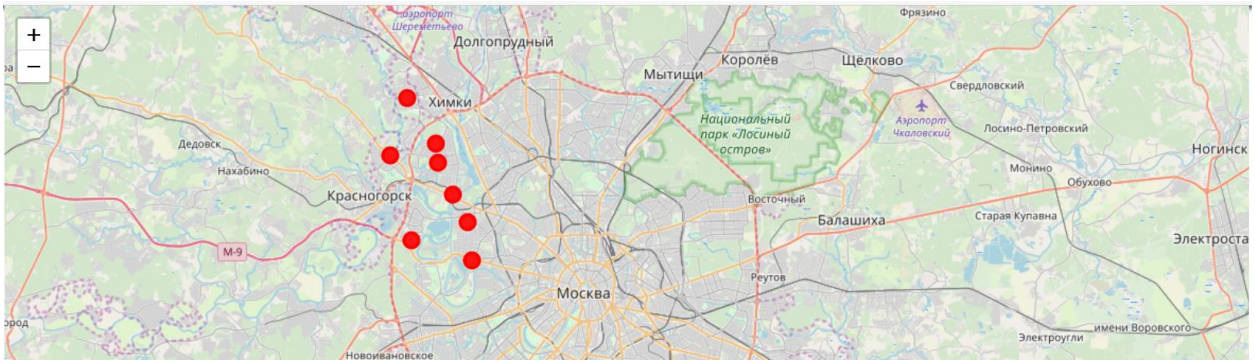
There are no public data for geo coordinates of the district so I had to create my own base. I used Google Maps to find the center of each area including to the Northwest Administrative District.

III. Methodology description.

For database of areas within the chosen district I created my own dataframe (using Google Maps for coordinates).

	District	Latitude	Longitude
0	Kurkino	55.891917	37.388824
1	Mitino	55.850973	37.366766
2	Pokrovskoe-Streshnevo	55.821969	37.448387
3	Strogino	55.789169	37.394399
4	North Tushino	55.859291	37.425743
5	South Tushino	55.845282	37.428557
6	Khoroshovo-Mnevniki	55.774161	37.473079
7	Shchukino	55.802158	37.467494

I used python folium library to visualize areas on the map using latitude and longitude.



Then I used Foursquare database to extract all venues by areas. I excluded from Categories the ‘Food’ (I want to see it on the separate map) and ‘Colleges and Universities‘ (I believe they will not have a great impact on the demand).

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kurkino	55.891917	37.388824	Modulus	55.894698	37.389286	Dance Studio
1	Kurkino	55.891917	37.388824	Melior	55.895828	37.391191	Dance Studio
2	Kurkino	55.891917	37.388824	Todes	55.897382	37.389333	Dance Studio
3	Kurkino	55.891917	37.388824	Центр Творчества И Обучения "Верный Курс"	55.896126	37.396305	Performing Arts Venue
4	Kurkino	55.891917	37.388824	Coral travel	55.897500	37.405043	Tour Provider

I drop duplicates and get the total number of venues within each area, in total – 77 venue for 7 areas.

District	Venue Category
Khoroshovo-Mnevniki	13
Kurkino	6
Mitino	9
North Tushino	10
Pokrovskoe-Streshnevo	10
Shchukino	18
South Tushino	8
Strogino	3

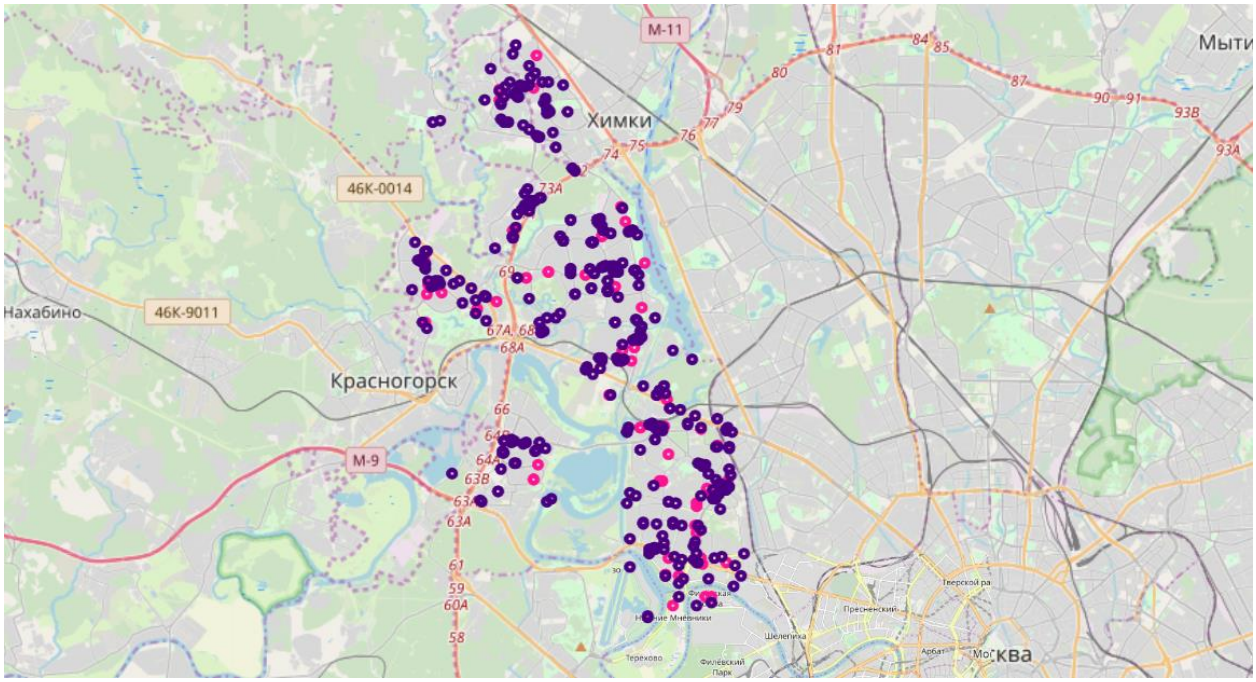
Now I do the same for category ‘Food’.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kurkino	55.891917	37.388824	Тутти и Джорни	55.888963	37.391325	Italian Restaurant
1	Kurkino	55.891917	37.388824	Арт Бар	55.889600	37.390103	Café
2	Kurkino	55.891917	37.388824	Мериталь	55.896054	37.396085	Café
3	Kurkino	55.891917	37.388824	БарБошко Loft 11	55.896259	37.397059	Restaurant
4	Kurkino	55.891917	37.388824	GianFranco	55.888773	37.393147	BBQ Joint

District	Venue Category
Khoroshovo-Mnevniki	46
Kurkino	45
Mitino	38
North Tushino	79
Pokrovskoe-Streshnevo	47
Shchukino	79
South Tushino	53
Strogino	24

To visualize how all venues are spread within areas I used the Folium library and create a new map, for Food venues I used the purple color, for all other venues – the pink color.

Even from the brief look it's clear that there are already lot's of different 'Food' venues and not so much infrastructure objects which can generate the high demand, so the competition could be quite rough.



For further analysis I'm using k-mean clustering. I used k=4, also I added top-5 venues for each area.

	District	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Kurkino	55.891917	37.388824	0	Dance Studio	Art Gallery	Performing Arts Venue	Tour Provider	Zoo Exhibit
1	Mitino	55.850973	37.366766	3	Tour Provider	Music Venue	Dance Studio	Movie Theater	Public Art
2	Pokrovskoe-Streshnevo	55.821969	37.448387	2	Dance Studio	Zoo Exhibit	Theater	Concert Hall	Music Venue
3	Strogino	55.789169	37.394399	0	Dance Studio	Tour Provider	Zoo Exhibit	Art Gallery	Arts & Entertainment
4	North Tushino	55.859291	37.425743	1	Multiplex	Tour Provider	Movie Theater	Theater	Arts & Entertainment
5	South Tushino	55.845282	37.428557	1	Concert Hall	Movie Theater	History Museum	Theater	Museum
6	Khoroshovo-Mnevniki	55.774161	37.473079	2	Dance Studio	Museum	Go Kart Track	Amphitheater	Art Gallery
7	Shchukino	55.802158	37.467494	2	Dance Studio	Multiplex	Art Gallery	Music Venue	Tour Provider

Cluster 1 - Strogino, Kurkino.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Kurkino	Dance Studio	Art Gallery	Performing Arts Venue	Tour Provider	Zoo Exhibit
3	Strogino	Dance Studio	Tour Provider	Zoo Exhibit	Art Gallery	Arts & Entertainment

Cluster 2 - North Tushino, South Tushino.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	North Tushino	Multiplex	Tour Provider	Movie Theater	Theater	Arts & Entertainment
5	South Tushino	Concert Hall	Movie Theater	History Museum	Theater	Museum

Cluster 3 - Pokrovskoe-Streshnevo, Khoroshovo-Mnevniki, Shchukino.

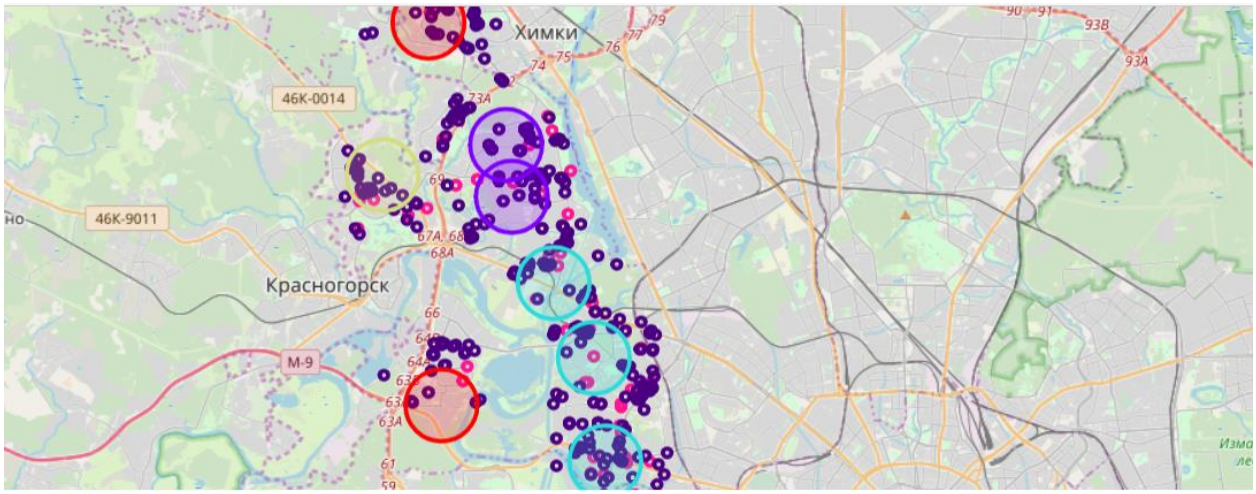
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Pokrovskoe-Streshnevo	Dance Studio	Zoo Exhibit	Theater	Concert Hall	Music Venue
6	Khoroshovo-Mnevniki	Dance Studio	Museum	Go Kart Track	Amphitheater	Art Gallery
7	Shchukino	Dance Studio	Multiplex	Art Gallery	Music Venue	Tour Provider

Cluster 4 - Mitino.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Mitino	Tour Provider	Music Venue	Dance Studio	Movie Theater	Public Art

IV. Results.

I created the new map to visualize clusters and the existing venues.



From this map and evaluation of each clusters I can give the recommendation to open the restaurant of the Cluster 3 (areas Pokrovskoe-Streshnevo, Khoroshovo-Mnevniki, Shchukino). Among the top-values are enough popular places such as museums, galleries etc. Comparing with other cluster it's not so far from the city center, which may give additional benefits.

V. Discussions.

As it was mentioned in the beginning this research was based only on data of the one district in Moscow - Northwest Administrative District which I believe to be the fastest growing. For better predictions it's necessary to get information for the whole city as well.

For the more proper research the potential investor should know the type of the restaurant he plans to open and the average pricing so all other categories could be dropped in Foursquare (or other sources).

Another point – for venues search it was used only Foursquare API source. The quantity of Non-Food venues which I get looks quite small for the area which means that probably not all mentioned on the Foursquare database. It means that it should be used additional statistical and geo databases (probably Google Nearby Search or some local sources) to get more results.

And the last but not least point is to have the statistical data for the profitability of existing competitors among areas and use it for clustering and segmentation as well, as it's a key point of each business work.

VI. Conslusion.

Making such type of the research should be the first thing which the investor do before creating any new business. It shows in the very simple way (even for someone without deep business knowledge) all main aspects which should be considered – the potential demand (and where it comes from), potential profitability and competitors.

VII. References.

Foursquare API - https://developer.foursquare.com/docs

Google maps for developers - https://developers.google.com/maps/documentation/geocoding/intro.

Moscow districts - http://mosopen.ru/regions.