Moscow City Business Recommendation Metrics

Timur Taumurzaev

January, 2020

1. Introduction

1.1 Background

Moscow is the most popular tourist center of Russia. The Kremlin, Red Square, Novodevichy Convent and Ascension Church in Kolomenskoye are included in the UNESCO World Heritage List. It is one of the most important transportation hub in Russia: the city is served by 6 airports, 9 railway stations, 3 river ports. Since 1935, the subway has been operating in Moscow. Moscow is the largest nationwide financial center, an international business center and a management center for a large part of the country's economy. For example, about half of the banks registered in Russia are concentrated in Moscow. In addition, most of the largest companies are registered and have central offices in Moscow, although their production can be located thousands of kilometers from it. According to 2008 data, Moscow was ranked 15th among the largest cities in the world in terms of GDP ($ 321 billion). If someone is looking to start their own business in Russia, Moscow is the first area that comes up to mind.

1.2 Problem

What is the major question that you have to deal while opening a business in Moscow? The key factor of success is a location. A client flow, number of existing same category venues in the neighborhood, distance of possible location from city center and rent costs are the factors that will influence our decision. Therefore our goal is to determine where exactly should you setup your office or open a restaurant or a beauty shop. This project aims to predict what location is going to be the

most appropriate for your business based on relevant data. Specifically, this report will be targeted to stakeholders interested in opening a Pub chain in Moscow. What is pub? A pub can be dated back to Roman taverns, that were known as Alehouses. The clients of pubs mostly include locals or regulars that come to relax and unwind after work. The pub takes its name from “public house,” a type of establishment with its origin in the culture of countries like Britain, Australia, Ireland, and New Zealand. The atmosphere of a pub is cozy and more relaxed. Along with the drinks, the pub also offers a wide range of food, so it is also called as a restaurant. People generally go to pubs for having a few drinks and quick bites with a familiar circle of friends.

2. Data acquisition

At first we have to decide what type of locations is going to be taken into consideration. As long as we are up to find an appropriate place for a pub, we need to set up some requirements to its location. Let us define our typical client wishes and possibilities. It should be easy to get to our pub by public transport, it should have a dense flow of customers nearby and rent cost should be taken into consideration as well. The most obvious emplacement at this point is a subway station. The majority of required data can be found at https://mosmetro.ru/. Number of pubs and their type and location near every station will be obtained using **Foursquare API.** Data with commercial real estate for rent in Moscow is confidential and cannot be obtained with free sources. But instead of it we will use residential properties rent costs, easy to see, they have a direct relationship. Thus, we can use it after some transformations.

Data downloaded or scraped from multiple sources were combined into one table.

3. Solving a problem

Since there are lots of pubs in Moscow we will try to detect locations that are not already crowded with pubs. We are also interested in areas with high passenger flow and acceptable rent cost. We would also prefer locations as close to city center as possible. In addition we will perform our analysis with *k-means* ML algorithm clustering so we can easily flex with requirements, and find any appropriate locations according to stakeholders wishes.

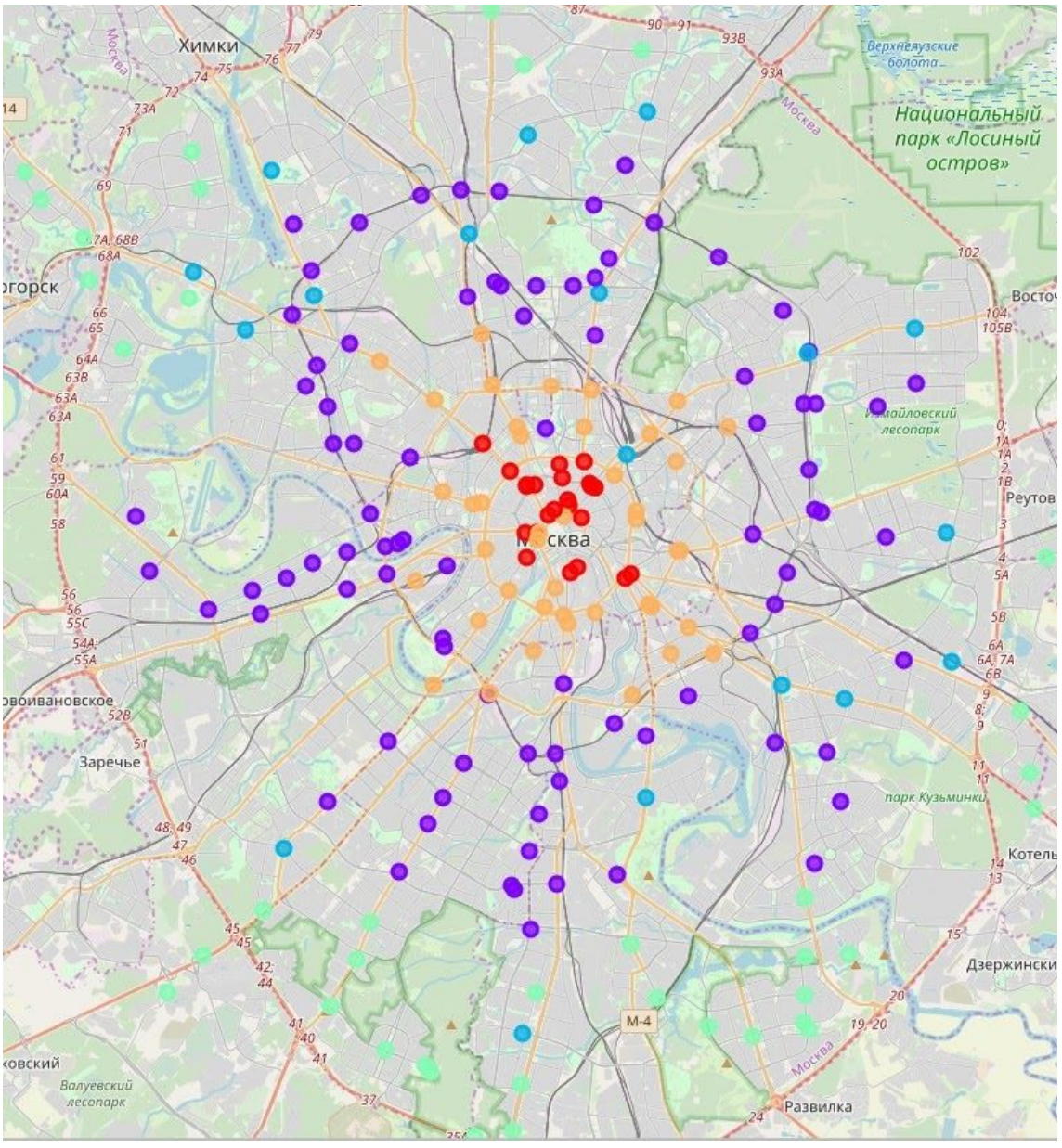

Figure 1. Cluster map distribution.

We will use our data science powers to generate a few most promising subway stations based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

4. Methodology and discussion.

K-means machine learning algorithm partitioned subway stations into 5 groups [fig.1]. As we can see they have distance based characterization, which is no surprise, because the distance between the stations and city center is in feature data. Also the other features brought some impact. Methodology of the research was the normalization and summarizing the resulting features which were estimated in 0, 1 interval, whereas 0 is ‘bad’ result and 1 is ‘good’. For example we have 1 in passenger flow feature for ‘Комсомольская’ station. The highest value in our sample.

This is easily explained by the fact that it has 3 railway terminals nearby and it is the largest transport hub in Moscow. However each cluster has its pros and cons [fig.2]. Such as Distance feature for cluster 0 or cont rent (‘PriceMSQ’) feature for cluster 3.

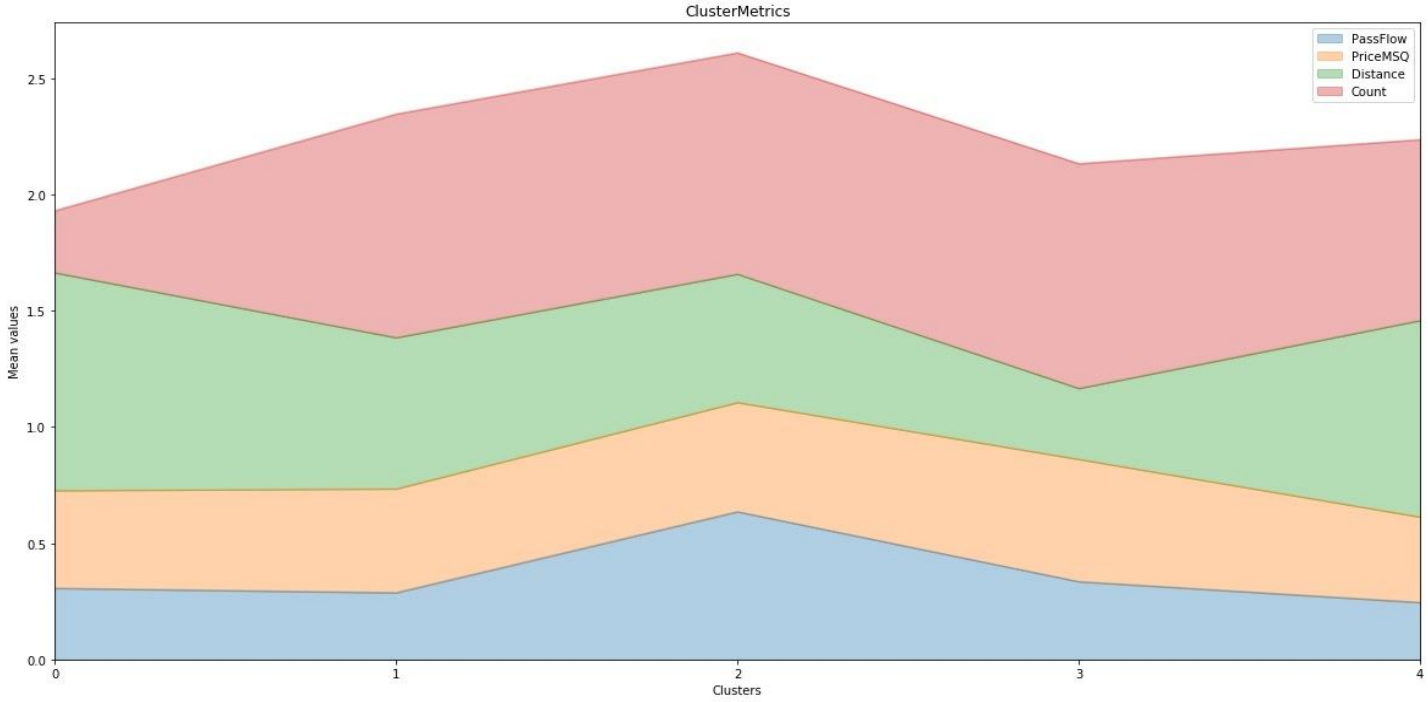

Figure 2. Metrics area plot.

Cluster Number	Color	Sum value
0	Red	1,929
1	Purple	2,345
2	Blue	2,609
3	Light green	2,132
4	Orange	2,235

Figure 3. Cluster metrics.

After summarizing the values for each cluster [fig. 2] we obtained that cluster number 2 has the greatest value among others. Thus, the following search for the best stations was performed among this group. Using the same summarizing method for the subgroup of stations we easily discover top 10 stations [fig. 4]. According to stakeholders specific requirements we can choose the appropriate station list.

	Station	Cluster Labels	PassFlow	PriceMSQ	Distance	Count	Total
44	Выхино	2	0.943880	0.564970	0.476871	1.000000	2.985721
66	Комсомольская	2	1.000000	0.238607	0.875619	0.866667	2.980892
170	Текстильщики	2	0.634955	0.490114	0.654738	1.000000	2.779807
125	Петровско-Разумовская	2	0.620524	0.501950	0.614195	1.000000	2.736669
105	Новогиреево	2	0.609300	0.544118	0.510839	1.000000	2.664257
79	Кузьминки	2	0.619989	0.519786	0.579828	0.933333	2.652937
198	Черкизовская	2	0.506146	0.507306	0.618293	1.000000	2.631745
208	Щелковская	2	0.663816	0.542543	0.487221	0.933333	2.626914
32	ВДНХ	2	0.715660	0.404593	0.702297	0.800000	2.622550
65	Коломенская	2	0.523250	0.451232	0.641165	1.000000	2.615646

Figure 4. Top ten stations.

5. Conclusion.

Purpose of this project was to cluster the potential locations for a pub chain in Moscow. We reduced the problem to identifying Moscow subway stations close to center with low number of pubs and bars nearby, high passenger flow and low commercial rent cost, in order to aid stakeholders in narrowing down the search for optimal location. With initial scrapped data about rent, passenger flow, calculated distance between stations and city center, also pubs distribution from Foursquare we have identified general cluster that satisfy our analysis. In fine we obtain the final exploration about the appropriate stations to open pub nearby according to stakeholder further requirements.

Thank you for your attention!