Choosing a city for relocation based on its availability of entreating facilities

Alexey Leontyev

May 08, 2020

1. Introduction

My name is Alexey and I live in Russia. I'm married and we have two children. Currently we live in a town called Stupino which is quite small (~40 000 people). I've received a job offer and we plan to relocate to other location. I will be working on a manufacturing plant situated in a town called **Protvino** ("Протвино" in Russian). There is a choice to which city to move, because several options exist. In 10 kilometers from Protvino, there is a bigger city called **Serpukhov** ("Серпухов" in Russian).

For me the choice is irrelevant because anyway I will spend majority of my time on a plant. However my wife is very sensitive to the choice of entertaining facilities accessible in a place of living. She enjoys spending time in coffee shops, shopping centers and similar places.

The problem i will be solving using data science and Foursquare service is the **choice of living place** between two cities - **Protvino** and **Serpukhov**.

The **target audience** of my data science project is the most important stakeholder in my life - **my wife**.

Main criterion for final desicion is the availability of entertaining servises and facilities.

2. Data requirements

It is important to align the choice of a city with my main stakeholder in advance. As far as I know her, she cares both about availability of facilities and its quality. Therefore I plan to utilize the following data:

1. Using the **geo package** I will obtain the coordinates of both cities.

2. I will find all the places in 10 km radius using **Foursquare places** service

3. I will filter these places and exclude everything that is not of a entertaining nature. For example places like banks, hotels etc should be ignored. On this stage it will be possible to compare the absolute number of entertaining facilities of two cities.

4. Using **Wikipedia** pages I will obtain the population of both cities and therefore it will be possible to calculate and then compare "per capita" availability of entertaining facilities. It is nessessary step because theoretically desired facilities might be overcrowded.

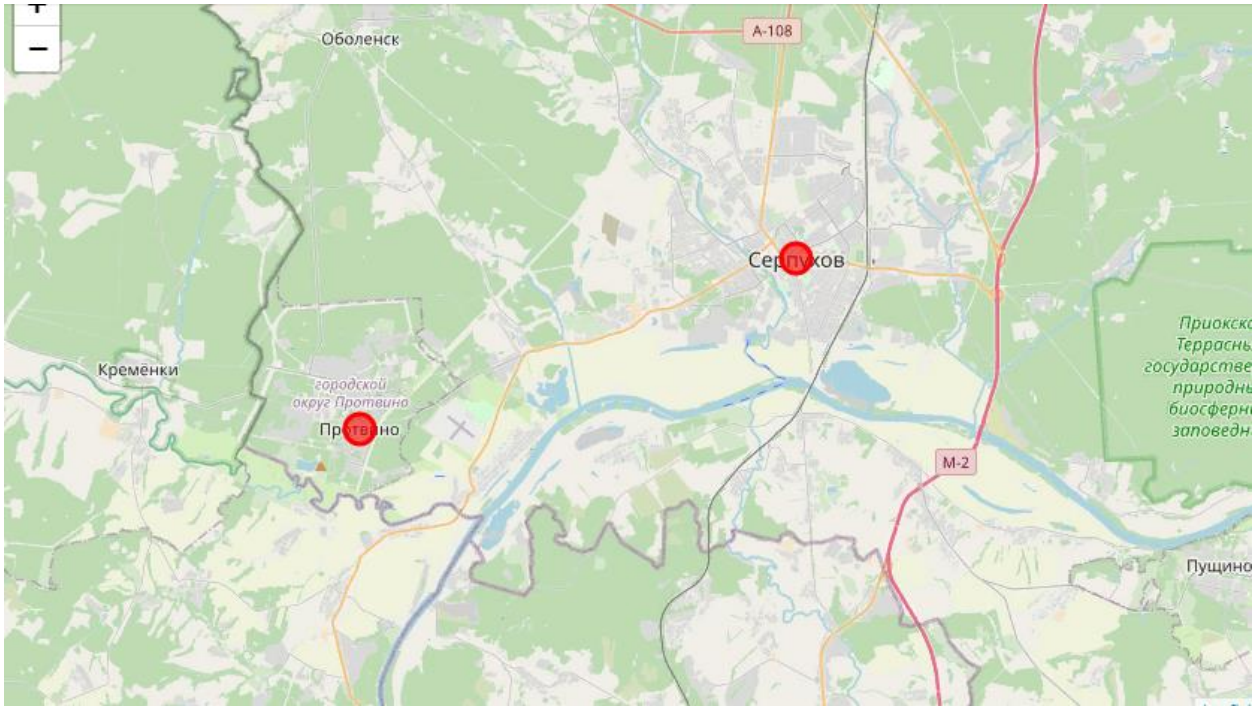
5. Not only availability is important, but also the quality. In order to access the quality of entertaining facilities, I will analyze the **Foursquare tips** related to the locations. I'm going to calculate the "average" rating for the entertaining facilities for both cities.

6. I will interview my main stakeholder and ask her about the importance of availability and quality of entertaining services. I will ask to assign the **weight of importance**.

7. Based "per capita" availability of entertaining facilities in both cities, its ratings based on tips and importance weights provided by main stakeholder, **I will finally calculate the rating for each city under consideration.**

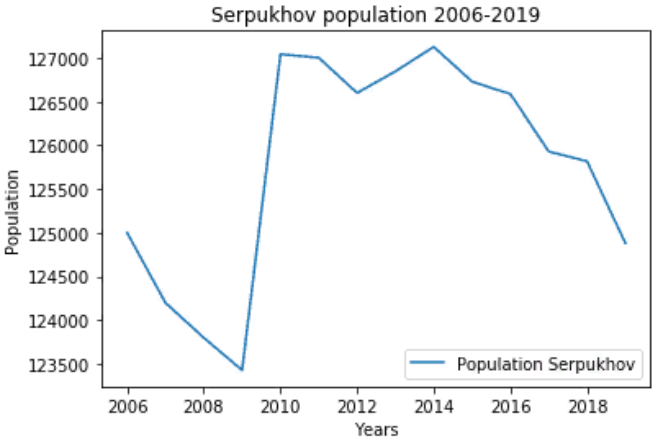
3. Methodology and Data analysis

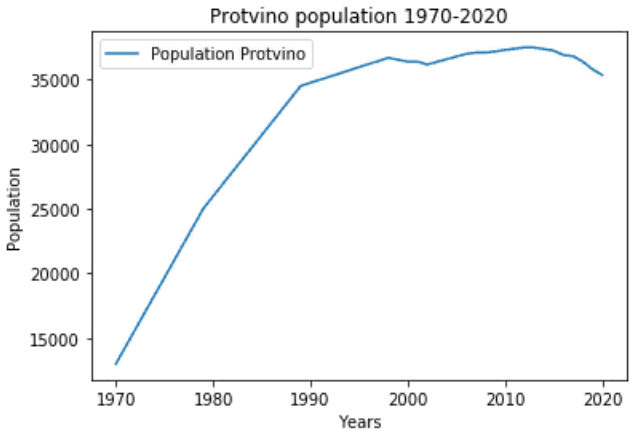
As a first step, let's look at a map to get first idea of how cities are located, based on coordinates obtained from Geo package



It is clear that cities are located at a close proximity. Another observation is that Serpukhov (on the right) appears to be wider (has a bigger square).

As we are looking for entertaining venues, it is important to check the population and its dynamics. Theoretically, bigger city should have more entertaining services. To access population, I scraped the Wikipedia data and transformed it into data frame.



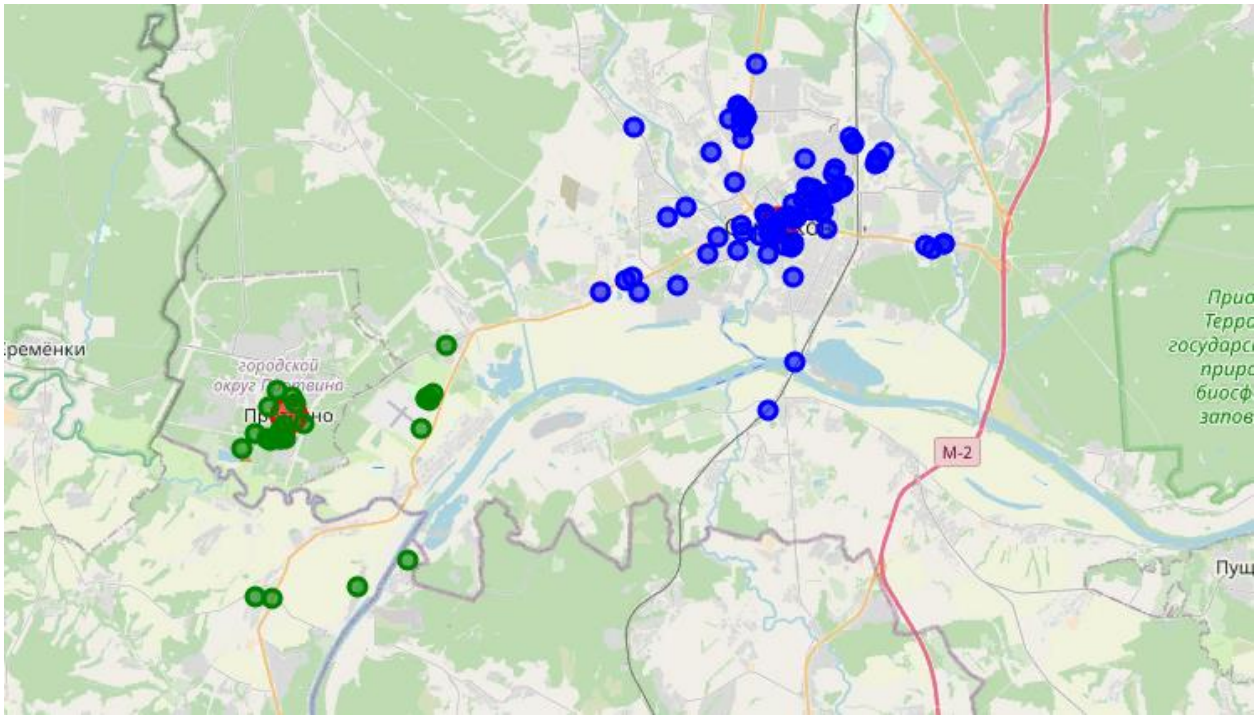


We do see that

1) Serpukhov is 5 timer bigger in terms of population
2) Both cities demonstrate decline since 2010-2012

Let's obtain a total number of venues for each city based on Foursquare data.

We see that for Serpukhov the number is 87 vs 27 for Protvino. This corresponds to the difference in square and population.



We see that for Serpukhov the venues are gradually spread around city center. For Protvino we can identify some remote venues, and it seems that they are wild nature related. They are close to a river or a lake.

However I need to exclude all the venues that are not entertaining related from the analysis, in other words I need to clean the data. For that purpose I found all the categories related to identified venues.

In total, I had 44 types of category. I exported this list to csv and then manually entered a feature whether the category is entertainment related or not. This csv file was imported back and become a dictionary.

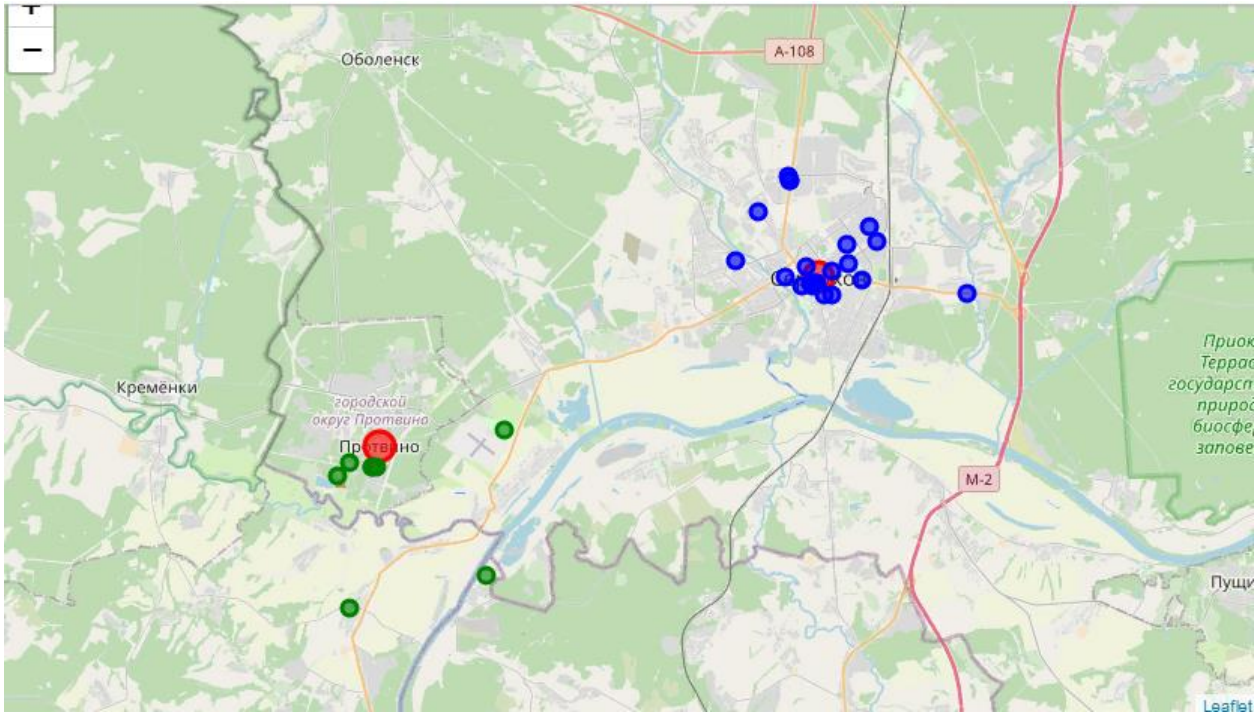
	categories	Entertainment
0	Beach	Y
1	Pub	Y
2	Spa	Y
3	Athletics & Sports	Y
4	Food	Y

This allowed me to add another column to venues' data frames for each city by merging the original table and dictionary.

	name	categories	lat	lng	id	Entertainment
0	Прожарка	Burger Joint	54.914220	37.418401	573c45b7498eca68d535e1d6	Y
1	Три Сковородки	Blini House	54.914231	37.415952	51b491f5498eb97867d54651	Y
2	Kitchen Market	BBQ Joint	54.913395	37.411720	5cfba0508496ca002c2566d9	Y
3	Шинок "Куманец"	Ukrainian Restaurant	54.913715	37.416524	505de15de4b08887c64c54ef	Y
4	Матрена	Tea Room	54.914256	37.418634	4c45949759d99c746525fcad	Y

After filtering venues' data using this new feature, I've obtained only venues of entertaining nature: 22 for Serpukhov and 8 for Protvino. For both cities it represent approximately ¼ of total number of venues.

The geographical spread looks as follows:



It is obvious that in Protvino there is clearly a lack of entertaining facilities inside the city.

name	categories	lat	lng	id	Entertainment
Протвинский карьер	Beach	54.862611	37.199485	4f93cb85e4b008256603318b	Y
Pub 501$	Pub	54.865128	37.217279	4ccb371072106dcb829f9299	Y
Дракино SPA	Spa	54.874877	37.276042	529998cc498e6622f7eef6b2	Y
ДС Импульс	Athletics & Sports	54.866369	37.205464	4f7dc68ae4b0208681e76297	Y
Малибу	Food	54.865202	37.215627	4e53c78eae606f5737a2cce8	Y
Боулинг-клуб "Малибу"	Sports Bar	54.865176	37.215633	4f5c39cb121d26e952f2d073	Y
загородный клуб «Гуляй-город»	RV Park	54.836176	37.267478	576e5c57498e85d8f001cbdd	Y
Volkov-Sky	Recreation Center	54.827754	37.205346	5388827b498e1afc1634fb01	Y

We see that only small number of venues are presented by typical places like pubs, cafés etc. Almost half consists of outdoor entertainments such as parks, beaches and recreation centers.

Another way of accessing the situation is check “per capita’ availability of entertaining facilities. We can do that by dividing total number by city population.

	Town	Entertaining venues	Population	Venues per 1000 people
0	Serpukhov	22	126273	0.174226
1	Protvino	8	35367	0.226200

This measure provides an interesting view because the figure for Protvino is higher. That means that lower number of facilities is compensated by lower population and as a result, entertaining facilities are less crowded compared with Serpukhov.

4. Results

We can conclude that Serpukhov city provides significantly higher access to entertaining facilities compared with Protvino: 22 vs 8. Therefore Serpukhov should be considered as a more suitable place to relocate to since it is a main criterion for my decision maker.

However there is a risk that entertaining facilities is Serpukhov might be overcrowded based on “per capita measure”

5. Discussion

Despite the choice of a city is clear, there are some other observations to mention:

1) Both cities demonstrate a decline in population which may result in a long term in a corresponding decline of entertaining venues. It might be beneficial looking for other cities that might prosper in the future
2) Quality criterion was not utilized during the analysis because free account on Foursquare limits the number of tips’ requests. In case of premium account it might be beneficial to access the ratings of entertaining facilities of both cities.

6. Conclusion

Performed exercise provided the answer to the problem, the city we should relocate to is Serpukhov.