

The final

Moscow coffee houses

Kirill Mokhniuk

March, 2021

Introduction

In this work I want to analyze amount of coffee houses per capita in different Moscow boroughs. I believe that it will be interesting for entrepreneurs who want to start own coffee house in Moscow, but don't know where there will be fewer competitors.

Data

For the work I need two parts of data.

1. Data about name, size, population, longitude, and latitude of Moscow boroughs.
2. Data about amount of coffee houses in every Moscow borough.

Methodology

First of all I scrapped data about Moscow boroughs from [Wikipedia article](#) (it is in Russian, but I translated data to English). Then clean it and receive a dataframe:

Out[94]:

	Borough	Area	Population
0	Akademicheskij	583	7110 038
1	Alekseevskij	529	780 634
2	Altufevskij	325	757 697
3	Arbat	211	736 308
4	Aeroport	458	779 541
...
141	Sosenskoe, poselenie	6707	733 096
142	Troick	1633	761 284
143	Filimonkovskoe, poselenie	3577	78017
144	SHChapovskoe, poselenie	8606	79601
145	SHCherbinka, gorodskoj okrug	762	754 008

146 rows x 3 columns

Then I used Yandex API (it is a good choice for Russia) for longitude and latitude for every borough, and added it to the dataframe:

Out[99]:

	Borough	Area	Population	Longitude	Latitude
0	Akademicheskij	583	110038	37.572684	55.688005
1	Alekseevskij	529	80634	37.650163	55.812949
2	Altufevskij	325	57697	37.582278	55.879849
3	Arbat	211	36308	37.590003	55.751138
4	Aeroport	458	79541	37.542599	55.803312
...
141	Sosenskoe, poselenie	6707	33096	37.465892	55.57833
142	Troick	1633	61284	37.295014	55.481546
143	Fillimonkovskoe, poselenie	3577	8017	37.322368	55.570014
144	SHCHapovskoe, poselenie	8606	9601	37.380848	55.391631
145	SHCHerbinka, gorodskoj okrug	762	54008	37.567653	55.507525

146 rows × 5 columns

Next step is receiving data about coffee houses in Moscow. I had a problem: different boroughs have different size of area, and I did know which radius I should choose for Foursquare API, but I had size of area! So I decided to consider each borough as a circle, and found radius for every borough. Yes, it is not very accurate. So after that I parsed data from Foursquare about venues in these boroughs.

Results

Unfortunately I had fatal problem. Foursquare hasn't a lot data about Moscow venues, and it is impossible to use it for good analysis. So at the final step I received this dataframe:

Out[104]:

	Borough	Area	Population	Longitude	Latitude	Number of places
0	Akademicheskij	5.83	110038	37.572684	55.688005	15
1	Alekseevskij	5.29	80634	37.650163	55.812949	3
2	Altufevskij	3.25	57697	37.582278	55.879849	2
3	Arbat	2.11	36308	37.590003	55.751138	1
4	Aeroport	4.58	79541	37.542599	55.803312	0
...
141	Sosenskoe, poselenie	67.07	33096	37.465892	55.57833	55
142	Troick	16.33	61284	37.295014	55.481546	4
143	Fillimonkovskoe, poselenie	35.77	8017	37.322368	55.570014	0
144	SHCHapovskoe, poselenie	86.06	9601	37.380848	55.391631	6
145	SHCHerbinka, gorodskoj okrug	7.62	54008	37.567653	55.507525	4

146 rows × 6 columns

For example, Foursquare has data just about 15 venues in Akademicheskij borough, and it is not only coffee houses, it is different types of venues: pharmacies, restaurants etc. It is impossible, because Akademicheskij borough is pretty close to city center, so in real life there are hundreds of venues there.

Discussion

This work shows us that Foursquare not very popular in Moscow (and I guess in Russia at all) and people don't add info about venues to Foursquare database, but if we choose other city, for example NY or Paris, where Foursquare has more data, we can do this analysis.

Conclusion

Unfortunately, I didn't reach the goal of this work: didn't find out amount per capita of coffee houses in Moscow boroughs. But negative result is still result, and I believe that this work has some value. Additionally, I used all skills from the specialization and showed you, that I can handle with data.

Thanks for reading, and have a nice day!