A. Introduction

A.1. Description & Discussion of the Background

Minsk is the capital and largest city of Belarus. As the capital, Minsk has a special administrative status in Belarus and is the administrative center of Minsk Region and Minsk District. The population in January 2018 was 1,982,444 making Minsk the 11th most populous city in Europe. As a resident of this city, I decided to use Minsk in my project. The city is divided into 9 boroughs and 64 neighborhoods in total.

In this project, I would like to raise the topic of a healthy lifestyle. Consider which neighborhoods of Minsk are most suitable for this (gyms, swimming pools, yoga studios q-ty). Additionally, I will attempt to answer the question for investor “Where should I open a gym?”

When we consider all these problems, we can create a map where each neighborhood is clustered according to the venue density.

A.2. Data Description

To consider the problem we can list the data as below:

- I didn’t found any useful file with Administrative Divisions. So I used different resources (wiki, Ato.by, realt.by) and created excel file with all boroughs and neighborhoods of Minsk. I cleaned the data and used **geopy** library to add latitude and longitude.
- I used **Forsquare API** to get the most common venues.
- I used **transliterate** library to convert all text from Russian

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components *Borough, Neighborhood, Full address* information of the city.

	Borough	Neighborhood	Full address
0	Фрунзенский	Сухарево	Минск, Фрунзенский район, Сухарево
1	Фрунзенский	Запад	Минск, Фрунзенский район, Запад
2	Фрунзенский	Красный Бор	Минск, Фрунзенский район, Красный Бор
3	Фрунзенский	Медвежино	Минск, Фрунзенский район, Медвежино
4	Фрунзенский	Кунцевщина	Минск, Фрунзенский район, Кунцевщина

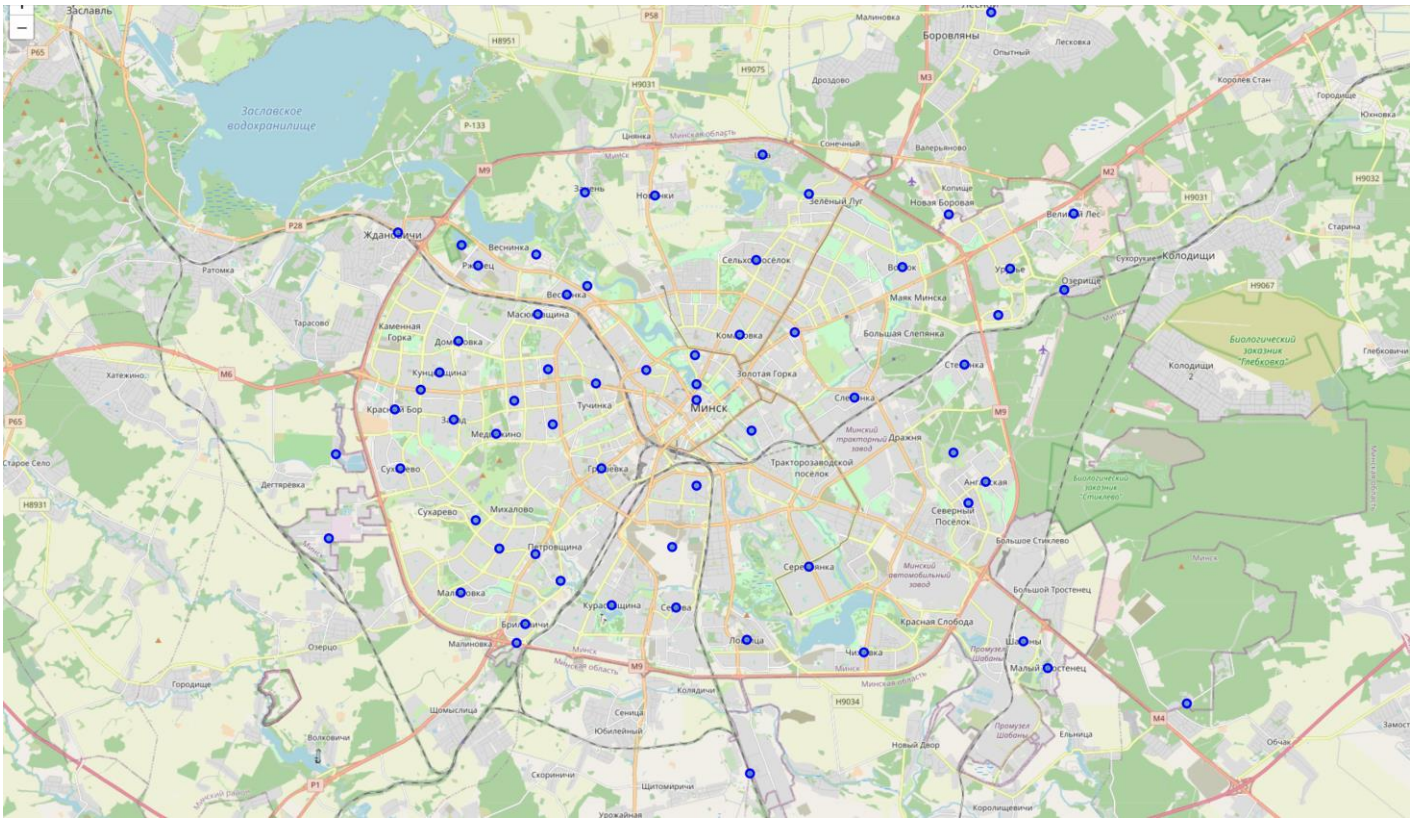
I used **geopy** library and Full address column to add longitude and latitude of all Neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Фрунзенский	Сухарево	53.8867	27.4271
1	Фрунзенский	Запад	53.8993	27.4503
2	Фрунзенский	Красный Бор	53.902	27.4245
3	Фрунзенский	Медвежино	53.8955	27.4689
4	Фрунзенский	Кунцевщина	53.9114	27.4441

I used **transliterate** library to convert from Cyrillic (Russian language) to Latin.

	Borough	Neighborhood	Latitude	Longitude
0	Frunzenskij	Suharevo	53.8867	27.4271
1	Frunzenskij	Zapad	53.8993	27.4503
2	Frunzenskij	Krasnyj Bor	53.902	27.4245
3	Frunzenskij	Medvezhino	53.8955	27.4689
4	Frunzenskij	Kuntsevschina	53.9114	27.4441

I used python **folium** library to visualize geographic details of Minsk and its boroughs/neighborhoods.



I utilized the Foursquare API to explore the neighborhoods and segment them. I designed the limit as **100 venue** and the radius **600 meter** for each neighborhood from their given latitude and longitude information. In summary of this data 1030 venues were returned by Foursquare. Here is a merged table of boroughs and venues. We can see 20 main venue categories:

```
print(minsk_venues.shape)
minsk_venues.head()
```

```
(1030, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Suharevo	53.88665	27.427141	фитнес-клуб «Старт+»	53.889615	27.428852	Gym / Fitness Center
1	Suharevo	53.88665	27.427141	Салон красоты "Карамель"	53.886722	27.430281	Yoga Studio
2	Suharevo	53.88665	27.427141	Банно-оздоровительный комплекс «Сухарево»	53.887432	27.422580	Bath House
3	Suharevo	53.88665	27.427141	Мила	53.888478	27.422119	Cosmetics Shop
4	Suharevo	53.88665	27.427141	Точка на Сухаревской	53.888686	27.422195	Beer Store

```
print(minsk_venues['Venue Category'].value_counts()[0:20])
```

```
Bus Stop                84
Café                    35
Coffee Shop             34
Gym / Fitness Center    29
Pizza Place             26
Food & Drink Shop       26
Gym                     25
Park                    24
Restaurant              23
Cosmetics Shop          23
Cocktail Bar            21
Hotel                   18
Pharmacy                17
Bar                     16
Soccer Field            15
Grocery Store           13
Spa                     12
Flower Shop             12
Electronics Store       12
Fast Food Restaurant    11
Name: Venue Category, dtype: int64
```

As we raised the topic of healthy lifestyle, let’s explore Venue categories (gyms, pools, yoga studios, etc.)

```
print(minsk_gyms.shape)
minsk_gyms.head()
```

```
(75, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Suharevo	53.886650	27.427141	фитнес-клуб «Старт+»	53.889615	27.428852	Gym / Fitness Center
1	Suharevo	53.886650	27.427141	Салон красоты "Карамель"	53.886722	27.430281	Yoga Studio
11	Zapad	53.899276	27.450313	Фрунзенский физкультурно-оздоровительный центр	53.894694	27.451526	Gym / Fitness Center
12	Zapad	53.899276	27.450313	Фрунзенский ФОЦ	53.902711	27.449563	Gym / Fitness Center
35	Kuntsevschina	53.911438	27.444141	Тонус	53.910778	27.448675	Gym

```
print(minsk_gyms['Venue Category'].value_counts()[0:20])
```

```
Gym / Fitness Center    29
Gym                     25
Athletics & Sports       8
Gym Pool                 4
Pool                     3
Yoga Studio              3
Gymnastics Gym           2
Sports Club              1
Name: Venue Category, dtype: int64
```

In summary 75 venues were returned by Foursquare. Then I created a table which shows list of top venue category for each neighborhood in below table.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Akademgorodok	Gymnastics Gym	Gym / Fitness Center	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gym Pool
1	Akademija nauk	Gym Pool	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym / Fitness Center
2	Angarskaja	Gym / Fitness Center	Pool	Yoga Studio	Sports Club	Gymnastics Gym	Gym Pool	Gym
3	Borovljany	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
4	Brilevichi	Pool	Yoga Studio	Sports Club	Gymnastics Gym	Gym Pool	Gym / Fitness Center	Gym

Then I created a data-frame with pandas one hot encoding for the venue categories. I used pandas groupby on neighborhood column and calculate the mean of the frequency of occurrence of each venue category.

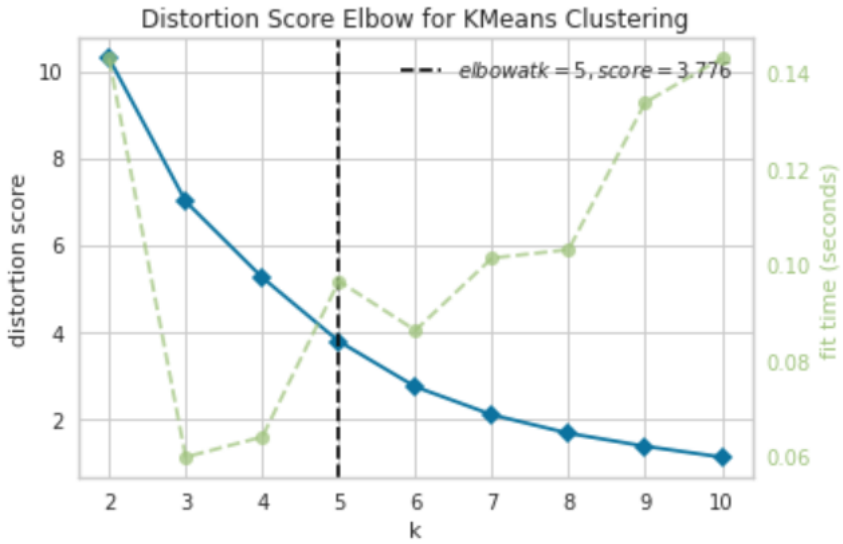
	Neighborhood	Athletics & Sports	Gym	Gym / Fitness Center	Gym Pool	Gymnastics Gym	Pool	Sports Club	Yoga Studio
0	Akademgorodok	0.333333	0.0	0.333333	0.0	0.333333	0.00	0.0	0.0
1	Akademija nauk	0.000000	0.5	0.000000	0.5	0.000000	0.00	0.0	0.0
2	Angarskaja	0.000000	0.0	0.750000	0.0	0.000000	0.25	0.0	0.0
3	Borovljany	1.000000	0.0	0.000000	0.0	0.000000	0.00	0.0	0.0
4	Brilevichi	0.000000	0.0	0.000000	0.0	0.000000	1.00	0.0	0.0

```
minsk_grouped.shape
```

```
(34, 9)
```

We have common venue categories in neighborhoods. In this reason I used unsupervised learning K-means algorithm to cluster them. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the boroughs into 5 clusters because when I analyze the K-Means with elbow method it ensured me the 5 degree for optimum k of the K-Means.

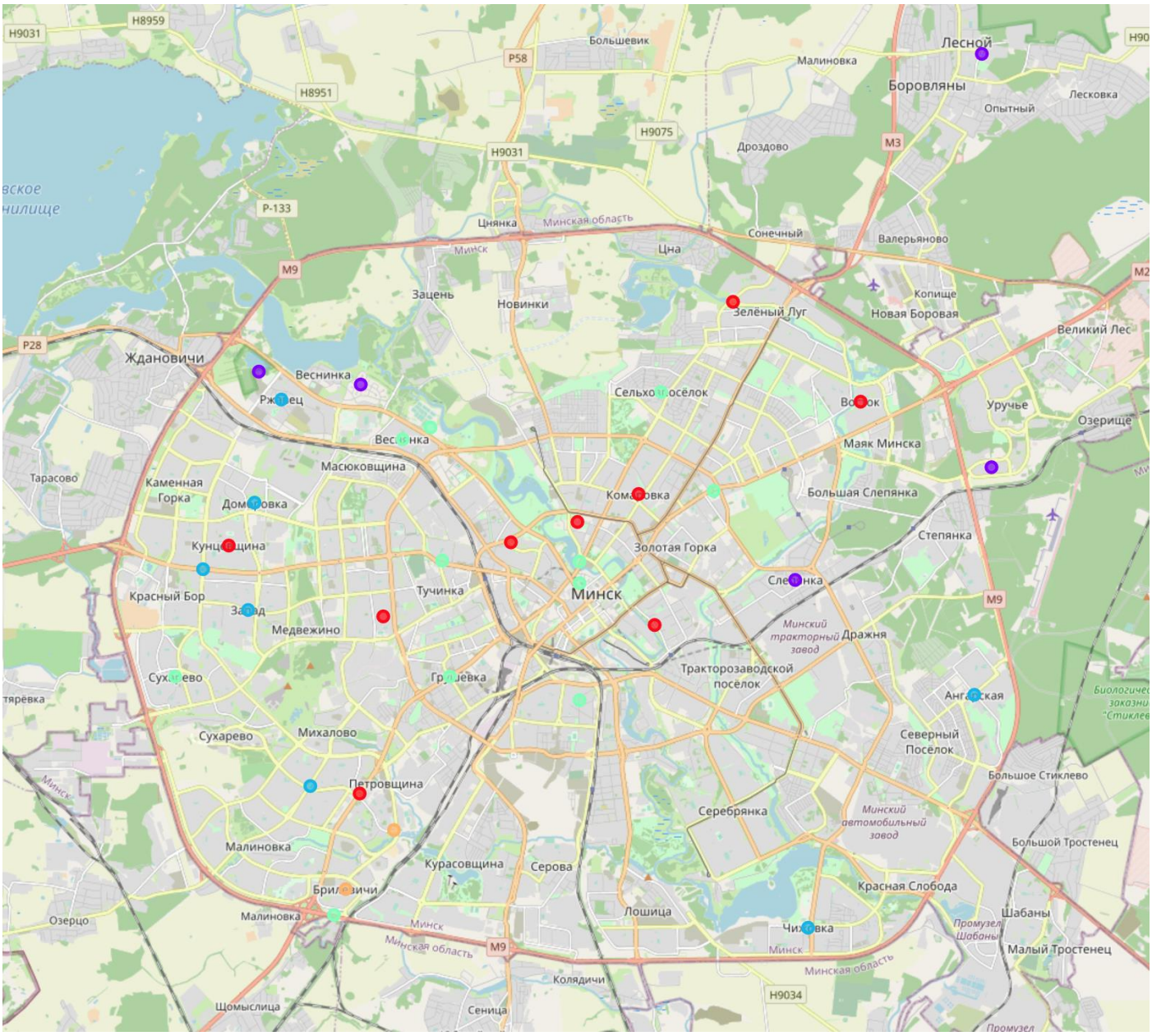


Here is my merged table with cluster labels for each neighborhood.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Frunzenskij	Suharevo	53.8867	27.4271	3	Yoga Studio	Gym / Fitness Center	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
1	Frunzenskij	Zapad	53.8993	27.4503	2	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
4	Frunzenskij	Kuntsevschina	53.9114	27.4441	0	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
5	Frunzenskij	Kamennaja Gorka	53.9069	27.4359	2	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
6	Frunzenskij	Dombrovka	53.9195	27.4526	2	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym

C. Results & Discussion

Now we can see a map with clustering.



We got a glimpse of the Healthy Lifestyle venues in Minsk and were able to find out some interesting insights which might be useful to the citizens as well as people with business interests.

We found that only **34 (53%)** from 64 neighborhoods have venues for healthy lifestyle. It means you need to think twice before choosing an area to stay.

The clustering is completely based on the most common venues obtained from Foursquare data.

However, in our analysis, we have ignored other factors like range of prices of houses/flats in Minsk, reviews of the venues and so on.

[32]:
```
minsk_merged.loc[minsk_merged['Cluster Labels'] == 0, minsk_merged.columns[[1] + list(range(5, minsk_merged.shape[1]))]]
```

[32]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
4	Kuntsevschina	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
12	Har'kovskaja	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
16	Petrovschina	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
36	Pulihova	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
43	Zaslavskaja	Gym	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool
47	Storozhovka	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
50	Zelenyj lug	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
51	Komarovka	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
58	Vostok	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center

[33]:
```
minsk_merged.loc[minsk_merged['Cluster Labels'] == 1, minsk_merged.columns[[1] + list(range(5, minsk_merged.shape[1]))]]
```

[33]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
33	Slepjanka	Gym Pool	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym / Fitness Center
45	Lebjazhij	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
46	poselok Raduzhnyj	Gym / Fitness Center	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool
54	Borovljany	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center
61	Akademgorodok	Gymnastics Gym	Gym / Fitness Center	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gym Pool

[34]:
```
minsk_merged.loc[minsk_merged['Cluster Labels'] == 2, minsk_merged.columns[[1] + list(range(5, minsk_merged.shape[1]))]]
```

[34]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
1	Zapad	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
5	Kamennaja Gorka	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
6	Dombrovka	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
21	Jugo-Zapad	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
27	Chizhovka	Gym / Fitness Center	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
29	Angarskaja	Gym / Fitness Center	Pool	Yoga Studio	Sports Club	Gymnastics Gym	Gym Pool	Gym
38	Rzhavets	Gym / Fitness Center	Gymnastics Gym	Yoga Studio	Sports Club	Pool	Gym Pool	Gym

[35]:
```
minsk_merged.loc[minsk_merged['Cluster Labels'] == 3, minsk_merged.columns[[1] + list(range(5, minsk_merged.shape[1]))]]
```

[35]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Suharevo	Yoga Studio	Gym / Fitness Center	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym
10	Kaskad	Gym / Fitness Center	Gym	Athletics & Sports	Yoga Studio	Sports Club	Pool	Gymnastics Gym
14	Grushevka	Gym / Fitness Center	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool
20	Filial BGU	Gym / Fitness Center	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool
25	Minsk-Siti	Sports Club	Gym	Athletics & Sports	Yoga Studio	Pool	Gymnastics Gym	Gym Pool
40	Vesnjanka	Gym / Fitness Center	Gym Pool	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym
41	Verhnij gorod	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool	Gym / Fitness Center	Gym
48	Troitskoe predmest'e	Yoga Studio	Gym / Fitness Center	Gym	Athletics & Sports	Sports Club	Pool	Gymnastics Gym
49	Tsnjanka	Gym / Fitness Center	Gym	Gym Pool	Yoga Studio	Sports Club	Pool	Gymnastics Gym
52	Sel'hozposelok	Gym / Fitness Center	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym Pool
62	Akademija nauk	Gym Pool	Gym	Yoga Studio	Sports Club	Pool	Gymnastics Gym	Gym / Fitness Center

[36]:
```
minsk_merged.loc[minsk_merged['Cluster Labels'] == 4, minsk_merged.columns[[1] + list(range(5, minsk_merged.shape[1]))]]
```

[36]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
18	Brilevichi	Pool	Yoga Studio	Sports Club	Gymnastics Gym	Gym Pool	Gym / Fitness Center	Gym
19	Semashko	Pool	Yoga Studio	Sports Club	Gymnastics Gym	Gym Pool	Gym / Fitness Center	Gym

D. Conclusion

The results can help traveler/citizen/investor to decide about the neighborhood that fit the most his needs.

I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major neighborhoods of Minsk and saw the results of segmentation of neighborhoods using Folium leaflet map.

Similarly, data can also be used to solve other problems, which most people face in metropolitan cities. Potential for this kind of analysis in a real-life problem is discussed in great detail. Also, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned.