BEST LOCATION FOR GYM ENTHUSIAST IN LONDON

CLUSTERING LONDON LOCATION WITH MOST NUMBER OF GYM IN SURROUNDING AREA

OUTLINE

1. INTRODUCTION

2. DATA ACQUISITION AND CLEANING

3. EXPLORATORY DATA ANALYSIS (METHODOLOGY & RESULTS)

4. CONCLUSIONS

5. DISCUSSION AND RECOMMENDATION FOR FURTHER STUDY

1. INTRODUCTION

Body Building is the process of putting on muscle through working out and dedicated yourself to a steady routine. There are two main factors of successful bodybuilding: workout consistently and consume good nutrition. As we can see, people who want to begin doing body building have to do workout routine to get the body they want. As well as people who has been on body building for a long time have to maintain their workout routine to stay in shape.

PROBLEM

To maintain their routine, instead of doing workout in their own home, it is likely to be more effective for them to do their work out in the gym / fitness center. If they go to gym/fitness center, not only they can get the best equipment they can use, but they can also meet a personal trainer or other bodybuilder to help their bodybuilding progress. Unfortunately, they can be reluctant to go to the gym regularly (especially for beginner) because of:

a) The gym is too far away, and

b) There are very little options of another gym, so if they do not like some gyms near their home, they does not have other options as alternative.

IDEAS

We are going to help the body building enthusiast, the beginners, and the experts, to choose the best place for them to live in a city, in this case is London Area. We will find the areas in London where have many options of gym around its small surrounding area. The methodology is as follow:

a. Retrieve nearby gym venues on each London Borough to be analyzed

b. Cluster gym venues using Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

c. Analyze the clusters whether it is already show the most dense gym venues in small area

d. If clusters is too large, find the subcluster to obtain the most suitable area which surrounded by many gym venues

2. DATA ACQUISITION AND DATA CLEANING

Data to be retrieved
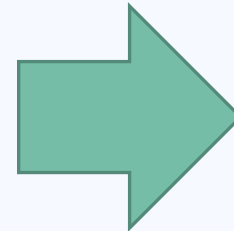
List of London Borough

- Retrieved from wikipedia
- Will be used for searching the coordinate of borough

Coordinate of each Borough

- Retrieved with Geopy Python library
- This Data will be used for searching nearby gym venues around the borough area

Gym Venues

- This data can be collected from Foursquare API
- Gym venues data will be clustered based on its location

DATA CLEANING AND PREPROCESSING

CLEAN GYM VENUES DATASET

	Venue	Venue Latitude	Venue Longitude	Venue Category	Borough
0	PureGym	51.539250	-0.143077	Gym / Fitness Center	Camden
1	Barry's Bootcamp	51.527075	-0.131056	Gym / Fitness Center	Camden
2	PureGym	51.554052	-0.144984	Gym / Fitness Center	Havering
3	Urban Kings	51.531300	-0.121950	Gym / Fitness Center	Camden
4	Somers Town Community Sports Centre	51.532768	-0.133157	Gym / Fitness Center	Camden
...	...	...	...	...	...
536	Blitz CrossFit	51.448928	-0.332068	Gym	Richmond upon Thames
537	Go-Gym	51.360355	-0.195039	Gym	Sutton
538	Leyton Leisure Centre	51.573975	-0.010304	Gym	Waltham Forest
539	Shoreditch House Gym	51.523687	-0.076177	Gym	City of London
540	Montcalm Gym	51.520826	-0.091617	Gym	City of London

541 rows × 5 columns

MAP OF GYM VENUES
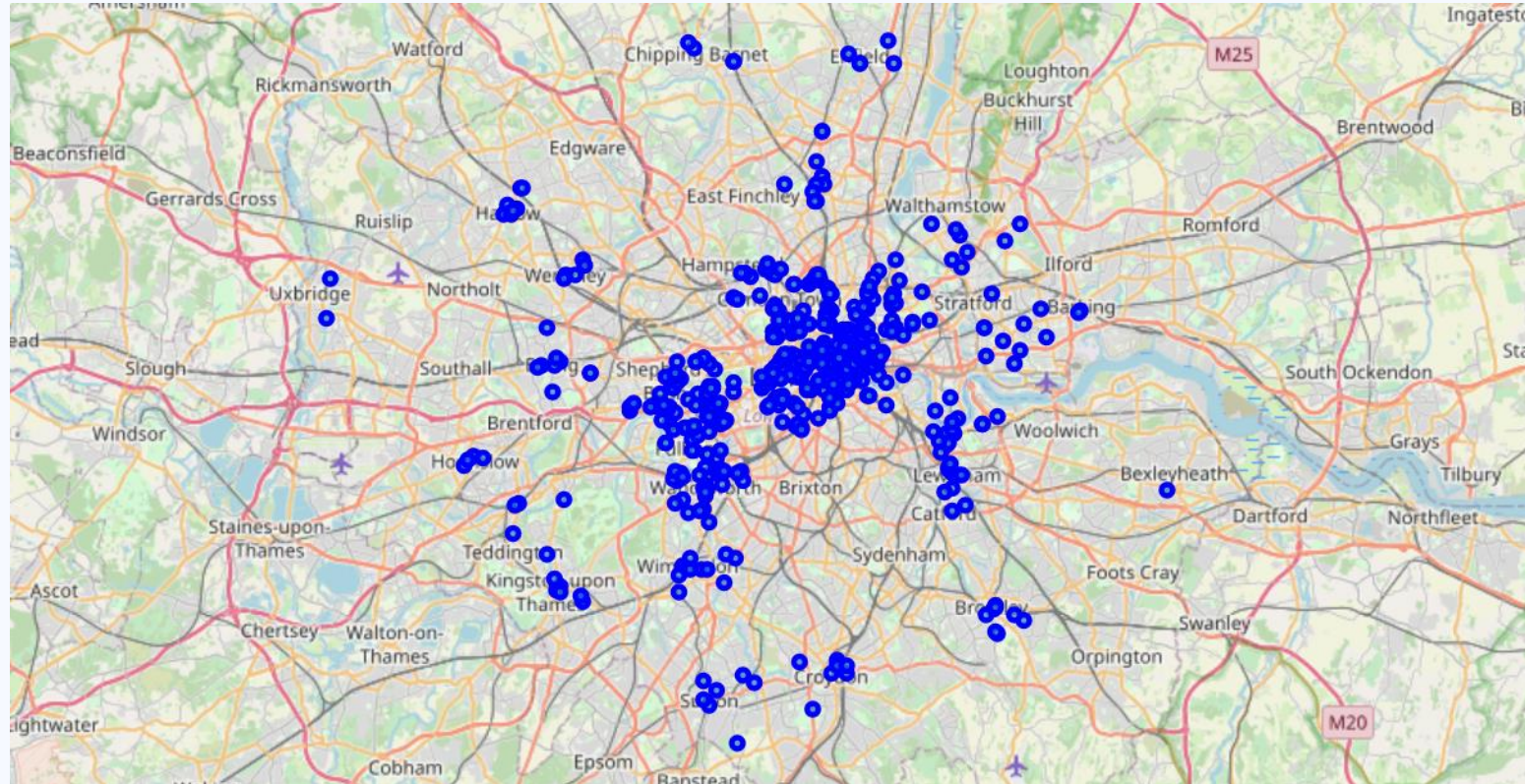
# 3. EXPLORATORY DATA ANALYSIS (METHODOLOGY & RESULTS)

| DBSCAN Clustering	Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Method
	Unsupervised learning methods that identify distinctive groups/clusters in the data, based on the idea that a cluster in data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density
	This method could find the area of London which has the high density of gym venues

CLUSTERING LEVEL 1
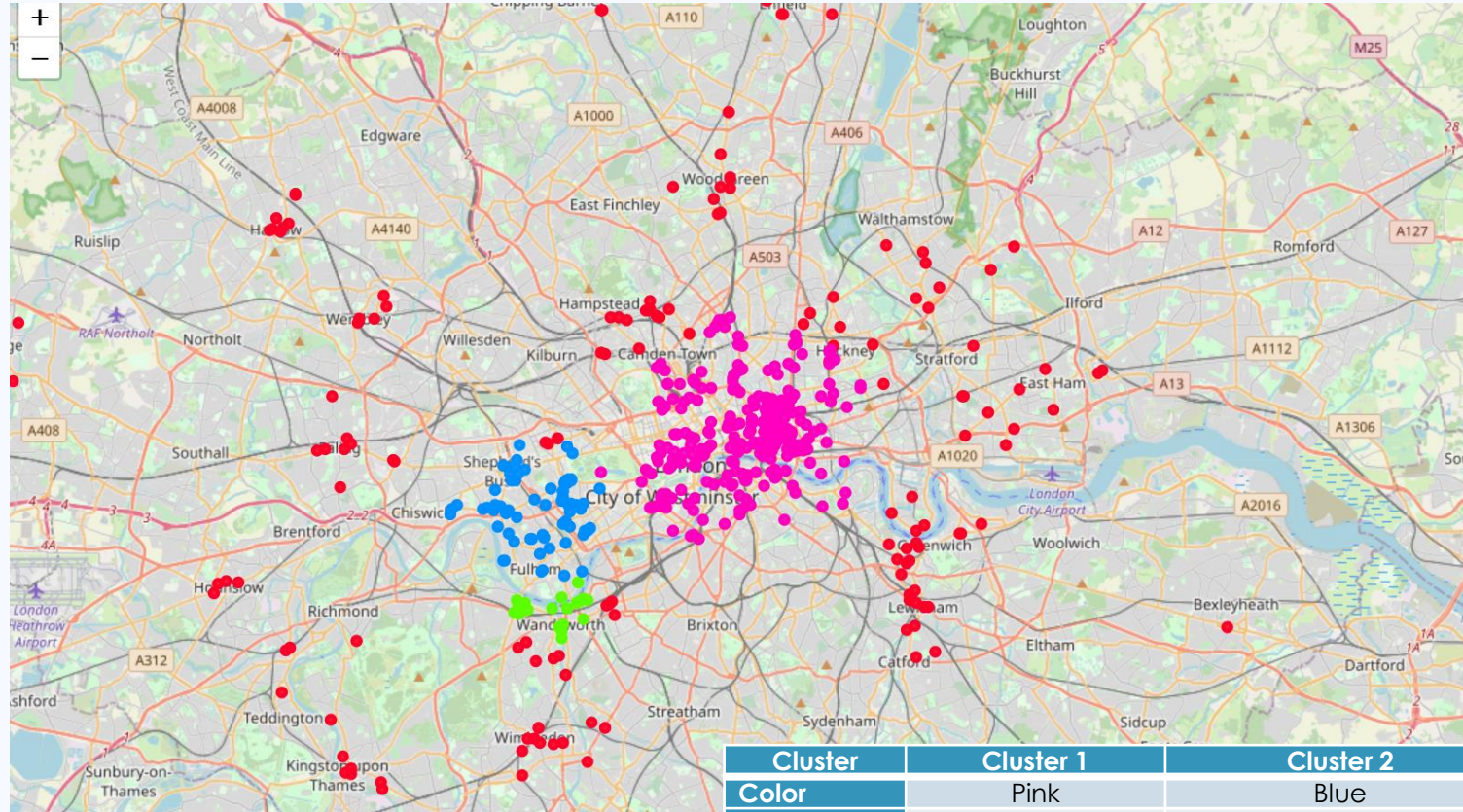
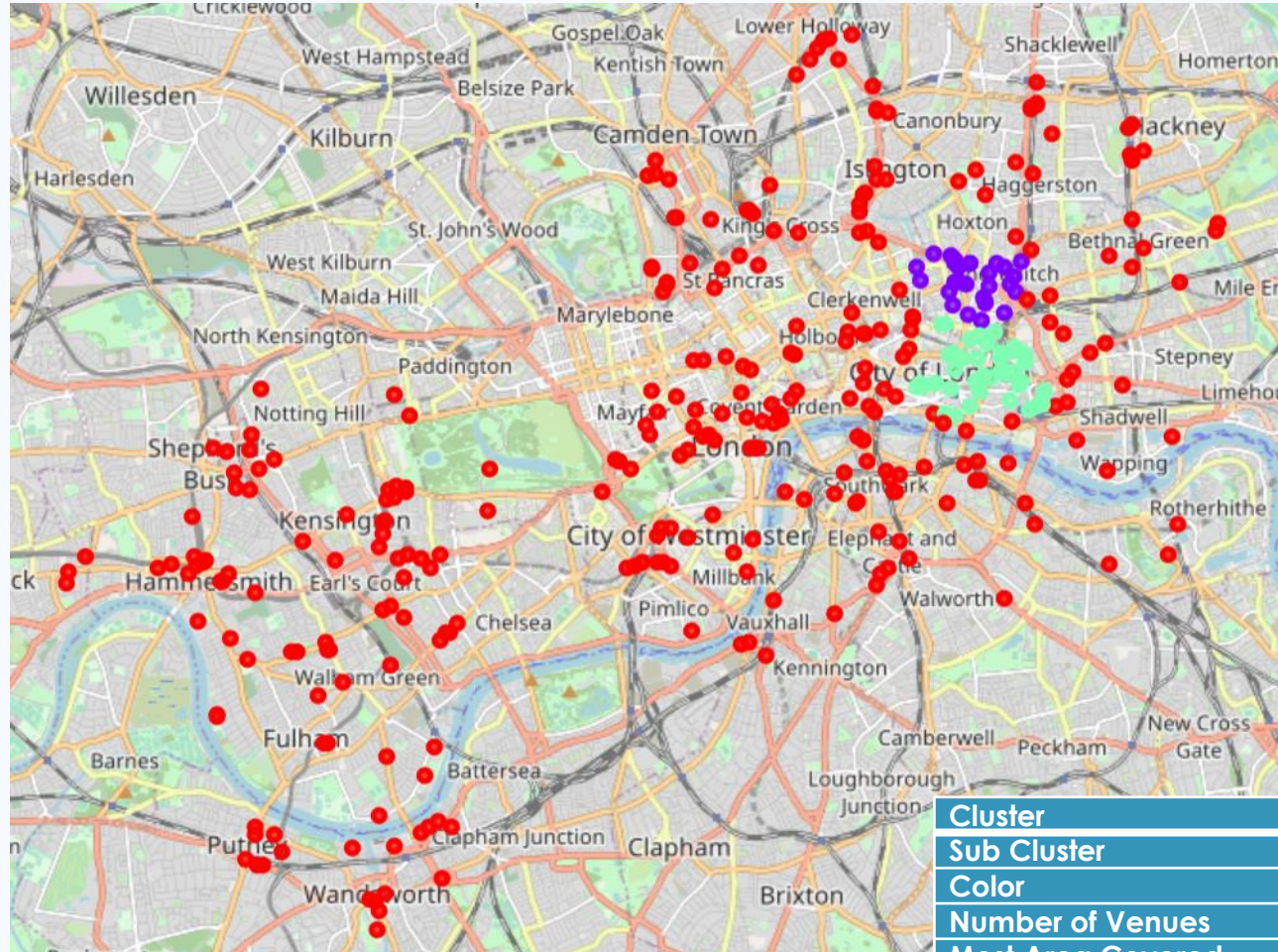| Cluster	Cluster 1	Cluster 2	Cluster 3	No Cluster (Outlier)
Color	Pink	Blue	Green	Red
Number of Venues	273	73	21	174
Most Area Covered	City of London, Westminster, Tower Hamlets, Islington	Hammersmith and Kensington and Chelsea	Wandsworth	Outer

ANALYSIS

The clustering result seems successfully separate groups of gym locations from others.

But the resulting dataset could not specify the little neighborhood area of the most densely populated area by gym venues in London. So, we should deploy next clustering method to find the subcluster in the clusters to identify the dense narrow area of gym venues.
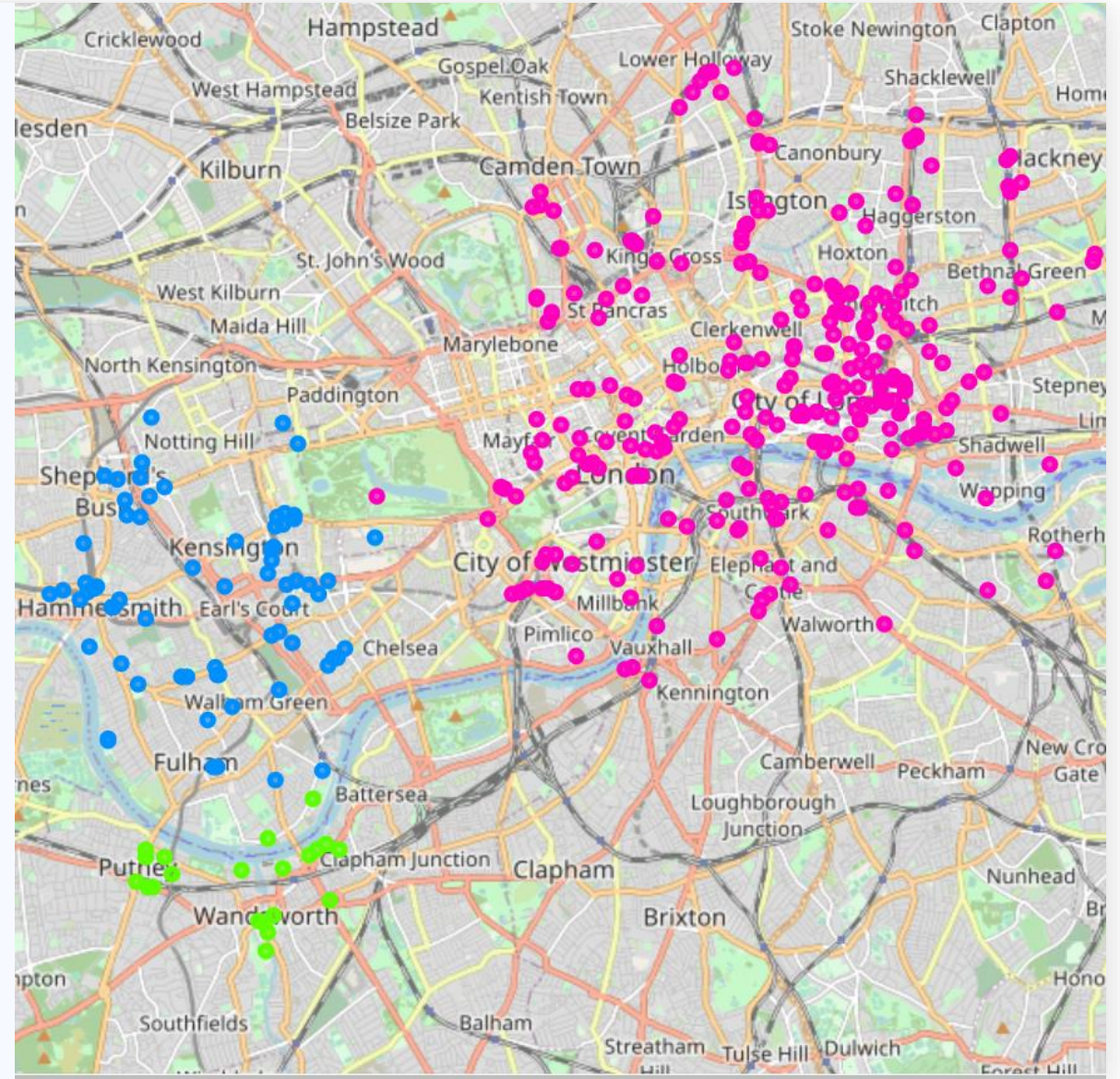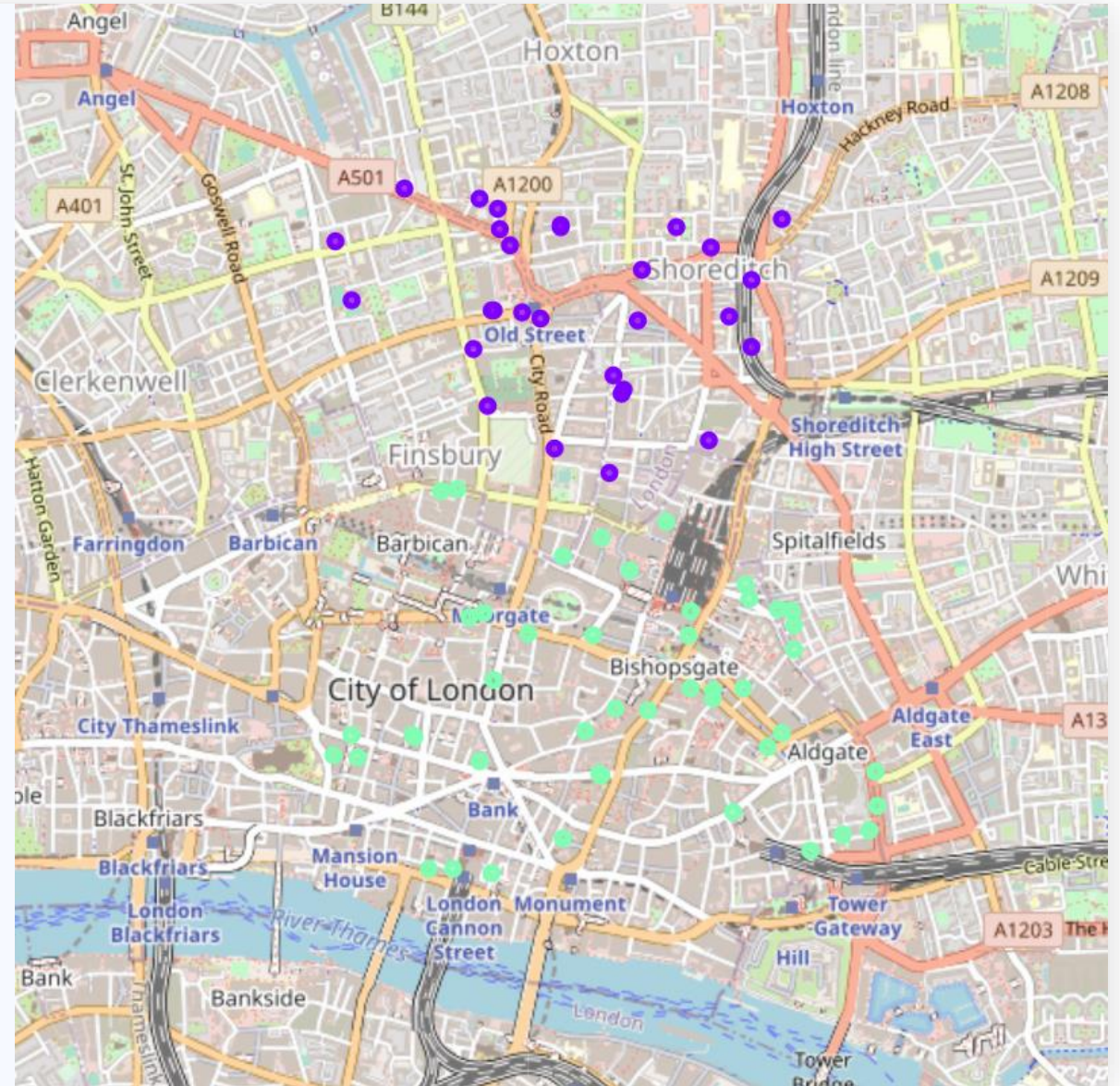
CLUSTERING LEVEL 2

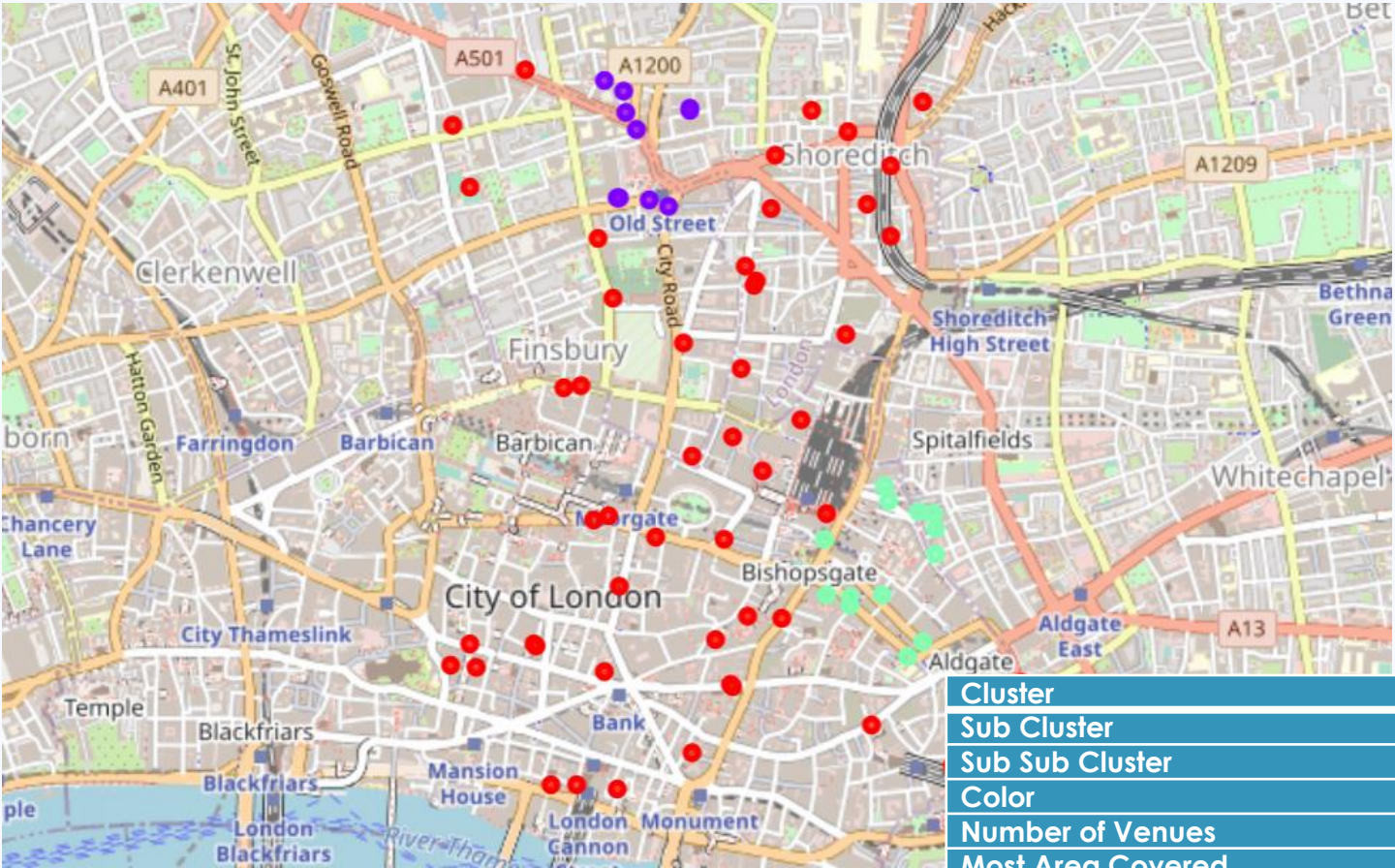| Cluster	Cluster 1	Cluster 1	Cluster 1,2,3
Sub Cluster	Sub Cluster 1	Sub Cluster 2	No Sub Cluster
Color	Purple	Green	Red
Number of Venues	29	46	292
Most Area Covered	City of London	Tower Hamlets	Other

ANALYSIS

This level 2 clustering method could obtain more narrow clusters compared to the first clustering.

Although this subclustering result are already show the more densely gym venues in narrower area, we could try to do the same clustering method of this subclusters, which means the level 3 clustering of gym venues to find whether there are some specific location that is more dense compared by other location in those two subclusters.
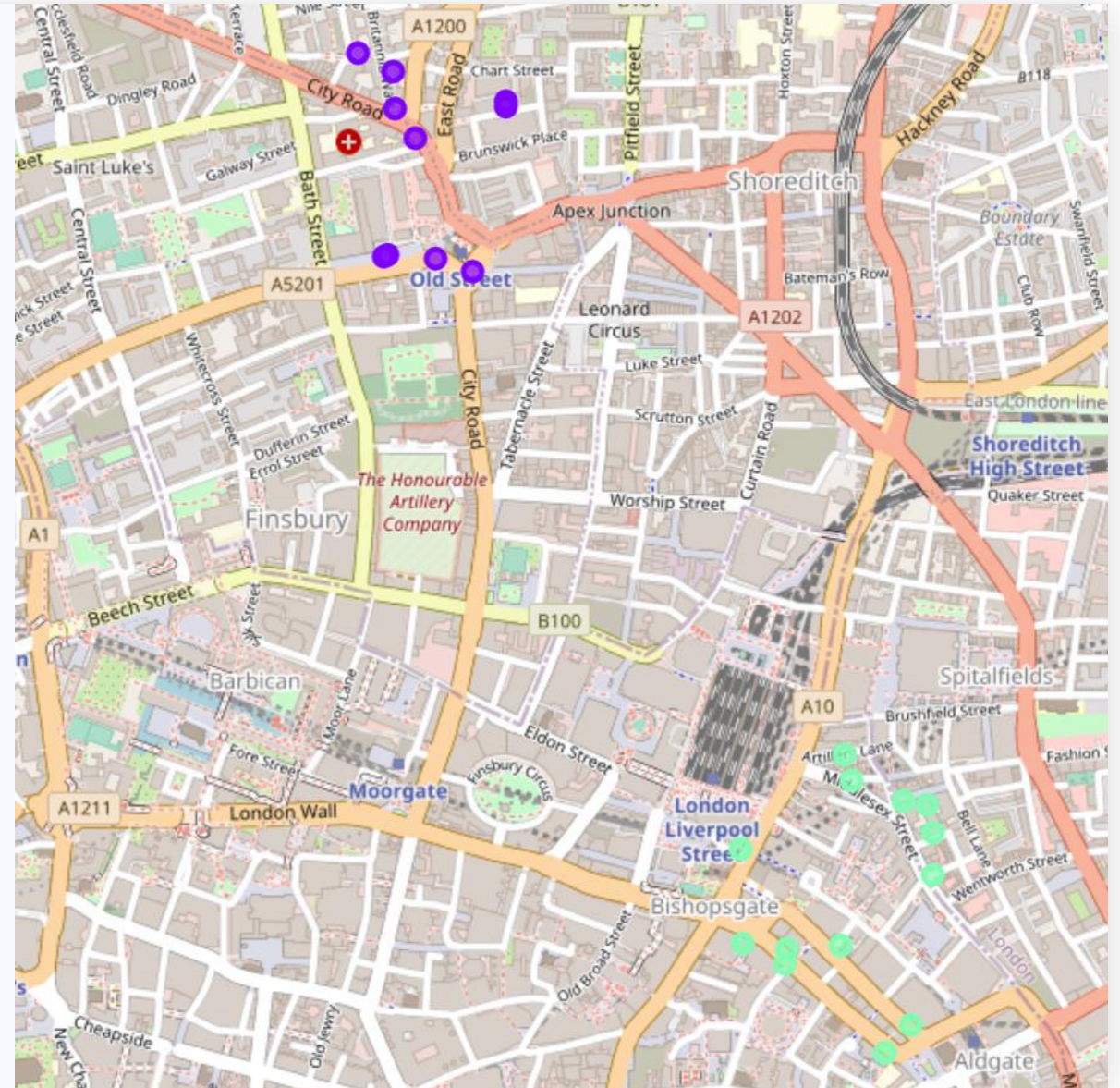
CLUSTERING LEVEL 3

Cluster	Cluster 1	Cluster 1
Sub Cluster	Sub Cluster 1	Sub Cluster 2
Sub Sub Cluster	Sub Sub Cluster 1	Sub Sub Cluster 2
Color	Purple	Green
Number of Venues	10	13
Most Area Covered	City of London around Old Street Station	Around Bishopgate, between City of London and Tower Hamlets

ANALYSIS

This level 3 clustering method has produced satisfying outcome to get the most densely populated area by gym venues in London.

As we expected, this level 3 clustering method could obtain more narrow clusters compared to the second clustering. Each Sub Sub Cluster represents the most densely populated by gym venues area of each Sub Cluster.

4. CONCLUSION

Clustering	Output						
Clustering Level 1	Cluster 1				Cluster 2	Cluster 2	Outlier
Clustering Level 2	Sub Cluster 1		Sub Cluster 2	Outlier	Outlier	Outlier	-
Clustering Level 3	Sub Sub Cluster 1	Outlier	Sub Sub Cluster 2	Outlier	-	-	-	-

DBSCAN Clustering Method has been successfully extract some clusters of areas in London based on gym venue's location data. In order to get the specific location on the most densely populated area by gym venues, we deploy 3 levels of DBSCAN Clustering method. The higher level of DBSCAN Clustering, is conducted to find the subclusters from retrieved clusters of previous clustering method.

RECOMMENDED AREA FOR BODY BUILDERS #1

Bishopgate, between City of London and Tower Hamlets *(Sub Sub Cluster 2)*

This is the most recommended area for body builders to live in London. The area consists of 13 gym venues on its surrounding. Which means, that if the body builders got bored or not liking one of the nearby gym venues, there are still 12 other gym venues to try.
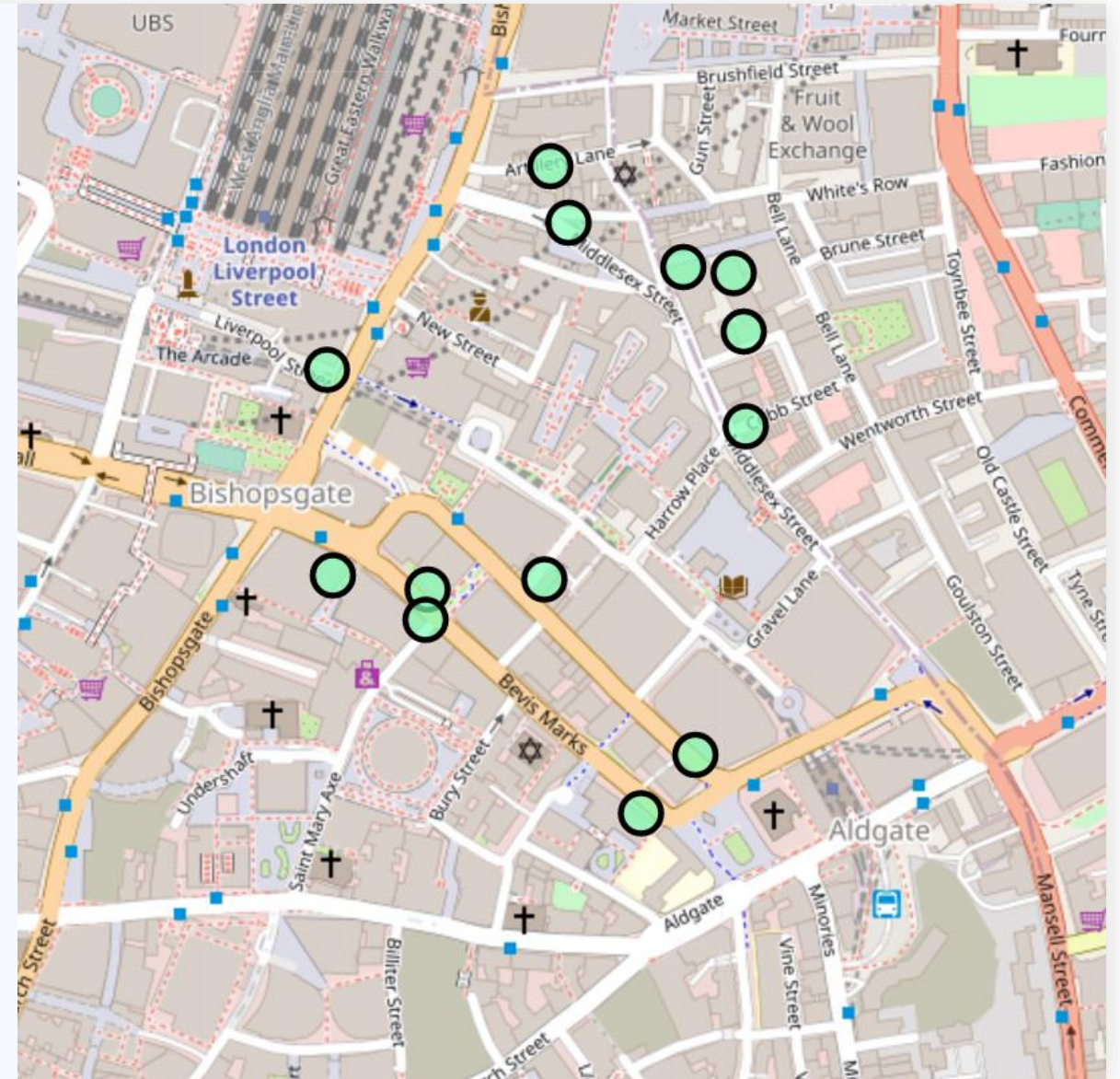
RECOMMENDED AREA FOR BODY BUILDERS #2

City of London, around Old Street Station (Sub Sub Cluster 1)

This area consists of 10 gym venues on its surrounding. Which means, that if the body builders got bored or not liking one of the nearby gym venues, there are still 9 other gym venues as other alternative. This is the second recommended area for body builders to live in London in this area.
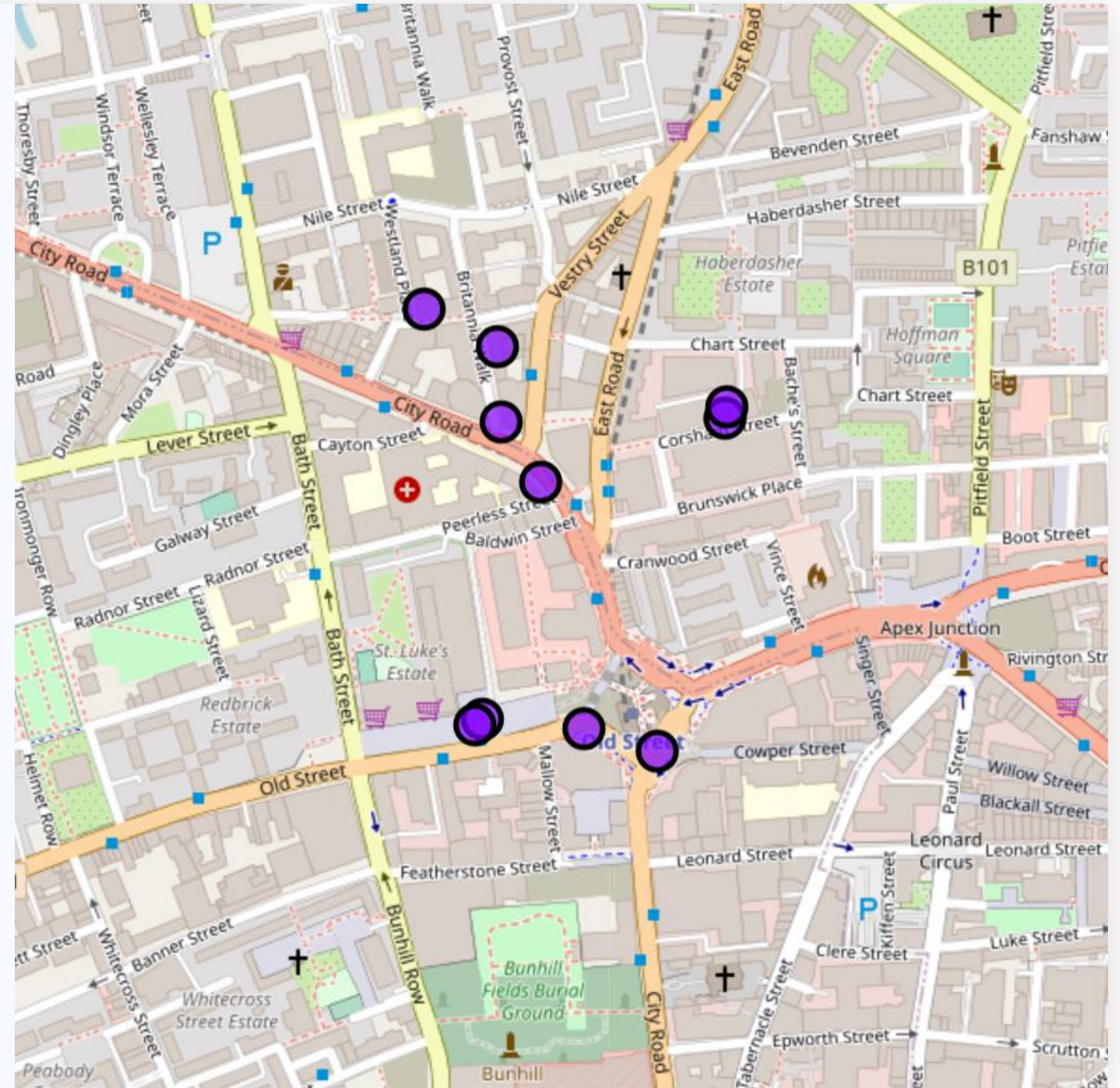
5. DISCUSSION AND RECOMMENDATION FOR FURTHER STUDY

The 3 levels of clustering method could obtain the specific location of the most densely populated area by gym venues in London. This result is satisfying enough as supporting data in order to recommend someone who has passion of body building finding the best location for him to live. However, this study is not including some of other component that could become consideration of finding place to live, such as housing price, public transportation access, etc. We hope that if someone want to conduct this kind of study in higher level which including more variables, those data (housing price, public transportation data, etc) could be included to the study in order to get more satisfying result.

THANK YOU

☺

References:

- IBM Professional Data Science Course
- https://bodyspartan.com/blogs/all-articles/what-is-bodybuilding
- https://en.wikipedia.org/wiki/London_boroughs
- https://www.kdnuggets.com/2020/04/dbscan-clustering-algorithm-machine-learning.html