

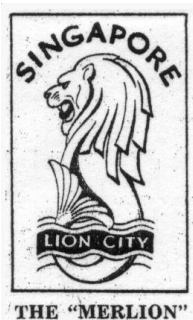
# Plan Model for Decathlon Business Expansion in Singapore

## (Market Research Analysis)

IBM Data Science Professional Certificate

Applied Data Science Capstone Project : “The Battle of Neighborhoods”

(February 2019)



Nishanth Priyandas V.N.

[nishanth.pvn@live.com](mailto:nishanth.pvn@live.com)

# Agenda

- ▶ Introduction
- ▶ Data
- ▶ Methodology
- ▶ Results
- ▶ Discussion
- ▶ Conclusion

# Introduction

# Introduction

- ▶ **Decathlon** is a French sporting goods retailer. With over 1500 stores in 49 countries, it is the largest sporting goods retailer in the world. It stocks a wide range of sporting goods for tennis rackets to advanced scuba diving equipment. Decathlon Group also owns over 20 brands with research and development facilities to develop the latest innovative designs registering up to 40 patents per year. Each brand represents a different sport or group of sports, with a dedicated product development and design team.
- ▶ In recent global management discussion, it was decided to expand their presence in South East Asia especially Singapore. Plans are underway to open more retail outlets all over island to improve business. In addition, it becomes necessity to understand each locality in Singapore better based on top outdoor/recreational activities available in surrounding areas so that planned new outlets could sell right product in right places.
- ▶ Marketing Consultants has determined that special attention needs to be given for top outdoor/recreational activities available in specific areas so that they can plan on products/brands which can be given importance in those areas as well as advertise them accordingly to have good reach among public.
- ▶ Our project will obtain information about neighborhoods in Singapore and make recommendations for Decathlon management to understand Singapore locality better for expanding their business.

Data

# Data

- ▶ Singapore regions information can be obtained from **Wikipedia**: [Regions of Singapore](#)
  - ▶ In order to make recommendations on suggested product brands in new retail outlets to be opened, Singapore segmentation will be made based on different neighborhoods.
  - ▶ Full list of neighborhoods can be obtained from Wikipedia under Regions of Singapore, but only their names. They must be geolocated in order to use Foursquare services for obtaining venues.
- ▶ For geolocation of neighborhoods **Python geocoder** will be used.
  - ▶ Geocoder returns latitude and longitude information for every neighborhood center, then it will be used as main Foursquare API input.
- ▶ In order to obtain top attractions/recreational facilities in each locality we will use [FOURSQUARE](#) API.
  - ▶ Using services provided by Foursquare we can obtain top outdoor / recreational facilities for every neighborhood. Such services requires input geolocalization (i.e.) latitude and longitude which we obtained in previous step.

# Methodology

# Methodology

## ► Build Neighborhoods List

A list of localities in Singapore based on regions is obtained from Wikipedia page. That list contains the names of the locality in every region.

As output a dataset containing a list of "*region,locality*" is build.

Example:

	neighborhood	latitude	longitude
0	Central,Bishan	NaN	NaN
1	Central,Bukit Merah	NaN	NaN
2	Central,Bukit Timah	NaN	NaN
3	Central,Downtown Core	NaN	NaN
4	Central,Geylang	NaN	NaN
5	Central,Kallang	NaN	NaN



# Methodology

## ► Neighborhoods Geolocation

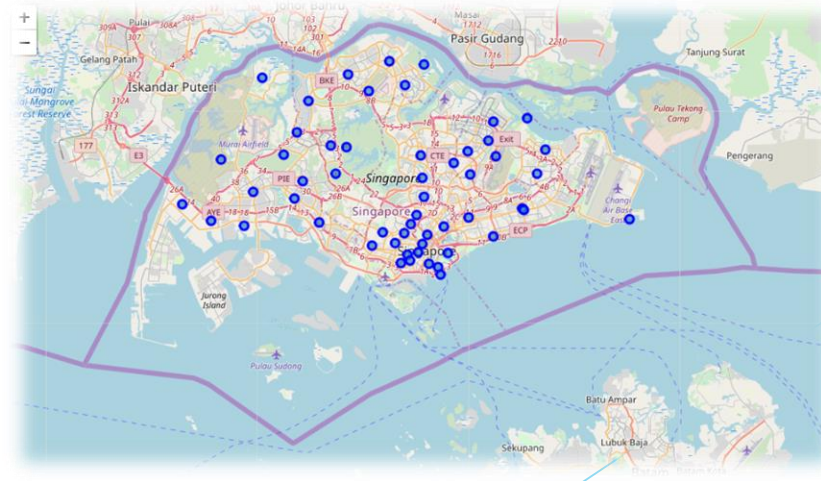
Every element in the neighborhoods dataset is geolocated using *Python Geolocator* and two new columns containing latitude and longitude were added for each locality.

Since geolocator service has frequent connection issues resulting in time out error, so to handle this information obtained is saved as a text file in CSV format.

Therefore this step can be run many times, invoking geolocation only for missing data (timed out errors in previous executions). After various executions all the neighborhoods geolocation is obtained and it's used in text file.

### Example:

	neighborhood	latitude	longitude
0	Central,Bishan	1.35146	103.848
1	Central,Bukit Merah	1.28063	103.831
2	Central,Bukit Timah	1.35469	103.776
3	Central,Downtown Core	1.28712	103.836
4	Central,Geylang	1.31819	103.887
5	Central,Kallang	1.31076	103.866



# Methodology

## ► Venues Compilation (Using Foursquare API)

As next step Foursquare API services are used for obtaining venues for every neighborhood locality. The output is a new dataset with top three outdoor/recreational place in every neighborhood locality.

Foursquare Developer (Personal Account) with 99,500 regular calls/per day is used to extract required information. In order to minimize the usage of Foursquare, the information which have been extracted is saved in a text (CSV) file. Here it's assumed that information gathered doesn't change in short period of time. When the analysis needs to be re-done after long while then it's suggested to delete existing file's and regenerate them by calling Foursquare API service

### Example:

	neighborhood	neighborhood_latitude	neighborhood_longitude	venue	venue_latitude	venue_longitude	venue_category
0	Central,Bishan	1.3514551	103.8482628	Gymm Boxx XL	1.3499089544875253	103.85068897477275	Gym
1	Central,Bishan	1.3514551	103.8482628	Bishan Swimming Complex	1.354225078025654	103.85044843987507	Pool
2	Central,Bishan	1.3514551	103.8482628	Bishan Active Park	1.3548245096703175	103.84768419495201	Park
3	Central,Bukit Merah	1.2806275	103.8305915	Yoga Movement - The Pod	1.2826514716467718	103.83029451970962	Yoga Studio
4	Central,Bukit Merah	1.2806275	103.8305915	YOGA INC.	1.2828363	103.8302	Yoga Studio

# Methodology

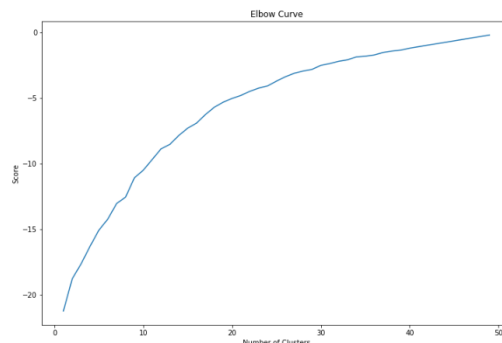
## ► Neighborhood Segmentation

The business case in hand falls under unsupervised machine learning approach so K-means clustering algorithm was chosen for analysis.

Taking in consideration that the venues information obtained from Foursquare is categorical, it must be previously processed in order to be handled by K-means algorithm. For this "pandas.getdummies" is used for dummies variables.

The list of dummy variables obtained is then grouped as features of every neighborhood locality.

- After executing K-means algorithm the "Elbow Curve" it's plotted in order to obtain the *best K*. Analyzing the change in the slope of the curve, it's determined that K=15 is a good value.
- K-means algorithm is then executed.
- In next step segmented data-frame is built which is composed of the top three venues for every neighborhood locality plus a segment label which is determined by K-means algorithm.



	neighborhood	1st	2nd	3rd
0	Central,Bishan	Pool	Gym	Park
1	Central,Bukit Merah	Yoga Studio	Athletics & Sports	Trail
2	Central,Bukit Timah	Mountain	Nature Preserve	Scenic Lookout
3	Central,Downtown Core	Yoga Studio	Pedestrian Plaza	Fishing Spot
4	Central,Geylang	Basketball Court	Soccer Field	Yoga Studio
5	Central,Kallang	Pool	Park	Yoga Studio

# Methodology

## ► Segment Cluster Analysis

Every segment is listed individually and it's further analyzed to derive meaningful insight as described in next section.

Example :

### Cluster 1

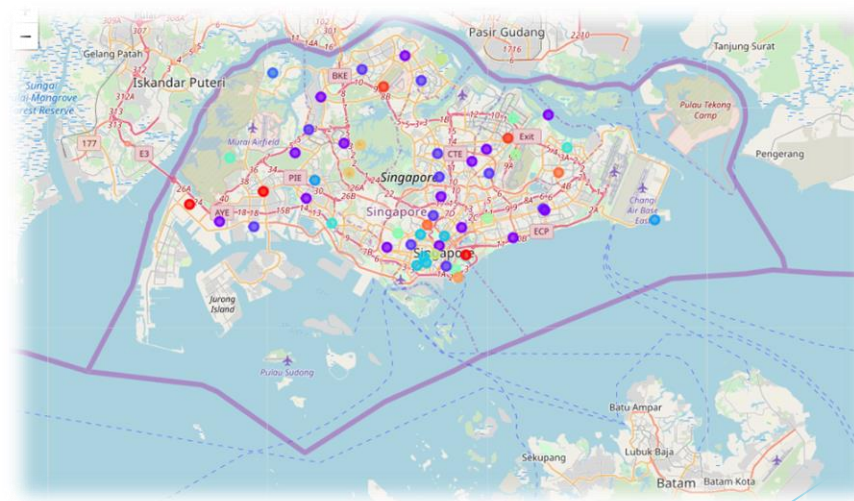
```
[24]: print_cluster_info(0)
```

	neighborhood	latitude	longitude	1st	2nd	3rd
6	Central, Marina East	1.28862	103.87	Golf Course	Park	Yoga Studio
50	West, Jurong West	1.33964	103.707	Gym / Fitness Center	Park	Playground
52	West, Tuas	1.32937	103.648	Golf Course	Gym / Fitness Center	Yoga Studio

### Cluster 6

```
[29]: print_cluster_info(5)
```

	neighborhood	latitude	longitude	1st	2nd	3rd
1	Central, Bukit Merah	1.28063	103.831	Yoga Studio	Athletics & Sports	Trail
3	Central, Downtown Core	1.28712	103.836	Yoga Studio	Pedestrian Plaza	Fishing Spot
12	Central, Orchard	1.30527	103.833	Yoga Studio	Gym / Fitness Center	Scenic Lookout
13	Central, Outram	1.28287	103.838	Yoga Studio	Sports Club	Farm
16	Central, Rochor	1.30398	103.853	Yoga Studio	Gym / Fitness Center	Playground



# Results

# Results

List of sports-product categories currently sold by Decathlon worldwide.

- ▶ **CYCLING** (City Bikes, Hybrid Bikes, Kids Bikes, Mountain Bikes, Road Bikes, Triathlon)
- ▶ **FITNESS SPORTS** (Bodybuilding, Cross Training, Dance, Fitness Cardio, Gymnastics, Pilates, Yoga)
- ▶ **RUNNING SPORTS** (Athletics, Fitness & Nordic Walking, Running, Trail Running, Triathlon)
- ▶ **URBAN SPORTS** (Inline Skates, Scooters, Skateboards)
- ▶ **TEAM SPORTS** (Baseball, Basketball, Cricket, Floorball, Football, Handball, Rugby, US Football, Volleyball & Beach Volley)
- ▶ **OUTDOOR SPORTS** (Camping, Canyoning, Climbing & Mountaineering, Fishing, Hiking & Trekking, Horse Riding, Ski & Snowboard, Wildlife Exploration)
- ▶ **WATER SPORTS** (Aqua Aerobics, Kayak, Kites, Sailing, Scuba Diving, Snorkeling, Stand-Up Paddle (SUP), Surf & Body board, Swimming, Water Sport Games)
- ▶ **RACKET SPORTS** (Badminton, Squash, Table Tennis, Tennis, Turn ball)
- ▶ **TARGET SPORTS** (Archery, Billiards, Darts, Golf, Petanque)
- ▶ **COMBAT SPORTS** (Boxing, Judo, Karate, Taekwondo)

Suggested sports-products (**Top-3, in-order of importance**) which can be focused at specific localities in Singapore based on segmentation result is described in next slides.



1. TARGET SPORTS
2. FITNESS SPORTS
3. RUNNING SPORTS

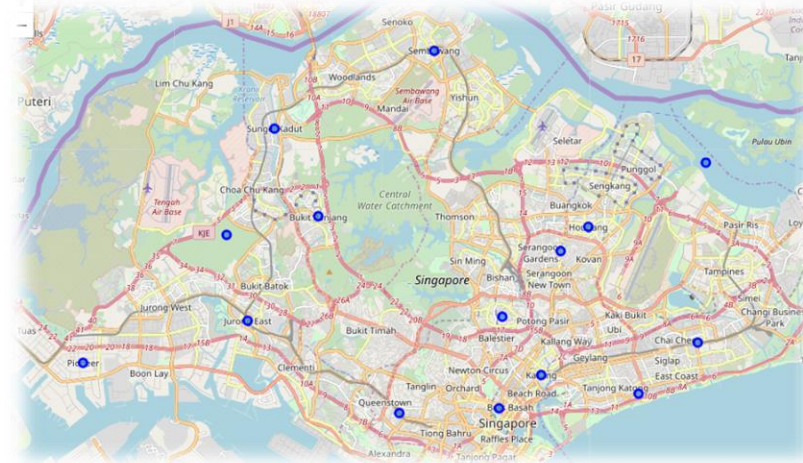


# Results

## Cluster 2

1. WATER SPORTS
2. RACKET SPORTS
3. OUTDOOR SPORTS

	neighborhood	latitude	longitude	1st	2nd	3rd
5	Central,Kallang	1.31076	103.866	Pool	Park	Yoga Studio
8	Central,Marine Parade	1.30269	103.907	Athletics & Sports	Playground	Waterfront
9	Central,Museum	1.29655	103.848	Yoga Studio	Park	Plaza
14	Central,Queenstown	1.29462	103.806	Pool	Skate Park	Park
21	Central,Toa Payoh	1.33539	103.85	Pool	Playground	Yoga Studio
23	East,Changi	1.32435	103.933	Pool	Gym / Fitness Center	Plaza
31	North,Sembawang	1.44806	103.821	Athletics & Sports	Park	Plaza
33	North,Sungei Kadut	1.415	103.753	Trail	Recreation Center	Playground
37	North-East,Hougang	1.37336	103.886	Athletics & Sports	Pool	Park
38	North-East,North-Eastern Islands	1.40078	103.936	Fishing Spot	Beach	Campground
42	North-East,Serangoon	1.36324	103.874	Soccer Field	Park	Playground
44	West,Bukit Panjang	1.37792	103.772	Park	Gym	Yoga Studio
46	West,Pioneer	1.3158	103.672	Harbor / Marina	National Park	Golf Course
49	West,Jurong East	1.3338	103.742	Trail	Gym / Fitness Center	Skating Rink
51	West,Tengah	1.37024	103.733	Pool	Basketball Court	Park

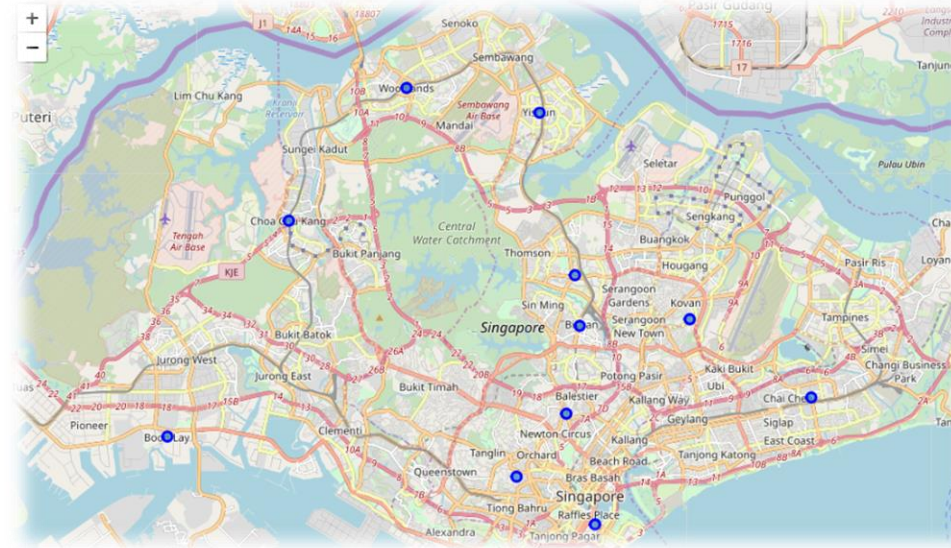




# Results

## Cluster 3

1. FITNESS SPORTS
2. TEAM SPORTS
3. URBAN SPORTS

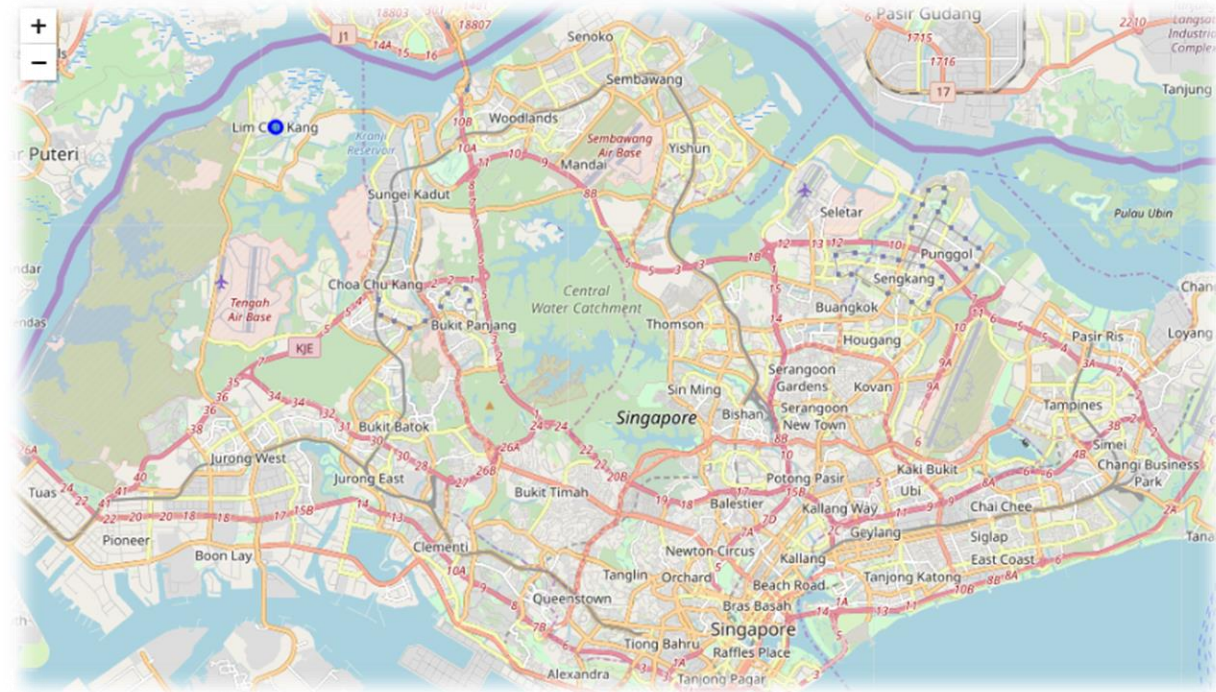


	neighborhood	latitude	longitude	1st	2nd	3rd
0	Central,Bishan	1.35146	103.848	Pool	Gym	Park
11	Central,Novena	1.32008	103.843	Gym / Fitness Center	Gym	Park
15	Central,River Valley	1.29728	103.826	Gym	Park	Canal
19	Central,Straits View	1.28006	103.854	Waterfront	Harbor / Marina	Gym
22	East,Bedok	1.32567	103.931	Pool	Gym	Plaza
25	East,Paya Lebar	1.3539	103.888	Gym / Fitness Center	Soccer Field	Bowling Green
34	North,Woodlands	1.4369	103.786	Park	Gym	Yoga Studio
35	North,Yishun	1.42814	103.834	Gym	Park	Hot Spring
36	North-East,Ang Mo Kio	1.36984	103.847	Gym	Park	Dog Run
45	West,Boon Lay	1.31156	103.7	Harbor / Marina	Scenic Lookout	Farm
47	West,Choa Chu Kang	1.38926	103.744	Pool	Park	Playground

# Results

## Cluster 4

1. OUTDOOR SPORTS
2. WATER SPORTS
3. FITNESS SPORTS



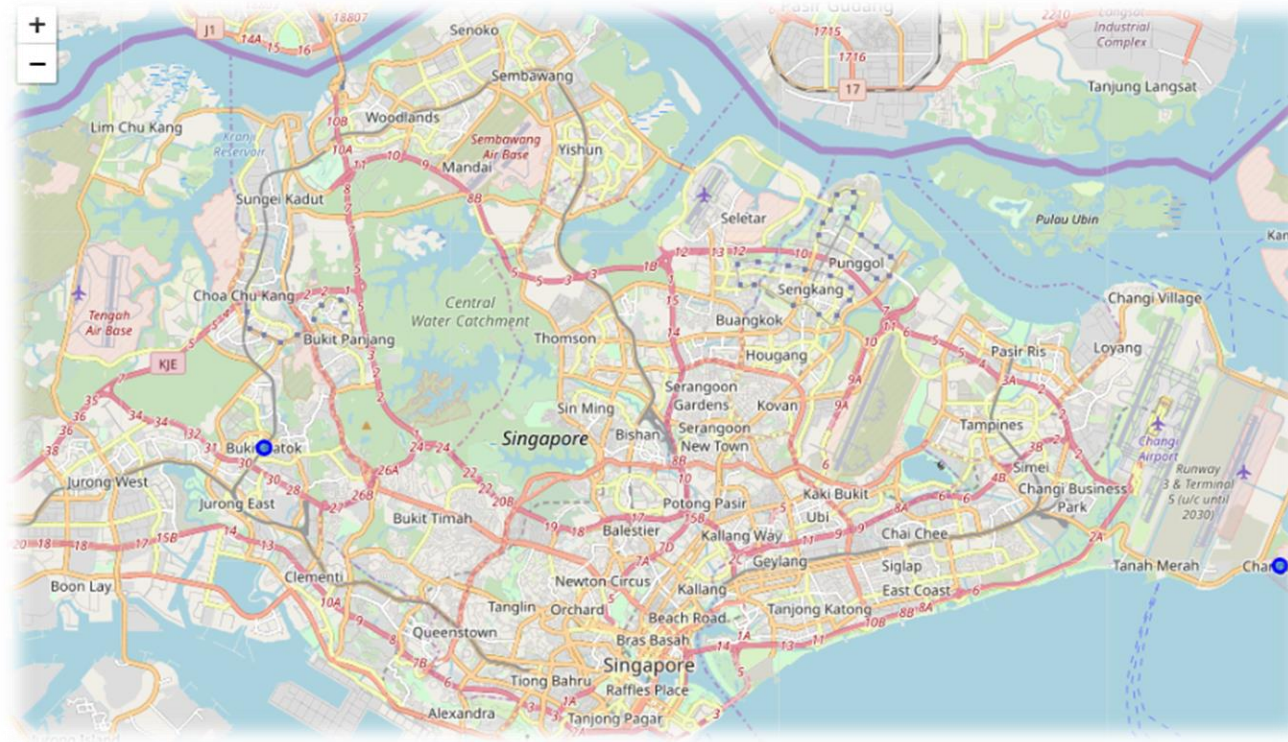
	neighborhood	latitude	longitude	1st	2nd	3rd
29	North,Lim Chu Kang	1.43422	103.715	Farm	Hot Spring	Gym Pool



# Results

## Cluster 5

1. WATER SPORTS
2. TEAM SPORTS
3. FITNESS SPORTS

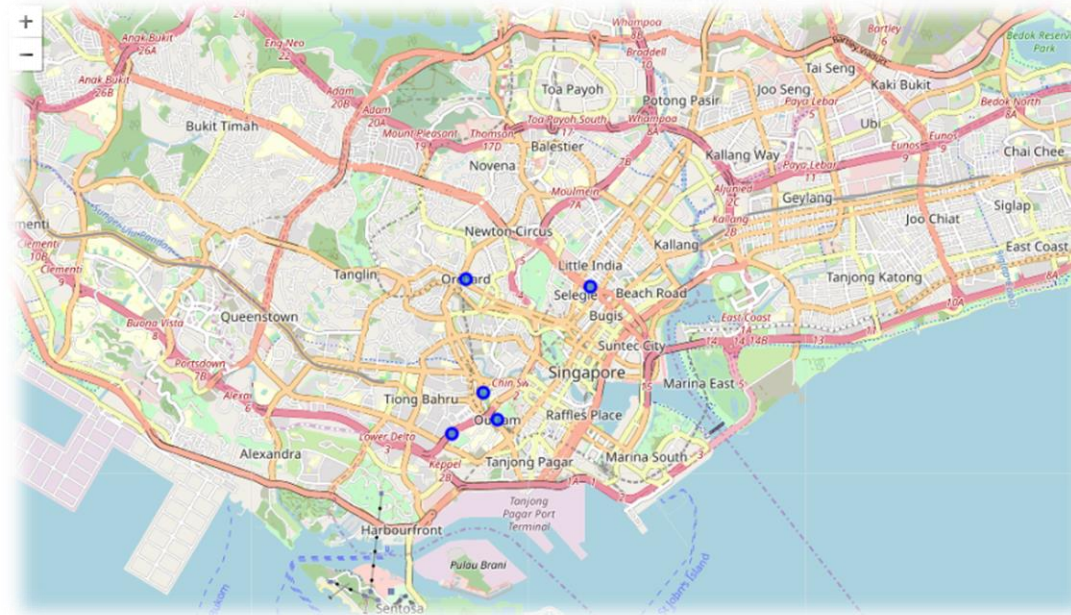


	neighborhood	latitude	longitude	1st	2nd	3rd
24	East,Changi Bay	1.31685	104.021	Harbor / Marina	Gym Pool	Hot Spring
43	West,Bukit Batok	1.34828	103.749	Pool	Basketball Court	Playground

# Results

## Cluster 6

1. FITNESS SPORTS
2. OUTDOOR SPORTS
3. RUNNING SPORTS



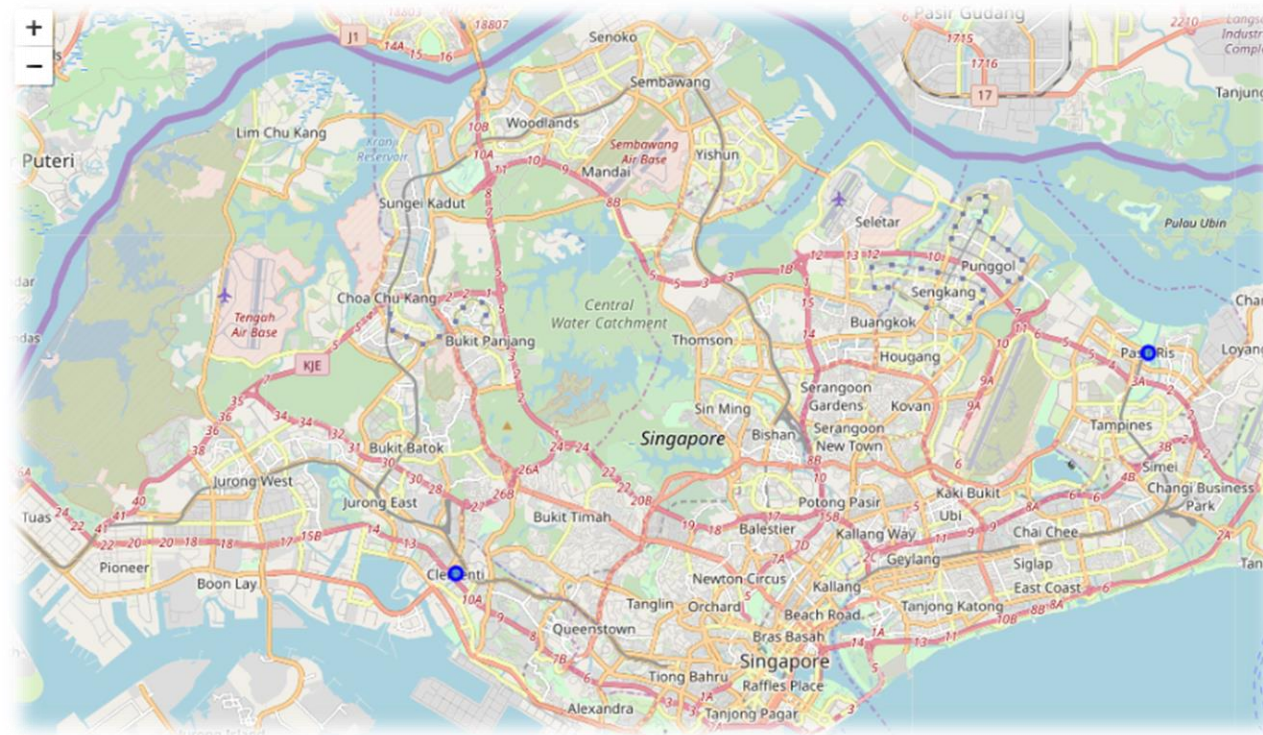
	neighborhood	latitude	longitude	1st	2nd	3rd
1	Central,Bukit Merah	1.28063	103.831	Yoga Studio	Athletics & Sports	Trail
3	Central,Downtown Core	1.28712	103.836	Yoga Studio	Pedestrian Plaza	Fishing Spot
12	Central,Orchard	1.30527	103.833	Yoga Studio	Gym / Fitness Center	Scenic Lookout
13	Central,Outram	1.28287	103.838	Yoga Studio	Sports Club	Farm
16	Central,Rochor	1.30398	103.853	Yoga Studio	Gym / Fitness Center	Playground



# Results

## Cluster 7

1. COMBAT SPORTS
2. FITNESS SPORTS
3. TEAM SPORTS



	neighborhood	latitude	longitude	1st	2nd	3rd
26	East,Pasir Ris	1.37422	103.951	Recreation Center	Gym	Park
48	West,Clementi	1.31403	103.762	Pool	Gym	Park

# Results

## Cluster 8

1. TARGET SPORTS
2. FITNESS SPORTS
3. WATER SPORTS



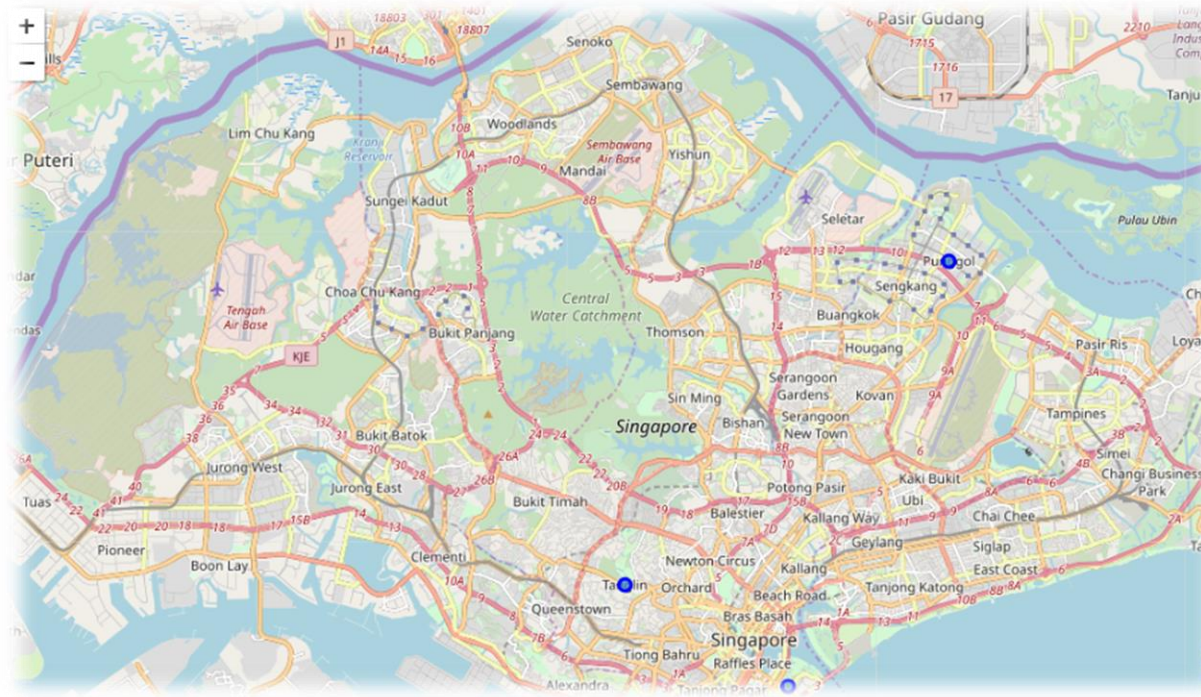
	neighborhood	latitude	longitude	1st	2nd	3rd
53	West, Western Water Catchment	1.36645	103.681	Gun Range	Yoga Studio	Hot Spring



# Results

## Cluster 9

1. RUNNING SPORTS
2. TEAM SPORTS
3. FITNESS SPORTS

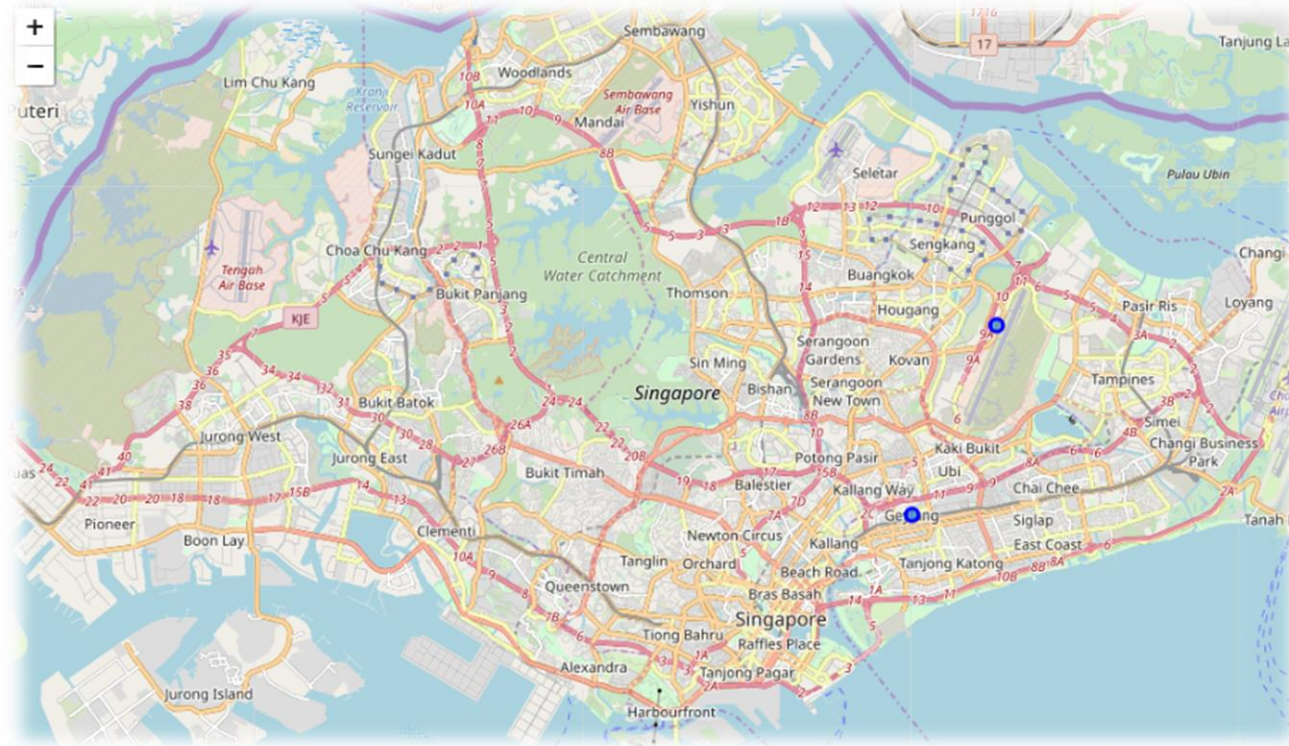


	neighborhood	latitude	longitude	1st	2nd	3rd
7	Central, Marina South	1.277	103.862	Garden	Park	Yoga Studio
20	Central, Tanglin	1.30604	103.815	Trail	Park	Garden
39	North-East, Punggol	1.39803	103.907	Basketball Court	Park	Garden

# Results

## Cluster 10

1. TEAM SPORTS
2. RUNNING SPORTS
3. FITNESS SPORTS



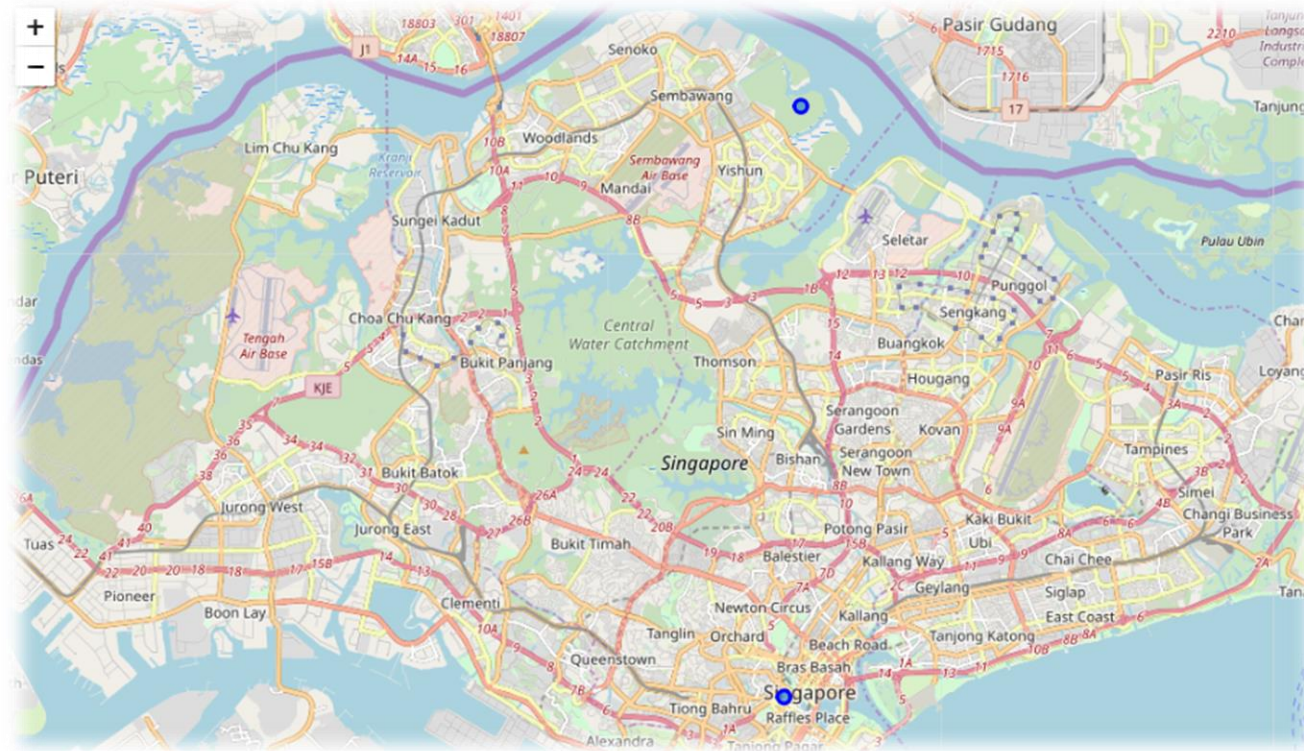
	neighborhood	latitude	longitude	1st	2nd	3rd
4	Central,Geylang	1.31819	103.887	Basketball Court	Soccer Field	Yoga Studio
40	North-East,Seletar	1.36889	103.91	Basketball Court	Dog Run	Yoga Studio



# Results

## Cluster 11

1. WATER SPORTS
2. OUTDOOR SPORTS
3. FITNESS SPORTS

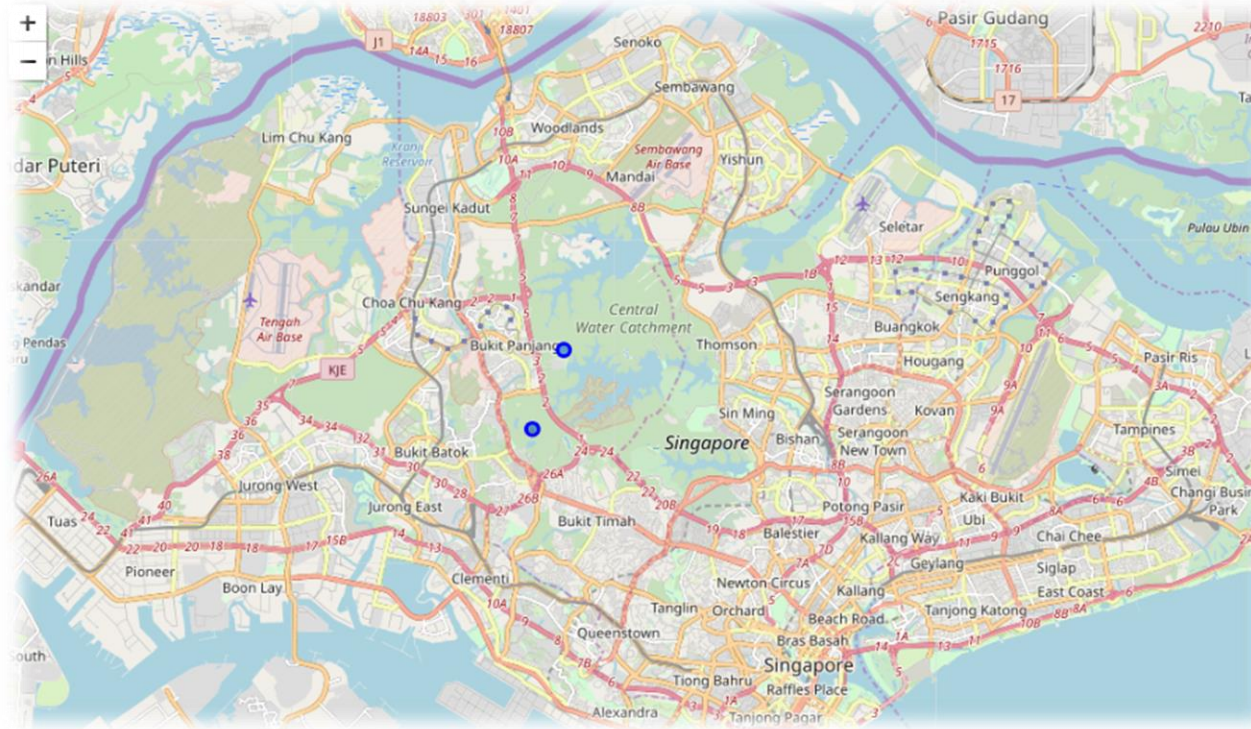


	neighborhood	latitude	longitude	1st	2nd	3rd
17	Central, Singapore River	1.28918	103.845	Waterfront	Canal	River
32	North, Simpang	1.4453	103.85	Outdoors & Recreation	River	Yoga Studio

# Results

## Cluster 12

1. OUTDOOR SPORTS
2. CYCLING
3. FITNESS SPORTS



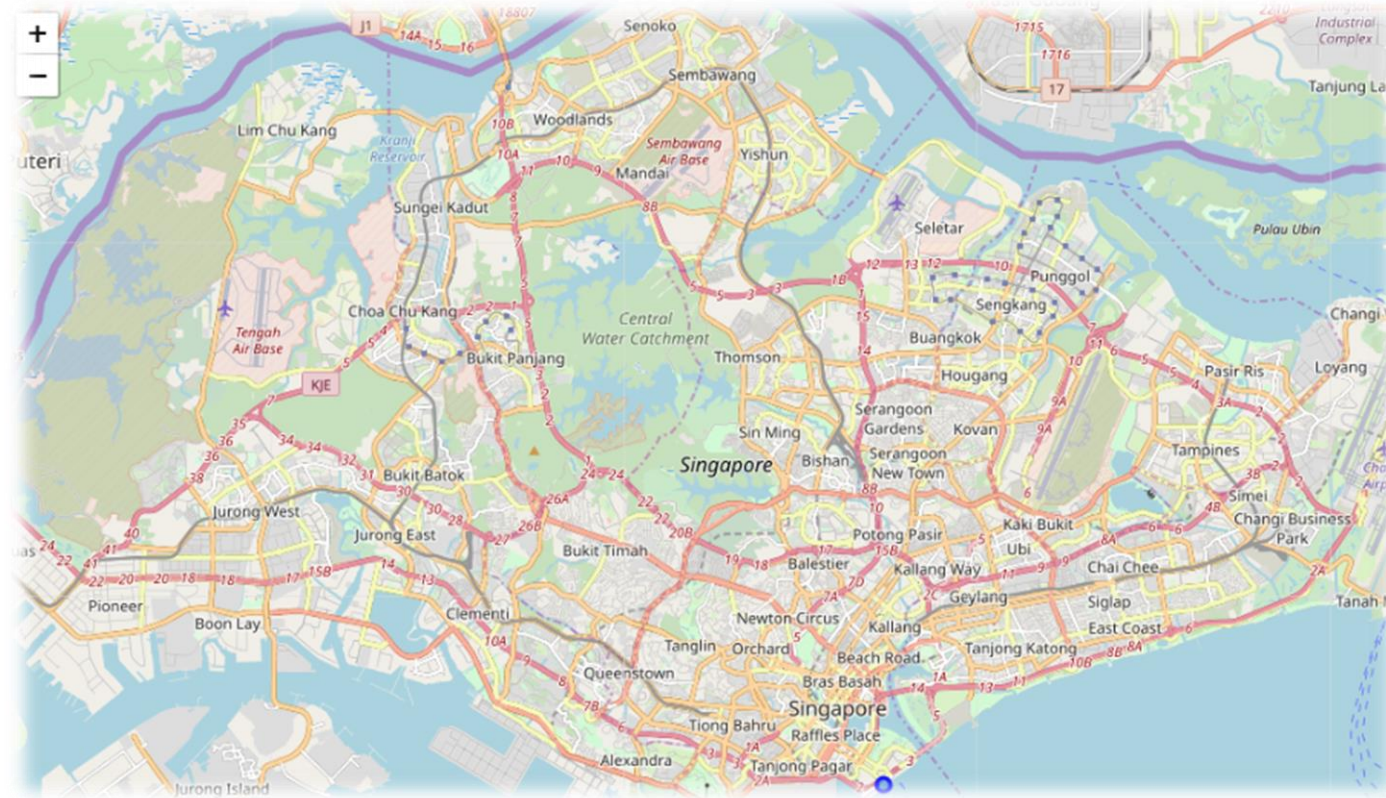
	neighborhood	latitude	longitude	1st	2nd	3rd
2	Central,Bukit Timah	1.35469	103.776	Mountain	Nature Preserve	Scenic Lookout
28	North,Central Water Catchment	1.37632	103.785	Nature Preserve	Bike Trail	Yoga Studio



# Results

## Cluster 13

1. TEAM SPORTS
2. FITNESS SPORTS
3. OUTDOOR SPORTS

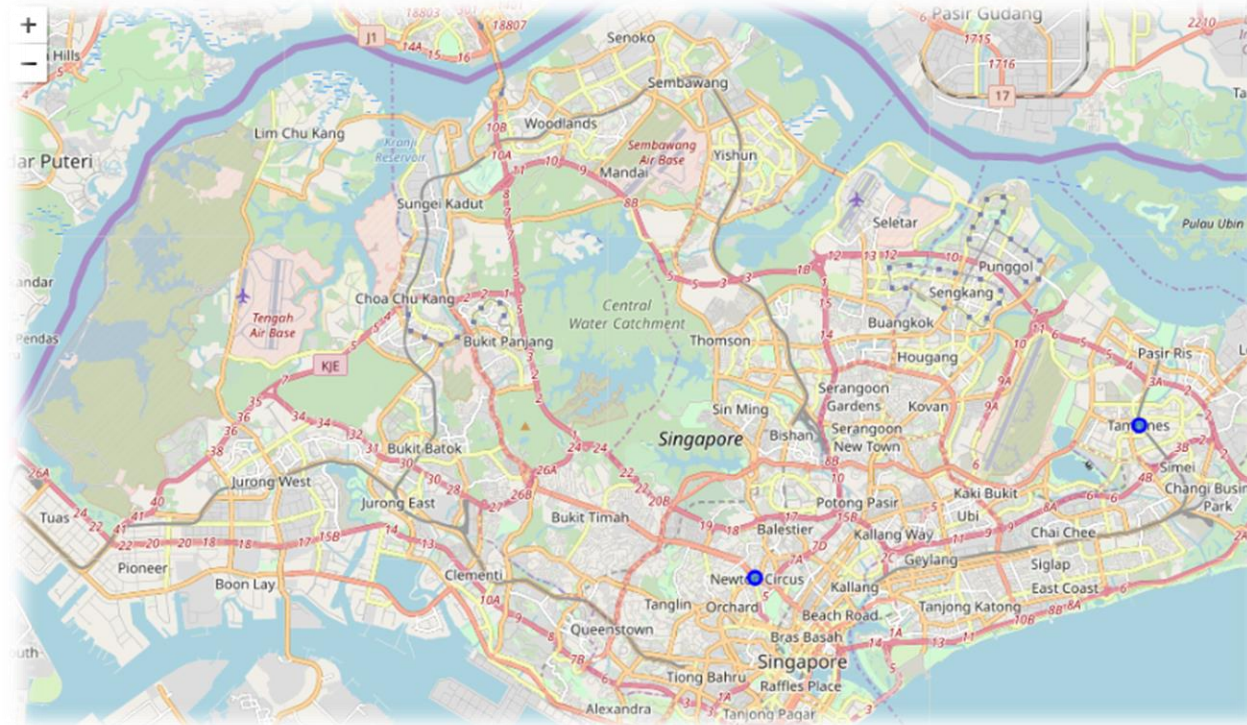


	neighborhood	latitude	longitude	1st	2nd	3rd
18	Central, Southern Islands	1.27109	103.863	Park	Yoga Studio	Fishing Spot

# Results

## Cluster 14

1. FITNESS SPORTS
2. RUNNING SPORTS
3. URBAN SPORTS



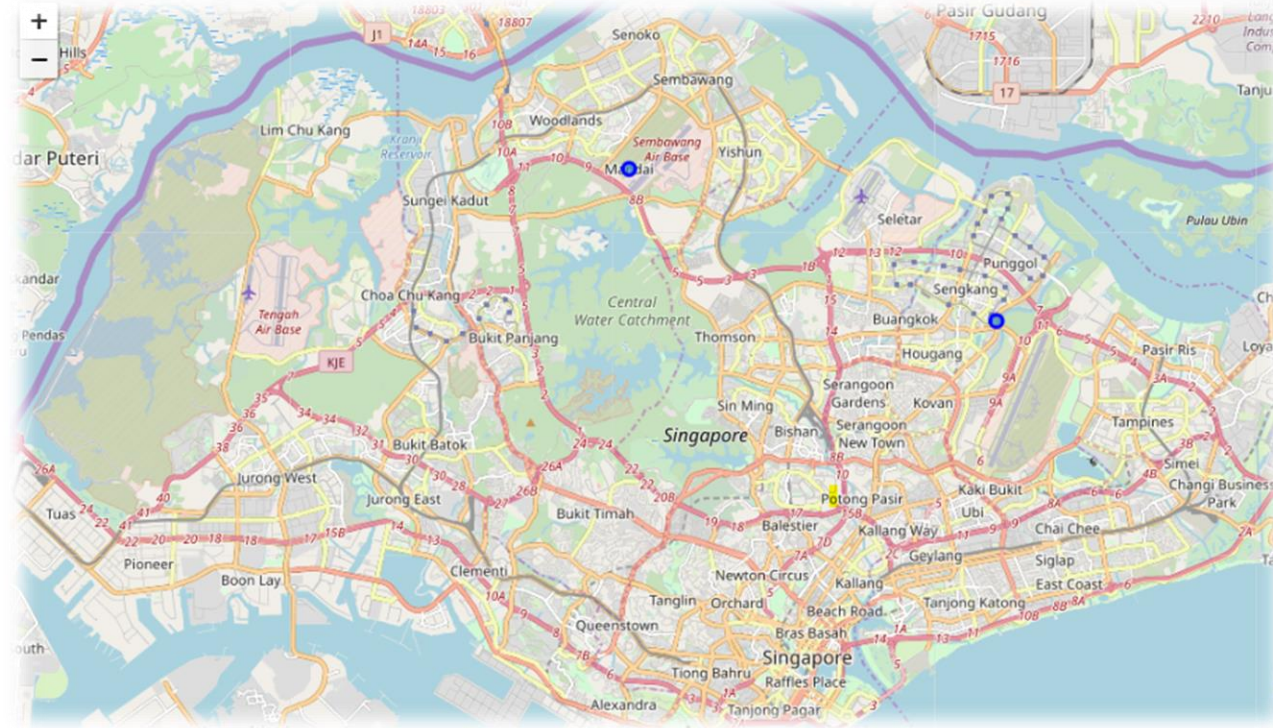
	neighborhood	latitude	longitude	1st	2nd	3rd
10	Central,Newton	1.31296	103.839	Gym	Gym / Fitness Center	Yoga Studio
27	East,Tampines	1.35465	103.944	Gym	Gym / Fitness Center	Yoga Studio



# Results

## Cluster 15

1. RUNNING SPORTS
2. FITNESS SPORTS
3. OUTDOOR SPORTS



	neighborhood	latitude	longitude	1st	2nd	3rd
30	North,Mandai	1.42353	103.803	Trail	Yoga Studio	Farm
41	North-East,Sengkang	1.38199	103.903	Trail	Park	Yoga Studio

# Discussion

# Discussion

- ▶ The objective of this project task is to find better localities where specific product brands and categories can be focused and advertise for better reach among public. Also it aids **Decathlon** management to consider opening new outlets serving specific region based on top suggested products to improve business.
- ▶ By applying **K-Means Algorithm** it was possible to achieve desired objective by segmenting localities in **Singapore** based on top venues thereby assisting to understand local market better.
- ▶ Observed findings are now ready to be presented to **Decathlon** management for assisting them to take progressive decisions on expanding their business in **Singapore** market.

# Conclusion



# Conclusion

- ▶ We have gathered data from trustworthy data sources, applied widely recognized clustering algorithm (K-Means) to perform segmentation to derive meaningful insight, hopefully it will be considered by top management for decision making with considerable level of confidence.
- ▶ Proposed idea can be easily applied in any domain to perform base studies on competitor insight, market research analysis etc., based on specific requirement.

**THANK YOU**  
நன்றி  
**TERIMA KASIH**  
謝謝

