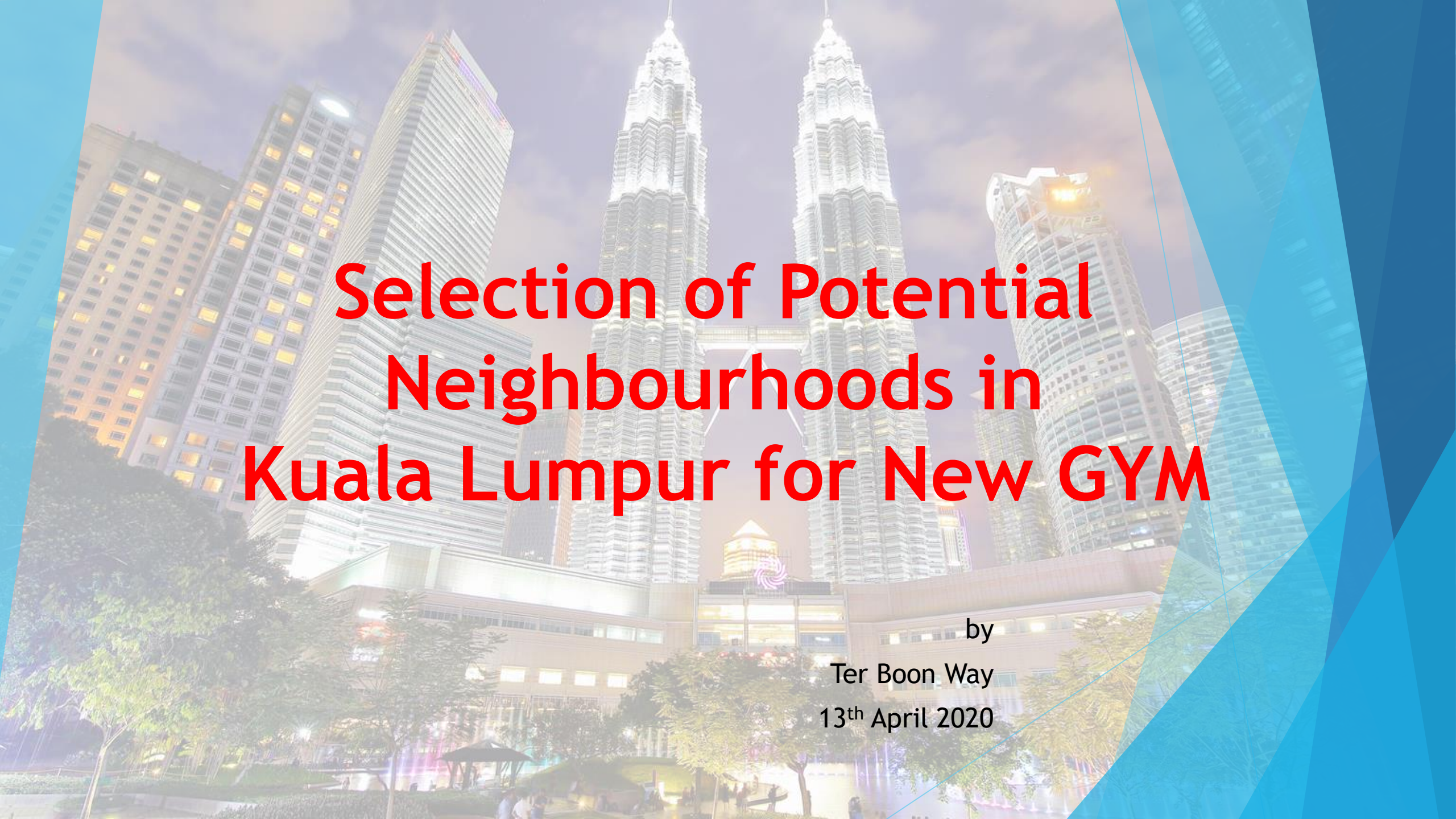


Selection of Potential Neighbourhoods in Kuala Lumpur for New GYM

by

Ter Boon Way

13th April 2020

INTRODUCTION

Business problem

Approach

Stakeholders

Data

Business Problem

- An investor is looking to open a new gym in Kuala Lumpur.
- Based on his previous experiences and marketing strategy, he would like to tap-into a mature neighbourhood with high traffic, but low competition.

- Area selection criteria:

- ✓ Area that has hotels or shopping mall or residential (apartments or condo) in vicinity.
- ✓ Area that is not already crowded with gyms

How?

Solution:

Data science approach.

The neighbourhoods data are collected, processed and clustered by Machine Learning Algorithm (Kmeans Clustering). Clusters of potential neighbourhoods are located.

Stakeholders

The audiences that might be interested on this project:

- Those that are interested in knowing high traffic attraction areas in Kuala Lumpur
- Those that are interested in knowing the area with sports amenities in Kuala Lumpur
- Those that have interest in using the exploration result from this project

Data

- **Neighbourhood list of Kuala Lumpur**

  From https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur

- **Coordinates of each neighbourhood**

  From python geocoding library

- **Venues in each neighbourhood**

  From Foursquare

METHODOLOGY

Pre-processing

Kmeans Clustering

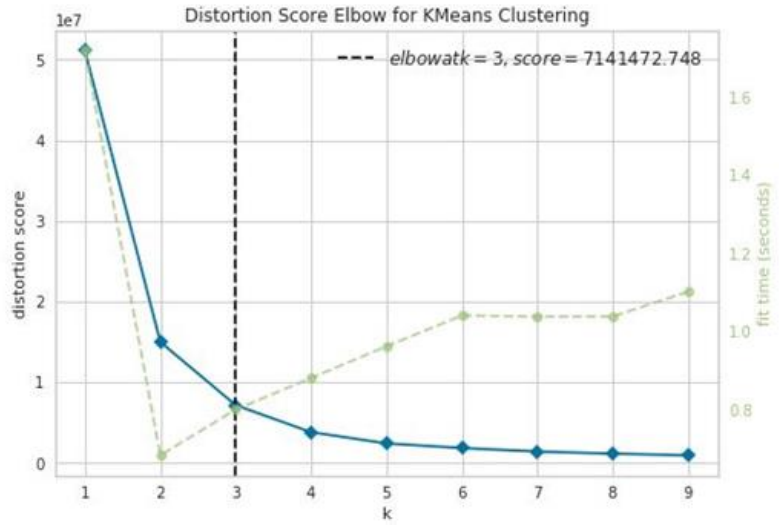
Pre-processing

- Data filtering
- One-hot encoding
- GroupBy Analysis

The data collected are pre-processed, and converted into form that are interpretable by machine learning algorithm.

Kmeans Clustering

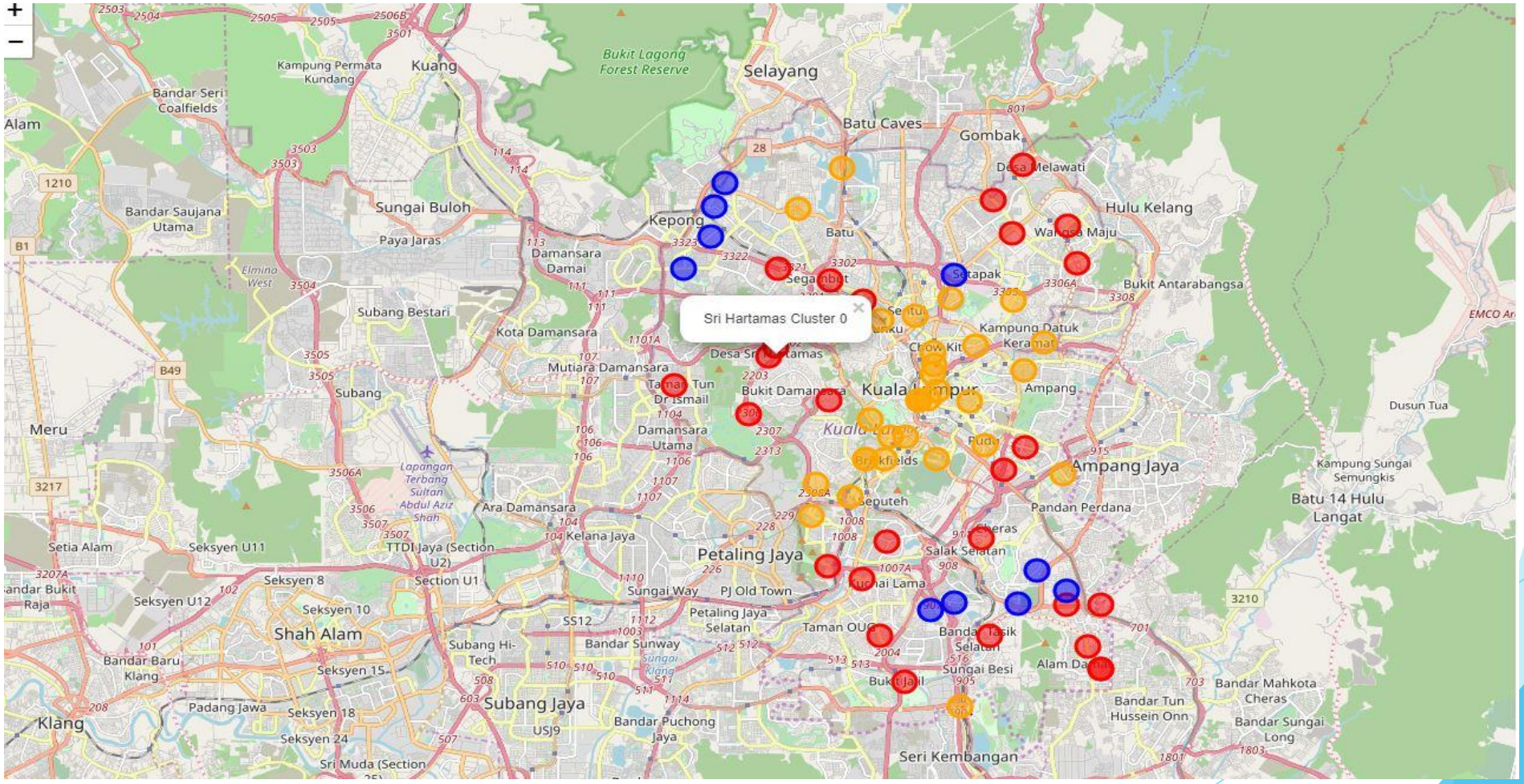
- The suitable numbers of cluster are determined through elbow method.



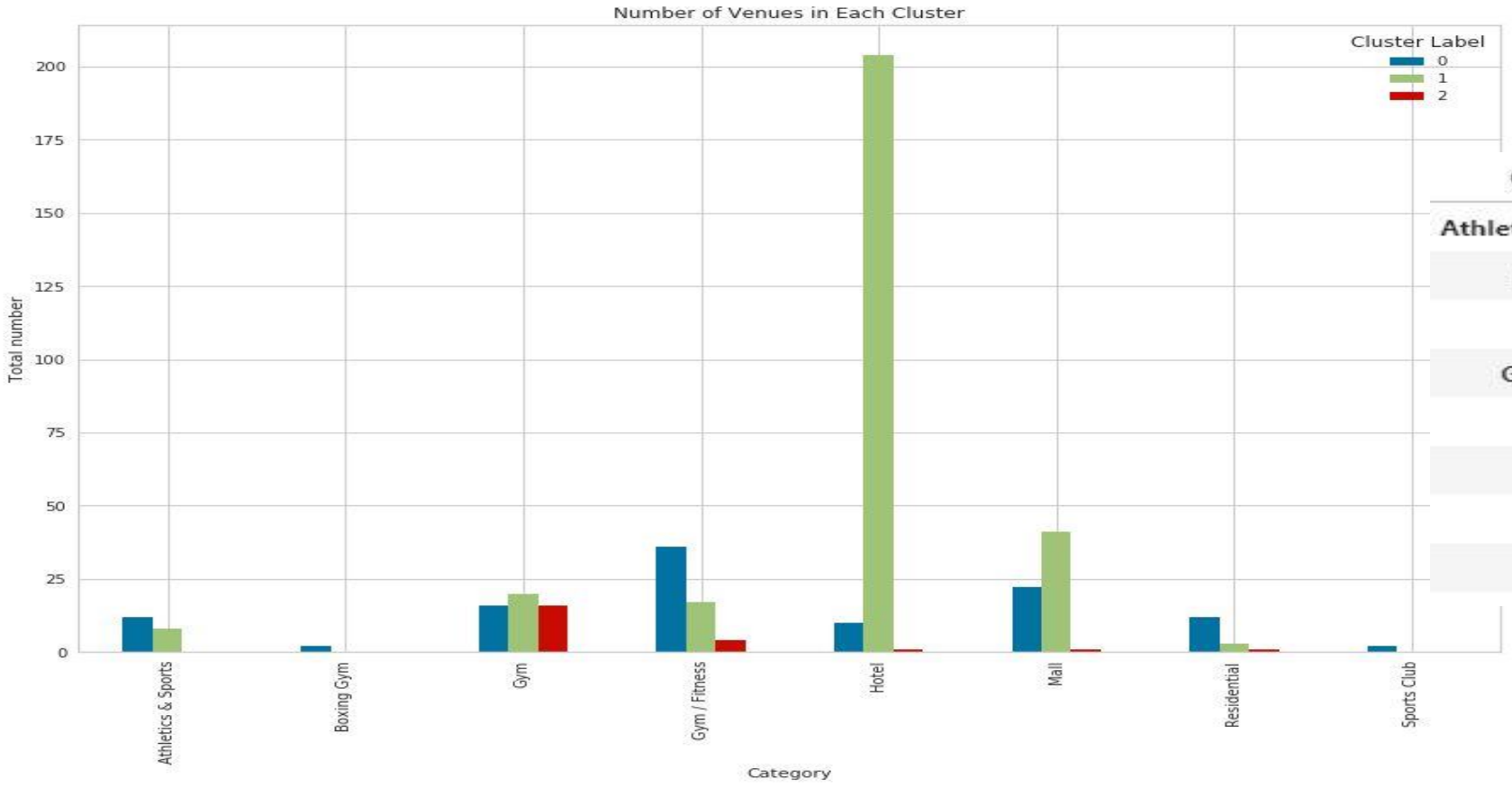
- The data are fit into machine learning algorithm, with selected number of cluster.

RESULT & DISCUSSION

Result:
3 clusters are identified



Examine the clusters



Cluster Label	0	1	2
Athletics & Sports	12	8	0
Boxing Gym	2	0	0
Gym	16	20	16
Gym / Fitness	36	17	4
Hotel	10	204	1
Mall	22	41	1
Residential	12	3	1
Sports Club	2	0	0

Examine the clusters cont..

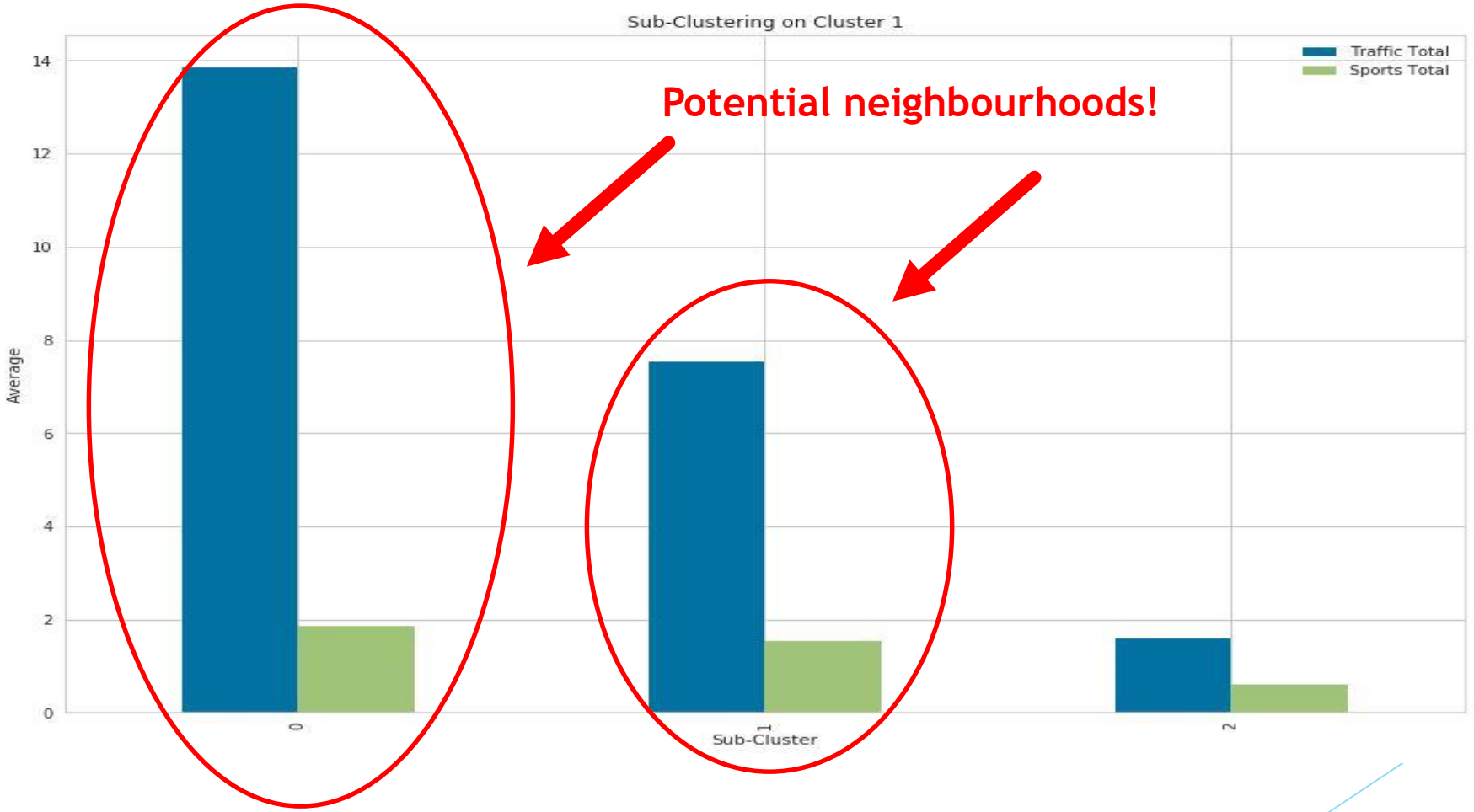
	Description
Cluster 0	Moderate number of traffic attraction venues high amount of sports amenities
Cluster 1	High amount of traffic attraction venues , moderate amount of sports amenities
Cluster 2	Low amount of traffic attraction venues, low amount of sports amenities

Examine the clusters cont..

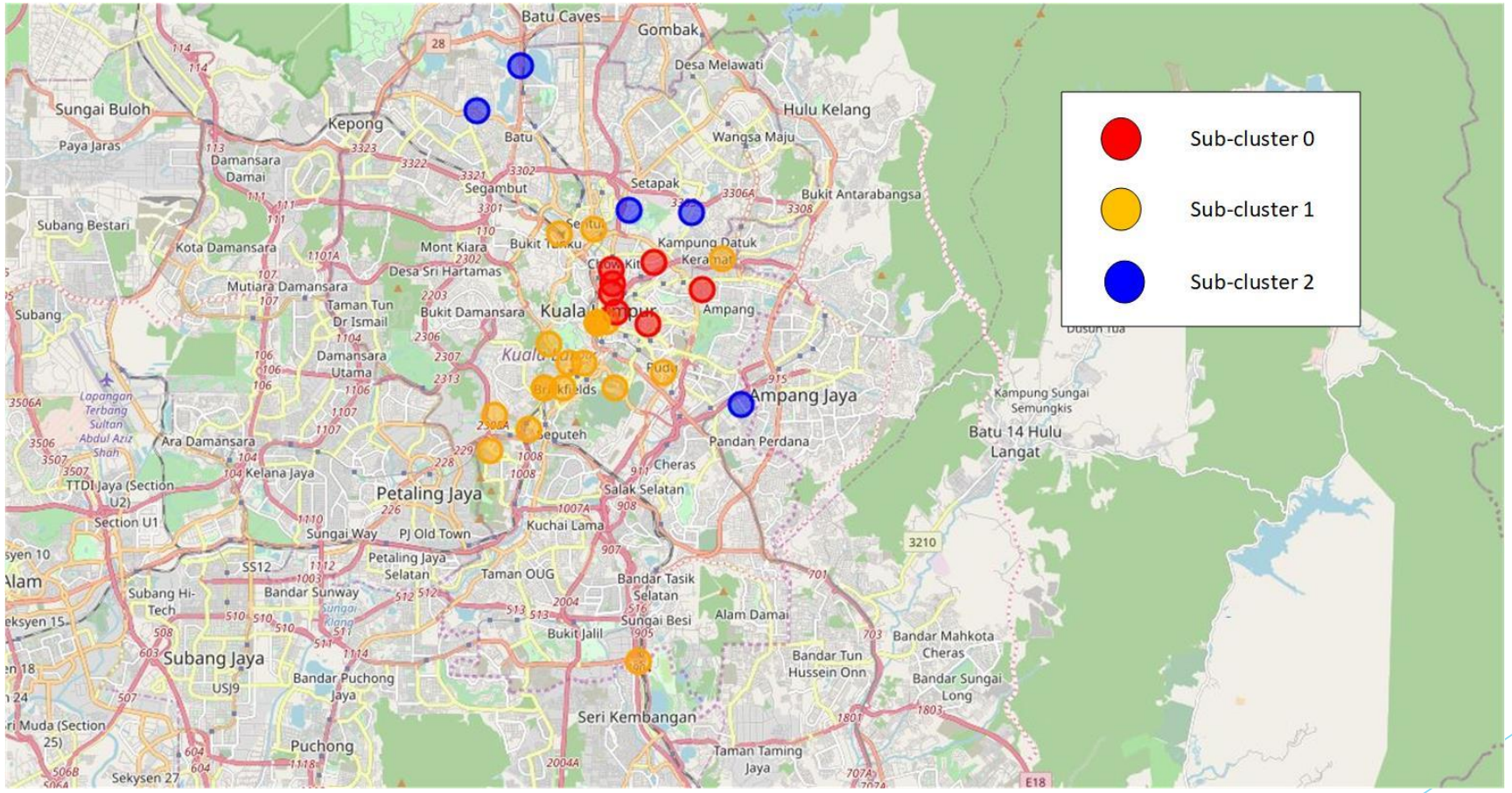
Cluster 1 looks interesting! Let’s explore cluster 1 in details.

- 30 neighbourhoods in total
- Could be further segmented into 3 sub-clusters

Sub-clustering on cluster 1



Sub-clustering on cluster 1



Cluster 1 - Sub cluster 0

7 neighbourhoods

	Neighborhood	Sub-Cluster label	Traffic Total	Sports Total	Latitude	Longitude
0	Bukit Bintang	0	15	1	3.147770	101.708550
1	Bukit Nanas	0	12	4	3.151142	101.699375
2	Chow Kit	0	14	1	3.163590	101.698110
3	Dang Wangi	0	15	3	3.156685	101.698077
4	Kampung Baru, Kuala Lumpur	0	12	0	3.165460	101.710280
5	Medan Tuanku	0	15	2	3.159260	101.698340
6	Taman U-Thant	0	14	2	3.157700	101.724520

Cluster 1 – Sub cluster 1

19 neighbourhoods

	Neighborhood	Sub-Cluster label	Traffic Total	Sports Total	Latitude	Longitude
0	Ampang, Kuala Lumpur	1	9	2	3.148494	101.696729
1	Bangsar	1	10	0	3.129200	101.678440
2	Bangsar Park	1	10	0	3.129200	101.678440
3	Bangsar South	1	6	2	3.111020	101.662830
4	Batu, Kuala Lumpur	1	7	2	3.147890	101.694050
5	Brickfields	1	10	2	3.129160	101.684060
6	Bukit Petaling	1	7	1	3.129290	101.698960
7	Bukit Tunku	1	6	2	3.173810	101.682760
8	Damansara Town Centre	1	8	2	3.136442	101.690296
9	Damansara, Kuala Lumpur	1	8	1	3.141906	101.679678
10	Federal Hill, Kuala Lumpur	1	9	2	3.136370	101.685640
11	KL Eco City	1	7	1	3.117130	101.673840
12	Kampung Datuk Keramat	1	6	2	3.166400	101.730460
13	Lembah Pantai	1	8	1	3.121202	101.663899
14	Maluri	1	7	2	3.147890	101.694050
15	Miharja	1	7	2	3.147890	101.694050
16	Pudu, Kuala Lumpur	1	6	0	3.133540	101.713070
17	Sentul, Kuala Lumpur	1	6	2	3.175080	101.693050
18	Sungai Besi	1	6	3	3.049970	101.706030

Conclusion

The potential neighbourhoods are located.

2 classes:

- Class 1 (Cluster 1 – sub cluster 0)
  Best fit
  High traffic attraction venues
  Moderate to low competition

- Class 2 (Cluster 1 – sub cluster 1)
  Moderate amount of traffic attraction venues
  Moderate to low competition