



# Coursera Capstone Project

*IBM Data Science Professional Certificate*

**Best location for opening a gym accessory retail outlet in  
Singapore**

By: Yang Lin

Date: 9<sup>th</sup> January 2021

# Business Problem and Objective

- High demand for gym and fitness lifestyle boom demand for gym peripheral products.
- Existing gym accessory retailers may want to open more outlets for business expansion.
- Potential investors want to seize the investment opportunities in gym accessory retails industry.
- Locations where to open gym accessory outlets become the vital factor for decision making.
- Where are the best neighborhood locations to open new gym accessory retail outlets in Singapore?

## **Objective:**

- Identify the best locations for opening new gym accessory outlets in Singapore.

# Data

- List of neighborhoods in Singapore. This defines the scope of this project that only within the Country of Singapore.
- Latitude and longitude coordinates of the neighborhoods. This is required in order to plot on the map as well as getting the venue data.
- Venue data, particularly data related to gym and fitness centers. The venue data will be used in the machine learning process to cluster the labels on neighborhoods.

# Sources of Data

- List of neighborhoods of Singapore from Wikipedia.
- Geographical Coordinates from Geocoder package.
- Venues data exploration by using free version of Foursquare API.
- Latitude and Longitude of venues.

# Methodology

## Data sourcing

- Web scraping Wikipedia page for neighbourhoods list
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data

## Data cleaning

- Group data by neighbourhood and taking the means of the frequency of occurrence of each venue category
- Filter venue category by Gym / Fitness Center

## Machine learning & Clustering

- Perform clustering on the data by using k-means clustering

## Data Visualization

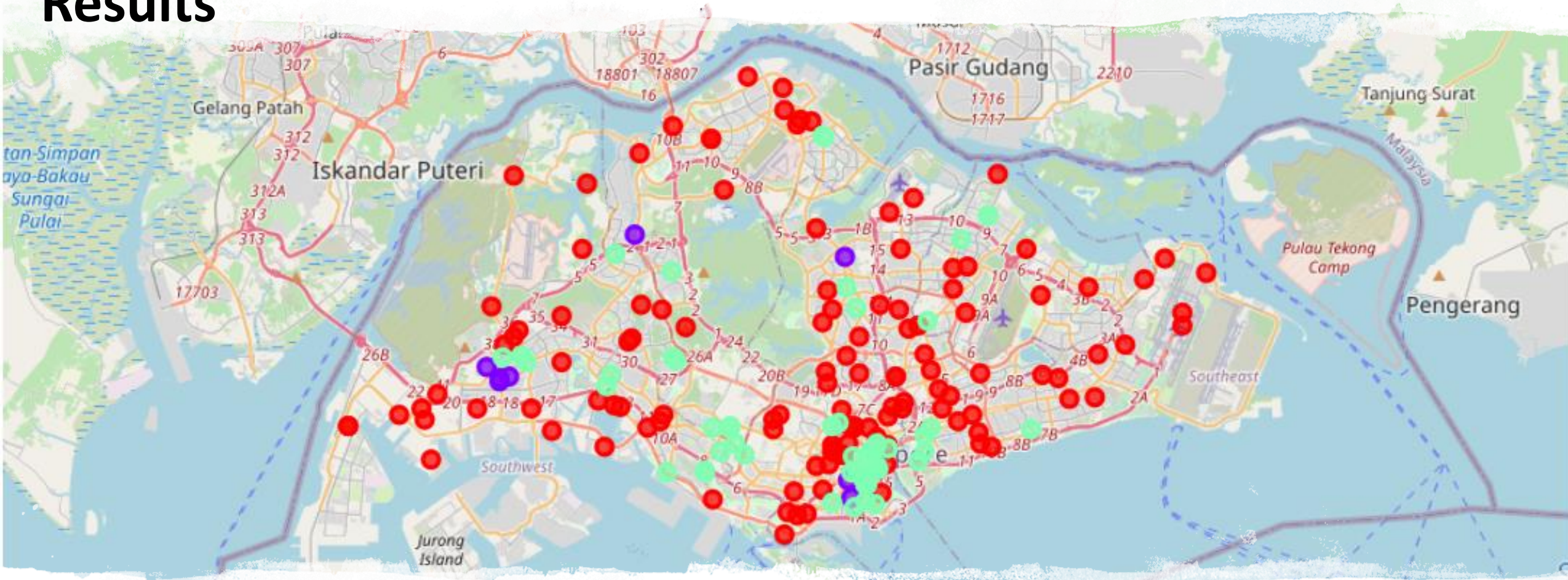
- Visualize the clusters in a map using Folium

## Result analysis

- Analyse each cluster and form the solutions to business question



# Results



- ❑ Cluster 0 (red): Neighborhoods with low frequency of occurrence and numbers of Gym and Fitness Centers.
- ❑ Cluster 1 (purple): Neighborhoods with high frequency of occurrence and numbers of Gym and Fitness Centers.
- ❑ Cluster 2 (green): Neighborhoods with moderate frequency of occurrence and numbers of Gym and Fitness Centers.

# Discussion and Recommendation

- 123 neighborhoods are in cluster 0 that gives lowest frequency of occurrence and therefore, not suitable for opening new outlets.
- 8 neighborhoods are in cluster 1 shows the highest group of frequency of occurrence. Neighborhoods in Cluster 1 and especially the central CBD locations are highly recommended for opening new outlets.
- The rest of 60 neighborhoods are moderate level in cluster 2. Certain locations with high number of outlets can be the next tier to consider for decision.

# Limitations

- Only two factor, frequency of occurrence and numbers of Gym and Fitness Centers are considered.
- Other factors such as population, lifestyle and shopping mode (on-line/off-line) are not taken into scope of this project.
- Limited calls and unstable result from free version Foursquare API.



# Conclusion

- The neighborhoods in cluster 1 are the most preferred locations and certain neighborhoods in cluster 2 are next tier to consider for opening a new outlet.
- The findings of this project will help the relevant stakeholders to capitalize on the opportunities in high potential locations while avoiding the low potential locations for business growth.

A rectangular chalkboard with a light brown wooden frame is centered in the image. The words "Thank" and "You" are written in white, serif, all-caps font, stacked vertically. The background is a rustic wooden surface with horizontal planks. In the top right corner, a green leaf is partially visible. In the bottom left corner, a portion of an orange object is visible.

Thank  
You