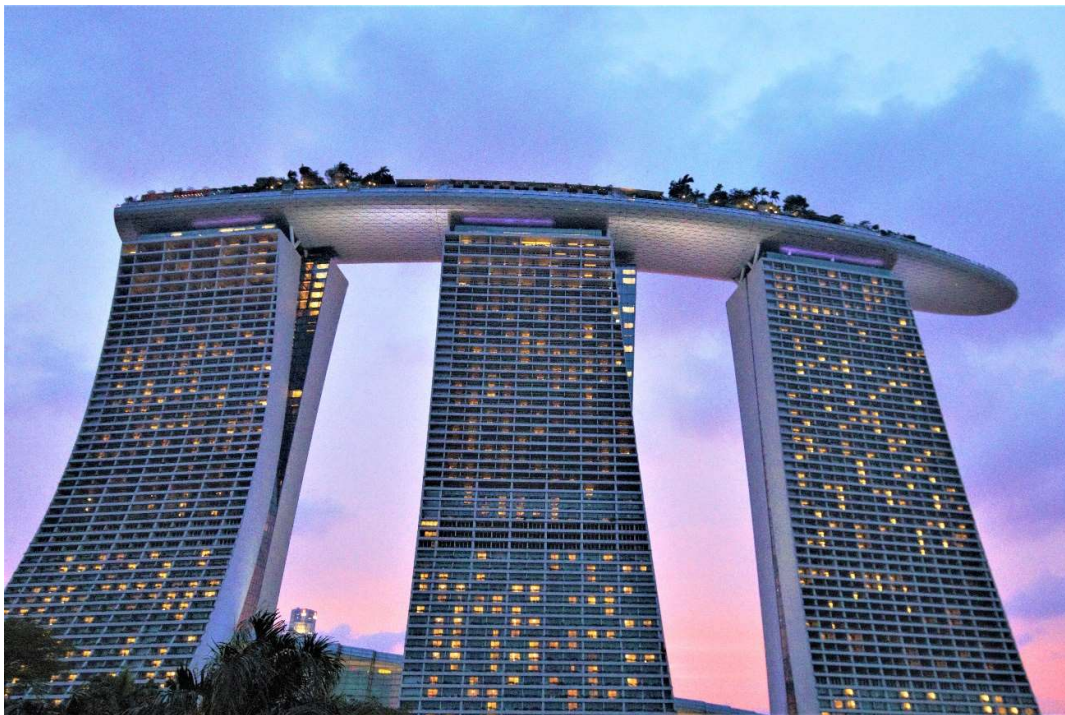


IBM Applied Data Science Project

Coursera Capstone

Potential Places to Establish Hospitals in Singapore



Vincentius Indra Sulaiman
March 2020

A. Introduction

A.1 Business Understanding

Due to the increasing median age of Singaporeans (see Figure 1), hospitals will become more important and be an integral part of Singaporeans. According to Statista, percentage of older population in Singapore is also found to be increasing over the years (see Figure 2). This phenomenon of ageing population will assert the importance of hospitals to cater the medical needs of Singaporeans. Singapore health care quality is known to be one of the best in South East Asia (International, 2020)¹. According to Ranking Web of Hospitals², there are 11 Singapore hospitals from the top 30 hospitals in South East Asia. This fact further supports the economic opportunity of building new hospitals in Singapore as the health care market is not only limited to Singaporeans. However, the scope of the present study will be limited to Singapore population, not taking into account the overseas patients.

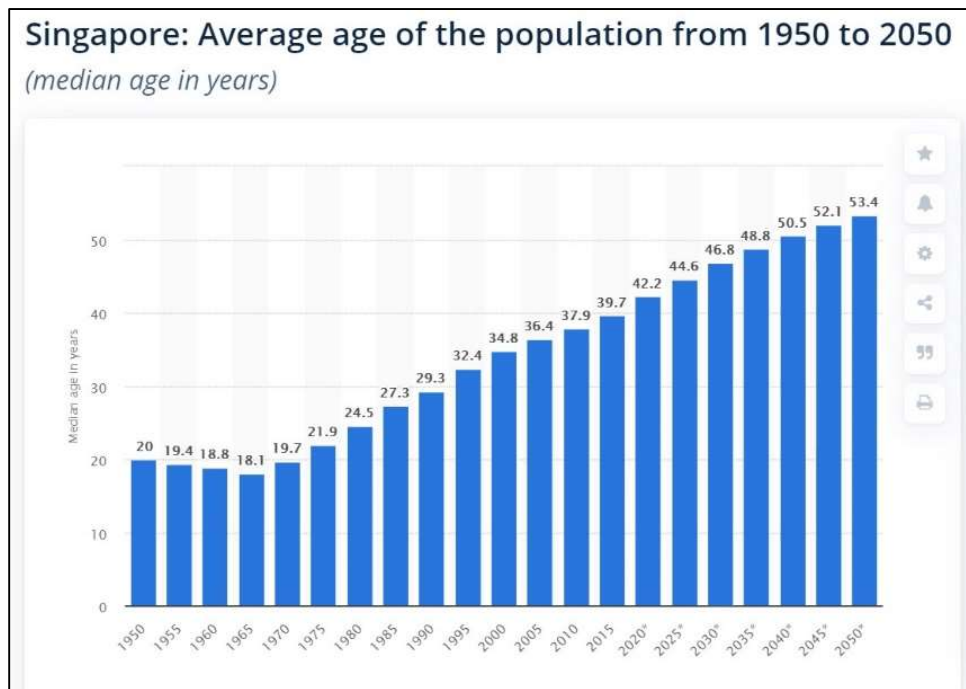


Figure 1. Median Age of Singaporeans (statista.com)

Based on prior knowledge, it is known that most hospitals or medical centres are located in central and southern areas of Singapore. However, Singapore population in general does not reside in those areas only (see Figure 3). Therefore, it becomes an interest to find which areas in Singapore that have relatively higher population density and less amounts

¹ (International, 2020)

² (South East Asia, n.d.)

of medical centres. The results of this investigation will be very handy and relevant for urban developers specialising in the construction of such medical centres of hospitals. Thus, valuable economic opportunities may be discovered by this study.

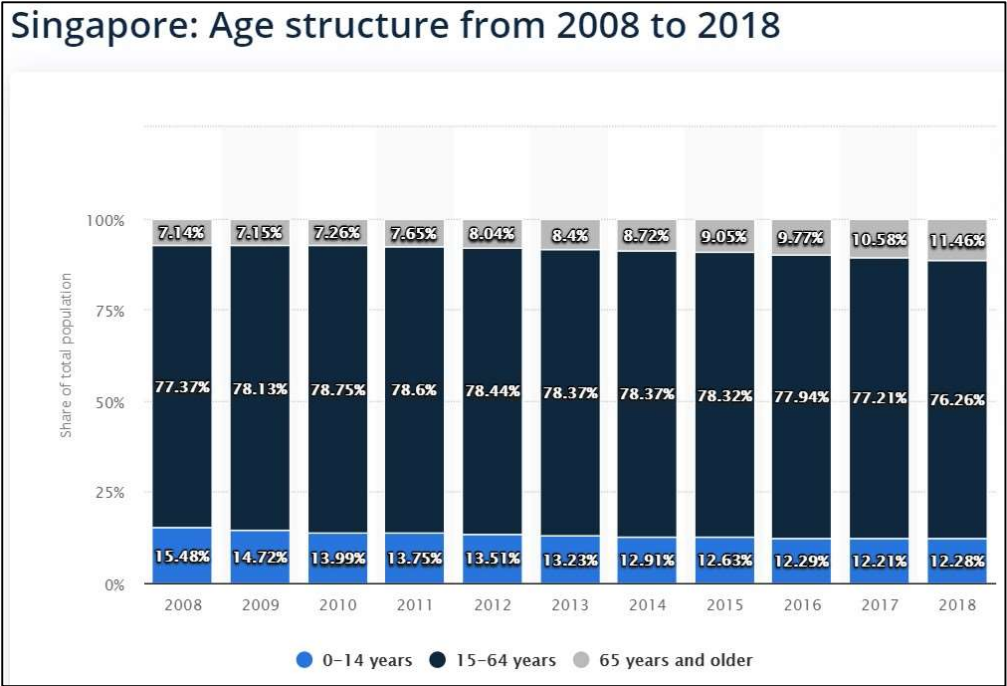


Figure 2 Singaporean Age Structure (statista.com)

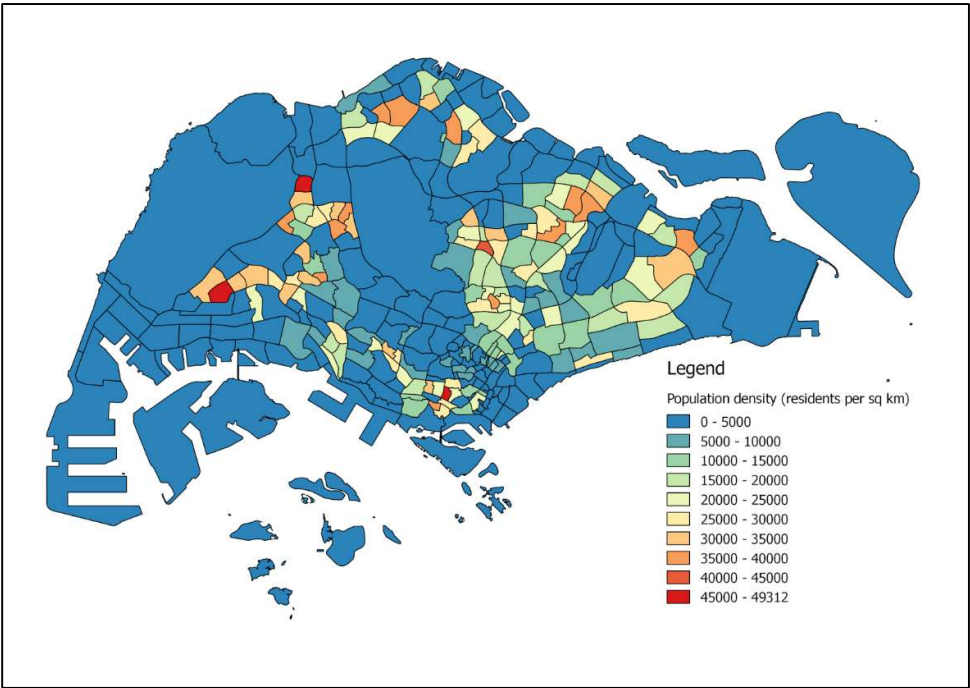


Figure 3 Singapore Population Density Heatmap (mothership.sg)

B. Study Objectives and Target Market

This study is specifically intended for:

1. Property developers and investors looking for opportunities to build new hospitals,
2. Government agencies, in particular Singapore Ministry of Health (MOH), which have interest in developing certain parts of Singapore related to healthcare industry, and
3. Individuals and families residing in certain areas of Singapore which currently have relatively lower density of hospital or medical centres despite the higher population density.

C. Data Requirements, Collection, and Preparation

C.1 Data Requirements

To fulfil the objectives of the current study, publicly available data is of upmost importance to provide valuable insights. As a whole, the required data are as follows:

1. List of neighbourhoods (towns) in Singapore. This dictates the scope of the present study which is the country of Singapore.
2. Location coordinates (latitude and longitude) of the neighbourhoods. This will be important to map the data.
3. Population of Singaporean residents in the neighbourhoods. This will be used to cluster Singapore towns.
4. Location of large hospitals and medical centres in Singapore. This will be used to cluster Singapore towns. Hospital size is determined by the presence of wards.

C.2 Methods of Data Collection

Based on the aforementioned required data, the following methods will be employed to collect them:

1. Scraping: https://en.wikipedia.org/wiki/New_towns_of_Singapore to obtain list of neighbourhoods with the total population and residential areas.
2. Utilising the geopy library to obtain geographical coordinates of Singapore towns.
3. Accessing Foursquare data through API to obtain crucial data on the locations of hospitals and medical centres in Singapore. Hospital wards fall under the foursquare category id: 58daa1558bbb0b01f18ec1f7.

C.3 Preparation of Data

After the data collection step has been done, the following measures will be undertaken to process the raw data:

1. Cleaning: to remove unnecessary parts of raw data. Cleaning is an iterative process: it is done in parts across the whole process, not only in the beginning section of the project.
 - a. Removing unnecessary columns from wikipedia table: name Chinese, Romanised Chinese name in Pinyin, name in Tamil, number of dwelling units, and projected number of ultimate dwelling units.
 - b. Modifying problematic data: town of Yishun that does not return geographical coordinates. The coordinates are searched on google and appended manually on the dataframe.
 - c. Renaming column names to ease analysis step.
 - d. Removing unnecessary duplicate columns that arise from hospital ward counting step.
 - e. Removing unnecessary columns that may be a hindrance to training algorithm.
 - f. Removing unnecessary data columns when analysing clusters.
2. Creating dataframes: to further simplify the processing step especially after extracting Foursquare data.
3. Scaling the data using MinMaxScaler from Sklearn to provide better machine learning algorithm.
4. Merging various information to selected dataframes: to provide the most useful information in the least amount of dataframes to further ease the analysis step.

D. Analysis of The Problem

The prepared data will be analysed in accordance with the following steps:

1. Data manipulation: arithmetic calculation of population density by dividing the total town population with town area.
2. Applying machine learning method: clustering. This is done by first initiating the KMeans training model from Sklearn, finding the optimal k-number for clustering the towns based on population density and amount of hospitals.
3. Mapping the data: hospital locations and the clusters in a Folium map.

By the implementation of clustering method of machine learning, the analysis step is expected to provide sufficient information and description on the clusters of Singapore towns. These findings will be the basis for the

readers to further comprehend the healthcare economic opportunity in Singapore.

E. Results and Discussion

Results from Machine Learning

1. Optimal number of k: the optimal number for KMeans machine learning is obtained through KElbowVisualizer from yellowbrick. The result is 6 and may be seen from the Figure 4 below.

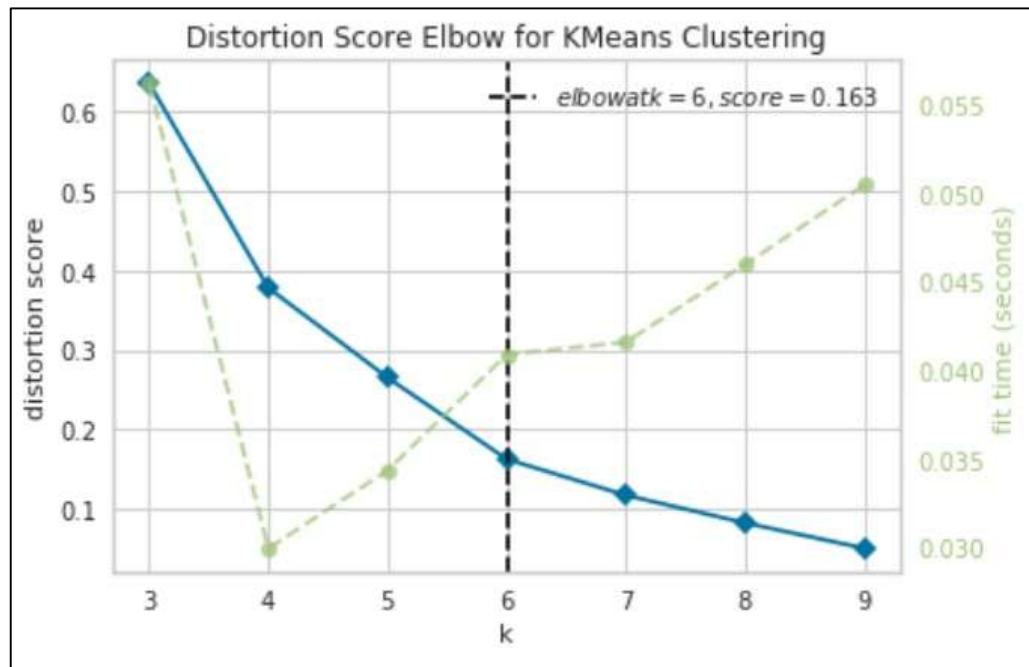


Figure 4 KElbowVisualizer from Yellowbrick

2. Clustering of Towns.
By taking k-number = 6, we obtain six different clusters of Singapore towns. The results are as the following Table 1 and mapped as in Figure 5 using Folium.

Table 1 Cluster and Members

Cluster	Towns
0	Toa Payoh
1	Bedok, Geylang, Hougang, Jurong East, Sengkang, Serangoon, Tampines, Yishun
2	Punggol, Sembawang
3	Ang Mo Kio, Bukit Panjang, Choa Chu Kang, Jurong West, Woodlands
4	Bishan, Bukit Batok, Clementi, Pasir Ris, Queenstown
5	Kallang/Whampoa

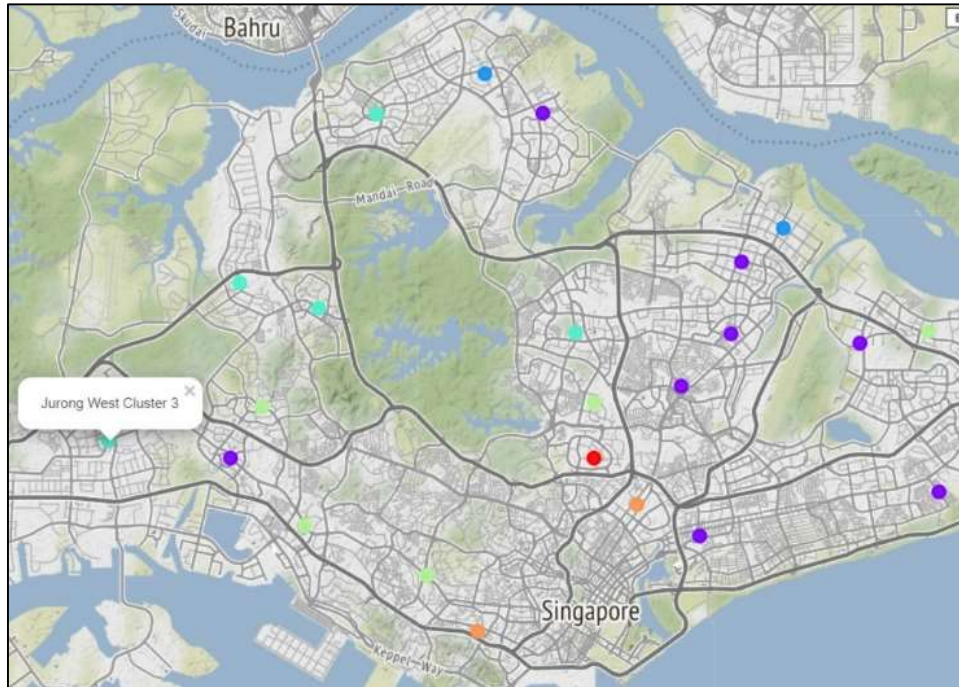


Figure 5 Map of Singapore with Cluster Markers

Map Legend

- Cluster 0: **Red**
- Cluster 1: **Violet**
- Cluster 2: **Blue**
- Cluster 3: **Turquoise**
- Cluster 4: **Pale Green**
- Cluster 5: **Orange**

3. Characteristics of Each Clusters

By analysing the average data obtained from the clustering machine learning (population density and hospital ward count in each town), characteristics of each clusters may be defined. The results are as in Table 2 below.

Table 2 Cluster Characteristics

Cluster	Average Population Density	Average Hospital Count	Ward	Characteristics
0	0.661177	1.000000		medium-high population density, highest hospital ward count
1	0.755956	0.058673		high population density, very low hospital ward count
2	0.075705	0.030612		lowest population density, very low hospital ward count
3	0.953241	0.000000		highest population density, no hospital ward count
4	0.480231	0.093878		medium population density, low hospital ward count
5	0.823461	0.642857		high population density, moderate ward count

4. Further Analysis

By investigating further on high priority clusters: cluster 3 and cluster 1
It is found that certain towns provide even more surprising new facts.

Those facts are reported as follows:

- Bukit Panjang has the highest population density across **Singapore** but no hospital wards.
- Jurong West has the highest population density after Bukit Panjang in cluster 3 but no hospital wards.
- All cluster 3 towns outclassed those of cluster 1 because all cluster 3 towns have higher population density and no hospital ward count.
- Hougang, Bedok, and Jurong East have the best potential from cluster 1 towns: the highest population density in cluster 1 but no hospital wards.

In order to provide preventive advices, we may also look on other clusters.
Other findings are also reported below:

- Clementi and Pasir Ris has moderate population density but zero and very low hospital ward respectively. These two towns may be prioritised after considering the high priority areas from cluster 3 and 1 first.
- Toa Payoh has the highest hospital ward count across **Singapore**. Thus, the reader is advised to set it aside.
- Sembawang has the lowest population density across **Singapore**, even though it has no hospital wards nearby. Thus, the reader is advised to set it aside.

F. Summary

Therefore, we can summarise the findings from this investigation as priorities to develop each cluster:

Table 3 Priority Recommendation

Priority	Clusters	Special Note
High	3, 1	Cluster 3 should be prioritised first, in particular Bukit Panjang and Jurong West towns.
Medium	4	Clementi and Pasir Ris are the better towns from Cluster 4.
Low	5, 2	Sembawang from cluster 2 should be set aside as this town has the lowest population density in Singapore.
Zero	0	Toa Payoh should be avoided as this town has the most hospital wards in Singapore.

References:

- <https://www.statista.com/statistics/378424/average-age-of-the-population-in-singapore/>
- <https://www.statista.com/statistics/378566/age-structure-in-singapore/>
- https://hospitals.webometrics.info/en/Asia_Pacifico/South%20East%20Asia
- <https://mothership.sg/2016/10/nifty-map-shows-which-parts-of-spore-has-densest-population-guess-which-is-no-1/>
- https://en.wikipedia.org/wiki/New_towns_of_Singapore
- <https://www.moh.gov.sg/>
- <https://www.coursera.org/learn/applied-data-science-capstone/home/welcome>
- <https://docs.python.org/3/>
- <https://pandas.pydata.org/docs/>
- <https://numpy.org/doc/>
- <https://developer.foursquare.com/>
- <https://matplotlib.org/3.1.3/contents.html>
- <https://scikit-learn.org/stable/>
- <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- <https://lxml.de/>
- <https://geocoder.readthedocs.io/>
- <https://www.scikit-yb.org/en/latest/quickstart.html>
- <https://python-visualization.github.io/folium/>