

# IBM Data Science Capstone Project: Analysis of Singapore's Premium Product Shop Placement

Kar Wai Lim

January 5, 2020

## Executive Summary

1. More than 70% of Singapore residents purchase premium products in store rather than online.
2. Using condominium rental data in Singapore, I find out the areas where Singapore residents are paying more for housing rental.
3. I use foursquare API to look for potential competitors in premium products.
4. Clustering algorithm is used to cluster 174,780 rental data into 100 groups based on location and rental per room.
5. Illustration using folium package allows us to determine the best location — where there are many rich residents but little competitors.

## 1 Introduction

Despite its small size, Singapore is one of the richest country in the world with more than 180,000 millionaires residing in the little red dot.<sup>1</sup> Interestingly, a market survey by The Nielsen Company shows that more than 70% of Singapore residents are purchasing premium products from local store rather than online.<sup>2</sup> This provides a great opportunity for businesses to set up physical premium stores in Singapore.

In this report, I will leverage the foursquare API to look for existing premium stores in Singapore and use condominium rental data to determine

---

<sup>1</sup><https://www.straitstimes.com/business/economy/number-of-millionaires-in-singapore-up-11-to-183737-in-year-to-mid-2018-credit>

<sup>2</sup><https://www.nielsen.com/sg/en/insights/article/2019/popularity-of-the-premium/>

potential areas for new businesses to set up their premium stores. This report will be of interest to premium stores retailer such as Hermes and Louis Vuitton.

## 2 Data

I will be using two kinds of data in this report:

1. Foursquare location data
2. Condominium rental data from Urban Redevelopment Authority (URA)<sup>3</sup>

### 2.1 Foursquare location data

In particular, I will only use the venues data from Foursquare, which is obtainable simply by using API calls to Foursquare. While the details can be found in the accompanying jupyter notebook,<sup>4</sup> I briefly describe the querying procedure.

Note that Foursquare limits the number of results returned from each API calls to 50, which is not enough to cover the whole Singapore. Thus, I use a grid based approach to split Singapore into 49 (7 X 7) regions and run an API call to each region (different latitude and longitude). After removing the duplicates, I obtain a total of 239 premium stores, some of these are located in Malaysia and Indonesia due to the use of large query radius. I left them in the dataset as it does not affect our analysis.

For simplicity, I use only the keywords ‘Premium’ and ‘Luxury’, though a more in depth studies would cover the major premium retailers in Singapore. More effort would be required in the additional steps of cleaning the data.

### 2.2 Condominium rental data

I queried the condominium data from the URA website to obtain the monthly rental associated with condominium units rented out for the past few years. The data consists of condominium name, street address, number of bedrooms, rental amount, lease date, and area. Additionally, I computed an additional row corresponding to the rental amount per number of bedrooms, as I believe this is a strong indicator to the willingness to purchase premium products (we need more data to verify this, but for the purpose of the course, we will assume this is the case).

---

<sup>3</sup><https://www.ura.gov.sg>

<sup>4</sup>[https://github.com/karwailim/Courser\\_Capstone/blob/master/Capstone\\_Project.ipynb](https://github.com/karwailim/Courser_Capstone/blob/master/Capstone_Project.ipynb)

I use Google map API to gather the condominiums latitude and longitude. Further cleaning is performed to remove the data that falls outside of Singapore (due to similar street address, missing data from Google map API, etc).

## **A   Appendix here**