

# Saving my rental in Singapore by applying Machine Learning

Le Bui Hung

May 16, 2019

## 1. Introduction

### 1.1. Background

Despite a fall of US\$500 (S\$678) in average monthly rental prices, rents paid by expatriates in Singapore are still among the top 10 highest in Asia, a study by ECA International has shown.

According to the expatriate management and human resource consultancy, the average rental price of an unfurnished, mid-market, 3-bedroom apartment in areas commonly inhabited by international executives in Singapore was US\$4,215 a month in September 2018.

Top 10 most expensive locations for expatriate rent – Asia

Location	2019 ranking	2018 ranking	Average monthly rent (USD)*
Hong Kong	1	1	10, 929
Tokyo, Japan	2	2	8, 668
Shanghai, China	3	3	5, 305
Seoul, Korea Republic	4	4	5, 245
Yokohama, Japan	4	7	4, 865
Beijing, China	6	6	4, 638
Mumbai, India	7	5	4, 342
Singapore	8	8	4, 215
Osaka, Japan	9	9	4, 061
Bangkok, Thailand	10	11	3, 880

There are many districts in Singapore (e.g Raffles Place, Cecil, Marina, People's Park, etc.). The list of Singapore district can be found here:

[https://en.wikipedia.org/wiki/Postal\\_codes\\_in\\_Singapore](https://en.wikipedia.org/wiki/Postal_codes_in_Singapore)

## 1.2. Problem

We as a young family really enjoy Tiong Bahru which is the oldest housing estate in Singapore.

When you visit this quaint community, you'll immediately notice an intriguing mix of old and new. While many older residents have lived in the shophouses and flats here for decades, the younger generation flocks here for the food, shops, and culture.

You can find more information about Tiong Bahru [here](#)

We want to live there but renting an apartment there is very expensive.

## 1.3. Interest

We need to look for another district which is similar to Tiong Bahru but the rental is cheaper. In order to do that, I will segment districts into different clusters based on Point of Interests (e.g. coffee shop, bakery, etc.)

In addition, I'm a frequent traveller. Thus, I want to stay somewhere not very far from the Changi airport.

## 2. Data acquisition and preparation

### 2.1. Data Sources

The list of districts can be found [here](#)

Postal district	Postal sector (1st 2 digits of 6-digit postal codes)	General location
01	01, 02, 03, 04, 05, 06	Raffles Place, Cecil, Marina, People's Park
02	07, 08	Anson, Tanjong Pagar
03	14, 15, 16	<a href="#">Bukit Merah, Queenstown, Tiong Bahru</a>
04	09, 10	Telok Blangah, Harbourfront
05	11, 12, 13	Pasir Panjang, Hong Leong Garden, Clementi New Town
06	17	High Street, Beach Road (part)
07	18, 19	Middle Road, Golden Mile
08	20, 21	<a href="#">Little India, Farrer Park, Jalan Besar, Lavender</a>
09	22, 23	Orchard, Cairnhill, River Valley
10	24, 25, 26, 27	Ardmore, Bukit Timah, Holland Road, Tanglin
11	28, 29, 30	Watten Estate, Novena, Thomson

## 2.2. Data Preparation

Data is downloaded and loaded into a data frame

Print out the result

```
In [154]: df.head()
```

```
Out[154]:
```

	postalDistrict	Neighbourhood
0	01	Raffles Place
1	01	Cecil
2	01	Marina
3	01	People's Park
4	02	Anson

The data is quite cleaned and organized. Thus, there is data cleaning required.

We will use the Geopy toolkit to get the latitude and longitude for each district

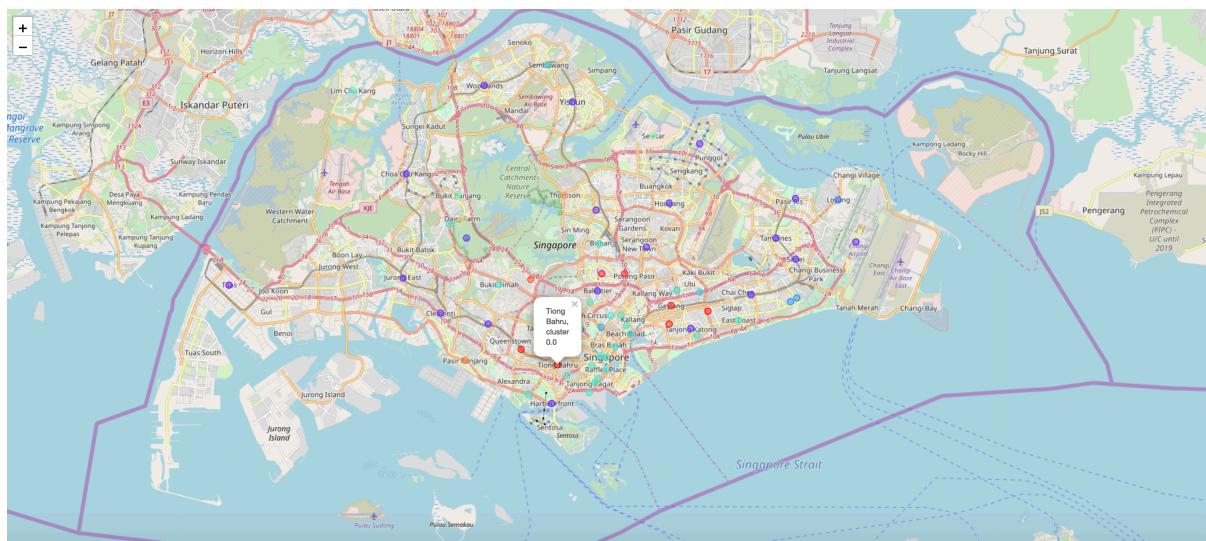
Print out the result - We should see the longitude and latitude

```
In [157]: df1.head()
```

```
Out[157]:
```

	postalDistrict	Neighbourhood	longitude	latitude
0	01	Raffles Place	103.851568	1.283595
1	01	Cecil	103.848094	1.279265
2	01	Marina	103.852036	1.290475
3	01	People's Park	103.844160	1.285810
4	02	Anson	103.846731	1.276088

A map is also created to visualize the input



From there we will use Foursquare service to get all the point of interests in each neighbourhood so that we can do our analytics

	<b>Neighbourhood</b>	<b>Neighbourhood Latitude</b>	<b>Neighbourhood Longitude</b>	<b>Venue</b>	<b>Venue Latitude</b>	<b>Venue Longitude</b>	<b>Venue Category</b>
<b>0</b>	Ang Mo Kio	1.369842	103.846609	Old Chang Kee	1.369094	103.848389	Snack Place
<b>1</b>	Ang Mo Kio	1.369842	103.846609	MOS Burger	1.369170	103.847831	Burger Joint
<b>2</b>	Ang Mo Kio	1.369842	103.846609	Subway	1.369136	103.847612	Sandwich Place
<b>3</b>	Ang Mo Kio	1.369842	103.846609	Bun Master	1.369242	103.849031	Bakery
<b>4</b>	Ang Mo Kio	1.369842	103.846609	NTUC FairPrice	1.371507	103.847082	Supermarket
<b>5</b>	Ang Mo Kio	1.369842	103.846609	FairPrice Xtra	1.369279	103.848886	Supermarket
<b>6</b>	Ang Mo Kio	1.369842	103.846609	Kam Jia Zhuang Restaurant	1.368167	103.844118	Asian Restaurant
<b>7</b>	Ang Mo Kio	1.369842	103.846609	PLAYe	1.369109	103.848225	Hobby Shop
<b>8</b>	Ang Mo Kio	1.369842	103.846609	POSB Ang Mo Kio Central Branch	1.370969	103.847341	Bank
<b>9</b>	Ang Mo Kio	1.369842	103.846609	Malaysia Boleh!	1.369441	103.849066	Food Court
<b>10</b>	Ang Mo Kio	1.369842	103.846609	Face Ban Mian 非板面 (Ang Mo Kio)	1.372031	103.847504	Noodle House
<b>11</b>	Ang Mo Kio	1.369842	103.846609	Collin's Grille . Bento	1.371713	103.847526	Modern European Restaurant
<b>12</b>	Ang Mo Kio	1.369842	103.846609	Express Teppan-Yaki	1.369137	103.848232	Japanese Restaurant
<b>13</b>	Ang Mo Kio	1.369842	103.846609	Mr Teh Tarik	1.372168	103.845602	Halal Restaurant
<b>14</b>	Ang Mo Kio	1.369842	103.846609	Starbucks	1.369230	103.848683	Café
<b>15</b>	Ang Mo Kio	1.369842	103.846609	True Fitness	1.372891	103.847661	Gym

We then group them by neighbourhood

	<b>Neighbourhood</b>	<b>1st Most Common Venue</b>	<b>2nd Most Common Venue</b>	<b>3rd Most Common Venue</b>	<b>4th Most Common Venue</b>	<b>5th Most Common Venue</b>	<b>6th Most Common Venue</b>	<b>7th Most Common Venue</b>	<b>8th Most Common Venue</b>	<b>9th Most Common Venue</b>	<b>10th Most Common Venue</b>
<b>0</b>	Amber Road	Hotel	Indian Restaurant	Café	Chinese Restaurant	Pub	Japanese Restaurant	Bakery	Dessert Shop	Bar	Noodle House
<b>1</b>	Ang Mo Kio	Coffee Shop	Fast Food Restaurant	Dessert Shop	Bubble Tea Shop	Food Court	Japanese Restaurant	Supermarket	Sandwich Place	Modern European Restaurant	Seafood Restaurant
<b>2</b>	Braddell	Seafood Restaurant	Gym / Fitness Center	Zoo Exhibit	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop	Food	Flower Shop
<b>3</b>	Bukit Panjang	Food Court	Park	Fruit & Vegetable Store	Noodle House	Miscellaneous Shop	Market	Food Stand	Food Truck	Food & Drink Shop	Farm
<b>4</b>	Bukit Timah	Trail	Mountain	Rest Area	Food	Farm	Fast Food Restaurant	Filipino Restaurant	Flea Market	Flower Shop	Food & Drink Shop

### 3. Neighbourhood Segmentation

We will use K-mean to segment the list into 20 clusters

```
In [170]: # Set number of clusters
k_clusters = 20

singapore_grouped_clustering = singapore_grouped.drop("Neighbourhood", 1)

# Run K-means clustering
kmeans = KMeans(n_clusters = k_clusters, random_state = 0).fit(singapore_grouped_clustering)

# Check cluster labels generated for each row in the dataframe
kmeans.labels_[0:20]

Out[170]: array([ 7, 19,  6, 10,  1,  7,  9,  9, 19, 19, 17, 19, 16,  7,  9,  7,  3, 19, 7, 14], dtype=int32)

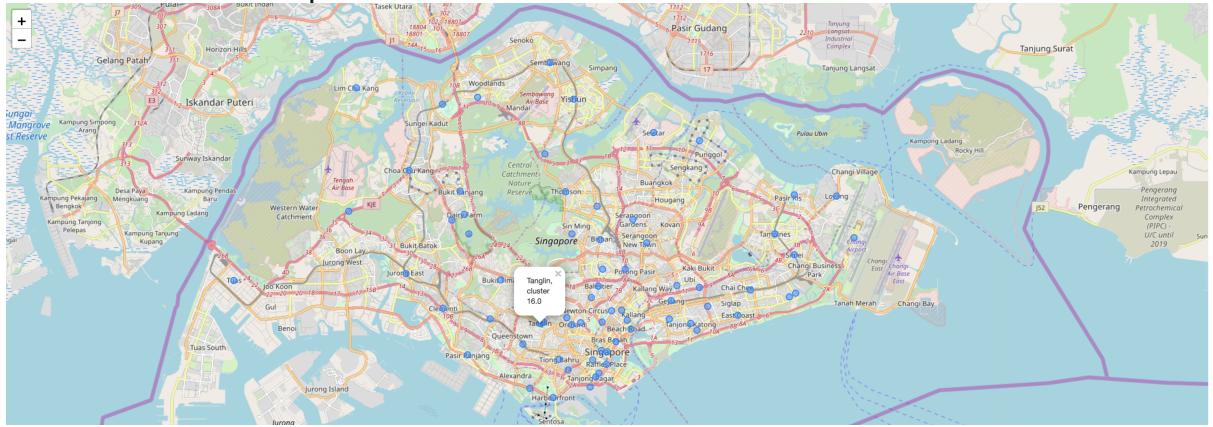
In [171]: # Add clustering labels
singapore_venues_sorted.insert(0, "Cluster Labels", kmeans.labels_)

singapore_venues_sorted.head()

# Merge Toronto grouped data with toronto data to add latitude and longitude for each borough
#
#singapore_merged.sample(8)

Out[171]: Cluster Labels Neighborhood 1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue 6th Most Common Venue 7th Most Common Venue 8th Most Common Venue 9th Most Common Venue 10th Most Common Venue
0 7 Amber Road Hotel Indian Restaurant Café Chinese Restaurant Pub Japanese Restaurant Bakery Dessert Shop Bar Noodle House
1 19 Ang Mo Kio Coffee Shop Fast Food Restaurant Dessert Shop Bubble Tea Shop Food Court Japanese Restaurant Supermarket Sandwich Place Modern European Restaurant Food Seafood Restaurant
2 6 Braddell Seafood Restaurant Gym / Fitness Center Zoo Exhibit French Restaurant Food Truck Food Stand Food Court Food & Drink Shop Flower Shop
3 10 Bukit Panjang Food Court Park Fruit & Vegetable Store Noodle House Miscellaneous Shop Market Food Stand Food Truck Food & Drink Shop Farm
4 1 Bukit Timah Trail Mountain Rest Area Food Farm Fast Food Restaurant Filipino Restaurant Flea Market Flower Shop Food & Drink Shop
```

And also create a map



From the result, we can clearly see Tiong Bahru is very similar to Geylang

```
In [179]: singapore_merged.loc[singapore_merged["Neighbourhood"] == "Tiong Bahru",
singapore_merged.columns[1] + list(range(5, singapore_merged.shape[1]))]

Out[179]: Neighborhood Cluster Labels 1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue 6th Most Common Venue 7th Most Common Venue 8th Most Common Venue 9th Most Common Venue 10th Most Common Venue
8 Tiong Bahru 14 Chinese Restaurant Coffee Shop Food Court Noodle House Bakery Café Bar Fried Chicken Joint Electronics Store Hainan Restaurant

As you can see Novena is in Cluster 15 (index = 14)
Let's narrow down to top 5 Most common venue

In [181]: singapore_merged.loc[singapore_merged["Cluster Labels"] == 14,
singapore_merged.columns[1] + list(range(5, 10))]

Out[181]: Neighborhood 1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
7 Queenstown Chinese Restaurant Noodle House Pool Stadium Seafood Restaurant
8 Tiong Bahru Chinese Restaurant Coffee Shop Food Court Noodle House Bakery
21 River Valley Chinese Restaurant Noodle House Food Court Bus Station Asian Restaurant
34 Geylang Chinese Restaurant Noodle House Vegetarian / Vegan Restaurant Food Court Dim Sum Restaurant
37 Joo Chiat Seafood Restaurant Grocery Store Asian Restaurant Malay Restaurant Food Court
```

### 4. Conclusion

By apply simple clustering, below neighbourhoods are similar to Tiong Bahru:

- Queentown
- River Valley
- Toa Payoh
- Geylang
- Joo Chiat

With the help of google I can see that Geylang will be a good choice as:

- It offers cheap rental (vs other neighborhood)
- Short distance to Changi Airport