

# IBM Applied Data Science Capstone

The Battle of Neighborhoods:

How Facilities Influence Singapore's HDB Pricing

# Introduction

- Singapore is a small city state with a land area of 728.6 km<sup>2</sup>, housing a total population of 5.76 million. This makes Singapore one of the most densely populated countries in the world with a population density of 8,041 ppl/km<sup>2</sup>. 81% of Singapore's population resides in public housing, known as Housing Development Board (HDB) flats. These flats are situated in 26 towns within Singapore.
- This project intends to study the neighbourhood facilities favoured by the population residing in HDB flats. This will be done by studying the price per square feet of HDB flats in the town versus the facilities available in these towns.

# Data acquisition

## Resale Flat Price Prices Based on Registration Date

Data will be retrieved from the Singapore Governments open dataset on Resale Flat Price Prices Based on Registration Date from data.gov.sg website.

Raw data retrieved from the government's dataset includes list of all transactions in all towns from 2017 onwards. Details on flat type, area of house, model, lease commence date and resale price are also provided for each transaction. There are a total of 103,221 entries.

## Singapore Town Location Data

Data coordinated of Towns are retrieved using GeoPy.

# Data acquisition

## Singapore Town Top Venue Recommendations

FourSquare API is used to retrieve the most common venue categories in each Town. This feature is used to group the neighbourhoods into clusters. The following data are retrieved on the first query:

- Venue ID
- Venue Name
- Coordinates
- Category Name

# Data Cleaning

## Resale Flat Price Prices Based on Registration Date

First data is grouped by town, and the block and street attributes are dropped as they are of no value to the assessment. Remaining lease attribute is calculated from taking the current year (2021) minus the lease commence date, following which the lease commence date is dropped.

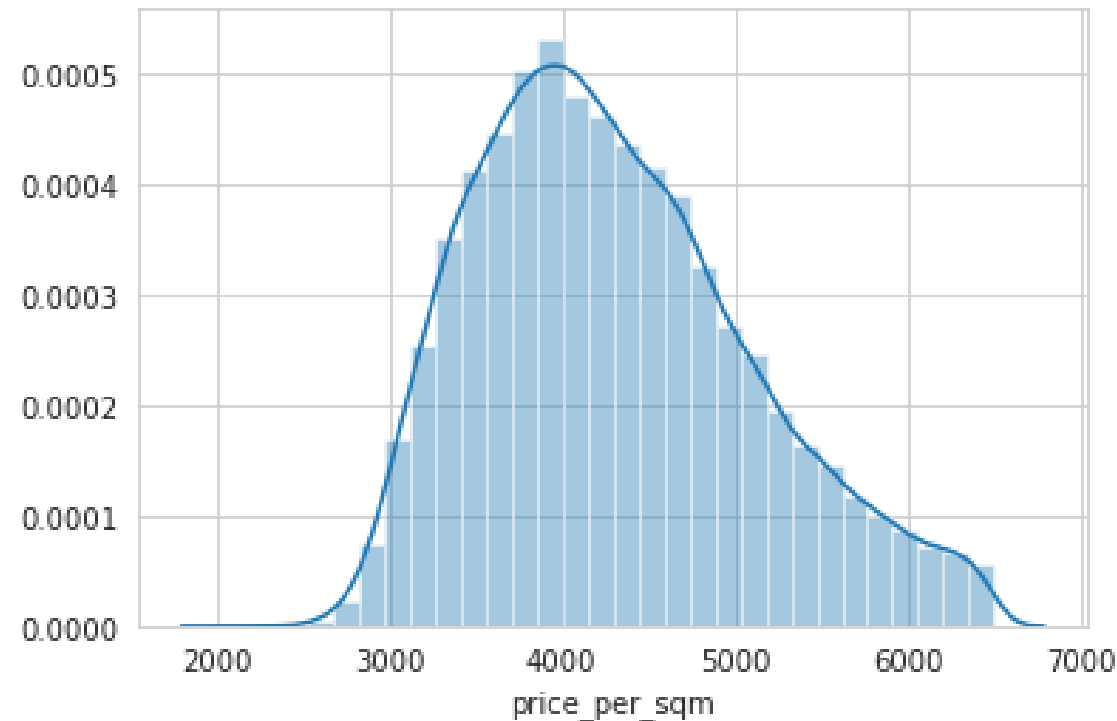
Price per square foot attribute is also calculated as the resale price divided by floor area of flat. Flat models Improved and Model A were selected and flats on higher floors were excluded as they were outliers in term of per square foot pricing. 3 Room, 4 Room and 5 Room flats were retained as they were the most prevalent. Per square foot pricing above \$6500 was also excluded to reduce outliers.

## Singapore Town Top Venue Recommendations

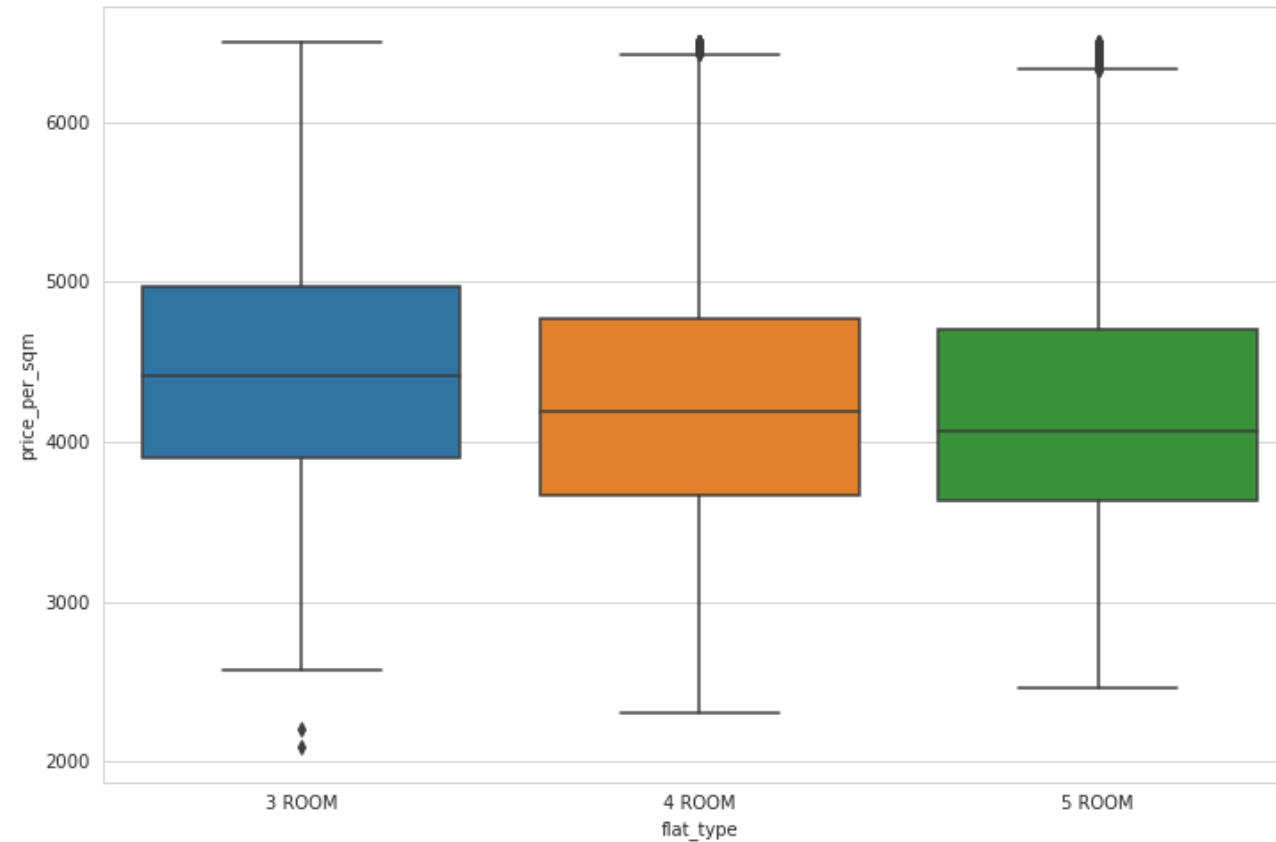
A total of 174 venue categories are retrieved from the Foursquare API. These are grouped into 9 categories, namely Casual Dining, Stores, Public Transport, Fitness, Recreation, Nature, Grocery, Bar and Amenities.

# Exploratory Data Analysis

- Price per square foot has a normal distribution. The 3 retained flat types (3 Room, 4 Room and 5 Room) have a mean square foot price of between \$4100 to \$4400.



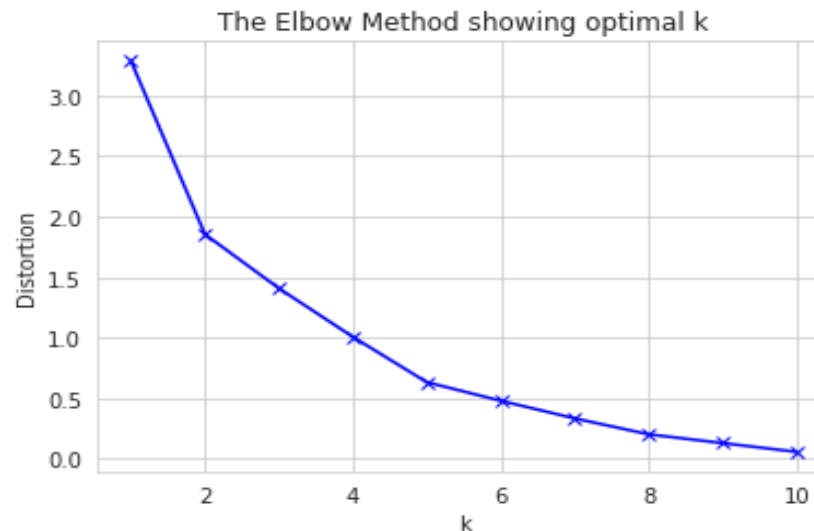
# Exploratory Data Analysis



# Results and Discussion

Foursquare API was used to acquire the venues at the different towns. These venues are grouped into 9 categories, namely Casual Dining, Stores, Public Transport, Fitness, Recreation, Nature, Grocery, Bar and Amenities.

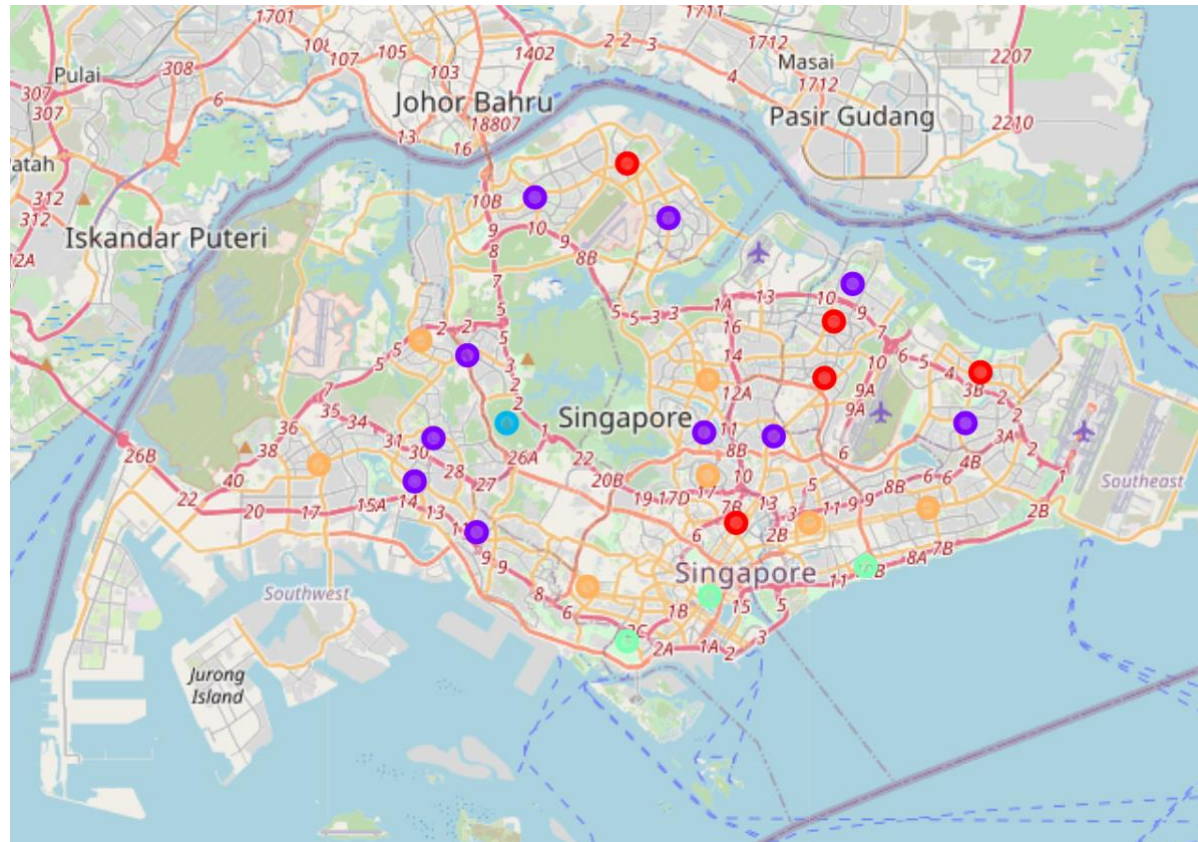
K-means clustering algorithm was used to group the towns into clusters based on the venue categories. The elbow method is used to determine the optimum value of K to perform K-means clustering. 5 clusters were selected for the evaluation.





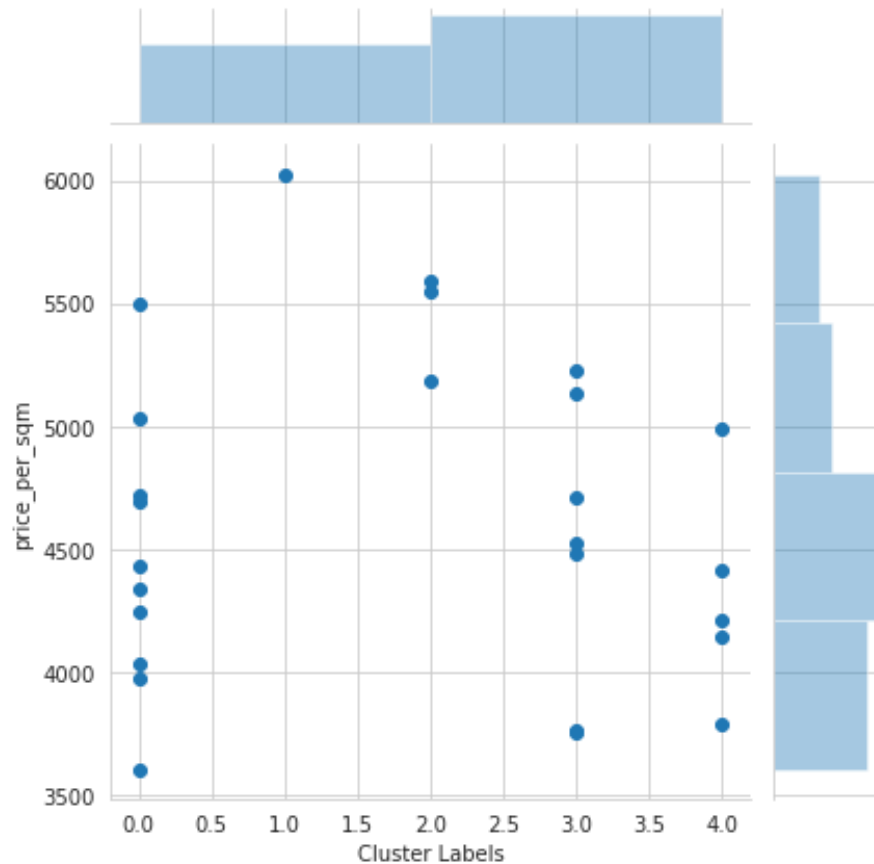
# Results and Discussion

Singapore Map Clusters:



# Results and Discussion

Cluster Labels versus Price per Square Foot:



Cluster 1 commands the highest price per square foot. This is followed by Cluster 2. Cluster 0, 3 and 4 have a mix of higher and lower pricing.

# Results and Discussion

## Cluster 0

	price_per_sqm	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 the most common venue	5 the most common venue
2	5497.0	Casual Dining	Stores	Grocery	Recreation	Nature
3	4032.0	Casual Dining	Recreation	Stores	Amenities	Bar
5	4247.0	Casual Dining	Stores	Grocery	Bar	Recreation
9	5028.0	Casual Dining	Stores	Recreation	Fitness	Bar
12	4334.0	Casual Dining	Stores	Recreation	Bar	Grocery
17	4722.0	Casual Dining	Stores	Recreation	Grocery	Nature
21	4693.0	Casual Dining	Stores	Nature	Grocery	Recreation
22	4433.0	Casual Dining	Stores	Fitness	Recreation	Grocery
24	3606.0	Casual Dining	Stores	Recreation	Grocery	Nature
25	3978.0	Casual Dining	Stores	Public Transport	Grocery	Recreation

# Results and Discussion

## Cluster 1

	price_per_sqm	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 the most common venue	5 the most common venue
6	6024.0	Nature	Recreation	Amenities	Bar	Grocery

## Cluster 2

	price_per_sqm	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 the most common venue	5 the most common venue
4	5186.0	Casual Dining	Nature	Recreation	Bar	Fitness
7	5545.0	Casual Dining	Recreation	Bar	Stores	Amenities
15	5592.0	Casual Dining	Recreation	Amenities	Stores	Bar

# Results and Discussion

## Cluster 3

	price_per_sqm	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 the most common venue	5 the most common venue
0	5132.0	Casual Dining	Stores	Grocery	Fitness	Amenities
1	4486.0	Casual Dining	Stores	Recreation	Bar	Grocery
8	3765.0	Casual Dining	Public Transport	Grocery	Recreation	Fitness
10	4711.0	Casual Dining	Stores	Recreation	Fitness	Amenities
13	3757.0	Casual Dining	Stores	Recreation	Amenities	Grocery
18	5223.0	Casual Dining	Recreation	Fitness	Public Transport	Amenities
23	4520.0	Casual Dining	Stores	Grocery	Recreation	Amenities

# Results and Discussion

## Cluster 4

	price_per_sqm	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4 the most common venue	5 the most common venue
11	4207.0	Casual Dining	Stores	Public Transport	Grocery	Fitness
14	4991.0	Casual Dining	Amenities	Stores	Grocery	Bar
16	4145.0	Casual Dining	Public Transport	Stores	Grocery	Nature
19	3788.0	Casual Dining	Stores	Grocery	Fitness	Amenities
20	4412.0	Casual Dining	Stores	Amenities	Grocery	Nature

From the breakdown above, the most distinct clusters are Cluster 1 and Cluster 2. Cluster 1 has nature and recreation as the 1<sup>st</sup> and 2<sup>nd</sup> most common venues. Cluster 2 has mix of nature and recreation as the 2<sup>nd</sup> most common venues. These two clusters also have the highest price per square foot.

# Conclusion

In conclusion, it can be derived from the high price per square foot for housing in Cluster 1 and 2 that there is a premium placed on nature and recreation facilities. These facilities are not as common in the other clusters. These two facilities have a significant influence in the pricing of the HDB flats in their area, compared to other venue categories.