

Capstone Project - Where to rent in Singapore?

Sam Williams

2019

Contents

Capstone Project - Where to rent in Singapore?	1
Introduction/Business Problem	2
Data.....	2
Foursquare location data (https://developer.foursquare.com/)	2
Data.gov.sg (https://data.gov.sg/)	3
Urban Redevelopment Authority (https://www.ur.gov.sg)	4
Methodology.....	4
Exploratory Data Analysis	4
Machine Learning.....	8
Results	8
Cluster 0 - Centre-East	9
Cluster 1 – Central Hotel	9
Cluster 2 – Outer reaches	9
Cluster 3 – Main Core.....	9
Cluster 4 – Green Life.....	10
Cluster 5 – Mid City.....	10
Discussion.....	10
Conclusion.....	10

Introduction/Business Problem

Singapore is a city state located at the southern tip of the Malay Peninsula approximately one degree north of the equator. Due to its location it has a tropical climate with uniform high temperatures and humidity, rainfall is abundant particularly in the two notable monsoon seasons.

Singapore is one of the global financial centres having a highly developed market economy based on its strategic location as a trading port within Asia. It retains a large amount of foreign investment as a result of its location, skilled workforce, low tax rates, advanced infrastructure and zero-tolerance against corruption. This promotes Singapore to be consistently ranked highly in best places to live and work.

The 5.64 million (approx.) multiracial and multicultural population of Singapore is made from 4 million residents and 1.64 million non-residents. This relates to approximately 44 percent of the workforce being non-Singaporean.

Despite Singapore growing its land mass through land reclamation from surrounding waters its limited size has caused property prices and developments to substantially increase in price. Combined with the laws regulating property ownership many non-Singaporean, non-residents rent during their time in Singapore. Ownership is allowed of condominium properties although pricing generally deters this option.

Prior to arrival and starting work in the city state many worldwide renters approach the internet and limited resources attempting a wide variety of research in consideration of which location to settle in. This can often lead to confusion and missed opportunities for a better homelife. This Project attempts to relieve the pressure of the 1.64 million renters and provide a structured answer to which neighbourhood (or district as they are referred in Singapore) they can focus property search efforts.

Data

The data used for this project will segment Singapore into its 28 postal districts for comparison. The use of postal codes is a general method adopted within Singapore to categorise properties and services; it is therefore deemed a suitable method to answer the main project query. Each of the districts will then be compared and clustered using venues, places, transportation and rental rates to determine outcomes of areas to consider.

Specific datasets used are:

Foursquare location data (<https://developer.foursquare.com/>)

Data Provider

Back in 2009, Foursquare invented the check-in. Today, those 13+ billion check-ins are the foundation of our powerful, proprietary Pilgrim technology that helps make sense of where phones go for the more than 150,000 partners who have registered to use our developer tools, including Uber, Tencent, Apple, Samsung, and Twitter. (extract from website <https://foursquare.com/about>)

Data Set

Location data provided by calls to Foursquare API based on a specific location (latitude, longitude) to demonstrate Amenities, Transport, Schools, Places, Venues etc.

<https://developer.foursquare.com/docs/api/endpoints>

Data.gov.sg (<https://data.gov.sg/>)

Data Provider

Data.gov.sg was first launched in 2011 as the government's one-stop portal to its publicly available datasets from 70 public agencies. (extract from website <https://data.gov.sg/>)

Data Set

Rentals of Non-Landed Residential Buildings, Quarterly

<https://data.gov.sg/dataset/rentals-of-non-landed-residential-buildings-with-at-least-10-rental-contracts-signed-in-1q2016>

This listing contains summary statistics (median, 25th percentile, 75th percentile, and number of rental contracts) of major non-landed private residential projects with at least 10 rental contracts signed in a quarter.

Whilst this does not contain minor developments it is deemed appropriate in answering the projected question as smaller developments are a low consideration and potentially not open to the target audience. In addition, minor development details are not readily available without cost.

This data is available in csv format and direct from website data API.

Schema

Name	Type	Keep/Drop	Reason
Quarter	String	Drop	Period holds no relevance for initial area selection
Project Name	String	Keep	
Postal District	String	Keep	
25th Percentile (S\$ psm pm)	Float	Keep	Using for reference and checking only
Median (S\$ psm pm)	Float	Keep	Focus
75th percentile (S\$ psm pm)	Float	Keep	Using for reference and checking only
Rental Contracts (Count)	Float	Drop	Total number of available units in each building is different; weighting would need to be considered

S\$ psm pm = Singapore Dollar per square meter, per month

The dataset is provided under licence:

[Contains information from { Rentals of Non-Landed Residential Buildings, Quarterly } accessed on { 17/11/2019 } from { <https://data.gov.sg/> } which is made available under the terms of the Singapore Open Data Licence version 1.0 { <https://data.gov.sg/open-data-licence>}]

Urban Redevelopment Authority (<https://www.ura.gov.sg>)

Data Provider

At URA, we adopt a long-term and comprehensive planning approach to formulate strategic plans such as the Concept Plan and the Master Plan to guide the physical development of Singapore in a sustainable manner. Our plans and policies are focused on achieving a quality living environment for Singapore. (extract from website [https:// www.ura.gov.sg](https://www.ura.gov.sg))

Data Set

District Code and District Map of Singapore

https://www.ura.gov.sg/realEstateIIWeb/resources/misc/list_of_postal_districts.htm

A table containing the postal codes in Singapore used to create latitude and longitude details for each respective district. Where a postal district covers multiple locations the first named location will be used as its target.

The data is available in a tabular format and will be pulled using Beautiful Soup.

Schema

Name	Type	Keep/Drop	Reason
Postal District	String	Keep	
Postal Sector	String	Keep	
General Location	String	Keep	

Methodology

Exploratory Data Analysis

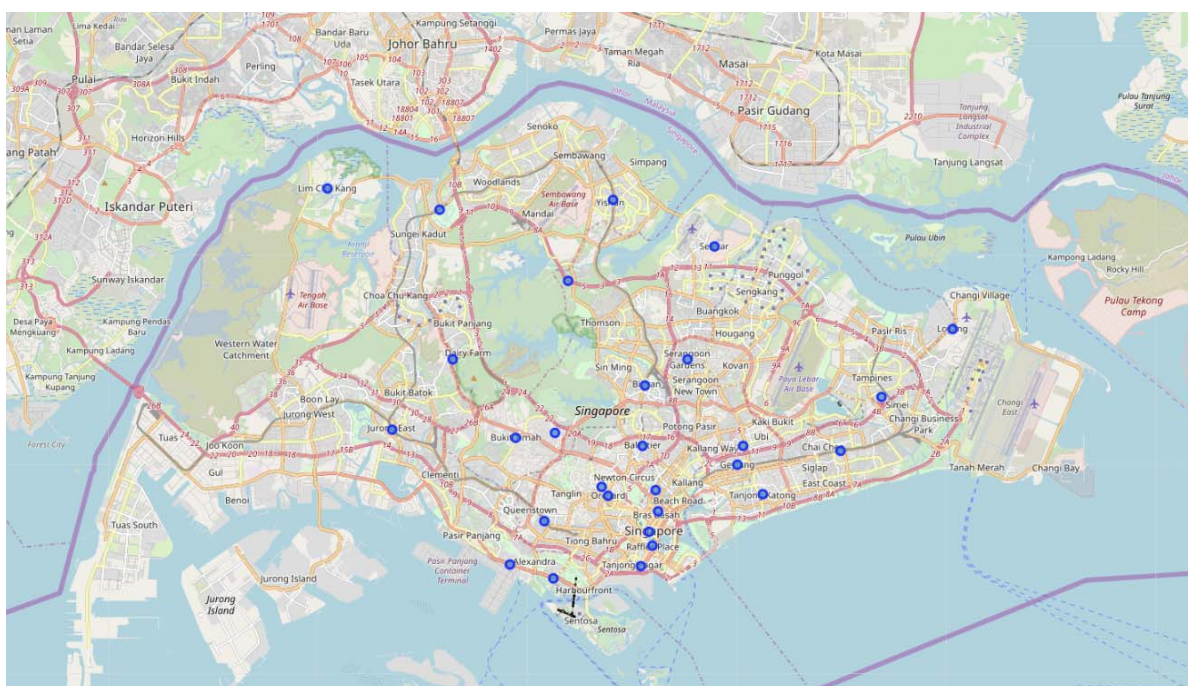
Location – Postal Districts

Initial investigation of the Postal Districts showed there was multiple General Locations quoted for each Postal District. It was decided within the cleaning only the first General location would be used as the basis for determining the Postal District latitude and longitude values by string.

List of Postal Districts

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	Queenstown, Tiong Bahru
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	Little India
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
12	31, 32, 33	Balestier, Toa Payoh, Serangoon
13	34, 35, 36, 37	Macpherson, Braddell
14	38, 39, 40, 41	Geylang, Eunos
15	42, 43, 44, 45	Katong, Joo Chiat, Amber Road
16	46, 47, 48	Bedok, Upper East Coast, Eastwood, Kew Drive
17	49, 50, 51	Loyang, Changi
18	51, 52	Tampines, Pasir Ris
19	53, 54, 55, 56	Serangoon Garden, Hougang, Ponggol
20	56, 57	Bishan, Ang Mo Kio
21	58, 59	Upper Bukit Timah, Clementi Park, Ulu Pandan
22	60, 61, 62, 63, 64	Jurong
23	65, 66, 67, 68	Hillview, Dairy Farm, Bukit Panjang, Choa Chu Kang
24	69, 70, 71	Lim Chu Kang, Tengah
25	72, 73	Kranji, Woodgrove
26	77, 78	Upper Thomson, Springleaf
27	75, 76	Yishun, Sembawang
28	79, 80	Seletar

Further graphical analysis of this in a mapped format demonstrates a good spread between the Postal District centres. It is noted that the centres also lie on or close to main road transportation and close the name locations the map provider produces. Where larger open spaces are evident there is less housing options so the removal of data is also not considered to have detrimental effect the outcome. Observations show the south west of Singapore does not have Postal Districts close by, with the closest being Jurong East. A web search determined this area is predominantly industrial type units and within consideration for rental housing.



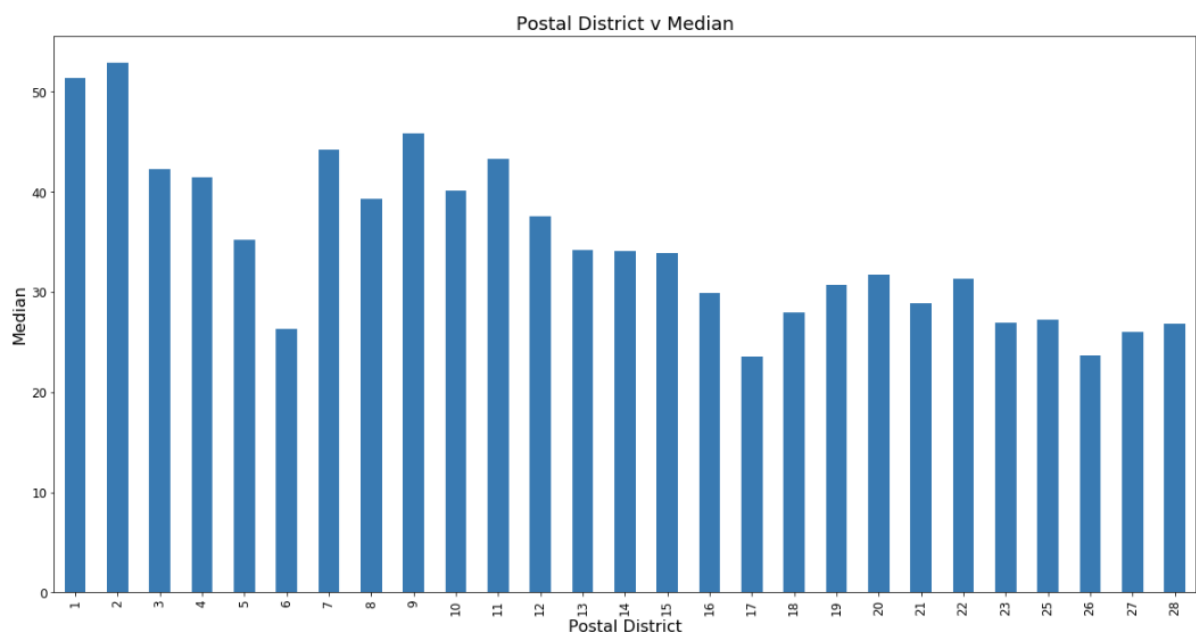
Rental Data

A simple approach to exploring the data was undertaken using the `.describe()` function. This allows quick assessment of the rental data showing there is a positive increment throughout the 25th percentile, 75th percentile and median so we do not expect and major outlying details.

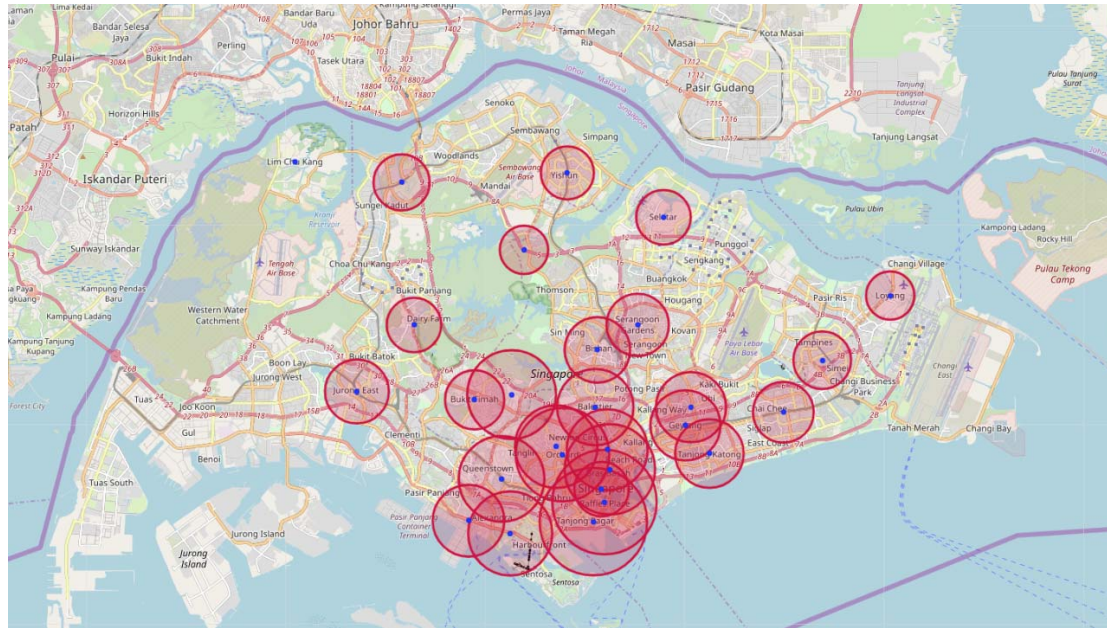
It is also decided to use the median as the focus to provide a simplified approach to further analysis. Whilst the full data gives a good range the question is to determine a suitable area to live based on multiple criteria with similar weight and not to focus heavily on the rental price.

	25th Percentile	Median	75th Percentile
count	604.000000	604.000000	604.000000
mean	33.306523	35.974156	39.159603
std	9.515605	10.339549	11.378834
min	16.260000	17.200000	18.770000
25%	26.170000	28.167500	30.575000
50%	31.710000	34.065000	37.390000
75%	39.155000	43.157500	46.135000
max	67.010000	70.710000	76.550000

Proceeding to chart the Postal District v Median exploration it is noted Postal District 24 does not contain any data.



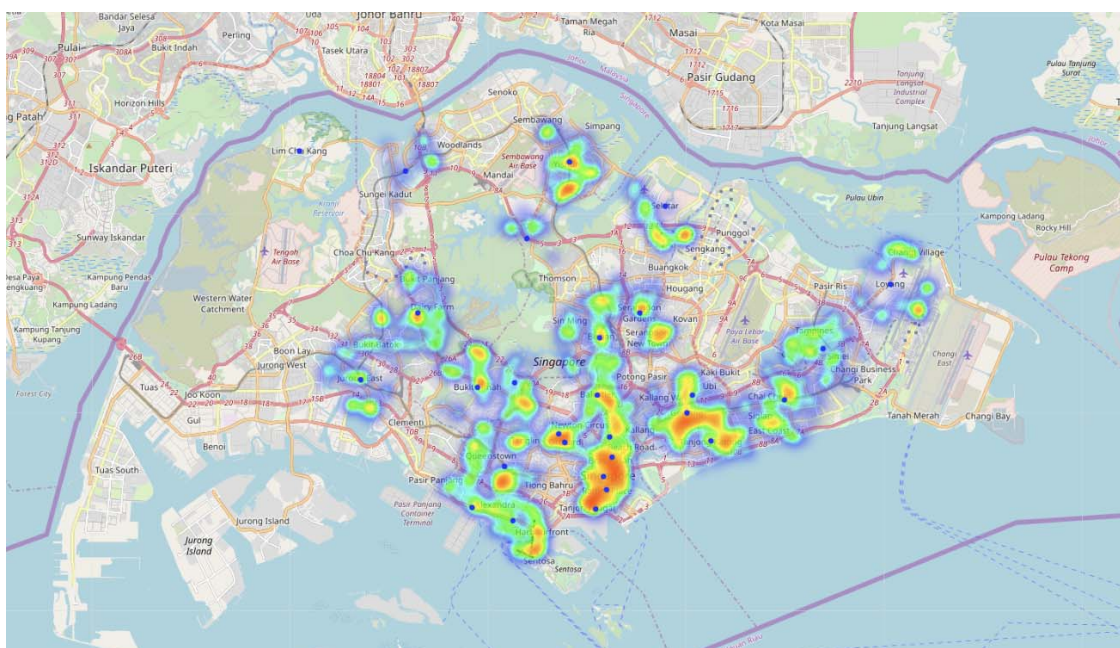
The data is then presented in a bubble plot to aid graphical representation. This clearly illustrates the central Singapore districts are most expensive, however many districts appear with very similar rates as the progress outwards.



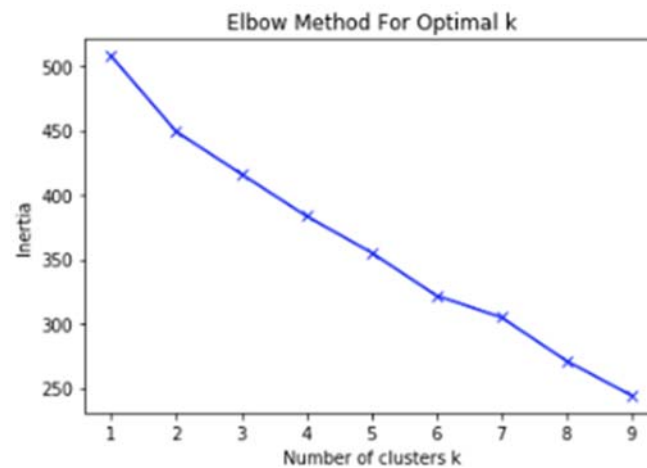
A further web search of the Lim Chu Kang area shows this to be mainly industrial, small scale farming units and nature reservations, without major rental projects. As this does not hold major relevance to the question it is deemed suitable to remove this Postal District from ongoing analysis.

Foursquare data

Initial pulling of data using 1000m from each Postal District was set which brought in a dataset of 1649 rows. Extending this to 2000m brought in 2461 rows ~70% larger. The greater quantity of data would provide a better base for analysis. In addition it was determined whilst the centres of Postal Districts are used it was not considered that all amenities, venues and rental properties were exactly central to each. In effort to better understand this data a heat map was built to locate the effective hot spots of venues imported from Foursquare data. It is seen that a denser heat is obvious in the central areas of Singapore. From the visualisation it can be construed there are two main channels running north to south and east to west.



Machine Learning

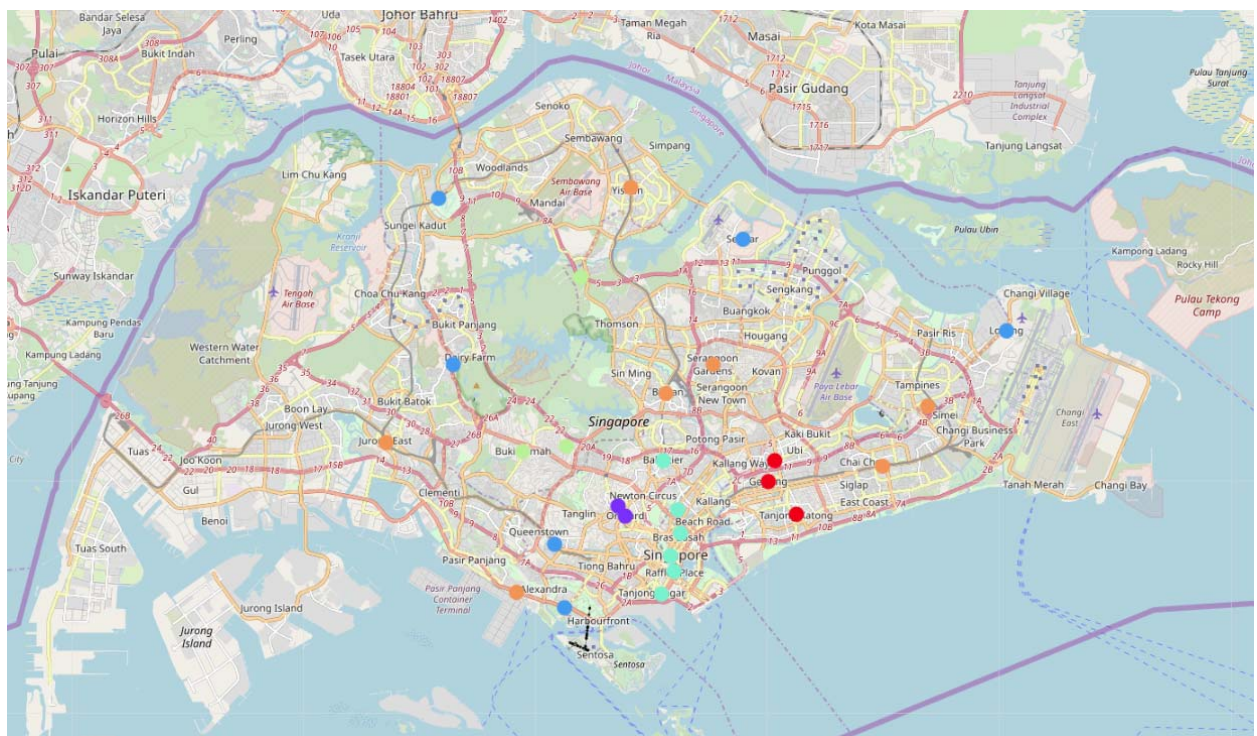


The decided machine learning is to use K-means as the form of unsupervised learning to cluster the Postal Districts into similar clusters, thus providing a set of locations to focus property search. It was the most sensible approach and answers the overall report question effectively.

The dataset was created with MinMaxScaler as the feature matrix. Using each value of k the k-means is initialised and the inertia attribute is used to identify proximity to nearest cluster centre. The following plot was achieved to show an elbow at k=6. Therefore, six clusters will be produced.

Results

The k-means clustering defines the city into six notable areas which any prospective renter can focus their search.



The following tables show the top 10 venues in each location to give prospective renters a more descriptive analysis of the main types of venues available in each district.

Cluster 0 - Centre-East

An area within close vicinity to the central business district, with a large number of eating establishments.

	General Location	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
13	Macpherson	Food Court	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	BBQ Joint	Snack Place	Noodle House	Café	Coffee Shop
14	Geylang	Chinese Restaurant	Asian Restaurant	Food Court	Seafood Restaurant	BBQ Joint	Noodle House	Vegetarian / Vegan Restaurant	Dim Sum Restaurant	Café	Vietnamese Restaurant
15	Katong	Chinese Restaurant	Hotel	Coffee Shop	Vietnamese Restaurant	Seafood Restaurant	Japanese Restaurant	BBQ Joint	Asian Restaurant	Noodle House	Multiplex

Cluster 1 – Central Hotel

These postal districts feature many hotels and boutiques, therefore expect the area to be highly catered for the tourist.

	General Location	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
9	Orchard	Hotel	Boutique	Japanese Restaurant	Shopping Mall	Coffee Shop	Bakery	Sushi Restaurant	Bubble Tea Shop	Cosmetics Shop	Chinese Restaurant
10	Ardmore	Hotel	Boutique	Japanese Restaurant	Chinese Restaurant	Sushi Restaurant	Bakery	Coffee Shop	Shopping Mall	Cosmetics Shop	Bubble Tea Shop

Cluster 2 – Outer reaches

Far reaching the bulk of these districts are on the edges of Singapore.

	General Location	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
3	Queenstown	Café	Chinese Restaurant	Coffee Shop	Park	Food Court	Indian Restaurant	Seafood Restaurant	Asian Restaurant	Ice Cream Shop	Korean Restaurant
4	Telok Blangah	Café	Chinese Restaurant	Scenic Lookout	Trail	Beach	Park	Food Court	Candy Store	Pool	Asian Restaurant
17	Loyang	Resort	Café	Asian Restaurant	Indian Restaurant	Bus Station	Waterfront	Restaurant	Coffee Shop	Grocery Store	Thai Restaurant
23	Hillview	Food Court	Trail	Fast Food Restaurant	Coffee Shop	Park	Nature Preserve	Italian Restaurant	Café	Sandwich Place	Shopping Mall
25	Kranji	Asian Restaurant	Grocery Store	Food Court	Trail	Department Store	Bar	Park	Hainan Restaurant	Bowling Alley	Go Kart Track
28	Seletar	Café	Airport	Food Court	Bus Stop	Pizza Place	Park	Restaurant	Noodle House	Bistro	Other Repair Shop

Cluster 3 – Main Core

The main core of the city again highly developed with the tourist in mind, whilst serving the attractions for locals alike.

	General Location	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
1	Raffles Place	Hotel	Waterfront	Italian Restaurant	Bar	Event Space	Cocktail Bar	Coffee Shop	Seafood Restaurant	Wine Bar	Concert Hall
2	Anson	Hotel	Japanese Restaurant	Coffee Shop	Café	Korean Restaurant	Bar	Ramen Restaurant	Spanish Restaurant	Yoga Studio	Beer Garden
6	High Street	Hotel	Japanese Restaurant	Waterfront	Restaurant	Coffee Shop	Event Space	Park	Plaza	French Restaurant	Cocktail Bar
7	Middle Road	Hotel	Café	Restaurant	Japanese Restaurant	Italian Restaurant	Event Space	Coffee Shop	Movie Theater	French Restaurant	Shopping Mall
8	Little India	Hotel	Indian Restaurant	Café	Chinese Restaurant	Restaurant	Ice Cream Shop	Italian Restaurant	Dessert Shop	Indonesian Restaurant	Whisky Bar
12	Balestier	Chinese Restaurant	Hotel	Bakery	Café	Dessert Shop	Food Court	Indian Restaurant	Coffee Shop	Ramen Restaurant	Supermarket

Cluster 4 – Green Life

Nestled around green areas the districts seem very cosmopolitan with European restaurants being higher on the list than local restaurants.

	General Location	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
11	Watten Estate	Bakery	Italian Restaurant	Japanese Restaurant	Café	Chinese Restaurant	Food Court	French Restaurant	Restaurant	Seafood Restaurant	Thai Restaurant
21	Upper Bukit Timah	Café	Italian Restaurant	Chinese Restaurant	Bakery	Pizza Place	Supermarket	Thai Restaurant	Restaurant	Soccer Field	Japanese Restaurant
26	Upper Thomson	Chinese Restaurant	Indian Restaurant	Golf Course	Café	Convenience Store	Supermarket	Gym	Lake	Thai Restaurant	Park

Cluster 5 – Mid City

This cluster feature districts in the mid distance from city centre. This is anticipated as more suburban living with a good quantity of venues available.

	General Location	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
5	Pasir Panjang	Chinese Restaurant	Coffee Shop	Food Court	Café	Indian Restaurant	Thai Restaurant	Bakery	Seafood Restaurant	Asian Restaurant	Scenic Lookout
16	Bedok	Chinese Restaurant	Coffee Shop	Food Court	Asian Restaurant	Seafood Restaurant	Noodle House	Fast Food Restaurant	Japanese Restaurant	Supermarket	Bakery
18	Tampines	Coffee Shop	Café	Dessert Shop	Fast Food Restaurant	Asian Restaurant	Noodle House	Food Court	Chinese Restaurant	Japanese Restaurant	Bakery
19	Serangoon Garden	Coffee Shop	Chinese Restaurant	Asian Restaurant	Noodle House	Food Court	Ice Cream Shop	Bakery	Dessert Shop	Café	Korean Restaurant
20	Bishan	Coffee Shop	Chinese Restaurant	Food Court	Ice Cream Shop	Seafood Restaurant	Bakery	Thai Restaurant	Asian Restaurant	Japanese Restaurant	BBQ Joint
22	Jurong	Chinese Restaurant	Coffee Shop	Food Court	Japanese Restaurant	Playground	Supermarket	Bar	Garden	Bubble Tea Shop	Vegetarian / Vegan Restaurant
27	Yishun	Coffee Shop	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Indian Restaurant	Food Court	Park	Sushi Restaurant	Fried Chicken Joint	American Restaurant

Discussion

The main venue types in Singapore are eating establishments and these far eclipse all other types. This is sensible being that there will be more restaurants than bus stops, or maybe more people wishing to share information about them.

Whilst transportation comes under venues, a better further analysis maybe to look at the flexibility of transportation links and time to travel from each point. Additionally, a greater look at the type of schooling available in each location may provide better results for potential users.

Conclusion

The report provides a basis for its intended audience to locate suitable Postal Districts to further property searches. It also helps to answer the main query and provide similar areas a user may not have initially considered. Whilst it does complete its goal a greater and more developed dataset would provide further insight to each of the districts and enhance the decision making process.