Common Venues, House Pricing and Rent in Singapore

Description and Discussion of the Background

Singapore is an island country off the southern tip of the Malay Peninsula in Southeast Asia which has a strategic location for Southeast Asian sea routes. As a foreigner who graduated from a university in Singapore recently and is trying to look for a job here, renting a house with fair price and learn more about surroundings would be something needed to be done. On the other hand, the population density in Singapore is 8358 per square kilometer which makes it one of the most crowded countries in the world. Therefore, in addition to rent and location, we also want to know the population density of the relevant area before making decision. At the end, we may also consider buy a house in the future when we are ready, although there are mainly 3 housing type HDB, Condo and apartment with totally different price distribution, regional average housing price would be a good start for us before we procced to next step.

Data Description

To consider the problem we can list the data as below:

1.Area Division: 1) Planning Areas, also known as DGP areas or DGP zones, are the main urban planning and census divisions of Singapore, therefore, I use planning areas as the division method in this project. List of Planning Areas can be found from Wikipedia. 2) The boundaries data of planning area (a KML file) is available on the Singapore official data portal: data.gov.sg. I transform the KML file to geojson format before I insert it to the code. And then use it to create choropleth map. 3) The latitude and longitude data of planning area and Singapore is obtained by using geopy library.

2.Common Venues: I use Foursquare API to get top 10 most common venues with respect to planning areas.

3.Population density: The population density of different planning areas can also be found from Wikipedia.

4.House pricing and rent: The relevant data I found come from an article on website ggg.sg, it is instructive although not quite official.

Methodology

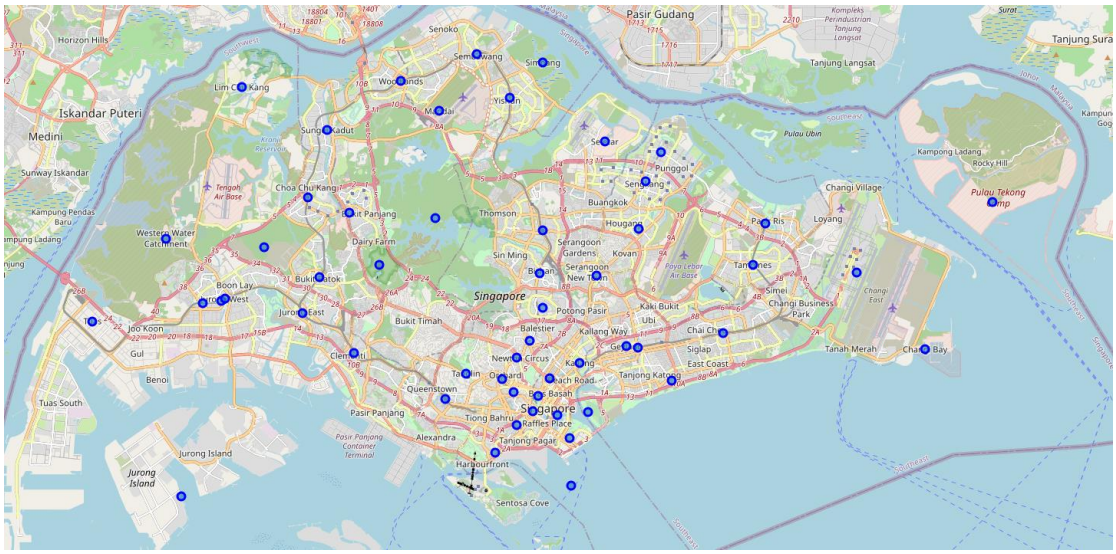
I use python to realize this project. The first step is data wrangling. After clear up the area division data from Wikipedia, the data keep the components of Planning area, Chinese (name), Region and Population density.

	Planning_Area	Chinese	Region	Population_Density/km2
0	Ang Mo Kio	宏茂桥	North-East	13400
1	Bedok	勿洛	East	13000
2	Bishan	碧山	Central	12000
3	Boon Lay	文礼	West	3.6
4	Bukit Batok	武吉巴督	West	14000
5	Bukit Merah	红山	Central	11000

Then I write a for loop to use geopy library to get the latitude and longitude values of Singapore and Planning Area and then merge them to the previous dataframe. We have our master data as below:

	Planning_Area	Chinese	Region	Population_Density/km2	Latitude	Longitude
0	Ang Mo Kio	宏茂桥	North-East	13400	1.370080	103.849523
1	Bedok	勿洛	East	13000	1.323976	103.930216
2	Bishan	碧山	Central	12000	1.350986	103.848255
3	Boon Lay	文礼	West	3.6	1.338550	103.705812
4	Bukit Batok	武吉巴督	West	14000	1.349057	103.749591
5	Bukit Merah	红山	Central	11000	1.270439	103.828318

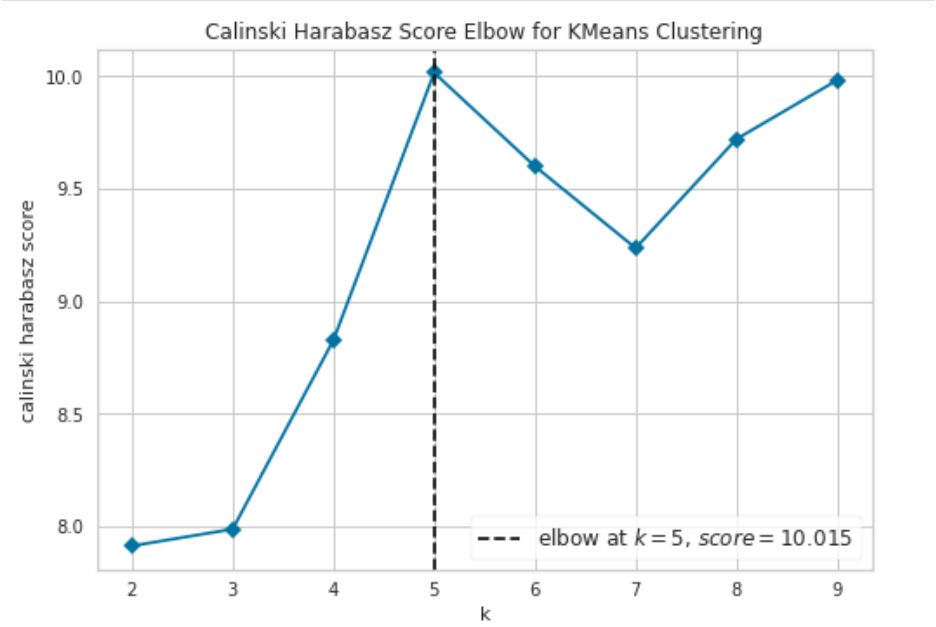
And a map of Singapore with Planning Areas superimposed on top can be created by using folium library.



I used the Foursquare API to explore the planning areas based on the latitudes and longitudes. The limit of venues is 100 and the radius is 500 meter for each planning area.

	Planning_Area	Region	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ang Mo Kio	North-East	1.37008	103.849523	Old Chang Kee	1.369094	103.848389	Snack Place
1	Ang Mo Kio	North-East	1.37008	103.849523	FairPrice Xtra	1.369279	103.848886	Supermarket
2	Ang Mo Kio	North-East	1.37008	103.849523	MOS Burger	1.369170	103.847831	Burger Joint
3	Ang Mo Kio	North-East	1.37008	103.849523	Face Ban Mian 非板面 (Ang Mo Kio)	1.372031	103.847504	Noodle House
4	Ang Mo Kio	North-East	1.37008	103.849523	A&W	1.369541	103.849043	Fast Food Restaurant
5	Ang Mo Kio	North-East	1.37008	103.849523	NTUC FairPrice	1.371507	103.847082	Supermarket
6	Ang Mo Kio	North-East	1.37008	103.849523	Subway	1.369136	103.847612	Sandwich Place
7	Ang Mo Kio	North-East	1.37008	103.849523	PLAYe	1.369109	103.848225	Hobby Shop
8	Ang Mo Kio	North-East	1.37008	103.849523	Phoon Huat & Co (Pte) Ltd	1.368318	103.851639	Miscellaneous Shop
9	Ang Mo Kio	North-East	1.37008	103.849523	Swensen's	1.369112	103.847718	Dessert Shop

One hot encoding is used then to transform the venue category to 0 & 1. So we can group the data by planning area and then proceed K-means cluster method. Where the set number of clusters k =5 are decided by using calinski harabasz elbow score.



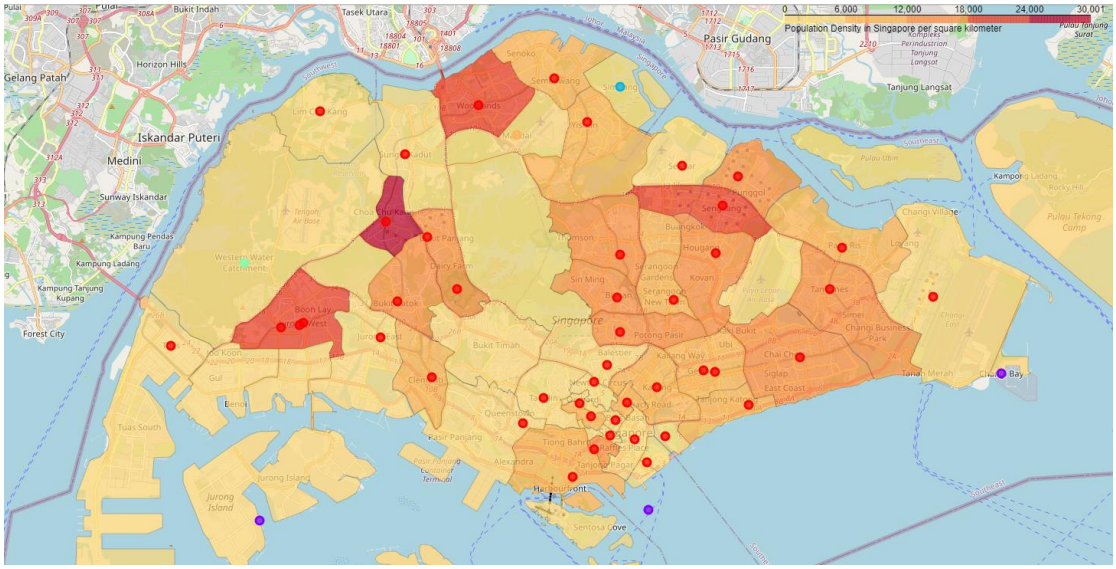
Merged table with cluster labels for each planning area is given below:

	Planning_Area	Chinese	Region	Population_Density/km2	Latitude	Longitude	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
0	Ang Mo Kio	宏茂桥	North-East	13400	1.370080	103.849523	0	Coffee Shop	Dessert Shop	Food Court	Bubble Tea Shop	Supermarket	Sandwich Place	Japanese Restaurant	Fast Food Restaurant	Gym / Fitness Center	Chinese Restaurant
1	Bedok	勿洞	East	13000	1.323976	103.930216	0	Food Court	Coffee Shop	Sandwich Place	Japanese Restaurant	Chinese Restaurant	Dessert Shop	Ice Cream Shop	Supermarket	Sushi Restaurant	Noodle House
2	Bishan	碧山	Central	12000	1.350986	103.848255	0	Food Court	Coffee Shop	Bubble Tea Shop	Café	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Supermarket	Cosmetics Shop	Ice Cream Shop
3	Boon Lay	文礼	West	3.6	1.338550	103.705812	0	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Chinese Restaurant	Dessert Shop	Coffee Shop	Supermarket	Karaoke Bar	Snack Place	Indian Restaurant
4	Bukit Batok	武吉巴督	West	14000	1.349057	103.749591	0	Coffee Shop	Fast Food Restaurant	Food Court	Chinese Restaurant	Department Store	Multiplex	Bus Station	Spa	Shopping Mall	Bowling Alley
5	Bukit Merah	红山	Central	11000	1.270439	103.828318	0	Pizza Place	Harbor / Marina	Wine Shop	Hotel	Juice Bar	Cafeteria	Zoo Exhibit	Flower Shop	Food	Food & Drink Shop
6	Bukit Panjang	武吉班让	West	15000	1.377903	103.763098	0	Sushi Restaurant	Bubble Tea Shop	Asian Restaurant	Fast Food Restaurant	Coffee Shop	Japanese Restaurant	Café	Shopping Mall	Bus Station	Supermarket
7	Bukit Timah	武吉知马	Central	4400	1.354690	103.776372	0	Trail	Scenic Lookout	Rest Area	Hill	Farmers Market	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Food Court
8	Changi	樟宜	East	80.62	1.351080	103.990064	0	Airport	Rest Area	Hotel	Bookstore	Zoo Exhibit	Fish & Chips Shop	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
9	Changi Bay	樟宜湾	East	0	1.316850	104.020649	1	Boat or Ferry	Pizza Place	History Museum	Gym Pool	Zoo Exhibit	Fish & Chips Shop	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
10	Choa Chu Kang	蔡厝港	West	30000	1.384749	103.744534	0	Fast Food Restaurant	Coffee Shop	Food Court	Asian Restaurant	Bus Station	Bus Stop	Café	Bookstore	Supermarket	Sandwich Place

At the end, we are able generate three choropleth map with cluster marks using the population density per square kilometer, housing price and rent.

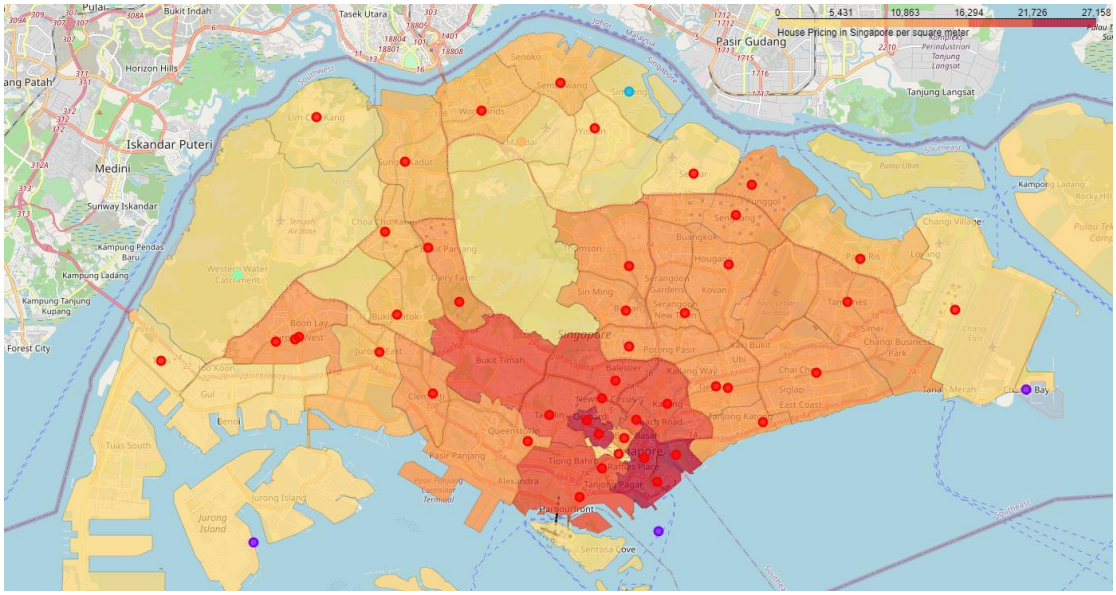
Results

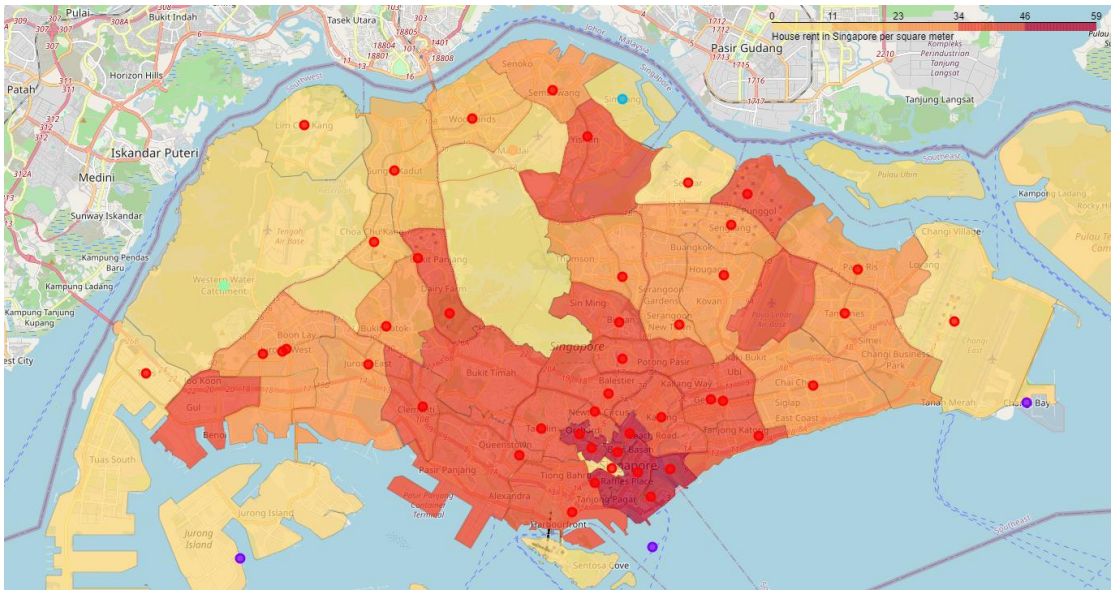
Based on the cluster labels and population density per square kilometer of each planning area, we have the following map:



Where we can find most of planning areas are belong to 0^{th} cluster which is not surprised since Singapore is a small squeezed island country, the central distance of some planning area is small which makes the exploration scope dose not have good independence. Also, food court culture is so popular in Singapore, and the diversity of food are observed in most of planning area. Except some planning areas locate in marginal sites, we are hardly to distinguish the other areas based on common venues cluster. On the other hand, we observe Choa Chu Kang, Woodlands, Jurong West and Sengang are the four planning areas with the highest population densities. Therefore, we may avoid to rent a house here due to potential foot crowd and traffic jam.

The house pricing per square meter and house rent per square meter in Singapore with common venues cluster labels are given below:





As we can see, the distribution of house pricing and rent are quite similar which is reasonable. The most expensive locations are mainly in central and southern regions with the city centre, large business districts and tourist attractions concentrated here, such as Orchard and Marina Bay. Also, we find the top 4 highest population density areas are all belong to second or third section, for house pricing (5,431-10,836) and (10,836-16,294), for rent (11,23) and (23,34). It is usual since lower rent or house pricing result higher buying and living trends. For the first section, it covers from 0, since pricing and rent data of some planning areas are missing in our dataset, and some of these sites are forest, marginal area or island.

Discussion

As mentioned before, Singapore is a small squeezed island country with a popular food court culture and food diversity, therefore, it's not surprised to observe most planning areas are get into same cluster based on common venues, which is a good result, because it means we don't need worry about the choices and diversity of venues when we are trying to decide where to live.

There are some missing data from the original dataset or the process using foursquare API, such as Central Water Catchment, North-Eastern Islands and Southern Islands. This outcome is also can be expected since all these areas are forest, marginal areas or island where people seldom inhabited. If we remove these areas, we may also observe the distribution of population density in the choropleth map is a reverse of the house pricing or rent distribution.

Conclusion

As a result, people like me, a foreigner pursuing a job in Singapore, or who want to learn more about this country, could have a basic understanding of the most common venues, house pricing and rent in Singapore based on different planning areas.