Rental fees and Venues Data Analysis of Singapore

Introduction

Description & Discussion of Background

Singapore is known to be one of the most expensive cities in the world for expats. This can be attributed to extremely high property prices. As a resident in this country-city, I have chosen to study home rental prices in Singapore. The small city of 720 square kilometres is divided into 25 neighbourhoods, and rental prices vary from neighbourhood to neighbourhood.

For an expat looking to rent a place, it may be difficult to differentiate the neighbourhoods and whether certain neighbourhoods are worth the high rental prices. This project aims to study the characteristics of each neighbourhood by looking at its surrounding venues and cluster similar neighbourhoods together. It will then compare the average rental prices of these different clusters and show the corresponding top venues of each cluster. This is to provide expats a simple overview of the different neighbourhood clusters in Singapore and decide for themselves which neighbourhood cluster to stay in. If there is indeed a clear distinction between the clusters' venues and rental prices, it will also seek to find key venues that can explain the relationship between venues and rental prices.

Data Description

1. Data from the Housing Development Board of Singapore is the main data source of this project. The latest data from 2020-Q1 is used. I clean the data by removing some small number of NaN values. The mean rental prices of a "4-Room Flat" of each neighbourhood is filtered to be used for the project.
2. Geocoder is used to find the coordinates of each neighbourhood listed by Housing Development Board.
3. Foursquare API is then used to get the most common venues of each neighbourhood using the coordinates from Geocoder.

Methodology

Rental prices data is downloaded from the official Housing Development Board website and saved as a csv file. After reading it into a dataframe, cleaning it and obtaining the geographical coordinates of each neighbourhood, the following dataframe is obtained.

	Neighborhood	Latitude	Longitude	median_rent
0	ANG MO KIO	1.370073	103.849516	2100.0
1	BEDOK	1.323976	103.930216	2000.0
2	BISHAN	1.350986	103.848255	2300.0
3	BUKIT BATOK	1.349057	103.749591	1900.0
4	BUKIT MERAH	1.270439	103.828318	2500.0
5	BUKIT PANJANG	1.378629	103.762136	1800.0

Python's Folium library is used to provide visualisation of Singapore and the different neighbourhoods. This is done by creating a map of Singapore and marking each neighbourhood individually.

Next, I used Foursquare API to explore the venues around each neighbourhood. The limit was set as 100 venues with radius of 1km for each neighbourhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ANG MO KIO	1.370073	103.849516	FairPrice Xtra	1.369279	103.848886	Supermarket
1	ANG MO KIO	1.370073	103.849516	Old Chang Kee	1.369094	103.848389	Snack Place
2	ANG MO KIO	1.370073	103.849516	Face Ban Mian 非板面 (Ang Mo Kio)	1.372031	103.847504	Noodle House
3	ANG MO KIO	1.370073	103.849516	MOS Burger	1.369170	103.847831	Burger Joint
4	ANG MO KIO	1.370073	103.849516	NTUC FairPrice	1.371507	103.847082	Supermarket

One hot encoding was then used on the **163** different venue categories. All the venues in each neighbourhood are grouped together and the frequency of each venue category was calculated to obtain the following dataframe.

	Neighborhood	ATM	Accessories Store	American Restaurant	Arcade	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	...	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Waterfi
0	ANG MO KIO	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	
1	BEDOK	0.0	0.0	0.016667	0.0	0.0	0.050000	0.0	0.0	0.0	...	0.0	0.0	0.016667	0.0	0.0	
2	BISHAN	0.0	0.0	0.000000	0.0	0.0	0.023256	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	
3	BUKIT BATOK	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	
4	BUKIT MERAH	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	

To better understand each neighbourhood, the top 10 venue categories of each neighbourhood was obtained by ranking the frequency of each venue category.

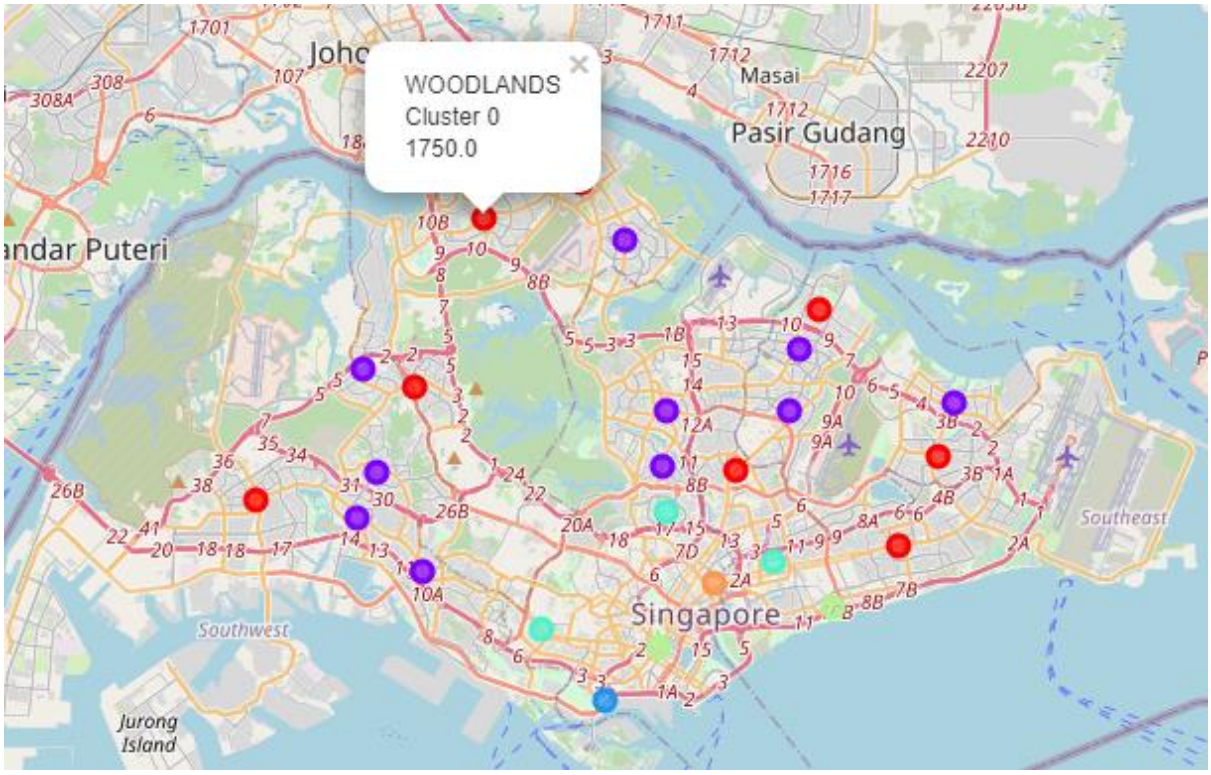
Now we have each neighbourhood's median rental prices as well as the top 10 venues.

	Neighborhood	median_rent	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	2100.0	Coffee Shop	Dessert Shop	Food Court	Sandwich Place	Snack Place	Supermarket	Japanese Restaurant	Fast Food Restaurant	Bubble Tea Shop	Modern European Restaurant
1	BEDOK	2000.0	Chinese Restaurant	Coffee Shop	Asian Restaurant	Sandwich Place	Food Court	Japanese Restaurant	Supermarket	Sushi Restaurant	Fast Food Restaurant	Karaoke Bar
2	BISHAN	2300.0	Food Court	Coffee Shop	Bubble Tea Shop	Ice Cream Shop	Japanese Restaurant	Supermarket	Café	Chinese Restaurant	Cosmetics Shop	Dumpling Restaurant
3	BUKIT BATOK	1900.0	Coffee Shop	Food Court	Chinese Restaurant	Bus Line	Frozen Yogurt Shop	Mobile Phone Shop	Grocery Store	Bowling Alley	Café	Shopping Mall
4	BUKIT MERAH	2500.0	Juice Bar	Coffee Shop	Chinese Restaurant	Hotel	Cafeteria	Flower Shop	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant

To try and detect patterns for neighbourhoods and its median rental prices, I used unsupervised learning K-means algorithm to cluster similar neighbourhoods. The following dataframe includes the cluster label for each neighbourhood.

	Neighborhood	Latitude	Longitude	median_rent	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	9t Co
0	ANG MO KIO	1.370073	103.849516	2100.0	1	Coffee Shop	Dessert Shop	Food Court	Sandwich Place	Snack Place	Supermarket	Japanese Restaurant	Fast Food Restaurant	Te
1	BEDOK	1.323976	103.930216	2000.0	0	Chinese Restaurant	Coffee Shop	Asian Restaurant	Sandwich Place	Food Court	Japanese Restaurant	Supermarket	Sushi Restaurant	Fa Res
2	BISHAN	1.350986	103.848255	2300.0	1	Food Court	Coffee Shop	Bubble Tea Shop	Ice Cream Shop	Japanese Restaurant	Supermarket	Café	Chinese Restaurant	Co
3	BUKIT BATOK	1.349057	103.749591	1900.0	1	Coffee Shop	Food Court	Chinese Restaurant	Bus Line	Frozen Yogurt Shop	Mobile Phone Shop	Grocery Store	Bowling Alley	
4	BUKIT MERAH	1.270439	103.828318	2500.0	2	Juice Bar	Coffee Shop	Chinese Restaurant	Hotel	Cafeteria	Flower Shop	Furniture / Home Store	Frozen Yogurt Shop	

To better visualise this, Folium map is used again to mark out these neighbourhoods in different colours according to their cluster labels. A pop-up to show the rental price of a neighbourhood when clicked was also included to improve user-friendliness.



Results

The median rental price of each neighbourhood cluster and the cluster's top venues are summarised in the dataframe below.

	Cluster Labels	median_rent	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	0	1943.750000	Fast Food Restaurant	Coffee Shop	Japanese Restaurant	Chinese Restaurant	Asian Restaurant	Café	Shopping Mall	Supermarket	Food Court	Clothing Store
1	1	2010.000000	Coffee Shop	Food Court	Fast Food Restaurant	Chinese Restaurant	Supermarket	Sandwich Place	Café	Shopping Mall	Asian Restaurant	Italian Restaurant
4	4	2325.000000	Hotel	Japanese Restaurant	Massage Studio	Multiplex	Indian Restaurant	Salad Place	Yoga Studio	Seafood Restaurant	Chinese Restaurant	Nightclub
3	3	2383.333333	Chinese Restaurant	Noodle House	Food Court	Vegetarian / Vegan Restaurant	Thai Restaurant	Asian Restaurant	Seafood Restaurant	Pool	Train Station	Café
5	5	2400.000000	Dessert Shop	Restaurant	Indian Restaurant	Coffee Shop	Supermarket	Hostel	Flower Shop	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
2	2	2500.000000	Juice Bar	Coffee Shop	Cafeteria	Chinese Restaurant	Hotel	Dumpling Restaurant	Electronics Store	Fast Food Restaurant	Donut Shop	Fish & Chips Shop

Discussion

It can be seen that different neighbourhoods clustered based on venues do have very different median rental prices. A simple inference made from the table above is that lower rental prices neighbourhood clusters have cheaper dining options like coffee shop and fast food as the most common venues. As we move towards higher rental prices, the most common venues become more expensive dining options like restaurants and juice bar.

Geographical location also plays an important a role in determining rental prices. This can be seen from the map visualisation where the more expensive neighbourhood clusters are all located near the central region.

One thing for sure, if expats are willing to payer higher rent prices in prime neighbourhood clusters, they should also be ready to pay more for their dining options.

Conclusion

This simple project clustered different neighbourhoods in Singapore and identified geographical location and dining options to be the key features related to rental prices. Expats can make use of the finding to compare top venues of different neighbourhoods and decide for themselves if they are ready to pay more for rental fees and dining options.