Capstone Project - The Battle of Neighborhoods

Select the Location for a New Chinese Restaurant in Singapore

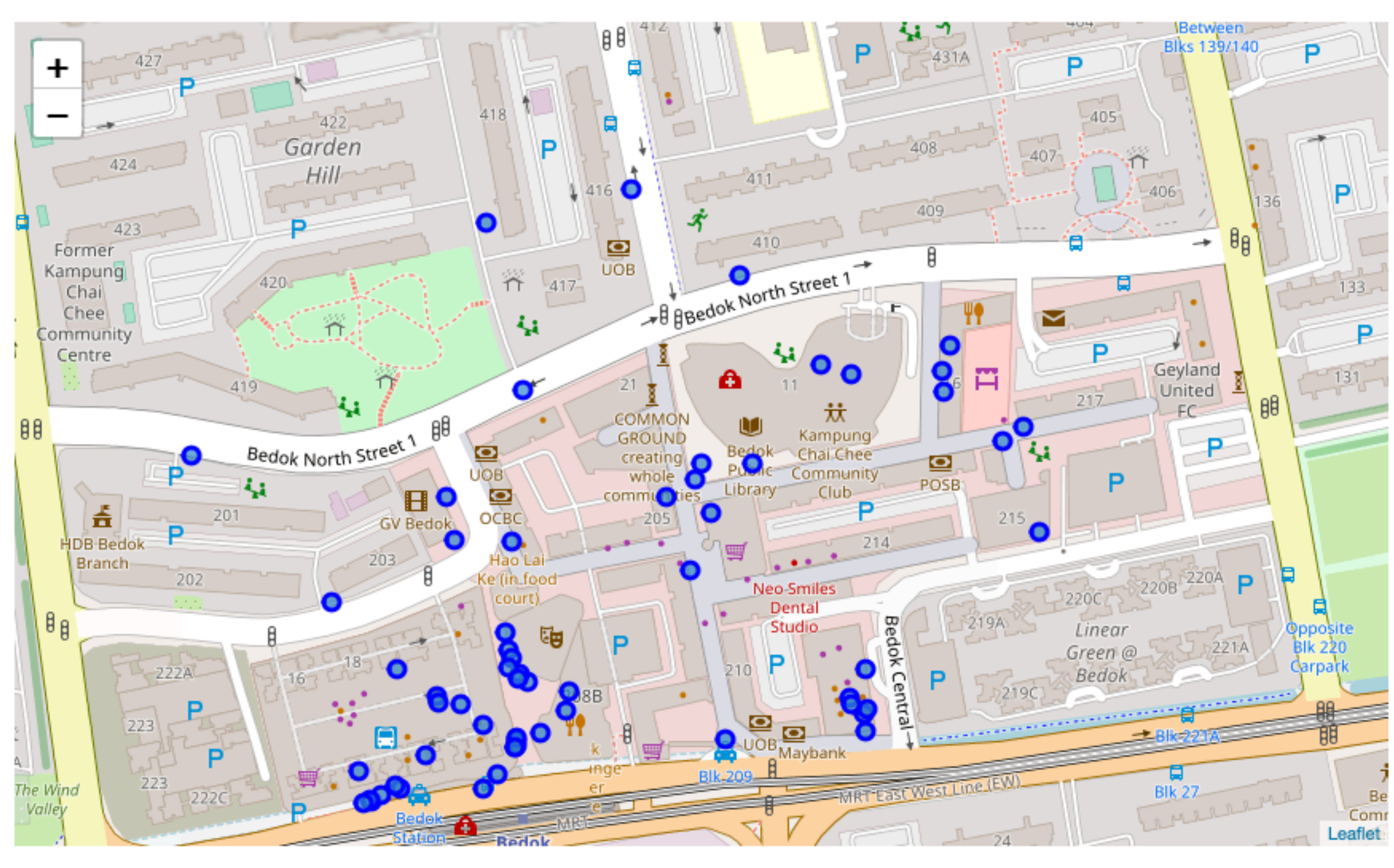
Introduction

Due to the on-going COVID19 pandemic, restaurants are closed for dining and people are staying at home. This is the time when people start to miss the past days when they could eat out and have some get-togethers with friends and families.

With the wish for everything back to normal very soon, I choose topic number 2 and will try to find a good location for a Chinese restaurant in Singapore.

In this project, I will be selecting a location for a new Chinese Restuarant. The **Foursquare** API will be called to gather data for various places in Singapore.

First I explored 200m around Bedok, Singaore, where I live. Using **geolocator**, the geograpical coordinate of Bedok are found to be 1.3239765, 103.930216. Using **folium**, we can plot the Bedok map shown below.



From the map it is clear that most of te venues are near the Bedok MRT staiton, so we will continue to explore all MRT stations in Singapore.

Method

The list of MRT stations in Singapore is available on wikipedia, the list contains all MRT stations including stations being built or planned to be built. The stations I am interested in are the MRT stations which are currently in operation, and are interchange stations.

The geography coordinates which are not included in the table can be obtained by searching using **geocoder** with all the station names.

With the station names and their coordinates, we can explore the interchange stations by calling **foursquare** API and get the most visited venues for each station. Finally we apply *kmeans clustering* and use **folium** to visualise the popular venue distributions among MRT interchange stations and select a area for a new Chinese restaurant.

Data pre-processing

In this part all Singapore MRT names and locations will be obtained by an easy web scraping using **pandas.read_html**, then I will extract the interchange MRT stations, search their coordinates, finally **Foursquare** API will be called to get the most visted venues for all interchange MRT stations in Singapore, and a dataframe will be created with all the information.

First I got a list of MRT stations from wikipedia and transferred it into a dataframe.

	Alpha-numeric code(s)	Alpha-numeric code(s).1	Station name	Station name.1	Station name.2	Opening	Name(s) during planning stages	Abbreviation	Location(s)	Connection(s) to other transport
1	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)	North South Line (NSL)
2	NS1 EW24	JE5	Jurong East	裕廊东	ஜூரோங் கிழக்கு	10 March 1990	Jurong East	JUR	Jurong East	Jurong East Temporary Bus Interchange
3	NS2	NaN	Bukit Batok	武吉巴督	புக்கிட் பாத்தோக்	10 March 1990	Bukit Batok South	BBT	Bukit Batok	Bukit Batok Bus Interchange
4	NS3	NaN	Bukit Gombak	武吉甘柏	புக்கிட் கோம்பாக்	10 March 1990	Bukit Batok North	BGB	Bukit Batok	NaN
5	NaN	NS3A	Brickland	红砖	பிரிக்லேண்ட்	Mid-2030s	Brickland	TBA	Exact location not yet known	NaN
...	...	...	...	...	...	...	...	...	...	...
250	NaN	CR13 TE7	Bright Hill	光明山	பிரைட் ஹில்	2029	Bright Hill	BRH	Bishan	NaN
251	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)	Punggol Extension (CRLe)
252	NaN	CP2	Elias	伊莱雅	இலியாஸ்	2031	Elias	TBA	Pasir Ris	NaN
253	PE4	CP3	Riviera	里维拉	ரிவியாரா	2031	Riviera	TBA	Punggol	NaN
254	NE17 PTC	CP4	Punggol	榜鹅	பொங்கோல்	2031	Punggol	PGL	Punggol	Punggol Temporary Bus Interchange

254 rows × 10 columns

The raw data can be processed into a dataframe with the information we need: interchange stations and their coordinates

	Station name	Station name in Chinese	Location	Latitude	Longitude
0	Jurong East	裕廊东	Jurong East	1.333115	103.742297
1	Choa Chu Kang	蔡厝港	Choa Chu Kang	1.384749	103.744534
2	Woodlands	兀兰	Woodlands	1.436897	103.786216
3	Bishan	碧山	Bishan	1.350986	103.848255
4	Newton	纽顿	Newton	1.313183	103.838040
5	Dhoby Ghaut	多美歌	Museum Planning Area	1.299353	103.845309
6	City Hall	政府大厦	Downtown Core	1.293027	103.852643
7	Raffles Place	莱佛士坊	Downtown Core	1.283542	103.851460
8	Marina Bay	滨海湾	Downtown Core, Straits View Planning Area	1.275559	103.854897
9	Tampines	淡滨尼	Tampines	1.354653	103.943571
10	Tanah Merah	丹那美拉	Bedok	1.327254	103.946505
11	Paya Lebar	巴耶利峇	Geylang	1.318218	103.893235
12	Bugis	武吉士	Downtown Core, Rochor	1.299953	103.855278
13	Outram Park	欧南园	Bukit Merah, Outram	1.280650	103.840279
14	Buona Vista	波那维斯达	Queenstown	1.307085	103.790579
15	Expo	博览	Tampines	1.335443	103.962085
16	HarbourFront	港湾	Bukit Merah	1.265395	103.822403
17	Chinatown	牛车水	Outram	1.283737	103.843798
18	Little India	小印度	Kallang, Rochor	1.306648	103.849269
19	Serangoon	实龙岗	Serangoon	1.349862	103.873729
20	Sengkang	盛港	Sengkang	1.391654	103.895364
21	Punggol	榜鹅	Punggol	1.405258	103.902330
22	Promenade	宝门廊	Downtown Core	1.292708	103.861124
23	MacPherson	麦波申	Geylang	1.326768	103.890302
24	Bayfront	海湾舫	Downtown Core	1.282572	103.859724

After exploring 200m around each MRT interchange station by making API calls with **Foursquare** developer account, we found 219 uniques categories among those MRT stations.

Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bayfront	45	45	45	45	45	45
Bishan	45	45	45	45	45	45
Bugis	100	100	100	100	100	100
Bukit Panjang	32	32	32	32	32	32
Buona Vista	46	46	46	46	46	46
Chinatown	100	100	100	100	100	100
Choa Chu Kang	23	23	23	23	23	23
City Hall	61	61	61	61	61	61
Dhoby Ghaut	77	77	77	77	77	77
Expo	63	63	63	63	63	63
HarbourFront	100	100	100	100	100	100
Jurong East	77	77	77	77	77	77
Little India	33	33	33	33	33	33
MacPherson	15	15	15	15	15	15
Marina Bay	12	12	12	12	12	12
Newton	28	28	28	28	28	28
Outram Park	61	61	61	61	61	61
Paya Lebar	72	72	72	72	72	72
Promenade	43	43	43	43	43	43
Punggol	45	45	45	45	45	45
Raffles Place	50	50	50	50	50	50
Sengkang	32	32	32	32	32	32
Serangoon	41	41	41	41	41	41
Tampines	84	84	84	84	84	84

Clustering and Data Analysis

The MRT stations and their top 10 most common venues can be found

	Station	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	Bayfront	Boutique	Scenic Lookout	Garden	Hotel	Waterfront	Roof Deck	Lounge	Bridge	Accessories Store	Park
1	Bishan	Coffee Shop	Food Court	Bubble Tea Shop	Cosmetics Shop	Pet Store	Café	Japanese Restaurant	Chinese Restaurant	Ice Cream Shop	Supermarket
2	Bugis	Bakery	Café	Hotel	Cocktail Bar	Dessert Shop	Chinese Restaurant	Japanese Restaurant	Thai Restaurant	Coffee Shop	Sandwich Place
3	Bukit Panjang	Fast Food Restaurant	Coffee Shop	Shopping Mall	Noodle House	Sushi Restaurant	Asian Restaurant	Café	Fried Chicken Joint	Gym	Supermarket
4	Buona Vista	Japanese Restaurant	Indian Restaurant	Food Court	Chinese Restaurant	Shopping Mall	Café	Bakery	Dessert Shop	Coffee Shop	Performing Arts Venue

station_venues_sorted.csv

After running clustering for 5 clusters, a new dataframe with clustering labels, station names, station Chinese names, station coordinates, and 1-10 most common venues is created

	Station name	Station name in Chinese	Location	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
0	Jurong East	裕廊东	Jurong East	1.333115	103.742297	0	Chinese Restaurant	Coffee Shop	Food Court	Café	Japanese Restaurant	Shopping Mall	Steakhouse	Multiplex
1	Choa Chu Kang	蔡厝港	Choa Chu Kang	1.384749	103.744534	0	Coffee Shop	Food Court	Portuguese Restaurant	Thai Restaurant	Playground	Fast Food Restaurant	Sandwich Place	Chinese Restaurant
2	Woodlands	兀兰	Woodlands	1.436897	103.786216	0	Café	Coffee Shop	Japanese Restaurant	Shopping Mall	Asian Restaurant	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant
3	Bishan	碧山	Bishan	1.350986	103.848255	0	Coffee Shop	Food Court	Bubble Tea Shop	Cosmetics Shop	Pet Store	Café	Japanese Restaurant	Chinese Restaurant
4	Newton	纽顿	Newton	1.313183	103.838040	0	Chinese Restaurant	Seafood Restaurant	Italian Restaurant	Hotel Bar	Hotel	Convenience Store	Grocery Store	Gym / Fitness Center
5	Dhoby Ghaut	多美歌	Museum Planning Area	1.299353	103.845309	3	Hotel	Café	Park	Japanese Restaurant	Cosmetics Shop	History Museum	Bubble Tea Shop	Theater
6	City Hall	政府大厦	Downtown Core	1.293027	103.852643	3	Hotel	Shopping Mall	Japanese Restaurant	Coffee Shop	French Restaurant	Steakhouse	Event Space	Concert Hall
7	Raffles Place	莱佛士坊	Downtown Core	1.283542	103.851460	3	Hotel	Café	Food Court	Gym	Cocktail Bar	Coffee Shop	Sandwich Place	Salad Place
8	Marina Bay	滨海湾	Downtown Core, Straits View Planning Area	1.275559	103.854897	4	Yoga Studio	Harbor / Marina	Spanish Restaurant	Plaza	Building	Gastropub	Mexican Restaurant	Government Building
9	Tampines	淡滨尼	Tampines	1.354653	103.943571	0	Bakery	Café	Coffee Shop	Fast Food Restaurant	Supermarket	Gym	Sushi Restaurant	Chinese Restaurant
10	Tanah Merah	丹那美拉	Bedok	1.327254	103.946505	0	Noodle House	Bus Station	Indian Restaurant	Malay Restaurant	Coffee Shop	Resort	Bus Line	Food Court
11	Paya Lebar	巴耶利峇	Geylang	1.318218	103.893235	0	Shopping Mall	Fast Food Restaurant	Coffee Shop	Food Court	Asian Restaurant	Supermarket	Halal Restaurant	Hotpot Restaurant
12	Bugis	武吉士	Downtown Core, Rochor	1.299953	103.855278	3	Bakery	Café	Hotel	Cocktail Bar	Dessert Shop	Chinese Restaurant	Japanese Restaurant	Thai Restaurant

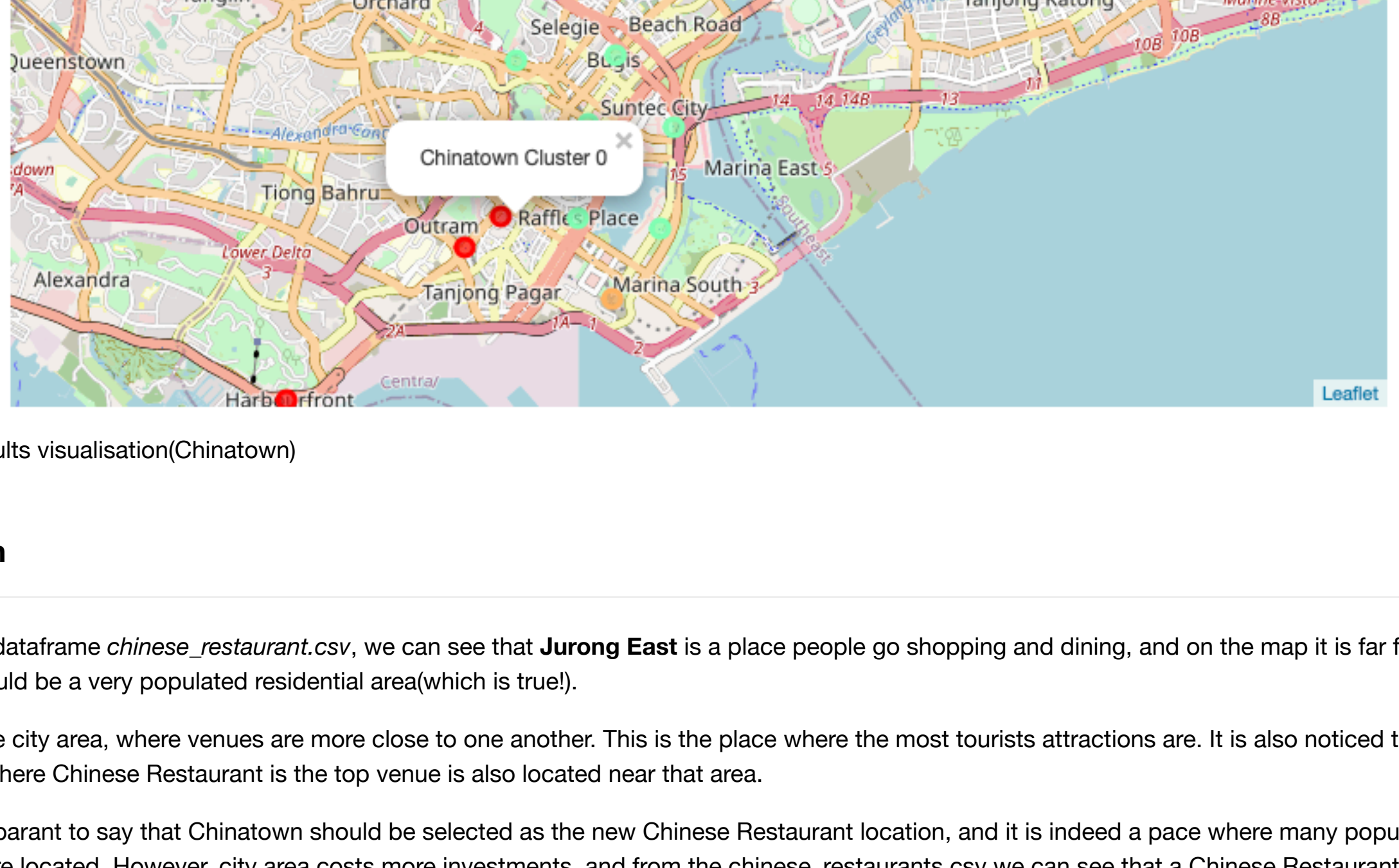
singapore_stations.csv

We can see that Chinese Restaurant is the top most common venues for several MRT stations, and they are all under Cluster 0.

	Station name	Station name in Chinese	Location	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
0	Jurong East	裕廊东	Jurong East	1.333115	103.742297	0	Chinese Restaurant	Coffee Shop	Food Court	Café	Japanese Restaurant	Shopping Mall	Steakhouse
1	Newton	纽顿	Newton	1.313183	103.838040	0	Chinese Restaurant	Seafood Restaurant	Italian Restaurant	Hotel Bar	Hotel	Convenience Store	Grocery Store
2	HarbourFront	港湾	Bukit Merah	1.265395	103.822403	0	Chinese Restaurant	Japanese Restaurant	Fast Food Restaurant	Toy / Game Store	Clothing Store	Multiplex	Coffee Shop
3	Chinatown	牛车水	Outram	1.283737	103.843798	0	Chinese Restaurant	Food Court	Hostel	Vegetarian / Vegan Restaurant	Italian Restaurant	Spa	Café

chinese_restaurant.csv

And singapore_station.csv can be visualised below:



Clustering results visualisation(Chinatown)

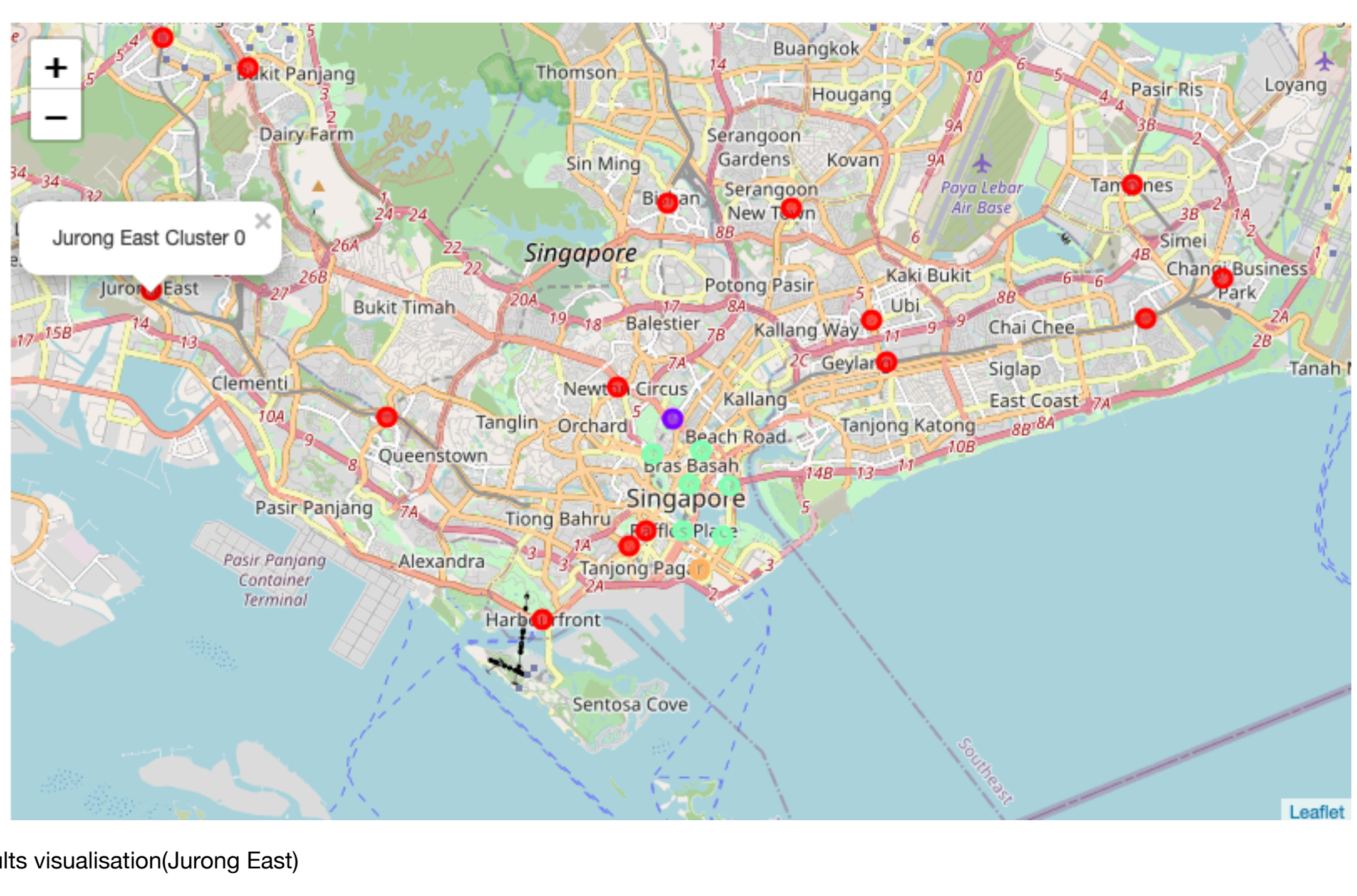
Discussion

From the last dataframe *chinese_restaurant.csv*, we can see that **Jurong East** is a place people go shopping and dining, and on the map it is far from the city area, so it should be a very populated residential area(which is true!).

Cluster 3 is the city area, where venues are more close to one another. This is the place where the most tourists attractions are. It is also noticed that **Chinatown**, where Chinese Restaurant is the top venue is also located near that area.

It might be apparant to say that Chinatown should be selected as the new Chinese Restaurant location, and it is indeed a pace where many popular Chinese Restaurants are located. However, city area costs more investments, and from the chinese_restaurants.csv we can see that a Chinese Restaurant in Chinatown tends to have more compitations not only from Chinese restaurants, but also from food courts and other Asian Restaurants.

Therefore I would apply *the Blue Ocean Strategy* and select **Jurong East** as the location to build a neighborhood Chinese Restaurant. **Jurong East** MRT interchange is a residential area with many shopping venues, Chinese Restaurants are popular there, and it is far from the compitations



Clustering results visualisation(Jurong East)

Conclusion

In this project, MRT stations in Singapre are gathered from wikipedia using **pandas.read_html**, all interchange stations are extracted and their coordinates are found. With those information, **Foursquare** API is called to explore 200ms around all the MRT interchange stations in Singapore, and their top 10 most common venues are found. Then new data frame is being created by filtering the MRT station where Chinese Restaurant is the top common venue. Finally clustering was applied on the MRT data all data is visualised on a map plotted by **folium**. With all the data analysed and visualised, a conclusion was drawn that the new Chinese Restaurant is to be located around **Jurong East** station.