Capstone Project - Land Prices & Venues Data Analysis of Tokyo Metropolis

1. Introduction: Business Problem

As my assignment, I would like to choose Tokyo Metropolis which is one of the largest cities in Asia Pacific. It has about over 13 million people and consists of 23 wards rural cities and islands. Tokyo Metropolis will have Summer Olympic game in 2020 and will have a lot of visitors.

At my assignment, I would analyze where is the best location to start a restaurant or to open an office for business from venue data view point and land price view point. In order to start new restaurant or new business, we will need to consider what kind of venues are available in Tokyo Metropolis. At the same time, in order for us to open a restaurant or an office, we will also need to choose appropriate area which has reasonable rental price of venues.

Among all cities and 23 wards (central area) of Tokyo Metropolis, I would like to focus on 23 wards which has more than 9 million people and they are the center of business, politics and all the cultural activities. In 23 wards, I would like to study venue information around major subway stations. Major subway stations are located in major area of 23 wards and as distance from station is essential to start new business in central Tokyo. Then, I would like to show major venue category and land price of central Tokyo area.

2. Data

In order to consider the problem, we will use the data as below:

1. I used land price information from Tokyo Metropolitan Government, Bureau of Finance (2018). (http://www.zaimu.metro.tokyo.jp/kijunchi/30nen/index.html). This land price will be reflected to rental price of office or restaurant and it will be the major part of the cost of new business.
2. I used Tokyo subway map and location information. (https://www.tokyometro.jp/en/subwaymap/)(https://www.javadrive.jp/google-maps-javascript/data/index2.html)
3. Geojson data is available for Tokyo Metropolis, and I used 23 ward parts to create choropleth map of average land price for each of 23 wards from the land price information. (https://github.com/dataofjapan/land/blob/master/tokyo.geojson)
4. To find our appropriate places to open a restaurant or an office, I used Foursquare API to find venues around major sub way station which are located in major part of Tokyo Metropolis 23 wards.

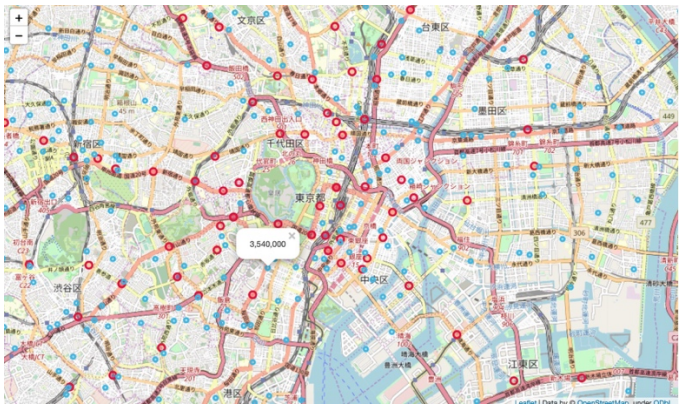
3. Methodology

I used python folium library to visualize geographic details of Tokyo Metropolis and its 23 wards. I selected 5 major subway lines (marunouchi, chiyoda, hanzomon, hibiya, yurakucho) which have total 104 stations. I created a map of Tokyo 23 wards with major subway stations superimposed on top. I used latitude and longitude values to get the visual as below:

In order to find land prices of Tokyo 23 wards. I used land price data from Tokyo Metropolitan Government, Bureau of Finance (2018). The data consists of ward, address (street), price (yen/m2), subway name, distance to the subway station, latitude, longitude. (total 678 locations for 23 wards)

	Location	Ward_id	Ward	Street	Price	Metro	Distance	latitude	longitude
0	1	13101	千代田区	隼町３－１９	2,800,000	半蔵門	180	35.68162	139.74202
1	2	13101	千代田区	大手町１－８－１	25,100,000	大手町	NaN	35.68778	139.76434
2	3	13101	千代田区	内神田１－１５－１０	3,700,000	小川町	250	35.69196	139.76558
3	4	13101	千代田区	岩本町３－７－１３	1,170,000	岩本町	240	35.69597	139.77777
4	5	13101	千代田区	丸の内３－３－１	26,300,000	東京	480	35.67756	139.76265

I also made another map which also have land price information (yen/m2)(yen for per square meter) and its locations. (This example shows 3, 540,000 yen/m2, which is about $33K/m2.)



I utilized the Foursquare API to explore the subway stations and its neighborhood. We designed the limit as 100 venue and the radius 500 meter for each subway station from their given latitude and longitude information. This example is shows venues around Chiyoda-line C01 (yoyogi-uehara are 35.669066, 139.67968)

	name	categories	lat	lng
0	Asterisque	Dessert Shop	35.668202	139.681458
1	haritts	Donut Shop	35.669316	139.682064
2	終日one	Gastropub	35.668281	139.678948
3	Yamato (山都)	Soba Restaurant	35.668070	139.681635
4	オトナノイザカヤ中戸川	Japanese Restaurant	35.669296	139.680633

I consolidated all the venues for 104 stations. C01, C02, C03 are showing stations of Chiyoda-lien and total venues are 87, 56, 100 (max) for example.

station_id	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
C01	87	87	87	87	87	87
C02	56	56	56	56	56	56
C03	100	100	100	100	100	100
C04	100	100	100	100	100	100
C05	83	83	83	83	83	83

I found 7722 venues and 308 unique categories for 104 stations of 5 major subway lines. I created a table which shows list of top 10 venue category for each subway station in below table.

	station_id	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 M Comn Ve
0	C01	Japanese Restaurant	Italian Restaurant	Convenience Store	Soba Restaurant	Café	Restaurant	Coffee Shop	Sake Bar	Deli / Bodega	Des S
1	C02	Café	Italian Restaurant	Coffee Shop	Chocolate Shop	Bakery	French Restaurant	Japanese Restaurant	Bistro	Japanese Curry Restaurant	Taiwan Restau
2	C03	Café	Boutique	Clothing Store	Sporting Goods Shop	Ramen Restaurant	Coffee Shop	Thai Restaurant	Toy / Game Store	Candy Store	Okonomi Restau
3	C04	Café	Boutique	Japanese Restaurant	Coffee Shop	Dessert Shop	Clothing Store	Tea Room	BBQ Joint	Bookstore	Chin Restau
4	C05	Japanese Restaurant	Italian Restaurant	Coffee Shop	Art Gallery	Café	Seafood Restaurant	Steakhouse	Chinese Restaurant	French Restaurant	Nightc

I have some common venue categories in subway stations, and I used unsupervised learning K-means algorithm to cluster the 104 subway stations. K-Means algorithm is one of the most common cluster methods of unsupervised learning. At this report, I used K as 5 for clustering. The following is top part of cluster – 0.

	station_id	sta_e	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
7	M08	shinjuku	0	Japanese Restaurant	Ramen Restaurant	Sake Bar	Coffee Shop	Dessert Shop	Record Shop	Chinese Restaurant	Café	Movie Theater	Bar
8	M09	shinjuku-sanchome	0	Café	Japanese Restaurant	Bar	BBQ Joint	Sake Bar	Ramen Restaurant	Pub	Japanese Curry Restaurant	Nightclub	Jazz Club
9	M10	shinjuku-gyoemmae	0	Café	Ramen Restaurant	Japanese Restaurant	Gay Bar	Chinese Restaurant	Bar	Italian Restaurant	Nightclub	Japanese Curry Restaurant	BBQ Joint
15	M16	ginza	0	Café	Japanese Restaurant	Boutique	Cocktail Bar	Sushi Restaurant	Clothing Store	Bookstore	Ramen Restaurant	Tempura Restaurant	Bakery
16	M17	tokyo	0	Café	Hobby Shop	Gift Shop	Dessert Shop	Japanese Restaurant	Ramen Restaurant	Tonkatsu Restaurant	Sushi Restaurant	Hotel	Lounge

The most common venues for each cluster are as follows.

Cluster 0: Café, Japanese Restaurant, Hotel, Italian Restaurant
Cluster 1: Sake Bar, Café, Convenience Store
Cluster 2: Convenience Store, Ramen Restaurant, Park, Hotel
Cluster 3: Convenience Store, Japanese Restaurant
Cluster 4: Japanese restaurant, Ramen Restaurant

I will also summarize land price of 23 ward. Based on land price map, I summarized average land price for 1 m2 (for per square meter) for each ward. These averages were made from total 678 locations for 23 wards. The following are average prices at 23 wards. (ward_id2 is geojson id.). Average land price for 23 wards varies according to the distance from central Tokyo (Tokyo Station) and at the same time, we understand land price varies according to the distance to nearby subway station in each ward.

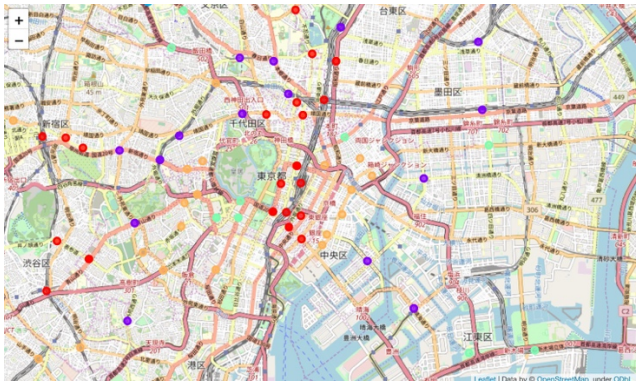
	Ward_id	Ward_id2	Ward	Price
0	13101	131016	千代田区	7059000
1	13102	131024	中央区	6010000
2	13103	131032	港区	3860000
3	13104	131041	新宿区	3762000
4	13105	131059	文京区	1184000
5	13106	131067	台東区	1297000
6	13107	131075	墨田区	556000
7	13108	131083	江東区	580000
8	13109	131091	品川区	920000
9	13110	131105	目黒区	1124000
10	13111	131113	大田区	548000
11	13112	131121	世田谷区	677000
12	13113	131130	渋谷区	3150000
13	13114	131148	中野区	819000
14	13115	131156	杉並区	657000
15	13116	131164	豊島区	1519000
16	13117	131172	北区	628000
17	13118	131181	荒川区	604000
18	13119	131199	板橋区	523000
19	13120	131202	練馬区	435000
20	13121	131211	足立区	371000
21	13122	131229	葛飾区	359000
22	13123	131237	江戸川区	422000

We can categories these 23 wards into four groups.

(1) 5,900,000 < 千代田区(Chiyoda), 中央区(Chou)
(2) 3,700,000 < 港区(Minato), 新宿区(Shinjuku)
(3) 1,500,000 < 渋谷区(Shibuya), 豊島区(Toshima)
(4) 350, 000 < Other wards

4. Results

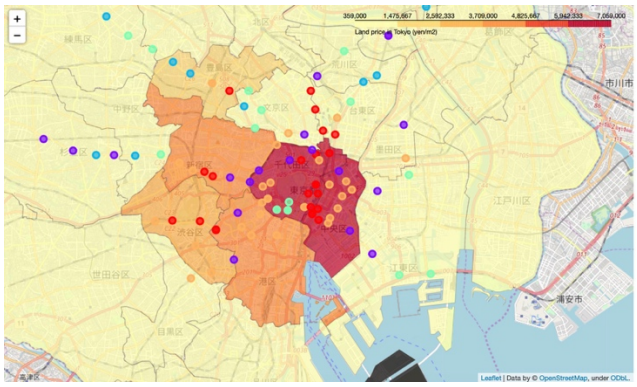
I will merge subway station map with related cluster information. This map has 104 stations with 5 cluster information.



In summary section, I would like to visualize average land price for per square meter choropleth style map. We are using geojson information for Tokyo 23 ward which we downloaded. I created choropleth map which also has the below information for each subway station:

- Station id
- Cluster name
- Land price information.

I finalized the study by visualizing the data and clustering information on geojson map.



5. Discussion

Tokyo has 9 million people in 23 wards. We used 104 subway stations for 5 major subway lines to find out good place to start business. Actually, Tokyo has mora than 12 subways, we may use another combination of subway line to expand our city coverage.

In order for us to assess various advantages and risks for opening new restaurant or for opening new office, we may also consider population or dynamics of population change in a day.

I used the K-means algorithm as part of this clustering study. At this report we used 5 clusters however we may change the number of K. We can also apply other clustering methods or classification methods.

I performed data analysis using the coordinates of subway station and land average prices. I used average land price for a ward. As we have more fine-grained data from Tokyo Metropolitan Government, we may use color map not by ward but by smaller city or town level.

6. Conclusion

When we think about starting new business (with restaurant or new office) it is better to start in the place with low cost (housing), in more populated area, and at the closer place to the major transportation as I studied at this report. This approach will show us the initial options which we can take for achieving the goal such as to start new restaurant or to start new office.

As next step, as we have dramatic improvement of telecommunication technology, of transportation, or as we have new type of services such as delivery services, therefore, we will also need to consider access for such new advantages and/or access to new platform where we can get such new services.

7. References

1. Tokyo Metropolitan Government (http://www.metro.tokyo.jp/english/)
2. Land price information from Tokyo Metropolitan Government, Bureau of Finance (2018). (http://www.zaimu.metro.tokyo.jp/kijunchi/30nen/index.html).
3. Tokyo subway map (https://www.tokyometro.jp/en/subwaymap/)
4. Tokyo subway geo coordinate (https://www.javadrive.jp/google-maps-javascript/data/index2.html)
5. Geojson data for Tokyo (https://github.com/dataofjapan/land/blob/master/tokyo.geojson)
6. Foursquare API (https://foursquare.com/)