BATTLE OF THE NEIGHBORHOODS

A Study of the Guangdong-Hong Kong-Macao Greater Bay Area

ABSTRACT

The Study leverages Foursquare location data to explore and compare the 11 constituents cities of the Guangdong-Hong Kong-Macao Greater Bay Area, with a view to identifying the strategic collaboration opportunities between the cities to maximise synergy of the development.

Vincent Hui

Coursera IBM Applied Data Science Capstone Project

Introduction

The Central Government of the People's Republic of China promulgated the Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area[1] (GBA) in 2019, which sets out the strategic development plan of the GBA with a view to developing an international first-class bay area ideal for living, working and travelling. The GBA comprises the two Special Administrative Regions of Hong Kong and Macao, and the nine municipalities of Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen and Zhaoqing in Guangdong Province. The total area is around 56 000 km^2. At end 2018, the total population is over 71 million, the GDP is USD 1,642.5 billion and GDP per capita is USD 23,342, presenting vast opportunities and prospects for future development. In the light of the development potential of the GBA, it would be of utmost importance to understand and harness the unique comparative advantages of each of the regions in order to maximise the synergy between the "9+2" cities within the GBA. Thus, this study undertakes to study the key cities within the GBA to identify the characteristics of individual cities and identify strategic clusters that could inform future collaborations between cities to take forward the mega-GBA development.

Data

2. The data used in the study is mainly obtained from two data courses. Firstly, the Foursquare developer API is used to obtain venue recommendations nearby the 11 GBA cities, with the “explore” endpoint. A radius of 10 km is adopted, and the first 100 venues are selected for the analysis.

3. In addition to the Foursquare location data, this study will also be utilising some demographic statistics obtained from the GBA thematic website maintained by the Government of the Hong Kong Special

[1] https://www.bayarea.gov.hk/filemanager/en/share/pdf/Outline_Development_Plan.pdf

Administrative Region[2] and the Hong Kong Trade Development Council[3], which includes information such as industry structure, GDP, population, area, trade value, etc. These are scrapped using the BeautifulSoup 4 package on Python. Besides, latitude and longitude values of the 11 cities are obtained online[4]. A summary of the demographic data obtained from other sources is at ***Table 1*** below. To avoid the different magnitude of data affecting the accuracy of the machine learning algorithm, standardisation of the demographic data (except that of latitude and longitude) is conducted with the use of the Standard Scaler of the sklearn.preprocessing package.

Table 1 Demographic Data of the GBA Cities

City	Land Area (km^2) (Area)	Population (million)	GDP per capita (US$) (GDP_pc)	GDP Share of tertiary industry (%) (Tertiary)	Export (US$ billion)	Foreign Direct Investment (US$ billion) (FDI)	Latitude	Longitude
Hong Kong	1,107	7.48	48,673	92.43	530.44	110.73	22.302711	114.177216
Macao	33	0.67	82,609	94.93	1.51	0.3753	22.198746	113.543877
Guangzhou	7,434	14.9	23,497	71.8	84.74	6.611	23.129110	113.264381
Shenzhen	1,997	13.03	28,647	58.8	245.94	8.203	22.543097	114.057861
Foshan	3,798	7.91	18,992	42.0	53.30	0.691	23.021479	113.121437
Dongguan	2,460	8.39	14,951	51.1	120.22	1.361	23.020674	113.751801
Huizhou	11,347	4.83	12,908	43.0	33.38	0.959	23.091181	114.400681
Zhongshan	1,784	3.31	16,711	49.3	27.23	0.527	22.527470	113.361526
Jiangmen	9,507	4.6	9,570	44.5	16.97	0.734	22.580391	113.080009
Zhuhai	1,736	1.89	24,100	49.1	28.52	2.391	22.270979	113.576675
Zhaoqing	14,891	4.15	8,050	38.6	3.59	0.143	23.047192	112.465091

4. With the latitude and longitude data, the study will identify the key venues of each cities in the GBA, utilising Foursquare, specifically those related to economic activities, to identify the more common types of venues in each city. In combination with the other demographic data, the study will perform clustering analysis on the cities to separate the 11 cities in total into 5 clusters; and identify each of their strategic strength area through analysis of their characteristics. The study will also use the folium map package to

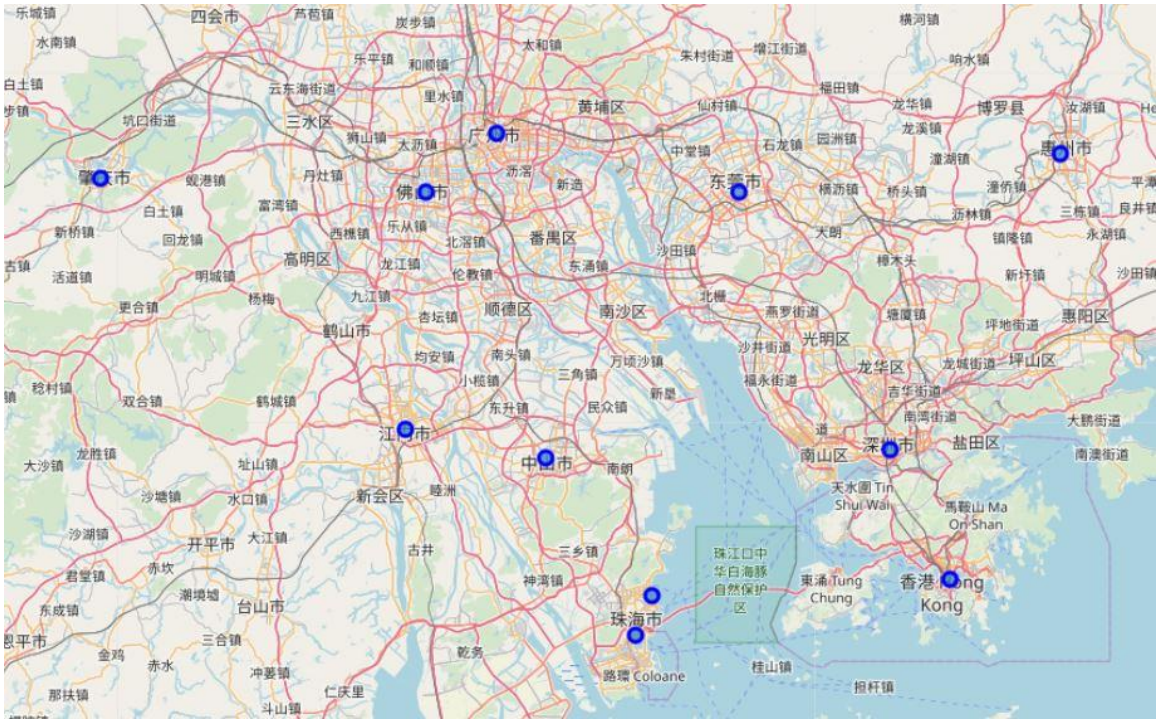
[2] https://www.bayarea.gov.hk/

[3] http://hong-kong-economy-research.hktdc.com/business-news/article/Guangdong-Hong-Kong-Macau-Bay-Area/Statistics-of-the-Guangdong-Hong-Kong-Macao-Greater-Bay-Area/bayarea/en/1/1X000000/1X0AE3Q1.htm

[4] https://github.com/Huirricane/Coursera_Capstone/blob/master/GBA_Coordinates.csv.

demonstrate the clustering result. A map showing the location of the 11 GBA cities is at ***Figure 1*** below.

Figure 1 Map of the GBA



Methodology

5. To utilise the Foursquare API, there is first a need to determine the parameters for using the “explore” end point, i.e. the limit number of venues and the radius of the search. As compared with the Manhattan neighborhoods project, this study warrants a larger search radius given that the locations are larger cities instead of small neighborhoods inside cities. A small radius would be prone to bias owing to the different selection of latitude and longitude to pinpoint that particular city. As such, there is a need to carry out trial and error to choose a suitable radius parameter. Having tested out the few options of 500m (same as Manhattan project), 5km and 10km, it is considered that 10km would be a more reasonable choice, given that the city with the least number of venues searched, i.e. Zhaoqing, has six venues. Otherwise, some cities may not be represented in the analysis with a choice

of a smaller radius parameter. As for the limit parameter, the same value of 100 in the Manhattan project is chosen.

6. After obtaining the venue recommendation data from the Foursqaure API, it is observed that 134 unique categories of venues are found among 656 venues in total in the 11 cities. The four largest and the most internationalised cities in the GBA, namely Hong Kong, Macao, Guangzhou and Shenzhen, each reaches the cap of 100 venue recommendations, while the number of venues recommended for smaller cities are fewer, with Jiangmen, Huizhou and Zhaoqing each having less than 20 recommendations. After calculating the average representation of the venues in the cities as well as the top ten venues in each city, to compare the similarity between different cities in the GBA and identify their strategic strengths, the study adopts the k-means clustering methodology to cluster the 11 cities into 5 clusters. Then, each of the clusters are analysed in accordance with their demographic features and venue types to identify their unique characteristics.

7. To evaluate the robustness of the results, the study also attempts various robustness tests, including conducting the analysis with non-standardised demographic data as well as using location data alone. It is found that these manoeuvres do affect the result to a certain extent and shall be elaborated in the section below.

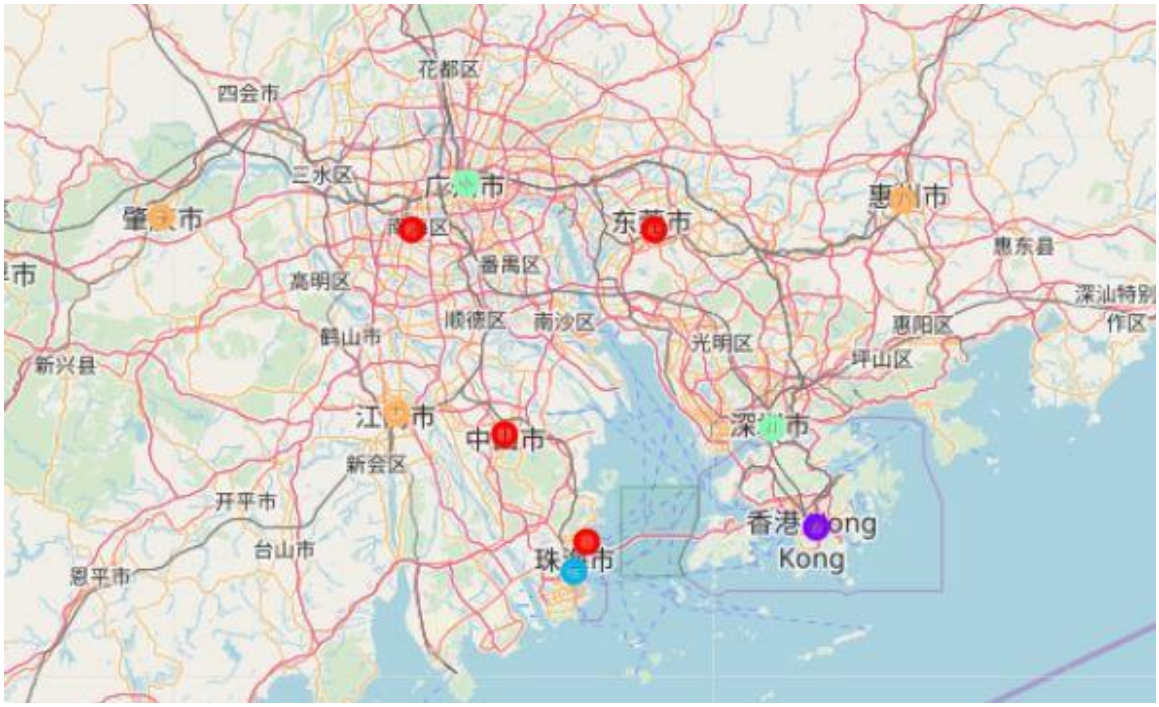
Results

8. Using k-means clustering with k set to 5, the resulting map demonstrating the 5 unique clusters of the 11 GBA cities is at ***Figure 2*** below, while a table summarising the clustering results and including the demographic and location features of the cities is at ***Table 2*** below.

9. Examining the five clusters one by one, attention is first drawn to the two clusters (number 1 and 2) with only one city in them, namely Hong Kong and Macao. This is understandable given that these are the only two Special Administrative Regions of China that are governed under “One Country, Two Systems” that gives them unique comparative advantages over the remaining GBA cities on the Mainland. Demographically, these two cities are

characterised by an exceptionally high contribution of the tertiary industry to their GDPs (over 90%), which signifies that both cities are highly developed service-oriented economies. In terms of venues, the top two venues are similarly hotels and cafes, while different types of restaurants are also prevalent throughout the ranks. However, it is evident that Hong Kong is more shopping-oriented with an abundance of clothing stores, electronic stores, and shopping malls, while Macao is more cultural-driven with historic sites, churches, plazas taking a place. In addition, Hong Kong has a much larger population and land area than Macao, with a much larger export and foreign direct investment value. This suits the positioning of Hong Kong as a centre for commerce and trade. On the contrary, the mainstay of Macao's development potential is in the tourism industry, in particular gambling, which contributes towards its exceptionally high GDP per capita level, almost double that of Hong Kong.

Figure 2 Map of the GBA City Clusters



10. Having settled the positioning of the two one-city clusters, the cluster 0 with four cities, Foshan, Donguan, Zhongshan, and Zhuhai, are analysed. In terms of demography, these cities have a moderate level of GDP per capita, at around US$15,000 – 24,000, with a comparable level of contribution from

Table 2 Clustering Result of the GBA Cities and their Features

Cluster	City	Area	Population	GDP_pc	Tertiary	Export	FDI	Most Common Venues									
								1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
0	Foshan	3,798.00	7.91	18,992.00	42.00	53.30	0.69	Coffee Shop	Hotel	Fast Food Restaurant	Shopping Mall	Furniture / Home Store	Pizza Place	Restaurant	Italian Restaurant	Dim Sum Restaurant	Diner
0	Dongguan	2,460.00	8.39	14,951.00	51.10	120.22	1.36	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Pizza Place	Hotel	Resort	Bar	Shopping Mall	Thai Restaurant	Sandwich Place
0	Zhongshan	1,784.00	3.31	16,711.00	49.30	27.23	0.53	Coffee Shop	Fast Food Restaurant	Hotel	Cantonese Restaurant	Shopping Mall	Spa	Park	Motel	Golf Course	Asian Restaurant
0	Zhuhai	1,736.00	1.89	24,100.00	49.10	28.52	2.39	Coffee Shop	Hotel	Chinese Restaurant	Café	Italian Restaurant	Portuguese Restaurant	Fast Food Restaurant	Church	Plaza	Steakhouse
1	Hong Kong	1,107.00	7.48	48,673.00	92.43	530.44	110.73	Hotel	Café	Dumpling Restaurant	Park	Clothing Store	Coffee Shop	Japanese Restaurant	Chinese Restaurant	Shopping Mall	Electronics Store
2	Macao	33.00	0.67	82,609.00	94.93	1.51	0.38	Hotel	Café	Portuguese Restaurant	Chinese Restaurant	Coffee Shop	Resort	Italian Restaurant	Historic Site	Church	Plaza
3	Guangzhou	7,434.00	14.90	23,497.00	71.80	84.74	6.61	Hotel	Park	Coffee Shop	Shopping Mall	Turkish Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Cantonese Restaurant	Restaurant	Cocktail Bar
3	Shenzhen	1,997.00	13.03	28,647.00	58.80	245.94	8.20	Hotel	Coffee Shop	Shopping Mall	Park	Japanese Restaurant	Lounge	Chinese Restaurant	Café	Electronics Store	Spa
4	Huizhou	11,347.00	4.83	12,908.00	43.00	33.38	0.96	Hotel	Shopping Mall	Coffee Shop	Fast Food Restaurant	Japanese Restaurant	Lake	Fish & Chips Shop	French Restaurant	Furniture / Home Store	Electronics Store
4	Jiangmen	9,507.00	4.60	9,570.00	44.50	16.97	0.73	Hotel	Bus Station	Fast Food Restaurant	Coffee Shop	Train Station	Shopping Mall	General Entertainment	Gym / Fitness Center	Grocery Store	Golf Course
4	Zhaoqing	14,891.00	4.15	8,050.00	38.60	3.59	0.14	Hotel	Playground	Park	Big Box Store	Zoo	Gym / Fitness Center	Grocery Store	Golf Course	Gift Shop	German Restaurant

the tertiary industry (around 50%). The level of FDI is also at a relatively low level. Population-wise, there is quite a bit of variation, ranging from 2 to 8 million. On the other hand, these cities have relatively similar distribution of venues. The most common venue type for all four cities is coffee shop, while the other ranks range from restaurants, hotels, parks, shops, etc. Hence, these cities can be categorised as moderately developed cities that probably caters for more domestic residents than foreign business or tourism.

11. Next, the cluster 3 comprises the two most populated cities in the GBA, Guangzhou and Shenzhen, each with population exceeding 10 million. They have also a relatively high level of FDI, only second to Hong Kong. Yet, their level of tertiary industry contribution to GDP are moderate, at around 60 – 70%, which is much lower than that of Hong Kong or Macao. This reflects that a sizeable proportion of their business activities are still manufacturing based. In terms of GDP per capita, they are at a slightly higher level in the mid US$20,000s, but still not comparable with that of Hong Kong, the same goes for export. It is also noteworthy that these two cities have quite drastic difference in their land areas, which does not affect them being clustered together. In terms of venues, they are still largely focused on hotel and restaurants. However, it is interesting to see some more niche and unique types of establishment in these two cities, such as Turkish restaurant, Cocktail bar, etc. This can possibly be explained by the fact these two cities are the major rapidly developing business centres of the Guangdong province, where the population is relatively young. Thus, this cluster can be described as the up-and-rising innovative manufacturing hub in the GBA.

12. Last but not least, the final cluster comprises Huizhou, Jiangmen and Zhaoqing, probably the three least heard of cities in the GBA. Yet, they are also the largest cities in terms of land area, all above 10 000 km^2. Their population is moderate but at a similar level of around 4 million. The GDP per capita in these regions are the lowest, with Jiangmen and Zhaoqing below US$10,000, with minimal tertiary industry contribution at around 30 – 40%. Their export and FDI levels are also low. This reflects that these cities are most likely more backward cities primarily focusing on lower-level manufacturing and agriculture. While the venue distribution of these cities looks relatively similar to others, with hotels being the top venue, more nature-related spots are also identified, such as parks, lakes and golf courses. Major transit

locations, such as bus stops and train stations, are also key venues. It is crucial to note that the number of venues recommended in these cities are very limited so it is likely that these tourist-related venues such as hotels and stations are more prominent in the database, which indirectly demonstrates the positioning of these cities as less developed agriculture and manufacturing cities that are mainly inward-looking.

Discussion

13. Having clustered the 11 GBA cities into 5 clusters following the prescribed methodology, this section discusses the results and identifies some of the limitations of the methodology.

14. Firstly, two robustness tests were conducted on the results, one without data standardisation and one without demographic data all together. This first case reveals that the demographic data would dominate the result, particularly the data on land area and population which are more diverse in magnitude. The clustering results show that Zhuhai and Macao grouped together, and Hong Kong, Guangzhou and Shenzhen each with their own cluster, while the remaining cities form the last cluster. This does not provide much insight into the differentiation of cities, as it is dominated by macro geographical factors. Thus, this sheds light on the importance of standardising data.

15. On the other hand, conducting clustering with only venue recommendation data from the Foursquare API, the clustering results are also quite drastically different, with only the Shenzhen-Guangzhou cluster remaining unchanged, while other cities are grouped into pairs with Jiangmen singled out. This is probably because Jiangmen is the only city with transport stations as the major venue recommendations. In view of the diverse results obtained from using different parts of the data set, it is reasonable to consider that a combination of standardised demographic data and venue data would present a more reasonable and interpretable results for our analysis, and thus this approach is chosen for the study.

16. Moreover, this also draws out a major limitation of the methodology of this study. While the dataset is undoubtedly small with only 11 cities, the more significant limitation is that Foursquare data may not be readily available for certain less developed cities in the GBA, with Zhaoqing only having 6 venues within its 10 km radius. The choice of centre point to locate the respective cities would also have sizeable implications on the exploration results. Therefore, the approach of using Foursqare API might not be ideal in this context. Nonetheless, the results presented in this study is still relatively reasonable and functions well as an exploratory venture for further studies.

Conclusion

17. This study utilises Foursquare API location data and demographic data obtained from other sources to cluster the 11 GBA cities into 5 distinct clusters with their only unique positioning and characteristics. Hong Kong and Macao, the two Special Administrative Regions earned their only special role as the key commerce and tourism centre respectively. While the two most populated cities, Guangzhou and Shenzhen, are the relatively more developed Mainland cities with a young and vibrant innovative manufacturing-based economy. For Zhuhai, Foshan, Donguan, and Zhongshan, they are moderately developed inland cities with focus on domestic residents, while the remaining Jiangmen, Zhaoqing, and Huizhou, are the least developed cities with more focus on agriculture with the largest land area. Having identified the positioning of the 11 cities, this exploratory study could shed light on how future studies and develop plan could harness the potential synergy between these diverse cities, with a view to taking forward the GBA development.