APPLIED DATA SCIENCE CAPSTONE

Capstone Project: The Battle of Neighborhoods

Self-designed topic: *The investigation of business opportunities by comparing models of top 10 best foreign-earned-income cities all around the world*

Author: Thai Quoc Tran

Date submission: March 14, 2020

*This report is available in https://github.com/thaitran7/Coursera_Capstone, which also includes this project's presentation and related jupyter notebook.

I. Introduction

1. How business opportunities are so important for entrepreneurs?

Most of cities around the world provides lot of business opportunities and motivities to investors and entrepreneurs, which also gives more jobs to promote economic development of nations. Many international visitors are mainly merchants and tourists boost income of local business. Also, many cities attract these persons to open widely their markets. However, to adapt in the highly competitive markets for earning income, entrepreneurs have to consider necessary factors as the demand markets and current business status in the attractive areas of business. Therefore, they should investigate the business venture and expansion needs. With the insights of business patterns, such people who want to open own business can get into the market with spectacular successes.

2. Problem statements and Interest

It is important for entrepreneurs or business investors to strategically plan for their survival in those competitive markets. They need some information about advantage points, current patterns of business to operate it successfully. To gather such information, the entrepreneurs have to determine the best cites, which were reported with high income from international visitors. Moreover, to evaluate of business plans and models, they have to learn the business patterns in multicultural countries to balance the demands and supplies in the markets where they will join.

II. Data acquisition and processing

1. Data sources

- The ranking list of cities which have the number of international visitors and income from them, was published in Wikidata (**https://en.wikipedia.org/wiki/List_of_cities_by_international_visitors**) without its coordinates. To fill the geographic data, I use the tool of Nominatim to retrieve the location of the cities in the world map. Another sources of related data is to explore the nearby venues by using the Foursquare API.

2. **Data cleaning**

   The data was downloaded from Wikidata had to be cleansed by sorting the best income cities (based on income value instead of ranking data, dropping not available data and creating new dataframe with only top 10 of the cities, its countries, and its income values (unit: billion US dollars). In addition, ranking value was set up with the income values, which is not from the original table.

3. **Feature selection**

   All the venues were retrieved by using Foursquare include coordinates (latitude and longitude), names, and categories. The analysis of venues is to perform statistics on their categories and then group the same distribution patterns of each explored location.

III. Methodology

1. **Exploring and locating places/venues**

   To get the cities' coordination, I used the Nominatim tool of geopy package by searching its names. And then the data was inserted into 'top 10 best income cities' dataframe with two columns of latitude and longitude.

   For exploring nearby venues and get the related information, the familiar tool were used as Foursquare API with personal access with CLIENT_ID, CLIENT_SECRET, and date version of it.

2. **Geographic Clusters of locations and groups of similar business patterns in each cluster**

   In this project, I perform k-mean clustering methods to cluster areas of high-density venues and business distribution patterns of each area. Importantly, optimization of k clusters had been recommended to find the best numbers of clusters, which were based on elbow graph. In the results of k-mean clustering, 1,000 venues are grouped into clusters by its density in each city and business distribution of cities' clusters are grouped as familiar business patterns.

3. **Mapping interesting places and clusters in world map**

   With the supports of folium package, it found easy to map the locations with interests as cities and locations of clusters with having coordinates as latitude and longitude. Moreover, I used available map from https://raw.githubusercontent.com/parulnith/Visualising-Geospatial-data-with-Python/master/world-countries.json to visualize the annotation in world map.

IV. Results

1. **The top 10 world cities, which have the best income from foreign customers (Table 1)**

Ranking	City	Country	Income_billion$	Latitude	Longitude
1	Dubai	United Arab Emirates	31.30	25.065700	55.171300
2	London	United Kingdom	19.76	51.507322	-0.127647
3	New York City	United States of America	18.52	40.712728	-74.006015
4	Bangkok	Thailand	14.84	13.754253	100.493087
5	Tokyo	Japan	13.48	35.682839	139.759455
6	Paris	France	12.88	48.856697	2.351462
7	Singapore	Singapore	12.54	1.357107	103.819499
8	Kuala Lumpur	Malaysia	11.34	3.151696	101.694237
9	Seoul	South Korea	11.30	37.566679	126.978291
10	Taipei	Taiwan	9.60	25.037520	121.563680

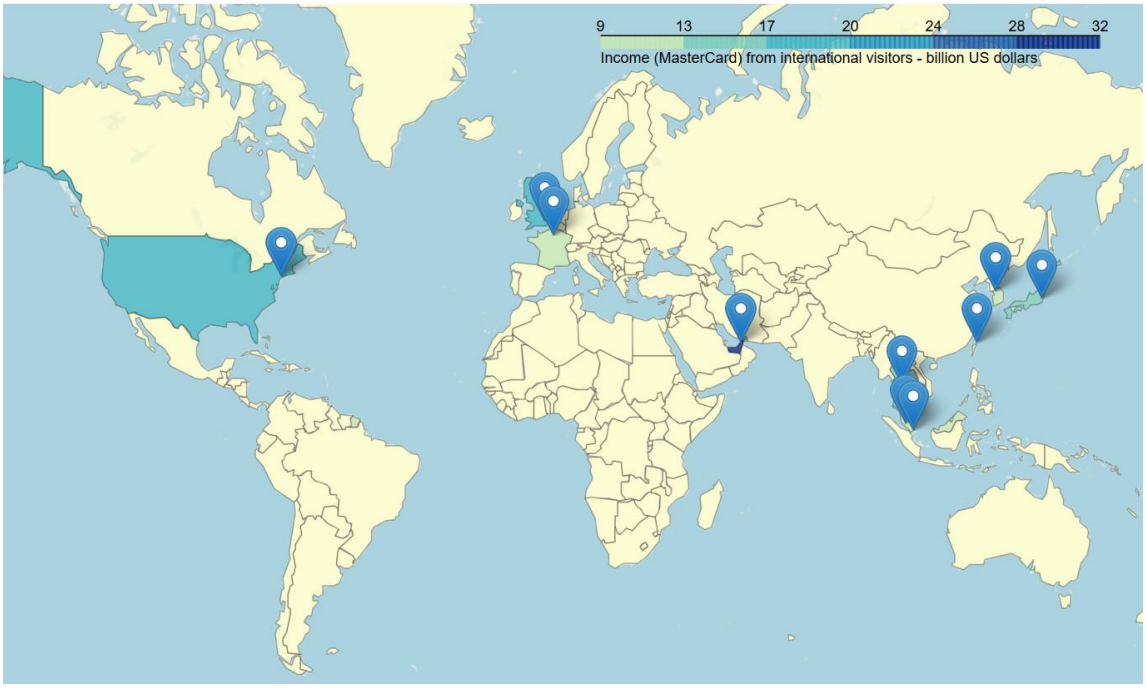


Figure 1. Location of these cities and their countries

2. The geographic clusters with the high density of venues in the cities (Table 2)

	Country	City	Advantage points	Adv Latitude	Adv Longitude
0	United Arab Emirates	Dubai	D1	25.075501	55.145859
1	United Arab Emirates	Dubai	D2	25.058710	55.179493
2	United Arab Emirates	Dubai	D3	25.088783	55.168106
3	United Kingdom	London	L1	51.510417	-0.125521
4	United Kingdom	London	L2	51.509073	-0.131431
5	United Kingdom	London	L3	51.506569	-0.126501
6	United States of America	New York City	NY1	40.710595	-74.007638
7	United States of America	New York City	NY2	40.714981	-74.007962
8	Thailand	Bangkok	B1	13.752736	100.498033
9	Thailand	Bangkok	B2	13.753480	100.490267
10	Thailand	Bangkok	B3	13.757981	100.494926
11	Japan	Tokyo	Tk1	35.680889	139.764140
12	Japan	Tokyo	Tk2	35.683652	139.756037
13	Japan	Tokyo	Tk3	35.684921	139.762806
14	France	Paris	P1	48.856740	2.356175
15	France	Paris	P2	48.857261	2.348593
16	France	Paris	P3	48.858968	2.352753
17	Singapore	Singapore	Sp1	1.369585	103.830633
18	Singapore	Singapore	Sp2	1.347414	103.813966
19	Singapore	Singapore	Sp3	1.353628	103.834504
20	Malaysia	Kuala Lumpur	Kl1	3.148660	101.696939
21	Malaysia	Kuala Lumpur	Kl2	3.154565	101.696474
22	Malaysia	Kuala Lumpur	Kl3	3.151855	101.693946
23	South Korea	Seoul	Se1	37.569466	126.978108
24	South Korea	Seoul	Se2	37.565283	126.980273
25	South Korea	Seoul	Se3	37.565958	126.974887
26	Taiwan	Tapei	Tp1	25.035146	121.565110
27	Taiwan	Tapei	Tp2	25.039692	121.566086

3. Ranking the most frequency venues in each cluster (Table 3)

	Advantage points	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	B1	Noodle House	Thai Restaurant	Asian Restaurant	Hotel	Café	Hostel	Dessert Shop	Bar	Palace	Buddhist Temple
1	B2	Noodle House	Palace	Café	Pharmacy	Thai Restaurant	History Museum	Bakery	Coffee Shop	Ice Cream Shop	Theater
2	B3	Bar	Noodle House	Café	Asian Restaurant	Hotel	Thai Restaurant	Massage Studio	Vegetarian / Vegan Restaurant	Bakery	History Museum
3	D1	Hotel	Coffee Shop	Gym	Spa	Café	Tram Station	Breakfast Spot	Burger Joint	Middle Eastern Restaurant	Indian Restaurant
4	D2	Coffee Shop	Hotel	Supermarket	Park	Café	Fast Food Restaurant	Shopping Mall	Restaurant	Convenience Store	Gym
5	D3	Hotel	Coffee Shop	Café	Bar	Gym / Fitness Center	Italian Restaurant	Asian Restaurant	Spa	French Restaurant	Middle Eastern Restaurant
6	KI1	Indian Restaurant	Café	Restaurant	Malay Restaurant	Hotel	Asian Restaurant	Food Truck	Convenience Store	South Indian Restaurant	Coffee Shop
7	KI2	Indian Restaurant	Hotel	Food Court	Malay Restaurant	Café	Coffee Shop	Cafeteria	Bakery	Shoe Store	Restaurant
8	KI3	Indian Restaurant	Coffee Shop	Malay Restaurant	Café	Hotel	Food Court	Boutique	South Indian Restaurant	Clothing Store	Food Truck
9	L1	Coffee Shop	Theater	Bakery	Wine Bar	Clothing Store	Ice Cream Shop	Pub	French Restaurant	Dessert Shop	Cosmetics Shop
10	L2	Theater	Hotel	Ice Cream Shop	Italian Restaurant	Cocktail Bar	Bookstore	Seafood Restaurant	Plaza	Bakery	Speakeasy
11	L3	Hotel	Theater	Monument / Landmark	Cocktail Bar	Wine Bar	Garden	Sandwich Place	Pub	Plaza	French Restaurant
12	NY1	Coffee Shop	Hotel	Pizza Place	Sandwich Place	Italian Restaurant	American Restaurant	Memorial Site	Wine Shop	Restaurant	Falafel Restaurant
13	NY2	Coffee Shop	Hotel	Gym / Fitness Center	Sandwich Place	Cocktail Bar	Indian Restaurant	French Restaurant	Café	Wine Shop	Gym
14	P1	French Restaurant	Clothing Store	Pastry Shop	Ice Cream Shop	Gourmet Shop	Wine Bar	Cosmetics Shop	Cocktail Bar	Garden	Coffee Shop
15	P2	French Restaurant	Plaza	Clothing Store	Ice Cream Shop	Art Gallery	Historic Site	Hotel	Cocktail Bar	Jazz Club	Gay Bar
16	P3	French Restaurant	Art Gallery	Plaza	Ice Cream Shop	Bakery	Coffee Shop	Cocktail Bar	Clothing Store	Pub	Tea Room
17	Se1	Korean Restaurant	Coffee Shop	Hotel	Café	Chinese Restaurant	Bookstore	Plaza	BBQ Joint	Sushi Restaurant	Burger Joint
18	Se2	Korean Restaurant	Hotel	Coffee Shop	Chinese Restaurant	Café	Japanese Restaurant	Plaza	Dessert Shop	Historic Site	Lounge
19	Se3	Coffee Shop	Korean Restaurant	Hotel	Café	Historic Site	Chinese Restaurant	Lounge	Noodle House	Plaza	Art Museum
20	Sp1	Food Court	Chinese Restaurant	Asian Restaurant	Park	Indian Restaurant	Café	Thai Restaurant	Ice Cream Shop	Seafood Restaurant	Restaurant
21	Sp2	Café	Chinese Restaurant	Thai Restaurant	Japanese Restaurant	Coffee Shop	Soccer Field	Bakery	Trail	Ice Cream Shop	Italian Restaurant
22	Sp3	Chinese Restaurant	Café	Thai Restaurant	Coffee Shop	Ice Cream Shop	Spa	Vegetarian / Vegan Restaurant	Sushi Restaurant	Food Court	Dessert Shop
23	Tk1	Japanese Restaurant	Café	Chinese Restaurant	French Restaurant	Ramen Restaurant	Sushi Restaurant	Gourmet Shop	Deli / Bodega	Hotel Bar	Dessert Shop
24	Tk2	Historic Site	Café	Art Museum	Hotel	Lounge	Japanese Restaurant	Sushi Restaurant	Garden	Italian Restaurant	Bridge
25	Tk3	Japanese Restaurant	Café	Historic Site	Hotel	Sushi Restaurant	Chinese Restaurant	Italian Restaurant	Lounge	Thai Restaurant	Hotel Bar
26	Tp1	Department Store	Coffee Shop	Chinese Restaurant	Lounge	Steakhouse	Hotel	Hotpot Restaurant	Gym / Fitness Center	Plaza	Bakery
27	Tp2	Department Store	Hotel	Japanese Restaurant	Café	Coffee Shop	Taiwanese Restaurant	Furniture / Home Store	Lounge	Chinese Restaurant	Hotpot Restaurant

4. Comparison of venues' frequency distribution as business models (Table 4)

	Country	City	Advantage points	Adv Latitude	Adv Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	United Arab Emirates	Dubai	D1	25.075501	55.145859	2	Hotel	Coffee Shop	Gym	Spa	Café
1	United Arab Emirates	Dubai	D2	25.058710	55.179493	2	Coffee Shop	Hotel	Supermarket	Park	Café
2	United Arab Emirates	Dubai	D3	25.088783	55.168106	2	Hotel	Coffee Shop	Café	Bar	Gym / Fitness Center
3	United Kingdom	London	L1	51.510417	-0.125521	5	Coffee Shop	Theater	Bakery	Wine Bar	Clothing Store
4	United Kingdom	London	L2	51.509073	-0.131431	5	Theater	Hotel	Ice Cream Shop	Italian Restaurant	Cocktail Bar
5	United Kingdom	London	L3	51.506569	-0.126501	5	Hotel	Theater	Monument / Landmark	Cocktail Bar	Wine Bar
6	United States of America	New York City	NY1	40.710595	-74.007638	2	Coffee Shop	Hotel	Pizza Place	Sandwich Place	Italian Restaurant
7	United States of America	New York City	NY2	40.714981	-74.007962	2	Coffee Shop	Hotel	Gym / Fitness Center	Sandwich Place	Cocktail Bar
8	Thailand	Bangkok	B1	13.752736	100.498033	8	Noodle House	Thai Restaurant	Asian Restaurant	Hotel	Café
9	Thailand	Bangkok	B2	13.753480	100.490267	8	Noodle House	Palace	Café	Pharmacy	Thai Restaurant
10	Thailand	Bangkok	B3	13.757981	100.494926	8	Bar	Noodle House	Café	Asian Restaurant	Hotel
11	Japan	Tokyo	Tk1	35.680889	139.764140	3	Japanese Restaurant	Café	Chinese Restaurant	French Restaurant	Ramen Restaurant
12	Japan	Tokyo	Tk2	35.683652	139.756037	3	Historic Site	Café	Art Museum	Hotel	Lounge
13	Japan	Tokyo	Tk3	35.684921	139.762806	3	Japanese Restaurant	Café	Historic Site	Hotel	Sushi Restaurant
14	France	Paris	P1	48.856740	2.356175	5	French Restaurant	Clothing Store	Pastry Shop	Ice Cream Shop	Gourmet Shop
15	France	Paris	P2	48.857261	2.348593	5	French Restaurant	Plaza	Clothing Store	Ice Cream Shop	Art Gallery
16	France	Paris	P3	48.858968	2.352753	5	French Restaurant	Art Gallery	Plaza	Ice Cream Shop	Bakery
17	Singapore	Singapore	Sp1	1.369585	103.830633	9	Food Court	Chinese Restaurant	Asian Restaurant	Park	Indian Restaurant
18	Singapore	Singapore	Sp2	1.347414	103.813966	3	Café	Chinese Restaurant	Thai Restaurant	Japanese Restaurant	Coffee Shop
19	Singapore	Singapore	Sp3	1.353628	103.834504	6	Chinese Restaurant	Café	Thai Restaurant	Coffee Shop	Ice Cream Shop
20	Malaysia	Kuala Lumpur	Kl1	3.148660	101.696939	7	Indian Restaurant	Café	Restaurant	Malay Restaurant	Hotel
21	Malaysia	Kuala Lumpur	Kl2	3.154565	101.696474	7	Indian Restaurant	Hotel	Food Court	Malay Restaurant	Café
22	Malaysia	Kuala Lumpur	Kl3	3.151855	101.693946	7	Indian Restaurant	Coffee Shop	Malay Restaurant	Café	Hotel
23	South Korea	Seoul	Se1	37.569466	126.978108	4	Korean Restaurant	Coffee Shop	Hotel	Café	Chinese Restaurant
24	South Korea	Seoul	Se2	37.565283	126.980273	4	Korean Restaurant	Hotel	Coffee Shop	Chinese Restaurant	Café
25	South Korea	Seoul	Se3	37.565958	126.974887	4	Coffee Shop	Korean Restaurant	Hotel	Café	Historic Site
26	Taiwan	Tapei	Tp1	25.035146	121.565110	1	Department Store	Coffee Shop	Chinese Restaurant	Lounge	Steakhouse
27	Taiwan	Tapei	Tp2	25.039692	121.566086	1	Department Store	Hotel	Japanese Restaurant	Café	Coffee Shop

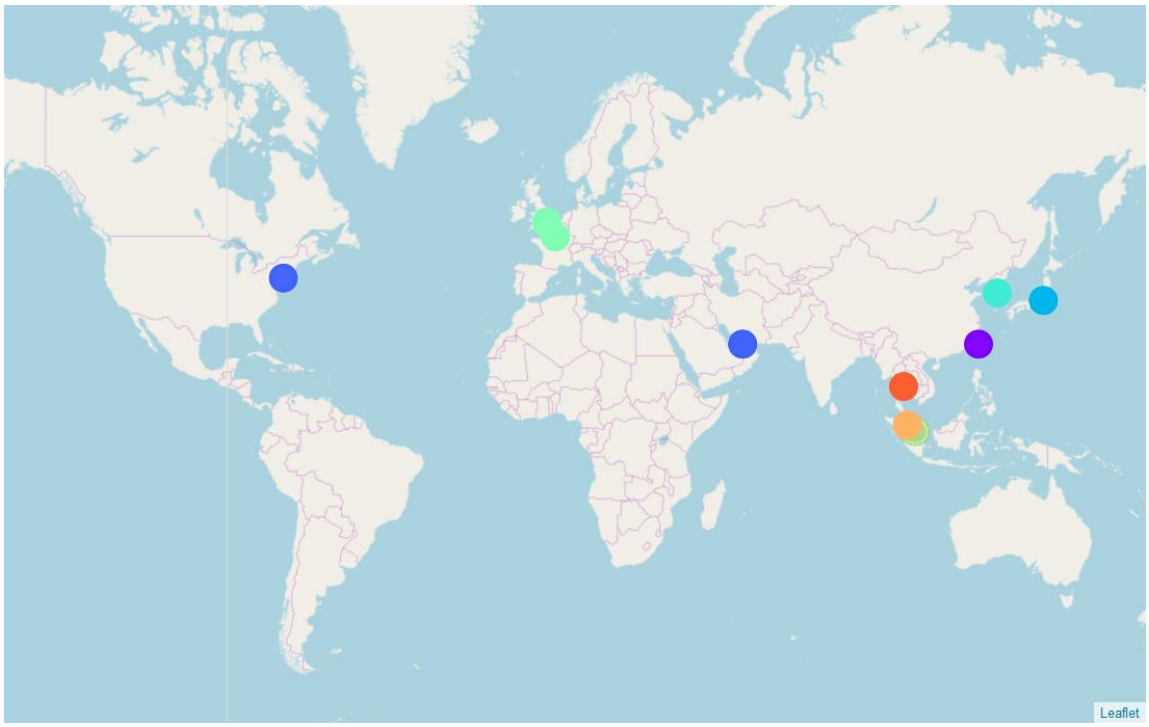
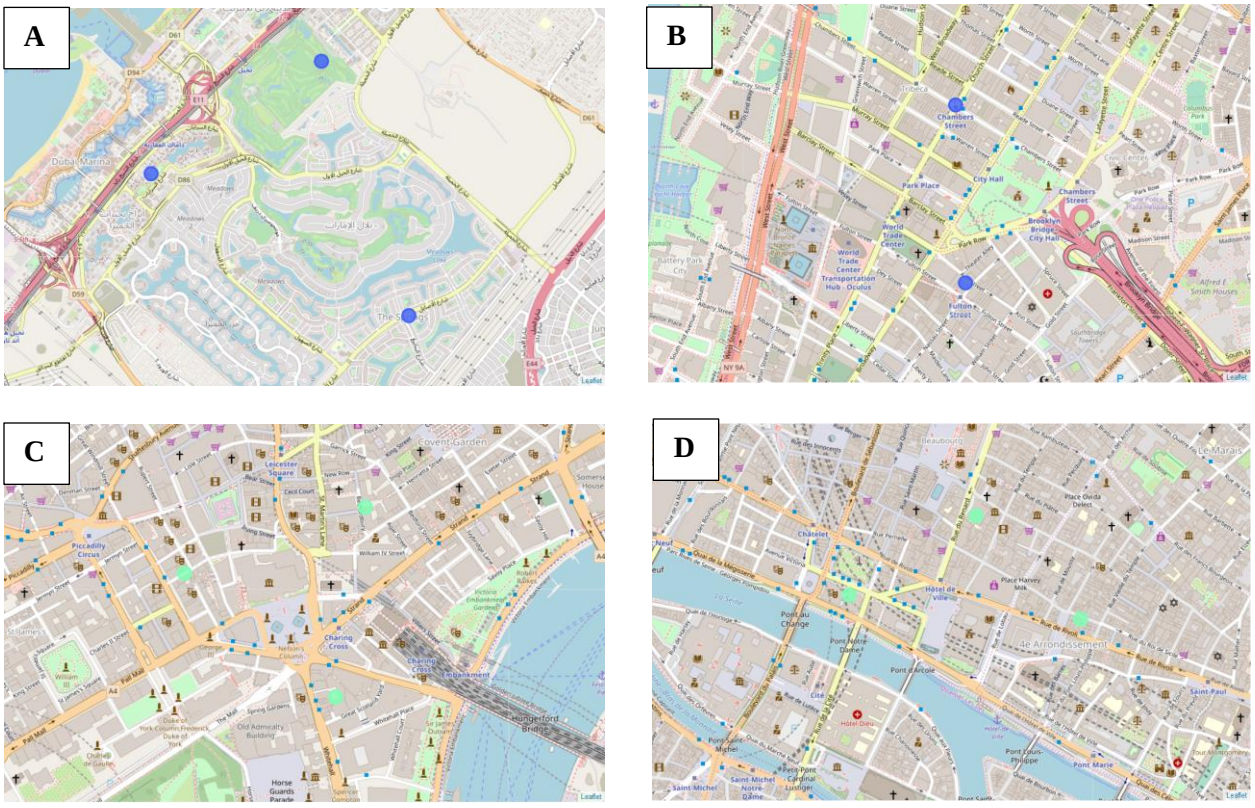


Figure 2. Comparison of venues' frequency distribution as business models in the GLOBAL MAP



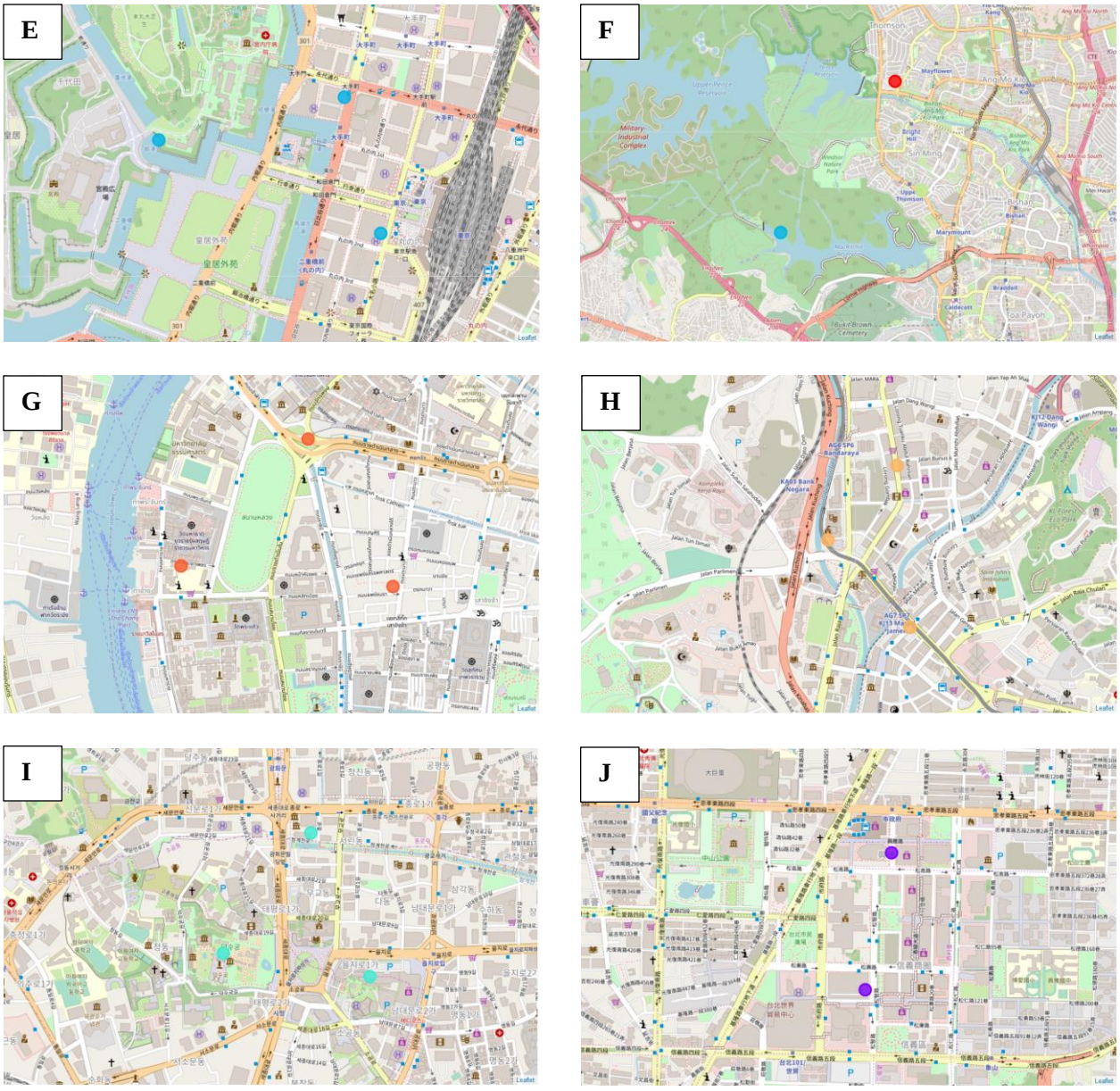


Figure 3. Comparison of venues' frequency distribution as business models in the LOCAL MAP. Dubai (A), New York City (B), London (C), Paris (D), Tokyo (E), Singapore (F), Bangkok (G), Kuala Lumpur (H), Seoul (I), and Taipei (J).

V. Discussion

The top 10 cities which have the best income from international visitors: with the first rank is Dubai and the tenth rank is Taipei. (Table 1 and Figure 1).

Around the advantage points of cities, the most common venues, include restaurants, hotels, and café, were found similar in the 1st, 2nd, and 3rd ranks. (Table 2, 3)

The business patterns of Dubai (1st rank) and New York City (3rd), London (2nd) and Paris (6th), Tokyo (5th) and one investigating areas of Singapore (7th) are similar, except which of Bangkok (4th), Kuala

Lumpur (8th), Seoul (9th), and Taipei (10th) are different to the others and unique. (Table 2-4 and Figure 2,3)

VI. Conclusion

The top 10 best income cities, which is dependent on international visitors, have the highest rank as Dubai. Moreover, there are similar features with common venues in distance cities (New York and Dubai), European ones (London and Paris), Asian ones (Tokyo and Singapore).

VII. Future prospects

- **From the research:**

Best income business pattern belongs to Dubai and New York can influence the trends of the others, which follows the modern western business.

- **Future prospects:**

The research needs to investigate more advantage areas in different cultural regions, which helps to suggest the better the models being applied. From each model, we can balance the demands of locals and supplies of current business, which provide us better recommendation to open own business with specific location.