

# City Advisor

Antonio Saccoman

# Objective, methodology, assumptions

## Objective

Recommendations around best areas in world metropolitan areas, with view to purchase property in the best area.

The key question to be answered is: based on available venues in the selected areas, which area best fits the client's personal preferences with regards to venue categories?

## Methodology

1. select a shortlist of cities;
2. subdivide city centre into 10 to 20 circular areas (radius = 1Km);
3. retrieve areas geographical coordinates (via Nominatim), and venues information for each area (via Foursquare's API);
4. explore venues data via data visualization tools and Folium mapping;
5. apply clustering via machine learning, to observe any similarity across different areas;
6. apply scoring functions to rank areas, in order to identify the most desirable ones. The scoring algorithm is based on venue *tiering*, i.e. a highly customizable feature which aims to reflect the assumed client's personal preferences with regards to venue categories.

## Assumptions

Assumptions over client's personal preferences are multiple: the client is considered to be neutral to features such as *language, currency, tax regime, local weather, geographical location*, etc. The client is also assumed to prefer a **central** area, with as many venues as possible; other assumptions will be made about the personal preferences over venue categories.

# Cities overview

Table sourced from Numbeo website.  
Various economic and social indicators for the selected cities.

	City	Country	\$ price per sqm*	Quality of Life Index	Cost of Living Index	Groceries Index	Restaurant Price Index	Health Care Index	Crime Index	Pollution Index	Weather Index	Gross Rental Yield**	Affordability Index
1	Dubai	UAE	3653	154.03	68.83	53.87	65.63	67.34	16.70	50.65	50.27	10.10	2.62
2	London	UK	14284	128.23	81.10	59.12	82.82	70.28	52.71	58.57	88.25	3.34	0.93
3	Milan	Italy	10214	117.27	82.37	71.36	79.45	71.57	43.19	66.06	88.12	3.16	0.87
4	New York	USA	15685	139.69	100.00	100.00	100.00	62.96	45.14	57.00	79.66	5.10	1.39
5	Paris	France	14105	117.69	91.01	89.96	77.63	78.58	51.85	64.23	88.39	2.36	0.80
6	S.Francisco	USA	13489	161.12	92.68	91.85	87.01	66.02	55.30	46.95	97.26	5.70	1.89
7	Singapore	Singapore	20086	140.47	82.51	71.56	57.91	71.07	32.10	33.26	57.45	2.35	0.71
8	Sydney	Australia	10136	175.79	84.63	77.99	71.35	77.75	33.32	26.82	97.07	4.07	1.39
9	Tokyo	Japan	11274	153.26	91.75	94.46	49.48	80.13	23.40	42.25	85.26	2.65	1.05
10	Zurich	Switzerland	13639	196.06	136.91	139.13	121.60	74.58	16.56	17.17	81.48	3.43	2.22





# Areas selection and venues info

Table of number of areas per city.

City	Dubai	London	Milan	New York	Paris	San Francisco	Singapore	Sydney	Tokyo	Zurich	Total Areas
#Areas	16	16	15	16	15	19	12	14	16	11	150

Examples of areas selection and venue count, and mapping (Sydney).

Venue		
Borough	City+	
Bondi	Sydney, Australia	94
Chippendale - Ultimo	Sydney, Australia	100
City (*)	Sydney, Australia	100
Clovelly	Sydney, Australia	48
Elizabeth Bay House	Sydney, Australia	100
Forest Lodge	Sydney, Australia	100
Lavender Bay	Sydney, Australia	100
Macdonaldtown	Sydney, Australia	100
Pyrmont	Sydney, Australia	100
Randwick	Sydney, Australia	76
Rozelle	Sydney, Australia	92
Surry Hills - Darlinghurst	Sydney, Australia	100
Waterloo	Sydney, Australia	100
Woollahra - Double Bay	Sydney, Australia	58

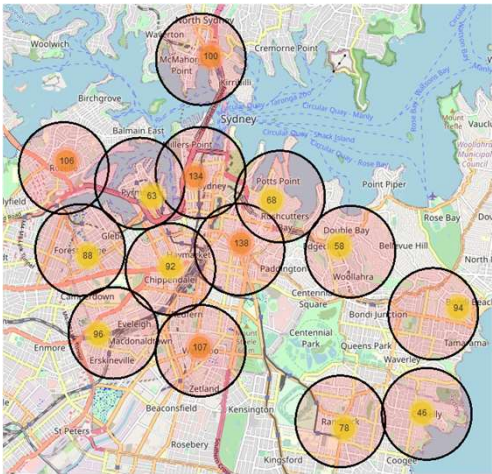


Table of most frequent venues per city.

Most frequent venue categories per city

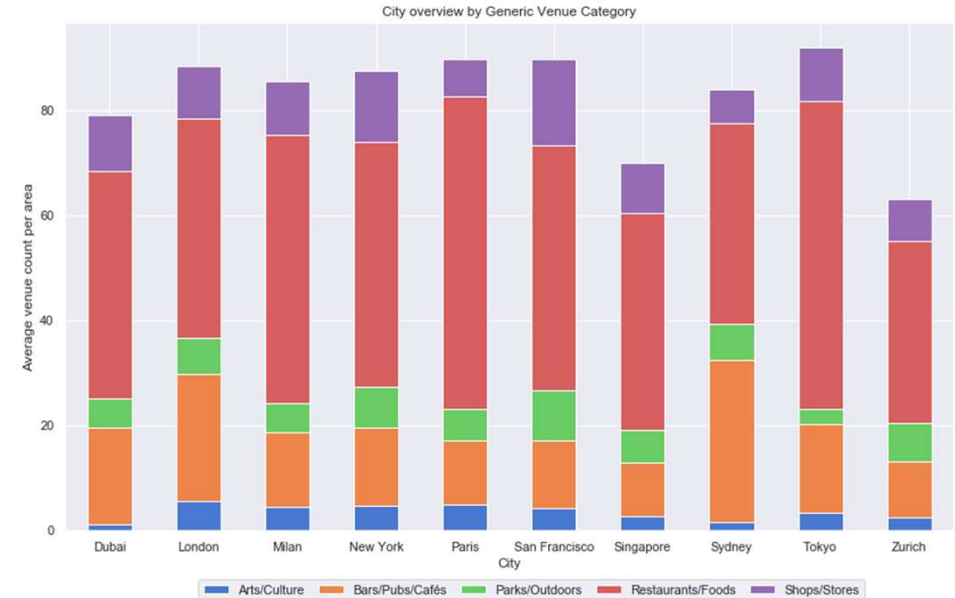
	Dubai	London	Milan	New York	Paris	San Francisco	Singapore	Sydney	Tokyo	Zurich
1	Hotel	Pub	Italian Restaurant	Coffee Shop	French Restaurant	Coffee Shop	Hotel	Café	Japanese Restaurant	Italian Restaurant
2	Café	Coffee Shop	Pizza Place	Italian Restaurant	Hotel	Park	Chinese Restaurant	Bar	Ramen Restaurant	Swiss Restaurant
3	Coffee Shop	Café	Café	Park	Italian Restaurant	Bakery	Coffee Shop	Park	Café	Café
4	Middle Eastern Restaurant	Hotel	Ice Cream Shop	American Restaurant	Bakery	Café	Café	Thai Restaurant	Coffee Shop	Hotel
5	Indian Restaurant	Italian Restaurant	Hotel	Bakery	Coffee Shop	Pizza Place	Japanese Restaurant	Pub	Sake Bar	Restaurant
6	Restaurant	Bakery	Plaza	Café	Plaza	Chinese Restaurant	Food Court	Coffee Shop	BBQ Joint	Bar
7	Asian Restaurant	Park	Japanese Restaurant	Mexican Restaurant	Bar	Sushi Restaurant	Bakery	Italian Restaurant	Italian Restaurant	Supermarket
8	Burger Joint	Gym / Fitness Center	Cocktail Bar	Gym	Japanese Restaurant	Wine Bar	Noodle House	Pizza Place	Chinese Restaurant	Bakery
9	Fast Food Restaurant	Theater	Restaurant	Pizza Place	Vietnamese Restaurant	Mexican Restaurant	Indian Restaurant	Bakery	Soba Restaurant	Tram Station
10	Gym / Fitness Center	French Restaurant	Seafood Restaurant	Grocery Store	Café	Vietnamese Restaurant	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Park

## Generic Venue Categories (GVC)

**GVC** = macro-categories to group Foursquare's venue categories (500+ types)

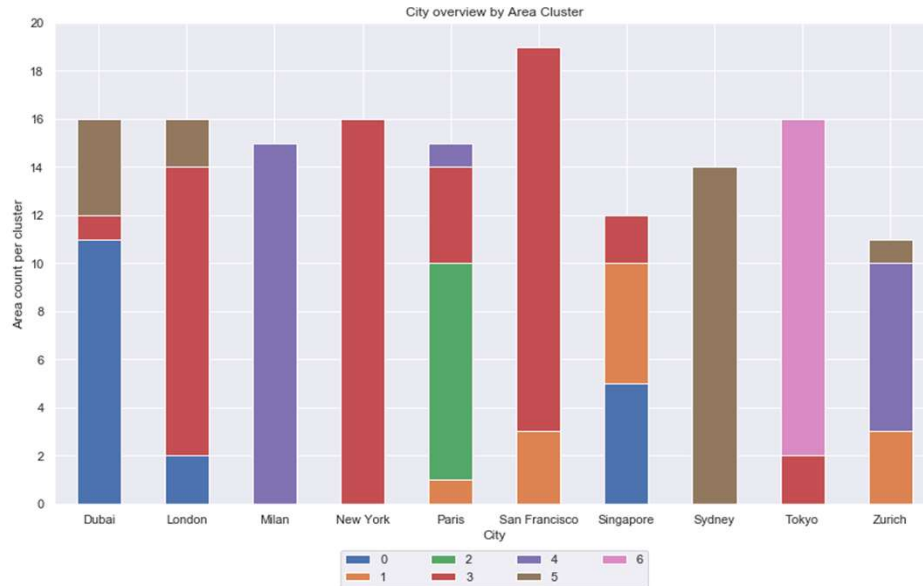
Most frequent venue categories per GVC

	Arts/Culture	Bars/Pubs/Cafés	Parks/Outdoors	Restaurants/Foods	Shops/Stores
1	Theater	Café	Park	Italian Restaurant	Supermarket
2	Art Gallery	Coffee Shop	Plaza	French Restaurant	Grocery Store
3	Art Museum	Bar	Garden	Bakery	Bookstore
4	Museum	Pub	Playground	Japanese Restaurant	Clothing Store
5	Movie Theater	Cocktail Bar	Pool	Pizza Place	Spa
6	Historic Site	Wine Bar	Scenic Lookout	Chinese Restaurant	Boutique
7	Performing Arts Venue	Juice Bar	Trail	Restaurant	Convenience Store
8	Concert Hall	Sake Bar	Beach	Ice Cream Shop	Cosmetics Shop
9	Music Venue	Beer Bar	Beer Garden	Indian Restaurant	Shopping Mall
10	Monument / Landmark	Hotel Bar	Dog Run	Thai Restaurant	Gift Shop

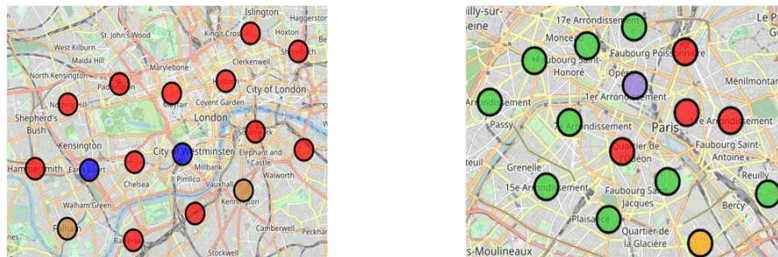


- ❑ **Sydney** dominates for Bars et simil venues, closely followed by London;
- ❑ **Paris** and **Tokyo** lead for Restaurants et simil, with Milan just behind;
- ❑ **San Francisco** tops the other cities for both Parks and Shops GVCs;
- ❑ **London** shows the highest Arts et simil average per area, then Paris and New York;
- ❑ **Singapore** and **Zurich** do not seem to be at the same level as other cities in terms of overall number of venues per area.

# Machine Learning (K-Means) Clustering



Examples of clustered areas mapping; London and Paris.



- ❑ **Cluster 0** (blue): composed prevalently of hotels, cafés, coffee shops, Middle Eastern and Indian restaurants. **Dubai** is mostly represented by this cluster; a good part of **Singapore** too. Total areas: **18**.
- ❑ **Cluster 1** (orange): displays multiple types of Asian restaurants (Chinese, Thai, Vietnamese, Korean, etc); some cafés, some bakeries. The remaining part of **Singapore** is mostly represented by this cluster; some parts of **San Francisco** and **Zurich** are to be mentioned too. Total areas: **12**.
- ❑ **Cluster 2** (green): mainly shows French restaurants, hotels, Italian restaurants, bakeries, bars. **Paris** (most of its areas) is the only city allocated to this cluster. Total areas: **9**.
- ❑ **Cluster 3** (red): this seems to be the “anything else” cluster, as compared to the others; there are no clear similarities among areas. This cluster displays a wide variety of restaurants of different cuisines, then cafes, coffee shops, bars, pubs, bakeries, parks and gardens. **New York** (in its entirety), then **San Francisco** and **London** are the main representatives of this “diversity” cluster. Total areas: **53**.
- ❑ **Cluster 4** (purple): most frequent venues are Italian restaurants and pizza places, hotels, cafes, bars, plazas, ice creams. **Milan** (in its entirety), then **Zurich** are the main representatives of this cluster. Total areas: **23**.
- ❑ **Cluster 5** (brown): predominantly includes cafés, coffee shops, bars, pubs, then some Asian and Italian restaurants, parks. **Sydney** (in its entirety) is the main representative of this cluster. Total areas: **21**.
- ❑ **Cluster 6** (pink): characterized mostly by the presence of Japanese restaurants (various local types), sake bars, BBQ joints, and some Italian restaurants. **Tokyo** (almost entirely) is the main representative of this cluster. Total areas: **14**.

## Venue Tiers

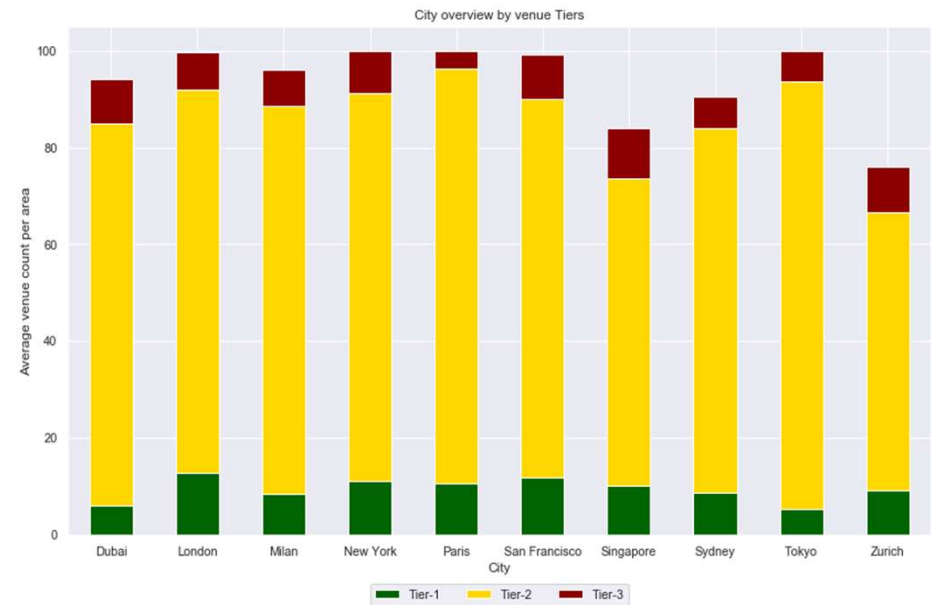
**Tiers** = features under which areas ranking is *customised* to each client, based on his/her personal preferences; as opposed to GVC, which is generic venue flag (macro-category).

Each of the 500+ unique venue categories from Foursquare API is assigned with a Tier representing the client's degree of appeal towards the given venue category:

- ❑ Tier-3 is assigned to any venue category client is *indifferent* to;
- ❑ Tier-2 is assigned to venue category that the client is assumed to consider good-to-have in the neighbourhood; and finally,
- ❑ Tier-1 is assigned to client's assumed most preferred venue categories, i.e. must-have in the area.

Most frequent venue categories per Tier

	Tier-1	Tier-2	Tier-3
1	Park	Café	Sandwich Place
2	Plaza	Coffee Shop	Burger Joint
3	Theater	Italian Restaurant	Vegetarian / Vegan Restaurant
4	Garden	Hotel	Yoga Studio
5	Deli / Bodega	French Restaurant	Fast Food Restaurant
6	Art Museum	Bakery	Nightclub
7	Food Court	Japanese Restaurant	Hostel
8	Museum	Pizza Place	Tram Station
9	Farmers Market	Bar	Salad Place
10	Pool	Chinese Restaurant	Bus Station



- ❑ **Tier-1** venues are on average most frequent in **London**, followed by San Francisco;
- ❑ Aggregate of **Tier-1 & Tier-2** venues is most frequent in **Paris**, then in Tokyo.

## Areas preliminary ranking; by Tier count

Areas Preliminary Ranking						
	Area	City	Tier-1	Tier-2	Tier-3	Total Venues
Position						
1	Southwark	London	27	65	8	100
2	Marina South	Singapore	25	54	13	92
3	Buena Vista	San Francisco	24	62	14	100
4	East Harlem - 104th St	New York	23	72	5	100
5	Invalides - La Tour-Maubourg	Paris	22	76	2	100
6	Anderson Bridge	Singapore	22	65	13	100
7	Riesbach	Zurich	21	70	6	97
8	Place Monge	Paris	20	80	0	100
9	Lincoln Square - 66 St	New York	20	75	5	100
10	National Library (Richelieu)	Paris	19	79	2	100

**Southwark** in London is the top ranked area thanks to highest number of Tier-1 venues; looking at the area venues, its high Tier-1 count is mostly due to the high concentration of Arts/Culture venues (14/27) - mostly due to Theatres (7/27) and Museums (4/27); plus Outdoors (7/27), and Food places other than regular restaurants (6/27).

Southwark , Tier-1 venues composition		
Generic Venue Category	Venue Category	Venue
Arts/Culture	Art Museum	3
	Concert Hall	1
	History Museum	1
	Performing Arts Venue	2
	Theater	7
Parks/Outdoors	Park	2
	Pedestrian Plaza	1
	Plaza	1
	Scenic Lookout	3
Restaurants/Foods	Deli / Bodega	1
	Farmers Market	2
	Street Food Gathering	3
Total Venues		27



## Areas final ranking; by Tier scoring

### Finalized points system

Venue Tiers	Tier-3	Tier-2	Tier-1
Points per Venue	1	3	7

### Areas Final Ranking

		Area	City	Tier-1	Tier-2	Tier-3	Total Venues	Total Score
Position								
1		Southwark	London	27	65	8	100	392
2		Invalides - La Tour-Maubourg	Paris	22	76	2	100	384
3		East Harlem - 104th St	New York	23	72	5	100	382
4		Place Monge	Paris	20	80	0	100	380
5		National Library (Richelieu)	Paris	19	79	2	100	372
6		Lincoln Square - 66 St	New York	20	75	5	100	370
7		Buena Vista	San Francisco	24	62	14	100	368
8		West Bermondsey	London	16	84	0	100	364
9		Riesbach	Zurich	21	70	6	97	363
10		Anderson Bridge	Singapore	22	65	13	100	362

### Some initial observations

- ❑ **Southwark** in London is confirmed as best area out of 150 selected world-wide, according to the described methodology; **Invalides** area in Paris and **East Harlem** in New York are the second and third best areas, respectively.
- ❑ Final top 10 areas list is just **30 points** wide in terms of score; meaning that just above 4 Tier-1 venues – or just 10 Tier-2 venues – make the difference between the 1st and the 10th area.
- ❑ **Paris** is the most frequent city in the top 10 areas, with 3 areas. Then both London and New York have 2 areas each.
- ❑ Only **6 cities** out of 10 are represented in the top 10 area ranking. The excluded cities are: Dubai, Milan, Sydney and Tokyo.
- ❑ The most frequent K-Means **cluster** (5 areas out of 10) is the number **3**.

## Top 10 areas summary tables

*Table of top 10 areas inclusive of both Tier counts and GVC counts.*

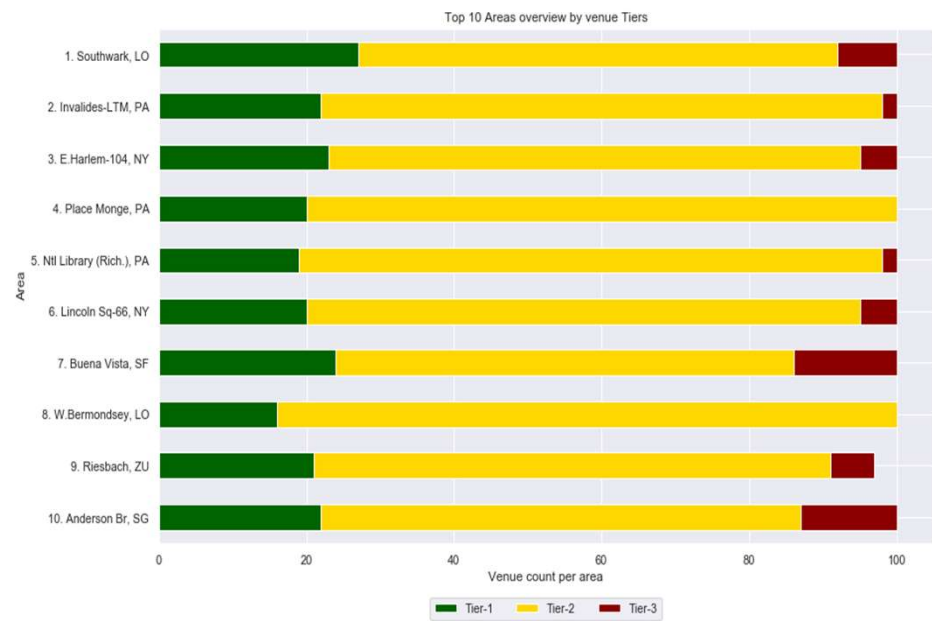
Area	Tier-1	Tier-2	Tier-3	Total Venues	Total Score	Arts / Culture	Bars / Pubs / Cafés	Parks / Outdoors	Restaurants / Foods	Shops / Stores
1.Southwark	27	65	8	100	392	17	19	8	42	6
2.Invalides	22	76	2	100	384	11	9	12	57	0
3.E.Harlem	23	72	5	100	382	4	14	17	45	8
4.PI Monge	20	80	0	100	380	10	15	11	55	2
5.Ntl Library	19	79	2	100	372	8	16	11	42	12
6.Lincoln Sq	20	75	5	100	370	14	12	9	29	13
7.Buena Vista	24	62	14	100	368	8	6	20	44	12
8.W.Bermonds	16	84	0	100	364	5	30	8	47	5
9.Riesbach	21	70	6	97	363	9	10	14	47	8
10.AndersonBr	22	65	13	100	362	11	11	12	30	4

*Table of top 10 areas most frequent venue categories; with K-Means cluster.*

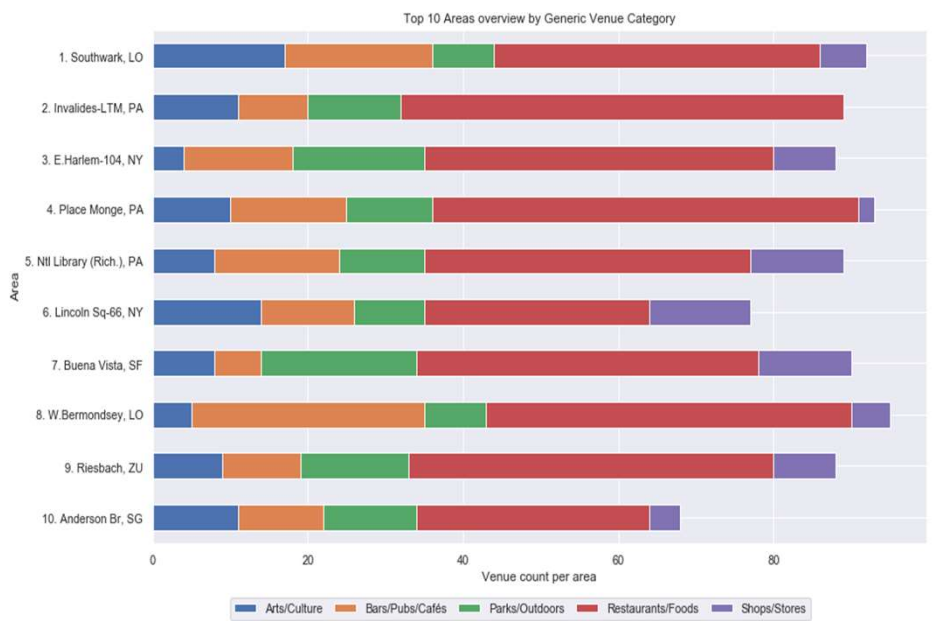
Area	City	Cluster#	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Southwark	London	3	Pub	Theater	Coffee Shop	Hotel	Italian Restaurant
Invalides - La Tour-Maubourg	Paris	2	French Restaurant	Hotel	Italian Restaurant	Plaza	Historic Site
East Harlem - 104th St	New York	3	Mexican Restaurant	Park	Gym	Coffee Shop	Deli / Bodega
Place Monge	Paris	2	French Restaurant	Coffee Shop	Italian Restaurant	Café	Hotel
National Library (Richelieu)	Paris	4	Hotel	Plaza	Japanese Restaurant	Historic Site	French Restaurant
Lincoln Square - 66 St	New York	3	Gym / Fitness Center	Italian Restaurant	Jazz Club	Plaza	Gym
Buena Vista	San Francisco	3	Seafood Restaurant	Park	Chocolate Shop	Historic Site	Bike Rental / Bike Share
West Bermondsey	London	3	Pub	Coffee Shop	Brewery	Italian Restaurant	Park
Riesbach	Zurich	4	Italian Restaurant	Swiss Restaurant	Café	Restaurant	Hotel
Anderson Bridge	Singapore	0	Hotel	Gym / Fitness Center	Waterfront	Event Space	Cocktail Bar

# Top 10 areas summary graphs

Graph of top 10 areas view by Tiers.



Graph of top 10 areas view by GVC.



## Final observations

- ❑ Arts & Culture venues (14) are the driving force behind **Southwark**'s 1st place. To complete the area Tier-1 composition: Parks & Outdoors venues (7), Restaurants & Foods venues (6). Good cuisine diversification with regards to Restaurants.
- ❑ In addition to Arts & Culture venues (10) and Parks & Outdoors venues (12) (both mostly due to proximity to Les Invalides complex), the actual Restaurants (45, under Tier-2) seem to be one of the reasons for **Invalides** 2nd place (see GVC graph). Yet the lack of cuisine diversification must be highlighted (78% of restaurants are French).
- ❑ Parks & Outdoors concentration helps **East Harlem** area to the 3rd place. There are 14 venues under this GVC; most of them due to proximity to one large park: Central Park. To complete the area Tier-1 composition: Arts/Culture venues (4). Good cuisine diversification with regards to Restaurants.
- ❑ As per Tier graph, highest count of aggregate (Tier-1+Tier-2) venues is reached by **Place Monge** in Paris (4th place) and by **West Bermondsey** in London (8th place).
- ❑ As per GVC graph, Parks & Outdoors are the most frequent in **Buena Vista** in San Francisco (20), thanks to combination of Fort Mason / Great Meadow park and coastal location; while Bars et simil are the most frequent in **West Bermondsey** in London (30).
- ❑ **Multiple parameters** other than available venues must be considered before property purchase (among which stamp duty tax). See for reference the Numbeo table of city indices in slide 2. Key index would obviously be the city's average property **Price per sqm**; other parameters too would be relevant, such as **Quality of Life, Cost of Living, Weather, Healthcare**, etc.

### A few suggestions for further developments

- set a higher max limit on Foursquare's API queries (e.g. 200+);
- add more cities added to the search;
- include more areas per city, and/or set smaller area radius (walking distance);
- account/adjust for Foursquare's venue data limitations (see end of par. 4);
- define more venue Tiers (5? 10?), to enhance customization;
- include negative weights for undesired venues if any ("negative Tiering");
- add cuisine diversification index to scoring;
- add Numbeo's indicators to scoring;
- experiment non-linear scoring functions (e.g. exponential), and/or alternative scoring point systems.