Micro-module A: Online Urban Data Gathering

A4- Social Media Data Gathering and Processing

The rapid advances in information and communication technology (ICT) are rapidly integrating with the built environment, resulting in the emergence of new urban science manifesting as a new infrastructure of sensing, data gathering, and urbanism analysis. Point of Interests (POIs) is a significant type of social media data, it provides finer-grained information of urban land use and vitality.

In this micromodule, you will learn how to obtain, filter and input POIs into QGIS based on Google Place API.

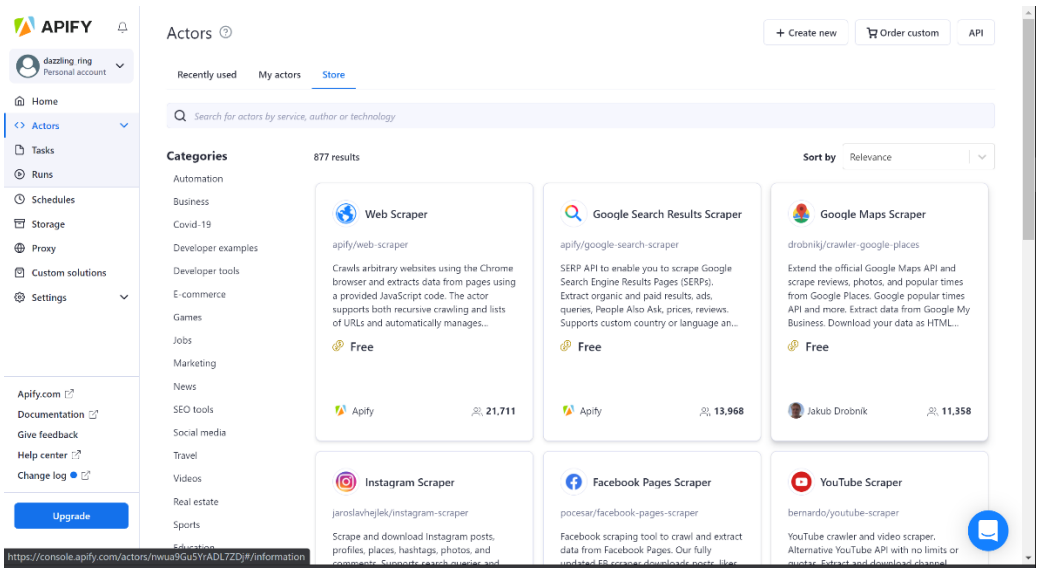
In the first part, this tutorial shows how to employ APIFY to automatically obtained selected functions of POIs in the selected research area, the output .csv file includes geo-location of POIs, review scores distribution, popular time and addresses.

In the second part, this tutorial introduces how to input .csv and filter POIs in the selected area through the 'clip' function.

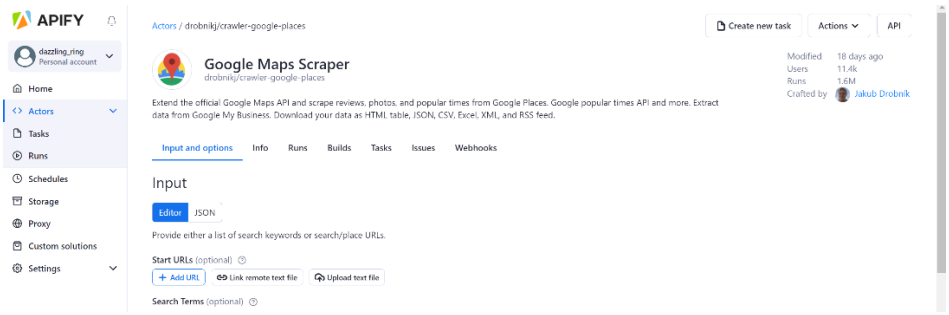
1. Obtaining the POIs

1.1 To Obtain POIs via APIFY

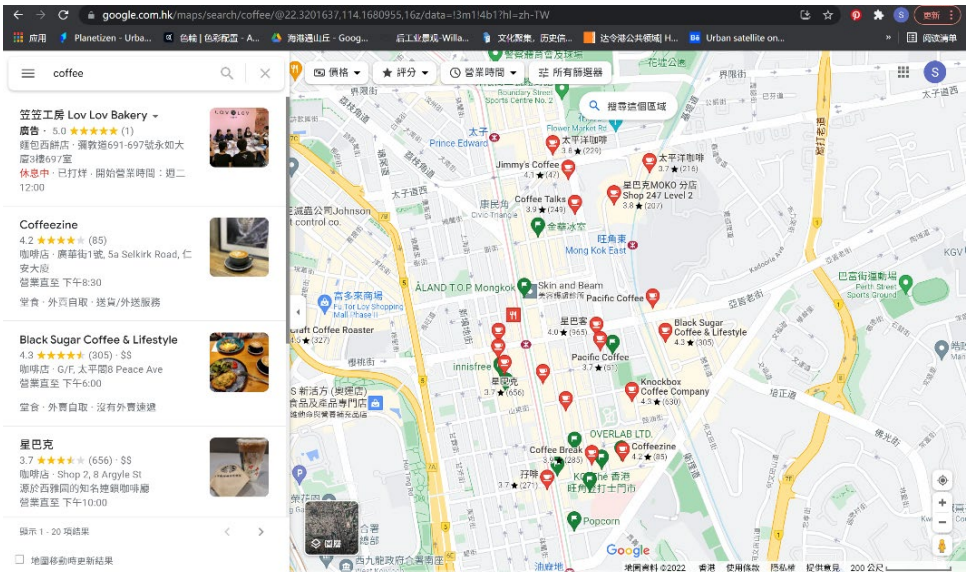
- Sign up to APIFY and find the actor called 'Google Map Scraper'. 'Google Map Scraper' extend the official Google Maps API and scrape reviews, photos, and popular times from Google Places. Google popular times API and more. Extract data from Google My Business. The output data can be downloaded as HTML table, JSON, CSV, Excel, XML, and RSS feed.



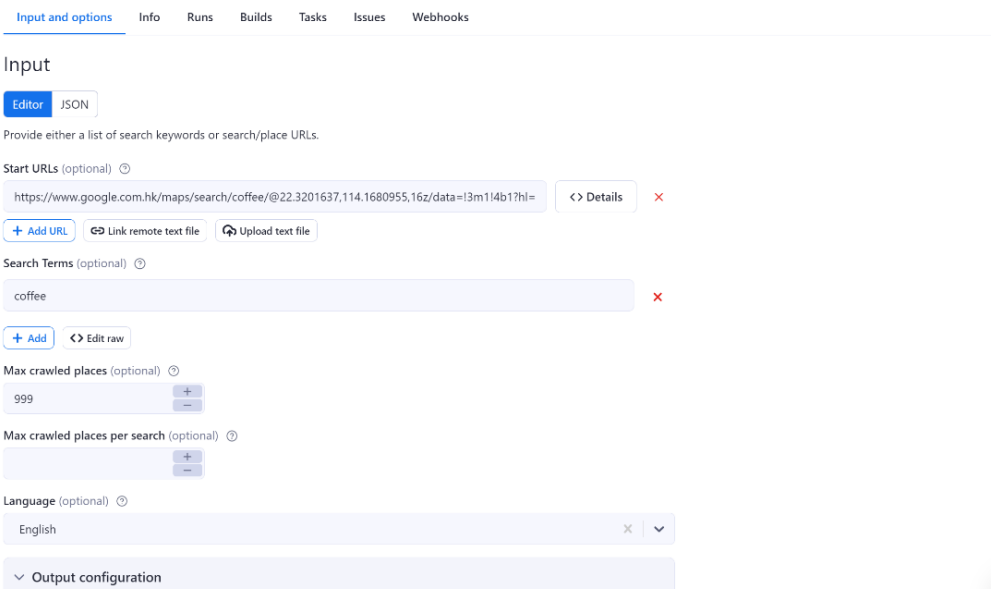
- By inputting the url of selected area and categories of POIs in Google map, the 'google map scraper' can automatically scrap POIs.



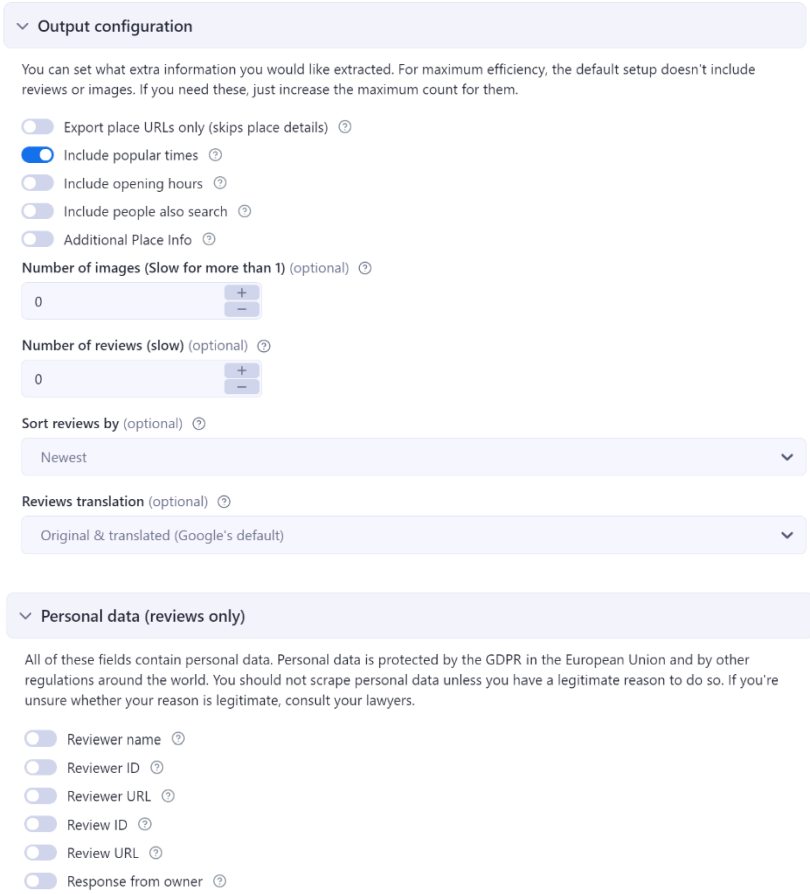
- You could zoom in to the specific area you want and search the category, for example, this tutorial search 'coffee' in Mong Kok area. You need to copy the url and paste in Google Map Scraper.



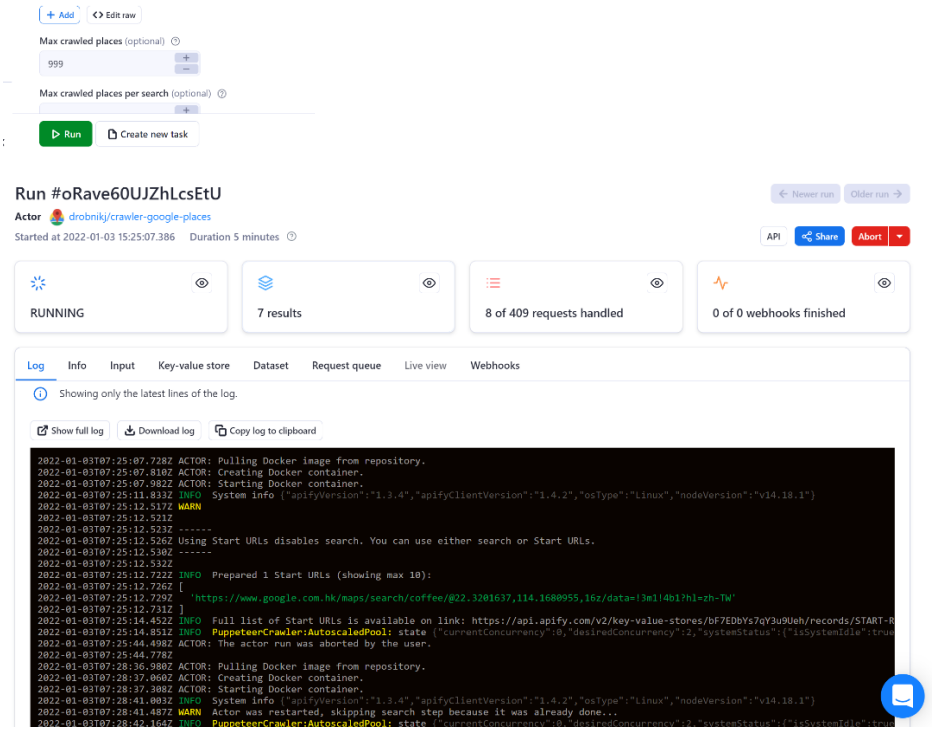
- The 'Search Terms' are array of strings to be searched. It is also possible to fill in Google Place IDs in the format place_id:ChIJp4JiUCNP0xQR1JaSjpW_Hms. Setting geolocation input fields is more accurate than including location in the search string. The 'Max crawled places' are the maximum number of places you scrape per whole run. If you want to scrape all available, set this to `9999999`.



- You could change the ‘Output configuration’ to include popular time, the dataset could be filtered to analysis amenity need pattern for a whole day. Moreover, you could change the ‘Personal data’ to decide if the output includes reviewers’ information.

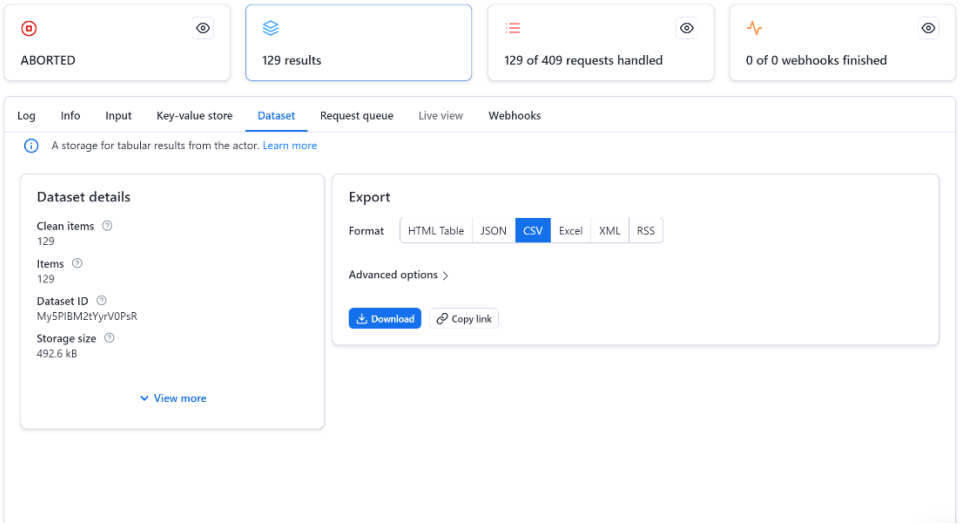


- Click ‘Run’, the scraper will start to crawl data. The dashboard shows the status of data-processing.

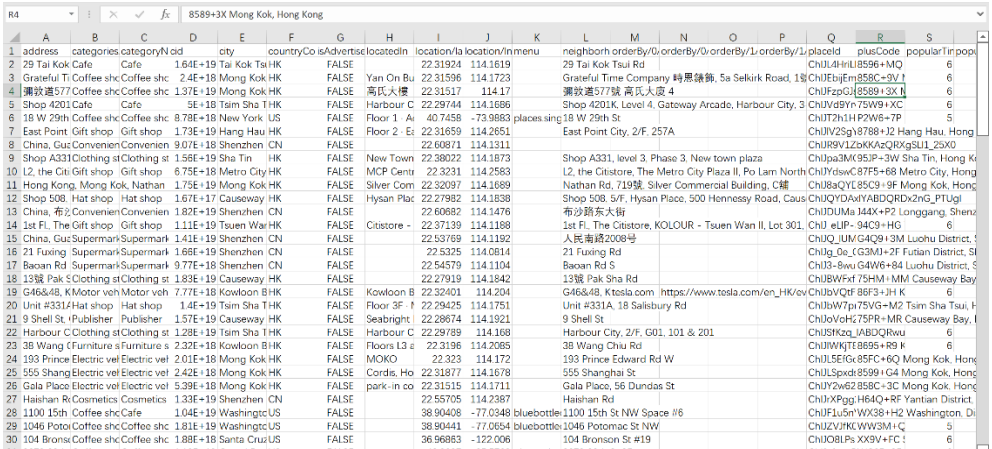


1.2 To Output data into .csv

- Once the data-processing is done, you could click ‘Dataset’ to choose the type of export data, in this case, we choose ‘CSV’ for it’s convenient to be input into QGIS in the following steps. Click ‘Download’.



- The output .csv data includes the address, categories, geo-location, popular time, reviews numbers, reviews score distribution.



address	categories	categoryN	cid	city	countryCo	isAdvertis	locatedIn	location/la	location/ln	menu	neighborh	orderBy/0	orderBy/1	orderBy/1	placeId	plusCode	popularTin	popu
29 Tai Kok	Cafe	Cafe	1.64E+19	Tai Kok Tsu	HK	FALSE		22.31924	114.1619		29 Tai Kok Tsui Rd				ChIJL4HriLJ	8596+MQ	6	
Grateful Ti	Coffee shc	Coffee shc	2.4E+18	Mong Kok	HK	FALSE	Yan On Bu	22.31596	114.1723		Grateful Time Company 時果鋪飾, 5a Selkirk Road, 1				ChIJEbijEm	858C+9V	6	
彌敦道577	Coffee shc	Coffee shc	1.37E+19	Mong Kok	HK	FALSE	高氏大樓	22.31517	114.17		彌敦道577號 高氏大廈 4				ChIJFzpGJ	8589+3X	6	
Shop 4201	Cafe	Cafe	5E+18	Tsim Sha T	HK	FALSE	Harbour C	22.29744	114.1686		Shop 4201K, Level 4, Gateway Arcade, Harbour City, 3				ChIJVd9Yn	75W9+XC	6	
18 W 29th	Coffee shc	Coffee shc	8.78E+18	New York	US	FALSE	Floor 1 · A	40.7458	-73.9883	places.sing	18 W 29th St				ChIJT2h1H	P2W6+7P	5	
East Point	Gift shop	Gift shop	1.73E+19	Hang Hau	HK	FALSE	Floor 2 · E	22.31659	114.2651		East Point City, 2/F, 257A				ChIJV2Sg	8788+J2	Hang Hau, Hong	
China, Guz	Convenien	Convenien	9.07E+18	Shenzhen	CN	FALSE		22.60871	114.1311						ChIJR9V1Z	bKKAzQRXgSlJ1_25X0		
Shop A331	Clothing st	Clothing st	1.56E+19	Sha Tin	HK	FALSE	New Town	22.38022	114.1873		Shop A331, level 3, Phase 3, New town plaza				ChIJpa3M	C95JP+3W	Sha Tin, Hong K	
L2, the Citi	Gift shop	Gift shop	6.75E+18	Metro City	HK	FALSE	MCP Centr	22.3231	114.2583		L2, the Citistore, The Metro City Plaza II, Po Lam North				ChIJYdswC	87F5+68	Metro City, Hong	
Hong Kong,	Mong Kok,	Nathan	1.75E+19	Mong Kok	HK	FALSE	Silver Com	22.32097	114.1689		Nathan Rd, 719號 Silver Commercial Building, C樓				ChIJ8aQYL	85C9+9F	Mong Kok, Hong	
Shop 508.	Hat shop	Hat shop	1.67E+17	Causeway	HK	FALSE	Hysan Plac	22.27982	114.1838		Shop 508. 5/F, Hysan Place, 500 Hennessy Road, Caus				ChIJQYDAx	IYABDQRDx2nG_PTUgI		
China, 布	Convenien	Convenien	1.82E+19	Shenzhen	CN	FALSE		22.60682	114.1476		布沙路东大街				ChIJDUMa	J44X+P2	Longgang, Shenz	
1st Fl, The	Gift shop	Gift shop	1.11E+19	Tsuen War	HK	FALSE	Citistore -	22.37139	114.1188		1st Fl., The Citistore, KOLOUR - Tsuen Wan II, Lot 301,				ChIJ eLIP-	94C9+HG	6	
China, Guz	Supermark	Supermark	1.41E+19	Shenzhen	CN	FALSE		22.53769	114.1192		人民南路2008号				ChIJQ_IUM	G4Q9+3M	Luohu District,	
21 Fuxing	Supermark	Supermark	1.68E+19	Shenzhen	CN	FALSE		22.5325	114.0814		21 Fuxing Rd				ChIJg_0e_	CG3MJ+2F	Futian District, S	
Baoan Rd	Supermark	Supermark	9.77E+18	Shenzhen	CN	FALSE		22.54579	114.1104		Baoan Rd S				ChIJ3-8wu	G4W6+84	Luohu District, S	
13號 Pak S	Clothing st	Clothing st	1.83E+19	Causeway	HK	FALSE		22.27919	114.1842		13號 Pak Sha Rd				ChIJBWFxf	75HM+MM	Causeway Bay	
G46&48, K	Motor veh	Motor veh	7.77E+18	Kowloon B	HK	FALSE	Kowloon B	22.32401	114.204		G46&48, Ktosla.com	https://www.tesla.com/en_HK/ov			ChIJbVQtF	86F3+JH K	6	
Unit #331A	Hat shop	Hat shop	1.4E+19	Tsim Sha T	HK	FALSE	Floor 3F ·	22.29425	114.1751		Unit #331A, 18 Salisbury Rd				ChIJbW7pi	75VG+M2	Tsim Sha Tsui, H	
9 Shell St,	Publisher	Publisher	1.57E+19	Causeway	HK	FALSE	Seabright	22.28674	114.1921		9 Shell St				ChIJoVcH	Z75PR+MR	Causeway Bay, I	
Harbour C	Clothing st	Clothing st	1.28E+19	Tsim Sha T	HK	FALSE	Harbour C	22.29789	114.168		Harbour City, 2/F, G01, 101 & 201				ChIJSfKzq	IABQQRwu	6	
38 Wang C	Furniture s	Furniture s	2.22E+18	Kowloon B	HK	FALSE	Floors 13 a	22.3196	114.2085		38 Wang Chiu Rd				ChIJWKjT	8695+R9 K	6	
193 Prince	Electric vel	Electric vel	2.01E+18	Mong Kok	HK	FALSE	MOKO	22.323	114.172		193 Prince Edward Rd W				ChIJL5EfG	85FC+6Q	Mong Kok, Hong	
555 Shang	Electric vel	Electric vel	2.42E+18	Mong Kok	HK	FALSE	Cordis, Ho	22.31877	114.1678		555 Shanghai St				ChIJLSpxd	8599+G4	Mong Kok, Hong	
Gala Place	Electric vel	Electric vel	5.39E+18	Mong Kok	HK	FALSE	park-in co	22.31515	114.1711		Gala Place, 56 Dundas St				ChIJY2w62	85BC+3C	Mong Kok, Hong	
Haishan R	Cosmetics	Cosmetics	1.33E+19	Shenzhen	CN	FALSE		22.55705	114.2387		Haishan Rd				ChIJrXPgg	H64Q+RF	Yantian District,	
1100 15th	Coffee shc	Cafe	1.04E+19	Washingtc	US	FALSE		38.90408	-77.0348	bluebottle	1100 15th St NW Space #6				ChIJF1u5n	WX38+H2	Washington, Di	
1046 Potor	Coffee shc	Coffee shc	1.81E+19	Washingtc	US	FALSE		38.90441	-77.0654	bluebottle	1046 Potomac St NW				ChIJZVJfK	WW3M+C	5	
104 Bronsc	Coffee shc	Coffee shc	1.88E+18	Santa Cruz	US	FALSE		36.96863	-122.006		104 Bronson St #19				ChIJO8LPs	XX9V+FC	6	
3679 29th	Coffee roa	Coffee roa	1.19F+19	Grand Rac	US	FALSE		42.9097	-85.5763	places.sing	3679 29th St SE				ChIJvfwz_	FWC6F+2F	6	

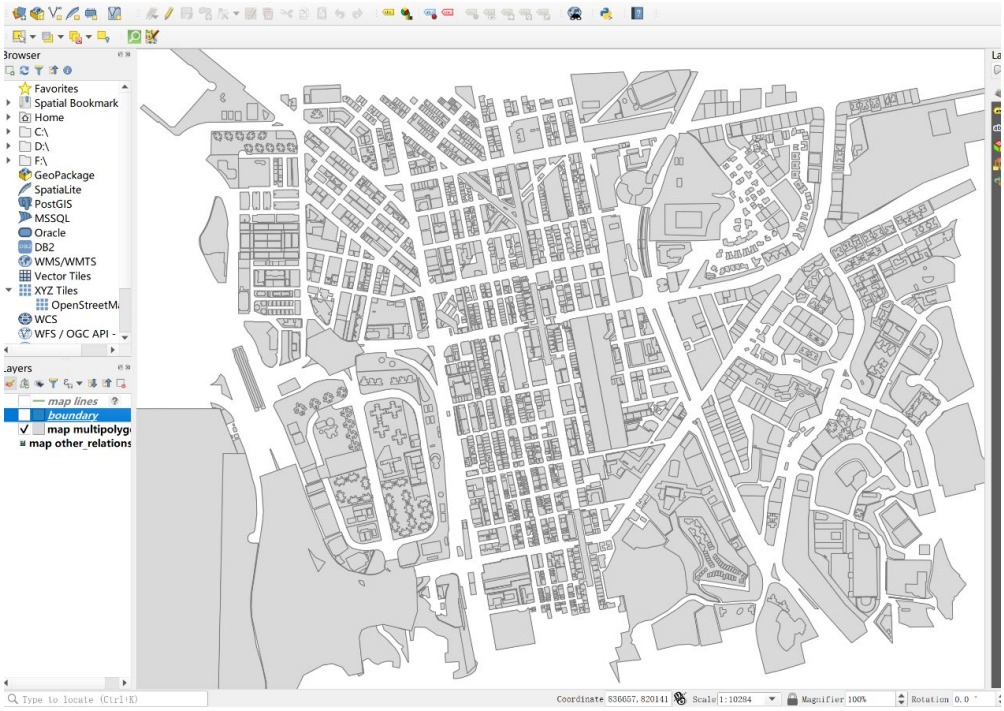
dataset_crawler-google-places_2

popularTin	popularTin	popularTin	postalCod	price	rank	reviewsCount	reviewsDistribution /fiveStar	reviewsDistribution /fourStar	reviewsDistribution/ oneStar	reviewsDistribution/ threeStar	reviewsDistribution /twoStar
	41	Less busy than usual		$$	3	327	210	87	3	17	10
		Now: Usually not too busy			2	85	40	29	2	11	3
	1	Not busy		$$	20	271	67	104	18	63	19
		Now: Usually a little b		$$	80	165	39	57	13	42	14
			10001	$$	417	1354	920	323	17	66	28
					360	4	1	1	2	0	0
					359	0	0	0	0	0	0
					358	4	1	1	0	2	0
					357	2	0	0	1	1	0
					320	0	0	0	0	0	0
					356	1	1	0	0	0	0
					355	0	0	0	0	0	0
		Now: Usually a little busy			354	2	0	0	0	1	1
					353	23	6	9	1	5	2
					352	14	5	5	1	3	0
			518010		351	31	14	7	0	9	1
					350	6	3	2	0	1	0
	100	Busier than usual			349	92	46	24	10	8	4
					348	0	0	0	0	0	0
					347	0	0	0	0	0	0
		Now: Usually a little b		$$$	346	1	0	1	0	0	0
	32	Not too busy			319	4272	1738	1726	87	617	104
					318	0	0	0	0	0	0
					317	0	0	0	0	0	0
					316	0	0	0	0	0	0
					380	0	0	0	0	0	0
			20005	$$	416	61	42	15	0	3	1
			20007	$$	415	755	505	163	21	51	15
			95062	$$	414	375	305	52	3	11	4

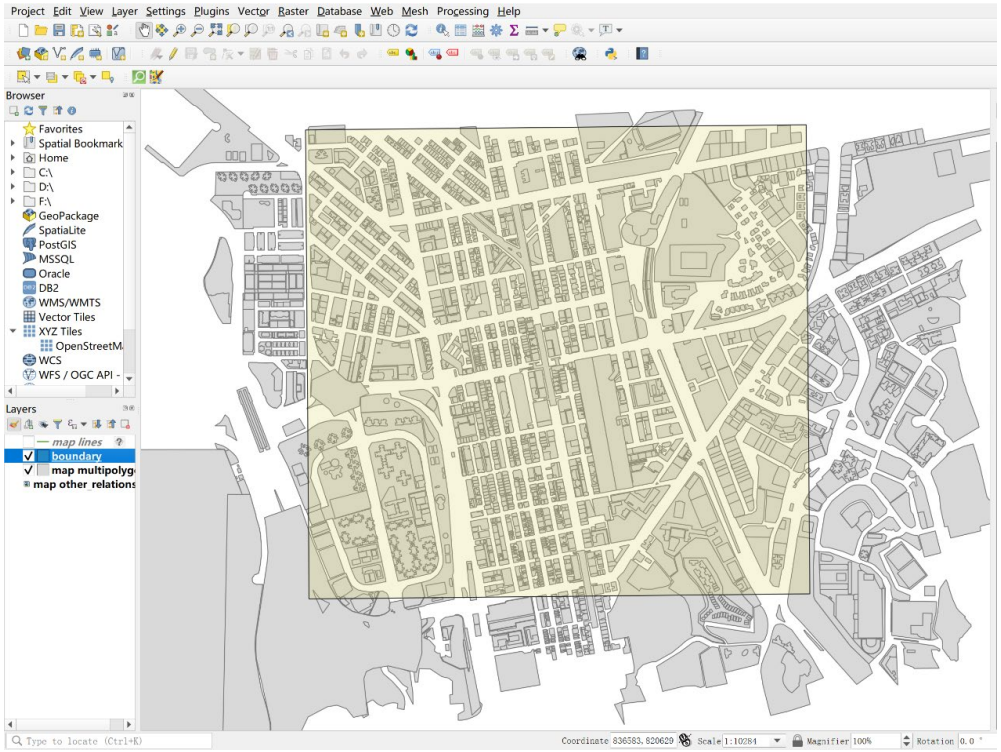
2. Processing the POIs in QGIS

2.1 To input the base map and boundary

- Input the base map through Open Street Map, following the tutorial in A3.



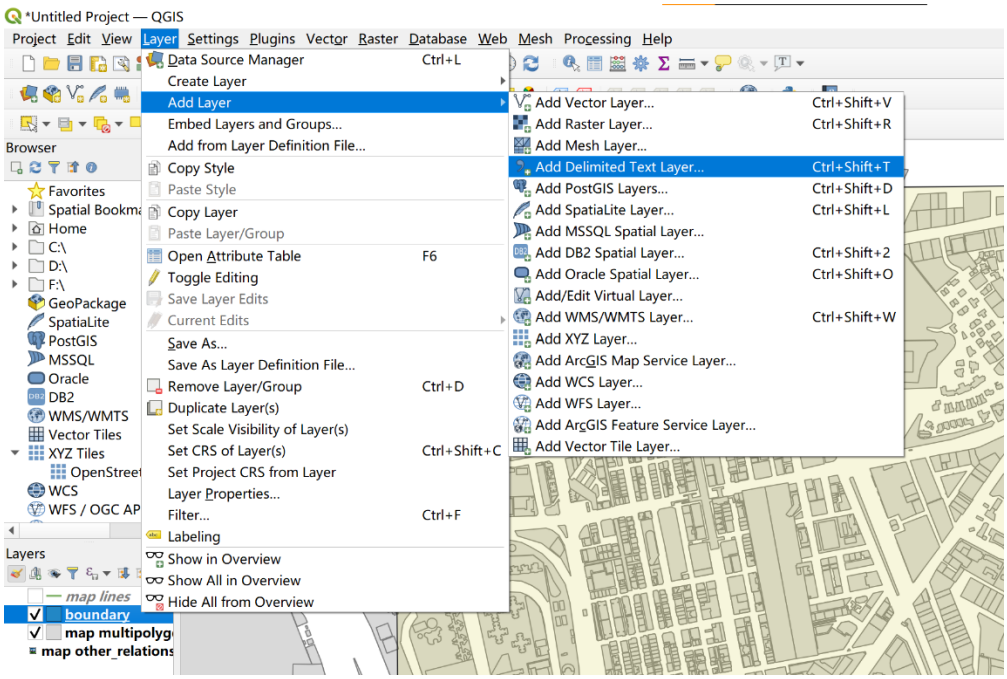
- Add a shapefile layer, draw a polygon, renamed as boundary.



2.2 To input the .csv into QGIS

Once the csv files are formatted properly, you can add them into QGIS.

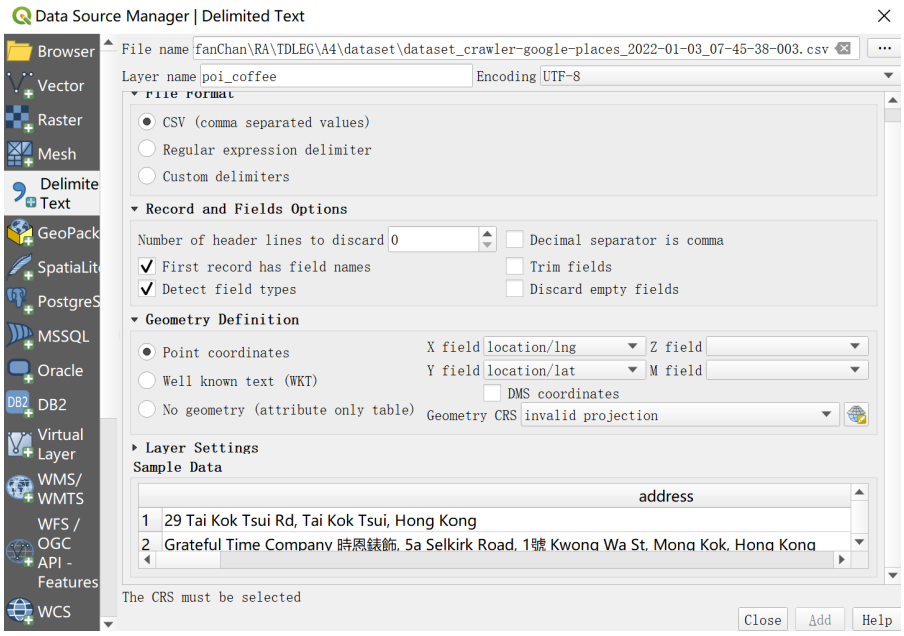
- Click on the "Layer" menu, mouse-over "Add Layer" and click on "Add Delimited Text Layer..." or click on the "Add Delimited Text Layer" icon in the left column of QGIS.



- The next GUI will have many different options you may need to change depending on the specific data set you have. Here is an outline of the most common fields needed to be changed.
- Browse - click on Browse and find the folder where the csv file is saved and open the file.

Layer name - the name of the CSV will show up here.

- File format - depending on the version of QGIS you are using, you may need to verify the file format.

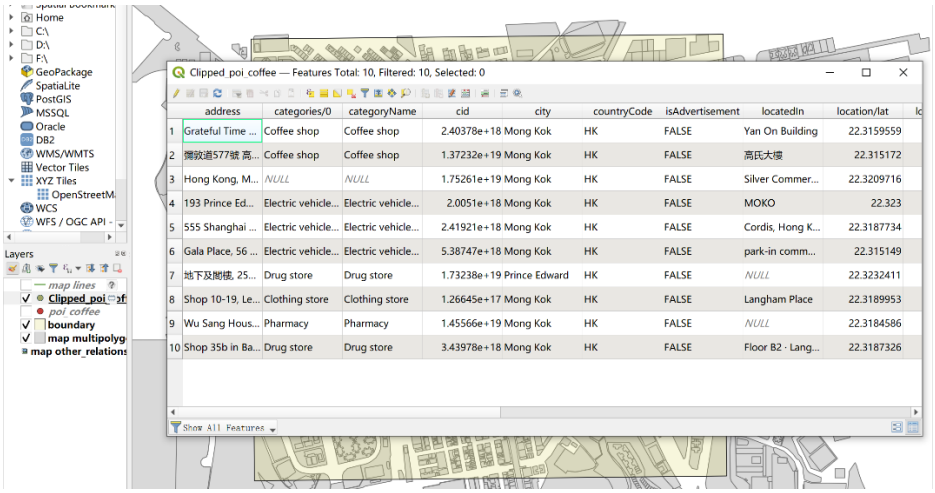


- Geometry definition - If you have x, y coordinates you will choose the "Point coordinates" option. Verify the X Field is pointing to your Longitude field and the Y Field is pointing to your Latitude Field. If you have a table with no x,y coordinates you will choose the "No Geometry" option.

- Layer settings - you will see a preview of the table. Verify that everything looks correct. Click OK.
- If you have a field with no x, y coordinates you are done importing the csv file. However, depending on the version of QGIS you are using, you may be prompted to define the coordinate reference system (CRS) of your x, y coordinates. Longitude/Latitude coordinates are unprojected, and you should choose the CRS of WGS 84 (EPSG:4326). If you have coordinates using something else, like meters in a UTM zone, search for that using the filter box in the CRS Selector Dialog. For instance, type in <utm 17> to get a short list that includes the UTM zone for Durham, 17N (e.g., NAD83 / UTM zone 17N, or EPSG:26917).

2.3 Clip/ Filter the POIs

After inputting the csv file named ‘poi_coffee’, you could clip the .shp with the boundary to filter POIs in the selected area.



After clipping, you could find the new .shp layer, open the attribute table, you could choose and filter the POIs as values you want.