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.

| 🚺 APIFY 🛛 📮 | Actors ⑦ | | | + Create new P Order custom API |
|--|---|--|--|---|
| O dazzling ring Personal account | Recently used My acto | rs Store | | |
| M Home ↔ Actors ✓ | Q Search for actors by serv | ice, author or technology | | |
| Tasks Runs | Categories | 877 results | | Sort by Relevance V |
| Schedules Storage | Business Covid-19 | Keb Scraper | Q Google Search Results Scraper | Google Maps Scraper |
| Proxy | Developer examples | apify/web-scraper | apify/google-search-scraper | drobnikj/crawler-google-places |
| Custom solutions | Developer tools | Crawls arbitrary websites using the Chrome browser and extracts data from pages using | SERP API to enable you to scrape Google Search Engine Results Pages (SERPs). | Extend the official Google Maps API and scrape reviews, photos, and popular times |
| ⊕ Settings ✓ | E-commerce Games | a provided JavaScript code. The actor supports both recursive crawling and lists of URLs and automatically manages | Extract organic and paid results, ads, queries, People Also Ask, prices, reviews. Supports custom country or language an | from Google Places. Google popular times API and more. Extract data from Google My Business. Download your data as HTML |
| | Jobs Marketing | Ø Free | Ø Free | 🦑 Free |
| Apify.com 📝 | News | | | |
| Documentation 🖄 | SEO tools | M Apify 🔍 21,711 | 🚺 Apify 🖉 13,968 | 📳 Jakub Drobník 🔍 11,358 |
| Give feedback | Social media | | | |
| Help center 🛛 | Travel | | | |
| Change log 🗢 🖄 | Videos | Instagram Scraper | Facebook Pages Scraper | YouTube Scraper |
| Upgrade | Real estate Sports | jaroslavhejlek/instagram-scraper | pocesar/facebook-pages-scraper | bernardo/youtube-scraper |
| https://console.apify.com/actor | Education s/nwua9Gu5YrADL7ZDj#/informati | Scrape and download instagram posts, profiles, places, hashtags, photos, and comments. Supports carch queries and | Facebook scraping tool to crawl and extract data from Facebook Pages. Our fully updated F8 scraper downloads posts liker | YouTube crawler and video scraper. Alternative YouTube API with no limits or guotar. Extract and download channel |

 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.

| ← → C 🚔 google.com.hk/map | s/search/coffee/@ | 22.3201637,114.1680955,16z/data=!3m1!4b1?hl=zh-TW | (날 ☆ 🧕 🗯 🕄 📼 :) |
|--|-------------------|---|--|
| 🏥 应用 🕴 Planetizen - Urba 🔍 色輪 | 色彩配置 - A 🛛 🕌 | 海港通山丘 - Goog 后工业景境-Willa 🍞 文化聚集,历史倍 📕 达令港公共领域 H 🚥 Urban satell | ite on × 🗐 网致清单 |
| ≡ coffee | Q X | ♥ 図 價格 ▼ ★ 詳分 ▼ ② 営業時間 □ 友祥無 明現時 ▼ ★ 詳分 ▼ ③ 営業時間 ▼ 詳 所有勝選議 ■ 中花並公告 - 昇間的 | · · · · · · · · · · · · · · · · · · · |
| 笠笠工房 Lov Lov Bakery → 廣告 - 5.0 ★★★★ (1) 嫁包西無馬 - 强教施691-697破永如大 魔魂穆697章 休念中 - 已打烊 - 開始營業時間:近二 12:00 | COV DIEV | 日本市 | |
| Coffeezine 4.2 ★★★★★ (85) 咖啡店。 廣華街 號 Sa Selkirk Road, 仁 安大慶 營業直至 下午8:30 堂食・外酉自取・送貨/外送販務 | | 正義高公元Johnson Weiling Autority Autori | And Same Same Same Same Same Same Same Same |
| Black Sugar Coffee & Lifestyle 4.3 ★★★★ (305) - \$\$ 咖啡店 · G/F, 太平岡8 Peace Ave 營業直至 下午6:00 堂合 · 外賣自取 · 沒有外賣速速 | | series Control Reparator (中) (22) 御田市 S 新活方 (時間の) 和品は書が目れる 本のので最もない。 S 新活方 (時間の) 和品は、新活方 (時間の) ののののののののののののののののののののののののののののののののののの | An and a second se |
| 星巴克 3.7 ★★★★★ (656) · \$S 咖啡店 · Shop 2, 8 Argyle St 源於西雅岡的知名煙銀咖啡廳 營業直至 下午10:00 | | Coffee Great → Coffe | |
| 顯示 1 - 20 項結果 □ 地圖移動時更新結果 | < > | | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 |

 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 `99999999`.

| Input and options | Info Ru | ns Builds | Tasks | Issues | Webhooks | | |
|------------------------------------|----------------------|------------------|------------|------------|--------------------------|------------|---|
| Input | | | | | | | |
| Editor JSON | | | | | | | |
| Provide either a list of s | earch keyword | s or search/plac | e URLs. | | | | |
| Start URLs (optional) |) | | | | | | |
| https://www.google.c | om.hk/maps/s | earch/coffee/@ | 22.3201637 | ,114.16809 | 55,16z/data=!3m1!4b1?hl= | <> Details | × |
| + Add URL GD Link | remote text file | ርጉ Upload t | ext file | | | | |
| Search Terms (optional) | 0 | | | | | | |
| coffee | | | | | | | × |
| + Add | | | | | | | |
| Max crawled places (op | tional) ③ | | | | | | |
| 999 | + - | | | | | | |
| Max crawled places per | search (optio | nal) () | | | | | |
| | + - | | | | | | |
| Language (optional) @ |) | | | | | | |
| English | | | | | | × | ~ |
| | | | | | | | |
| Output configu | ration | | | | | | |

- 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.

| Output configu | uration | |
|--|--|-------|
| You can set what ext reviews or images. If | tra information you would like extracted. For maximum efficiency, the default setup doesn't includ If you need these, just increase the maximum count for them. | e |
| Export place U | URLs only (skips place details) 📀 | |
| Include popul | ılar times ③ | |
| Include openi | ning hours ③ | |
| Include peopl | ole also search ③ | |
| Additional Pla | ace Info ③ | |
| Number of images (| (Slow for more than 1) (optional) ③ | |
| 0 | + | |
| Number of reviews | (slow) (optional) ③ | |
| 0 | + | |
| Sort reviews by (opt | ational) ③ | |
| Newest | | ~ |
| Reviews translation | (optional) ③ | |
| Original & transla | ated (Google's default) | ~ |
| | | |
| Personal data (| (reviews only) | |
| All of these fields co regulations around unsure whether you | ontain personal data. Personal data is protected by the GDPR in the European Union and by other the world. You should not scrape personal data unless you have a legitimate reason to do so. If y ur reason is legitimate, consult your lawyers. | ou're |
| Reviewer nam | me 💿 | |
| Reviewer ID | 0 | |
| Reviewer URI | RL 💿 | |
| C Review ID @ | 0 | |
| Review URL | 0 | |
| Response fro | om owner ③ | |

- Click 'Run', the scraper will start to crawl data. The dashboard shows the status of dataprocessing.

| 0 | | | | | | | |
|--|---------------------------------------|--|-------------------|--|------------|-----------------------|----------------|
| + Add C Edit raw | | | | | | | |
| Max crawled places (optic | onal) () | | | | | | |
| 999 | + | | | | | | |
| Max crawled places per s | earch (optional) ③ | | | | | | |
| | + | | | | | | |
| D Run Create | new task | | | | | | |
| n #oRave60U | 17hl csFtU | | | | | ← New | |
| r 🐣 drobniki/crawler | | | | | | | |
| ed at 2022-01-03 15:25: | 07.386 Duration 5 | minutes ③ | | | | API | Share Abort |
| | | | | | | | |
| \$ (* | ۲ | 8 | ۲ | = | ۲ | ~ | (|
| UNNING | | 7 results | | 8 of 409 requests handled | | 0 of 0 webhooks fi | inished |
| showing only the la | nest lines of the log. | | | | | | |
| Show full log | ownload log | py log to clipboard | | | | | |
| 2022-01-03T07:25:07 2022-01-03T07:25:07 | .728Z ACTOR: Pull 810Z ACTOR: Crea | ing Docker image from | | | | | |
| 2022-01-03T07:25:07 2022-01-03T07:25:11 | .982Z ACTOR: Star .833Z INFO Syste | ting Docker container m info {"apifyVersion | | | | | |
| 2022-01-03T07:25:12 2022-01-03T07:25:12 | .517Z WARN .521Z | | | | | | |
| 2022-01-03T07:25:12 2022-01-03T07:25:12 | .523Z .526Z Using Start | URLs disables search | . You can use eit | ner search or Start URLs. | | | |
| 2022-01-03T07:25:12 2022-01-03T07:25:12 | .530Z | | | | | | |
| 2022-01-03T07:25:12 2022-01-03T07:25:12 | .722Z INFO Prepa .726Z [| red 1 Start URLs (sho | wing max 10): | | | | |
| 2022-01-03T07:25:12 2022-01-03T07:25:12 | .729Z 'https:// | | | | | | |
| 2022-01-03T07:25:14 2022-01-03T07:25:14 | 452Z INFO Full | list of Start URLs is | available on lin | <pre>c: https://api.apify.com/v2/key prentConcurrency":0."desiredCon</pre> | value-stor | es/bF7EDbYs7qY3u9Ueh/ | records/START- |
| 2022-01-03T07:25:44 | .498Z ACTOR: The | actor run was aborted | by the user. | | | | |
| 2022-01-03T07:28:36 2022-01-03T07:28:36 | .980Z ACTOR: Pull | ing Docker image from | | | | | |
| 2022-01-03107:28:37 | .308Z ACTOR: Star | ting Docker container | | | | | |
| 2022-01-03107:28:41 2022-01-03T07:28:41 | .487Z WARN Actor | was restarted, skipp | ing search step b | cause it was already done | | | |

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'.

| O ABORTED | 129 results | i≡ 129 of 409 requests handled 0 of 0 webhooks finished | | | | | |
|--|---|--|--|--|--|--|--|
| Log Info Input Key-value store | Dataset Request queue Live view | Webhooks | | | | | |
| Dataset details Clean items © 129 Items © 129 Dataset ID © MySPIBM2XYyrV0PsR Storage size © 492.6 kB | Export Format HTML Table Advanced options > | JSON CSV Excel XML RSS | | | | | |
| ✓ View more | | | | | | | |

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

| R | 4 | * E > | < fx | 858 | 9+3X N | long Kol | , Hong Ko | ng | | | | | | | | | | | | | Ý | , |
|----|------------------------|-------------|----------------|--------|---------|----------|-----------|-----------|--------------|-------------|-------------|-------------|---------------|--------------|--------------|--------------|--------------|-----------|------------|------------|------------|---|
| 1 | A | В | С | D | | E | F | G | н | 1 | J | K | L | м | N | 0 | Р | Q | R | s | | - |
| 1 | address | categorie | s categoryN | cid | city | | countryCo | sAdvertis | elocatedIn | location/la | location/Ir | menu | neighborh | orderBy/(|)/orderBy/0 | //orderBy/1 | orderBy/1 | placeld | plusCode | populari | inpopu | |
| 2 | 29 Tai Kol | Cafe | Cafe | 1.64E+ | 19 Tai | Kok Tst | HK | FALSE | | 22.31924 | 114.1619 | | 29 Tai Kok | Tsui Rd | | | | ChUL4Hril | .18596+MQ | | 6 | |
| 3 | Grateful T | i Coffee sh | cCoffee shc | 2.4E+ | 18 Mo | ng Kok | HK | FALSE | Yan On Bu | 22.31596 | 114.1723 | | Grateful Tir | me Comp | any 時思議 | 飾, 5a Selki | rk Road, 1§ | ChUEbijEr | 1858C+9V | 1 | 6 | |
| 4 | 潮敦道573 | 7 Coffee sh | c Coffee shc | 1.37E+ | 19 Mo | ng Kok | HK | FALSE | 高氏大樓 | 22.31517 | 114.17 | | 彌敦道577 | 號 高氏大 | 廈 4 | | | ChUFzpG. | 8589+3X | N | 6 | |
| 5 | Shop 420. | 1 Cafe | Cafe | 5E+ | 18 Tsin | n Sha T | HK | FALSE | Harbour C | 22.29744 | 114.1686 | | Shop 4201 | K, Level 4. | Gateway A | vrcade, Hari | bour City, 3 | ChUVd9Y | h75W9+X0 | | 6 | |
| 6 | 18 W 29th | Coffee sh | cCoffee shc | 8.78E+ | 18 Nev | v York | JS | FALSE | Floor 1 - Ar | 40.7458 | -73.9883 | places.sing | 18 W 29th | St | | | | ChUT2h1 | 1P2W6+7F | | 5 | |
| 7 | East Point | Gift shop | Gift shop | 1.73E+ | 19 Har | ig Hau | HK | FALSE | Floor 2 - Ea | 22.31659 | 114.2651 | | East Point | City, 2/F, 2 | 257A | | | ChUIV2Sg | \8788+J2 H | lang Hau | Hong | _ |
| 8 | China, Gu | Convenie | n Convenien | 9.07E+ | 18 She | nzhen | CN | FALSE | | 22.60871 | 114.1311 | | | | | | | ChUR9V1 | ZbKKAzQR. | gSLI1_25 | X0 | |
| 9 | Shop A33 | 1Clothing s | st Clothing st | 1.56E+ | 19 Sha | Tin | HK | FALSE | New Town | 22.38022 | 114.1873 | | Shop A331 | , level 3, F | hase 3, Ne | w town pla: | za | ChUpa3M | (95JP+3W | Sha Tin, F | long Kr | |
| 10 | L2, the Cit | i Gift shop | Gift shop | 6.75E+ | 18 Met | tro City | HK | FALSE | MCP Centr | 22.3231 | 114.2583 | | L2, the Citis | store, The | Metro City | Plaza II, Po | Lam North | ChUYdsw | C87F5+68 | Metro City | , Hong | |
| 11 | Hong Kor | ig, Mong K | lok, Nathan | 1.75E+ | 19 Mo | ng Kok | HK | FALSE | Silver Com | 22.32097 | 114.1689 | | Nathan Rd | . 719號. S | Iver Comm | ercial Build | ing, C舖 | ChIJ8aQY | E85C9+9F | Mong Kol | , Hong | |
| 12 | Shop 508. | Hat shop | Hat shop | 1.67E+ | 17 Cau | seway | HK | FALSE | Hysan Plac | 22.27982 | 114.1838 | | Shop 508. | 5/F, Hysai | n Place, 500 | Hennessy | Road, Caus | ChUQYD/ | XIYABDQR | Dx2nG_P1 | Ugl | |
| 13 | China, 布; | Convenie | n Convenien | 1.82E+ | 19 She | nzhen | CN | FALSE | | 22.60682 | 114.1476 | | 布沙路东大 | 、街 | | | | ChUDUM: | a J44X+P2 | .onggang | Shenz | |
| 14 | 1st FL, The | Gift shop | Gift shop | 1.11E+ | 19 Tsu | en War | HK | FALSE | Citistore - | 22.37139 | 114.1188 | | 1st FL, The | Citistore, | KOLOUR - | Tsuen Wan | II, Lot 301, | ChU_eLIP- | 94C9+HG | | 6 | |
| 15 | China, Gu | Supermar | rkSupermark | 1.41E+ | 19 She | nzhen | CN | FALSE | | 22.53769 | 114.1192 | | 人民南路2 | 008号 | | | | ChUQ_IUI | /G4Q9+3N | 1 Luohu D | istrict, 1 | |
| 16 | 21 Fuxing | Superman | rk Supermark | 1.66E+ | 19 She | nzhen | CN | FALSE | | 22.5325 | 114.0814 | | 21 Fuxing I | Rd | | | | ChUg_0e_ | (G3MJ+2F | Futian Dis | strict, SI | |
| 17 | ' Baoan Rd | Superman | kSupermark | 9.77E+ | 18 She | nzhen | CN | FALSE | | 22.54579 | 114.1104 | | Baoan Rd S | S | | | | ChU3-8w | 1G4W6+84 | Luohu D | istrict, S | |
| 18 | 13% Pak | Clothing a | stClothing st | 1.83E+ | 19 Cau | seway | нк | FALSE | | 22.27919 | 114.1842 | | 13號 Pak S | ha Rd | | | | ChUBWFx | f75HM+M | M Causev | ray Bay | |
| 19 | G46&48, I | Motor vei | h Motor veh | 7.77E+ | 18 Kov | vloon B | HK | FALSE | Kowloon B | 22.32401 | 114.204 | | G46&48, K | tesla.com | https://wv | ww.tesla.cor | n/en_HK/ev | ChUbVQt | F86F3+JH | < | 6 | |
| 20 |) Unit #331. | Hat shop | Hat shop | 1.4E+ | 19 Tsin | n Sha T | HK | FALSE | Floor 3F · N | 22.29425 | 114.1751 | | Unit #331A | , 18 Salist | oury Rd | | | ChUbW7p | 175VG+M2 | Tsim Sha | Tsui, F | |
| 21 | 9 Shell St, | Publisher | Publisher | 1.57E+ | 19 Cau | seway | HK | FALSE | Seabright | 22.28674 | 114.1921 | | 9 Shell St | | | | | ChlJoVoH | 275PR+MR | Causewa | y Bay, (| |
| 22 | Harbour G | Clothing s | stClothing st | 1.28E+ | 19 Tsin | n Sha T | HK | FALSE | Harbour C | 22.29789 | 114.168 | | Harbour C | ity, 2/F, GI | 01, 101 & 20 | 01 | | ChUSfKzg | _IABDQRw | | 6 | |
| 23 | 38 Wang | Furniture | s Furniture s | 2.32E+ | 18 Kov | vloon B | HK | FALSE | Floors L3 a | 22.3196 | 114.2085 | | 38 Wang C | hiu Rd | | | | ChUIWKjT | E8695+R9 | ĸ | 6 | |
| 24 | 193 Prince | Electric ve | HElectric veh | 2.01E+ | 18 Mo | ng Kok | HK | FALSE | MOKO | 22.323 | 114.172 | | 193 Prince | Edward R | W b | | | ChUL5EfG | c85FC+6Q | Mong Ko | k, Hong | |
| 25 | 555 Shang | Electric ve | Electric veł | 2.42E+ | 18 Mo | ng Kok | нк | FALSE | Cordis, Ho | 22.31877 | 114.1678 | | 555 Shang | hai St | | | | ChULSpxc | 1:8599+G4 | Mong Kol | k, Hong | |
| 26 | Gala Place | Electric vo | Electric veł | 5.39E+ | 18 Mo | ng Kok | HK | FALSE | park-in co | 22.31515 | 114.1711 | | Gala Place, | 56 Dund | as St | | | ChUY2w6 | 2858C+3C | Mong Ko | k, Hong | |
| 27 | ¹ Haishan R | Cosmetics | s Cosmetics | 1.33E+ | 19 She | nzhen | CN | FALSE | | 22.55705 | 114.2387 | | Haishan Ro | ł | | | | ChUrXPgg | :H64Q+RF | Yantian [| District, | |
| 28 | 1100 15th | Coffee sh | c Cafe | 1.04E+ | 19 Wa | shingto | JS | FALSE | | 38.90408 | -77.0348 | bluebottle | (1100-15th) | St NW Sp | ace #6 | | | ChUF1u5r | °WX38+H | Washing? | ton, Di | |
| 29 | 1046 Poto | Coffee sh | cCoffee shc | 1.81E+ | 19 Wa | shingto | JS | FALSE | | 38.90441 | -77.0654 | bluebottle | 1046 Potor | mac St NV | v | | | ChUZVJfK | CWW3M+0 | 2 | 5 | |
| 30 | 104 Brons | Coffee sh | c Coffee shc | 1.88E+ | 18 San | ta Cruz | JS | FALSE | | 36.96863 | -122.006 | | 104 Bronso | on St #19 | | | | ChUO8LP | xX9V+FC | 1 | 6 | |
| 31 | 3679 29th | Coffee ro | a Coffee roa | 1 19F+ | 19 Gra | nd Rac | JS | FALSE | | 42,9097 | -85 5763 | olaces sinc | :3679 29th | St.SF | | | | ChUvfwz | EWC6E+2E | | 6 | * |

| JW | JX | JY | JZ | KA | KB | KC | KD | KE | KF | KG | KH |
|------------|-------------|-------------|-----------------|--------|------|--------------|---------------------|---------------------|----------------------|----------------------|---------------------|
| popularTir | npopularTir | popularTir | postalCod | price | rank | reviewsCount | reviewsDistribution | reviewsDistribution | reviewsDistribution/ | reviewsDistribution/ | reviewsDistribution |
| | 41 | Loss husy | than usual | \$\$ | 3 | 327 | 210 | 210015tai | 3 | 17 | 10 |
| | | Now: Usua | ally not too | busy | 2 | 85 | 40 | 29 | 2 | 11 | 10 |
| 1 | | Not husy | iny not too | SS | 20 | 271 | 67 | 104 | 18 | 63 | 19 |
| | | Now Usua | ally a little h | \$\$ | 80 | 165 | 39 | 57 | 13 | 42 | 14 |
| | | | 10001 | \$\$ | 417 | 1354 | 920 | 323 | 17 | 66 | 28 |
| | | | | | 360 | 4 | 1 | 1 | 2 | C | 0 |
| | | | | | 359 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 358 | 4 | 1 | 1 | 0 | 2 | C |
| | | | | | 357 | 2 | 0 | 0 | 1 | 1 | C |
| | | | | | 320 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 356 | 1 | 1 | 0 | 0 | C | 0 |
| | | | | | 355 | 0 | 0 | 0 | 0 | C | 0 |
| | | Now: Usua | ally a little b | usy | 354 | 2 | 0 | 0 | 0 | 1 | . 1 |
| | | | | | 353 | 23 | 6 | 9 | 1 | 5 | 2 |
| | | | | | 352 | 14 | 5 | 5 | 1 | 3 | C |
| | | | 518010 | | 351 | 31 | 14 | 7 | 0 | 9 | 1 |
| | | | | | 350 | 6 | 3 | 2 | 0 | 1 | . C |
| | 100 | Busier than | n usual | | 349 | 92 | 46 | 24 | 10 | 8 | 4 |
| | | | | | 348 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 347 | 0 | 0 | 0 | 0 | C | C |
| | | Now: Usua | ally a little b | \$\$\$ | 346 | 1 | 0 | 1 | 0 | C | C |
| | 32 | Not too bu | Jsy | | 319 | 4272 | 1738 | 1726 | 87 | 617 | 104 |
| | | | | | 318 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 317 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 316 | 0 | 0 | 0 | 0 | C | C |
| | | | | | 380 | 0 | 0 | 0 | 0 | C | C |
| | | | 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.



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.

Micro-module A-A4: Social Media Data Gathering and Processing

| 🔇 *Untitled Project – | - QGIS | | | | _ |
|-----------------------------------|--|---------------------------------|---|--------------|-------------|
| Project <u>E</u> dit <u>V</u> iew | <u>Layer</u> <u>S</u> ettings <u>P</u> lugins Vect <u>o</u> r <u>R</u> ast | er <u>D</u> atabase <u>W</u> eb | <u>M</u> esh Pro <u>c</u> essing <u>H</u> elp | | |
| i 🗋 🚞 🖥 🔂 😫 | 🐙 <u>D</u> ata Source Manager | Ctrl+L | 🖢 🔁 🗉 🔍 🛅 🧱 🌞 \Sigma 🛲 🗸 🥐 🍭 🗸 🖵 | • | |
| 👘 🐼 V° 🖉 📖 | Create Layer | • | | | |
| | Add Layer | Þ | V _D Add Vector Layer | Ctrl+Shift+V | |
| 🖳 🔻 📑 👻 🌄 🖛 🗖 | Embed Layers and Groups | | 📲 Add Raster Layer | Ctrl+Shift+R | |
| Browser | Add from Layer Definition File | | 🕌 Add Mesh Layer | | |
| l C T 🕈 🛛 | Copy Style | | 🤊 Add Delimited Text Layer | Ctrl+Shift+T | 7 |
| ☆ Favorites | Paste Style | | 🧠 Add PostGIS Layers | Ctrl+Shift+D | |
| III Spatial Bookma | Copy Layer | | 🎤 Add SpatiaLite Layer | Ctrl+Shift+L | |
| Home | Paste Layer/Group | | Madd MSSQL Spatial Layer | | |
| C:\ | Open <u>Attribute</u> Table | F6 | 🖳 Add DB2 Spatial Layer | Ctrl+Shift+2 | IIII |
| | / Toggle Editing | | Add Oracle Spatial Layer | Ctrl+Shift+O | 1 2395 |
| GeoPackage | Save Layer Edits | | 💹 Add/Edit Virtual Layer | | And the set |
| SpatiaLite | // Current Edits | • | 🧐 Add WMS/WMTS Layer | Ctrl+Shift+W |) a at |
| PostGIS | Save As | | Add XYZ Layer | | |
| 🗯 MSSQL | Save As Laver Definition File | | Add Arc <u>G</u> IS Map Service Layer | | 1//-60 |
| Oracle | Remove Laver/Group | Ctrl+D | Add WCS Layer | | |
| DB2 | Duplicate Laver(s) | curre | Add WFS Layer | | C MADA |
| WMS/WMTS | Set Scale Visibility of Laver(s) | | Add ArcGIS Feature Service Layer | | Lanon EP |
| Vector Liles | Set CRS of Laver(s) | Ctrl+Shift+C | Add Vector Tile Laver | 1 | 6 |
| | Set Project CPS from Laver | curronnere | | | THE |
| WCS | Laver Properties | | | | |
| WFS / OGC AP | Filter | Ctrl+F | | | IHA 2 |
| 1 | Filter | Cultr | | | -10- |
| Lavers | Cabeling | | | HLEN | V FT |
| 😺 🕼 👁 🍸 🖏 🔻 🚺 | Show In Overview | | | 向用 二 | |
| — map lines | Show All In Overview | | | | ARIT |
| ✓ boundary | - Hide All from Overview | A SEL Mas | | HEREY LA | BBX |
| ✓ map multip | olyg | 1 673 2 | | PAYA | |
| map other_rela | tions | | | KITC | |

- 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.

| Q Data Source | e Manager Delimited Text | | × |
|--|---|--|----------------------------------|
| 📛 Browser 🔶 | File name fanChan\RA\TDLEG\A4\dataset\dataset | taset_crawler-google-places_2022-01-03_07-45-38-003.csv 🖾 | |
| V Vector | Layer name poi_coffee | Encoding UTF-8 | • |
| Raster Mesh Delimite Text | CSV (comma separated values) Regular expression delimiter Custom delimiters Regular explade Options | | |
| 😪 GeoPack 🥖 SpatiaLite | Number of header lines to discard 0 ✓ First record has field names ✓ Detect field types | Decimal separator is comma Trim fields Discard empty fields | |
| MSSQL Oracle | Geometry Definition Point coordinates Well known text (WKT) No geometry (attribute only table) | X field location/lng V Z field Y field location/lat V M field DMS coordinates Geometry CRS invalid projection | ▼ ▼ |
| WITUAI Layer WMS/ WMTS WFS / | Layer Settings Sample Data 29 Tai Kok Tsui Rd, Tai Kok Tsui, Hoi Control Time Company HERGES 5 | address ng Kong a Callide Dand 19# Kunang Wa Ch. Mang Kalu Hang Kang | |
| API - Features | The CRS must be selected | Close Add | Help |

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 <utmlefted rule 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.

| h R Hama | | <hr/> | | | | | | | | | | | |
|---------------------------|----------|-------|---------|------------------|----------------------|---------------------|-------------------|-----------------|-------------|-----------------|-----------------|--------------|-----|
| IO Home | | | 1 | | | | | | 11 | | | | |
| | | | | 1 | | | ~ ~ | | | | TTM T | | |
| ET EA | | | 0 | | | and the second | 明 面 胸 習 四 | El D / Py | | T | EQ353 1991 | u n | 1 |
| 😵 GeoPackage | 12 | 1 | Q | Clipped_poi_co | ffee — Features T | otal: 10, Filtered: | 10, Selected: 0 | | | | | - 0 | × |
| / SpatiaLite | | | 1 | | < 0 0 1 6 = 5 | I 🔍 🝸 🗉 🏷 😥 I | 15 (15 🗶 28) 🖷 (| = e . | | | | | |
| POSTGIS | | 4 | | address | categories/0 | category/Name | cid | city | countryCode | isAdvertisement | locatedin | location/lat | Ic |
| Oracle | | | 1 | Grateful Time | Coffee shop | Coffee shop | 2.40378e+1 | 8 Mong Kok | нк | FALSE | Yan On Building | 22.31595 | 59 |
| @ WMS/WMTS | | | 2 | 彌敦道577號 高 | Coffee shop | Coffee shop | 1.37232e+1 | 9 Mong Kok | нк | FALSE | 高氏大樓 | 22.3151 | 72 |
| * XYZ Tiles | | | 3 | Hong Kong, M | NULL | NULL | 1.75261e+1 | 9 Mong Kok | нк | FALSE | Silver Commer | 22.32097 | 16 |
| WCS | | 7 | 4 | 193 Prince Ed | Electric vehicle | Electric vehicle | 2.0051e+1 | 8 Mong Kok | нк | FALSE | моко | 22.3 | 23 |
| WFS / OGC API - | * | | 5 | 555 Shanghai | Electric vehicle | Electric vehicle | 2.41921e+1 | 8 Mong Kok | нк | FALSE | Cordis, Hong K | 22.31877 | 34 |
| Layers a | 8 | | 6 | Gala Place, 56 | Electric vehicle | Electric vehicle | 5.38747e+1 | 8 Mong Kok | нк | FALSE | park-in comm | 22.3151 | 49 |
| | | | 7 | 地下及閣槵 25 | Drug store | Drug store | 1.73238e+1 | 9 Prince Edward | нк | FALSE | NULL | 22.32324 | 11 |
| ✓ ○ Clipped_poi □: | f | | 8 | Shop 10-19, Le | Clothing store | Clothing store | 1.26645e+1 | 7 Mong Kok | нк | FALSE | Langham Place | 22.31899 | 53 |
| V boundary | | | 9 | Wu Sang Hous | Pharmacy | Pharmacy | 1.45566e+1 | 9 Mong Kok | нк | FALSE | NULL | 22.31845 | 36 |
| a map other_relation | 9' ns | | 10 | Shop 35b in Ba | Drug store | Drug store | 3.43978e+1 | 8 Mong Kok | нк | FALSE | Floor B2 · Lang | 22.31873 | 26 |
| | | | | | | | | | | | | | |
| | | | 4 ₹: | Show All Feature | 5 . | | | | | | | | 3 1 |
| | | | | 4 | | | | | Y | | <u>j</u> řk | | |

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