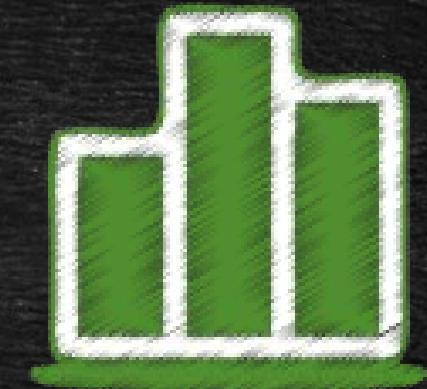


SPARK STREAMING PROJECT

Analyzing Instagram Trending Data

Presented by :
Jeanderson Almeida



About Project

Analyzing Instagram stream
trend data and publishing it to
real-time dashboard

Our analysis subjects;

1. The most profiles per city?
2. The most posts per city?
3. The most posts per profile?
4. The most likes per profile?

Data Source

instagram_locations.csv

Field	Description
SID	Sequence ID
ID	Instagrams ID for that could be used on the website ex: ID=230466055 the url is https://www.instagram.com/explore/locations/230466055
Name	Locations Name
Street	Street Address, may contain '\n'
ZIP	Zip code
City	City Name
Region	Region
CD	Country Code
Phone	The phone in format as on the Instagram
aj_exact_city_match	The Instagrams Internal key
aj_exact_country_match	The Instagrams Internal key
blurb	Description of the place, may contain '\n'
dir_city_id	The Instagrams internal City ID
dir_city_name	city Name
dir_city_slug	City tag (sortof)
dir_country_id	Country ID
dir_country_name	country
lat	Latitude
lng	Longitude
primary_alias_on_fb	Bool Flag
slug	
website	The URL to web site, may contain more then 1 URL, may contain '\n'
cts	Timestamp when the location was visited

Data Source

instagram_profile.csv

Profiles

Field	Description
SID	Sequence ID
profile_id	Instagrams ID
profile_name	profile name
firstname_lastname	firstname lastname
description	May contain '\n'
following	Number of following profile at the moment it was visited
Followers	Number of followers at the moment it was visited
n_posts	Number of posts at the moment it was visited
URL	Url
cts	Timestamp when the profile was visited
is_business_account	Boolean flag if that profile was makred by the owner as business account

```
"sid" "profile_id" "profile_name" "firstname_lastname" "description" "following" "followers" "n_posts" "url" "cts" "is_business_account"
4184446 4721049632 "jphillip033" "John Pierce" """Document Everything Always"" 250 146 170 2019-08-01 14:38:55.394 -0400 "false"
4184457 590583002 "mama haas" "Deanna" "trying to enjoy the simple things in life. Knitting, crocheting, canning, baking, gardening, taking care of my family!" 534 1145
4184460 1074146658 "yellowlightbreen" "Eliza Gray" "Maine is for Lovers" 469 324 431 "elizajanegray.com" 2019-08-01 14:39:54.407 -0400 "false"
4184461 1472038842 "tecl025" "Thomas Clark" 2019-08-01 14:40:06.472 -0400
4184462 3531420803 "luckyfluff" "" 2019-08-01 14:40:07.806 -0400
4184465 145064167 "sabahike" "Sarah bahlke" "" 266 192 590 2019-08-01 14:40:16.443 -0400 "false"
4184471 2061867521 "masslivehs" "MassLive High School Sports" "Your spot for the best Western Mass. high school sports coverage. Submit your photos and videos by tagging @masslivehs" 2019-08-01 14:40:19.771 -0400 "false"
4184472 1446650654 "hvcanes" "Hoosac Valley" 2019-08-01 14:40:52.635 -0400
4184475 1743726268 "will_jay_k" "William Kramer" "I'd rather die a big death than live a small life" 115 183 37 2019-08-01 14:40:59.969 -0400 "false"
4184476 5455198037 "ashley_downing722" "Ashley Downing" 2019-08-01 14:41:12.826 -0400
4184479 531147579 "touristiers" "tori <3>" "% fulton mo \n\ fhs 20 \n\ @breezebridal ambassador" 679 2034 39 2019-08-01 14:41:19.771 -0400 "false"
4184482 303800693 "valeriesublett" "val" 2019-08-01 14:41:33.642 -0400
4184483 1689604658 "lillian.drew" "L i 1" 2019-08-01 14:41:35.002 -0400
4184486 304139153 "katapani413" "kp!!!! a" "cowley college volleyball\n3.18.19 @garrett.lange" 1543 3008 115 "@sco.co/kpani" 2019-08-01 14:41:46.967 -0400 "true"
4184488 49002935 "emma.abbott_" "E m m a A b b o t t" 2019-08-01 14:41:59.226 -0400
4184489 366754332 "kelsieskinner" "Kelsie Skinner" 2019-08-01 14:42:00.579 -0400
4184491 928132400 "kyleighdavis_" "Kyleigh Davis" 2019-08-01 14:42:01.915 -0400
4184500 1259258021 "aniferres10" "Anioler Ferres" "DEFENSAR LA TERRA NO ÉS CAP DELICTE" 862 723 13 2019-08-01 14:42:09.115 -0400 "true"
4184504 471102662 "ferranfafajula_" "Ferran Fajula G" 2019-08-01 14:42:21.371 -0400
4184505 493386517 "ponchete_9" "eduardoponce" 2019-08-01 14:42:22.720 -0400
4184506 889666287 "cesarpozas" "Cesar Pozas" 2019-08-01 14:42:24.057 -0400
4184507 1201060776 "nilpalcome" "Nil Palomé" 2019-08-01 14:42:25.422 -0400
4184509 1400407075 "laaaauuuusss" "LAUUU" 2019-08-01 14:42:27.373 -0400
4184510 1638211029 "carlespd" "Carles Domenech" 2019-08-01 14:42:28.715 -0400
4184466 10978291950 "maison_exclusive_cv" "Maison Exclusive" "" 590 299 99 2019-08-01 14:40:29.506 -0400 "true"
4184468 8559205476 "maxime_grauls" "Maxime G" 2019-08-01 14:40:41.458 -0400
4184474 3957867328 "thyago_limirio" "Thyago Limirio" 2019-08-01 14:40:59.822 -0400
4184478 6095905332 "gerardalleaume" "ALLEAUME" "" 21 21 730 2019-08-01 14:41:08.475 -0400 "false"
4184481 567655114 "paige_hockman" "Paige Hockman" "missouri state" 729 1044 54 2019-08-01 14:41:28.506 -0400 "false"
4184484 21966754 "carolinesassner" "Caroline Sasser" 2019-08-01 14:41:40.672 -0400
4184487 235174956 "tucker_caswell" "Tucker Caswell" "DU Wrestling" 941 1753 295 2019-08-01 14:41:47.557 -0400 "false"
4184490 252820589 "devinfenero" "Devin ferro" 2019-08-01 14:42:00.921 -0400
4184492 273602282 "hunterwilson_20" "Hunter Wilson" 2019-08-01 14:42:02.281 -0400
4184493 375192247 "ylg.cam" "Cam-Ron" 2019-08-01 14:42:04.242 -0400
4184494 383613016 "m.a.k.y.g.h" "Makygh Galbreath" 2019-08-01 14:42:05.596 -0400
4184495 390135644 "j0rdanshaw" "J-Shaw" 2019-08-01 14:42:06.953 -0400
4184497 400341604 "jacobkern_25" "Jacob" 2019-08-01 14:42:08.297 -0400
4184498 402600158 "solomon_huff" "Boo Huff" 2019-08-01 14:42:09.658 -0400
4184499 428631777 "coltonbrisistol" "Colton Bristol" 2019-08-01 14:42:11.007 -0400
4184500 513453897 "na3ir_lowlife3" "" 2019-08-01 14:42:12.370 -0400
4184501 1213515450 "ylg.chey" "YLG Chey Q&Q" 2019-08-01 14:42:13.711 -0400
4184502 1729393571 "tre.vaughn1" "Tre Vaughn" 2019-08-01 14:42:15.062 -0400
4184508 283994482 "destridungan" "Destri" "JTH//MSU" 1109 1217 27 2019-08-01 14:42:21.888 -0400 "false"
4184515 134083906 "marvelized52" "Nicole Efferson" 2019-08-01 14:42:34.278 -0400
4184517 189634647 "farmerdow1" "farmerdow1" "Trying to run a business and fall in love ☺" 2566 1808 1324 2019-08-01 14:42:45.900 -0400 "false"
4184528 195744263 "raffaela_maiello" "Raffaela Maiello" "There is no way to happiness. Happiness is the way..." 285 291 794 2019-08-01 14:43:05.635 -0400 "false"
4184536 3683070404 "ccap969" "Claire Capdet-Tisseyre" "@France Pays Catalan \nUne photo prise ce jour ou lors d'un voyage ambiance nature... @" 268 157 225 2019-08-01
4184539 543192226 "azijaff" "Arazu Jaff" "| Mummy | Wife | ❤ Traveling | Entrepreneur | Beverly's" 285 655 614 2019-08-01 14:43:42.471 -0400 "true"
4184557 4729769915 "boxofgoth" "Box of Goth" 2019-08-01 14:44:58.126 -0400
```

Data Source

instagram_posts.csv

Posts

Field	Description
SID	Sequence ID
sid_profile	Sequence ID of the profile from ## Profiles tabke
post_id	Instagrams ID
profile_id	Instagrams ID may be null
location_id	Instagrams ID
cts	Timestamp when the Post was created
post_type	1 - Photo, 2 - Video, 3 - multy
description	May contain '\n'
numbr_likes	Number of Likes at the moment it was visited
number_comments	Number of comments at the moment it was visited

sid	sid_profile	post_id	profile_id	location_id	cts	post_type	description	numbr_likes	number_comments
28370919	3496776	BXdjjU1gcqg	2237947779	1022366247837915	2017-08-06 20:06:57.000	2	Wreckloose! Deevalley bike park laps on the @intensecycles recluse!\n@alfjonescycles\n@saddleback_1 ...	80	0
28370932	-1	BVg0pb0lYBC	5579335020	457426771112991	2017-06-19 09:31:16.000	1	believe in ya dreams just like I believe in male lace rompers and boneless watermelon\n\nT ...	25	1
28370933	-1	BRgkjcXFp3Q	313429634	457426771112991	2017-03-11 20:05:03.000	1	#meraviglia #incensi #the #candele #profumo #amicizia #primavera #starshine	9	0
28370934	-1	BTKKeNhjEA7	1837592700	457426771112991	2016-09-13 16:27:16.000	1	#teatime #scorpiun #friends #love #mountains #september	4	0
28370935	-1	8-NQrv0YLX	1131527143	457426771112991	2015-10-18 10:19:27.000	1	thE sky gave a #constEllation	8	0
28370964	-1	BrYDPJeABJQ	16262389	282618748	2018-12-14 18:16:15.000	1	#beautiful #Christmas #lights	138	15
28370957	-1	BmQ418mgSnL	35673872	282618748	2018-08-09 17:53:46.000	1	As per your suggestion. #themerryriotakeslondon	389	10
28370960	-1	Bm1YQ3Hn1oJ	840873396	282618748	2018-08-23	1	-Blue-♥	198	23

Methodology

Flow Management



StreamSets

Reading csv file and produce record to Kafka Topic
ETL jobs (CSV to JSON, JSON to SQL etc.)
After Spark Analyzing sessions sink data to Nosql and RDBMS

Stream Processing



kafka

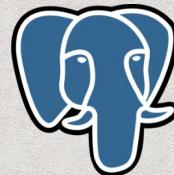


Listen to Kafka Topic
Analyzing a stream data from kafka Topic using Scala-Spark
Writing to a different Kafka Topic after analysis

Databases



elasticsearch



Importing Analyzed data to NoSQL and RDBMS systems.

Querying by running KQL over Elasticsearch

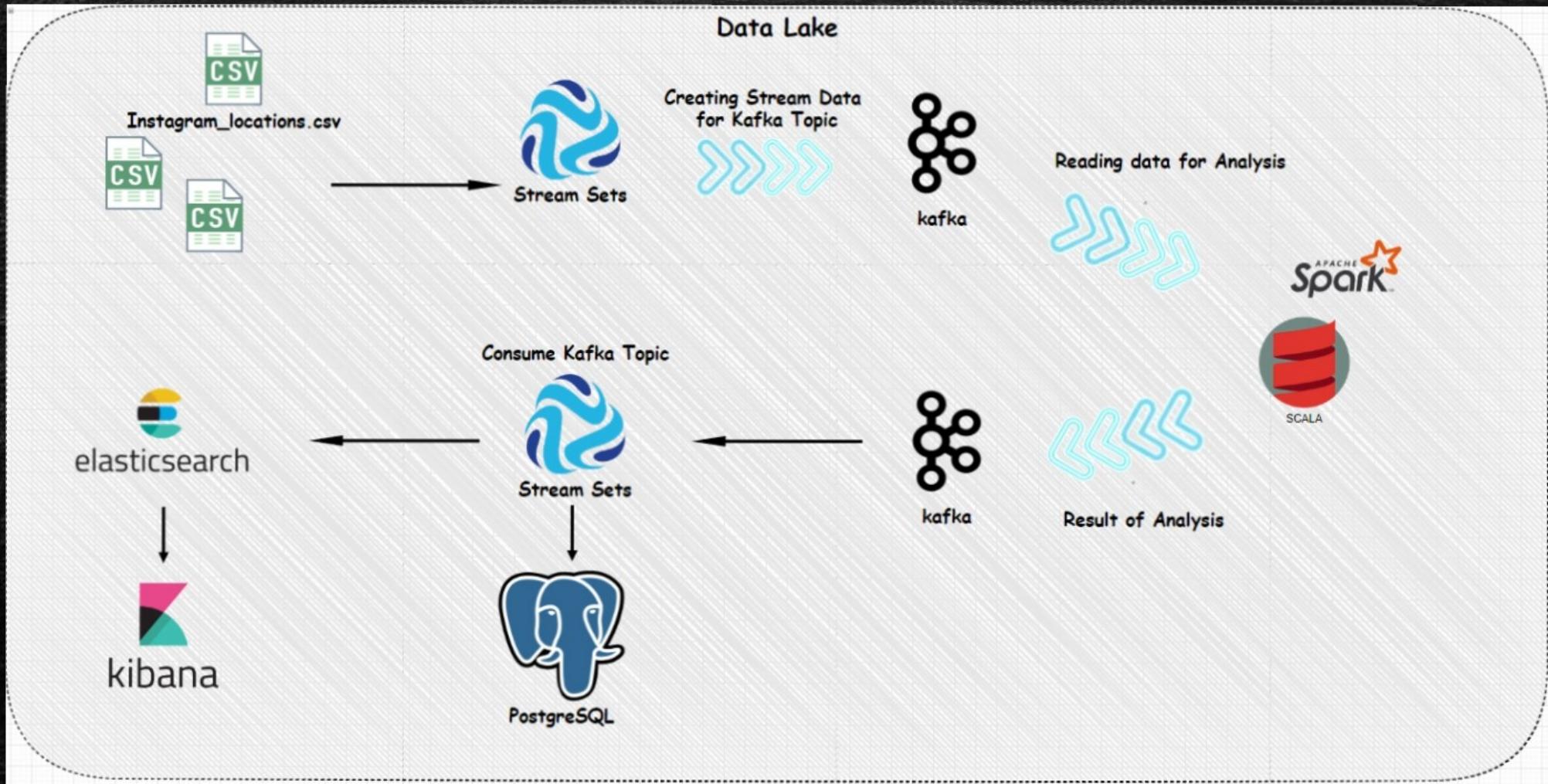
Dashboard



kibana

Creating a dashboard every 1 second over real-time data

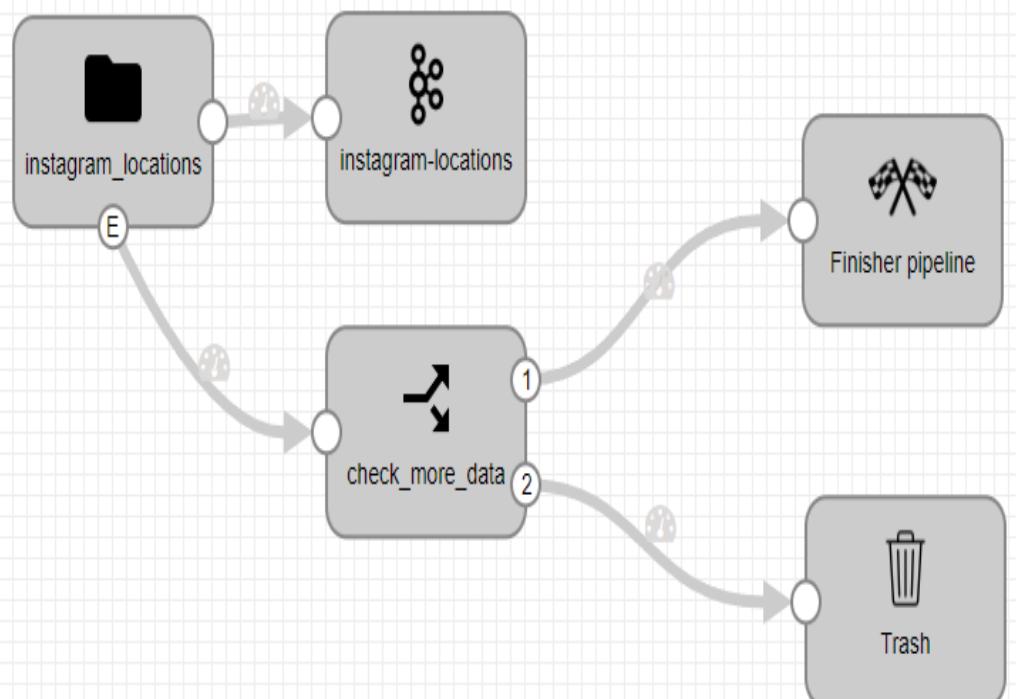
Project Overview



Project Details

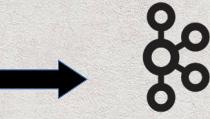


Pipelines / instagram_pipeline



→

{ json }



instagram_locations.csv

```
{"id": "772979", "id": "417805423", "name": "Emap - Porto Do Itaqui - Ponta Da Espera", "street": "", "zip": "65060-641", "city": "S\u00e3o Luis, Brazil", "region": "", "cd": "BR", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "-44.354091", "lng": "-2.536156"}, {"primary_alias_on_fb": "", "slug": "emap-porto-do-itaqui-ponta-da-espera", "website": "", "cts": "2019-05-31 08:45:22.611"} {"id": "391884", "id": "1076049002496141", "name": "Amedkar Park", "street": "", "zip": "226010", "city": "Lucknow, Uttar Pradesh", "region": "", "cd": "IN", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "c2797174", "dir_city_name": "Monti Nagar", "dir_country_id": "IN", "dir_country_name": "India", "lat": "81.002330", "lng": "26.853130"}, {"primary_alias_on_fb": "", "slug": "amedkar-park", "website": "", "cts": "2019-04-07 12:48:22.991"} {"id": "772953", "id": "415198296", "name": "Кугузавский огурец ЗАГС", "street": "Кугузавский проспект", "zip": "121151", "city": "Moscow, Russia", "region": "", "cd": "RU", "phone": "+99-2493927", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "c1996942", "dir_city_name": "Krasnaya Presnya", "dir_country_id": "RU", "dir_country_name": "Russia", "lat": "55.744773", "lng": "37.548606"}, {"primary_alias_on_fb": "", "slug": "", "website": "http://tags.mos.ru/", "cts": "2019-05-31 08:43:43.581"} {"id": "772963", "id": "4186993987613888", "name": "Ресторан 'Как в детстве'", "street": "Невский проспект 63", "zip": "191106", "city": "Saint Petersburg, Russia", "region": "", "cd": "RU", "phone": "+89633090359", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "Новый уютный ресторан в сердце северной столицы", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "30.352850", "lng": "59.931690"}, {"primary_alias_on_fb": "", "slug": "", "website": "", "cts": "2019-05-31 08:44:20.919"} {"id": "772984", "id": "19904390793061", "name": "Munch Box", "street": "", "zip": "", "city": "", "region": "", "cd": "", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "Best Breakfast and Burger Joint\nin the Heart of London", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "0.040684", "lng": "51.531280"}, {"primary_alias_on_fb": "", "slug": "munch-box", "website": "http://www.munchboxhalal.com", "cts": "2019-05-31 08:45:39.226"} {"id": "772986", "id": "181682893", "name": "澳门新濠天地/水舞间", "street": "新濠天地", "zip": "9000", "city": "Macao, China", "region": "", "cd": "MO", "phone": "+853 88686680", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "13.562817", "lng": "22.149910}, {"primary_alias_on_fb": "", "slug": "macao-new-hotel", "website": "", "cts": "2019-05-31 08:45:46.487"} {"id": "772999", "id": "15190600", "name": "Mamoshino", "street": "", "zip": "", "city": "", "region": "", "cd": "", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "c2014507", "dir_city_name": "Mamoshino", "dir_country_id": "RU", "dir_country_name": "Russia", "lat": "36.418166", "lng": "55.928961"}, {"primary_alias_on_fb": "", "slug": "", "website": "", "cts": "2019-05-31 08:46:36.153"} {"id": "772998", "id": "1818667248", "name": "Praça da Harmonia", "street": "Praça da Harmonia, S/N", "zip": "20221-160", "city": "Rio de Janeiro, Rio de Janeiro", "region": "", "cd": "BR", "phone": "+552140633999", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "22.92126", "lng": "-43.192126"}, {"primary_alias_on_fb": "", "slug": "", "website": "", "cts": "2019-05-31 08:46:36.153"} {"id": "772996", "id": "191268620", "name": "Uncle Cheffy Magnolia", "street": "4/F, Two E-Com Building, Mall of Asia Complex, Ocean Drive", "zip": "3000", "city": "Quezon City, Philippines", "region": "", "cd": "PH", "phone": "+6329474069", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "Located in Robinsons Magnolia Town Center, Dona Hemady cor Aurora Blvd, New Manila, Quezon City. Contact us at 029881249 and 09277418987 :)", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "14.038446", "lng": "121.641300"}, {"primary_alias_on_fb": "", "slug": "unclecheffymagnolia", "website": "http://www.unclecheffygroup.com", "cts": "2019-05-31 08:46:20.282"} {"id": "772988", "id": "41876425", "name": "SundaySky", "street": "229 W 36th St, Fl 2nd", "zip": "10019", "city": "New York, New York", "region": "", "cd": "US", "phone": "+1212 929-9111", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "Advancing the audience of one with personalized video engagement.", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "40.753010", "lng": "73.990852}, {"primary_alias_on_fb": "", "slug": "sundayskyvideo", "website": "http://www.sundaysky.com", "cts": "2019-05-31 08:45:53.745"} {"id": "391901", "id": "4156752395", "name": "Georgina Lake", "street": "", "zip": "L4P", "city": "Keswick, Ontario", "region": "", "cd": "CA", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "44.4980146", "lng": "79.480146"}, {"primary_alias_on_fb": "", "slug": "georgina-lake", "website": "", "cts": "2019-04-07 12:49:39.330"} {"id": "391906", "id": "4386158486", "name": "Angies Firehouse Tavern", "street": "703 Waterlyet Ave", "zip": "45420", "city": "Dayton, Ohio", "region": "", "cd": "US", "phone": "+1937 252-5125", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "Angie's Firehouse Tavern", "dir_city_id": "c2751700", "dir_city_name": "Belmont, Dayton", "dir_country_id": "", "dir_country_name": "", "lat": "39.731470", "lng": "84.142660}, {"primary_alias_on_fb": "", "slug": "angies-firehouse-tavern", "website": "https://www.groupon.com/biz/dayton/angies-firehouse-tavern", "cts": "2019-04-07 12:49:50.451"} {"id": "773004", "id": "191227720", "name": "Stade P\u00e8re-J\u00e9go", "street": "", "zip": "", "city": "Casablanca, Morocco", "region": "", "cd": "MA", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "c1402574", "dir_city_name": "Oasis", "dir_country_id": "MA", "dir_country_name": "Morocco", "lat": "33.954743", "lng": "7.651486"}, {"primary_alias_on_fb": "", "slug": "stade-pere-jego", "website": "", "cts": "2019-05-31 08:46:50.585"} {"id": "773006", "id": "191523047", "name": "Hoofddorppleinbuurt", "street": "Hoofddorpplein", "zip": "1058/1059", "city": "Amsterdam, Netherlands", "region": "", "cd": "NL", "phone": "", "aj_exact_city_match": "false", "aj_exact_country_match": "false", "blurb": "", "dir_city_id": "", "dir_city_name": "", "dir_country_id": "", "dir_country_name": "", "lat": "4.849700", "lng": "52.351100}, {"primary_alias_on_fb": "", "slug": "hoofddorppleinbuurt", "website": "", "cts": "2019-05-31 08:46:58.414"} {"id": "772992", "id": "18629418556989", "name": "The Life Centre", "street": "i Britannia Row", "zip": "N1 8QH", "city": "London, United Kingdom", "region": "", "cd": "GB", "phone": "+442077040
```