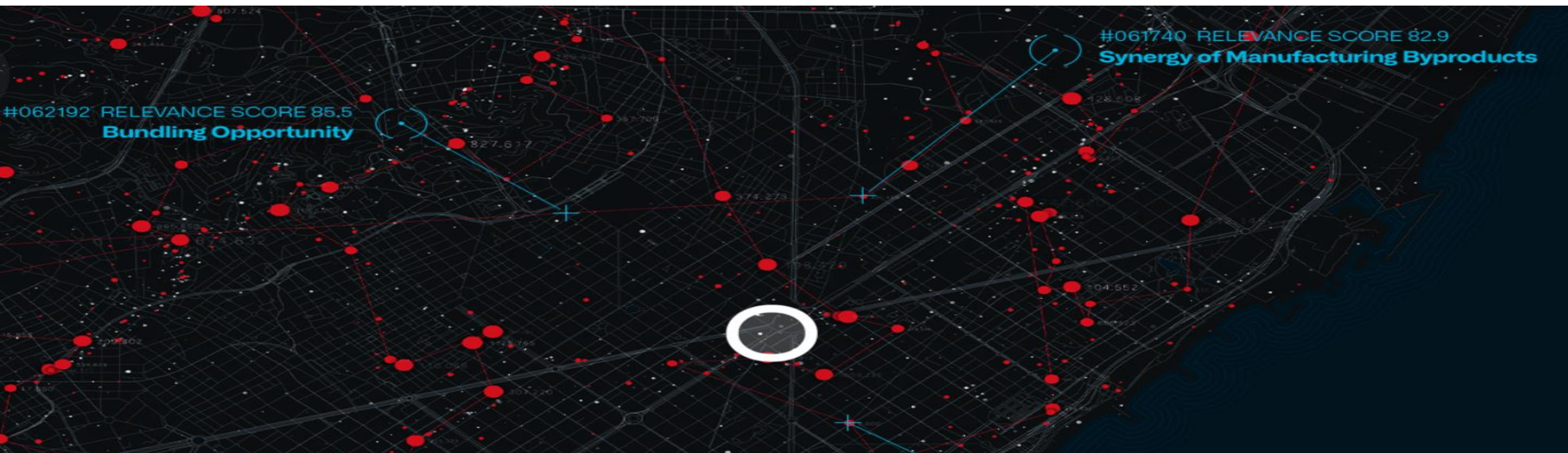


# HOW TO ENRICH INTERNAL DATABASE via GEOSPATIAL INFORMATIONS for Retailers and FMCG



**Matching**

**TripAdvisor**



## Matching Step

---

- What?

Enriching the Customer Master Data (CMD) by searching for each HoReCa outlet its corresponding entry in TripAdvisor.

- How?

Each outlet from the CMD is compared against all the TripAdvisor entries in a 2km radius.

The comparison is done based on 3 parameters:

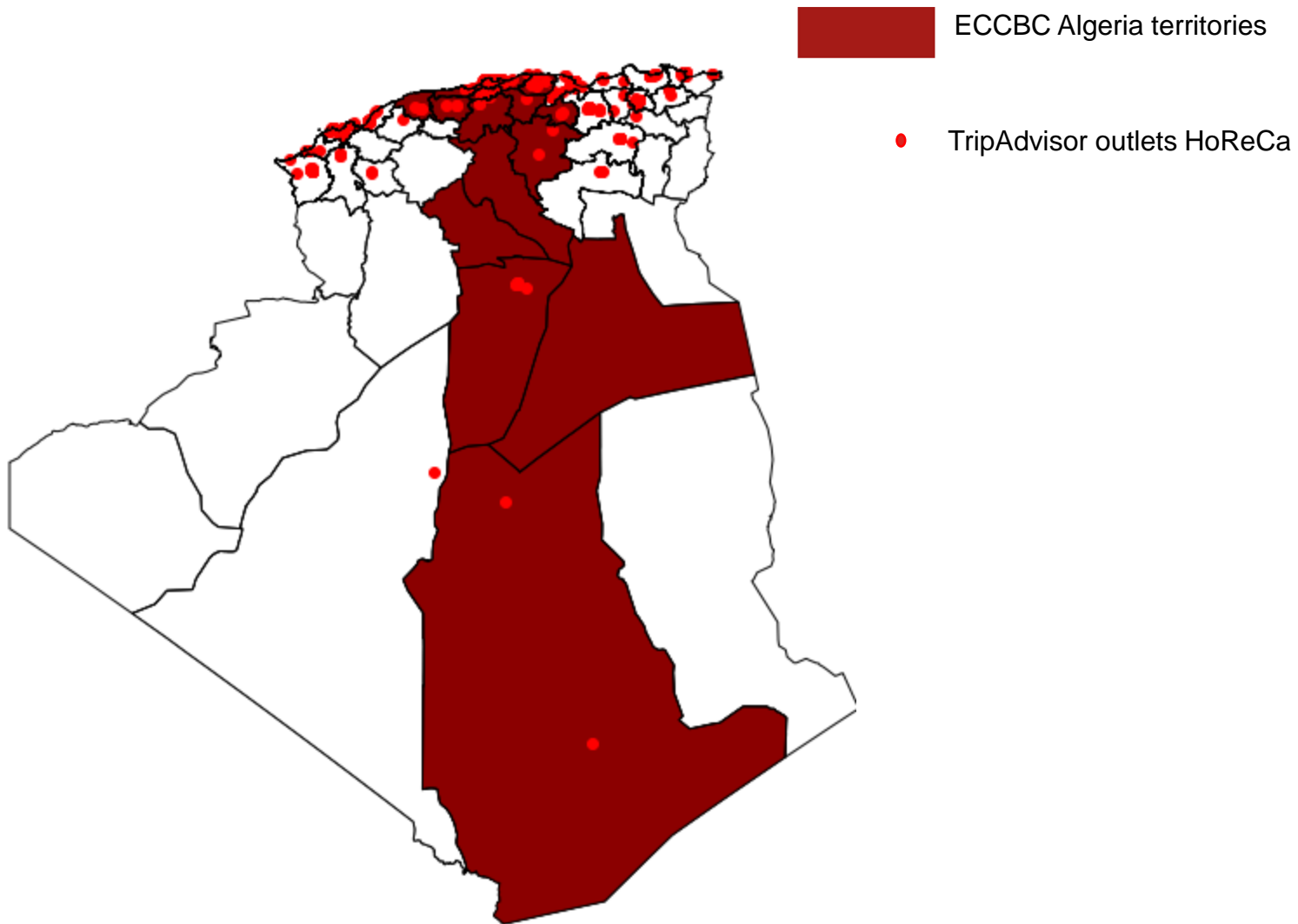
1. Name (weight :75%)
2. Address (weight :15%)
3. Distance based on the geolocation – GPS coordinate (weight :10%)

- Based on this, an overall similarity score is calculated, and the TripAdvisor entries are ranked based on it. The entry with the highest similarity is the potential candidate.
- In some cases, even the potential candidate isn't a proper match, hence, we have defined a similarity threshold above which the candidate is considered a match.



## TripAdvisor data

### Scraping Coverage



Wilaya	count TripAdvisor poi horeca
Alger	458
Tizi Ouzou	27
Blida	23
Boumerdes	13
Tipaza	9
Ghardaia	9
Chlef	5
Bordj Bou Arrer	3
Ain-Defla	3
Tamanrasset	2
MSila	2
Medea	2
Bouira	1
Oran	203
Annaba	46
Setif	44
Bejaia	41
Tlemcen	37
Constantine	34
Mostaganem	22
Skikda	14
Batna	13
Biskra	11
Jijel	7
Sidi Bel Abbes	5
Ain-Temouchent	4
Guelma	3
Saida	2
Mila	2
El-Tarf	2
Oum El Bouaghi	1
Relizane	1
Adrar	1



## TripAdvisor data

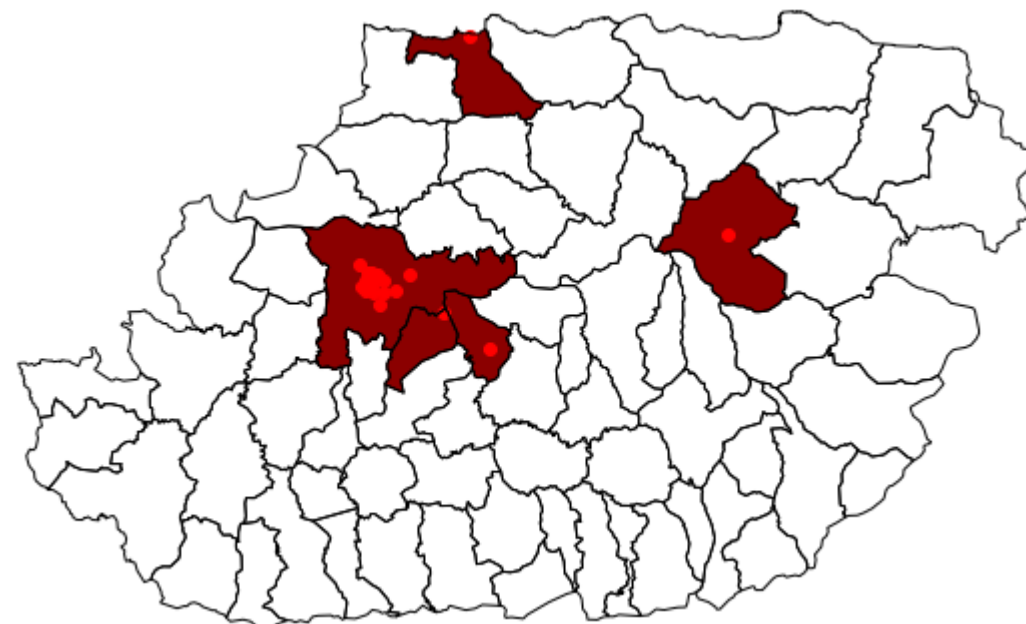
### Scraping Coverage

	Snack	HoReCa	scrapping	scraping HoReCa coverage
Count in Algeria	5483	5483	1237	23%
Count in Tizi Ouzou	1480	1006	27	3%

	HoReCa	OTHER	SNACK
Gold	1%	0%	2%
Bronze	67%	0%	9%
Silver	30%	0%	13%
Tin	1%	100%	75%

Wilaya Tizi Ouzou : 27 outlets in TripAdvisor

Availability of information about consumption occasions & type of services however, TripAdvisor outlets in Tizi Ouzou is limited for the matching step



## However, apart from the limitations of the TripAdvisor data, we've got some learnings on our own data

---

- **Address:** in some cases, the address field in the CMD isn't filled in with a real address but more with indications on how to find the outlet. These indications cannot be matched to any address in TripAdvisor and reduce the chances of matching.
  - "ARRÊTE DE BUS"
  - "SUD OUEST LAVILLA"
  - "SORTIR DE BOUKHALFA"

### Recommendation:

The Address field in the CMD should always be populated with the real address. We'd require to ask the pre-sellers to confirm the addresses as they visit the stores.

- **Name:** in most of cases, the name in the CMD is not entered with the name of the restaurant but rather with the name of the owner / manager. Therefore, name match does not work with TripAdvisor.. Hence, the matching on name doesn't work with TripAdvisor.

<ul style="list-style-type: none"><li>• "BEN ADJAL MOHAMED"</li><li>• "DAHOU MOHAMED"</li></ul>	<div>VS</div>	<ul style="list-style-type: none"><li>• Restaurant L'Ambassade</li><li>• Restaurant Le Bagdad</li><li>• Restaurant Hermes lounge cafe</li></ul>
---	---------------	---

### Recommendation:

The Name field in the CMD should always be populated with the public name of the restaurant. We'd require to ask the pre-sellers to confirm the names as they visit our customers.

# Online Data sources

## HORECA



*Bottling Happiness in Africa  
Since 1989*

## According to Commercial team feedbacks, TripAdvisor have not a good coverage in Algeria

---

The purpose of the following slides is to see the different possibilities to records online data sources for horeca first of all but also for grosseries if there is.

- Open Source Data : OpenStreetMaps, Google Maps, Facebook restaurants data
- Data aggregator | Company



# Carto premium POI

	Count POI	
	Tizi Ouzou	Algeria
Eating place/Restaurant	565	15 731
Drinking Places/Cafe	68	1 224
Hotel/ Motel	97	2 766
<b>All</b>	<b>730</b>	<b>19 721</b>

## Useful columns in datasets and missing value (%)

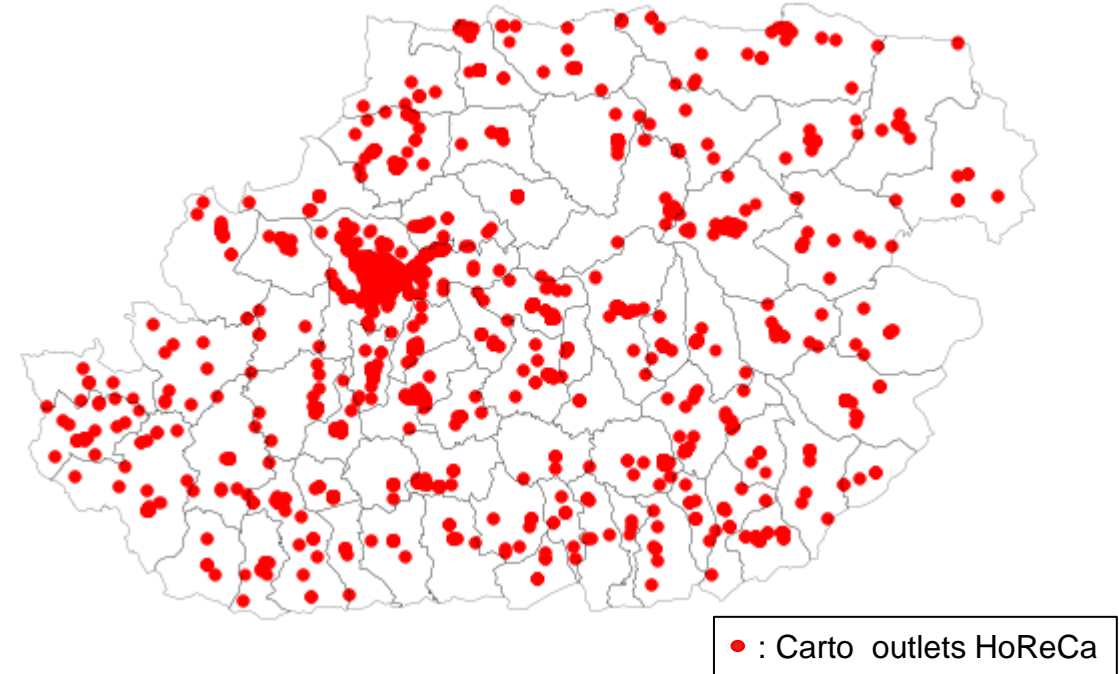
- Latitude & longitude 0%
- Adresses 0%
- Outlets names 0%
- Website : 92%
- Open\_24 : 99,99%

396 Outlets in All Algeria and 16 outlets in Wilaya Tizi Ouzou  
among Carto HoReCa poi use facebook as website

Non-availability of informations about consumption occasions & type of services :

→Carto POI data cannot be used for matching

Carto HoReCa POI répartitions in Tizi Ouzou



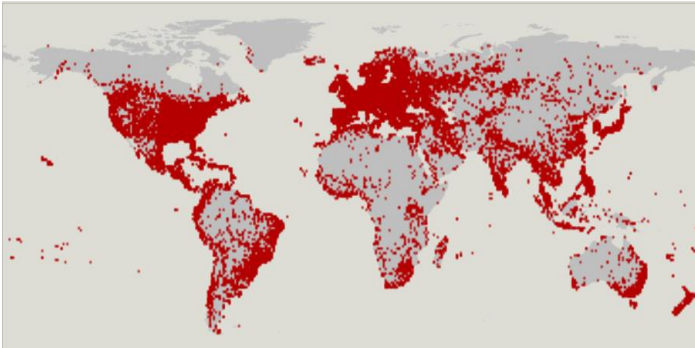
## OpenStreetMap data

great Open Source of spatial data.  
But, is OpenStreetMap cove

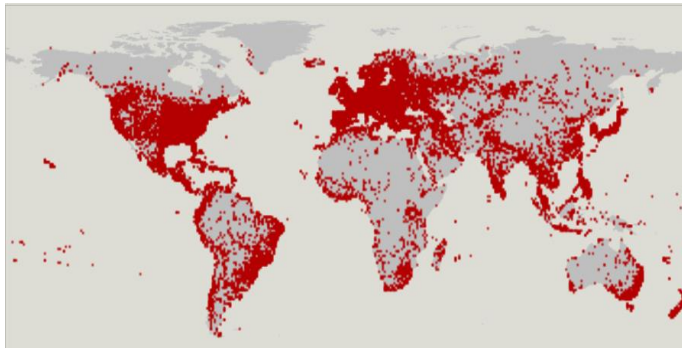
Tag/Keys	Description	Availability in Algeria
<b>name</b>	name	
<b>website</b>	website	
<b>Latitude/Longitude</b>	geolocation	
<b>cuisine</b>	type of cuisine or food	
<b>opening_hours</b>	opening hours	
<b>takeaway</b>	yes/no/only whether the place offers takeaway	
<b>delivery</b>	yes/no whether the place offers delivery	
<b>outdoor_seating</b>	whether the place has an outdoor seating area	
<b>reservation</b>	whether reservation is possible/required	
<b>addr:city</b>	address	
<b>addr:full</b>	address	
<b>addr:street</b>	address	
<b>addr:places</b>	address	
<b>operator</b>	operator	
<b>Highway</b>	key used for identifying any kind of road, street	
<b>wheelchair</b>	wheelchair accessibility	
<b>diet:vegetarian</b>	yes/no/only	
<b>diet:vegan</b>	yes/no/only	
<b>diet:halal</b>	yes/no/only	
<b>start_date</b>	start date	
<b>air_conditioning</b>	yes/no whether there is air conditioning	
<b>capacity</b>	number of seats at tables	
<b>internet_access</b>	Internet connection	
<b>payment:method</b>	methods of payment	
<b>smoking</b>	smoking / non smoking	
<b>image</b>	photo	
<b>reusable_packaging</b>		
<b>website</b>	website	
<b>phone</b>	phone number	
<b>bar</b>	the presence of a bar with alcoholic beverages	
<b>microbrewery</b>	if a microbrewery is located in the restaurant	
<b>brand</b>	brand	
<b>level</b>	floor number, if located in a multi-story building	
<b>stars</b>	restaurant with Michelin star	
<b>changing_table</b>	availability of an infant changing table	
<b>kids area</b>	area where kids can play in a children's corner	

## OSM tags and keys availability

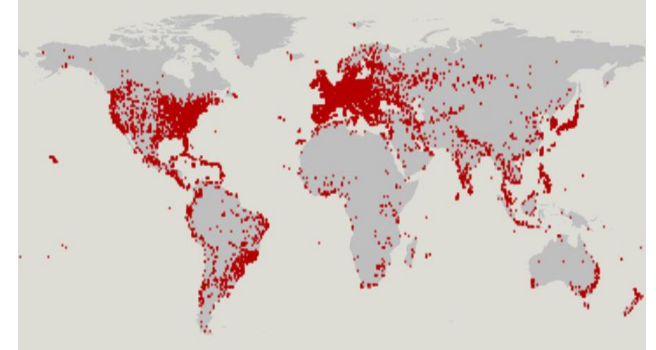
*Fast Food*



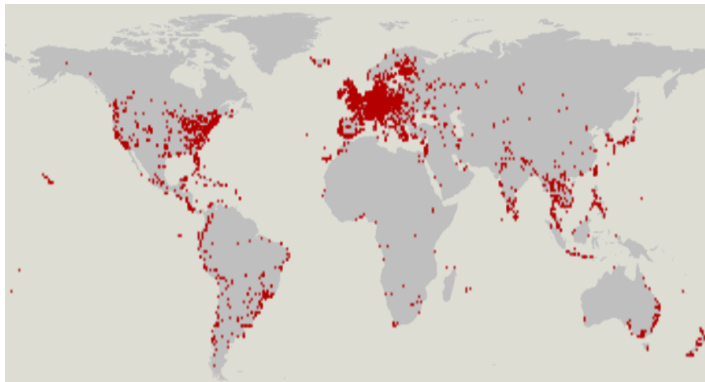
*addr:street*



*Delivery*



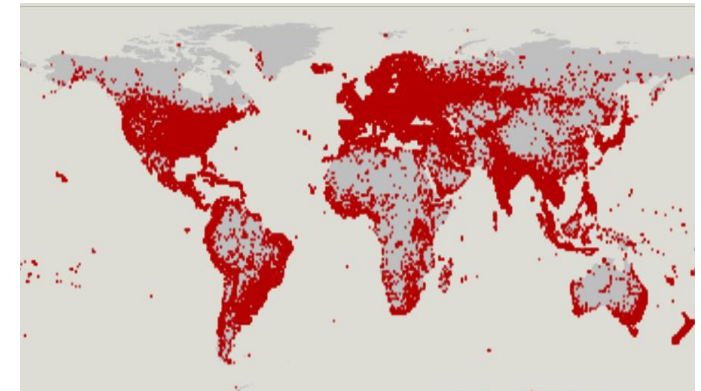
*vegetarian diets*



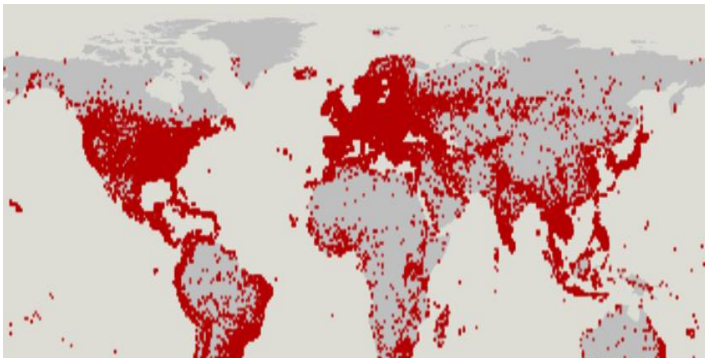
*dinner, lunch*



*Open hours*



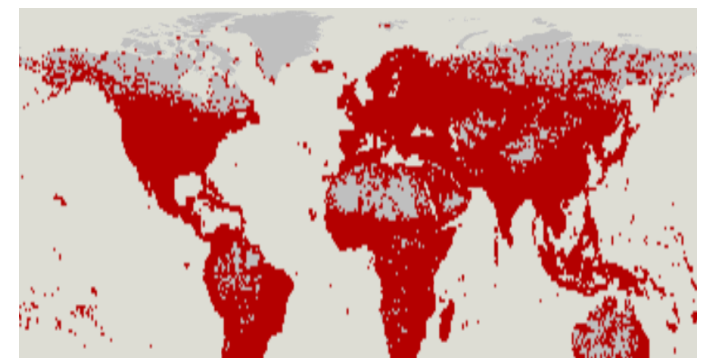
*payment:cash*



*Cuisine types*

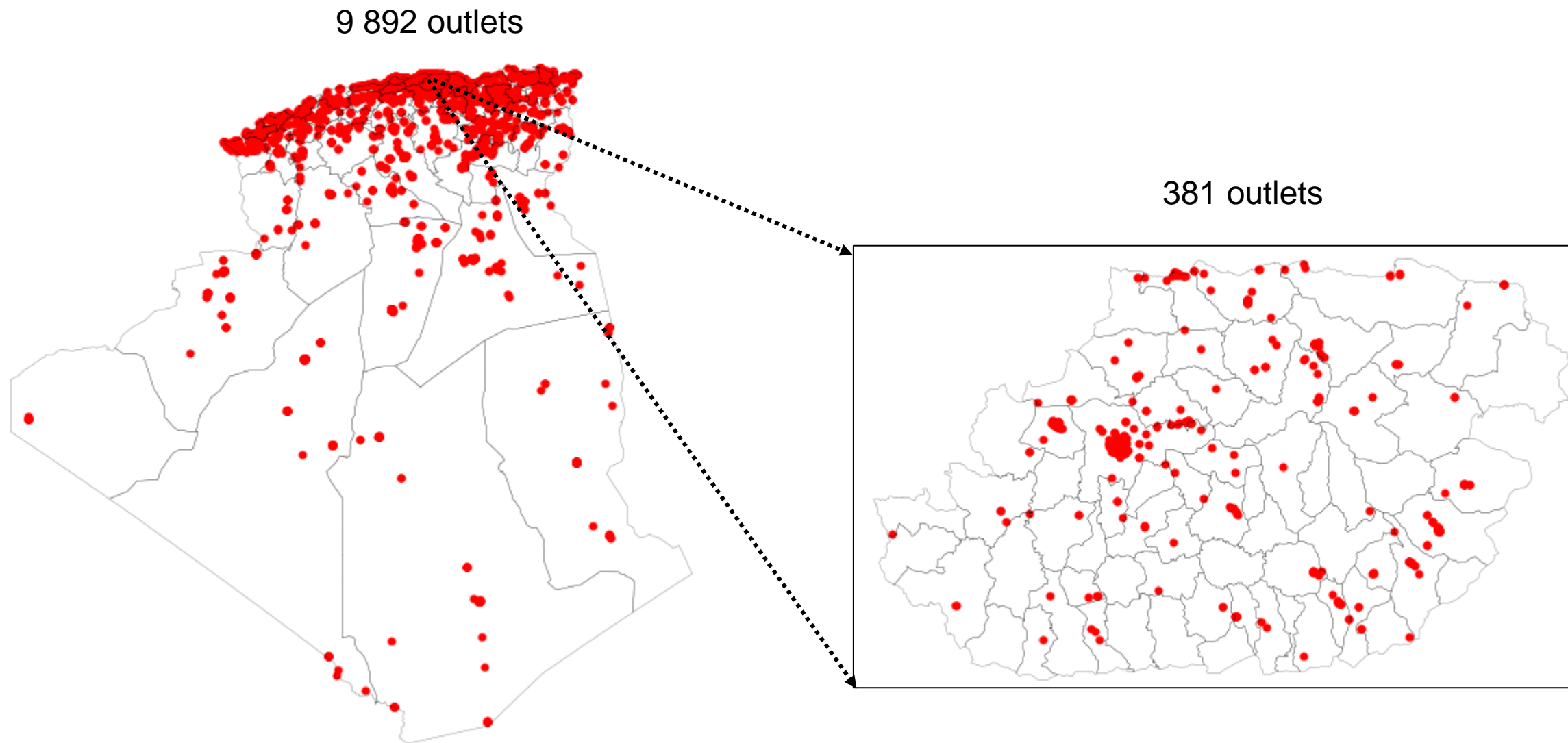


*Highway=residential*



**OSM tags : Fast food| Restaurant| Cafe| Hotel| Alcohol| bar| Nightclub| kiosk**

---



# OSM Features availability in Algeria

Features	missing value(%)
osm_id	0
latitude	0
longitude	0
amenity	0
website	99,38
name	32,64
addr_street	90,69
opening_hours	94,38
internet_access	96,19
outdoor_seating	98,98
takeaway	99,07
delivery	99,07
highway	99,98
place	99,99
Smoking	98,44
cuisine	87,85
wheelchair	99,72
capacity	99,70
email	99,37
diet_halal	99,10
payment_cash	99,94
bar	98,95
alcohol	99,58
level	99,83
air_conditioning	99,97
diet_vegetarian	99,99
payment_cards	99,99
lunch	99,98
name_fr	63,52

Possibility to extract meaningfull features but we are faced with a lot of missing values  
→OSM amenities services cannot be used for matching

## Data appeal Company Features common to TripAdvisor

Name of outlet

Address

Geolocation coordinates

Nr. of reviews

Overall rating

Price range

Outlet rankings

Types of cuisine

Meal types

Opening hours

Website

Phone nr.

Reviews lists

Type of Payment

## Additional features

Delivery

Accessibility (wheelchair)

Health & Safety messages

Sentiment score

## Features to validate

Convenience : good  
for family,



**is one of the go-to-market partners for Google Cloud**

---

**Searce uses google maps API to extract and store data in legal way.They can provide us the features below**

Name of outlet

Address

Geolocation coordinates

Nr. of reviews

Overall rating

Price range

Opening hours

Website

Phone nr.

Access to 5 Reviews



# Google maps data features

0 Geolocation

1 NAME

2 Totat Reviews

3 Rating

4 Dine in service

5 Takeaway service

6 Delivery service

7 Address

8 Open and Close Time

9 Health message

10 Facebook

11 Contact

12 Reviews data



1 La Presqu'île Restaurant  
Poissonnerie Traiteur

3 4.6 ★★★★★ 44 reviews 2  
Restaurant



Directions



Save



Nearby



Send to your  
phone



Share

✓ Dine-in · ✓ Takeaway · ✓ Delivery >

4 5 6  
3 Chem. Sidi Yahia, Bir Mourad Raïs, Algeria 7

Open now ^

Wednesday 8:30AM–11PM

Thursday 8:30AM–11PM

Friday 8:30AM–12:30PM  
2–11PM

Saturday 8:30AM–11PM 8

Sunday 8:30AM–11PM

Monday 8:30AM–11PM

Tuesday 8:30AM–11PM

[Suggest an edit](#)



Health and safety: Mask required · Staff wear masks



facebook.com 10



+213 557 85 70 17 11



P2QW+63 Bir Mourad Raïs, Algeria



Manage this listing



Identifies as women-led



Houssam Kassimi  
2 reviews 12

★★★★★ 10 months ago

Je peux rien dire à part que c un delice  
Top du top

(Translated by Google)

I can't say anything except that it's a delight  
Top of the top



# Benchmark TripAdvisor Scraped Data & Google maps API data

---



## Common data

- Geolocation
- Name
- Address
- Open and Close Time
- Total Reviews
- Rating
- Price level



## Not available on Google Maps API

- Outlet certificates
- Ratings per category
- Meal types
- Type of cuisines
- Outlets ranking

# Google API places pricing

## Basic data

Address Component  
business\_status,  
formatted\_address  
geolocation  
name  
permanently\_closed,  
Photo,  
Place id  
type  
url  
vicinity.

## Contact data

Phone Number  
International Phone Number  
Opening Hours  
Website

## Atmosphere data

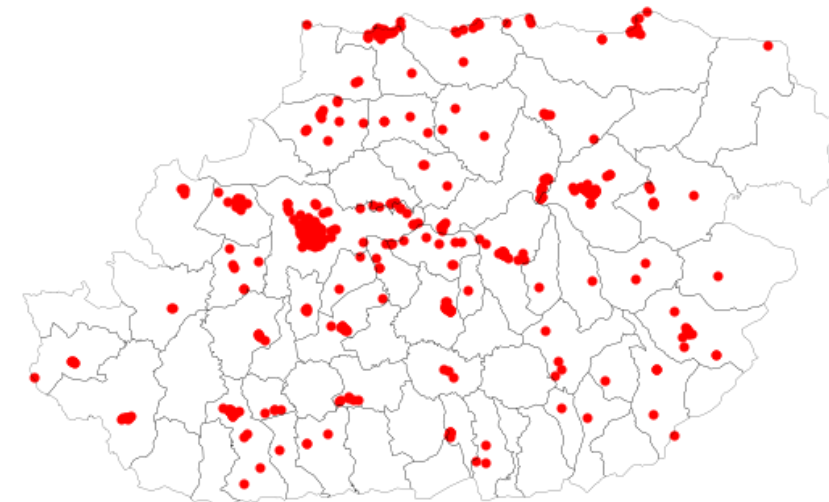
Price Level  
Rating  
User Ratings Total  
Reviews

	MONTHLY VOLUME RANGE		
	(Price per CALL)		
	0-100,000	100,001-500,000	500,000+
	place request cost	place request cost	place request cost
Basic data	0 USD	0 USD	0 USD
Contact data	0.003 USD per each	0.0024 USD per each	Contact Sales for volume pricing
Atmosphere data	0.005 USD per each	0.004 USD per each	Contact Sales for volume pricing

## Experimentation in Wilaya Tizi Ouzou

- 100 0001-500 000 calls
- 0.55 USD/outlets
- Cost : 500 USD
- 817 outlets recorded

Google maps Restaurants distribution in Wilaya Tizi Ouzou



# Ad hoc scrapping tools : Instant Data Scraper

Extracts data from web pages and exports it as Excel or CSV files.

Google chrome extension.

An automated data extraction tool for any website. It uses AI to predict which data is most relevant on a HTML



chrome web store



[Accueil](#) > [Extensions](#) > Instant Data Scraper



## Instant Data Scraper

Proposé par : [webrobots.io](https://webrobots.io)

★★★★★ 1258 | [Productivité](#) | • 100 000+ utilisateurs

Supprimer de Chrome

No geoloactions(latitude/ longitude)



10 / 13 features disponibles



With tag **restaurants/ fastfood/ snack/ pizzeria/ hotel** near algeria: 1085 outlets



# Ad hoc scraping tools : Instant Data Scraper

Extracts data from web pages and exports it as Excel or CSV files.

Google chrome extension.

An automated data extraction tool for any website. It uses AI to predict which data is most relevant on a HTML

name	sub_channel	rating	total_review	locality	services	Open	rlt_details_6	if66xd	BI0Dve 2
سناك حمد شواي	- Restaurant	1.0	(1)	29.4 km · Berrechid	Takeaway	NaN	NaN	NaN	NaN
Snack Nour	- Fast Food	5.0	(1)	7.6 km · 38 Rue d'Alger	Dine-in	- Opens 9AM	Closed	.	Kerbside pickup
Restaurant Eldjazair	- Restaurant	3.8	(26)	1,037.4 km · Algeria	Dine-in	- Opens 11AM	Closed	.	Takeaway
Restaurant Le Bearnais	- Restaurant	4.0	(47)	1,036.7 km · El Djazair, Algeria	Dine-in	NaN	NaN	.	Takeaway
Chic Chic Restaurant	- Restaurant	3.7	(170)	1,034.4 km · Algeria	Dine-in	- Opens 6:01AM	Closed	.	Kerbside pickup
...	...	...	...	...	...	...	...	...	...

No geoloactions(latitude/ longitude)



10 / 13 features disponibles



With tag **restaurants/ fastfood/ snack/ pizzeria/ hotel** near algeria: 1085 outlets



# Build our own scraper

---

Legal issue



No geoloactions(latitude/ longitude)



All available feature



Time



# Ad hoc scraping tools : Webautomations Google Maps Scraper

With geoloactions(latitude/ longitude)



All favailable features



Not free with trial period

