



# Geometry



SafeGraph's Geometry data provides POI footprints and spatial hierarchy metadata for places all over the globe. It maps the size and area of physical locations for the POIs using polygons and includes data for understanding location boundaries and relationships between POIs.

## Contents:

- [Schema](#)
- [Key Concepts](#)
- [Column Details](#)
- [Data Issues or Artifacts](#)
- [Video Overview](#)

## Geometry Schema

[geometry.csv]

POI footprints and spatial hierarchy metadata (Geometry metadata not provided for closed POIs).

Column Name	Description	Type	Example
placekey	Unique and persistent ID tied to this POI. See the <a href="#">Placekey Key Concept</a> for details on placekey design.	String	<a href="#">[email protected]</a>
parent_placekey	If place is encompassed by a larger place (e.g. mall, airport), this lists the placekey of the parent place; otherwise null . See more on parent-child relationships in <a href="#">Spatial Hierarchy</a> .	String	<a href="#">[email protected]</a>
location_name	The name of the place of interest.	String	Salinas Valley Ford Lincoln
brands	If this POI is an instance of a larger brand that we have explicitly identified, this column will contain that brand name. See more details in <a href="#">brands</a> .	List	Ford, Lincoln
latitude	<a href="#">Latitude</a> coordinate of the place of interest.	Float	36.714767
longitude	<a href="#">Longitude</a> coordinate of the place of interest.	Float	-121.662912
street_address	<a href="#">Street address</a> of the place of interest.	String	1100 Auto Center Circle
city	The <a href="#">city</a> of the point of interest.	String	Irvine
region	The state, province or county of the place of interest. See <a href="#">region</a> for more details.	String	CA
postal_code	The <a href="#">postal code</a> of the place of interest.	String	92602
iso_country_code	The 2 letter <a href="#">ISO 3166-1 alpha-2 country code</a> . Expected values are US , CA , and GB .	String	US
polygon_wkt	The shape of the place of interest, formatted as Well-Known Text (WKT). See <a href="#">polygon_wkt</a> for more details.	String	Polygon ((-121.663310 36.715207, ...,

Column Name	Description	Type	Example
			-121.663310 36.715207))
polygon_class	The classification of the polygon: 1) OWNED_POLYGON : only one POI maps to this distinct polygon. 2) SHARED_POLYGON : at least two POIs share the same polygon. See <a href="#">Shared Polygons</a> for more details.	String	OWNED_POLYGON
includes_parking_lot	Whether or not the polygon includes the parking lot. See <a href="#">includes_parking_lot</a> for more details.	Boolean	false
is_synthetic	If true then this is not a precise POI footprint polygon, but instead is an inferred polygon from an accurate centroid and category-based radius. See <a href="#">is_synthetic</a> for more details.	Boolean	false
building_height	Due to sourcing and fill rate challenges, this column is null until further notice and is excluded from new shop downloads and enterprise deliveries.	String	null
enclosed	If true, then this POI is completely enclosed indoors by its parent and is only accessible by entering the parent structure. See <a href="#">Influence on Patterns</a> for more information on visit attribution to enclosed POI.	Boolean	false
wkt_area_sq_meters	The calculated area of the polygon_wkt in square meters. Enterprise customers: Please contact your customer success manager to have this column added to your deliveries.	Float	184.363874629

## Key Concepts

### Spatial Hierarchy ( parent\_placekey )

Some POIs are characterized by a broader footprint and cannot be represented by the outline of a single building. These types of POIs often encompass smaller POIs within their borders, and we try to flag where these overlapping relationships exist in the real world by setting the `parent_placekey` of the smaller POI equal to the `placekey` of the larger, encompassing POI. We colloquially refer to the larger, containing POI as the "parent" and the smaller POI as the "child."

If a POI is not contained by an overlapping polygon, the `parent_placekey` will be `null`. Only POIs of particular categories can qualify as "parent" POIs with the exception of brands wholly containing other brands (ex: a Subway within Wal-Mart). Below is a table of all `sub_categories` and corresponding `naics_codes` that have the potential to be parents:

Schools	Medical Facilities	Large Outdoor Spaces	Places for Leisure	Other
Elementary and Secondary Schools (611110)	General Medical and Surgical Hospitals (622110)	Skiing Facilities (713920)	Sports Teams and Clubs (711211)	Other Airport Operations (488119)
Colleges, Universities, and Professional Schools (611310)	Family Planning Centers (621410)	Nature Parks and Other Similar Institutions (712190)	Amusement and Theme Parks (713110)	Correctional Institutions (922140)
Junior Colleges (611210)	All Other Outpatient Care Centers (621498)	Golf Courses and Country Clubs (713910)	Casinos (except Casino Hotels) (713210)	Hotels (except Casino Hotels)

Schools	Medical Facilities	Large Outdoor Spaces	Places for Leisure	Other
				and Motels (721110)
Other Technical and Trade Schools (611519)	Freestanding Ambulatory Surgical and Emergency Centers (621493)	Promoters of Performing Arts, Sports, and Similar Events with Facilities (711310)	Malls (531120)	Gasoline Stations with Convenience Stores (447110)
	Kidney Dialysis Centers (621492)		Casino Hotels (721120)	

In rare cases, a parent POI can have a parent. Examples include:

- Starbucks > Airport terminal > Airport
- Subway > Walmart > Shopping center
- Physician's office > Outpatient care center > Regional medical campus

## Shared Polygons

In dense environments, such as indoor malls or multi-story buildings, we may not be confident about a POI's true shape, so we provide the overall structure polygon instead. This results in several POIs sharing the same polygon. In other cases, we simply may not have a unique polygon for each POI, so several POIs end up sharing the same polygon. In each of these cases, the POIs would be classified as "SHARED\_POLYGON" in the `polygon_class` column.

If a single POI maps to a distinct polygon (excluding that POI's children), then the POI is classified as "OWNED\_POLYGON" in the `polygon_class` column. We exclude children from influencing a POI's `polygon_class` because in cases where a unique polygon is not available for a child POI, the child POI most likely maps to the parent POI's polygon; however, that does not mean the polygon is not a good representation of the parent itself.

For example, a Nike store inside of a shopping mall. If we don't have a good polygon for the Nike store, then the Nike store may share the same polygon as the mall, but the polygon for the mall is still representative of the mall's shape and size. For more details on parent-child relationships, see the Spatial Hierarchy section above.

If you need to differentiate unique stores within a shared polygon, you should use the POI centroids (the `latitude` and `longitude` columns). Since user GPS signals often drift inside of large structures, for use cases such as determining places visited by a user, we have found that user distance to centroid is a good substitute for distance to polygon.

## Influence on Patterns

Within major indoor structures, a POI's true footprint can be hard to discern, and the horizontal accuracy of GPS data deteriorates dramatically. For these reasons, we are reluctant to assign visits to `enclosed` POIs and instead roll the visits up to the parent POI ([see here](#) for more on visits to parent POIs). We have always tracked `enclosed` POIs internally, and we are externalizing this concept for complete transparency around when to expect visits at a given POI. The children of the following parent POIs are currently set to `enclosed = "TRUE:"`

- Hotels (`naics_code = 721110 OR 721120`) when the child POI is a restaurant or bar (`naics_code = 722514 OR 722515 OR 722513 OR 722511`)
- Stadium/arena (`naics_code = 713910 OR 711310`)
- Large medical facilities (`naics_code = 621498 OR 622110 OR 621410 OR 621493 OR 621492`)
- Airports and airport terminals (`naics_code = 488119`)
- Unless the child is an airport terminal POI
- Indoor shopping malls (`naics_code = 531120`). To be clear, children of open-air, outdoor shopping centers will have `enclosed = "FALSE:"`

Other than `naics_code` relationships, we track `enclosed` through known brand relationships where Brand A exists completely within Brand B. A canonical example: Brand A = Subway (`SG_BRAND_04a8ca7bf49e7ecb4a32451676e929f0`) and Brand B = Walmart (`SG_BRAND_de80593878cb1673c62a7f338dc7e4e1`). If a Walmart and Subway are co-located, the `parent_safegraph_place_id` for Subway = the Walmart's `safegraph_place_id`, and `enclosed = "TRUE"` for Subway.

## Column Name Detailed Descriptions

## latitude , longitude

- In general, latitude and longitude are defined by our best knowledge of the POI location. It is not designed to specifically locate the front door of the business, but rather defines the general center of the business.
- Latitude and longitude still attempt to identify the individual business even if that business and others have the same polygon (e.g. strip mall).

## street\_address

- We implement a number of steps to clean, validate and standardize `street_address` .
- You should expect `street_address` to be title-cased, consistent, and friendly for human reading. Please send us your feedback if you see otherwise.
- If you care about street addresses as much as we do, we also have more specific address columns to split out address components. These are optional and available upon request for future deliveries.
  - `primary_number`
  - `street_predirection`
  - `street_name`
  - `street_postdirection`
  - `street_suffix`

## city

- In the US, all centroids (latitudes/longitudes) are referenced against a geospatial file of city boundaries as defined by the US Census Bureau ([browse the boundaries here](#)). In edge cases, the preferred city name in the address line reflects a [pre-annexed](#) city name, and we try our best to preserve those city names where possible.
- In Canada, city names are the output of normalized address strings from POI sources.
- In Great Britain, city names are the output of normalized address strings from POI sources, but in edge cases, we allow POIs to have a null city name as long as `region` is populated. The `region` column in Great Britain refers to county boundaries, and counties are a decent alternative to cities for geographic filtering.

## region

- When `iso_country_code == US` , then this is the US state or territory.
- When `iso_country_code == CA` , then this is the Canadian Province or territory.
- When `iso_country_code == GB` , then this is the [United Kingdom county](#).

## postal\_code

- When `iso_country_code == US` , then this is the US 5 digit zip code.
- When `iso_country_code == CA` , then this is the [Canadian postal code](#) in the form of a 3 digit Forward Sortation Area (FSA), a space, and the 3 digit Local Delivery Unit (LDU).
- When `iso_country_code == GB` , then this is the British postal code. [Learn more about Great Britain postal code precision here](#).

## polygon\_wkt

- Spatial reference used: [EPSG:4326](#)
- WKT stands for [Well-Known-Text](#). It's a simple way to define a polygon/shape and is the standard format for polygons in SafeGraph Places.
- Other geospatial file formats you may utilize include Shapefile and GeoJSON. WKT can easily be converted to these formats and file conversions are available by request.

## includes\_parking\_lot

- In some cases, our polygons intentionally include the parking lot ([POIs with parent qualifying naics codes](#) and car dealerships). The value of the `includes_parking_lot` column is to make explicit to our customers when the `polygon_wkt` does or does not include the parking lot. There are three possible values `true` , `false` , and `null` (null when we are not sure whether a parking lot is included in the geometry).

## is\_synthetic

- We strive for precise polygons for nearly all of our places, but in some cases, we have not yet sourced an accurate polygon and will instead infer a synthetic polygon from an accurate centroid, category-based radius, and heuristics like avoiding overlap with roads. In these cases, `is_synthetic == "true."` For some categories, it does not make sense to provide a precise polygon. Those categories are listed below:
  - Cemeteries and Crematories (812220)