



# **Preparación de contenidos docentes a partir de Big Data geolocalizado**

---

**Joaquín Osorio Arjona – Universidad Nacional de Educación a Distancia, Departamento de Geografía.**

# Contenidos del curso

---

- ❖ Objetivo del curso: crear y subir material docente a partir de nuevas fuentes de datos.
- ❖ Caso práctico: uso de datos de telefonía móvil para crear mapa web para capítulo de Demografía de la asignatura de Geografía Humana.
- ❖ Análisis demográfico de España durante un día de invierno y durante un día de verano (los datos de población que solemos tener son la población en residencias de noche, ¿pero como está distribuida la población de día durante su actividad laboral?).

# Contenidos del curso

---

## TALLER 1

- Introducción al Big Data geolocalizado
- Ejemplos de usos de nueva fuentes de datos geolocalizados
- Fuentes de datos
- Descarga y almacenamiento de datos
- Procesado de datos

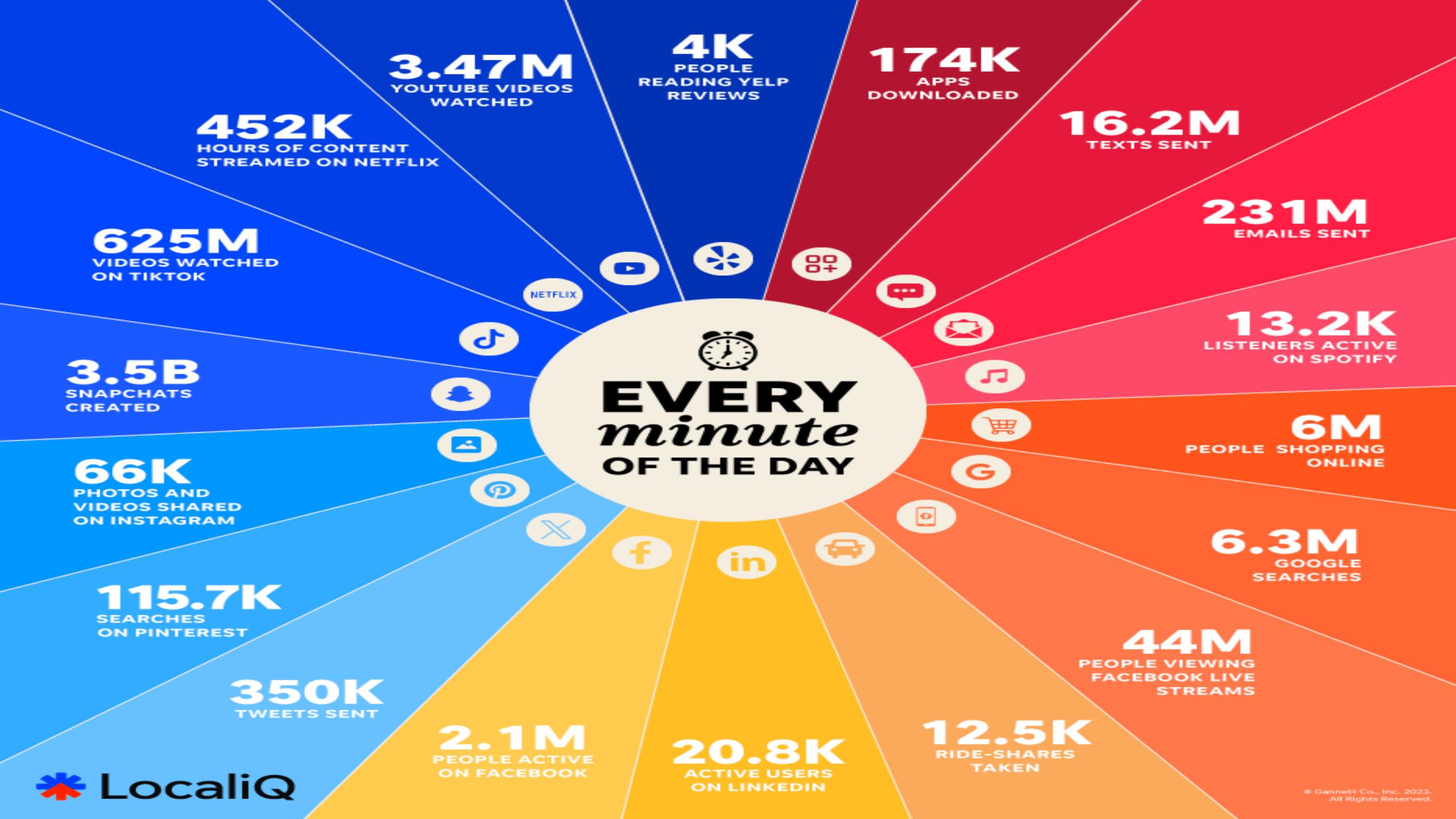
## TALLER 2

- Procesado de datos
- Creación de contenido docente
- Cartografía de material docente

# Introducción al Big Data geolocalizado

---

- ❖ Las actividades humanas en internet dejan una huella digital en el espacio y el tiempo.
- ❖ Producción de datos masivos a un ritmo antes insospechado.
- ❖ Una gran parte de estos datos están georreferenciados y pueden ser procesados en un SIG.



# De los Productos a los servicios

Servicios



Cambio a formato Digital



música



Productos

1920

1930

1940

1950

1960

1970

1980

1990

2000

2010

# Evolución de las Tecnologías IT

1

1780 - Primera Revolución Industrial



Producción mecánica con energía hidráulica y vapor

Maquinas [Trabajo]

2

1870 - Segunda Revolución Industrial



División de trabajo, producción en masa y electricidad

Energía [Producción]

3

1970 - Tercera Revolución Industrial



Sistemas electrónicos y sistemas IT

Sistemas [Conocimiento]

4

Hoy - Cuarta Revolución Industrial



Sistemas ciberfísicos con las tecnologías 4.0

Comunicación [Relaciones]

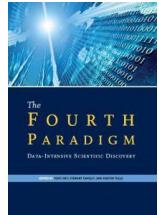


Imprenta



Internet

Información

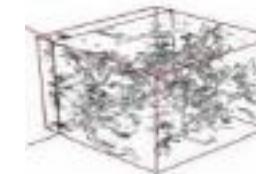


## El poder de los datos en la investigación: *el cuarto paradigma* (J. Gray, 2007)

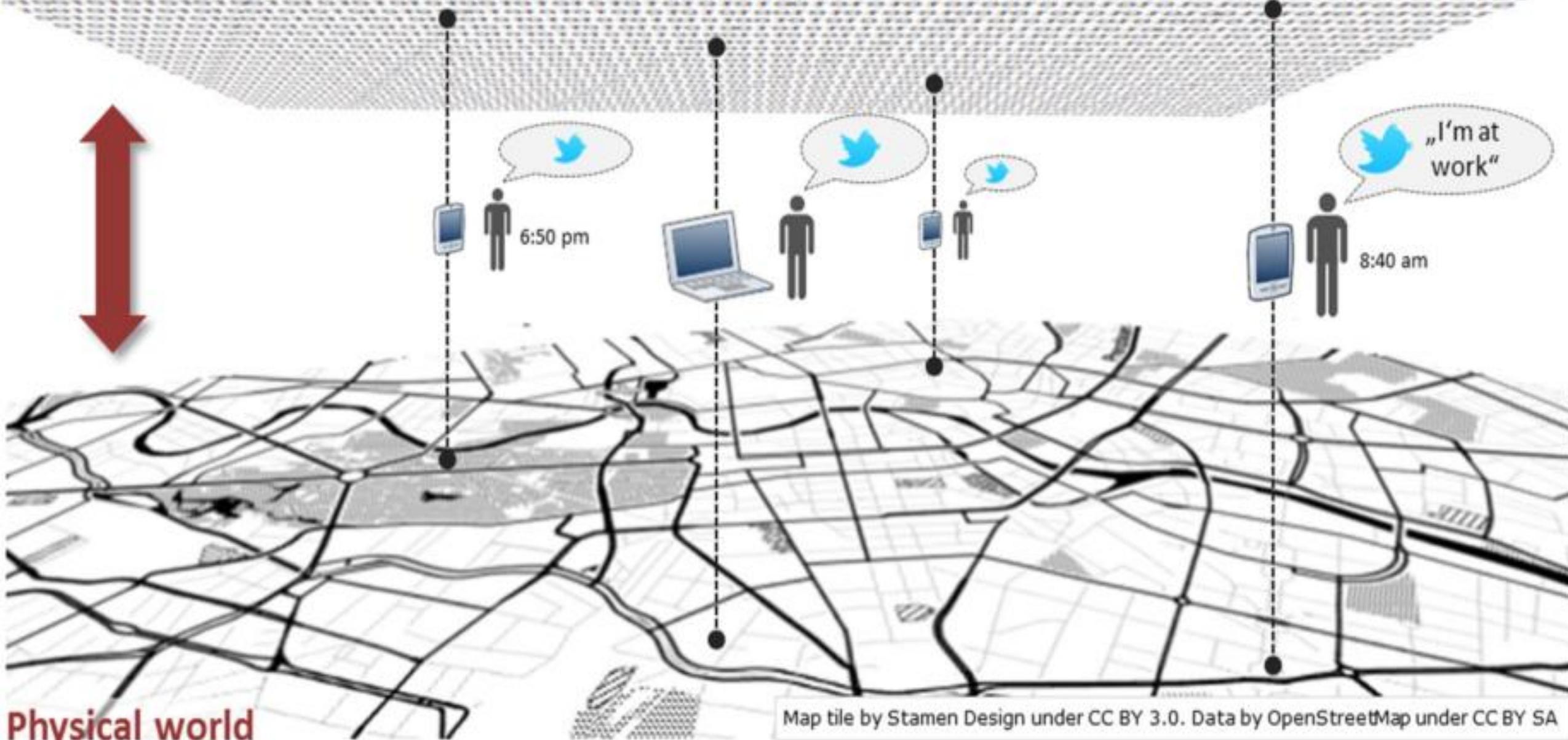
- Hace mil años, la **ciencia era empírica**: describía los fenómenos naturales.
- Los últimos siglos, pasó a ser una **ciencia teórica**: usando modelos, haciendo generalizaciones
- En las últimas décadas, llegó la **era computacional**: simulación de complejos fenómenos
- Hoy, el cuarto paradigma se apoya en la **exploración de datos**:
  - Científicos abrumados con conjuntos de datos de muchas fuentes diferentes
  - eScience es el conjunto de herramientas y tecnologías para apoyar una "*ciencia intensiva en datos*":
    - unificando la teoría, experimentación y simulación
    - apoyada en manejo de datos y análisis estadístico
    - peso de tareas de almacenaje, visualización, colaboración...



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K \frac{c^2}{a^2}$$



# Virtual world



# Physical world

JAN  
2024

# SPAIN

## OVERVIEW OF THE ADOPTION AND USE OF CONNECTED DEVICES AND SERVICES

**NOTE:** SIGNIFICANT REVISIONS TO SOURCE DATA MEAN THAT FIGURES SHOWN HERE ARE **NOT COMPARABLE** WITH PREVIOUS REPORTS. SEE THE IMPORTANT NOTES AT THE START OF THIS REPORT FOR DETAILS.



SPAIN

TOTAL  
POPULATION



we  
are  
social

**47.50**  
MILLION

YEAR-ON-YEAR CHANGE

**-0.09%**  
**-43 THOUSAND**

URBANISATION

**81.7%**

CELLULAR MOBILE  
CONNECTIONS



Meltwater

**60.44**  
MILLION

YEAR-ON-YEAR CHANGE

**+3.2%**  
**+1.9 MILLION**

TOTAL vs. POPULATION

**127.2%**

INDIVIDUALS USING  
THE INTERNET



Meltwater

**45.58**  
MILLION

YEAR-ON-YEAR CHANGE

**-0.09%**  
**-41 THOUSAND**

TOTAL vs. POPULATION

**96.0%**

SOCIAL MEDIA  
USER IDENTITIES



**39.70**  
MILLION

YEAR-ON-YEAR CHANGE

**-2.5%**  
**-1.0 MILLION**

TOTAL vs. POPULATION

**83.6%**

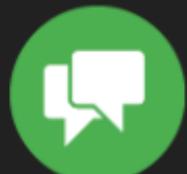
**SOURCES:** U.N.; GOVERNMENT AUTHORITIES; GSMA INTELLIGENCE; ITU; EUROSTAT; CNNIC; KANTAR & IMAI; PLATFORM RESOURCES; OCDH; BETA RESEARCH CENTER; KEPiOS ANALYSIS. **ADVISORY:** SOCIAL MEDIA USER IDENTITIES MAY NOT REPRESENT UNIQUE INDIVIDUALS. **COMPARABILITY:** SOURCE CHANGES AND BASE REVISIONS. FIGURES ARE **NOT COMPARABLE** WITH PREVIOUS REPORTS. GLOBAL DATASETS MAY USE DIFFERENT SOURCES vs. COUNTRY AND REGIONAL DATA, SO SUMS MAY NOT MATCH. **IMPORTANT:** NEGATIVE VALUES MAY INDICATE SOURCE DATA CORRECTIONS, AND MAY NOT REPRESENT DECREASES IN THE RELEVANT METRIC. WHERE YEAR-ON-YEAR CHANGE IS "[N/A]", COMPARISONS WITH HISTORICAL DATA WILL PRODUCE INACCURATE RESULTS. PLEASE SEE [NOTES ON DATA](#).

## MEDIA USE

THE PERCENTAGE OF INTERNET USERS AGED 16 TO 64 WHO CONSUME EACH MEDIA TYPE

USING THE INTERNET  
VIA A MOBILE PHONE

GWI.

**98.9%**YEAR-ON-YEAR CHANGE  
**+0.6% (+55 BPS)**USING THE INTERNET VIA A  
LAPTOP, DESKTOP, OR TABLETwe  
are  
social**94.5%**YEAR-ON-YEAR CHANGE  
**-0.9% (-89 BPS)**USING  
SOCIAL MEDIA

GWI.

**95.4%**YEAR-ON-YEAR CHANGE  
**+0.6% (+55 BPS)**WATCHING LINEAR  
AND BROADCAST TV**94.4%**YEAR-ON-YEAR CHANGE  
**-0.3% (-26 BPS)**WATCHING STREAMING  
AND ON-DEMAND TV**78.0%**YEAR-ON-YEAR CHANGE  
**-1.5% (-118 BPS)**READING ONLINE  
PRESS CONTENT

Meltwater

**81.2%**YEAR-ON-YEAR CHANGE  
**-2.7% (-226 BPS)**READING PHYSICAL  
PRESS CONTENT

GWI.

**60.2%**YEAR-ON-YEAR CHANGE  
**-3.9% (-246 BPS)**LISTENING TO  
BROADCAST RADIO

KEPIOS

**77.6%**YEAR-ON-YEAR CHANGE  
**-2.9% (-228 BPS)**LISTENING TO MUSIC  
STREAMING SERVICES

GWI.

**64.7%**YEAR-ON-YEAR CHANGE  
**+2.1% (+135 BPS)**LISTENING  
TO PODCASTS**63.4%**YEAR-ON-YEAR CHANGE  
**+0.9% (+54 BPS)**

**SOURCE:** GWI (Q3 2023). FIGURES REPRESENT THE FINDINGS OF A BROAD SURVEY OF INTERNET USERS AGED 16 TO 64. SEE [GWI.COM](#). **NOTES:** PERCENTAGE CHANGE VALUES REPRESENT RELATIVE CHANGE (I.E. AN INCREASE OF 20% FROM A STARTING VALUE OF 50% WOULD EQUAL 60%, NOT 70%). "BPS" VALUES REPRESENT BASIS POINTS, AND INDICATE ABSOLUTE CHANGE. VALUE FOR BROADCAST RADIO DOES NOT INCLUDE INTERNET RADIO. **COMPARABILITY:** METHODOLOGY CHANGES. SEE [NOTES ON DATA](#). **POTENTIAL MISMATCHES:** THE VALUES SHOWN HERE ARE BASED ON THE TIME THAT PEOPLE SAY THEY SPEND CONSUMING EACH MEDIA TYPE, AND MAY DIFFER FROM SIMILAR DATA POINTS SHOWN ELSEWHERE IN THIS REPORT THAT REFLECT OTHER BEHAVIORS SUCH AS DEVICE USAGE.

JAN  
2024

# DEVICES USED TO ACCESS THE INTERNET

PERCENTAGE OF INTERNET USERS AGED 16 TO 64 WHO USE EACH KIND OF DEVICE TO ACCESS THE INTERNET



MOBILE  
PHONE (ANY)



**96.1%**

YEAR-ON-YEAR CHANGE  
**+3.7% (+340 BPS)**

GWI.

LAPTOP OR  
DESKTOP (ANY)



**77.1%**

YEAR-ON-YEAR CHANGE  
**-1.3% (-100 BPS)**



SMART  
PHONE

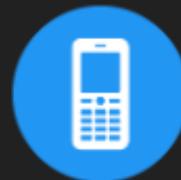


**94.1%**

YEAR-ON-YEAR CHANGE  
**+3.7% (+340 BPS)**

GWI.

FEATURE  
PHONE



**3.5%**

YEAR-ON-YEAR CHANGE  
**-18.6% (-80 BPS)**



TABLET  
DEVICE



**40.1%**

YEAR-ON-YEAR CHANGE  
**-0.5% (-20 BPS)**



PERSONAL LAPTOP  
OR DESKTOP



**68.1%**

YEAR-ON-YEAR CHANGE  
**-4.1% (-290 BPS)**

WORK LAPTOP  
OR DESKTOP



GWI.

**28.9%**

YEAR-ON-YEAR CHANGE  
**+1.0% (+30 BPS)**

CONNECTED  
TELEVISION



SMART HOME  
DEVICE



GWI.

**23.4%**

YEAR-ON-YEAR CHANGE  
**+8.8% (+190 BPS)**

GAMES  
CONSOLE



**16.7%**

YEAR-ON-YEAR CHANGE  
**+1.2% (+20 BPS)**



SOURCE: GWI (Q3 2023). FIGURES REPRESENT THE FINDINGS OF A BROAD SURVEY OF INTERNET USERS AGED 16 TO 64. SEE [GWI.COM](https://GWI.COM). NOTES: "MOBILE PHONE (ANY)" INCLUDES USERS WHO ACCESS VIA A SMARTPHONE OR A FEATURE PHONE. "LAPTOP OR DESKTOP (ANY)" INCLUDES USERS WHO ACCESS VIA THEIR OWN COMPUTER OR A COMPUTER PROVIDED BY THEIR EMPLOYER. PERCENTAGE CHANGE VALUES REFLECT RELATIVE CHANGE. "BPS" VALUES SHOW THE CHANGE IN BASIS POINTS, AND REFLECT ABSOLUTE CHANGE. COMPARABILITY: METHODOLOGY CHANGES. SEE [NOTES ON DATA](#).

JAN  
2024

# FAVOURITE SOCIAL MEDIA PLATFORMS

PERCENTAGE OF ACTIVE SOCIAL MEDIA USERS AGED 16 TO 64 WHO SAY THAT EACH OPTION IS THEIR "FAVOURITE" SOCIAL MEDIA PLATFORM



SPAIN

WHATSAPP

39.1%

DATAREPORTAL GWI.

INSTAGRAM

24.3%

TIKTOK

9.6%

FACEBOOK

9.5%

X (TWITTER)

5.5%

TELEGRAM

2.9%

PINTEREST

1.8%

LINKEDIN

1.1%

DISCORD

0.7%

FACEBOOK MESSENGER

0.6%

# Introducción al Big Data geolocalizado

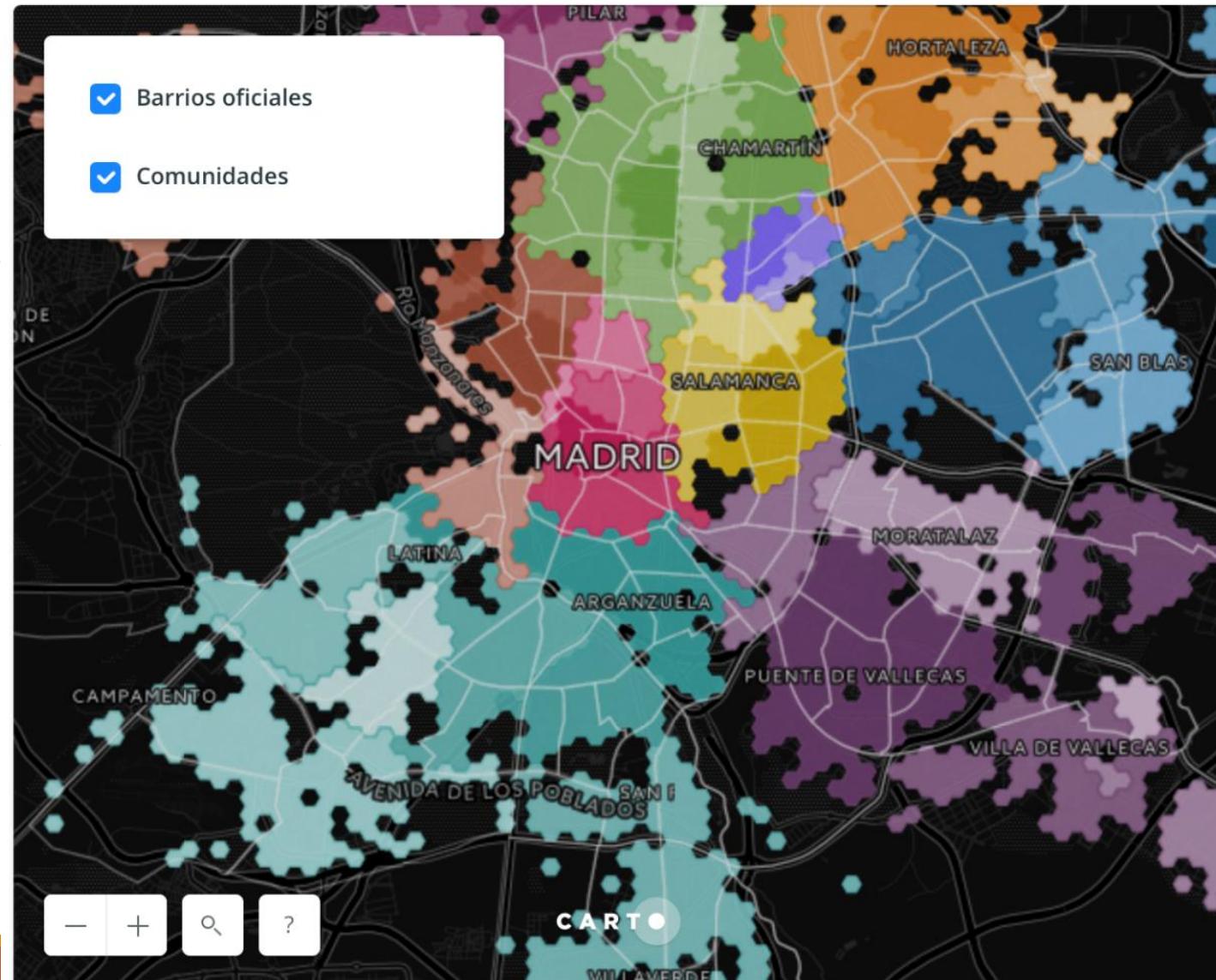
Así cambian las tarjetas de crédito los mapas de las ciudades

Por Montse Hidalgo Pérez

BBVA reinventa los mapas de Madrid, Barcelona y México con el análisis 413 millones de transacciones

26 SEP 2017 - 07:43 CEST

C aducaron los mapas. Las divisiones administrativas -comarcas, distritos, barrios, códigos postales...- quedan obsoletas cuando entra en juego la actividad de sus habitantes. **Esta es la realidad** que se destapa cuando se pintan sobre un mapa **413 millones de transacciones** con tarjeta, concretamente las realizadas en Madrid, Barcelona y México en el curso de 2016.



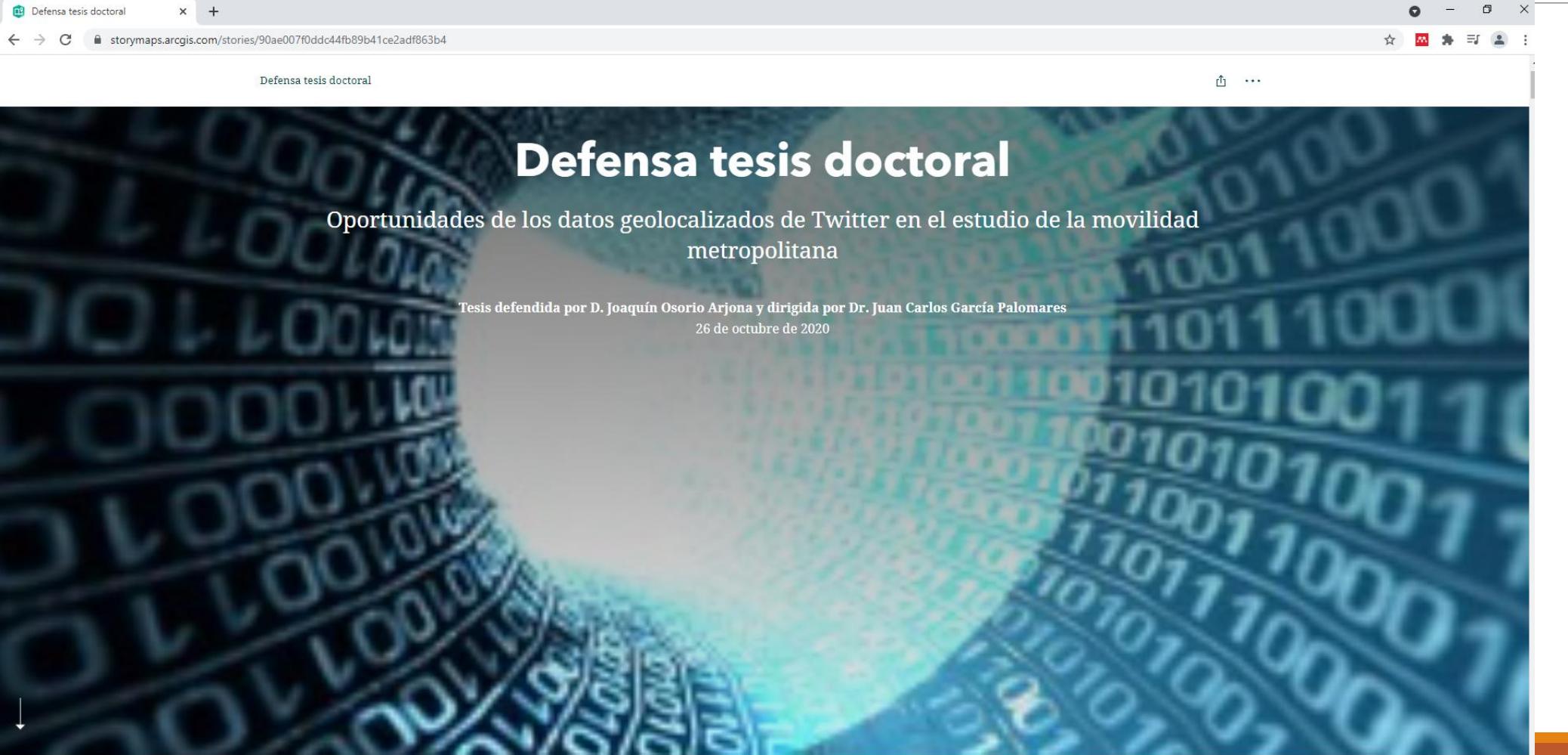
- Actividad
  - Redes sociales: Twitter, Facebook, Foursquare, etc.
  - Tarjetas de crédito: API BBVA, Travel club, tarjetas de fidelización
  - Llamadas/sms/conexiones a internet: SmartSteps



Fuente: (Moro, 2016).

# Ejemplos de uso de fuentes de datos

<https://storymaps.arcgis.com/stories/90ae007f0ddc44fb89b41ce2adf863b4>

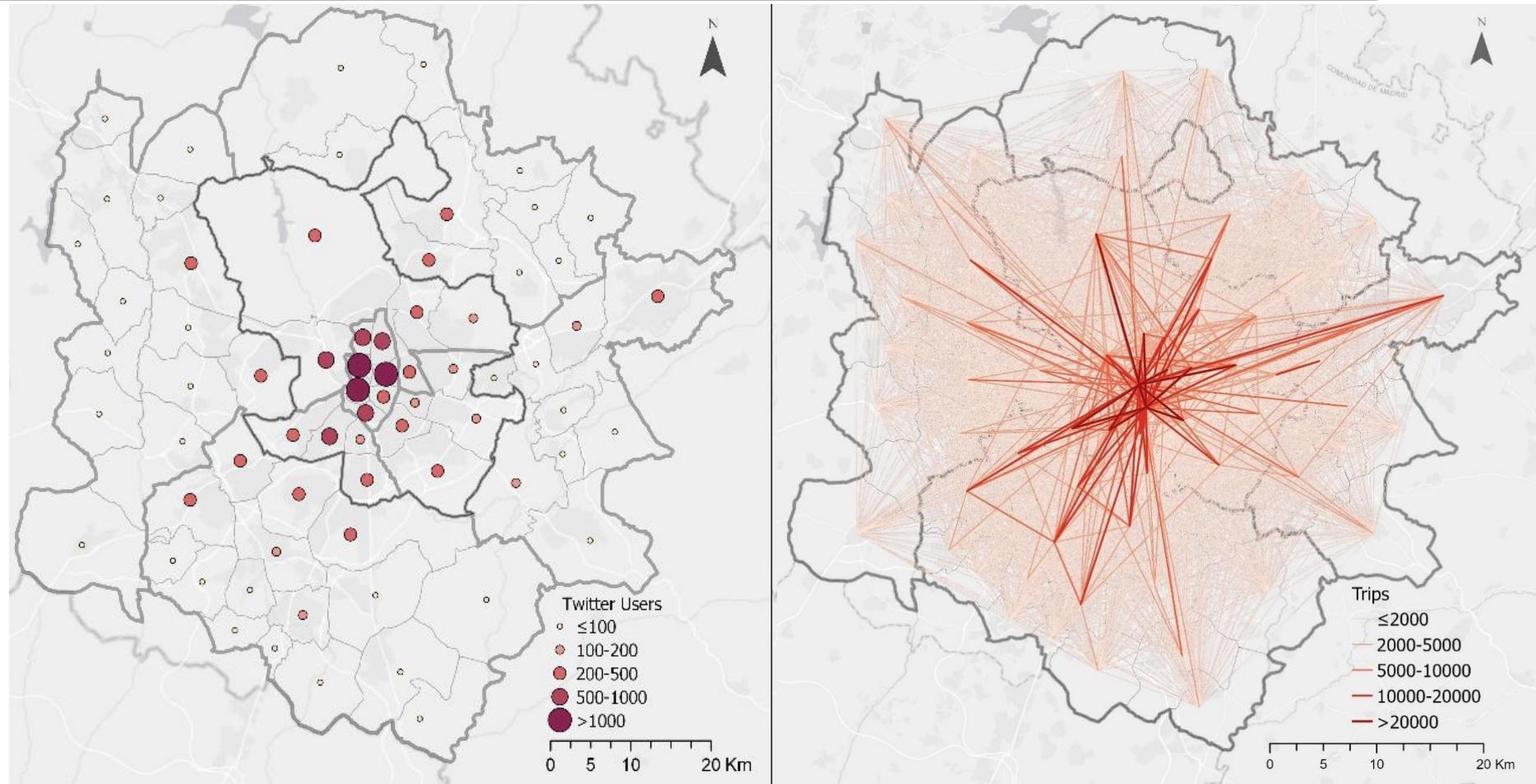


The screenshot shows a StoryMap titled "Defensa tesis doctoral" displayed in a web browser. The title is prominently displayed in white text against a background of blurred binary code. Below the title, the subtitle reads "Oportunidades de los datos geolocalizados de Twitter en el estudio de la movilidad metropolitana". Further down, it states "Tesis defendida por D. Joaquín Osorio Arjona y dirigida por Dr. Juan Carlos García Palomares" and the date "26 de octubre de 2020". At the bottom of the page, there is a navigation bar with links to "INTRODUCCIÓN", "MARCO TEÓRICO", "ÁREA ESTUDIO, DATOS Y METODOLOGÍA", "CASOS DE ESTUDIO", and "CONCLUSIONES".

# Ejemplos de uso de fuentes de datos

- Twitter como fuente de datos para estudiar la distribución de la población y la movilidad de una ciudad, región o país.
- Construcción de matrices OD.

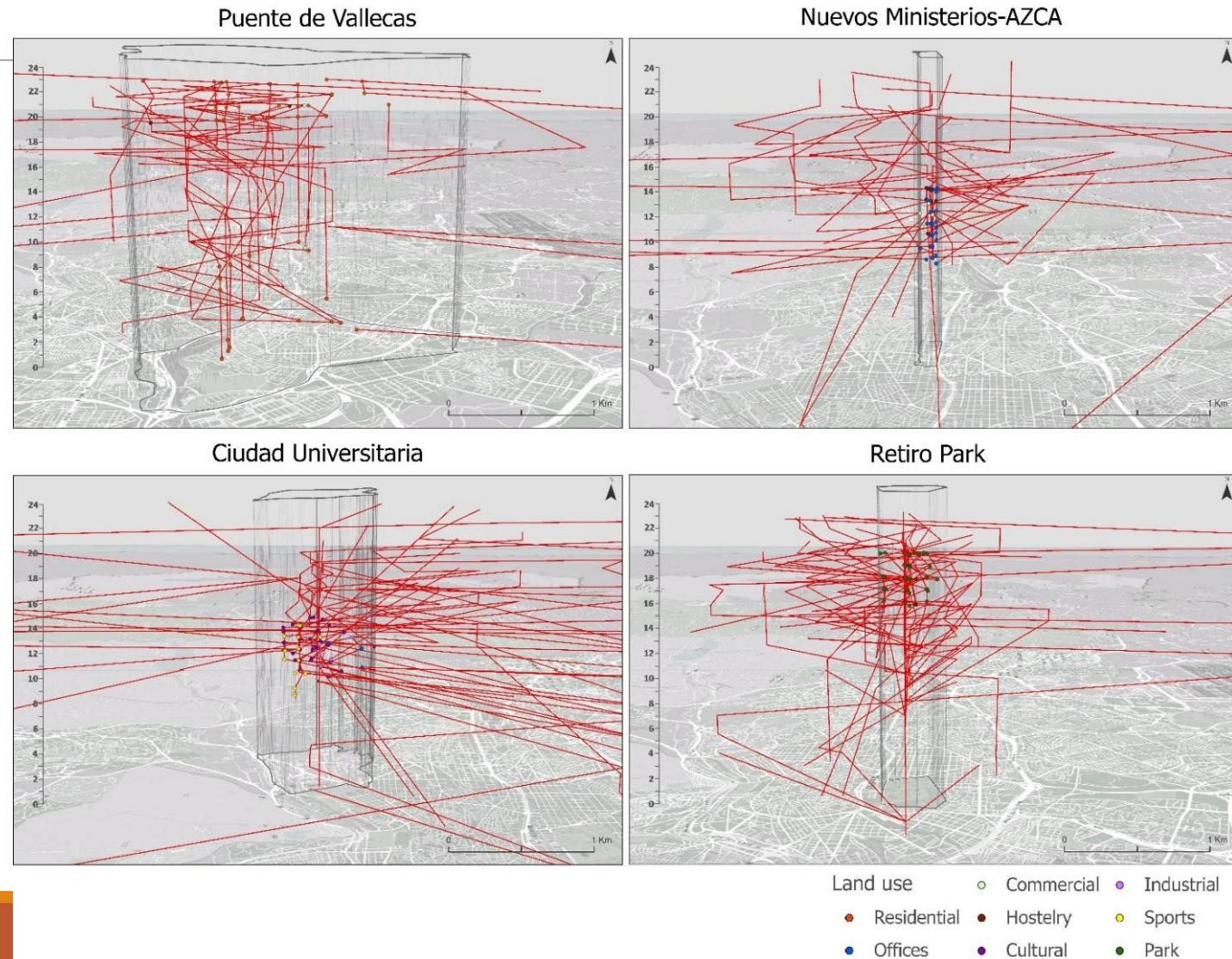
Osorio-Arjona, J. y García-Palomares J.C. (2019). Social media and urban mobility: Using Twitter to calculate home-work travel matrices. *Cities*, 89 (June 2019), 268-280 (DOI: 10.1016/j.cities.2019.03.006)



# Ejemplos de uso de fuentes de datos

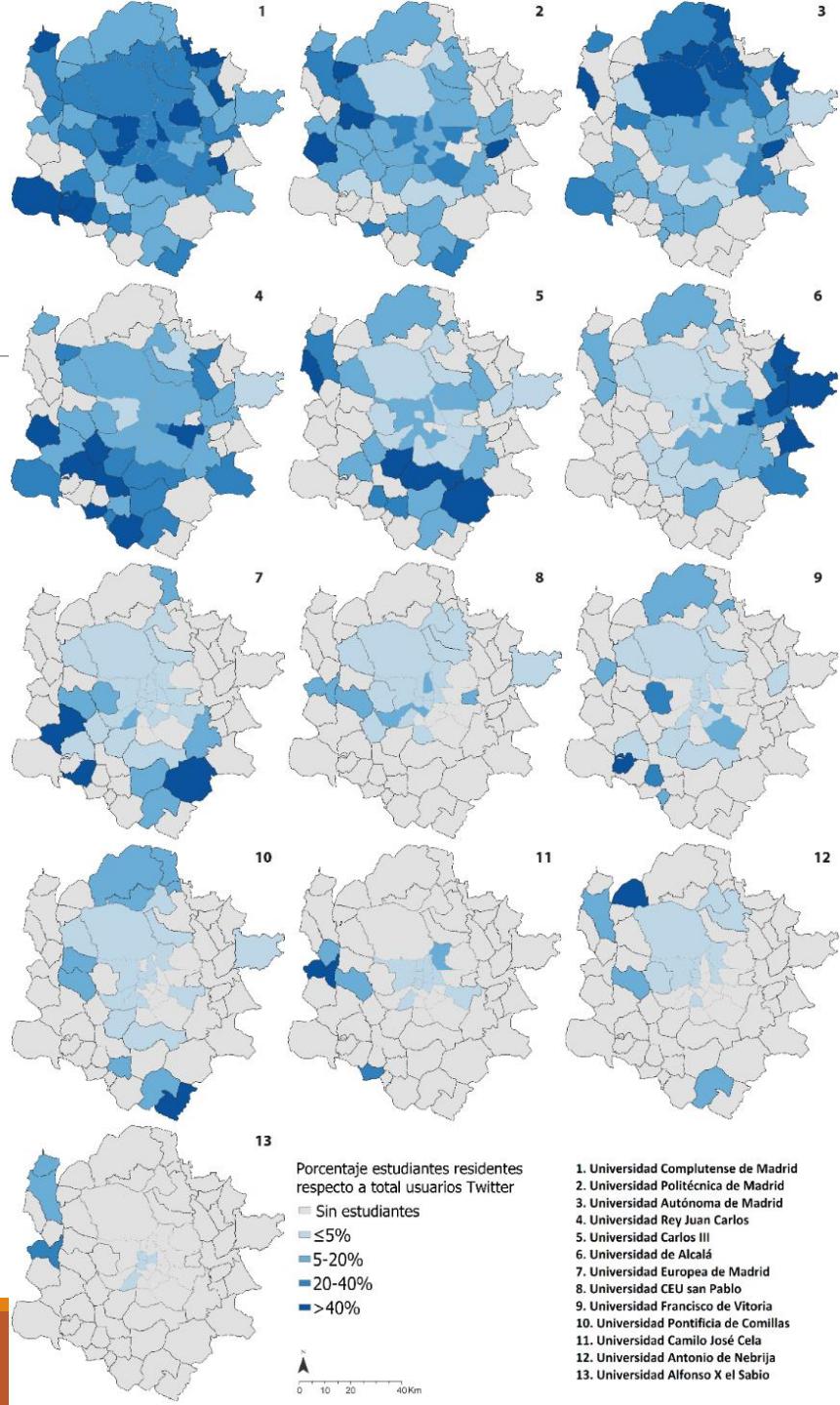
- Twitter como fuente de datos para visualizar como se mueve la población durante el día y como afectan a la movilidad los comportamientos de diferentes zonas de la ciudad
- Construcción de caminos espacio-temporales.

Osorio-Arjona, J. y García-Palomares J.C. (2020). Spatio-Temporal mobility and Twitter: 3D visualization of mobility flows. *Journal of Maps*, 16(1), 153-160 (DOI: 10.1080/17445647.2020.1778549).



# Ejemplos de uso de fuentes de datos

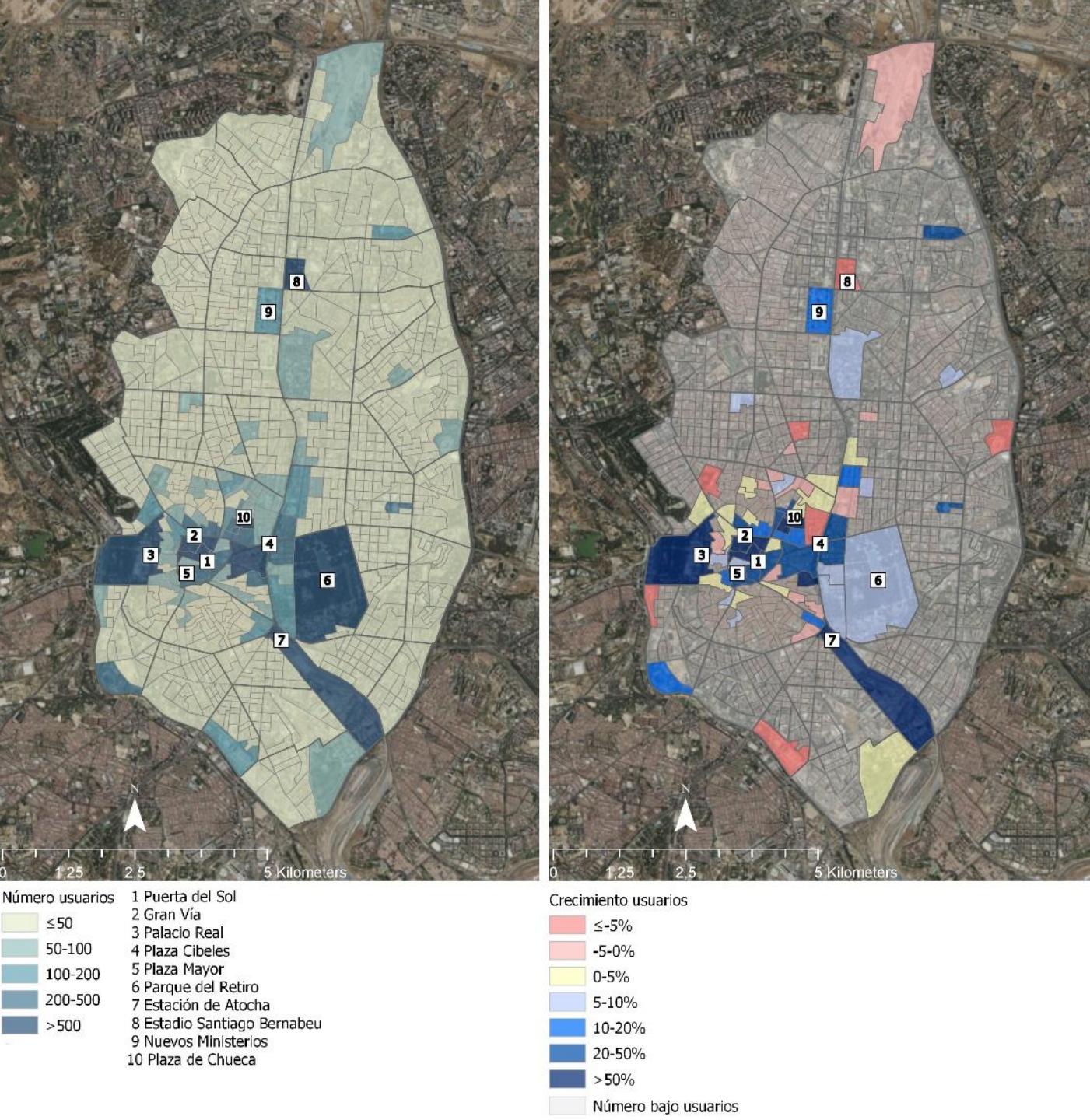
- Twitter como fuente de datos para estudiar la distribución espacial y la movilidad de determinados grupos sociales (como los estudiantes universitarios)
- Visualización de áreas de influencia de las universidades en función del número de usuarios que atraen y donde residen los estudiantes.



Osorio Arjona, J., and García Palomares, J.C. (2019). Big Data y universidades: análisis de movilidad de los estudiantes universitarios a partir de datos de Twitter. Geofocus, 24, 37-57. ISSN: 1578-5157 (DOI: 10.21138/GF.648).

# Ejemplos de uso de fuentes de datos

- Twitter como fuente de datos para estudiar el comportamiento espacial de una ciudad o país durante la celebración de un macroevento o festival.
- Visualización del crecimiento de la actividad en determinadas áreas en el espacio y tiempo durante un festival, o visualización de los países o regiones de procedencia de los turistas.

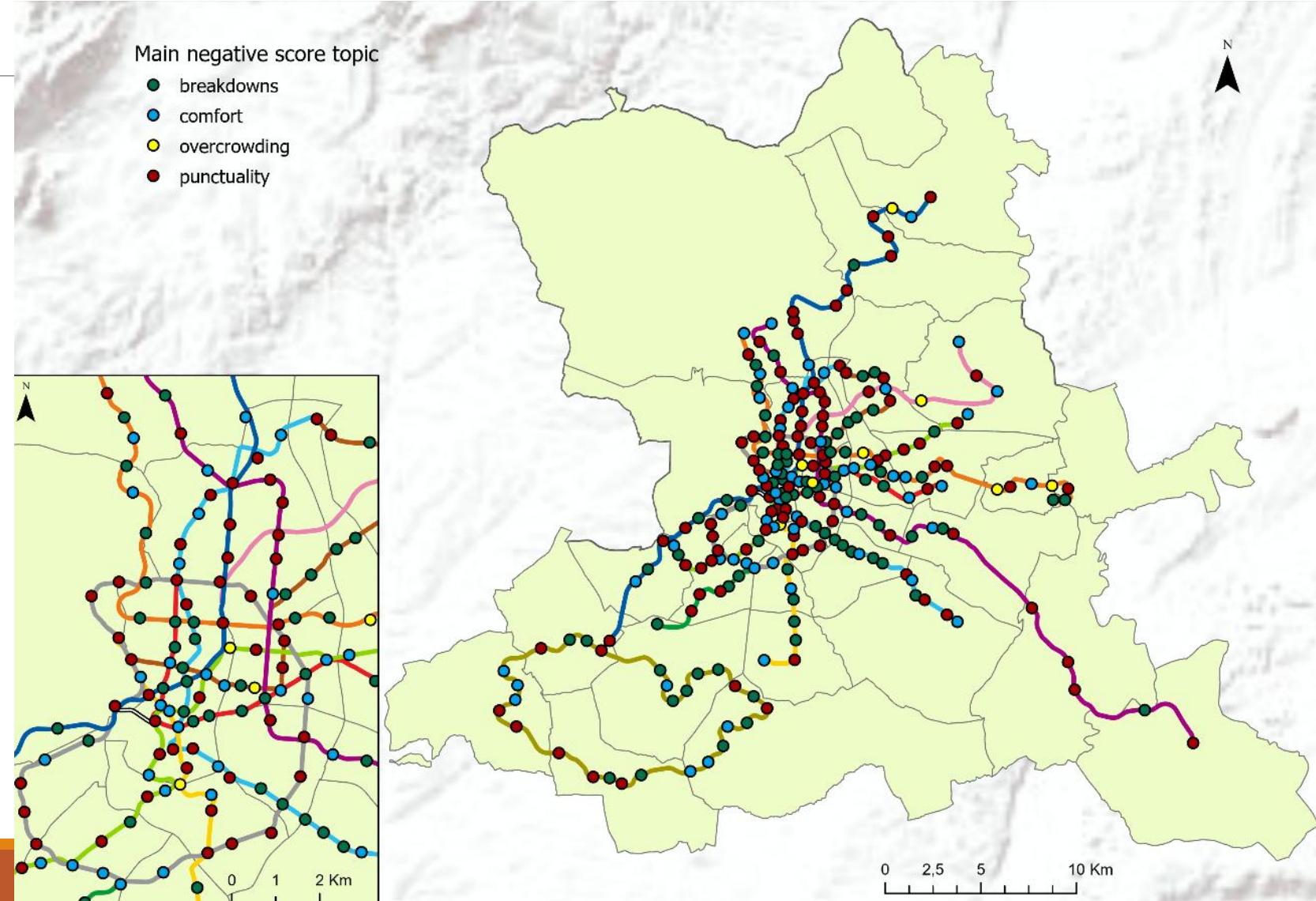


Osorio-Arjona, J. (2020). Análisis de los patrones espacio-temporales de eventos a partir de datos de Twitter: el caso de la World Pride 2017 en Madrid. *Estudios Geográficos*, 81 (288). ISSN: 0014-1496 (DOI:10.3989/estgeogr.202047.027).

# Ejemplos de uso de fuentes de datos

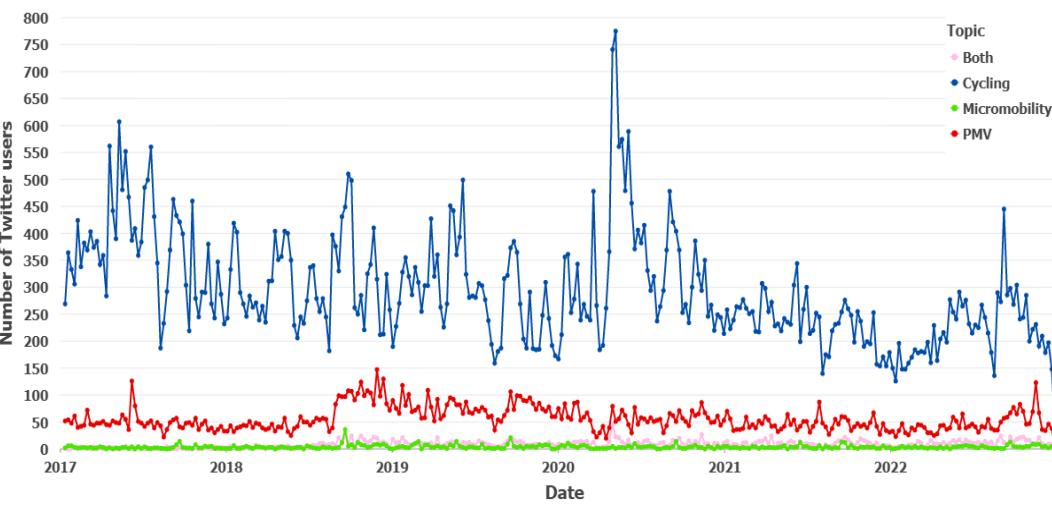
- Twitter como fuente de datos para analizar el comportamiento de infraestructuras vinculadas con el espacio y la movilidad.
- Análisis de los textos de los tweets para la obtención de opiniones relacionadas con un servicio y visualización de la distribución espacial de estas opiniones.

Osorio-Arjona, J., Horak, J., Svoboda, R., García-Ruiz, Y. (2021). Social media semantic perceptions on Madrid Metro system: using Twitter data to link complaints to space. *Sustainable Cities and Society*, 64 (102530) (DOI: 10.1016/j.scs.2020.102530).

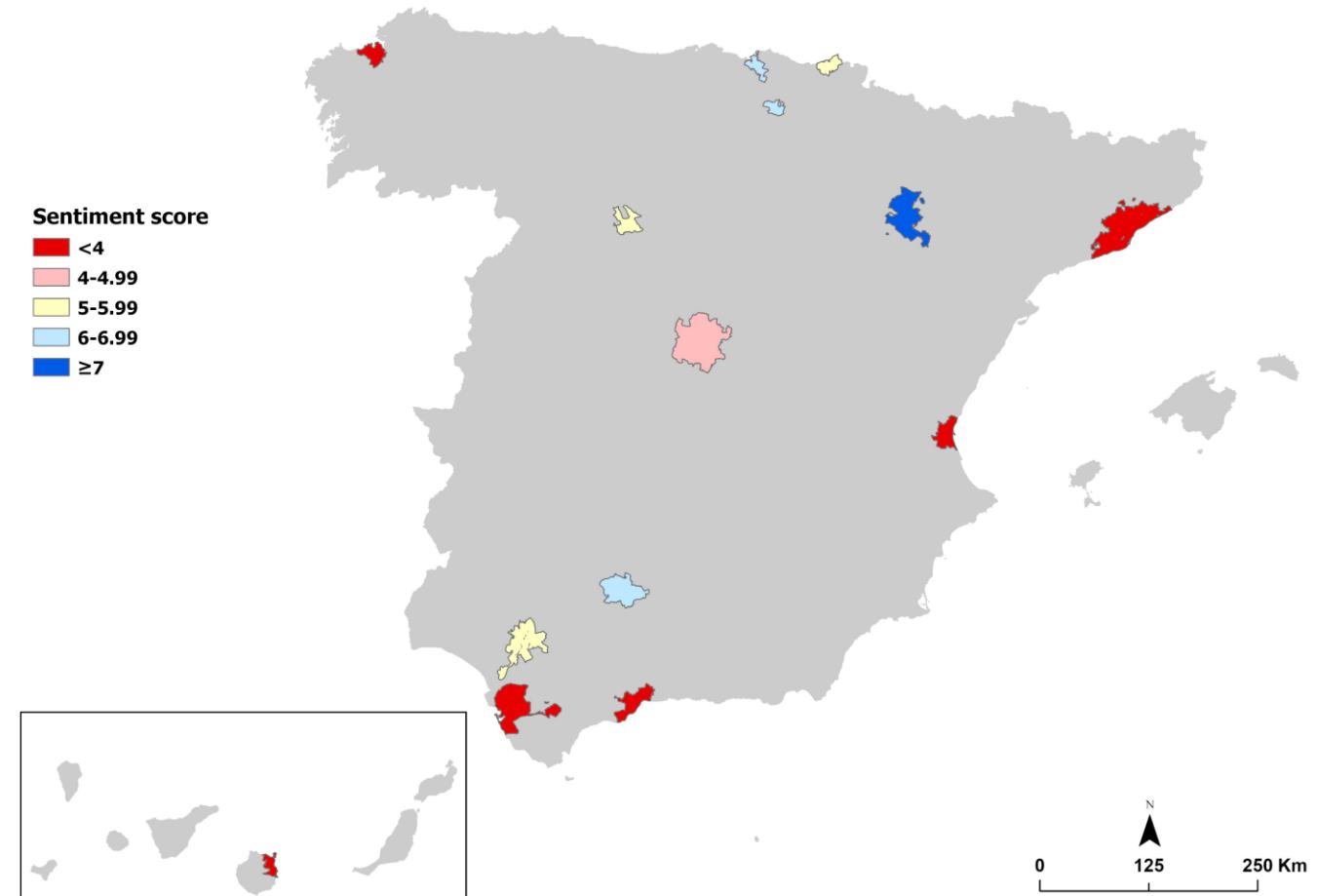


# Ejemplos de uso de fuentes de datos

-Twitter como fuente de datos para analizar la percepción de la micromovilidad de las ciudades principales de España.

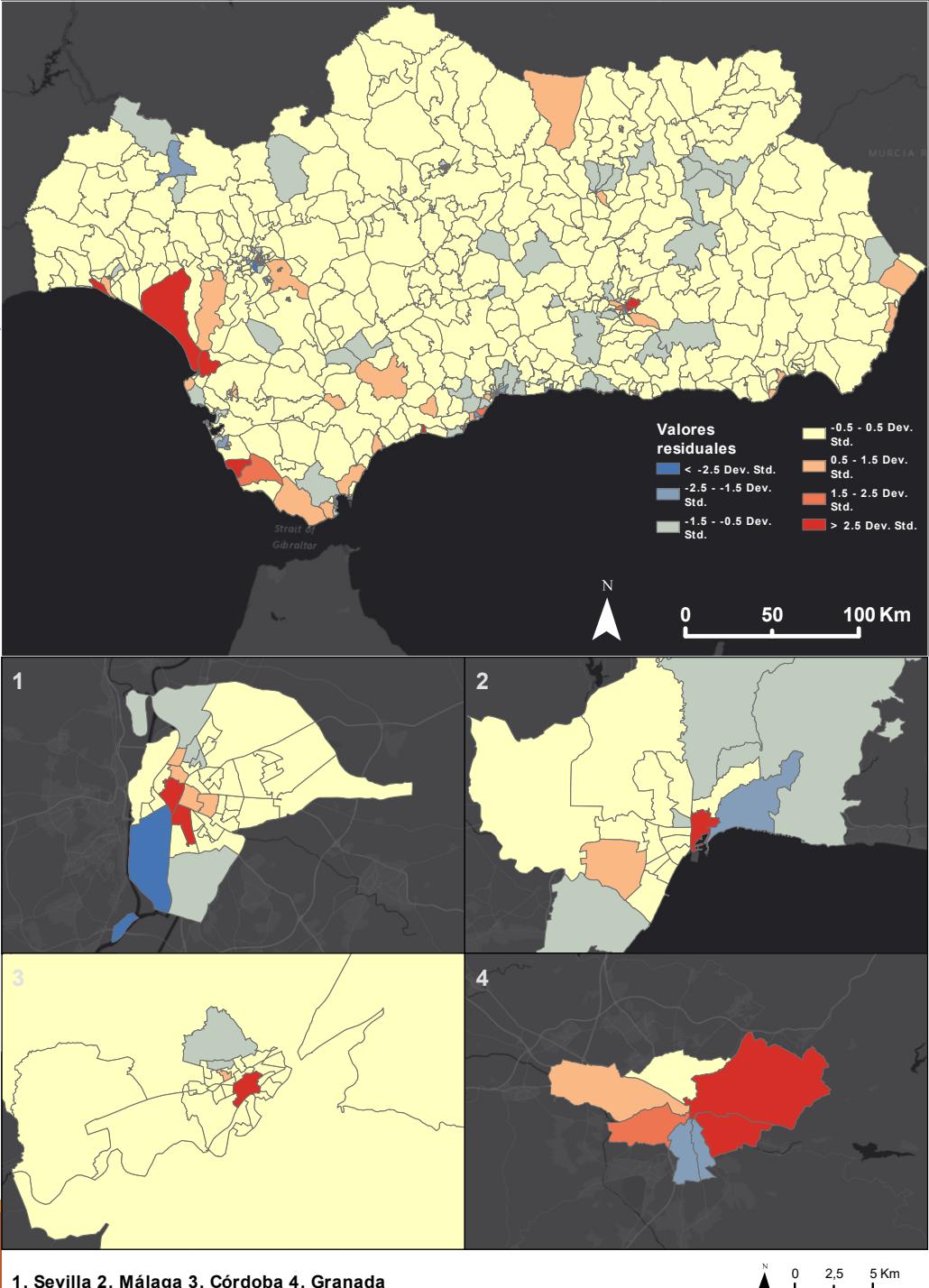


Osorio-Arjona, J., Arias-Molinares, D. Comparing Social Media Sentiments with Cycling Infrastructure Surveys: a Comprehensive Analysis of Micromobility Perceptions Across Multiple Spanish Urban Areas. Major Revision en Journal of Cycling and Micromobility Research.



# Ejemplos de uso de fuentes de datos

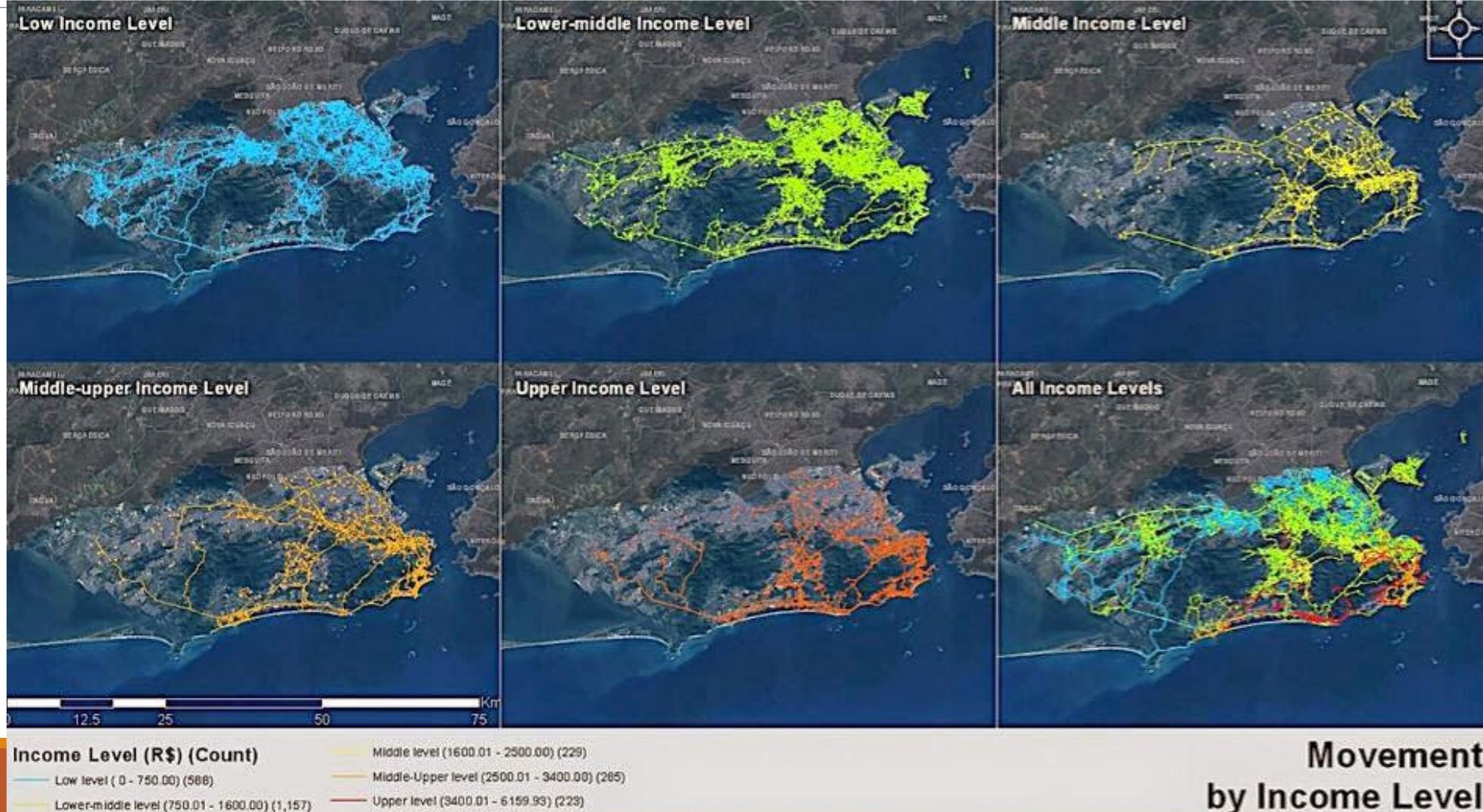
-Combinación de datos de Twitter y de OpenStreetMap para el análisis del grado de integración de las ciudades de los 15 minutos en múltiples ciudades de una misma región.



Osorio-Arjona, J. Análisis del desarrollo de las ciudades de 15 minutos en Andalucía a partir de datos de Twitter y OpenStreetMap. Enviado a Revista de Estudios Andaluces.

# Ejemplos de uso de fuentes de datos

Cartografía de espacios de segregación a partir de la combinación de datos de Twitter con datos socioeconómicos de otras fuentes de datos.

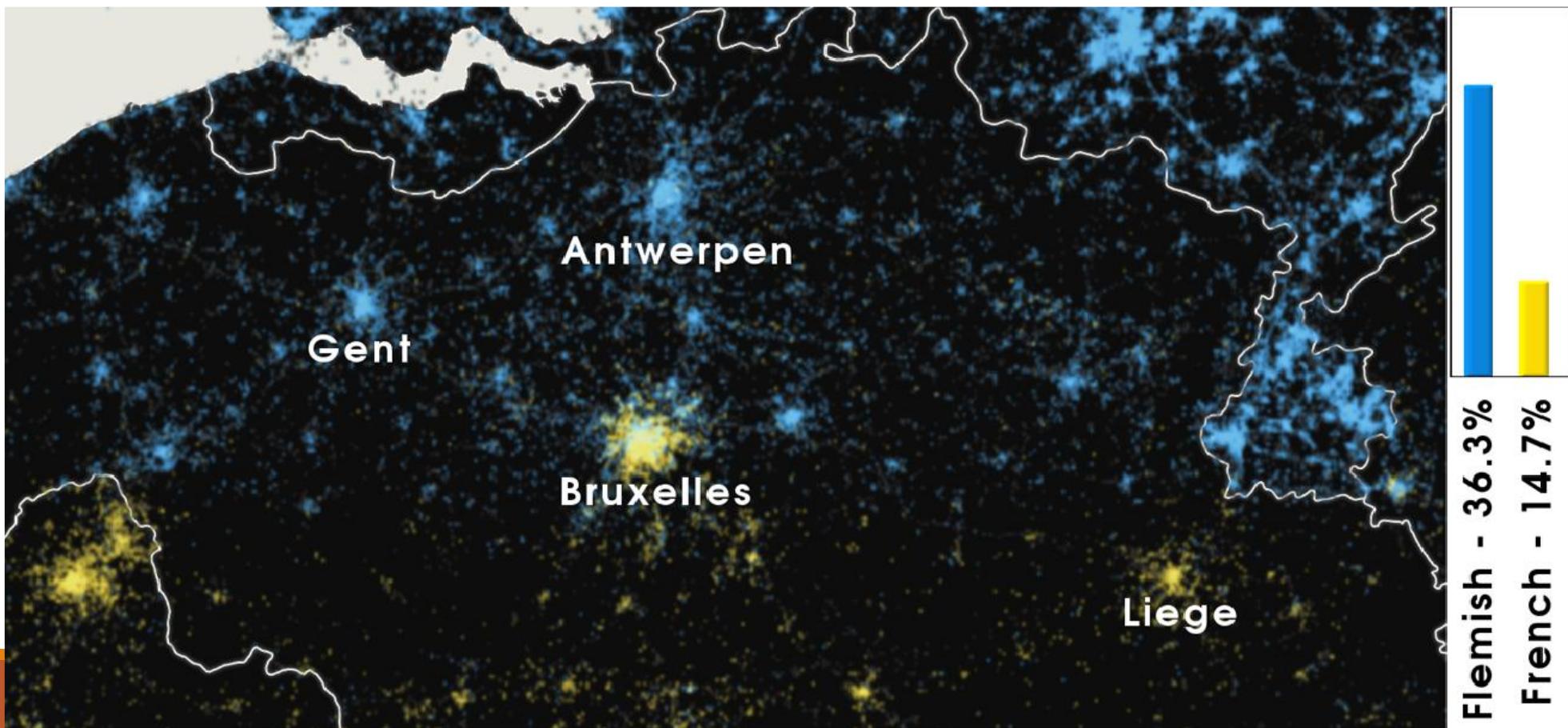


<http://urbandemographics.blogspot.co.uk/2015/02/finding-networks-of-segregation-through.html>

# Ejemplos de uso de fuentes de datos

---

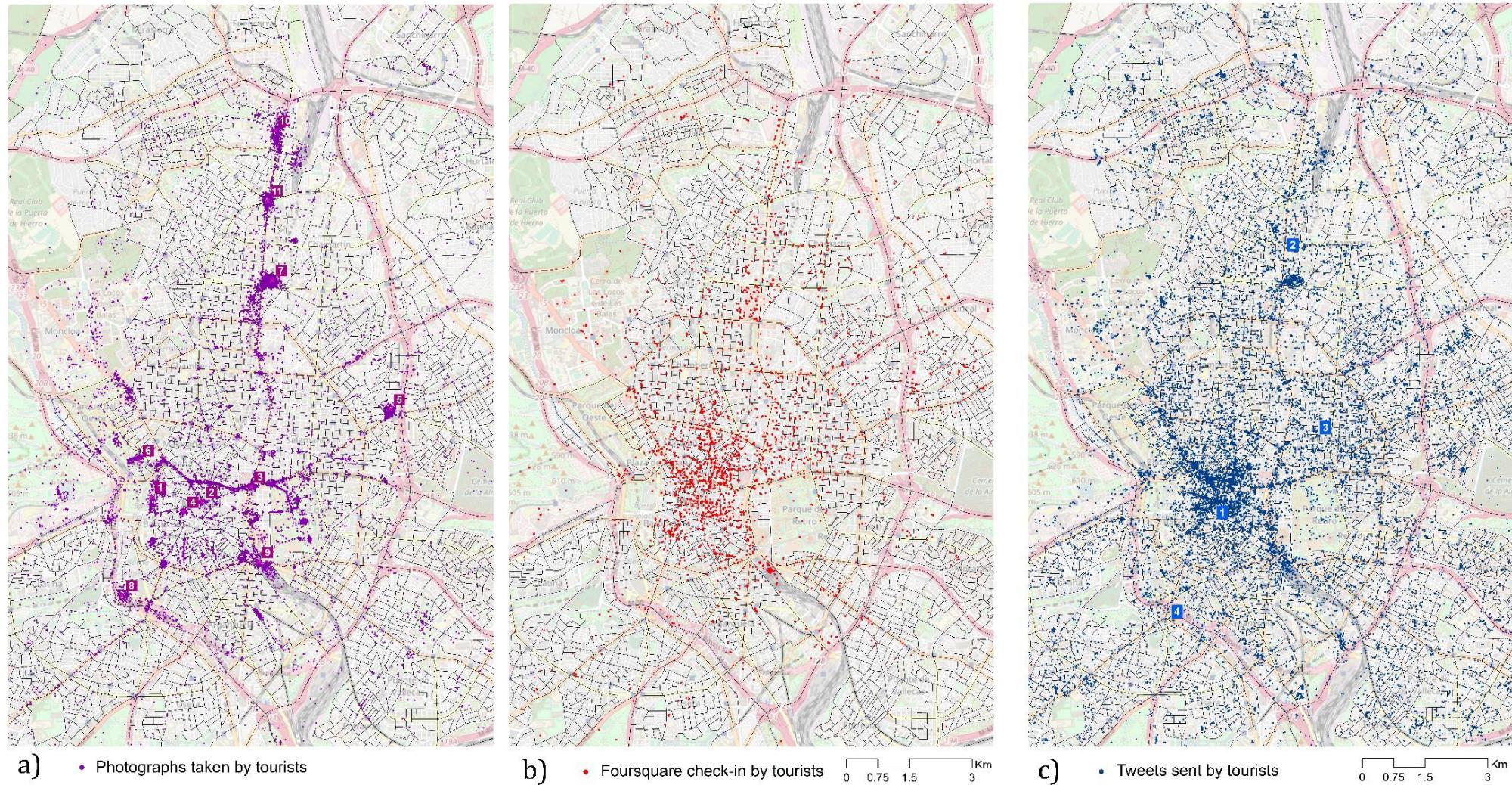
Cartografía de fronteras  
lingüísticas.



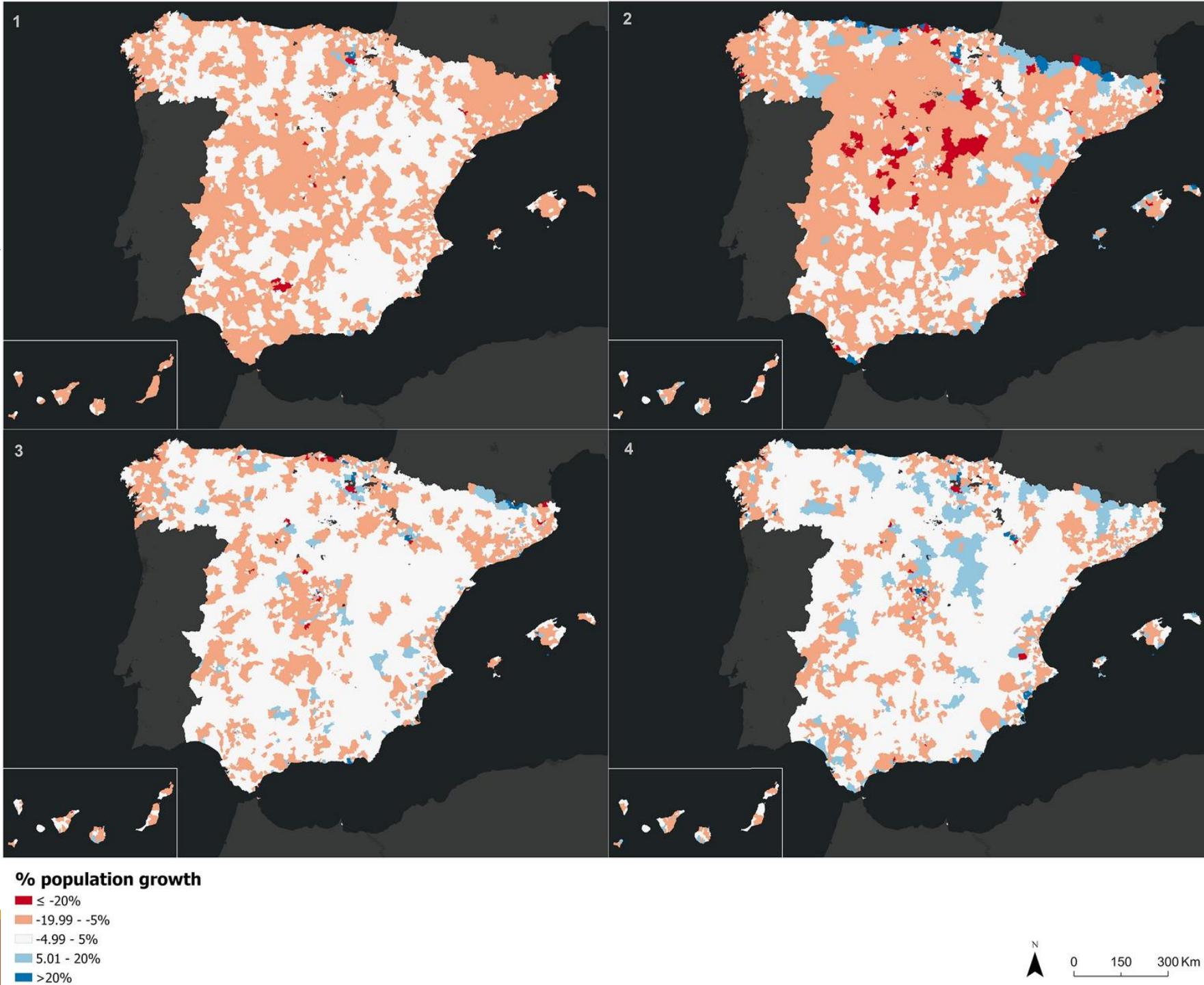
# Ejemplos de uso de fuentes de datos

Turismo a partir de redes sociales: ¿Dónde hacen fotografías los turistas? ¿En qué sitios tienden a comer o consumir? ¿Dónde twittean estos turistas? ¿Qué similitudes y diferencias se observan entre estas tres fuentes?

Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C., Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13-25



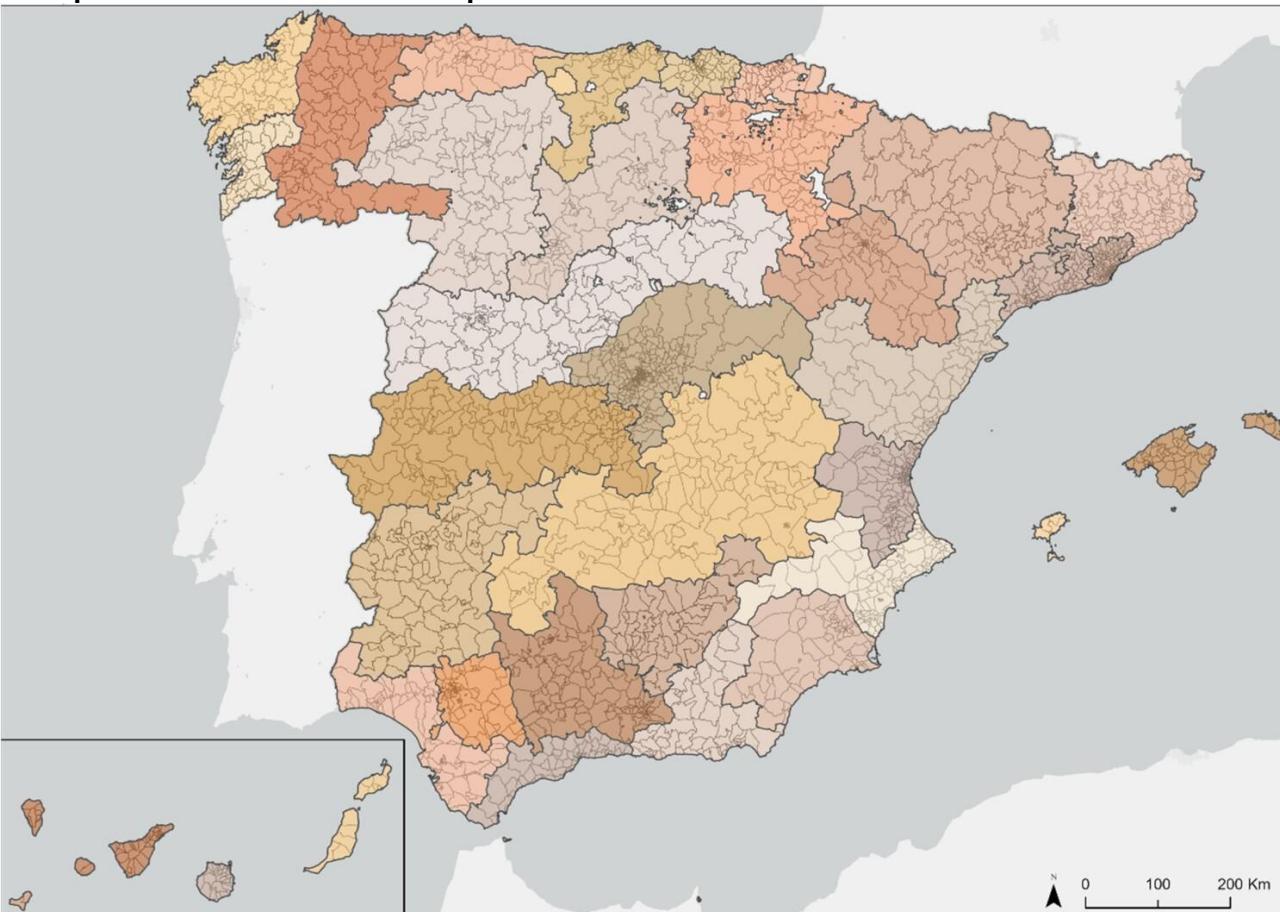
# Uso de datos de telefonía móvil para cartografiar cambios en la demografía y la movilidad de la población de España durante cuatro diferentes fases de la pandemia causada por el COVID-19.



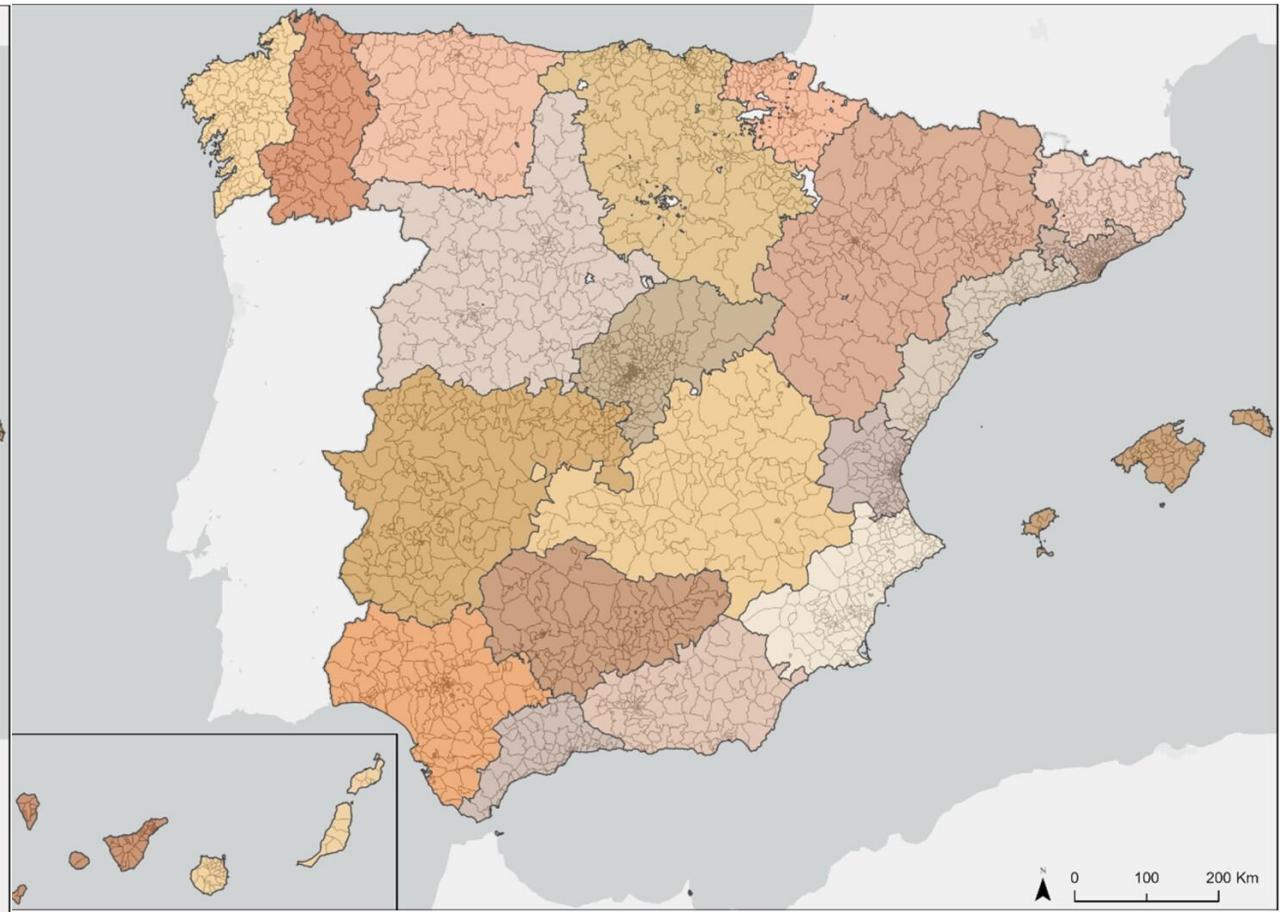
Osorio-Arjona, J., Ruiz-Santacruz, J.S., de las Obras Loscertales-Samperiz, J.(2023). Mapping of functional areas in Spain based on mobile phone data during different phases of the COVID-19 pandemic. *Journal of Maps*, 18 (4), 1-10  
(DOI:10.1080/17445647.2023.2214804).

# Ejemplos de uso de fuentes de datos

-Datos de telefonía móvil para la cartografía de áreas funcionales en España durante cuatro fases diferentes de la pandemia causada por el COVID-19.

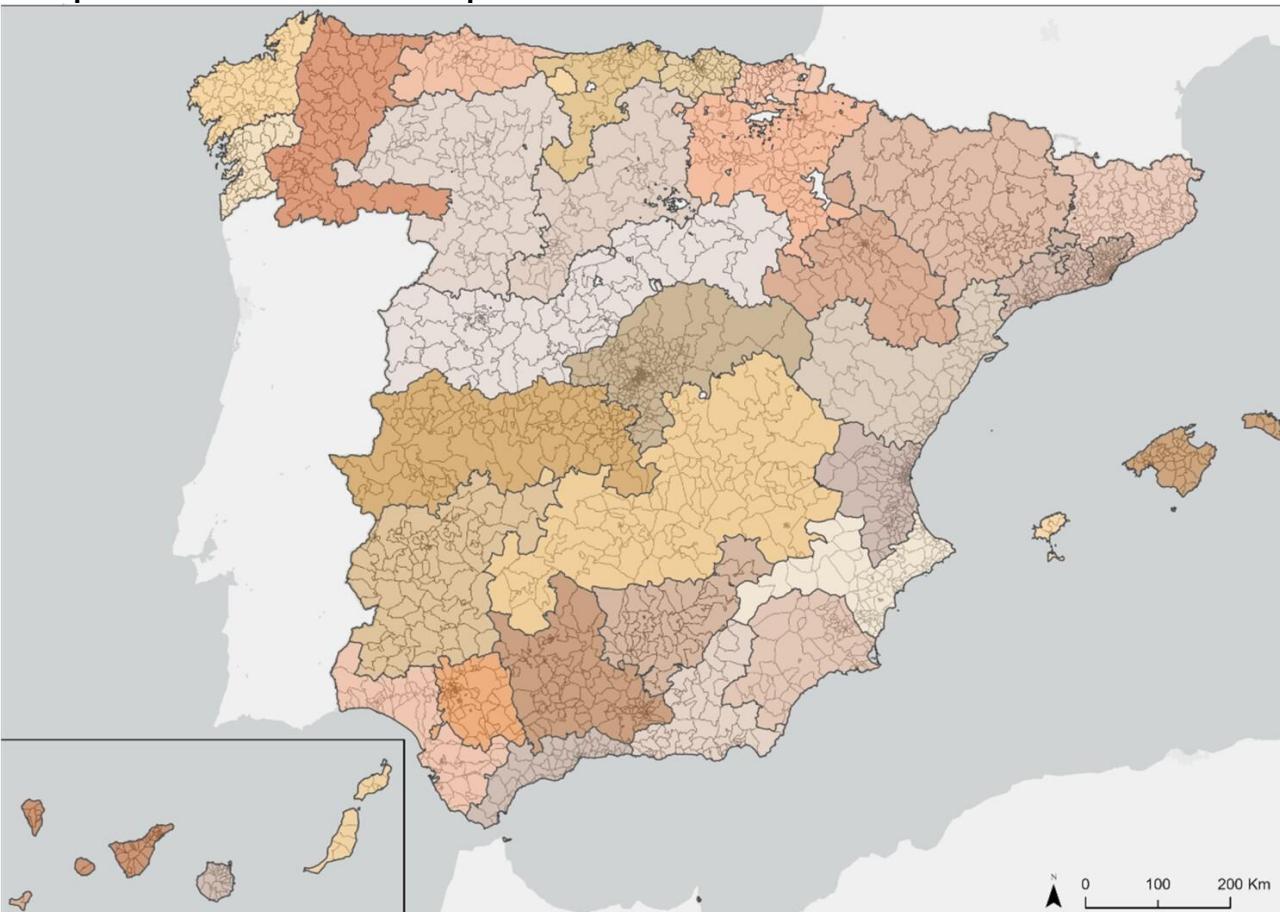


Osorio-Arjona, J., Horak, J., Svoboda, R., García-Ruiz, Y. (2021). Social media semantic perceptions on Madrid Metro system: using Twitter data to link complaints to space. *Sustainable Cities and Society*, 64 (102530) (DOI: 10.1016/j.scs.2020.102530).

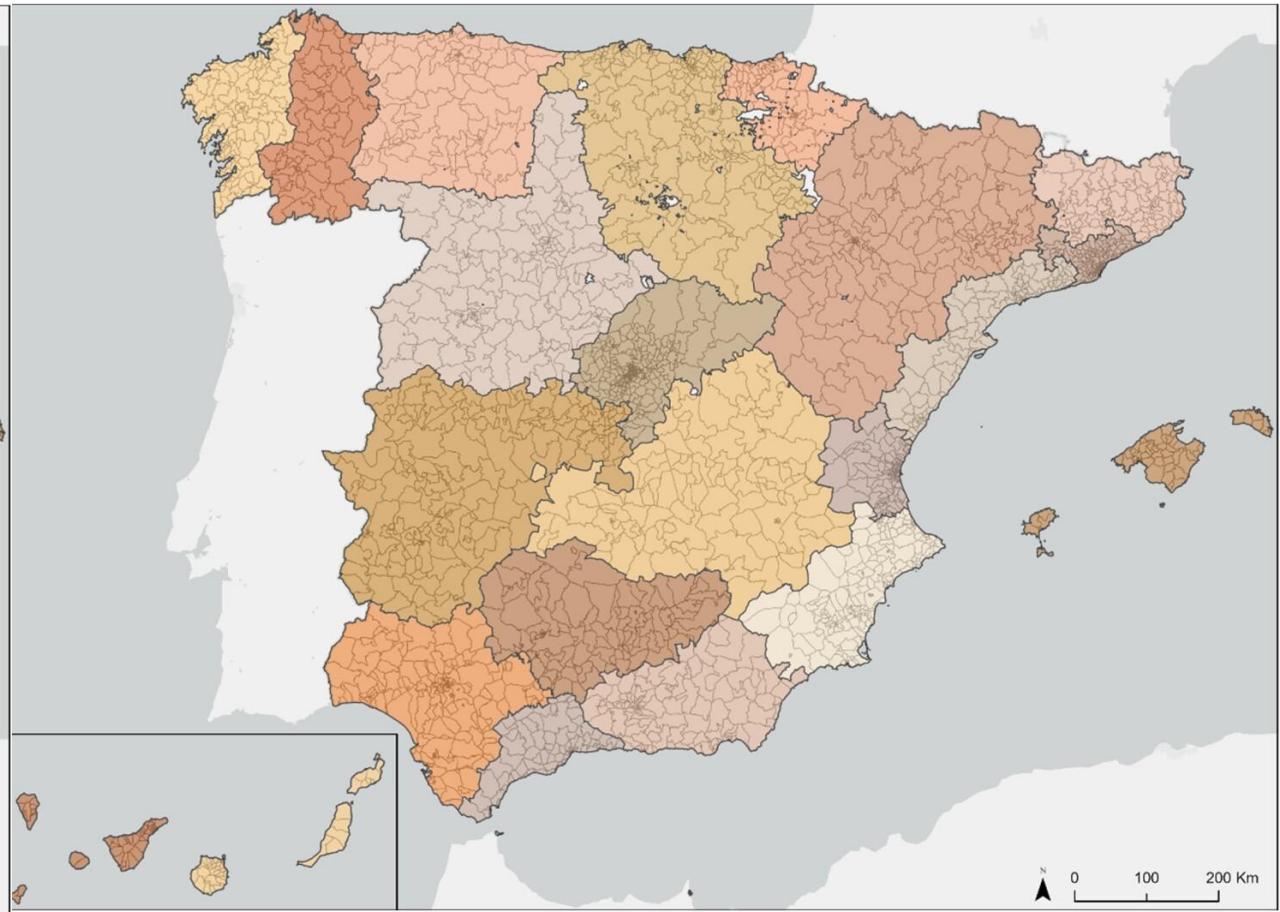


# Ejemplos de uso de fuentes de datos

-Datos de telefonía móvil para la cartografía de áreas funcionales en España durante cuatro fases diferentes de la pandemia causada por el COVID-19.

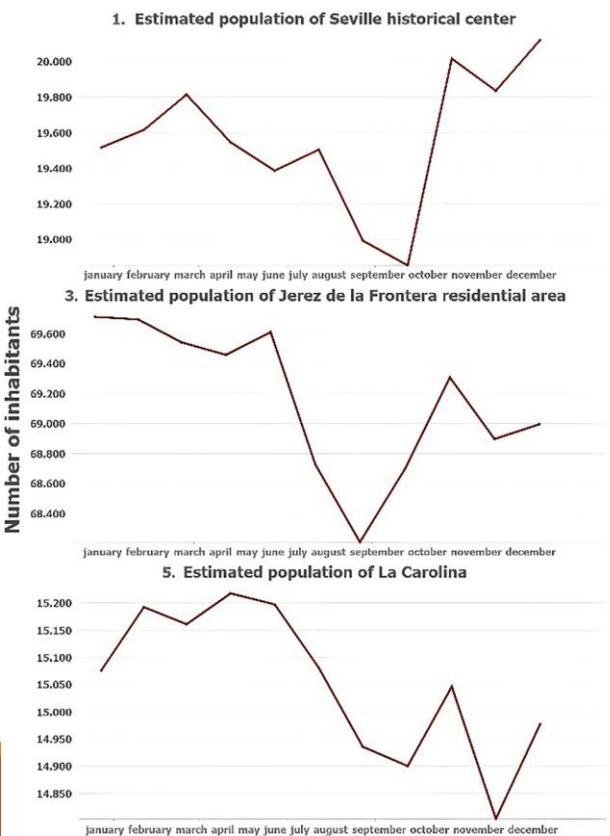


Osorio-Arjona, J., Horak, J., Svoboda, R., García-Ruiz, Y. (2021). Social media semantic perceptions on Madrid Metro system: using Twitter data to link complaints to space. *Sustainable Cities and Society*, 64 (102530) (DOI: 10.1016/j.scs.2020.102530).

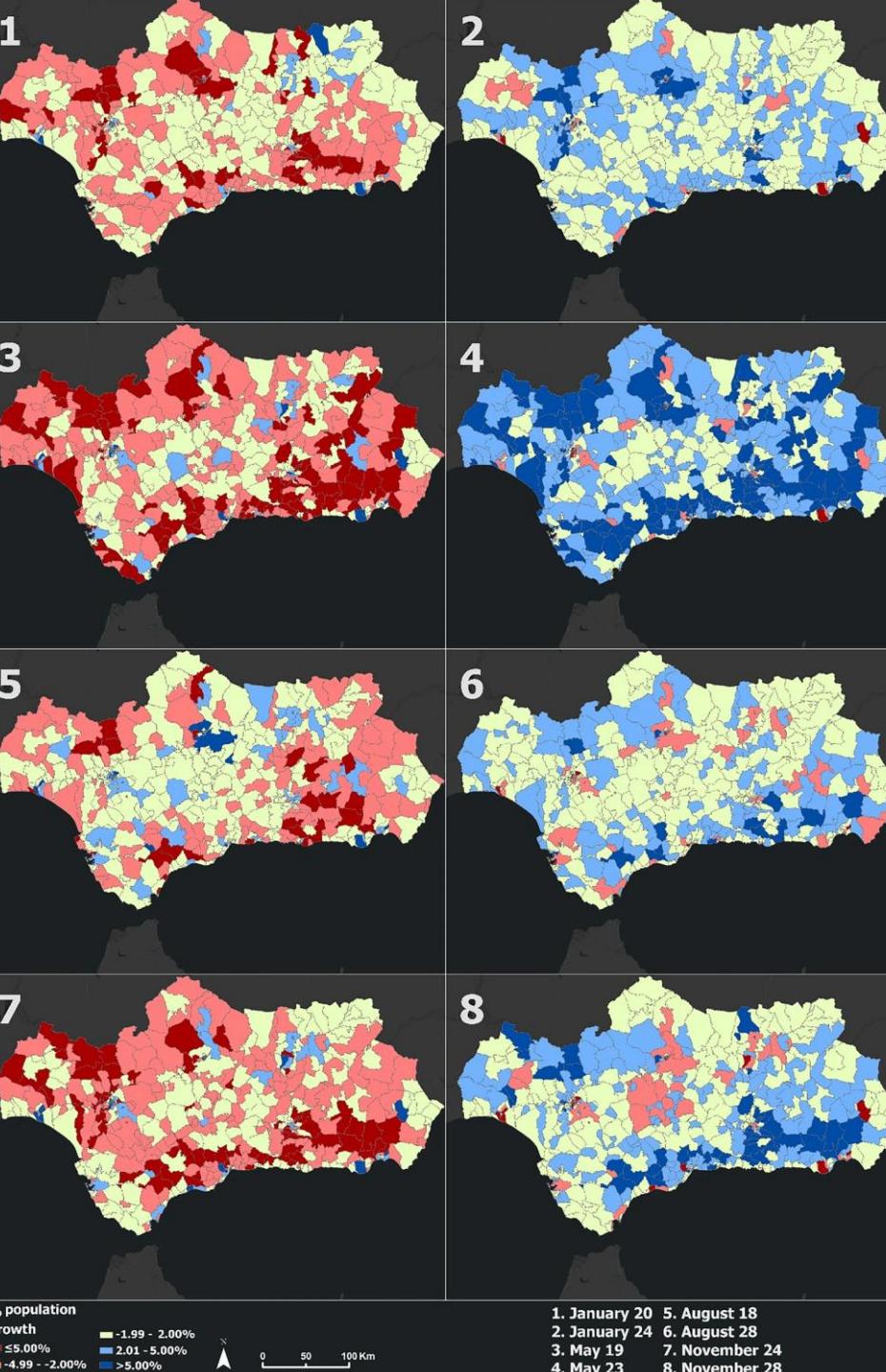


# Ejemplos de uso de fuentes de datos

Datos de telefonía móvil para la creación de una población sintética a partir de los cambios de movilidad de una región durante un año.



-Osorio-Arjona, J. (2024). Analyzing post-COVID-19 demographic and mobility changes in Andalusia using mobile phone data. *Scientific Reports*, 14(14828), 1-15 (DOI:10.1038/s41598-024-65843-2)



Aplicación de datos de AirBNB en la asignatura Estadística Aplicada del Máster de Humanidades Digitales de la UNED.

```
In [1]: #Los tres ficheros de la práctica han sido descargados a partir del portal de datos abiertos de AirBNB. Consisten en datos de alquileres de viviendas en la ciudad de Sevilla.  
#El fichero calendar tiene información sobre las fechas de alojamientos desde el 24 de septiembre del año 2023 al 22 de septiembre del año 2024.  
#El fichero listings cuenta con información sobre el alojamiento y las valoraciones que recibe.  
#El fichero reviews tiene información acerca de las valoraciones individuales de cada review.  
  
#Para responder a las preguntas, escribid la respuesta utilizando el comando print. Los valores con decimales tienen que tener 2 decimales.  
import pandas as pd  
import matplotlib.pyplot as plt
```

```
In [2]: #Ejercicio 1  
#Cread una tabla a partir del fichero listings.  
#Esta tabla tiene que tener los siguientes campos: id, neighbourhood_group_cleansed, latitude, longitude, beds, price, review_scores_accuracy, review_scores_cleanliness, review_scores_rating, review_scores_value.  
  
listings = pd.read_csv("./practica/listings.csv", sep=",", encoding='latin-1')  
listings = listings[['id', 'neighbourhood_group_cleansed', 'latitude', 'longitude', 'beds', 'price', 'review_scores_accuracy', 'review_scores_cleanliness', 'review_scores_rating', 'review_scores_value']]  
listings
```

Out[2]:

		id	neighbourhood_group_cleansed	latitude	longitude	beds	price	review_scores_accuracy	review_scores_cleanliness	review_scores_rating	review_scores_value
0		32347	Casco Antiguo	37.393580	-5.999750	2.0	\$99.00	4.91	4.96	4.96	4.96
1		49287	Casco Antiguo	37.398980	-5.995330	1.0	\$80.00	4.86	4.97	4.97	4.97
2		108236	Casco Antiguo	37.397940	-5.997950	3.0	\$94.00	4.85	4.83	4.83	4.83
3		108568	Casco Antiguo	37.399410	-5.993790	4.0	\$113.00	4.63	4.64	4.64	4.64
4		116703	Casco Antiguo	37.385510	-5.994970	2.0	\$97.00	4.69	4.68	4.68	4.68
...		...	...	...	...	...	...	...	...	...	...
7403	984498235656406767		Triana	37.381564	-5.997593	4.0	\$509.00	NaN	NaN	NaN	NaN
7404	984715946098475857		Triana	37.385759	-6.007944	4.0	\$170.00	NaN	NaN	NaN	NaN
7405	984839164085370227		Casco Antiguo	37.392893	-5.986302	4.0	\$700.00	NaN	NaN	NaN	NaN
7406	985289624228114477		Casco Antiguo	37.391933	-5.989434	3.0	\$150.00	NaN	NaN	NaN	NaN
7407	985517327675529213		Sur	37.368590	-5.9777971	6.0	\$305.00	NaN	NaN	NaN	NaN

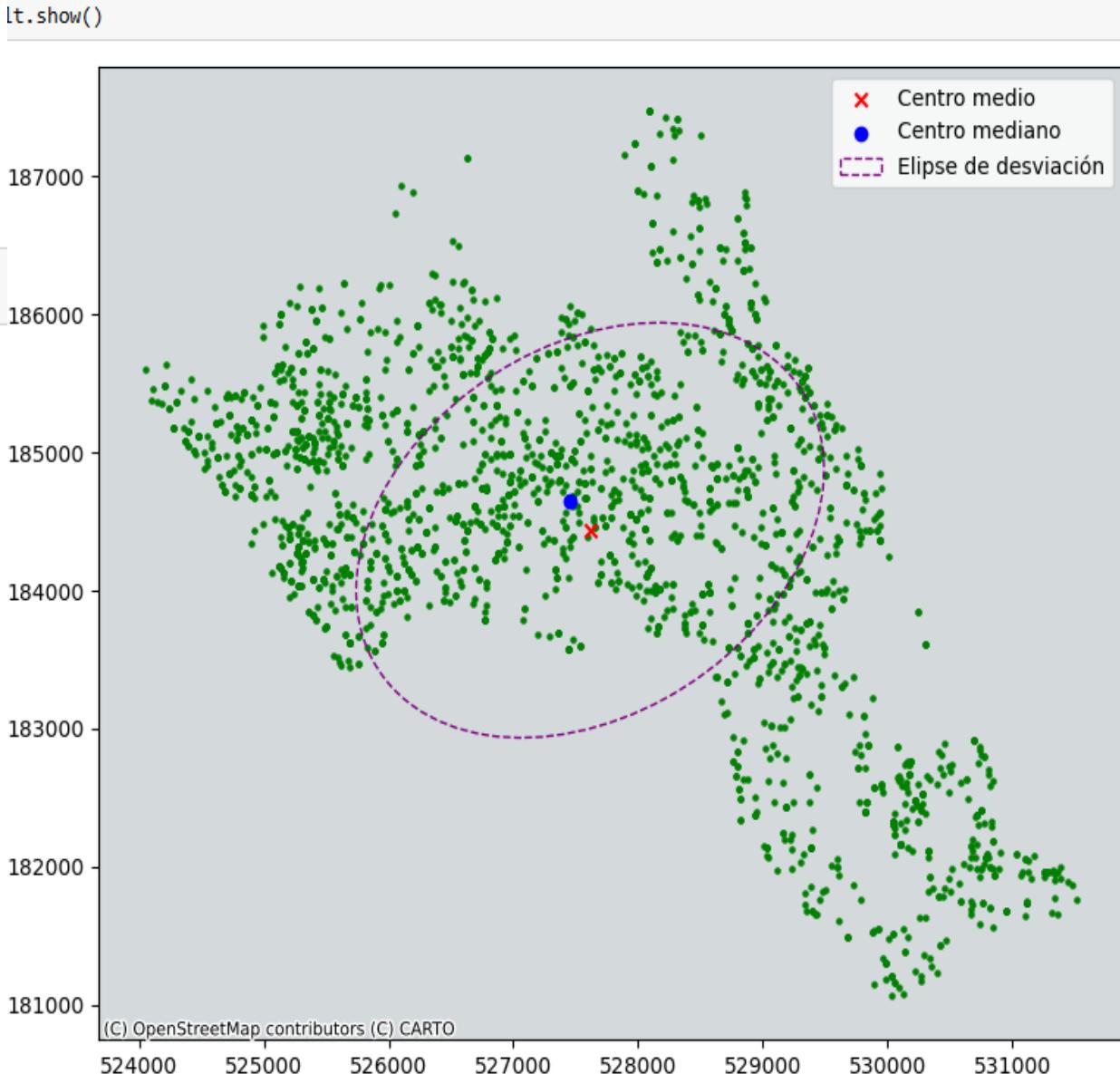
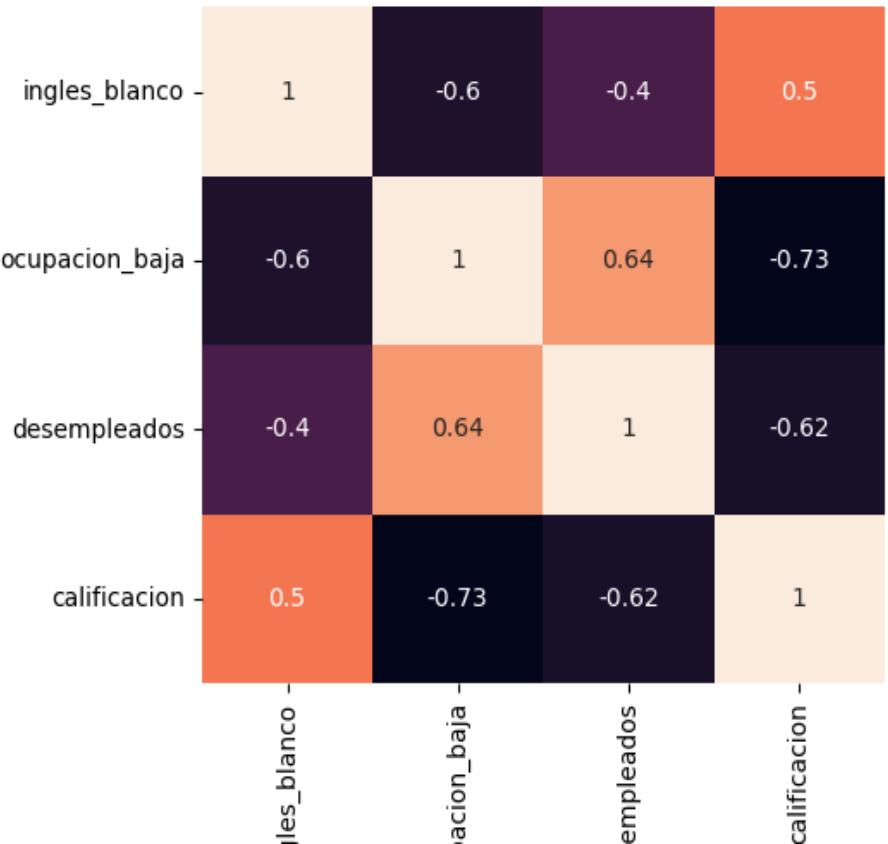
7408 rows × 11 columns

```
In [19]: #En el caso de Desempleados y Variables, el valor apenas ha cambiado, pero en otra parte
#¿Qué método es mejor para observar la relación entre calificación e inglés bajo Display
#Vamos a trabajar con la primera matriz de correlación, la que hemos calculado con Add basemap
cor_mat.round(3)
contextily.add_basemap(
    ax, source=contextily.providers.CartoDB.Positron
```

```
Out[19]:
```

	ingles_blanco	ocupacion_baja	desempleados	calificacion
ingles_blanco	1.000	-0.601	-0.398	0.499
ocupacion_baja	-0.601	1.000	0.641	-0.735
desempleados	-0.398	0.641	1.000	-0.624
calificacion	0.499	-0.735	-0.624	1.000

```
In [20]: #Ahora vamos a crear una matriz de correlación con La Librería seaborn.
sns.heatmap(cor_mat, annot=True);
```



# Ejemplos de uso de fuentes de datos

Se pueden crear contenidos docentes a partir de Big Data geolocalizado para una gran multitud de asignaturas de geografía:

- ❖ Geografía Física
- ❖ Geografía Humana
- ❖ Geografía de España
- ❖ Cartografía
- ❖ Estadística Aplicada
- ❖ Sistemas de Información Geográfica
- ❖ Programación
- ❖ Geografía Urbana
- ❖ Geografía Regional de España
- ❖ Paisaje, Patrimonio y Turismo
- ❖ Geografía de la Población
- ❖ Fuentes para la investigación
- ❖ Etc.

## Ventajas del Big Data geolocalizado

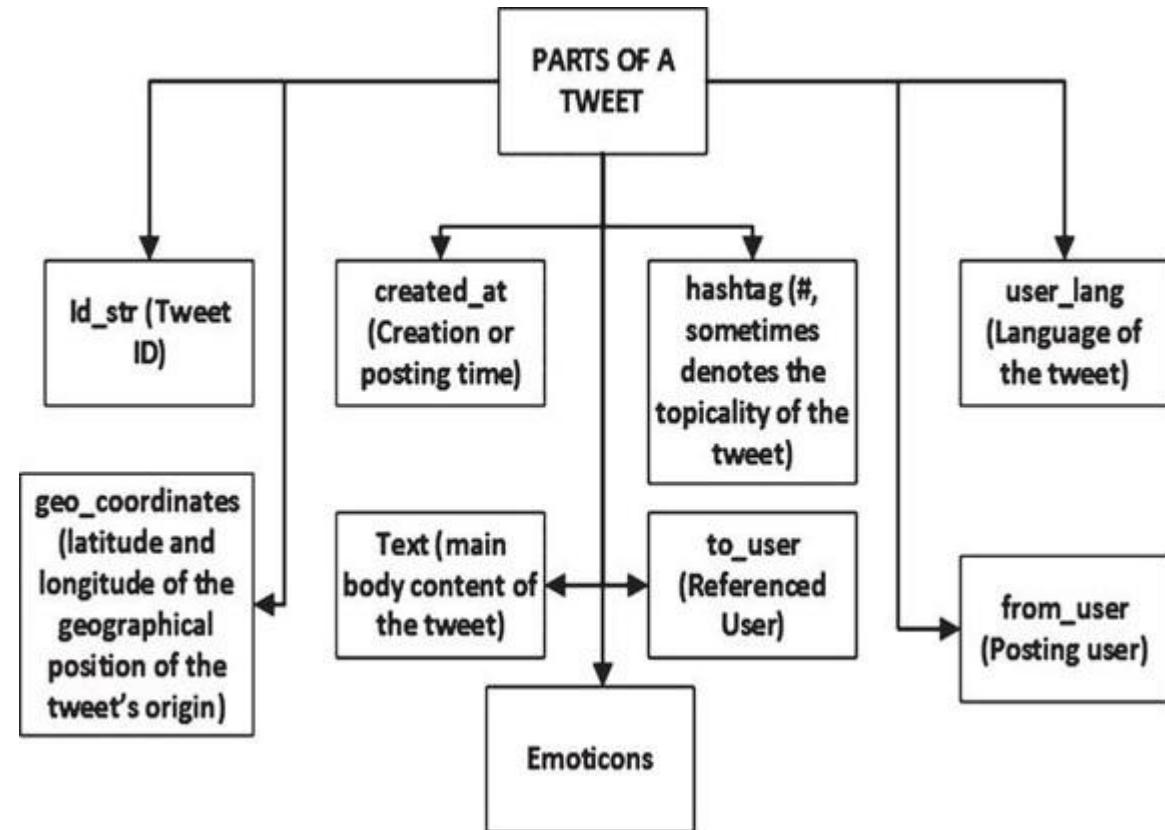
Cobertura global	En muchas fuentes
Recogida pasiva	Verosimilitud: lo que la gente hace, no lo que dice que hace
Grandes volúmenes de datos	Muestras de tamaño mucho mayor que en muchas fuentes Compleción (tarjeta transporte y bicicletas públicas)
Resolución temporal alta	Datos continuos: <ul style="list-style-type: none"><li>- información siempre actualizada</li><li>- monitorización en tiempo real o casi real</li><li>- secuencias temporales: día, semana, mes, año</li><li>- períodos atípicos</li></ul>
Resolución espacial alta – media/alta	Localización precisa: <ul style="list-style-type: none"><li>- GPS</li><li>- registros teléfonos móviles (antenas – triangulación)</li></ul>
Coste	Bajo coste de algunas fuentes: disponibles en internet (APIs y open data)

## Limitaciones del Big Data geolocalizado

Sesgo	Variable según fuentes de datos: - compleción (tarjetas transporte / registros bicicletas públicas) - bajo en registros de telefonía (penetración de la compañía y filtrado) - medio en redes sociales
Dificultad para limpieza, almacenamiento y proceso de datos	Erros Datos masivos – Tecnologías Big Data
Información pobre sobre los usuarios	Enriquecimiento de datos - por variables de los registros (directa o indirecta) - por patrones de comportamiento (domicilio-trabajo) - por cruce espacial – GIS
Accesibilidad a los datos	Dependencia de empresas para ciertas fuentes de datos

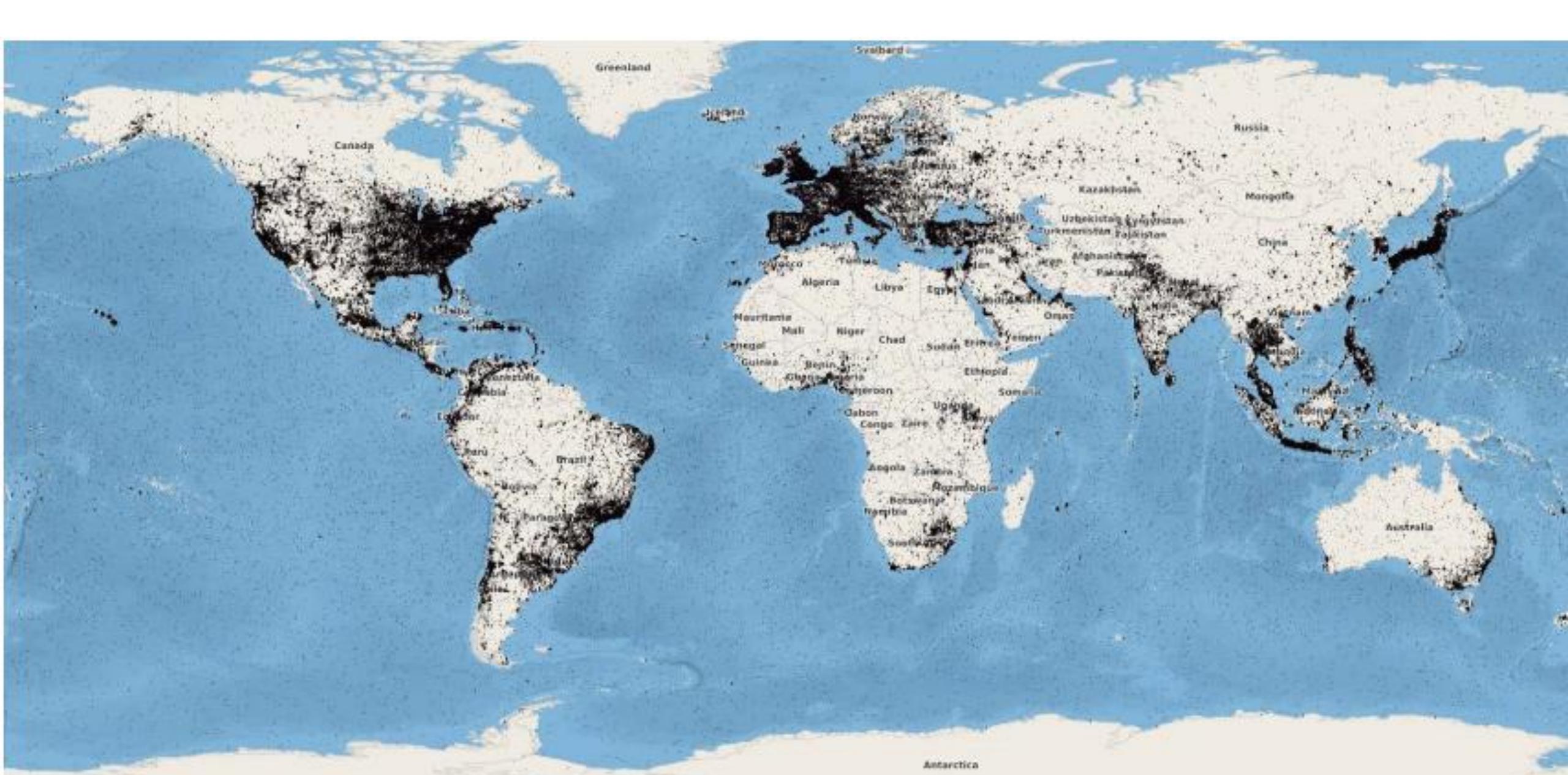
# Fuentes de datos

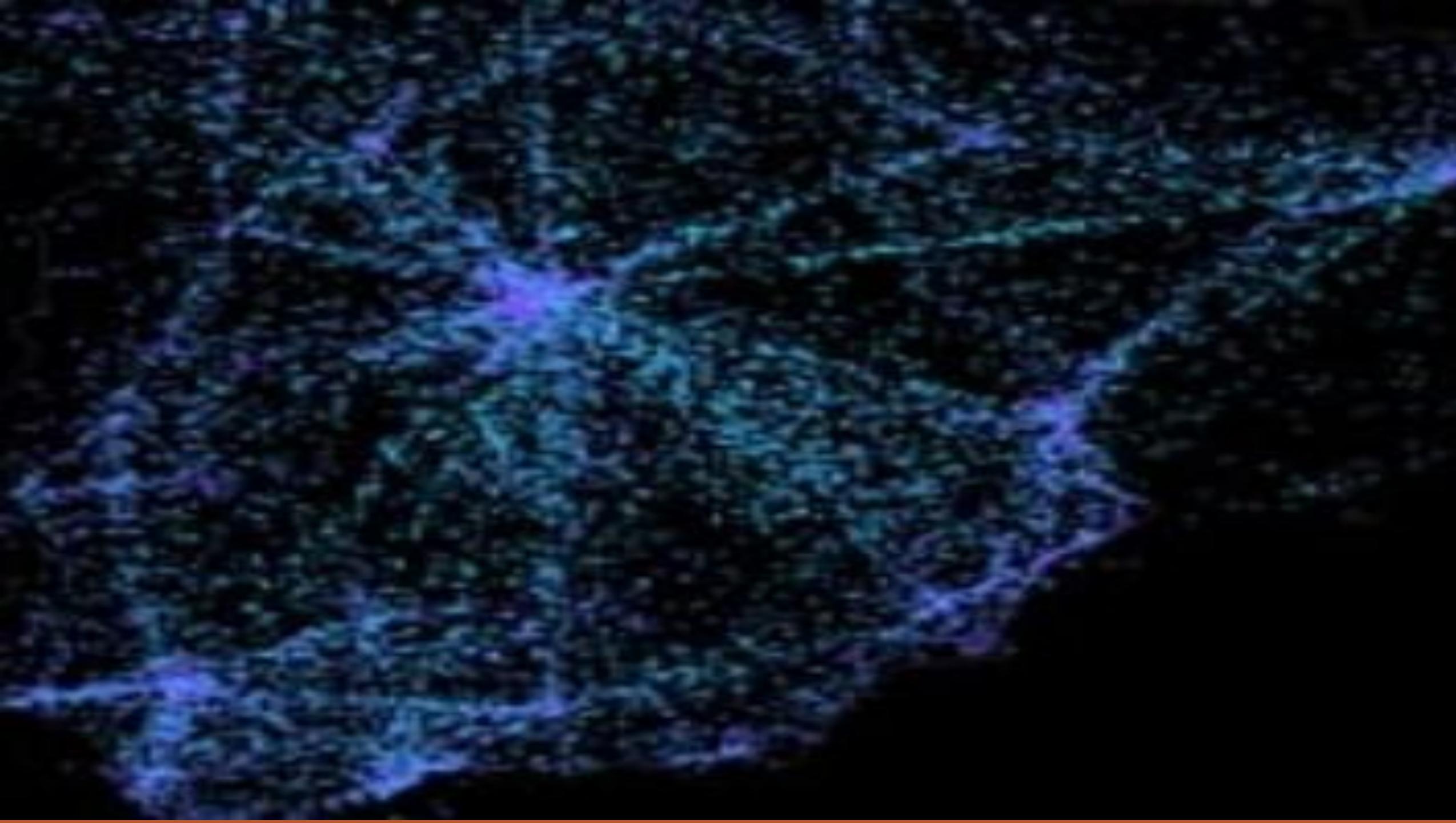
- ❖ Aunque los datos de telefonía presentan mayor volumen de datos y mayor detalle temporal, la estructura de **Twitter** y la geolocalización a partir de coordenadas GPS le confiere de un mayor detalle espacial y mayor valor semántico que los datos de telefonía móvil.
- ❖ Twitter tiene una gran penetración, permitiéndonos trabajar fácilmente con grandes cantidades de datos.
- ❖ Los datos se pueden agregar por usuario, espacio y tiempo, confiriendo una gran flexibilidad a los análisis y permitiendo la elaboración de cartografía temporal animada.
- ❖ Posibilidad de realizar filtros temáticos (por ejemplo, descargar tweets que contengan palabras relacionadas con vino o con bicicletas).



# Fuentes de datos

lang character varying	textFix character varying	user_id character varying	created_at timestamp without time zone
es	creamos una visita virtual a tu vivienda para que los compradores	306695960	2020-06-04 19:44:31
en	we must try not to sink beneath our anguish, harry, but battle on	128613823	2020-06-04 19:45:00
ca	son tres quarts de vuit del vespre	1384110594	2020-06-04 19:45:01
en	ding ding ding!!!!!!	584661634	2020-06-04 19:45:01
es	extranaba venir aqui en dehesa de la villa <a href="https://t.co/ekqchrncu3">https://t.co/ekqchrncu3</a>	75459867	2020-06-04 19:45:05
es	la coloracion elumenated se obtiene gracias a los pure pigments que envuelven de un halo, el color profundo e inten... <a href="https://t.c...">https://t.c...</a>	2951950750	2020-06-04 19:45:21
es	~que empiece la cuenta atras	1172855312	2020-06-04 19:45:39
es	in the dark en el plantio, madrid, spain <a href="https://t.co/6cfeou6qew">https://t.co/6cfeou6qew</a>	526544988	2020-06-04 19:46:02
es	#fase2, dia 11. mientras la mayoria esperamos con ansia la fase 3 y nos preguntamos que podremos hacer entonces, a... <a href="https://t.c...">https://t.c...</a>	203777391	2020-06-04 19:46:14
es	estuches rigidos a medida para tus instrumentos.	900818644799811587	2020-06-04 19:46:17
es	emorzar de forquilla #food #emorzar #family #friend en pineda de mar <a href="https://t.co/raq7aeauzn">https://t.co/raq7aeauzn</a>	1267796652540203014	2020-06-04 19:46:28
es	mizuno by la imperial granadina en cenes de la vega, spain <a href="https://t.co/mxyj6iyktv">https://t.co/mxyj6iyktv</a>	1153629245496844293	2020-06-04 19:46:32
es	#bisuteria ideal #ss20	1192612129	2020-06-04 19:46:39
en	happy @ barcelona, spain <a href="https://t.co/i8sclrjqkp">https://t.co/i8sclrjqkp</a>	727918546538369024	2020-06-04 19:46:40
und	#volkswagen #eos #vweos #vwcabriolet #vw #cabrio #cabriolet #descapotavel #serradaarrabida #setubal #portugal... <a href="https://t.c...">https://t.c...</a>	1168808994	2020-06-04 18:46:56
es	la mente humana no es capaz de concebir la cuarta dimension, asi que, como puede concebir a dios? para quien mil a... <a href="https://t.c...">https://t.c...</a>	156370892	2020-06-04 19:47:27
en	life becomes more meaningful when you realize the simple fact that youll never get the same moment twice @ barce... <a href="https://t.c...">https://t.c...</a>	132706743	2020-06-04 19:48:08
en	be good to people for no reason always, no matter our color, our religion, our believes. love will win a marsei... <a href="https://t.co/1kmd...">https://t.co/1kmd...</a>	117453887	2020-06-04 19:48:19
en	fell in love with those trees.	40958400	2020-06-04 18:48:28





# Fuentes de datos

- ❖ Desgraciadamente, desde que Elon Musk compró la red social, ha implementado una privatización de esta, y desde junio de 2022 los datos de Twitter dejaron de ser gratuitos.
- ❖ Lo que antes era gratuito, ahora cuesta 5000 euros al mes.

## Find the right access for you

### Free

- For write-only use cases and testing the X API
- Rate-limited access to v2 posting and media upload endpoints
  - 1,500 posts per month - posting limit at the user level
  - 1,500 posts per month - posting limit at the app level
  - 1 app ID
  - Login with X
  - Access to Ads API
  - Free

[Get started](#)

### Basic

- For hobbyists or prototypes
- Rate-limited access to suite of v2 endpoints
  - 3,000 posts per month - posting limit at the user level
  - 50,000 posts per month - posting limit at the app level
  - 10,000 posts per month - read limit at the app level
  - 2 app IDs
  - Login with X
  - Access to Ads API
  - \$100 per month

[Subscribe now](#)

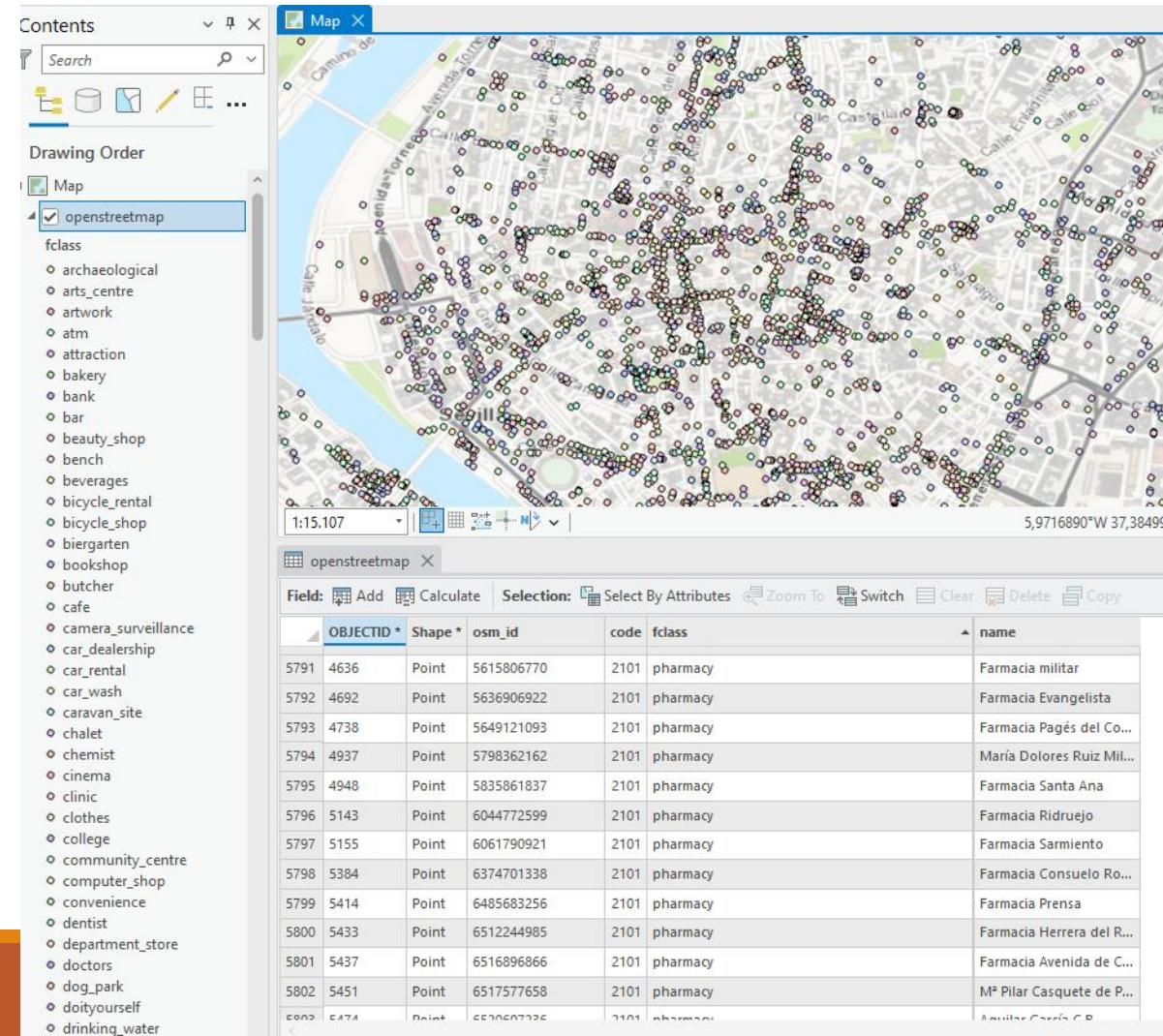
### Pro

- For startups scaling their business
- Rate-limited access to suite of v2 endpoints, including search and filtered stream
  - 288,000 posts per month - posting limit at the user level
  - 300,000 posts per month - posting limit at the app level
  - 1,000,000 posts per month - read limit at the app level
  - 3 app IDs
  - Login with X
  - Access to Ads API
  - \$5,000 per month

[Subscribe now](#)

# Fuentes de datos

- ❖ OpenStreetMap ofrece de forma gratuita todos sus datos, los cuales se actualizan todos los días.
- ❖ Estos datos son creados de forma voluntaria por ciudadanos que publican una nueva tienda, farmacia, hospital, banco, etc.
- ❖ Posibilidad de descargar los datos a diferentes escalas espaciales (continente, país, comunidad autónoma) en formato shapefile.
- ❖ <https://download.geofabrik.de/>



Do you find OpenStreetMap data useful? Consider giving something back in the OSMF's [funding campaign!](#)

## GEOFABRIK downloads

Download OpenStreetMap data for this region:

# Andalucía

[\[one level up\]](#)

The OpenStreetMap data files provided on this server do **not** contain the user names, user IDs and changeset IDs of the OSM objects because these fields are assumed to contain personal information about the OpenStreetMap contributors and are therefore subject to data protection regulations in the European Union.

[Extracts with full metadata](#) are available to OpenStreetMap contributors only.



## Commonly Used Formats

- [andalucia-latest.osm.pbf](#), suitable for Osmium, Osmosis, imposm, osm2pgsql, mkgmap, and others. This file was last modified 9 hours ago and contains all OSM data up to 2024-10-28T21:21:30Z. File size: 137 MB; MD5 sum: [dee6b9cd7aedcb407e9a095ae252365a](#).
- [andalucia-latest-free.shp.zip](#), yields a number of ESRI compatible shape files when unzipped. ([Format description PDF](#)) This file was last modified 6 hours ago. File size: 251 MB; MD5 sum: [3d325111efcd35d9439256915b43a75c](#).

## Other Formats and Auxiliary Files

- [andalucia-latest.osm.bz2](#), yields OSM XML when decompressed; use for programs that cannot process the .pbf format. [Deprecated](#). This file was last modified 76 days ago. File size: 235 MB; MD5 sum: [ee946f543abe19be385230201dd5998a](#).
- [andalucia/internal.osh.pbf](#) The history file contains personal data and is available on the [internal server](#) only. See notice above for further information.
- [.poly\\_file](#) that describes the extent of this region.
- [experimental vector tile package](#) conforming to [Shortbread schema](#) for use with MapLibre and other MVT capable software
- [.osc.gz files](#) that contain all changes in this region, suitable e.g. for Osmosis updates
- [Taginfo statistics for this region](#)
- [raw directory index](#) allowing you to see and download older files

 Not what you were looking for? Geofabrik is a consulting and software development firm based in Karlsruhe, Germany specializing in OpenStreetMap services. We're happy to help you with data preparation, processing, server setup and the like. [Check out our website](#) and contact us if we can be of service.

## Sub Regions

No sub regions are defined for this region.

 Nicht das Richtige dabei? Die Geofabrik ist ein auf OpenStreetMap spezialisiertes Beratungs- und Softwareentwicklungsunternehmen in Karlsruhe. Gern helfen wir Ihnen bei der Datenaufbereitung, Datenkonvertierung, Serverinstallation und ähnlichen Aufgaben. [Besuchen Sie unsere Webseite](#) und sprechen Sie mit uns, wenn wir Ihnen helfen können.

# Fuentes de datos

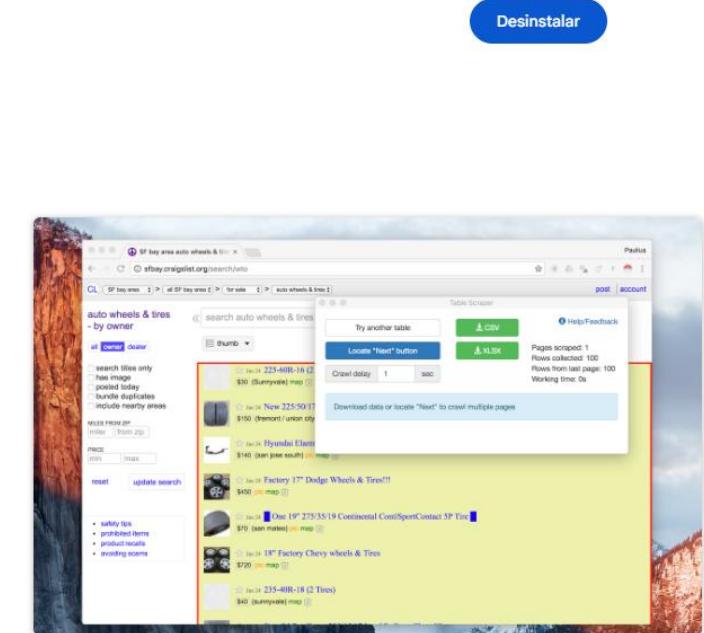
❖ Se puede hacer web scraping de **Google Maps**.

❖ Para ello se puede usar el plugin Instant Data Scraper.

❖ Con este plugin se pueden descargar datos del nombre de un establecimiento, valoración, número de reseñas, etc.

❖ Las coordenadas del sitio están dentro de la url por lo que hará falta sacarlas mediante un script de Python.

❖ [https://github.com/joaquinosorioarjona/curso\\_uned/blob/main/Web%20Scrappling%20Google%20Maps.ipynb](https://github.com/joaquinosorioarjona/curso_uned/blob/main/Web%20Scrappling%20Google%20Maps.ipynb)



## Descripción general

Instant Data Scraper extracts data from web pages and exports it as Excel or CSV files

mercadona - Google Maps

google.com/maps/search/mercadona/@37.3899091,-5.9960435,13z?entry=ttu&g\_ep=EgoYMDI0MTAyNy4wIKXMDSoASAFQAw%3D%3D

Todos los marcadores

mercadona

Compra en tienda · A domicilio

**Unide Market**  
supermercados  
3,8 ★★★★★ (583) ⓘ  
Supermercado · ⚡ Av. Cardenal Tabera, 3  
Abierto · Cierra a las 14:00 · Próxima apertura: 17:30 · 920 34 11 72  
A domicilio

**Sitio web** **Cómo llegar**

**Mercadona**  
4,1 ★★★★★ (2.246) ⓘ  
Supermercado · ⚡ Ctra. Burgohondo, 1  
Abierto · Cierra a las 21:30 · 920 25 08 21  
Compra en tienda · A domicilio

**Sitio web** **Cómo llegar**

**Mercadona**  
4,1 ★★★★★ (2.251) ⓘ  
Supermercado · ⚡ C. Albéniz, 35  
Abierto · Cierra a las 21:30 · 952 26 76 85  
Compra en tienda · A domicilio

**Sitio web** **Cómo llegar**

**Mercadona**  
4,3 ★★★★★ (3.714) ⓘ  
Supermercado · ⚡ Av. Santa Rosa de Lima, 20  
Abierto · Cierra a las 21:30 · 951 30 42 41  
Compra en tienda · A domicilio

**Sitio web** **Cómo llegar**

Actualizar los resultados cuando se mueva el mapa

Horario · Todos los filtros

Buscar en esta zona

Monasterio de San Isidoro del Campo

Río Guadalquivir

A-8006 A-8081

A-8002

A-8005

A-8008

Costco Wholesale Sevilla

Hilton Garden Inn Sevilla

Valdekorras

Aeropuerto de Sevilla

El Pino D

UF D

Buscar en esta zona

Monasterio de San Isidoro del Campo

Kids Park Sevilla

Mirador de Santa Brígida Atalaya

Parque del Alamillo

Carrefour

NORTE

Parque de Miraflores

Fibes - Palacio de Congresos y...

Centro Comercial Los Arcos

SEVILLA ESTE

Real Alcázar de Sevilla

Inundaciones en el sur Andalucía

Plaza de España

Parque de María Luisa

El Cerro del AgUILA

Cerro Amate

Palmete

Leroy Merlin Sevilla Alcalá

San Rafael

Bormujos

Tomares

Castilleja de la Cuesta

Setas de Sevilla

TRIANA

URB. VISTAHERMOSA

San Juan de Aznalfarache

Estadio Benito Villamarín

Hospital Universitario Virgen del Rocío

Universidad Pablo de Olavide Sevilla

Montequinto

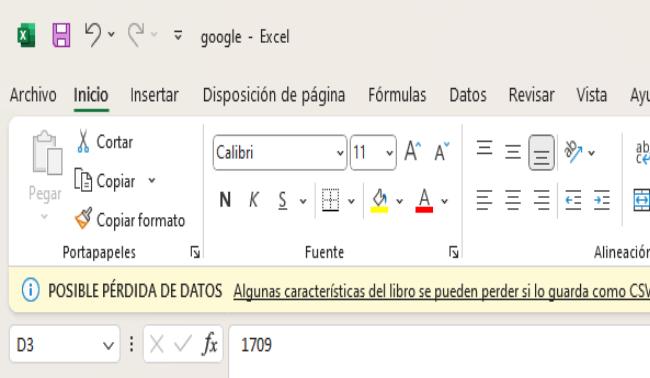
Centro Comercial Los Alcores

Alcalá Guadaíra

Parque Oromana

Capas

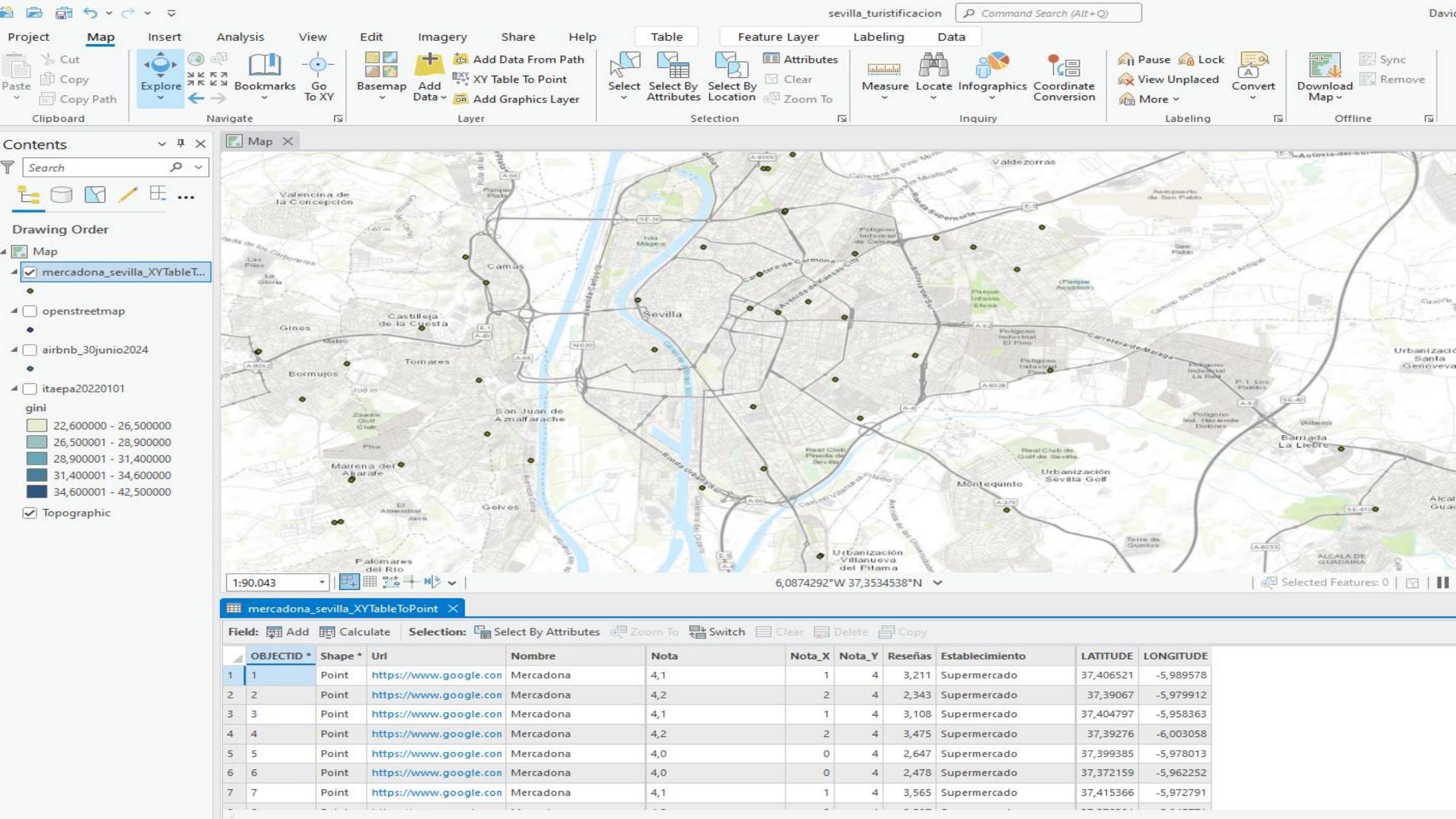
Datos del mapa ©2024 Inst. Geogr. Nacional España Términos Privacidad Envíar comentarios sobre el producto 1 km



```
In [167]: def extract_lat_lon(url):
    if pd.isna(url):
        return np.nan, np.nan
    # Split the URL by '!' and filter out parts that start with '3d' and '4d' for Lat and Lon
    parts = url.split('!')
    lat = None
    lon = None
    for part in parts:
        if part.startswith('3d'):
            lat = part[2:]
        elif part.startswith('4d'):
            lon = part[2:]
    return lat, lon
```

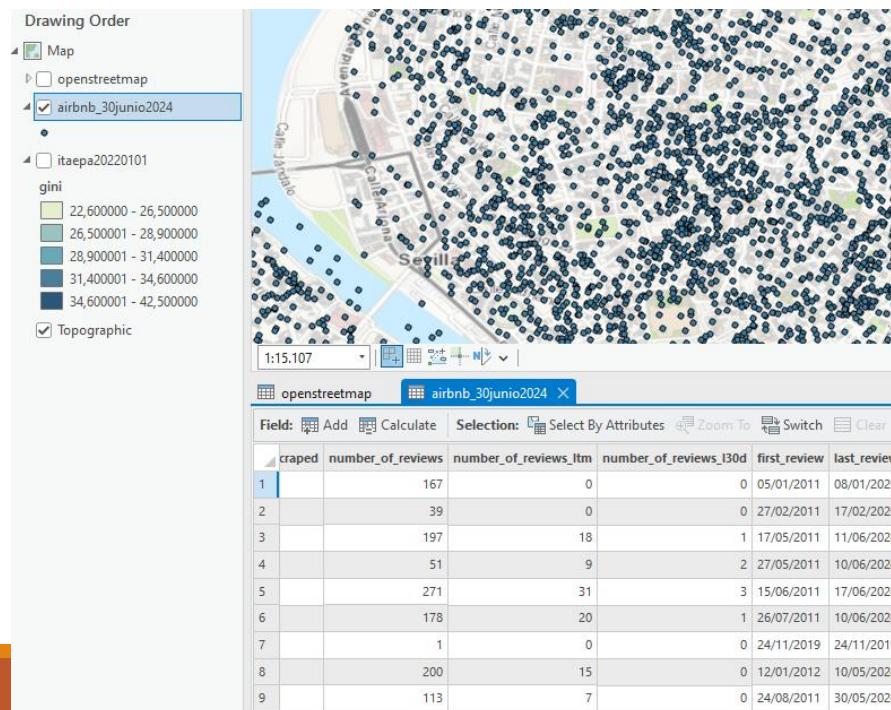
```
In [170]: df[['LATITUDE', 'LONGITUDE']] = df.apply(lambda row: extract_lat_lon(row['Url']), axis=1, result_type='expand')
df
```

	Url	Nombre	Nota	Reseñas	Establecimiento	LATITUDE	LONGITUDE
0	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,1	3.211	Supermercado	37.4065215	-5.9895782
1	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,2	2.343	Supermercado	37.39067	-5.979912
2	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,1	3.108	Supermercado	37.4047967	-5.9583632
3	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,2	3.475	Supermercado	37.39276	-6.0030582
4	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,0	2.647	Supermercado	37.399385	-5.9780131
...	...	...	...	...	...	...	...
59	https://www.google.com/maps/place/Aparcamiento...	Aparcamiento Mercadona	3,0	1.000	Aparcamiento	37.3351812	-6.0655404
60	https://www.google.com/maps/place/Aparcamiento...	Aparcamiento Clientes Supermercado Mercadona	NaN	NaN	Aparcamiento	37.3795278	-6.0811562
61	https://www.google.com/maps/place/Mercadona/da...	Mercadona	4,1	396.000	Supermercado	37.5171649	-5.9858796
62	https://www.google.com/maps/place/Parking+subt...	Parking subterráneo Mercadona	4,5	2.000	Aparcamiento cubierto	37.4266609	-5.9771242
63	https://www.google.com/maps/place/Supermercado...	Supermercado	3,8	29.000	Supermercado	37.3923212	-5.9679441



# Fuentes de datos

- ❖ AirBNB cuenta con un portal donde ofrece la descarga de sus datos de alojamiento en formato de punto.
- ❖ <https://insideairbnb.com/get-the-data/>



Inside Airbnb  
Adding data to the debate

Data About Support Organise Donate!

Seattle

neighbourhoods.csv

Seattle

neighbourhoods.geojson

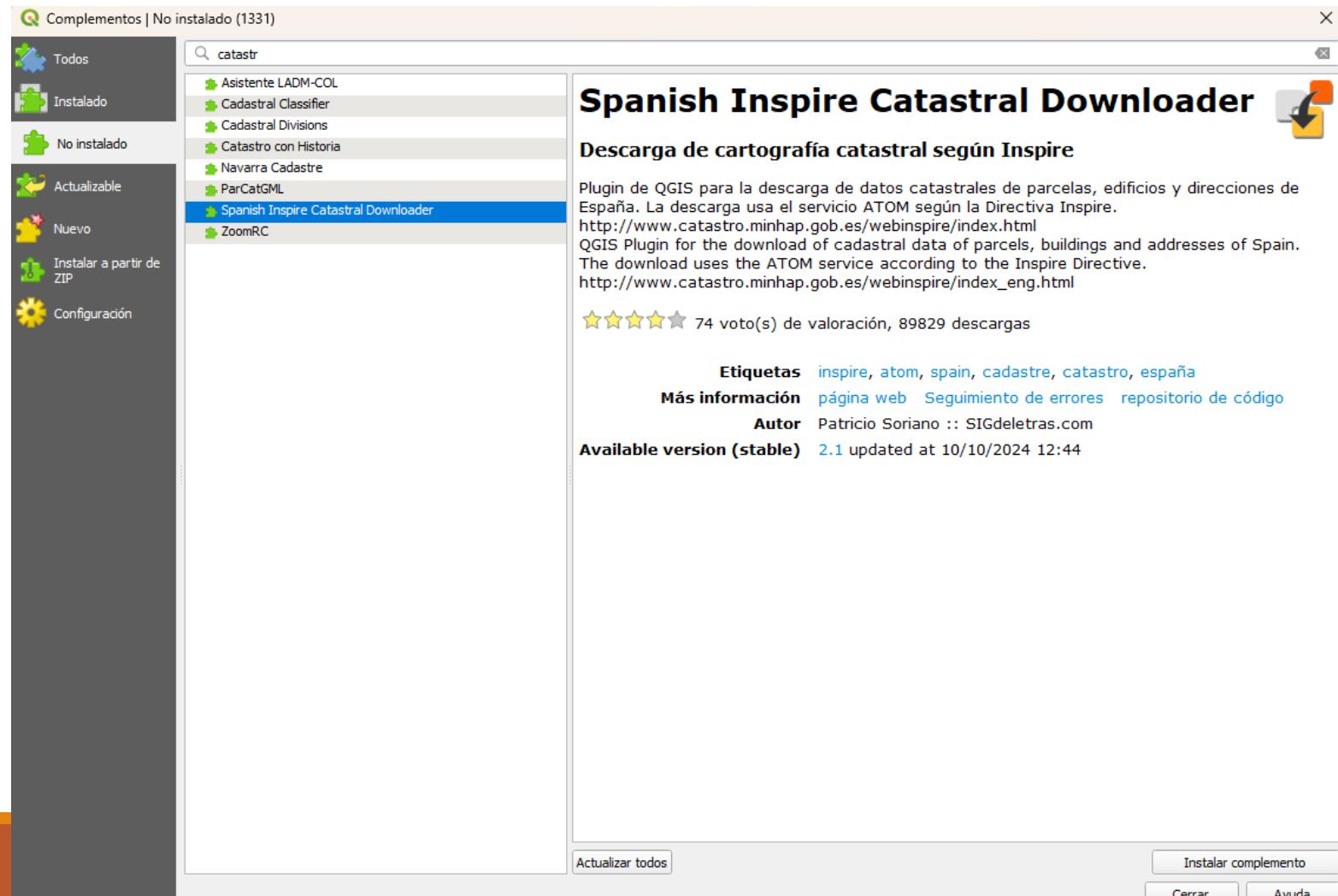
Sevilla, Andalucía, Spain

30 June, 2024 (Explore)

Country/City	File Name
Sevilla	listings.csv.gz
Sevilla	calendar.csv.gz
Sevilla	reviews.csv.gz
Sevilla	listings.csv
Sevilla	reviews.csv
Sevilla	neighbourhoods.csv
Sevilla	neighbourhoods.geojson

# Fuentes de datos

- ❖ En QGIS hay un plugin que permite la descarga de datos del catastro.
- ❖ Estos datos incluyen usos del suelo, antigüedad del edificio, altura o número de pisos.





## Spanish Inspire Catastral Downloader

Provincia

29 - MÁLAGA

 Parcelas catastrales

Municipio

29015 - ANTEQUERA

 Construcciones Direcciones

Carpeta de descarga

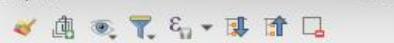
Cerrar

0%

Descargar datos de catastro

Añadir capas al proyecto QGIS

Capas

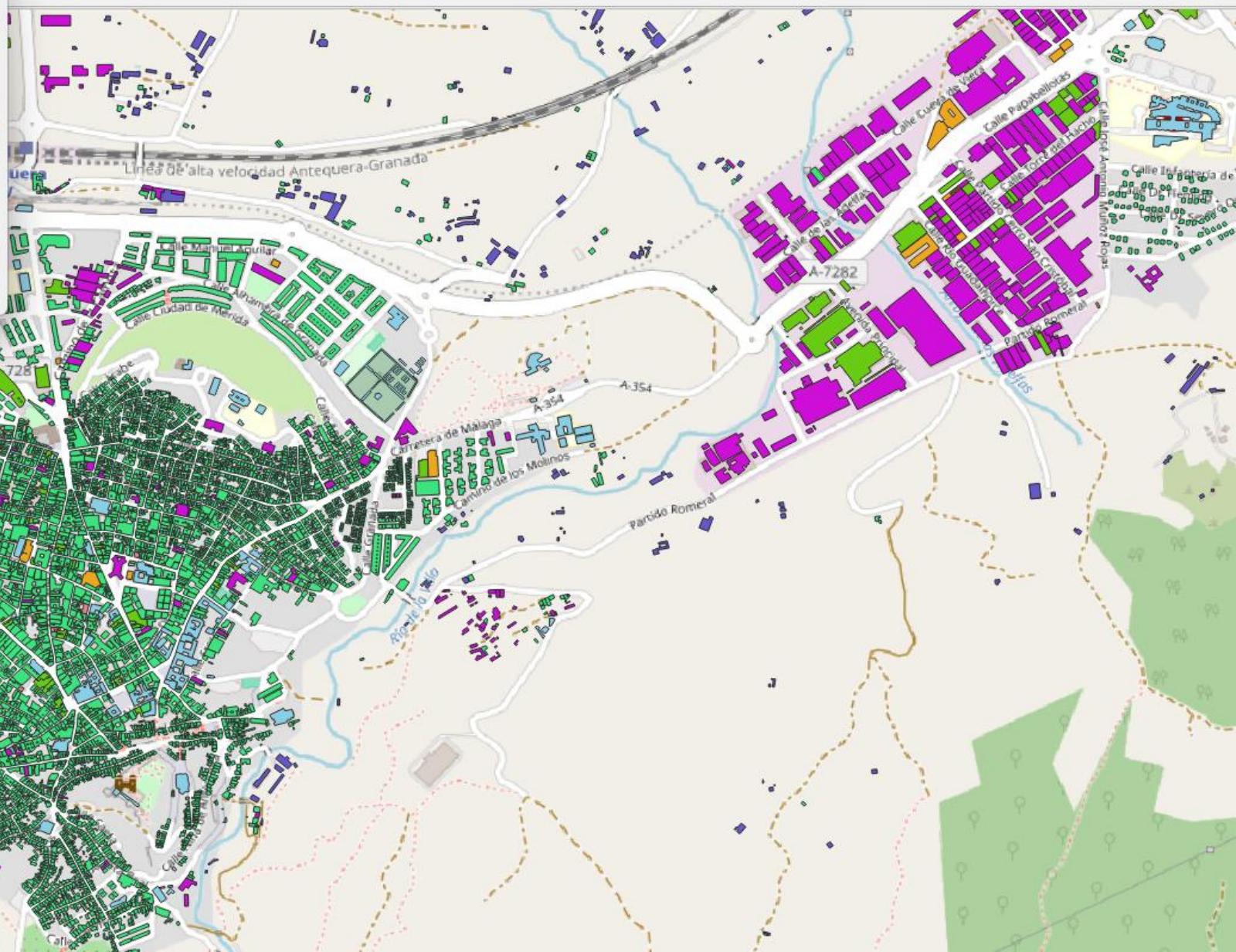


- A.E.S\_SDGC.BU.29015.building — I
  - 1\_residential
  - 2\_agriculture
  - 3\_industrial
  - 4\_1\_office
  - 4\_2\_retail
  - 4\_3\_publicServices
  -
- A.E.S\_SDGC.BU.29015.buildingpar
- A.E.S\_SDGC.AD.29015 — Address
- A.E.S\_SDGC.AD.29015 — AdminUnitN

Editor de vértices

Right click on an editable feature to show its table of vertices.

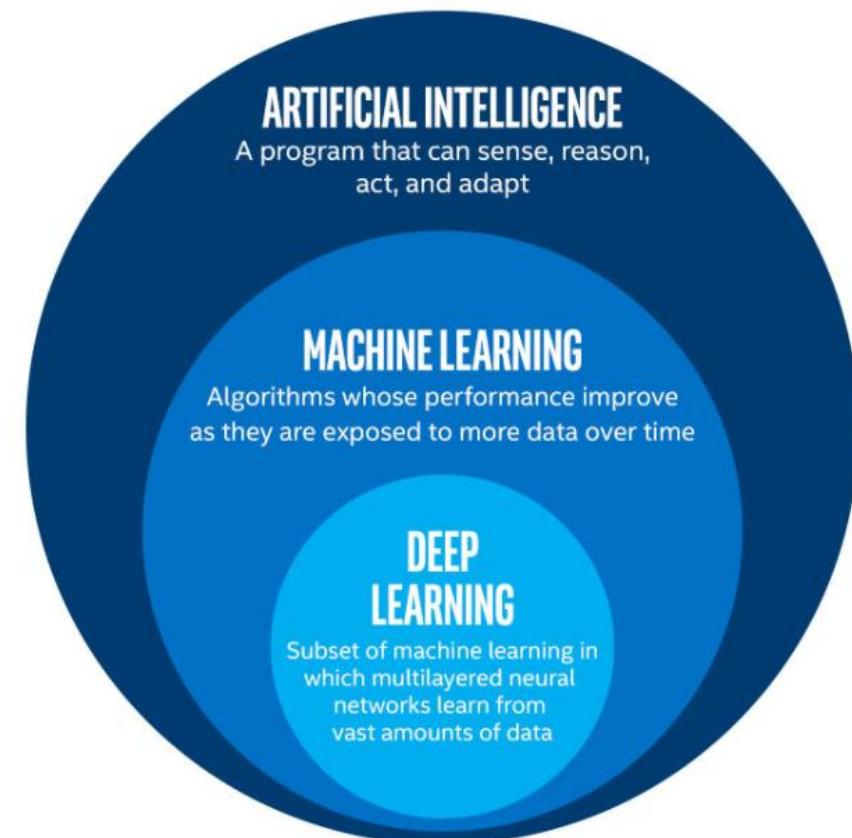
When a feature is bound to this panel, dragging a rectangle to select vertices on the canvas will only select those of the bound feature.



# Fuentes de datos

---

- ❖ El Big Data se está convirtiendo en parte imprescindible de la teledetección.
- ❖ Gran valor para la geografía física.
- ❖ Necesidad de desarrollar algoritmos de Machine Learning y Deep Learning para el tratamiento de datos masivos de teledetección y LIDAR.
- ❖ <https://dataspace.copernicus.eu/>



# Fuentes de datos

2

## DATOS DE OBSERVACIÓN REMOTA: PRINCIPALES PROGRAMAS CON DATOS ABIERTOS

GRATIS



Proveedor	Sensor	Cobertura	Resolución espectral	Resolución espacial	Resolución temporal
NASA	MODIS	2330 Km	Ms: 36 bandas (visible e infrarrojo cercano a térmico)	250, 500 y 1000 m	2 diarias (4 combinada TERRA/AQUA)
NASA	LANDSAT 8, 9	185 Km	Pan y Ms 10 bandas (visible, infrarrojo cercano a térmico)	15 (Pan), 30 y 100 m	16 / 8 días
NASA	ASTER	60 Km	Ms 14 bandas (visible, infrarrojo cercano a térmico)	15, 30, 90 m	16 días
ESA	SENTINEL1	80, 250, 400 Km	Banda C-SAR (polarizaciones VV, HH, VV+VH, HH+HV)	5°5, 5°20, 20°40 m.	6 días
ESA	SENTINEL2	290 Km	Ms 13 bandas (visible e infrarrojo cercano y medio)	10, 20, 60 m	10 / 5 días
ESA	SENTINEL3	1270 Km	Ms 21 bandas (OLCI), SLSTR y SRAL	300 m	1 día

## DATOS DE OBSERVACIÓN REMOTA: PROGRAMA COPERNICUS

**LIBRE Y ABIERTO**

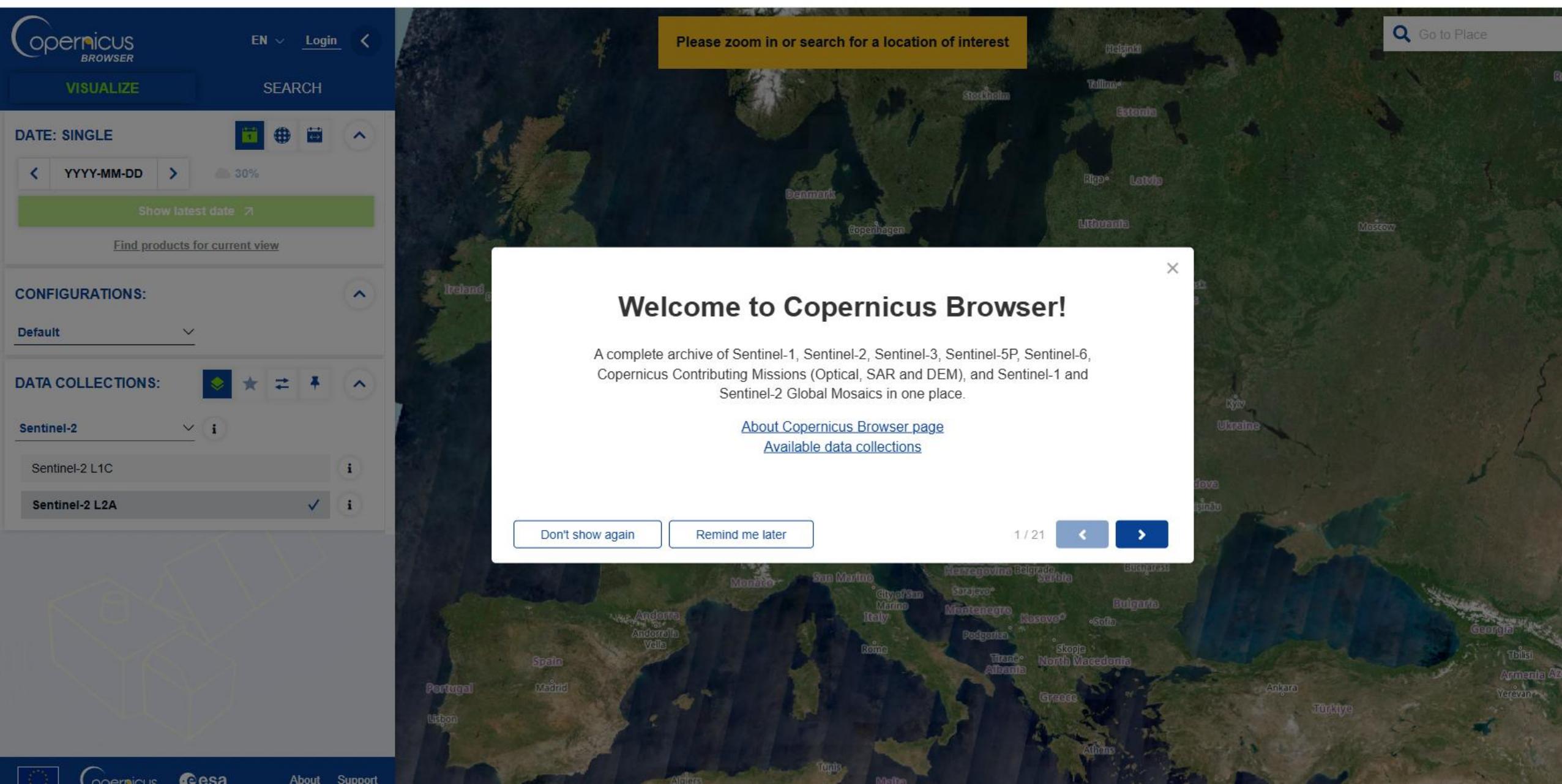
**CONSTELACIÓN SENTINELS**

*Constelación Sentinel. Estado*

Características	SENTINEL-1:	SENTINEL-2:	SENTINEL-3:	SENTINEL-4:	SENTINEL-5p:	SENTINEL-5:	SENTINEL-6:
Órbita polar. Imagen radar también con nubes, de día y de noche	9-40m resolución, 6 días revisita en ecuador SI-A y -B en órbita	10-60m resolución, 5 días revisita S2-A y -B en órbita	300-1200m resolución, <2 días revisita S3-A y -B en órbita	8km resolución, 60 min revisita Primer lanzamiento en 2020	7-68km resolución, 1 día revisita S-5p lanzado 13/10/2017	7.5-50km resolución, 1 día revisita Primer lanzamiento en 2021	10 días revisita Primer lanzamiento en 2020

**Copernicus Satellites**

15 17



# Nuevos estudios de movilidad a partir de datos de telefonía móvil

**INE**  
Instituto Nacional de Estadística

Métodos y proyectos / Nuevos ... / Estadísti... / **Estudios de movilidad a partir de la telefonía móvil**

English

Escriba el texto para buscar

Censo Electoral Sede electrónica Compartir

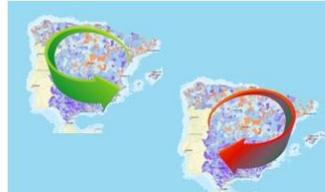
**e**xperimental

**Estudios de movilidad a partir de la telefonía móvil**

En 2019 el INE realizó un primer **estudio de movilidad 2019** a partir de datos de la telefonía móvil dentro de los trabajos preparatorios del Censo de Población y Viviendas 2021.

Con motivo del brote de COVID-19 en marzo de 2020, se ha ido renovando el estudio en distintas etapas. Actualmente el INE está publicando datos de movilidad cada semana, relativos a dos días concretos de la semana anterior (miércoles y domingo). El **estudio de movilidad 2020-2021** se mantendrá previsiblemente durante todo el año 2021.

[Nota de prensa](#) [Proyecto técnico](#)



- ¿Adonde va la gente?
  - ¿De donde viene la gente?
  - ¿Cómo se mueve la gente durante un día laboral?
  - ¿Cómo se mueve la gente durante un fin de semana?
  - ¿Cómo se mueve la gente durante un mes laboral?
  - ¿Cómo se mueve la gente durante el verano?
- Creación de 3216 celdas geográficas en función de la población:
- Municipios agregados
  - Grandes ciudades desagregadas

[https://www.ine.es/experimental/movilidad/experimental\\_em.htm](https://www.ine.es/experimental/movilidad/experimental_em.htm)

## Estudios de movilidad de la población a partir de la telefonía móvil 2020-2021

Este proyecto forma parte de los Estudios de movilidad de la población a partir de la telefonía móvil que el INE viene realizando desde 2019. El objetivo de este estudio es dar continuidad a la serie iniciada en 2019. Actualmente el INE está publicando datos de movilidad cada semana, relativos a dos días concretos de la semana anterior (miércoles y domingo). El estudio de movilidad se mantendrá previsiblemente durante todo el año 2021.

**Movilidad cotidiana:** se ofrecen datos de todos los miércoles y domingos desde el marzo de 2020. Se actualiza cada semana, los jueves no festivos. Además, se ofrecen mapas para algunas fechas concretas.

**Movilidad estacional:** se ofrecen datos de cuatro días concretos de 2020 (18 de julio y 15 de agosto, 22 de noviembre y 25 de diciembre), que se pueden comparar con los días similares de 2019, antes de la aparición del Covid-19 lo que permite una primera aproximación a conocer movimientos turísticos o estacionales. Estos datos se completarán con los correspondientes a fechas similares de 2021.

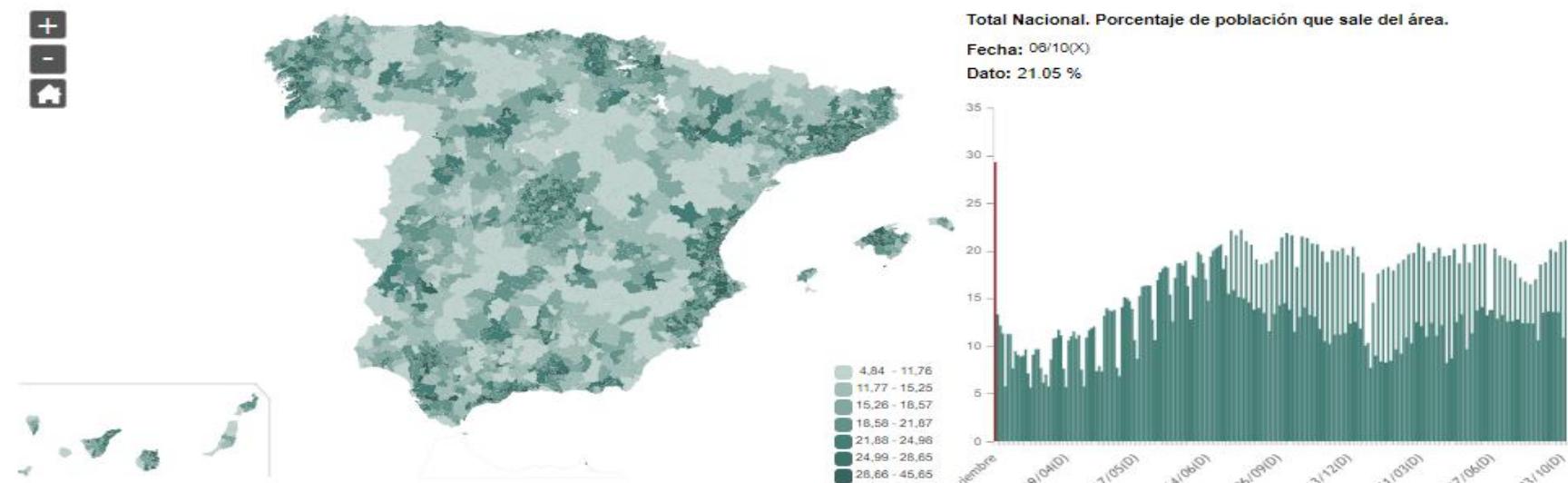


[Nota de prensa](#) [Visor de mapas](#) [Proyecto técnico](#) [Descarga de datos](#)

Selecciona el ámbito geográfico:  Comunidades y Ciudades Autónomas  Provincias  Áreas de movilidad

[Ver Total Nacional](#)

Área: Santanyí (11.76 %)



Pulse en cualquier región del mapa para ver sus datos.

[Descarga de datos](#)

Descripción de las áreas de movilidad y su población

Límites geográficos de las áreas de movilidad formato shape

Movilidad estacional

Movilidad cotidiana

[https://www.ine.es/experimental/movilidad/experimental\\_em.htm](https://www.ine.es/experimental/movilidad/experimental_em.htm)

# Fuentes de datos

---

- ❖ Desde el 1 de enero del año 2022, el Ministerio de Transportes y Movilidad Sostenible comparte datos diarios de telefonía móvil.
- ❖ Estos datos tienen información de la hora en la que fueron generados, información sociodemográfica del usuario, motivo de viaje, modo de transporte, etc.
- ❖ Datos de gran valor para estudios de geografía humana, demografía y población.
- ❖ <https://www.transportes.gob.es/ministerio/proyectos-singulares/estudio-de-movilidad-con-big-data>

# Fuentes de datos

---

- ❖ Desde el 1 de enero del año 2022, el Ministerio de Transportes y Movilidad Sostenible comparte datos diarios de telefonía móvil.
- ❖ Estos datos tienen información de la hora en la que fueron generados, información sociodemográfica del usuario, motivo de viaje, modo de transporte, etc.
- ❖ Datos de gran valor para estudios de geografía humana, demografía y población.
- ❖ <https://www.transportes.gob.es/ministerio/proyectos-singulares/estudio-de-movilidad-con-big-data>

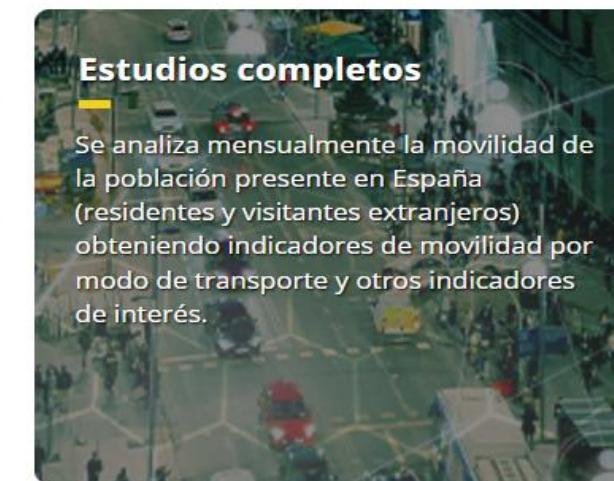


## Estudio de la movilidad con Big Data

Hoy en día, las nuevas tecnologías de Big Data e inteligencia artificial ofrecen unas posibilidades para la medida de la movilidad que eran inimaginables hace solo unos años. El ministerio utiliza desde 2017 estas tecnologías para el estudio de la movilidad a nivel nacional, compartiéndolos como **datos abiertos** para su reutilización.

Los estudios de demanda son una de las piedras angulares del ministerio para la monitorización, evaluación y planificación de mejoras en el Sistema Nacional de Transportes en favor de una **movilidad sostenible**.

Desde el 1 de enero de 2022, se está llevando a cabo un estudio de movilidad diario continuo, mediante la aplicación de tecnologías Big Data y de inteligencia artificial a los registros generados por los terminales de **telefonía móvil**, con el fin de conocer a fondo la movilidad en nuestro país.



### Metodología



El informe metodológico es el documento donde se recoge la descripción de los datos utilizados, la metodología empleada y el alcance del trabajo realizado. Su lectura es aconsejable para una correcta interpretación de los resultados obtenidos en el estudio.

[Ver metodología](#)

# Descarga y almacenamiento de datos

Explorador

mitma-movilidad-v2

Carpetas

Ver todo

7

Mostrar 50 registros

Buscar:

Archivo	↓	Fecha	↑	Tamaño
estudios_basicos				
estudios_completos				
estudios_rutas				
zonificacion				
LEEME_Especificaciones_indicadores_OpenDataMovilidad.xlsx		2024-10-15 08:49:22		123 KB
LICENCIA de datos abiertos del MITMA 20201203.pdf		2022-12-19 14:32:07		158 KB
RSS.xml		2024-10-24 12:21:52		4 MB

Mostrando registros del 1 al 7 de un total de 7 registros

Anterior

1

Siguiente

# Descarga y almacenamiento de datos

- ❖ Necesitaremos primero el shapefile de zonificación y los datos de población de cada distrito. Iremos a la carpeta de zonificación para ello.
- ❖ Descargaremos tanto la capa de distritos como la de centroides que usaremos luego.
- ❖ Si queremos hacer matrices OD para cartografiar flujos de movilidad, se crearán a partir de la capa de centroides.

 nombres_distritos.csv
 poblacion_distritos.csv
 zonificacion_distritos.cpg
 zonificacion_distritos.dbf
 zonificacion_distritos.prj
 zonificacion_distritos.qmd
 zonificacion_distritos.qpj
 zonificacion_distritos.shp
 zonificacion_distritos.shx
 zonificacion_distritos_centroides.cpg
 zonificacion_distritos_centroides.dbf
 zonificacion_distritos_centroides.prj
 zonificacion_distritos_centroides.qmd
 zonificacion_distritos_centroides.qpj
 zonificacion_distritos_centroides.shp



PostgreSQL

- datos\_telefonia
- metro\_sevilla
- postgres
- valle\_guadalquivir
- zonificaciones

SAP HANA

MS SQL Server

Oracle

WMS/WMTS

Vector Tiles

XYZ Tiles

Manzen\_Global\_Terrain

Capas

zonicion\_distritos\_centroides

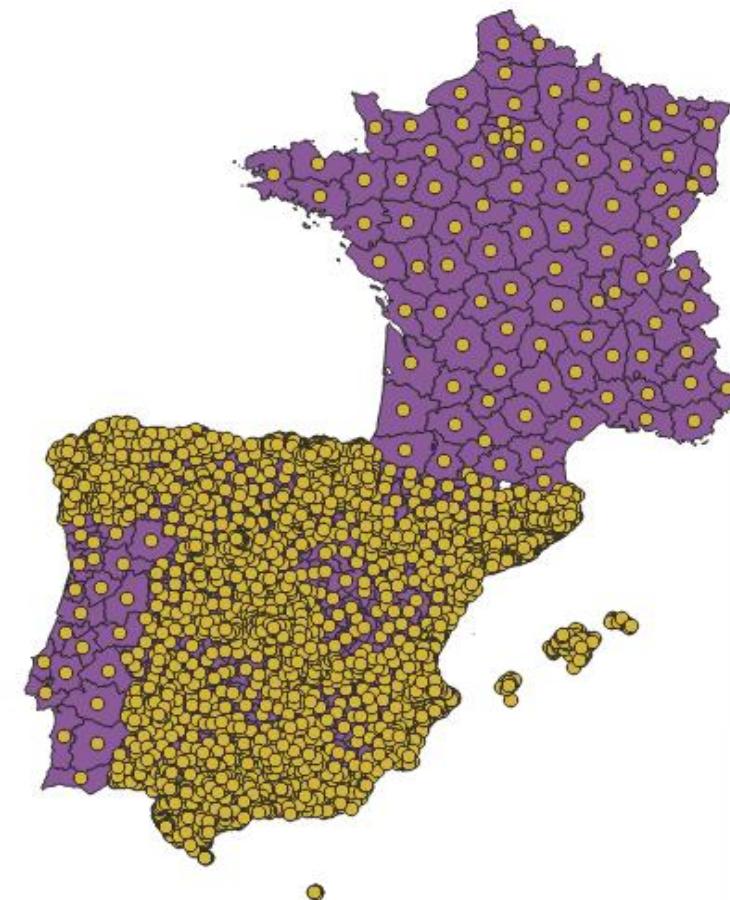
zonicion\_distritos

poblacion

Editor de vértices

Right click on an editable feature to show its table of vertices.

When a feature is bound to this panel, dragging a rectangle to select vertices on the canvas will only select those of the bound feature.

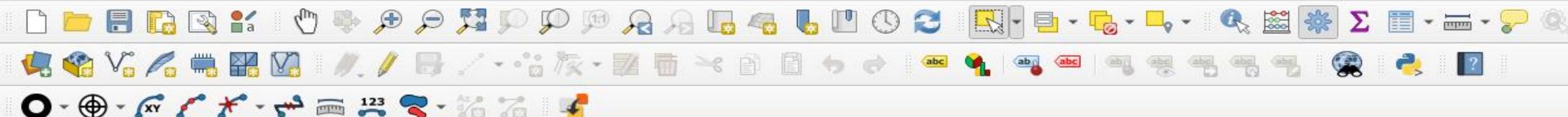


zonificacion\_distritos— Objetos Totales: 3909, Filtrados: 3909

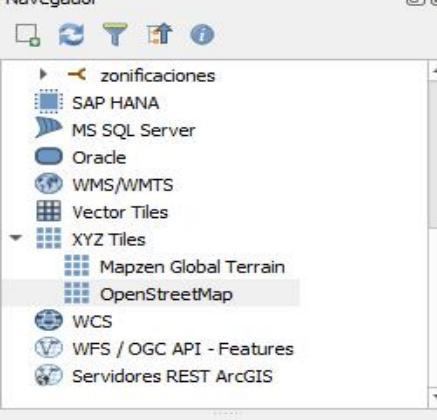
ID

ID
1 01001
2 01002
3 01004_AM
4 01009_AM
5 01010
6 01017_AM
7 01028_AM
8 01036
9 01043
10 01047_AM
11 01051
12 01054_AM
13 01058_AM
14 0105901
15 0105902

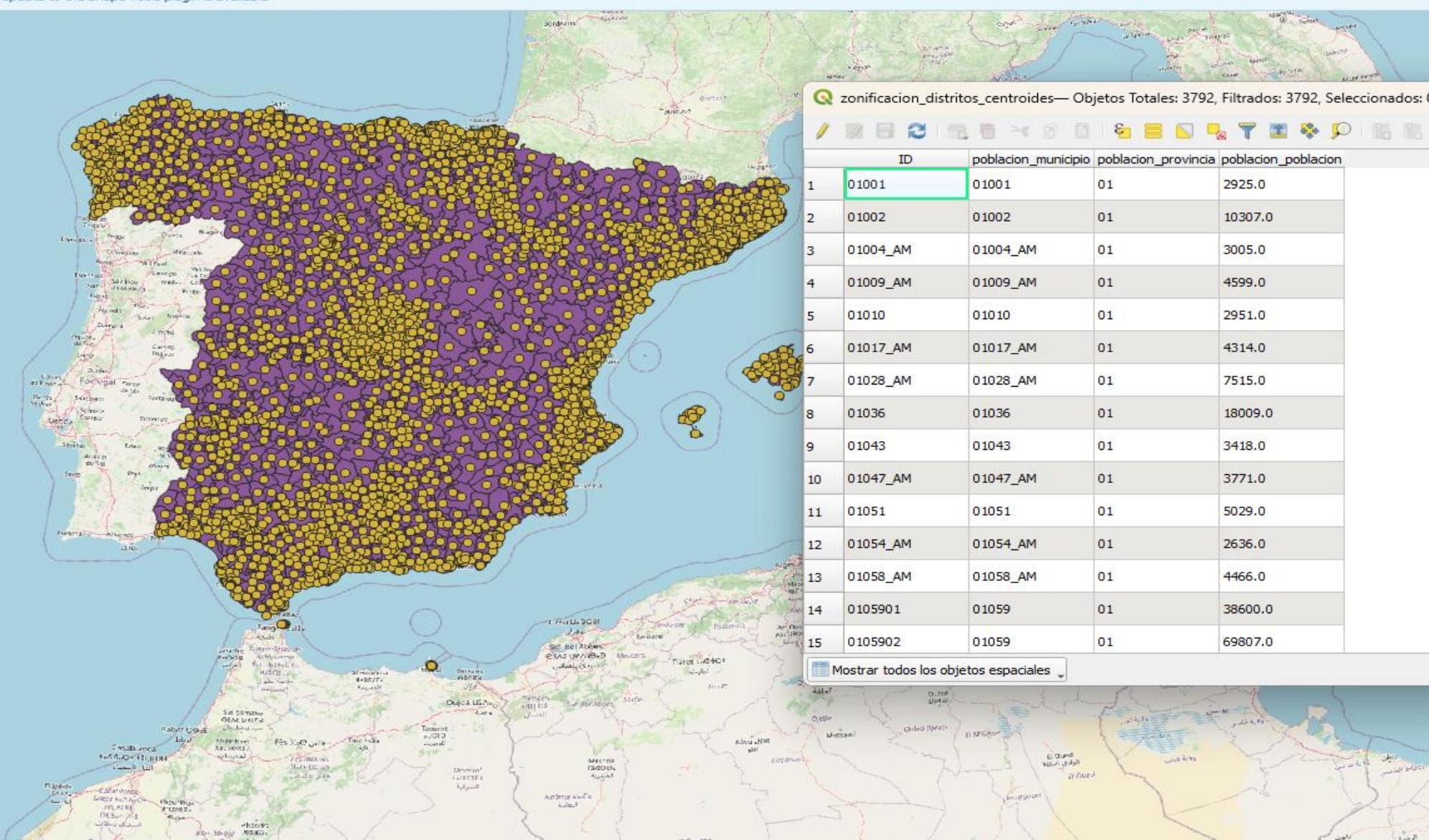
Mostrar todos los objetos espaciales



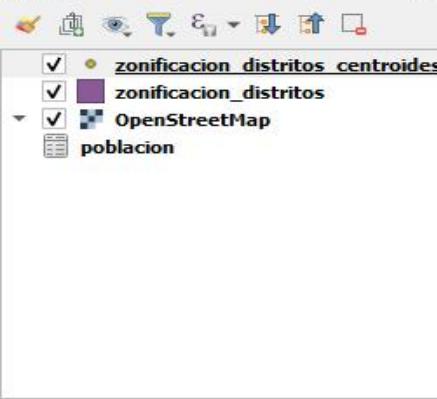
Navegador



An update to the Shape Tools plugin is available.



Capas



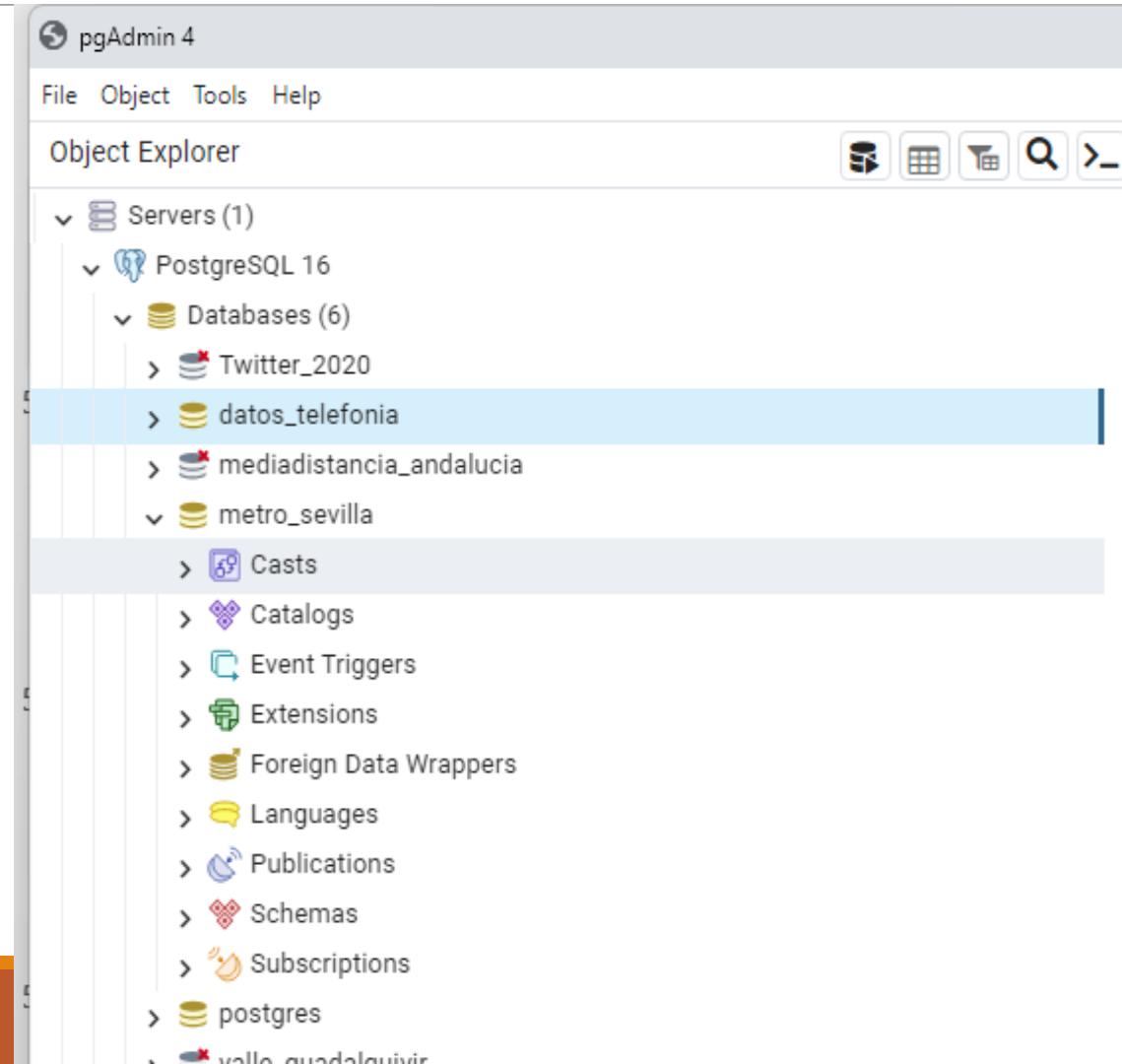
## Editor de vértices

Right click on an editable feature to show its table of vertices.

When a feature is bound to this panel, dragging a rectangle to select vertices on the canvas will only select those of the bound feature.

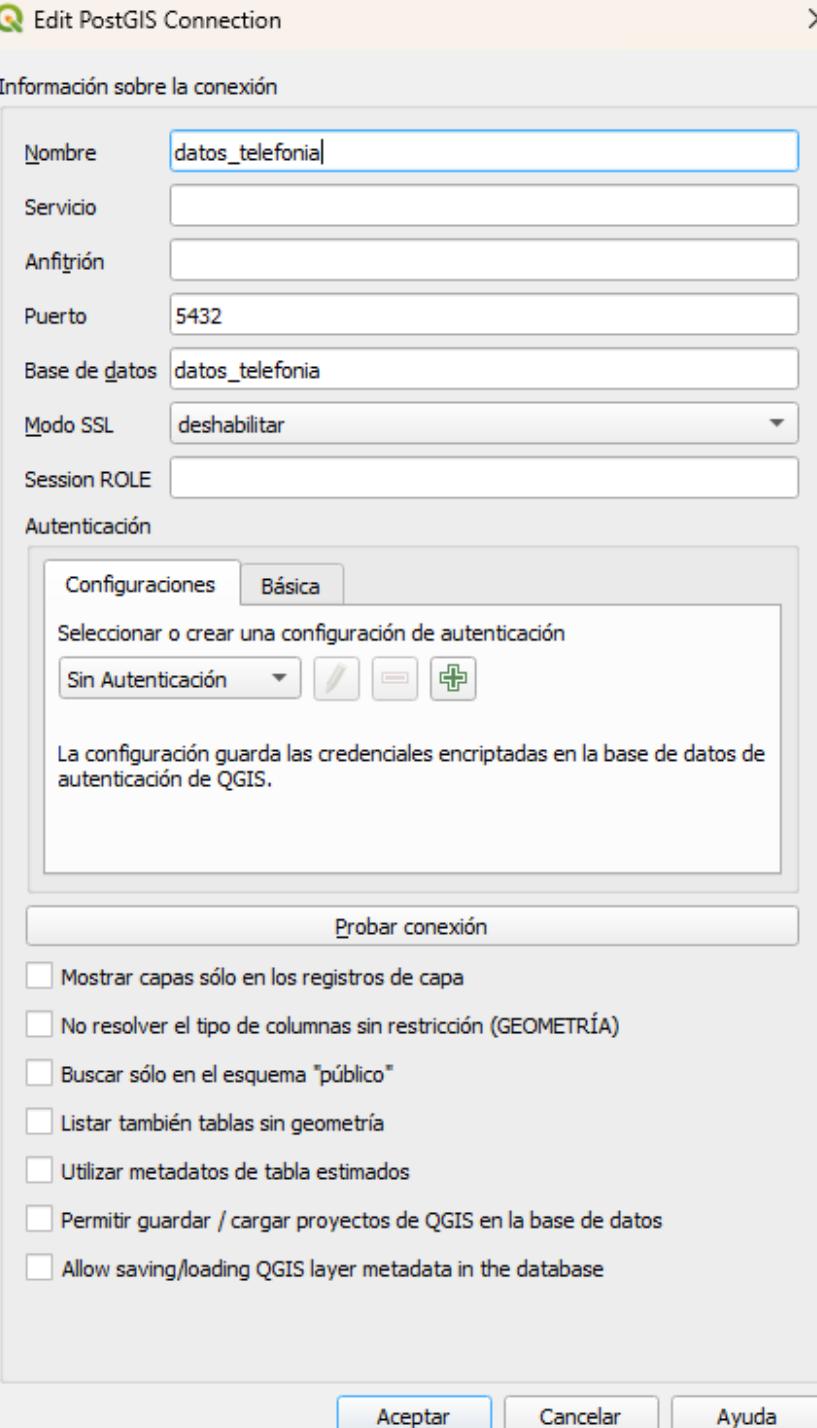
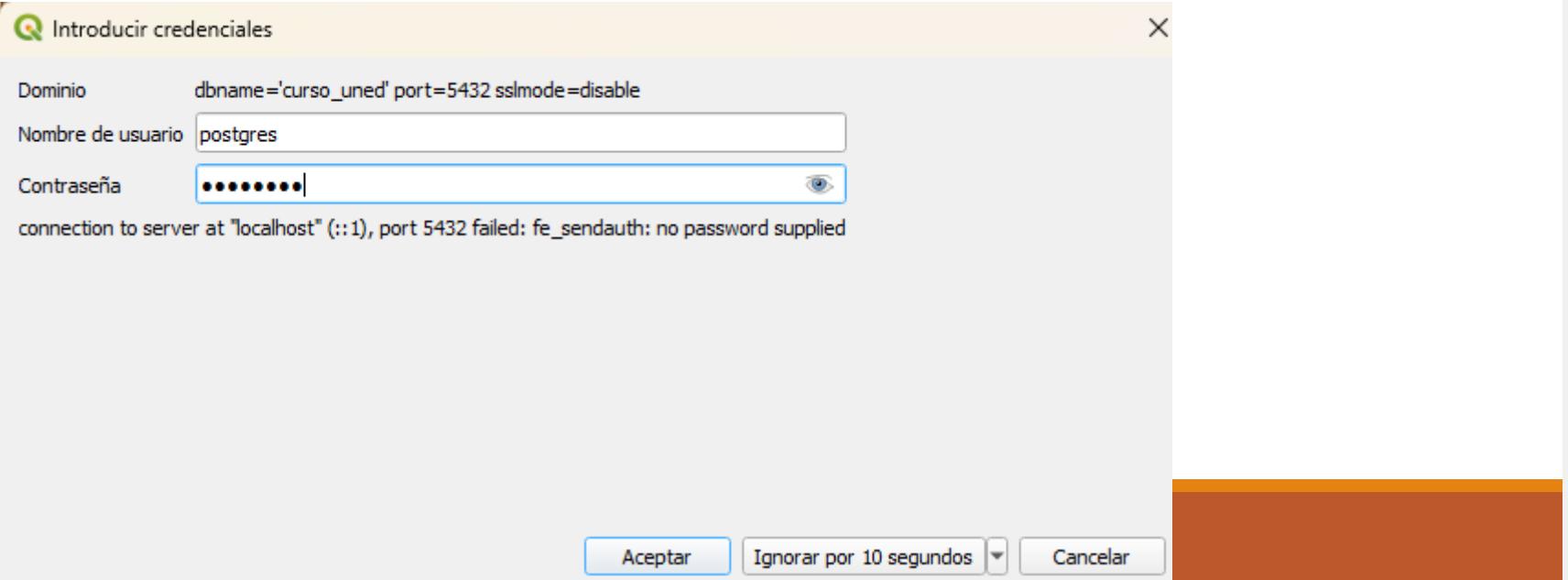
# Descarga y almacenamiento de datos

- ❖ Tras unir los datos de población a la capa de distritos, vamos a ir almacenando los datos en una base de datos de PostGIS llamada datos\_telefonia
- ❖ Cuando instaléis PostgreSQL, no olvidéis instalar la extensión de PostGIS ni el nombre de usuario (postgres o localhost por defecto) ni la contraseña (postgres por defecto).
- ❖ No olvidar escribir “create extensión postgis;” al crear la tabla.
- ❖ Si no está activado en QGIS, hay que activar el complemento DB Manager, para poder importar de QGIS a PostGIS nuestros datos.



# Descarga y almacenamiento de datos

- ❖ Antes de usar DB Manager hay que crear y configurar una nueva conexión de PostGIS con QGIS, de modo que nuestra base de datos quede registrada.



The screenshot shows the QGIS interface with a database schema viewer open. The schema viewer displays the 'public' schema from the 'curso\_uned' database. It includes sections for 'Info', 'Tabla', and 'Vista preliminar'. The 'Info' tab shows the schema details: Propietario: pg\_database\_owner and Comentario: standard public schema. The 'Tabla' tab shows the tables: geography\_columns, geometry\_columns, spatial\_ref\_sys, datos\_telefonia, metro\_sevilla, postgres, valle\_guadalquivir, and zonificaciones. The 'Vista preliminar' tab shows a preview of the 'public' schema.

**Importar capa vectorial**

Entrada: zonificacion\_distritos

Importar sólo objetos espaciales seleccionados

Tabla de salida:

Esquema: public

Tabla: distritos

Opciones:

- Clave primaria: id
- Columna de geometría: geom
- SRID de origen: EPSG:3042 - ETRS89 / UTM zone 30N (N-E)
- SRID de destino: EPSG:3042 - ETRS89 / UTM zone 30N (N-E)
- Codificación: Automática
- Sustituir la tabla de destino (si existe)
- No estimular a multi parte
- Pasar nombres de campos a minúsculas
- Crear índice espacial
- Comentario:

Aceptar Cancelar



Servers (1)



Databases (7)



Twitter\_2020



curso\_uned



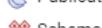
Casts



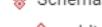
Catalogs



Event Triggers



Extensions



Foreign Data Wrappers



Languages



Publications



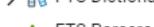
Schemas (1)



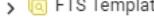
public



Aggregates



Collations



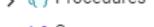
Domains



FTS Configurations



FTS Dictionaries



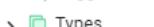
FTS Parsers



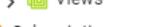
FTS Templates



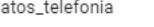
Foreign Tables



Functions



Materialized Views



Operators



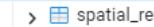
Procedures



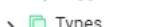
Sequences



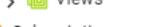
Tables (2)



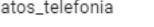
distritos



spatial\_ref\_sys



Trigger Functions



Types



Views



Subscriptions



datos\_telefonia



mediadistancia\_andalucia



metro\_sevilla



Casts



Catalogs



```
1 SELECT * FROM public.distritos
2 ORDER BY id ASC
```

You are currently running version 7.6 of pgAdmin

however the current version is 8.6.

Scratch pane

Please click here for more information.



		poblacion_municipio character varying	poblacion_provincia character varying	poblacion_poblacion character varying
1		01001	01	2925.0
2		01002	01	10307.0
3		01004_AM	01	3005.0
4		01009_AM	01	4599.0
5		01010	01	2951.0
6		01017_AM	01	4314.0
7		01028_AM	01	7515.0
8		01036	01	18009.0
9		01043	01	3418.0
10		01047_AM	01	3771.0
11		01051	01	5029.0
12		01054_AM	01	2636.0
13		01058_AM	01	4466.0
14		01059	01	38600.0
15		01059	01	69807.0
16		01059	01	54114.0
17		01059	01	33042.0
18		01059	01	52432.0
19		01059	01	5098.0

# Descarga y almacenamiento de datos

- ❖ Estudios básicos: personas, pernoctaciones y viajes.
- ❖ Estudios completos: solo están disponibles de forma abierta ficheros de movilidad obligada (número de personas que hacen un número de viajes en un periodo de dos semanas).
- ❖ Estudios de rutas: viajes agregados a la red de carreteras de España.
- ❖ Zonificación: shapefiles de unidades espaciales.

Archivo	↓
estudios_basicos	
estudios_completos	
estudios_rutas	
zonificacion	

# Descarga y almacenamiento de datos

- ❖ Los datos están disponibles en tres zonificaciones.
- ❖ Por distritos: la que vamos a usar por defecto. 3909 distritos con población homogénea (los municipios poco poblados se unen en un distrito mayor, y las ciudades muy pobladas son divididas en distritos o barrios).
- ❖ Por municipios.
- ❖ Por GAU: misma zonificación que municipios, pero los municipios pertenecientes a un área metropolitana son agregados en una sola unidad.

Explorador mitma-movilidad-v2 / estudios\_basicos

Mostrar 50 registros

Archivo

↑ [Volver a la carpeta padre]

📁 calidad

📁 por-distritos

📁 por-GAU

📁 por-municipios

Mostrando registros del 1 al 4 de un total de 4 registros

A screenshot of a file explorer interface. At the top, it says 'Explorador mitma-movilidad-v2 / estudios\_basicos'. Below that is a search bar with 'Mostrar 50 registros'. The main area shows a tree view of files and folders under 'Archivo'. The visible items are: 'calidad', 'por-distritos', 'por-GAU', and 'por-municipios'. There are arrows at the top right of the tree view. At the bottom, it says 'Mostrando registros del 1 al 4 de un total de 4 registros'.

# Descarga y almacenamiento de datos

- ❖ Pernoctaciones: número de personas que han pernoctado en un distrito y el distrito de residencia de esas personas.
- ❖ Personas: número de personas pernoctando en un distrito, con datos sociodemográficos básicos (edad (agregada en franjas de edad) y sexo).
- ❖ Viajes: número de personas que viajan de un distrito a otro. Los datos incluyen periodo horario, distancia, actividad en la zona de origen (casa, trabajo o estudio, actividad frecuente, actividad no frecuente), actividad en la zona de destino, provincia de residencia, renta (agregada en tres grupos) edad y sexo.

The screenshot shows a data exploration interface with a blue header bar. The header contains the text "Explorador" and a breadcrumb navigation path: "mitma-movilidad-v2 / estudios\_basicos / por-distritos". Below the header, there is a dropdown menu labeled "Mostrar 50 registros". The main content area is titled "Archivo" and contains three entries, each represented by a folder icon:

- [Volver a la carpeta padre]
- pernoctaciones
- personas
- viajes

At the bottom of the content area, a footer message reads "Mostrando registros del 1 al 3 de un total de 3 registros".

# Descarga y almacenamiento de datos

---

- ❖ Los estudios completos deberían ofrecer los datos de viajes completos: viajes de estudios básicos con la siguiente información agregada: día de la semana (indicando si es día normal o especial), modo de transporte (carretera, ferroviario, aéreo o marítimo), medio de transporte (ferrocarril convencional, AVE, autobús o vehículo privado), distancia del viaje, duración del viaje, y país de residencia del viajero (España, Alemania, Francia, Italia, Reino Unido, Portugal, Bélgica, Países Bajos, Países Nórdicos, Resto de Europa o Resto del Mundo).
- ❖ Estos datos pueden estar en etapas (cada etapa de un mismo trayecto es una fila) o viajes completos (todo el trayecto es una fila y se incluye un campo de itinerario, un campo de número de transbordos, un campo de modo de transporte principal, un campo de medio transporte principal y dos campos de secuencia ordenada de modos y medios de las etapas).
- ❖ Sin embargo, según el Ministerio, estos datos solo están para uso interno.

# Descarga y almacenamiento de datos

---

- ❖ Vamos a descargar ahora los datos de viajes con la zonificación por distritos.
- ❖ Los datos se ofrecen para meses completos y para días individuales.
- ❖ Los datos mensuales de pernoctaciones y personas no son pesados, pero los datos mensuales de viajes son muy pesados debido a su riqueza.
- ❖ Para este taller cogeremos los ficheros del 15 de enero y el 15 de agosto de 2023 (¡cada día individual pesa más de 1 GB!).

<a href="#"> 20230112_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:21	189 MB
<a href="#"> 20230113_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:29	195 MB
<a href="#"> 20230114_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:35	172 MB
<a href="#"> 20230115_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:39	135 MB
<a href="#"> 20230116_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:45	170 MB
<a href="#"> 20230117_Viajes_distritos.csv.gz</a>	2023-02-08 11:44:54	171 MB
<a href="#"> 20230118_Viajes_distritos.csv.gz</a>	2023-02-08 11:45:01	170 MB
<a href="#"> 20230119_Viajes_distritos.csv.gz</a>	2023-02-08 11:45:13	163 MB
<a href="#"> 20230120_Viajes_distritos.csv.gz</a>	2023-02-08 11:45:25	169 MB



	fecha	periodo	origen	destino	distancia	actividad_origen	actividad_destino	studio_origen_posib	tudio_destino_posib	residencia	renta	edad	sexo	viajes	viajes_km
1	20230115	00	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	3.243	12.651
2	20230115	00	01001	01001	2-10	trabajo_estudio	casa	no	no	01	10-15	NA	NA	3.271	13.547
3	20230115	01	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	2.311	8.141
4	20230115	02	01001	01001	2-10	trabajo_estudio	casa	no	no	01	10-15	NA	NA	6.507	25.611
5	20230115	03	01001	01001	2-10	no_frecuente	no_frecuente	no	no	48	>15	25-45	hombre	3.716	16.704
6	20230115	05	01001	01001	2-10	casa	no_frecuente	no	no	01	10-15	NA	NA	6.059	22.536
7	20230115	06	01001	01001	0.5-2	casa	frecuente	no	no	01	10-15	NA	NA	2.732	4.469
8	20230115	06	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	6.507	29.196
9	20230115	07	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	2.732	10.717
10	20230115	08	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	12.482	51.853
11	20230115	08	01001	01001	2-10	casa	trabajo_estudio	no	no	01	10-15	NA	NA	2.732	10.199
12	20230115	08	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	2.732	10.717
13	20230115	09	01001	01001	2-10	casa	trabajo_estudio	no	no	01	10-15	NA	NA	6.507	25.611
14	20230115	09	01001	01001	2-10	no_frecuente	no_frecuente	no	no	48	>15	25-45	hombre	3.716	16.704
15	20230115	10	01001	01001	2-10	casa	trabajo_estudio	no	no	01	10-15	NA	NA	6.057	24.682
16	20230115	10	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	8.835	38.474
17	20230115	11	01001	01001	2-10	casa	no_frecuente	no	no	01	10-15	NA	NA	6.507	24.536
18	20230115	11	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	5.097	19.94
19	20230115	11	01001	01001	2-10	trabajo_estudio	casa	no	no	01	10-15	NA	NA	5.518	21.334
20	20230115	12	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	15.352	62.814
21	20230115	12	01001	01001	2-10	casa	trabajo_estudio	no	no	01	10-15	NA	NA	2.786	12.34
22	20230115	12	01001	01001	2-10	frecuente	trabajo_estudio	no	no	01	>15	45-65	hombre	3.192	14.218
23	20230115	12	01001	01001	2-10	no_frecuente	casa	no	no	01	10-15	NA	NA	6.507	24.536
24	20230115	13	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	5.114	21.077
25	20230115	13	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	6.059	25.169
26	20230115	13	01001	01001	2-10	no_frecuente	casa	no	no	01	10-15	NA	NA	6.059	22.536
27	20230115	14	01001	01001	2-10	casa	frecuente	no	no	01	10-15	NA	NA	2.786	10.035
28	20230115	14	01001	01001	2-10	casa	trabajo_estudio	no	no	01	10-15	NA	NA	2.732	10.199
29	20230115	14	01001	01001	2-10	frecuente	casa	no	no	01	10-15	NA	NA	11.621	46.86

# Procesado de datos

---

- ❖ Vamos a procesar la tabla en Jupyter.
- ❖ Eliminamos los datos NA de renta y sexo
- ❖ Vamos a transformar el valor numérico de la provincia de residencia en el nombre de la provincia
- ❖ Vamos a añadir el nombre del municipio y de la Gran Área Urbana al distrito tanto de origen como de destino, para facilitar la identificación y poder hacer consultas espaciales más precisas
- ❖ Vamos también a darle un formato fecha al campo correspondiente (añadiendo también la hora) y crear además campo de día y mes

# Procesado de datos

- ❖ Lo primero que vamos a hacer antes del procesado en Jupyter es unir en QGIS los nombres de municipio y GAU a la capa de centroides.
- ❖ [https://github.com/joaquinosorioarjona/curso\\_uned/blob/main/nombr es%20unidades%20espaciales.csv](https://github.com/joaquinosorioarjona/curso_uned/blob/main/nombr es%20unidades%20espaciales.csv)
- ❖ Una vez hecha la unión, exportamos la capa de centroides a PostGIS.

zonificación\_distritos\_centroides— Objetos Totales: 3792, Filtrados: 3792, Seleccionados: 0

ID	diccionario_municipio	diccionario_provincia	diccionario_gau	diccionario_nombre_provincia	diccionario_nombre_municipio
1	01001	01001	1	01001	Álava
2	01002	01002	1	01002	Álava
3	01004_AM	01004_AM	1	01004_AM	Álava
4	01009_AM	01009_AM	1	01009_AM	Álava
5	01010	01010	1	01010	Álava
6	01017_AM	01017_AM	1	01017_AM	Álava
7	01028_AM	01028_AM	1	01028_AM	Álava
8	01036	01036	1	01036	Álava
9	01043	01043	1	01043	Álava
10	01047_AM	01047_AM	1	01047_AM	Álava
11	01051	01051	1	01051	Álava
12	01054_AM	01054_AM	1	01054_AM	Álava
13	01058_AM	01058_AM	1	01058_AM	Álava
14	0105901	01059	1	GAU Vitoria/Gasteiz	Álava
15	0105902	01059	1	GAU Vitoria/Gasteiz	Álava
16	0105903	01059	1	GAU Vitoria/Gasteiz	Álava
17	0105904	01059	1	GAU Vitoria/Gasteiz	Álava
18	0105905	01059	1	GAU Vitoria/Gasteiz	Álava
19	0105906	01059	1	GAU Vitoria/Gasteiz	Álava
20	01063_AM	01063_AM	1	01063_AM	Álava
21	01901	01901	1	01901	Álava
22	0200301	02003	2	GAU Albacete	Albacete
23	0200302	02003	2	GAU Albacete	Albacete
24	0200303	02003	2	GAU Albacete	Albacete
25	0200304	02003	2	GAU Albacete	Albacete

# Procesado de datos

---

- ❖ Para el procesado y enriquecimiento de nuestra tabla de datos de telefonía móvil vamos a necesitar un script de Python.
- ❖ Vamos a usar ChatGIS para ir creando este script.
- ❖ El consejo es ir paso a paso para obtener el script final (si ya tenemos un script de partida, el proceso es más fluido).

Tengo en PostGIS una capa de centroides almacenada donde cada centroide tiene el nombre del distrito, del municipio y de la GAU. Editame el script anterior para incorporar en la tabla csv de datos de telefonía móvil un campo de nombre de distrito de origen, nombre de distrito de destino, municipio de origen, municipio de destino, GAU de origen y GAU de destino.



Para incorporar los nombres de distrito, municipio y GAU a los datos de telefonía móvil desde una capa de centroides en PostGIS, puedes realizar una consulta espacial en PostgreSQL y luego cargar los resultados en el DataFrame de Pandas. Esto implica hacer un *join* espacial entre los datos de telefonía móvil (con los IDs de distrito de origen y destino) y la capa de centroides en PostGIS, y luego importar los resultados en Pandas.

Voy a dividir el proceso en dos partes:

1. **Paso 1:** Ejecutar una consulta SQL en PostGIS para unir los datos de telefonía móvil con la capa de centroides y obtener la información adicional de nombres de distrito, municipio y GAU.
2. **Paso 2:** Cargar los resultados en un DataFrame de Pandas y continuar con el procesamiento en Python.

### Paso 1: Consulta SQL en PostGIS

Aquí tienes una consulta SQL que asume que tienes dos tablas en tu base de datos:

- `telefonia_movil`: la tabla de datos de telefonía móvil, con columnas `id_distrito_origen` y `id_distrito_destino`.
- `centroides`: la capa de centroides con columnas `id_distrito`, `nombre_distrito`, `nombre_municipio` y `gau`.

La consulta enlazará cada ID de distrito en la tabla `telefonia_movil` con la información de la tabla `centroides`.

```
sql Copiar código
SELECT
    tm.*,
    co.nombre_distrito AS nombre_distrito_origen,
    co.nombre_municipio AS municipio_origen,
    co.gau AS gau_origen,
    cd.nombre_distrito AS nombre_distrito_destino,
    cd.nombre_municipio AS municipio_destino,
    cd.gau AS gau_destino
FROM
    telefonia_movil AS tm
LEFT JOIN
    centroides AS co ON tm.id_distrito_origen = co.id_distrito
LEFT JOIN
    centroides AS cd ON tm.id_distrito_destino = cd.id_distrito;
```

La tabla de datos de telefonía móvil no la tengo en PostGIS. Quiero hacer la unión en el script de Jupyter.



Si deseas hacer la unión en el script de Jupyter y tienes los datos de centroides en una tabla de PostGIS, puedes realizar lo siguiente:

1. Cargar los datos de telefonía móvil en un DataFrame de Pandas desde el CSV.
2. Conectar a la base de datos de PostGIS desde Jupyter usando `sqlalchemy` y `geopandas` para cargar la capa de centroides.
3. Hacer la unión de los datos de Pandas (datos de telefonía móvil) con los datos de PostGIS (centroides) en el propio script de Python.

Aquí tienes un script que implementa estos pasos:

### Paso 1: Conectar a PostGIS y cargar los datos de centroides

Primero, necesitas instalar las librerías necesarias si aún no las tienes:

bash

```
pip install sqlalchemy psycopg2-binary geopandas Copiar código
```

### Paso 2: Escribir el script en Jupyter

A continuación, este es el script para realizar la conexión, cargar los datos de PostGIS y hacer la unión:

python

```
# Importamos Las Librerias necesarias
import pandas as pd
import geopandas as gpd
from sqlalchemy import create_engine

# Configura los datos de conexión a tu base de datos PostGIS
db_host = "tu_host"
db_port = "5432" # Puerto por defecto de PostgreSQL
db_name = "tu_base_datos"
db_user = "tu_usuario"
db_password = 'tu_contraseña'

# Conectamos a La base de datos usando SQLAlchemy
engine = create_engine(f"postgresql://{db_user}:{db_password}@{db_host}:{db_port}/{db_name}")

# Cargamos La capa de centroides desde PostGIS en un GeoDataFrame
query_centroides = "SELECT id_distrito, nombre_distrito, nombre_municipio, gau FROM centroideos"
centroideos_df = gpd.read_postgis(query_centroides, con=engine)

# Cargamos el archivo CSV en un DataFrame
df_telefonia = pd.read_csv("datos_telefonia.csv")
```

# Procesado de datos

---

- ❖ Paso 1: pedir conversión del campo fecha a formato fecha y extraer campo de hora y de día.
- ❖ Paso 2: pedir la unión espacial con la capa de centroides que hemos almacenado en PostGIS, obtención de campos de nombre de distrito, municipio y GAU tanto para origen como para destino.
- ❖ Paso 3: pedir la eliminación de registros con valor NA en los campos edad y sexo.
- ❖ Paso 4: sustituir la id de provincia de residencia por el nombre de la provincia.
- ❖ Paso 5: (si trabajamos con una capa de matrices): pedir la creación de un campo de flujo que combine los nombres del distrito de origen y distrito de destino (`id_origen – id_destino`).
- ❖ Paso 6 (optativo): filtrar los registros con un determinado lugar de origen o destino y/o con un determinado valor de edad, sexo, renta o actividad.

# Procesado de datos

Editame el script para que el valor del campo id de provincia sea sustituido por el nombre de la provincia a partir de este diccionario:

- 1: 'Álava',
- 2: 'Albacete',
- 3: 'Alicante',
- 4: 'Almería',
- 5: 'Ávila',

51: 'Ceuta',  
52: 'Melilla'



Para reemplazar los valores de `id de provincia` por el nombre de la provincia en el DataFrame de Pandas, puedes utilizar el método `map`, que permite hacer una sustitución directa basada en un diccionario de mapeo. A continuación, aquí está el script modificado con esta funcionalidad añadida:

python

Copiar código

```
# Importamos Las Librerías necesarias
import pandas as pd
import geopandas as gpd
from sqlalchemy import create_engine

# Configura Los datos de conexión a tu base de datos PostGIS
db_host = 'tu_host'
db_port = '5432' # Puerto por defecto de PostgreSQL
db_name = 'tu_base_datos'
db_user = 'tu_usuario'
db_password = 'tu_contraseña'

# Conectamos a La base de datos usando SQLAlchemy
engine = create_engine(f"postgresql://{db_user}:{db_password}@{db_host}:{db_port}/{db_name}")

# Cargamos La capa de centroides desde PostGIS en un GeoDataFrame
query_centroides = "SELECT id_distrito, nombre_distrito, nombre_municipio, gau FROM centro"
centroides_df = gpd.read_postgis(query_centroides, con=engine)

# Cargamos el archivo CSV en un DataFrame
df_telefonia = pd.read_csv("datos_telefonia.csv")

# Eliminamos registros con valores NA en Los campos 'edad' y 'sexo'
df_telefonia.dropna(subset=['edad', 'sexo'], inplace=True)

# Diccionario de mapeo de ID de provincia a nombre de provincia
provincia_dict = {
    1: 'Álava', 2: 'Albacete', 3: 'Alicante', 4: 'Almería', 5: 'Ávila', 6: 'Badajoz', 7: 'Burgos', 10: 'Cáceres', 11: 'Cádiz', 12: 'Castellón', 13: 'Ciudad Real', 14: 'Córdoba', 16: 'Cuenca', 17: 'Girona', 18: 'Granada', 19: 'Guadalajara', 20: 'Gipuzkoa', 21: 'Huelva', 23: 'Jaén', 24: 'León', 25: 'Lleida', 26: 'La Rioja', 27: 'Lugo', 28: 'Madrid', 29: 'Málaga', 31: 'Navarra', 32: 'Ourense', 33: 'Asturias', 34: 'Palencia', 35: 'Las Palmas', 36: 'Pontevedra', 38: 'Tenerife', 39: 'Cantabria', 40: 'Segovia', 41: 'Sevilla', 42: 'Soria', 43: 'Tarazona', 45: 'Toledo', 46: 'Valencia', 47: 'Valladolid', 48: 'Vizcaya', 49: 'Zamora', 50: 'Zaragoza', 51: 'Ceuta', 52: 'Melilla'}
```

# Procesado de datos

- ❖ Vamos probando paso a paso el script para comprobar que se ha hecho bien.
- ❖ [https://github.com/joaquinosorioarjona/curso\\_uned/blob/main/procesado%20datos%20telefonia.ipynb](https://github.com/joaquinosorioarjona/curso_uned/blob/main/procesado%20datos%20telefonia.ipynb)
- ❖ Creamos una carpeta donde metemos las tablas.
- ❖ Solo vamos a tener una tabla por carpeta (ejecutamos el script con la tabla del día de enero como único archivo de la carpeta, luego repetimos con el día de agosto).
- ❖ Las tablas resultado se almacenan directamente en PostGIS.
- ❖ Optativo: probad a meter en la carpeta los siete ficheros de una semana completa y ejecutad el script. Tardará un rato, pero tendréis la tabla procesada con todos los datos de una semana entera.

```
In [3]: db_url = "postgresql://postgres:postgres@localhost:5432/datos_telefonia"
engine = create_engine(db_url)

In [4]: query = "SELECT id, nombre_mun, nombre_pro, gau FROM centroides"
centroides_df = pd.read_sql(query, engine)

centroides_df_origen = centroides_df.rename(columns={
    'id': 'origen',
    'nombre_mun': 'municipio_origen',
    'gau': 'gau_origen',
    'nombre_pro': 'provincia_origen'
})
centroides_df_destino = centroides_df.rename(columns={
    'id': 'destino',
    'nombre_mun': 'municipio_destino',
    'gau': 'gau_destino',
    'nombre_pro': 'provincia_destino'
})

In [5]: def procesar_y_cargar_csv(csv_path, table_name, centroides_df_origen, centroides_df_destino):
    # Leer el CSV
    df = pd.read_csv(csv_path, sep="|", encoding='UTF-8')

    # Aplicar transformaciones
    df = df.dropna(subset=['actividad_origen', 'actividad_destino', 'renta', 'edad', 'sexo'])
    df['fecha'] = pd.to_datetime(df['fecha'].astype(str), format='%Y%m%d')
    df['mes'] = df['fecha'].dt.month
    df['dia'] = df['fecha'].dt.day
    df.rename(columns={'periodo': 'hora'}, inplace=True)
    df['fecha hora'] = df['fecha'] + pd.to_timedelta(df['hora'], unit='h')
    df['origen'] = df['origen'].astype(str)
    df['destino'] = df['destino'].astype(str)
    df = df[df['origen'].str.contains('externo') & ~df['destino'].str.contains('externo')]
    df['flujo'] = df['origen'] + ' - ' + df['destino']
    df['residencia_provincia'] = df['residencia'].map(provincia_por_valor)

    df = df.merge(centroides_df_origen, on='origen', how='left')
    df = df.merge(centroides_df_destino, on='destino', how='left')
    df = df.dropna(subset=['municipio_origen', 'municipio_destino'])

    new_order = [
        'fecha hora', 'mes', 'dia', 'hora', 'origen', 'destino', 'flujo',
        'municipio_origen', 'municipio_destino', 'gau_origen',
        'gau_destino', 'residencia_provincia', 'provincia_origen', 'provincia_destino',
        'distancia', 'actividad_origen', 'actividad_destino',
        'renta', 'edad', 'sexo', 'viajes', 'viajes_km'
    ]
    df = df[new_order]

    df = df[(df['edad'] == '65-100') & (df['gau_origen'].isin(['GAU Sevilla', 'GAU Madrid', 'GAU Barcelona'])) & (df['gau_destino'].isin(['GAU Valencia', 'GAU Zaragoza', 'GAU Bilbao', 'GAU Pamplona', 'GAU León', 'GAU Oviedo', 'GAU Santiago', 'GAU Coruña']))]
    df.to_sql(table_name, engine, if_exists='append', index=False)
```



Dashboard Properties SQL Statistics Dependencies Dependents Processes public.distritos... public.envejeci... curso\_uned/po... public.zonifica... public.dis < >

public.viajes\_distritos\_15\_enero/datos\_telefonia/postgres@Po...



100 rows



You are currently running version 7.6 of pgAdmin

however the current version is 8.6.

Please click [here](#) for more information.

Query Query History

```
1 SELECT * FROM public.viajes_distritos_15_enero
2 LIMIT 100
3
```

Data Output Messages Notifications



	_hora	mes	dia	hora	origen	destino	flujo	distrito_origen	distrito_destino	municipio_origen	municipio_destino	gau_origen
	timestamp without time zone	integer	integer	bigint	text	text	text	text	text	text	text	text
1	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
2	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
3	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
4	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
5	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
6	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
7	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
8	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
9	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
10	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
11	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
12	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
13	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
14	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
15	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
16	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
17	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
18	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona
19	-01-15 20:00:00		1	15	20	0812103	0812104	0812103 - 0812104	Mataró distrito 03	Mataró distrito 04	Mataró	GAU Barcelona

- ❖ Vamos a suponer que preparamos el mismo material pero para la asignatura de Geografía Regional de España o Geografía de la Población.
- ❖ Como ejemplo, vamos a sacar la población diurna de los grupos de renta baja del Área Metropolitana de Sevilla en horario laboral (de 7 a 20).
- ❖ Podemos incluir directamente el filtro en el apartado de procesado que hemos hecho antes, o podemos filtrar sobre la tabla de PostGIS ya almacenada.
- ❖ Vamos a hacer esto último y exportar los resultados filtrados a una nueva tabla de PostGIS.
- ❖ [https://github.com/joaquinosorioarjona/curso\\_uned/blob/main/filtrado\\_datos\\_telefonia.ipynb](https://github.com/joaquinosorioarjona/curso_uned/blob/main/filtrado_datos_telefonia.ipynb)

Tengo este script para hacer un filtro de datos de Postgis y exportarlo a un dataframe de Pandas. Quiero almacenar este dataframe en una tabla nueva de la misma base de datos de Postgis import pandas as pd  
import psycopg2  
from sqlalchemy import create\_engine

```
db_url =  
'postgresql://postgres:postgres@localhost:5432/datos_telefonia'  
engine = create_engine(db_url)
```

query = """"  
SELECT \*  
FROM viajes\_distritos\_15\_enero  
WHERE gau\_origen = 'GAU Sevilla'  
AND gau\_destino = 'GAU Sevilla'  
AND renta = '<10'  
AND hora > 6 AND hora < 21  
"""

```
conn = psycopg2.connect(db_url)  
  
df_viajes = pd.read_sql_query(query, conn)  
df_viajes  
  
conn.close()
```



Para almacenar el `DataFrame` resultante en una nueva tabla dentro de la misma base de datos de PostGIS, puedes utilizar el motor de conexión `engine` que has creado con `SQLAlchemy`. En lugar de cerrar la conexión y terminar el proceso después de leer los datos, puedes utilizar el método `to_sql` de `Pandas` para exportar el `DataFrame` a una tabla en la base de datos.

Aquí te muestro cómo modificar el script para crear una nueva tabla en la base de datos de PostGIS a partir de tu `DataFrame`:

```
python  
  
import pandas as pd  
import psycopg2  
from sqlalchemy import create_engine  
  
# Parámetros de conexión a la base de datos  
db_url = 'postgresql://postgres:postgres@localhost:5432/datos_telefonia'  
engine = create_engine(db_url)  
  
# Definimos la consulta SQL para filtrar los datos  
query = """  
SELECT *
```

```
In [1]: import pandas as pd  
import psycopg2  
from sqlalchemy import create_engine
```

```
In [2]: db_url = 'postgresql://postgres:postgres@localhost:5432/datos_telefonia'  
engine = create_engine(db_url)
```

```
In [12]: query = """  
    SELECT *  
    FROM viajes_2023  
    WHERE municipio_origen ILIKE '%Málaga%'  
    AND gau_destino = 'GAU Málaga'  
    AND renta = '10-15'  
    AND fecha >= '2023-02-03 00:00:00' AND fecha < '2023-02-04 00:00:00'  
    AND hora > 3 AND hora < 12  
    """
```

```
In [4]: conn = psycopg2.connect(db_url)
```

```
In [13]: # Ejecutar la consulta y guardar los resultados en un DataFrame de pandas  
df_viajes = pd.read_sql_query(query, conn)  
df_viajes
```

C:\Users\Usuario\AppData\Local\Temp\ipykernel\_29332\2158733781.py:2: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please consider using SQLAlchemy.  
df\_viajes = pd.read\_sql\_query(query, conn)

Out[13]:

	fecha	mes	dia	hora	origen	destino	flujo	municipio_origen	municipio_destino	gau_origen	...	provincia_origen	provincia_destino	distancia
0	2023-02-03	2	3	5	2906701	2906702	2906701 - 2906702	Málaga distrito 01	Málaga distrito 02	GAU Málaga	...	Málaga	Málaga	0.5-2
1	2023-02-03	2	3	4	2906701	2906703	2906701 - 2906703	Málaga distrito 01	Málaga distrito 03	GAU Málaga	...	Málaga	Málaga	0.5-2
2	2023-02-03	2	3	4	2906701	2906704	2906701 - 2906704	Málaga distrito 01	Málaga distrito 04	GAU Málaga	...	Málaga	Málaga	0.5-2
3	2023-02-03	2	3	4	2906701	2906705	2906701 - 2906705	Málaga distrito 01	Málaga distrito 05	GAU Málaga	...	Málaga	Málaga	0.5-2

Dashboard Properties SQL Statistics Dependencies Dependents Processes public.distritos... public.envejeci... curso\_uned/po... public.zonifica... public.dis < >

public.viajes\_sevilla\_15\_enero/datos\_telefonia/postgres@Post...

No limit ▾ E

You are currently running version 7.6 of pgAdmin  
however the current version is 8.6.

Please click [here](#) for more information.

Query History

```
1 SELECT * FROM public.viajes_sevilla_15_enero
2
```

Data Output Messages Notifications

	distrito_origen	distrito_destino	municipio_origen	municipio_des	gau_origen	gau_destino	residencia_pr	provincia_origen	provincia_des	distancia	actividad_origen	actividad_des	renta	edad	sexo
1	403	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	0.5-2	frecuente	frecuente	<10	0-25	hombre
2	403	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Huelva	Sevilla	0.5-2	frecuente	frecuente	<10	45-65	mujer
3	301	Alcalá de ...	Dos Hermanas dis...	Alcalá de ...	Dos Herm...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	no_frecue...	frecuente	<10	45-65	mujer
4	404	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	0.5-2	frecuente	frecuente	<10	0-25	hombre
5	1	Alcalá de ...	Camas	Alcalá de ...	Camas	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	10-50	no_frecue...	casa	<10	25-45	mujer
6	404	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	frecuente	trabajo_e...	<10	25-45	hombre
7	404	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	trabajo_e...	frecuente	<10	45-65	mujer
8	1	Alcalá de ...	Camas	Alcalá de ...	Camas	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	10-50	no_frecue...	no_frecue...	<10	25-45	mujer
9	404	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	0.5-2	frecuente	frecuente	<10	0-25	mujer
10	404	Alcalá de ...	Alcalá de Guadaír...	Alcalá de ...	Alcalá de ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	frecuente	frecuente	<10	45-65	hombre
11	302	Alcalá de ...	Dos Hermanas dis...	Alcalá de ...	Dos Herm...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	frecuente	frecuente	<10	0-25	mujer
12	302	Alcalá de ...	Dos Hermanas dis...	Alcalá de ...	Dos Herm...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	no_frecue...	casa	<10	45-65	hombre
13	304	Alcalá de ...	Dos Hermanas dis...	Alcalá de ...	Dos Herm...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	no_frecue...	frecuente	<10	45-65	hombre
14	302	Alcalá de ...	Mairena del Alcor ...	Alcalá de ...	Mairena d...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	frecuente	casa	<10	0-25	mujer
15	102	Alcalá de ...	Rinconada, La dist...	Alcalá de ...	Rinconad...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	no_frecue...	casa	<10	45-65	hombre
16	5	Alcalá de ...	San Juan de Azná...	Alcalá de ...	San Juan ...	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	10-50	no_frecue...	casa	<10	25-45	mujer
17	104	Alcalá de ...	Sevilla distrito 04	Alcalá de ...	Sevilla	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	10-50	no_frecue...	casa	<10	45-65	mujer
18	104	Alcalá de ...	Sevilla distrito 04	Alcalá de ...	Sevilla	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	2-10	no_frecue...	casa	<10	45-65	mujer
19	109	Alcalá de ...	Sevilla distrito 09	Alcalá de ...	Sevilla	GAU Sevilla	GAU Sevilla	Sevilla	Sevilla	10-50	frecuente	no_frecue...	<10	65-100	mujer

Total rows: 1000 of 42579 Query complete 00:00:00.171 Mostrar escritorio

# Procesado de datos

---

- ❖ Como hemos visto, los datos diarios pesan individualmente más de 1GB. Con el script que hemos utilizado antes podemos unir todas las tablas dentro de una carpeta en una sola tras eliminar los registros con valores NA, lo cual ayuda a reducir el peso, pero sigue siendo una cantidad enorme de datos.
- ❖ Es posible unir todas las tablas de un mes en una sola, pero el script tardará en realizarse.
- ❖ Es muy poco viable unir todas las tablas de un año. Consumiría mucho espacio, tiempo y recursos.
- ❖ Si es viable crear ficheros mensuales de una comunidad autónoma (e incluso ficheros anuales de una provincia), pero el MITMA no ofrece la descarga de los datos de las regiones que queremos.

# Procesado de datos

---

- ❖ Será necesaria la descarga de los ficheros completos y un proceso de filtrado de datos y de unión de los datos filtrados en un solo fichero.
- ❖ Vamos a hacer un ejemplo descargando del MIMA los ficheros del 23 al 29 de abril (Feria de Abril de Sevilla de 2023) y agregando las siete tablas dentro de una misma carpeta.
- ❖ Al no estar enriquecidos los datos, para hacer el filtro, tendremos que hacer manualmente un diccionario con los códigos de los municipios que pertenecen a la GAU de Sevilla.
- ❖ [https://github.com/joaquinosorioarjona/curso\\_uned/blob/main/union\\_tablas\\_telefonia.ipynb](https://github.com/joaquinosorioarjona/curso_uned/blob/main/union_tablas_telefonia.ipynb)



ChatGPT



Data Analyst

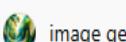


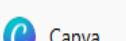
image generator



Write For Me



Scholar GPT



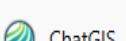
Canva



DALL-E



Code Copilot



ChatGIS

... Ver menos ^

Explorar GPT

Hoy

Copiar tabla PostgreSQL

Exportar DataFrame a PostGIS

Guardar DataFrame en script

Filtrado y unión de CSV ...

Procesar CSV con Pandas

Transformación y enriquecimiento

7 días anteriores

He actualizado el script para que cargue todos los archivos CSV de la carpeta `datos_telefonia` en lugar de especificarlos manualmente. A continuación te muestro el código modificado:

```
python

import os
import gc
import pandas as pd

# Definimos el conjunto de IDs de Sevilla para los filtros
ids_sevilla = ["4100401", "4100402", "4100403", "4100404", "4100701", "4100702", "41010",
                "41029", "4103401", "4103402", "4103403", "4103801", "4103802", "4103803",
                "41040", "41044", "41047", "4105801", "4105802", "4105803", "41059", "41069",
                "4106903", "4106904", "41070", "41079", "4108101", "4108102", "41086", "410902",
                "4109102", "4109103", "4109104", "4109105", "4109106", "4109107", "4109108",
                "4109110", "4109111", "41093", "4109501", "4109502", "4109503", "4109504",
                "41096", "41102", "41904"]

# Ruta a la carpeta con los archivos CSV
carpeta = "./datos_telefonia"

# DataFrame final para almacenar todos los datos procesados
julio = pd.DataFrame()

# Iterar sobre cada archivo CSV en la carpeta
for archivo in os.listdir(carpeta):
    if archivo.endswith(".csv"):
        # Construir la ruta completa al archivo
        ruta_archivo = os.path.join(carpeta, archivo)

        # Cargar el CSV
        df = pd.read_csv(ruta_archivo, sep="|", encoding='UTF-8')
```



Envía un mensaje a ChatGPT



python

```

import geopandas as gpd
from sqlalchemy import create_engine
from geoalchemy2 import Geometry, WKTElement
from shapely.geometry import LineString
import os

# Ruta de la capa de centroides
centroides_path = os.path.expanduser("~/Documentos/centroides.shp")

# Cargar los centroides como un GeoDataFrame
centroides_gdf = gpd.read_file(centroides_path)

# Verificar el sistema de coordenadas
if centroides_gdf.crs is None:
    raise ValueError("La capa de centroides no tiene un sistema de coordenadas definido.")

# Crear líneas origen-destino
lines = []
for i, origen in centroides_gdf.iterrows():
    for j, destino in centroides_gdf.iterrows():
        if i != j: # Evitar líneas a sí mismo
            line = LineString([origen.geometry, destino.geometry])
            lines.append({"origen_id": origen["id"], "destino_id": destino["id"], "geometry": line})

# Crear un nuevo GeoDataFrame con las líneas
lineas_gdf = gpd.GeoDataFrame(lines, geometry="geometry", crs=centroides_gdf.crs)

# Configurar la conexión a la base de datos PostGIS
db_connection_url = "postgresql+psycopg2://usuario:contraseña@localhost:5432/datos_telefonia"
engine = create_engine(db_connection_url)

# Guardar la capa de líneas en PostGIS
table_name = "lineas_origen_destino"

# Convertir geometrías a formato WKT para insertar en PostGIS
lineas_gdf["geometry"] = lineas_gdf["geometry"].apply(lambda geom: WKTElement(geom.wkt, srid=4326))

# Crear la tabla en PostGIS
lineas_gdf.to_postgis(name=table_name, con=engine, if_exists="replace", index=False, dtype={})

print(f"Capa de líneas origen-destino guardada en la tabla '{table_name}' en la base de datos")

```

**Descripción del script**

1. **Cargar la capa de centroides:** Usamos `geopandas` para cargar el shapefile de los centroides ubicado en la carpeta `Documentos`.
2. **Crear las líneas origen-destino:** Generamos líneas `LineString` conectando cada punto con todos los demás puntos, excluyendo conexiones de un punto consigo mismo.
3. **Crear el GeoDataFrame de líneas:** Cada línea incluye un ID de origen, un ID de destino y la geometría de la línea.
4. **Conectar con la base de datos PostGIS:** Configuramos la conexión a la base de datos usando `SQLAlchemy`.
5. **Guardar el GeoDataFrame en PostGIS:** Insertamos las líneas en la base de datos en una tabla nueva llamada `lineas_origen_destino`.

# Procesado de datos

ChatGIS ▾

❖ ChatGPT como herramienta para la creación de scripts de Python.

❖ ChatGPT orientado al SIG: ChatGIS.



❖ Ejemplo de uso: generación de script para la creación de una capa de matriz OD.

❖ [https://github.com/iaquinosorioarjona/cu\\_rso\\_uned/blob/main/matriz\\_od\\_telefonia.ipynb](https://github.com/iaquinosorioarjona/cu_rso_uned/blob/main/matriz_od_telefonia.ipynb)

Créame un script de Python para la creación de una capa de líneas origen-destino a partir de una capa de centroides llamada centroides y almacenada en la carpeta Documentos. La capa de líneas origen-destino la quiero almacenar directamente en una base de datos de PostGIS llamada datos\_telefonia. El script lo ejecutaré en Jupyter.

A continuación, te proporciono un script en Python que crea una capa de líneas origen-destino a partir de una capa de centroides en formato shapefile (almacenada en tu carpeta de Documentos). El script utiliza `GeoPandas` para cargar los datos de los centroides y `SQLAlchemy` junto con `GeoAlchemy2` para insertar la nueva capa de líneas en una base de datos PostGIS. Este script asume que cada punto de origen se conecta a todos los demás puntos como destino.

## Requisitos previos

Antes de ejecutar el script, asegúrate de tener instaladas las siguientes bibliotecas de Python:

- `geopandas`
- `sqlalchemy`
- `geoalchemy2`
- `psycopg2`

Puedes instalarlas con:

bash

Copiar código

```
pip install geopandas sqlalchemy geoalchemy2 psycopg2
```

```
In [13]: for geom1 in geoms:
    for geom2 in geoms:
        if geom1[1] != geom2[1]: # Evitar Líneas desde un punto hacia sí mismo
            linea = LineString([geom1[0].centroid, geom2[0].centroid])
            id_origen = geom1[1]
            id_destino = geom2[1]
            id_origen_destino = f'{id_origen} - {id_destino}'
            origen_data = centroides.loc[centroides['ID'] == id_origen].iloc[0]
            destino_data = centroides.loc[centroides['ID'] == id_destino].iloc[0]
            municipio_origen = origen_data['municipio']
            municipio_destino = destino_data['municipio']
            gau_origen = origen_data['gau']
            gau_destino = destino_data['gau']
            provincia_origen = origen_data['provincia']
            provincia_destino = destino_data['provincia']
            lineas.append((linea, id_origen, id_destino, id_origen_destino,
                           municipio_origen, municipio_destino,
                           gau_origen, gau_destino,
                           provincia_origen, provincia_destino))
            line_geoms.append(linea)
```

```
In [9]: gdf_lines = gpd.GeoDataFrame({
    'geometry': line_geoms,
    'origen': [line[1] for line in lineas],
    'destino': [line[2] for line in lineas],
    'flujo': [line[3] for line in lineas],
    'municipio_origen': [line[4] for line in lineas],
    'municipio_destino': [line[5] for line in lineas],
    'gau_origen': [line[6] for line in lineas],
    'gau_destino': [line[7] for line in lineas],
    'provincia_origen': [line[8] for line in lineas],
    'provincia_destino': [line[9] for line in lineas]
}, crs=sr)
```

```
In [14]: user = 'postgres'
password = 'postgres'
host = 'localhost'
port = '5432'
database = 'datos_telefonia'
conn_str = f"postgresql://{user}:{password}@{host}:{port}/{database}"
engine = create_engine(conn_str)
```

```
In [44]: gdf_lines.to_postgis('matriz', engine, if_exists='replace', index=False)
```

File Object Tools Help

Object Explorer

Dashboard Properties SQL Statistics Dependencies Dependents Processes public.matriz/c... public.matriz/curso\_uned/postgres@PostgreSQL 16

public.matriz/curso\_uned/postgres@PostgreSQL 16

No limit

You are currently running version 7.6 of pgAdmin, however the current version is 8.6.

Please click [here](#) for more information.

Query History

```
1 SELECT * FROM public.matriz
2
```

Data Output Messages Notifications

	geometry	origen	destino	flujo	distrito_origen	distrito_destino
	geometry	text	text	text	text	text
1	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241468447E75E521841922F01139D154F...	3304408	2909402_AD	3304408 - 2909402_AD	Oviedo distrito 08	Vélez-Málaga distrito 02 agreg
2	0102000020E20B00000200000125150A4666F1041DDF9BE94135652417518475E879A1841EAB70F04AA084...	3304408	2909403	3304408 - 2909403	Oviedo distrito 08	Vélez-Málaga distrito 03
3	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241EF015D5C49201841486F132077074F...	3304408	2909404	3304408 - 2909404	Oviedo distrito 08	Vélez-Málaga distrito 04
4	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241C7966E168FEF1641F652A8EE1C684...	3304408	29095_AM	3304408 - 29095_AM	Oviedo distrito 08	Villanueva de Algaidas agreg
5	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241275748D86917174164971D5D5841...	3304408	29096	3304408 - 29096	Oviedo distrito 08	Villanueva del Rosario
6	0102000020E20B00000200000125150A4666F1041DDF9BE941356524122ED7D9CBC5A1741748BB2371B484...	3304408	29097	3304408 - 29097	Oviedo distrito 08	Villanueva del Trabuco
7	0102000020E20B00000200000125150A4666F1041DDF9BE94135652415E6C149FE6F41341BA35A750C0074...	3304408	29100	3304408 - 29100	Oviedo distrito 08	Yunquera
8	0102000020E20B00000200000125150A4666F1041DDF9BE94135652419E3E5384E13E1641041F52E1D5EE4E...	3304408	29901	3304408 - 29901	Oviedo distrito 08	Torremolinos
9	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241A65E8BD92C3416414C720B708F304...	3304408	29902	3304408 - 29902	Oviedo distrito 08	Villanueva de la Concepción
10	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241377468A3807F24414AB0961BDC225...	3304408	3000101	3304408 - 3000101	Oviedo distrito 08	Abanilla distrito 01
11	0102000020E20B00000200000125150A4666F1041DDF9BE941356524146776DE8EF752441C0D02EACAE285...	3304408	3000102	3304408 - 3000102	Oviedo distrito 08	Abanilla distrito 02
12	0102000020E20B00000200000125150A4666F1041DDF9BE941356524156F00795AFBD2341B0022C2D00285...	3304408	30002	3304408 - 30002	Oviedo distrito 08	Abarán
13	0102000020E20B00000200000125150A4666F1041DDF9BE94135652414E5BA5E73C1023419967443F05A54...	3304408	30003	3304408 - 30003	Oviedo distrito 08	Águilas
14	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241012D391C95FF23414B36E00CA0095...	3304408	30005	3304408 - 30005	Oviedo distrito 08	Alcantarilla
15	0102000020E20B00000200000125150A4666F1041DDF9BE941356524131377BA62EE22341566ED31AD4125...	3304408	30007	3304408 - 30007	Oviedo distrito 08	Alguzas
16	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241B13E2C24F996234195F8FC146DF24...	3304408	30008	3304408 - 30008	Oviedo distrito 08	Alhama de Murcia
17	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241FCC587CEDFD12341F342CB3FAB185...	3304408	30009	3304408 - 30009	Oviedo distrito 08	Archena
18	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241E5B7C23314A324415108786103105...	3304408	30010	3304408 - 30010	Oviedo distrito 08	Beniel
19	0102000020E20B00000200000125150A4666F1041DDF9BE9413565241BF3B101ED6CE234195DE1ECEA8235...	3304408	30011	3304408 - 30011	Oviedo distrito 08	Blanca

Total rows: 1000 of 14367890 Query complete 00:00:22.699 Ln 1, Col 1

# Procesado de datos

---

- ❖ La unión de la tabla con geometría y la tabla con datos de telefonía móvil la podemos realizar ahora en PostGIS (para deshacer la unión, cerramos y reabrimos PgAdmin):

sql

Copiar código

SELECT

m.\*,

v.\*

FROM

matriz AS m

LEFT JOIN

viajes\_distritos\_15\_enero AS v

ON

m.flujo = v.flujo;

ERROR: la columna «origen» fue especificada más de una vez

SQL state: 42701



El error ocurre porque ambas tablas (`matriz` y `feria_abril_2023`) tienen columnas con el mismo nombre, y al utilizar `m.*` y `f.*` se están incluyendo duplicados de las mismas columnas, como `origen` y posiblemente otras. Para solucionar esto, selecciona explícitamente solo los campos necesarios de cada tabla y evita duplicar nombres de columnas en el `SELECT`.

Aquí tienes una consulta corregida:

sql

`CREATE TABLE matriz_feria_abril AS  
SELECT  
 m.id, -- Campo de ejemplo de la tabla matriz  
 m.gau_origen, -- Campo específico de matriz  
 m.gau_destino, -- Campo específico de matriz  
 m.flujo, -- Campo común utilizado en la unión  
 m.geom, -- Campo de geometría en matriz  
 f.campo_adicional_1, -- Campo específico de feria_abril_2023  
 f.campo_adicional_2 -- Otro campo adicional de feria_abril_2023  
FROM  
 matriz AS m  
JOIN  
 feria_abril_2023 AS f  
ON  
 m.flujo = f.flujo  
WHERE  
 m.gau_origen = 'GAU Sevilla' AND`

Copiar código



Envía un mensaje a ChatGIS





Object Explorer

- curso\_uned
  - Casts
  - Catalogs
  - Event Triggers
  - Extensions
  - Foreign Data Wrappers
  - Languages
  - Publications
  - Schemas (1)
    - public
      - Aggregates
      - Collations
      - Domains
      - FTS Configurations
      - FTS Dictionaries
      - FTS Parsers
      - FTS Templates
      - Foreign Tables
      - Functions
      - Materialized Views
      - Operators
      - Procedures
      - Sequences
    - Tables (10)
      - centroides
      - distritos
      - feria\_abril\_2023
      - matriz
      - matriz\_feria\_abril
      - spatial\_ref\_sys
      - viajes\_distritos\_15\_agosto
      - viajes\_distritos\_15\_enero
      - viajes\_sevilla\_15\_agosto
      - viajes\_sevilla\_15\_enero
    - Trigger Functions
    - Types
    - Views
  - Subscriptions
- datos\_telefonia

public.matriz\_feria\_abril/curso\_uned/postgres@PostgreSQL 16

Query History

```
1 SELECT * FROM public.matriz_feria_abril
2
```

Scratch Pad

Data Output Messages Notifications



	geometry geometry	origen text	destino text	flujo text	distrito_origen text	distrito_destino text
1	0102000020E20B000002000000EA2F45D53A8A0D41846CDC72E2804F4155ECE6EF27AC0E41184F15B8D0834...	4103801	4100404	4103801 - 4100404	Dos Hermanas distrito 01	Alcalá de Guadaíra distrito
2	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
3	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
4	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
5	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
6	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
7	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
8	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
9	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
10	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
11	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
12	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
13	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
14	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
15	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
16	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
17	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
18	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito
19	0102000020E20B00000200000055ECE6EF27AC0E41184F15B8D0834F4104C2DDA4E12E0E41432DF4B6678D4...	4100404	4100402	4100404 - 4100402	Alcalá de Guadaíra distrito 04	Alcalá de Guadaíra distrito

Total rows: 1000 of 2063846 Query complete 00:00:04.716

Ln 1, Col 1

