

# Informe de estado: Clasificación tipológica de barrios

Procesamiento del Lenguaje Natural

Jesús Martínez Leal & Isaac Martí García

16/05/2024

# Distribución de tareas



- Empezar memoria / Representaciones de clusters a partir de los datos estructurados



- Recolección de más opiniones (por distritos)



- Recolección de más opiniones (por distritos) / Uso de Transformers como modelos / Empezar memoria / Entradas para el Blog



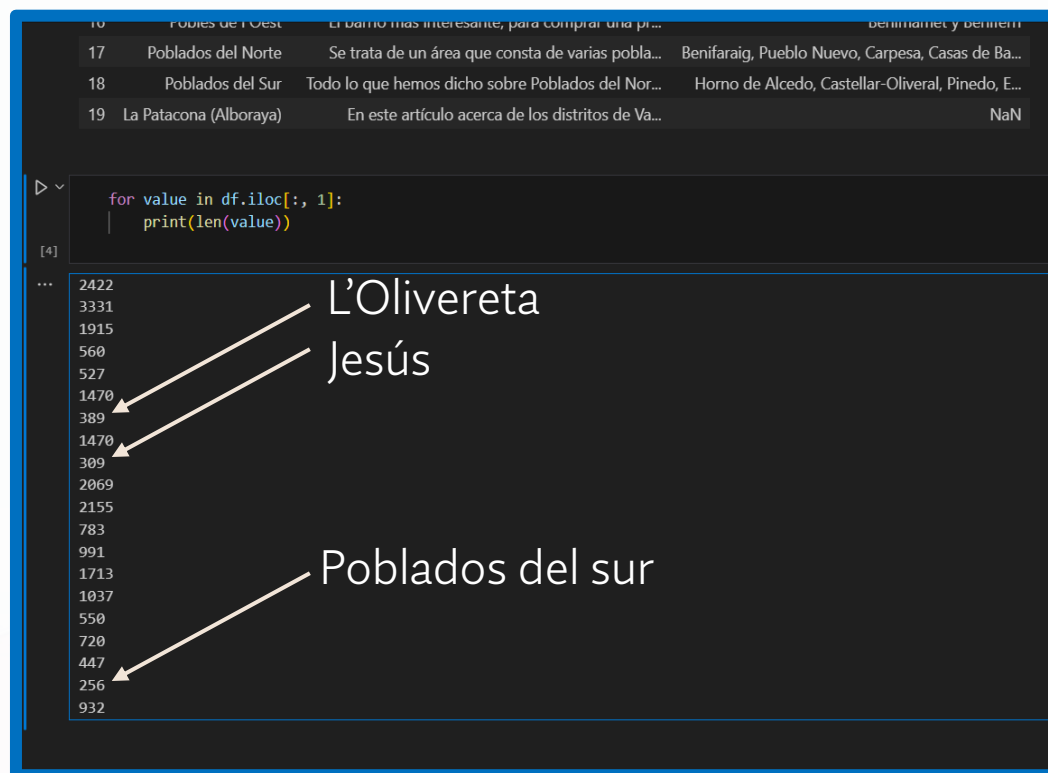
- Uso de Transformers como modelos / Investigación para el futuro...



- Entradas para el Blog / Empezar memoria

# Recolección de más opiniones

- **Objetivo:** Coger información adicional de los distritos con menos información.
- L'Olivereta, Jesús, Poblats del Sur.
- Poca información en general, páginas específicas para cada uno.
- Northleg y Valenciasecreta. En algunos casos información poco relevante o inexistente. (Caso distrito Jesús)



Longitud de cada descripción. Las que menos tienen son las mencionadas anteriormente

## L'Olivereta, el nuevo barrio de moda según Airbnb

Lidera el ranking de barrios españoles de auge entre turistas europeos.

VALENCIA SECRETA · NOVIEMBRE 15, 2018

COMPARTE EL ARTÍCULO



Alejado de toda la atención mediática, el **barrio de L'Olivereta** se hace popular entre los turistas. Así lo afirman datos facilitados por la plataforma de alquiler Airbnb, que recoge los barrios más populares de España de este otoño.

Valencia gusta entre los turistas europeos y cuando vienen, se centran últimamente en dos barrios: este, el de L'Olivereta, con un crecimiento del 186% y el **barrio de Jesús**, cerca de la estación del AVE Joaquín Sorolla. En España solo le siguen de cerca San Pablo y Santa Justa en Sevilla (ambos con un crecimiento del +142%).



**Las Gaviotas Apartments**

Your accommodation in the Canary Islands



lasgaviotas.es

Extracción de información de este tipo de páginas.



```

from urllib import request
import requests

# Fetch the webpage content
url = "https://es.northleg.com/valencia/barrios/barrios-populares/pobles-sud/"
response = requests.get(url)
html_content = response.content

# Parse the HTML content
soup = BeautifulSoup(html_content, "html.parser")

# Find all <p> tags
paragraphs = soup.find_all("p")

# Ignore the last paragraph and concatenate the text content
texto = ""
for i, paragraph in enumerate(paragraphs[::-1]):
    if i != 0: # Skip the first paragraph
        texto += paragraph.get_text(strip=True)

# Print the variable texto
print(texto)

```

Pobles del Sud es un gran barrio al sureste de Valencia, que abarca unos 20 km de terreno costero justo al sur de la ciudad. A pesar de que es el barrio más

```

df2.loc[df2['Zona'] == "Poblados del Sur", 'Información'] += texto
df2["Información"][18]

```

Todo lo que hemos dicho sobre Poblados del Norte, también se puede atribuir al área de Poblados del Sur. Las poblaciones pertenecientes a esta zona, se com

Concatenación de la nueva información con la anterior. CSV más completo



```

df = pd.read_csv("Distritos_v3.csv")
df

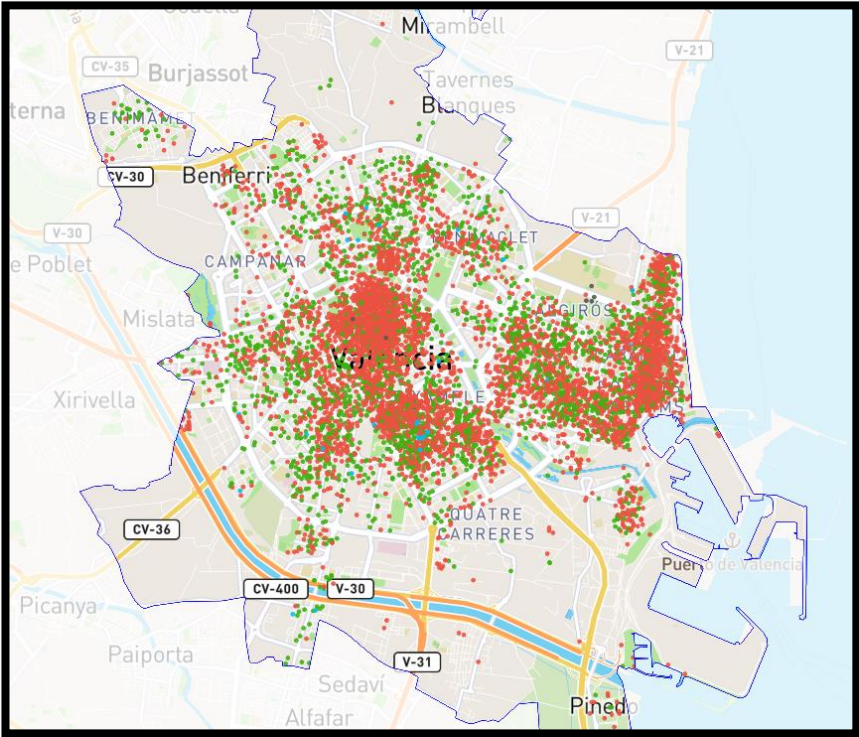
```

	Zona	Información	Barrios
0	Ciutat Vella	Este es el casco antiguo de Valencia, la zona ...	La Seu, La Xerea, El Carmen, El Pilar (Vellute...
1	L'Eixample	Es el distrito más de moda de Valencia. Es esp...	Ruzafa, El Pla del Remei y Gran Vía.
2	Extramurs	Aquí se encuentra el famoso complejo deportivo...	El Botànic, La Roqueta, La Pebina y Arrancapins.
3	Campanar	Pros: Cerca hay un supermercado Carrefour muy ...	Campanar, Les Tendetes, El Calvari y Sant Pau.
4	La Zaidia	Esta área está ubicada entre Campanar y El Pla...	Marxalenes, Morvedre, Trinitat, Tormos y Sant ...
5	El Pla del Real	Esta área está situada al lado del grande y he...	Exposició, Mestalla, Jaume Roig y Ciutat Unive...
6	L'Olivereta	Este distrito está bastante alejado del centro...	Nou Moles, Soternes, Tres Forques, La Fuensant...
7	Patraix	El distrito de Patraix está ubicado lejos del ...	Patraix, Sant Isidre, Vara de Quart, Safranar ...
8	Jesús	Un área de Valencia con inmuebles de bajo cost...	La Raiosa, L'Hort de Senabre, La Creu Coberta,...
9	Quatre Carreres	Las mejores propiedades hay que buscarlas en e...	Monteolivete, En Corts, Malilla, Fuente de San...
10	Poblats Maritims	La mejor área para comprar propiedades en este...	El Grao, Cabañal-Cañamelar, Malvarrosa, Beteró...
11	Camins al Grau	Se pueden encontrar muy buenas propiedades en ...	Ayora, Albors, La Creu del Grau, Camí Fondo y ...
12	Algirós	Algirós es un área estudiantil, especialmente ...	L'illa Perduda, Ciutat Jardí, Amistat, La Bega...
13	Benimaclet	Buena área para vivir, con excelente infraestr...	Benimaclet y Camí de Vera.
14	Rascanya	Este distrito es muy diverso. Aquí, en el barr...	Els Orriols, Torrefiel y Sant Llorenç.
15	Benicalap	Este distrito está muy alejado del centro y de...	Benicalap y Ciutat Fallera.
16	Pobles de l'Oest	El barrio más interesante, para comprar una pr...	Benimámet y Beniferri
17	Poblados del Norte	Se trata de un área que consta de varias pobla...	Benifaraig, Pueblo Nuevo, Carpesa, Casas de Ba...
18	Poblados del Sur	Todo lo que hemos dicho sobre Poblados del Nor...	Horno de Alcedo, Castellar-Oliveral, Pinedo, E...
19	La Patacona (Alboraya)	En este artículo acerca de los distritos de Va...	NaN

CSV Resultante, con Distrito, Información y sus correspondientes barrios.

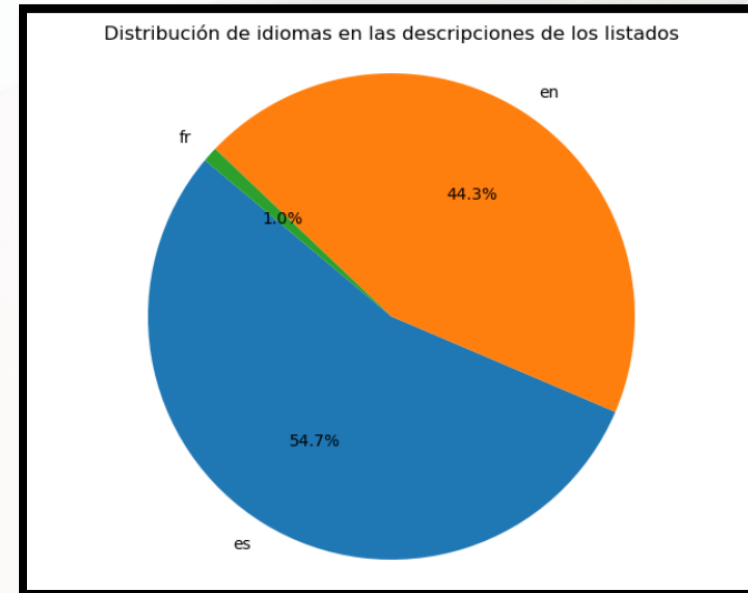
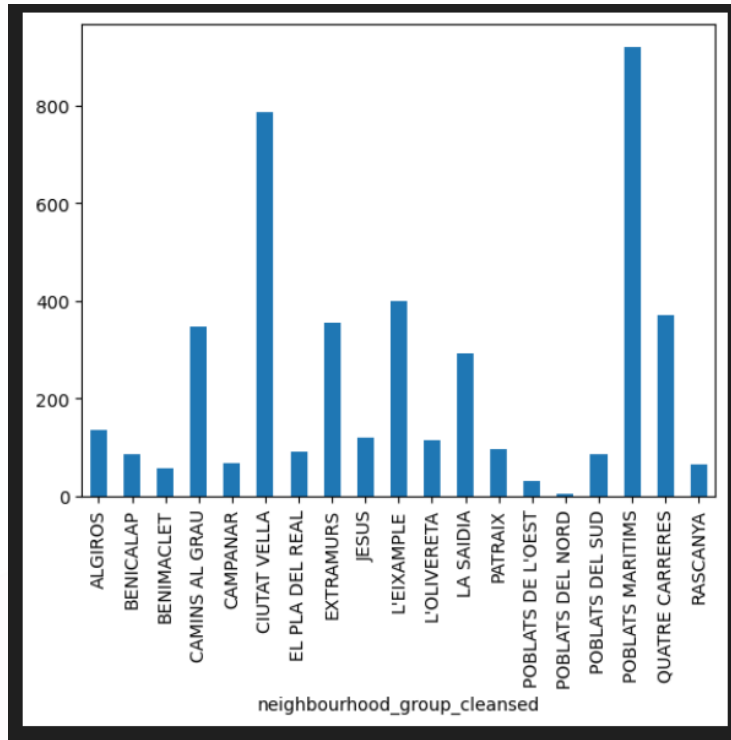
# Opiniones de Inside Airbnb por distrito

**Inside Airbnb**  
Adding data to the debate

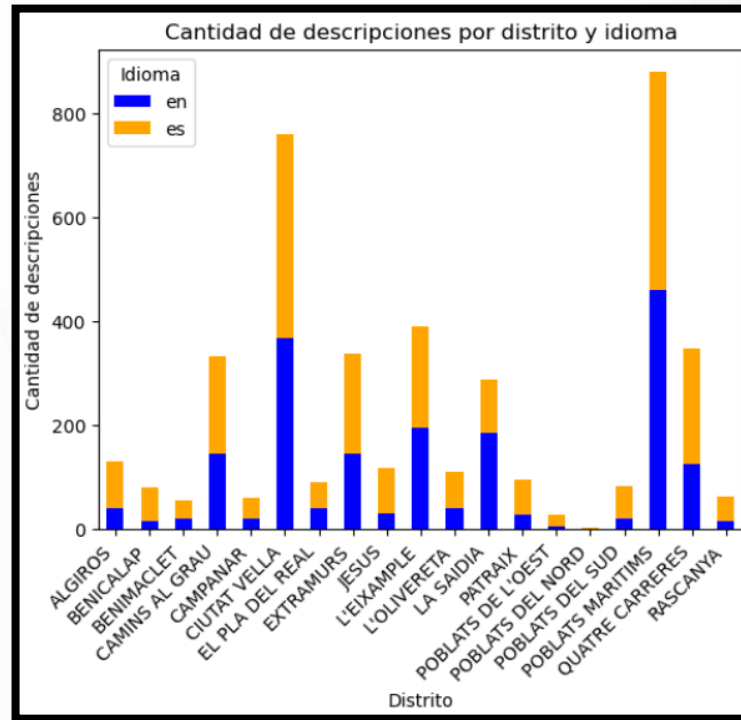


Country/City	File Name	Description
Valencia	<a href="#">listings.csv.gz</a>	Detailed Listings data
Valencia	<a href="#">calendar.csv.gz</a>	Detailed Calendar Data
Valencia	<a href="#">reviews.csv.gz</a>	Detailed Review Data
Valencia	<a href="#">listings.csv</a>	Summary information and metrics for listings in Valencia (good for visualisations).
Valencia	<a href="#">reviews.csv</a>	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).

# Opiniones de Inside Airbnb por distrito



# Opiniones de Inside Airbnb por distrito





# Clasificación de distritos

- ¿Histórico, lujo, multicultural, familiar, universitario, turístico, ¡suburbano!, comercial?

## Con opiniones de Inside Airbnb...



	BoW	TF-IDF	TF-IDF N-gram(1,3)	zero-shot-classification
ALGIROS	<u>suburbano (0.27)</u> , turístico (0.23)	suburbano (0.15), universitario (0.13)	suburbano (0.05), universitario (0.04)	multicultural (0.21), familiar (0.14)
BENICALAP	<u>suburbano (0.20)</u> , comercial (0.13)	suburbano (0.11), comercial (0.07)	suburbano (0.02), comercial (0.01)	histórico (0.18), familiar (0.15)
BENIMACLET	<u>suburbano (0.20)</u> , comercial (0.20)	universitario (0.11), suburbano (0.10)	suburbano (0.03), universitario (0.03)	multicultural (0.27), familiar (0.18)
CAMINS AL GRAU	<u>suburbano (0.33)</u> , comercial (0.25)	suburbano (0.19), comercial (0.14)	suburbano (0.06), comercial (0.05)	turístico (0.19), histórico (0.14)
CAMPANAR	<u>suburbano (0.26)</u> , comercial (0.20)	suburbano (0.15), comercial (0.13)	comercial (0.04), suburbano (0.04)	histórico (0.23), comercial (0.14)
CIUTAT VELLA	<u>suburbano (0.33)</u> , turístico (0.23)	suburbano (0.19), turístico (0.14)	suburbano (0.06), turístico (0.05)	turístico (0.21), lujo (0.19)
EL PLA DEL REAL	<u>suburbano (0.27)</u> , comercial (0.20)	suburbano (0.15), comercial (0.10)	suburbano (0.04), comercial (0.03)	turístico (0.25), histórico (0.19)
EXTRAMURS	<u>suburbano (0.32)</u> , comercial (0.24)	suburbano (0.19), comercial (0.13)	suburbano (0.06), comercial (0.04)	familiar (0.19), multicultural (0.15)

Todavía más evidente el problema que teníamos en los modelos de conteo

## Con opiniones de Inside Airbnb... Si quitamos **suburbano**

	BoW	TF-IDF	TF-IDF N-gram(1,3)	zero-shot-classification
ALGIROS	turístico (0.23), comercial (0.22)	universitario (0.13), turístico (0.12)	universitario (0.04), turístico (0.04)	multicultural (0.21), familiar (0.14), comerci...
BENICALAP	comercial (0.13), turístico (0.10)	comercial (0.07), turístico (0.05)	comercial (0.01), multicultural (0.01)	histórico (0.17), familiar (0.15), comercial (...)
BENIMACLET	comercial (0.20), universitario (0.20)	universitario (0.11), turístico (0.10)	universitario (0.03), turístico (0.03)	multicultural (0.27), familiar (0.18), univers...
CAMINS AL GRAU	comercial (0.25), turístico (0.21)	comercial (0.14), turístico (0.11)	comercial (0.05), turístico (0.04)	turístico (0.18), alternativo (0.13), históric...
CAMPANAR	comercial (0.20), turístico (0.13)	comercial (0.13), turístico (0.06)	comercial (0.04), turístico (0.02)	histórico (0.23), comercial (0.13), turístico ...
CIUTAT VELLA	turístico (0.23), comercial (0.23)	turístico (0.14), comercial (0.13)	turístico (0.05), comercial (0.05)	turístico (0.20), lujo (0.18), histórico (0.17...
EL PLA DEL REAL	comercial (0.20), turístico (0.17)	comercial (0.11), turístico (0.09)	comercial (0.03), turístico (0.02)	turístico (0.24), histórico (0.18), multicultu...
EXTRAMURS	comercial (0.24), turístico (0.22)	comercial (0.13), turístico (0.12)	comercial (0.04), turístico (0.04)	familiar (0.18), multicultural (0.14), turísti...

Surge el problema de **comercial**... ☹


Las opiniones de Inside Airbnb son demasiado complicadas de balancear para una buena clasificación.

Optamos por quedarnos por las opiniones de *Isaac* de datos **no estructurados** (los actualizados)



# Zero-Shot Classification

```
classifier = pipeline("zero-shot-classification")
```

 FacebookAI/**roberta-large-mnli**



HUGGING FACE

zero-shot-classification	
Ciutat Vella	turístico (0.22), histórico (0.19)
Extramurs	familiar (0.22), histórico (0.17)
Patraix	multicultural (0.24), comercial (0.17)
Jesús	multicultural (0.24), familiar (0.20)
Benimaclet	multicultural (0.18), familiar (0.16)
Benicalap	suburbano (0.19), familiar (0.17)

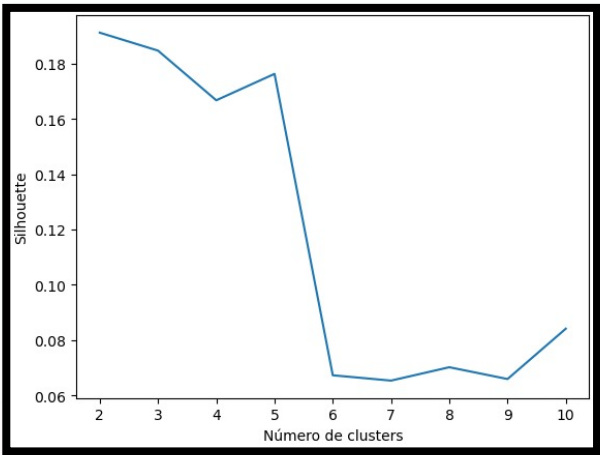




# Visualización de hallazgos

- A partir de datos estructurados (unión de todas las variables del Ayuntamiento y los del Barómetro tras su preprocesado)
- Turístico, multicultural, familiar, comercial, histórico

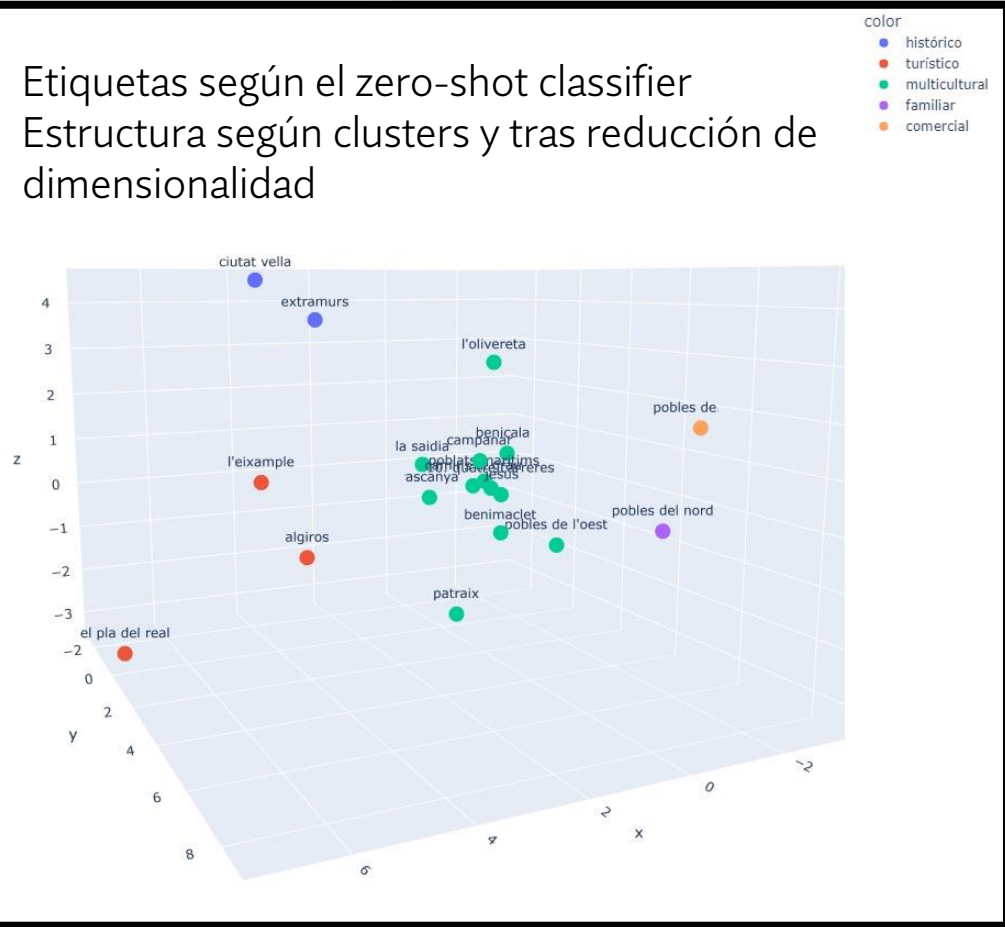
```
# junta los dos dataframes por orden de filas
df_conjunto = pd.concat([df, df_jesus], axis=1)
```



5 clases de manera natural en un clustering con KMeans() con el df\_conjunto

## Visualización tras PCA

Etiquetas según el zero-shot classifier  
Estructura según clusters y tras reducción de dimensionalidad



# Conclusiones

- Las opiniones que halló Isaac parecen **más concisas** y muestran la esencia verdadera del distrito, respecto a las de *Inside Airbnb* que pueden estar algo sesgadas.
- El **número de clusters** que detectaba Silhouette coincidía bien con la **estructura** a la hora de clasificar los **distritos**, tal y como mostraba el plot anterior



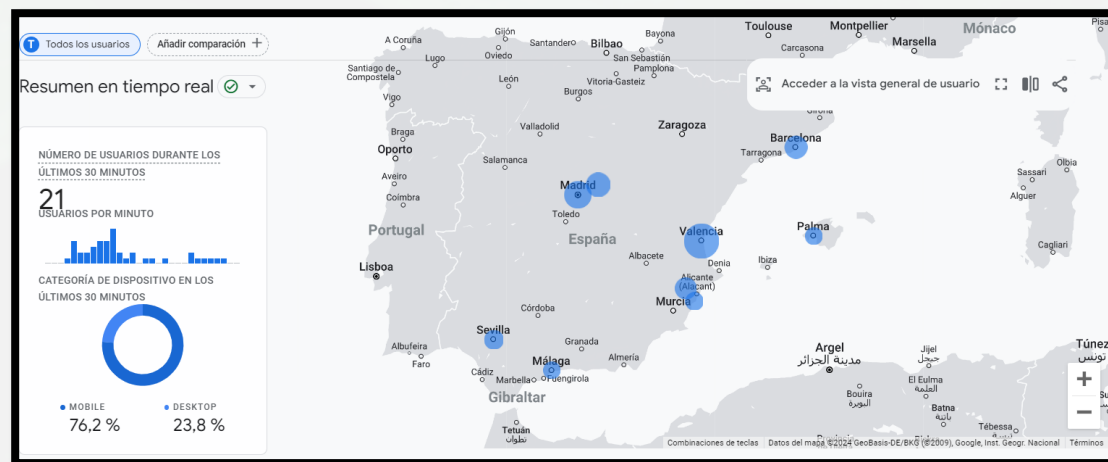
# Actualización de nuestro blog

[\(consultatubarrio.blogspot.com\)](http://consultatubarrio.blogspot.com)





# ¡Nuestro blog está en auge! Lo hemos estado difundiendo



Federico Díaz 15 de mayo de 2024, 17:01

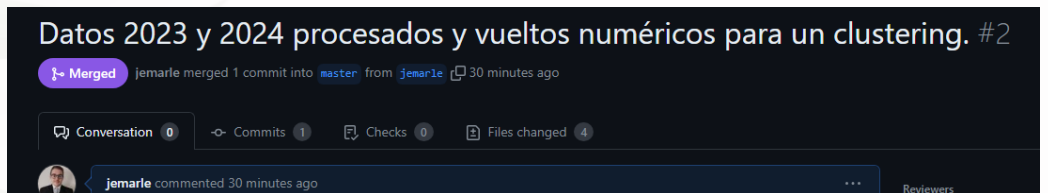
Descubrí el Blog por un viejo amigo mío, José. Desde el primer momento me quedé fascinado ante la narración tan absolutamente increíble que posee. Aparecen temas de actualidad que a todos nos carcomen por dentro, si sabes a lo que me refiero, jeje. Un abrazo desde Palma!

[RESPONDER](#) [ELIMINAR](#)

# GitHub



- Cada uno fue trabajando en su rama
- Se hacía un *pull request* de datos necesarios a rama *master*





**imargar3** Actualización de información de distritos



87a7f84 · 2 days ago



**Añadida exportacion de similitud a mapa. Añadidos distritos al mapa**



esedesam committed 4 hours ago



**nuevos cambios!**

jemarle committed 5 hours ago

Modificado el notebook de Samu para ver cómo quedan las cosas con las opiniones que saqué de InsideAirbnb. También he probado a quitar la clase 'suburbano' para las predicciones.

jemarle committed 16 hours ago

Commits on May 14, 2024

**Añado la descarga de reviews de barrios/distritos de Inside Airbn**

jemarle committed 2 days ago



**j**

Carsanpo committed 48 minutes ago

**probando cosas de zs-c**

Carsanpo committed 48 minutes ago

**info completa con Jesus. Ahora toca zero-shot**

Carsanpo committed 2 hours ago

**datos conjuntos con Jesus**

Carsanpo committed 3 hours ago



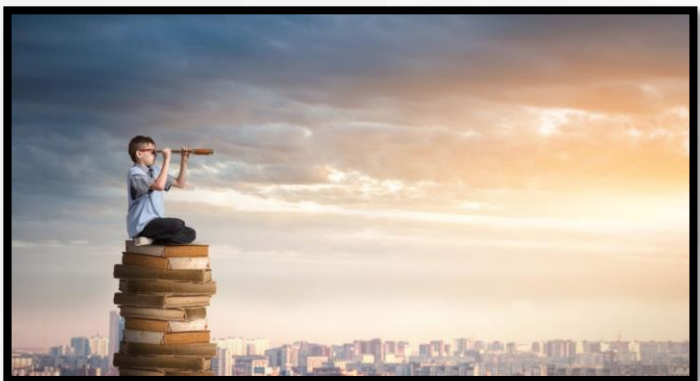
mayo 06, 2024

**¡CONOCE LOS MEJORES BARRIOS PARA UNA VIDA EN FAMILIA!**

Compartir · 3 comentarios



# Pensemos en el futuro...



- Interfaz gráfica geoespacial interactiva

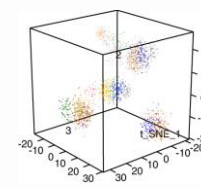


Sí, también se podrá entrar desde Android :D

- ¿Informe final?



- Clasificación definitiva de los barrios + reorganización





# Gracias

P. Sherman

Calle Wallaby 42

<https://consultatubarrio.blogspot.com>



No nos quites la beca

	zero-shot-classification
Ciutat Vella	turístico (0.29), histórico (0.25)
L'Eixample	multicultural (0.22), familiar (0.22)
Extramurs	familiar (0.30), histórico (0.24)
Campanar	comercial (0.41), multicultural (0.20)
La Zaidia	familiar (0.32), multicultural (0.27)
El Pla del Real	histórico (0.37), multicultural (0.22)
L'Olivereta	turístico (0.35), multicultural (0.35)
Patraix	multicultural (0.32), comercial (0.23)
Jesús	multicultural (0.32), familiar (0.27)