



CLASIFICACIÓN DE DISTRITOS DE VALENCIA

ANTES DE NADA... ¡PARTICIPA!

¿Te atreves a ser el experto?

Rellena la **encuesta** y clasifica tú mismo los distritos de Valencia.



<https://qrco.de/distritosvlc>



¡Tu atención tiene premio!

Escucha atentamente la presentación y responde a las **preguntas**.



<https://www.menti.com/aldphi573ycw>



ÍNDICE



INTRODUCCIÓN



DATOS USADOS



METODOLOGÍA



RESULTADOS



CONCLUSIONES

DISTRIBUCIÓN DE CONTENIDOS

- OBJETIVOS
- ORGANIZACIÓN DEL TRABAJO
- GITHUB



ÍNDICE



INTRODUCCIÓN



DATOS USADOS



METODOLOGÍA



RESULTADOS



CONCLUSIONES

DISTRIBUCIÓN DE CONTENIDOS

- TIPOS
- LIMITACIONES
- PRIVACIDAD



ÍNDICE



INTRODUCCIÓN



DATOS USADOS



METODOLOGÍA



RESULTADOS



CONCLUSIONES

DISTRIBUCIÓN DE CONTENIDOS

- DIAGRAMA DE FLUJO
- MÉTODOS
- VENTAJAS Y DESVENTAJAS



ÍNDICE



INTRODUCCIÓN



DATOS USADOS



METODOLOGÍA



RESULTADOS



CONCLUSIONES

DISTRIBUCIÓN DE CONTENIDOS

- CLUSTERS Y CATEGORÍAS
- CLASIFICACIÓN
- COMPARACIÓN Y DISCUSIÓN



ÍNDICE



INTRODUCCIÓN



DATOS USADOS



METODOLOGÍA



RESULTADOS



CONCLUSIONES

DISTRIBUCIÓN DE CONTENIDOS

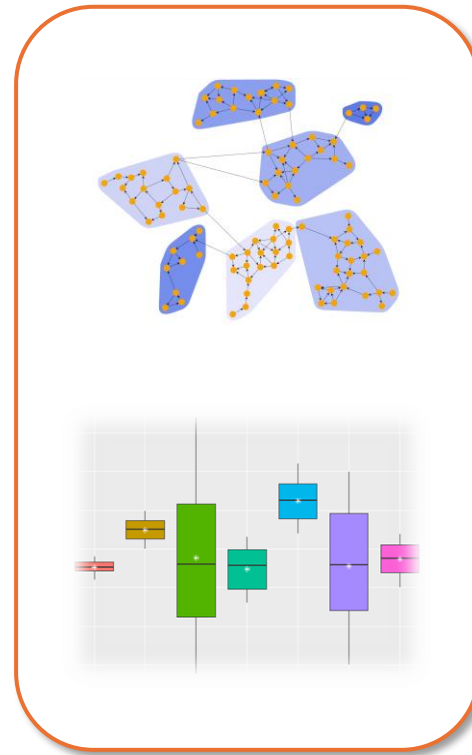
- ALGORITMOS Y MÉTODOS
- CLASIFICACIÓN FINAL



OBJETIVOS



Recopilación de datos

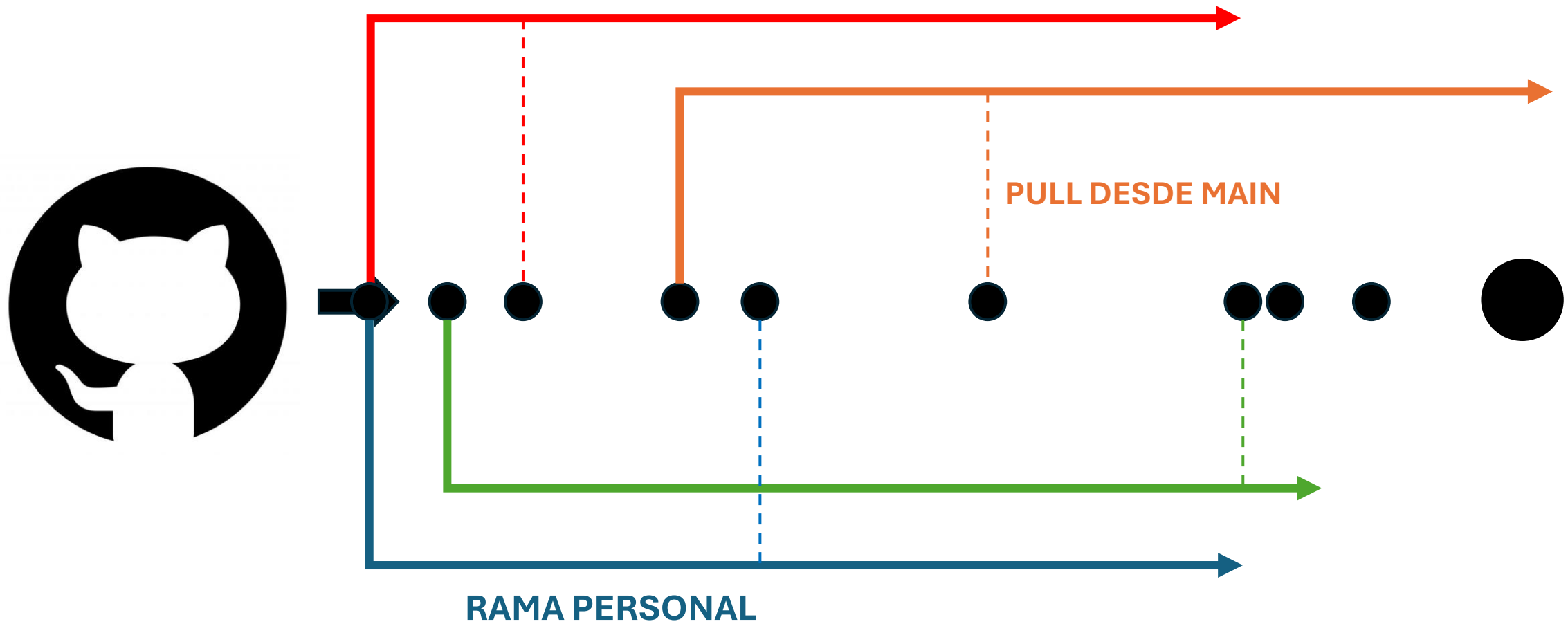


Representación y
modelado

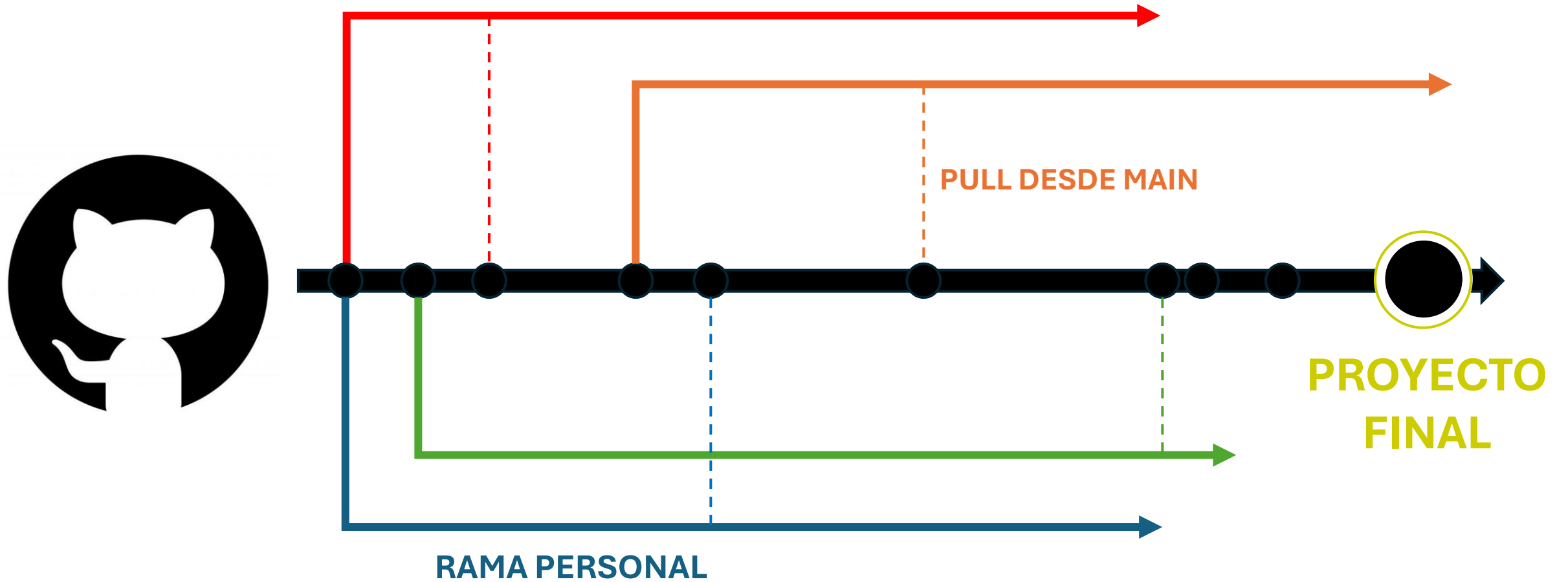


Clasificación de
distritos

ORGANIZACIÓN



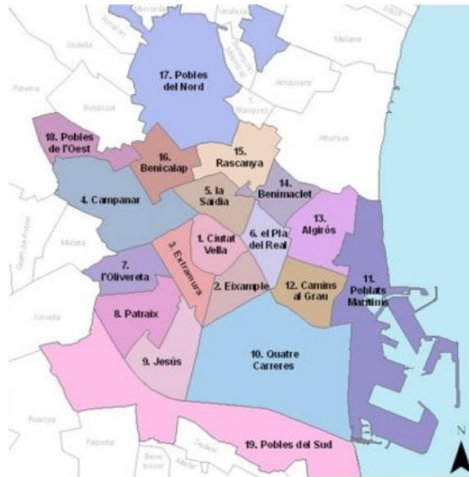
ORGANIZACIÓN



DATOS ESTRUCTURADOS



**OFICINA DE
ESTADÍSTICA**



Distrito:

Seleccione el año deseado:



LIMITACIONES

- Datos distintos por distrito
- Exhaustiva revisión de documentos individuales
- Datos faltantes
- Errores en los cálculos de estadísticos



SEGURIDAD

Pregunta 19.1

EN UNA ESCALA DEL 0 AL 10 DONDE 0 ES “MUY INSEGURO” Y 10 “MUY SEGURO”, ¿CÓMO VALORARIA USTED SU BARRIO DE RESIDENCIA?

	Porcentaje													(N)	Desv.		
	0	1	2	3	4	5	6	7	8	9	10	No sabe	No contesta		Media	Típ.	Mediana
Total	1,4	0,6	1,3	3,1	5,2	10,4	14,2	24,1	25,5	9,8	4,4	0,0	0,0	(2.356)	6,8	1,9	7
1. Ciutat Vella	0,0	0,9	0,9	0,9	3,5	4,4	20,2	28,1	21,1	13,2	7,0	0,0	0,0		7,2	1,6	7
2. l'Eixample	0,0	0,0	0,8	0,8	1,6	1,6	15,9	27,0	34,1	16,7	1,6	0,0	0,0		7,4	1,3	8
3. Extramurs	0,8	0,0	3,1	2,3	3,1	4,6	5,3	18,3	29,8	20,6	12,2	0,0	0,0		7,5	2,0	8
4. Campanar	0,0	0,0	0,8	0,8	4,2	12,7	18,6	16,9	28,0	17,8	0,0	0,0	0,0		7,0	1,6	7
5. la Saïdia	3,9	0,0	3,1	9,4	14,8	21,1	20,3	13,3	8,6	2,3	3,1	0,0	0,0		5,4	2,1	5
6. el Pla del Real	0,0	0,0	0,0	0,0	1,7	4,3	14,5	16,2	53,0	10,3	0,0	0,0	0,0		7,5	1,1	8
7. l'Olivereta	2,3	0,0	1,5	3,1	5,4	6,9	20,0	25,4	23,1	5,4	6,9	0,0	0,0		6,7	2,0	7
8. Patraix	0,0	0,0	0,0	0,7	1,5	4,4	14,7	33,8	28,7	11,8	4,4	0,0	0,0		7,3	1,3	7
9. Jesús	2,3	0,0	1,5	2,3	6,0	9,0	9,8	31,6	27,1	7,5	3,0	0,0	0,0		6,7	1,9	7

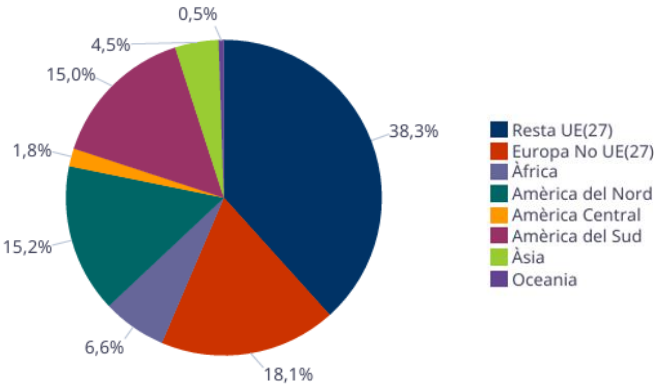
Padrón a 01/01/2023. Barrio 1.1. la Seu

1.3. Población por sexo y edad (grandes grupos)

1.3. Població per sexe i edat (grans grups)

	Total	0-15	16-64	65 y más
Total	3,151	350	2,183	618
Hombres	1,514	177	1,065	272
Mujeres	1,637	173	1,118	346

Varones / Homes



1.5. Población según lugar de nacimiento y sexo

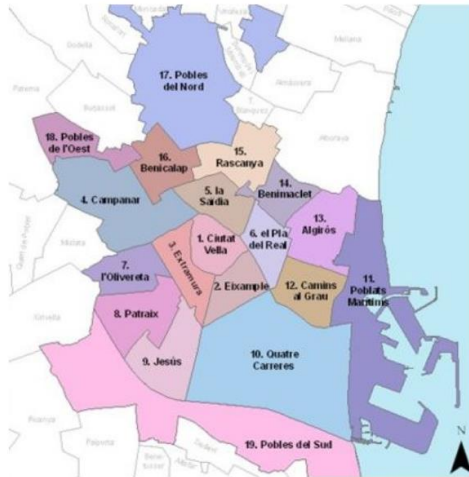
1.5. Població segons lloc de naixement i sexe

	Total	València	Resto de l'Horta	Resto de la Comunidad	Resto del Estado	Extranjero
Total	3,151	1,404	106	311	368	962
Hombres	1,514	672	46	160	194	442
Mujeres	1,637	732	60	151	174	520

DATOS ESTRUCTURADOS



**OFICINA DE
ESTADÍSTICA**



Distrito:

Seleccione el año deseado:



LIMITACIONES

- Datos distintos por distrito
- Exhaustiva revisión de documentos individuales
- Datos faltantes
- Errores en los cálculos de estadísticos

DATOS NO ESTRUCTURADOS

* VALENCIA SECRETA

Suscribirse



QUÉ HACER NOTICIAS GASTRONOMÍA CULTURA ESCAPADAS NATURALEZA Y BIENESTAR GUÍAS SECRETAS

Los 8 mejores barrios de Valencia para vivir

Ruzafa, Arrancapins o Patraix son algunos de los barrios de Valencia que forman parte de nuestra selección de mejores barrios para vivir en la ciudad.

The screenshot shows the Indomio website interface. At the top, there's a search bar with 'Ciutat Vella • València' entered. Below it, a row of filters includes 'Ventas', 'Casas - Pisos', 'Precio', 'Superficie', and 'Habitaciones'. The main content area shows '190 resultados para: casas y pisos en venta València' with 'Ciutat Vella' highlighted in a red box. Below this, a property listing is shown with a photo of a street, a 'NUEVO' badge, and a price of '€ 250.000' in a red box. The listing title is 'Piso de tres habitaciones a reformar, planta baja, El Carme, València'. A table of features is shown below the title, with a red box around it: '4 habitacion...', '106 m²', '1 baño', 'Planta B', 'No Ascensor', 'Terraza', and 'Trastero'. At the bottom, there's a 'MENSAJE' button and icons for a trash can and a heart.

4 habitacion...	106 m²	1 baño	Planta B
No Ascensor	Terraza	Trastero	

LIMITACIONES

- Gran falta de datos relacionados con la temática
- Políticas de privacidad de organizaciones
- Sesgo y subjetividad

PRIVACIDAD

Hay que asegurarse de hacer un uso correcto y legal de los datos

El Ayuntamiento de València permite reutilizar todos los datos e informaciones difundidos por medio de este web, siempre que no se indique lo contrario, con las condiciones siguientes:

- No alterar el contenido de la información.
- No desnaturalizar el sentido de la información.
- Mencionar la fuente de la información.
- No dar a entender que el Ayuntamiento patrocina o apoya la actividad en la que se reutiliza la información.



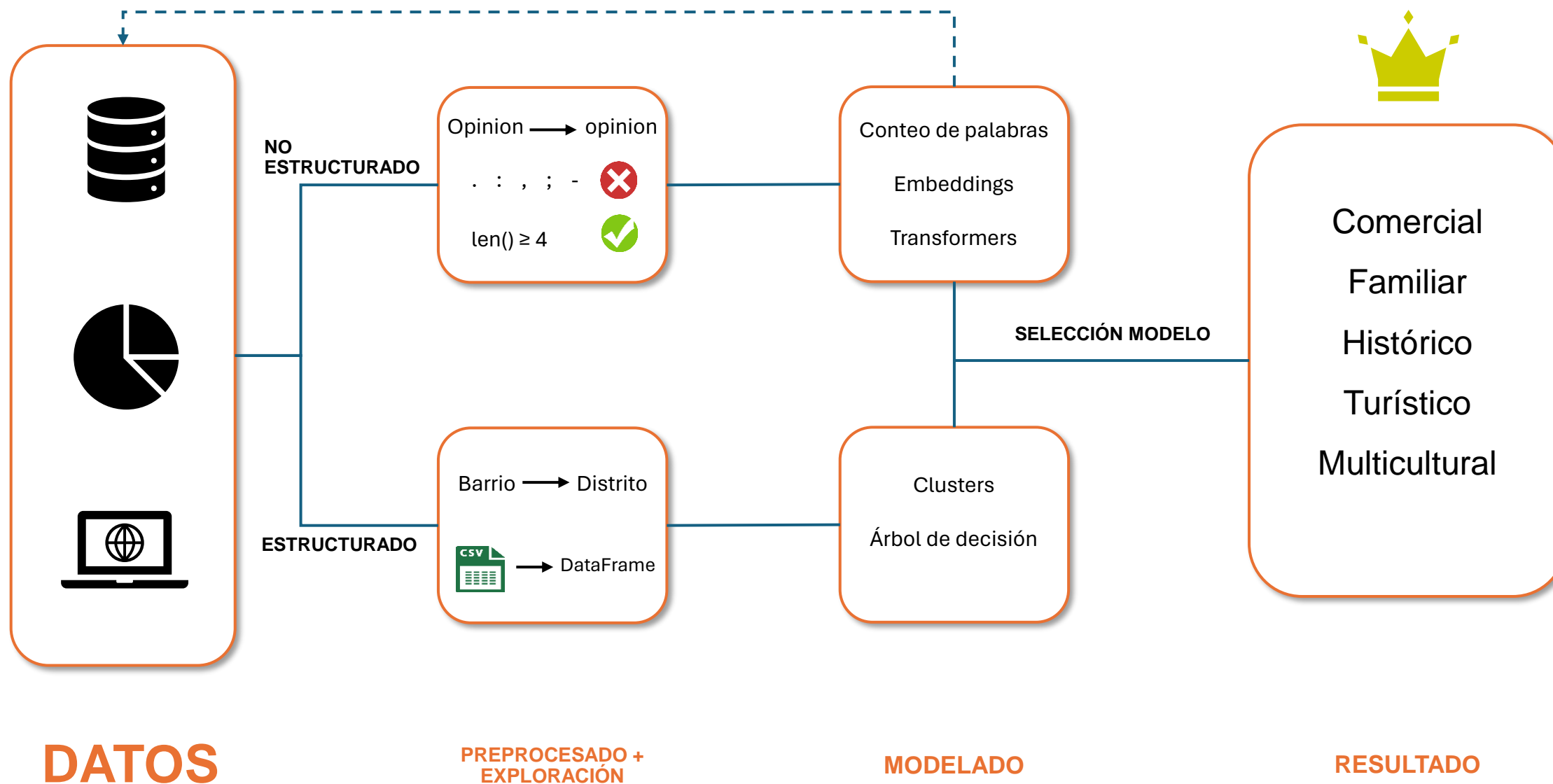
2. DERECHOS DE PROPIEDAD INTELECTUAL E INDUSTRIAL

El sitio web, incluyendo a título enunciativo, pero no limitativo su programación, edición, compilación y demás elementos necesarios para su funcionamiento, los diseños, logotipos, texto y/o gráficos, son propiedad del RESPONSABLE o, si es el caso, dispone de licencia o autorización expresa por parte de los autores. Todos los contenidos del sitio web se encuentran debidamente protegidos por la normativa de propiedad intelectual e industrial, así como inscritos en los registros públicos correspondientes.

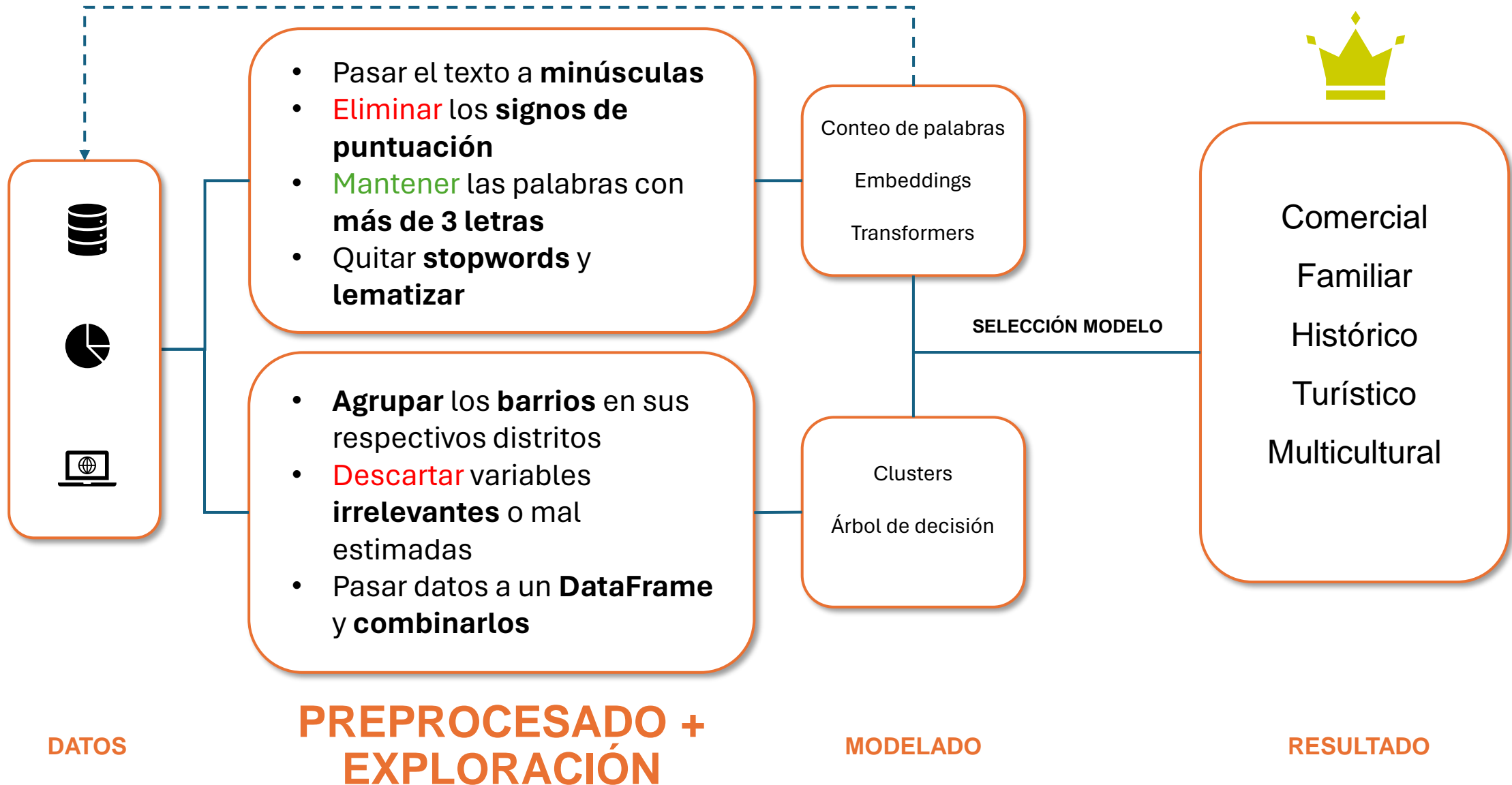
Independientemente de la finalidad para la que fueran destinados, la reproducción total o parcial, uso, explotación, distribución y comercialización, requiere en todo caso la autorización escrita previa por parte del RESPONSABLE. Cualquier uso no autorizado previamente se considera un incumplimiento grave de los derechos de propiedad intelectual o industrial del autor.



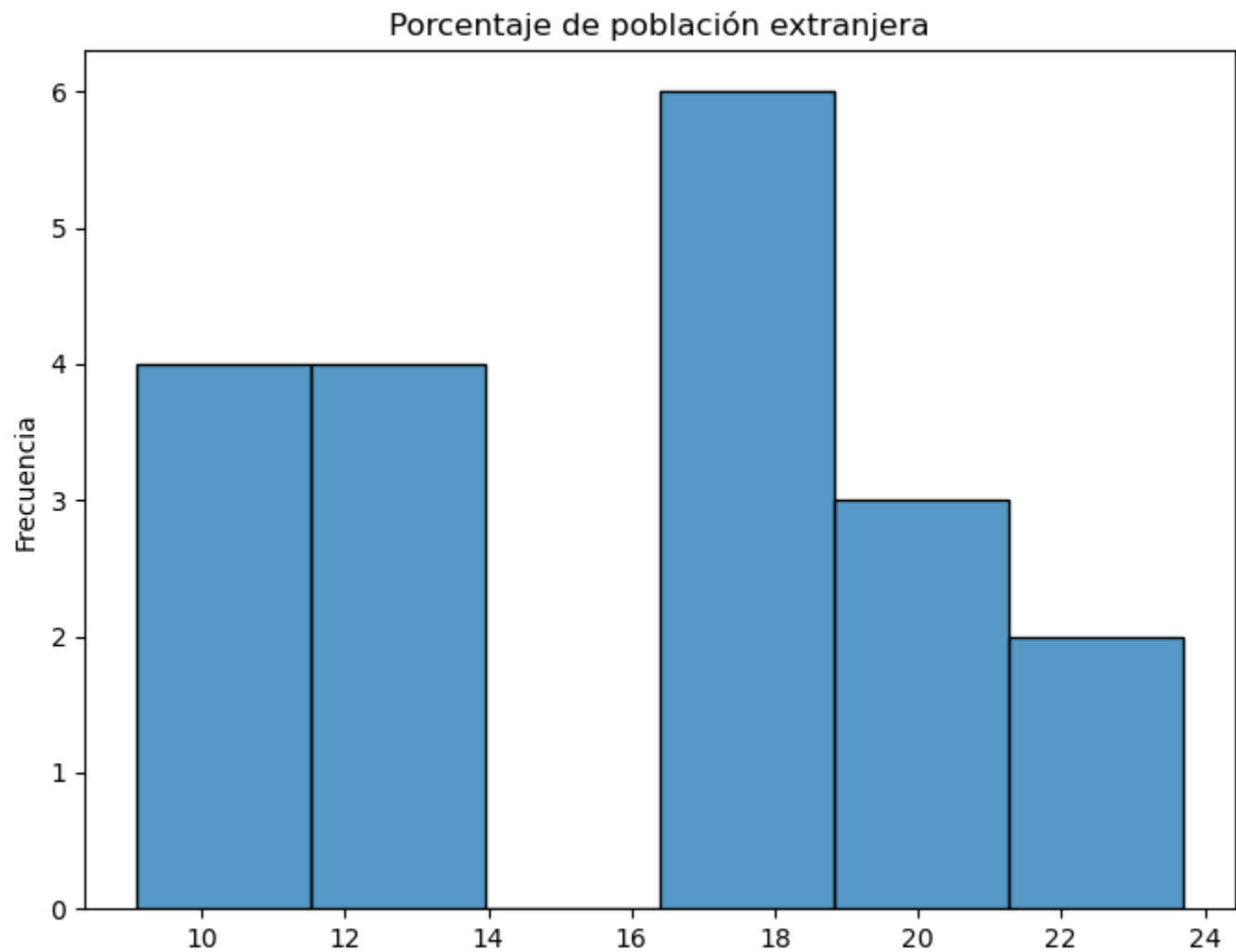
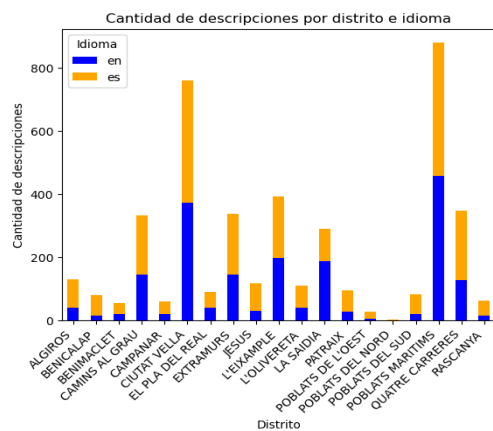
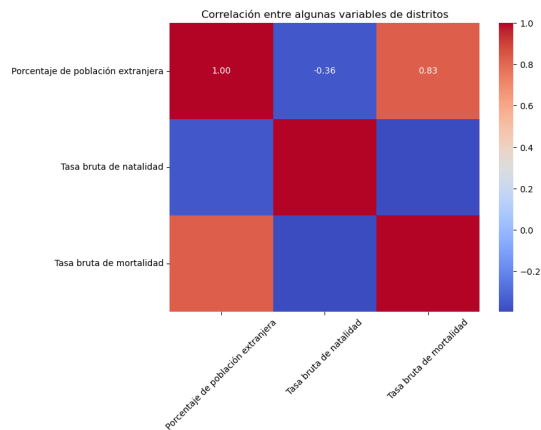
FLUJO DE TRABAJO



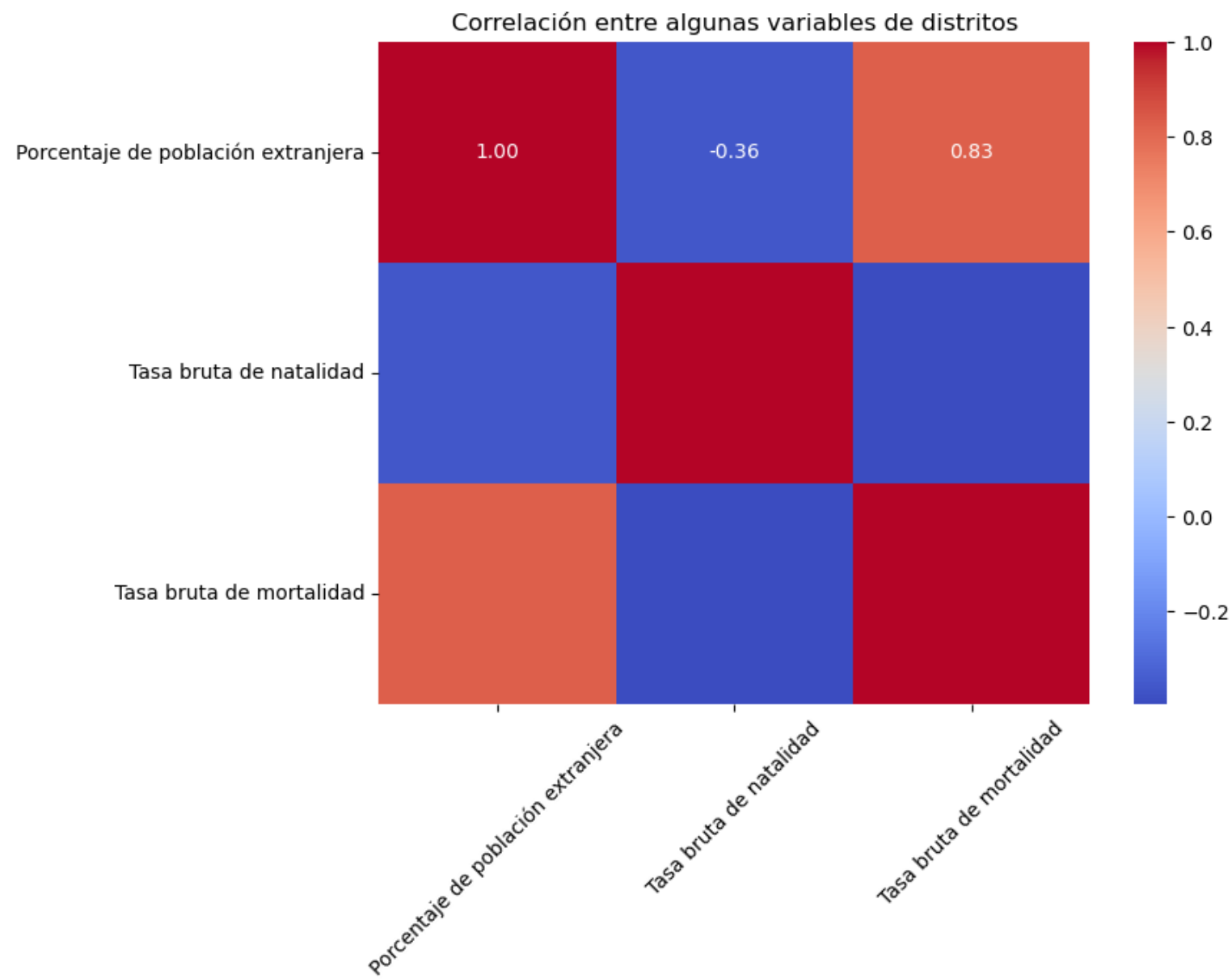
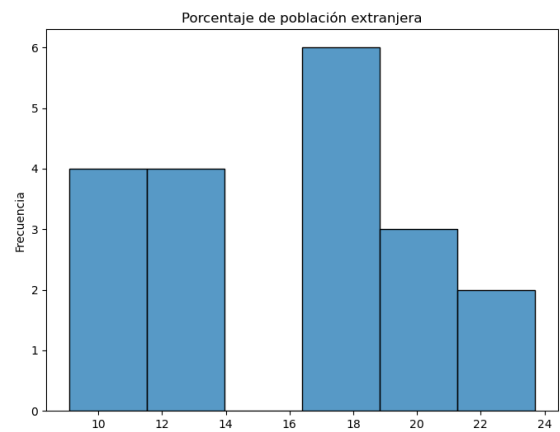
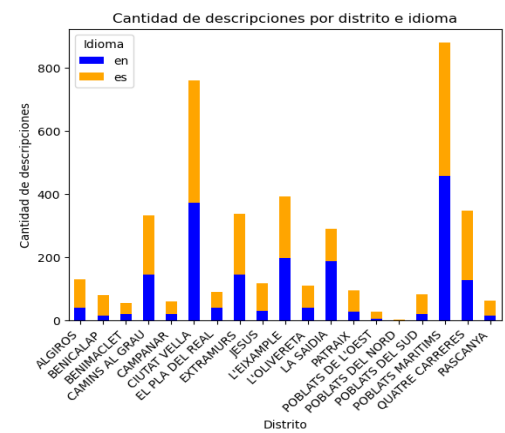
FLUJO DE TRABAJO



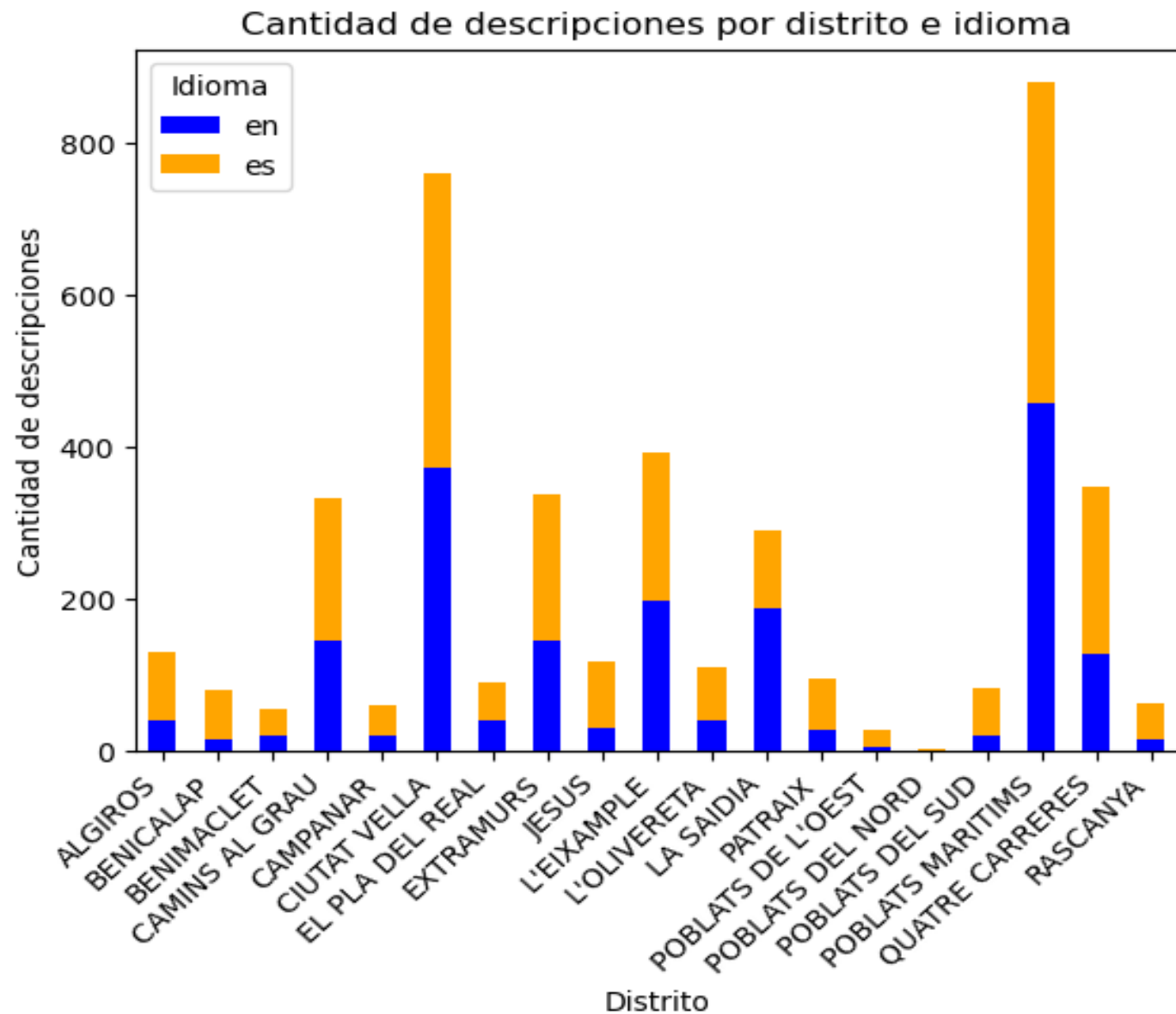
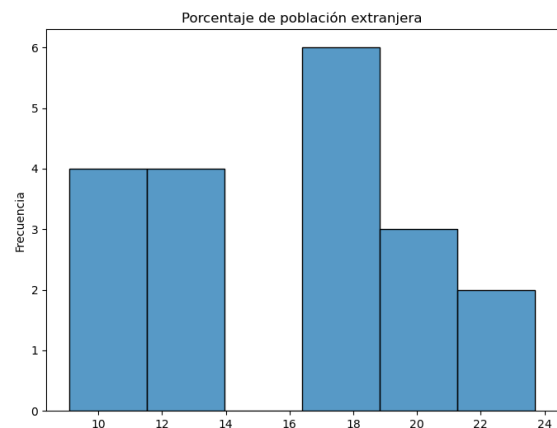
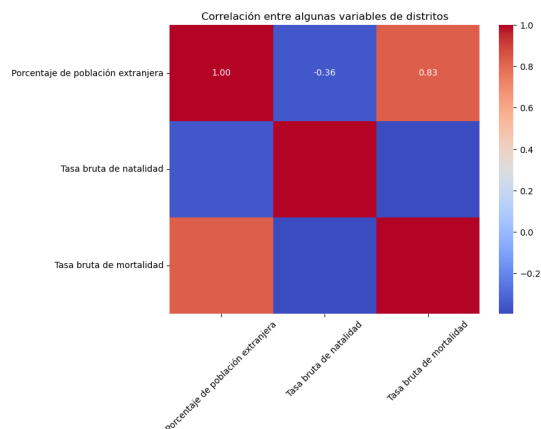
PREPROCESADO + EXPLORACIÓN



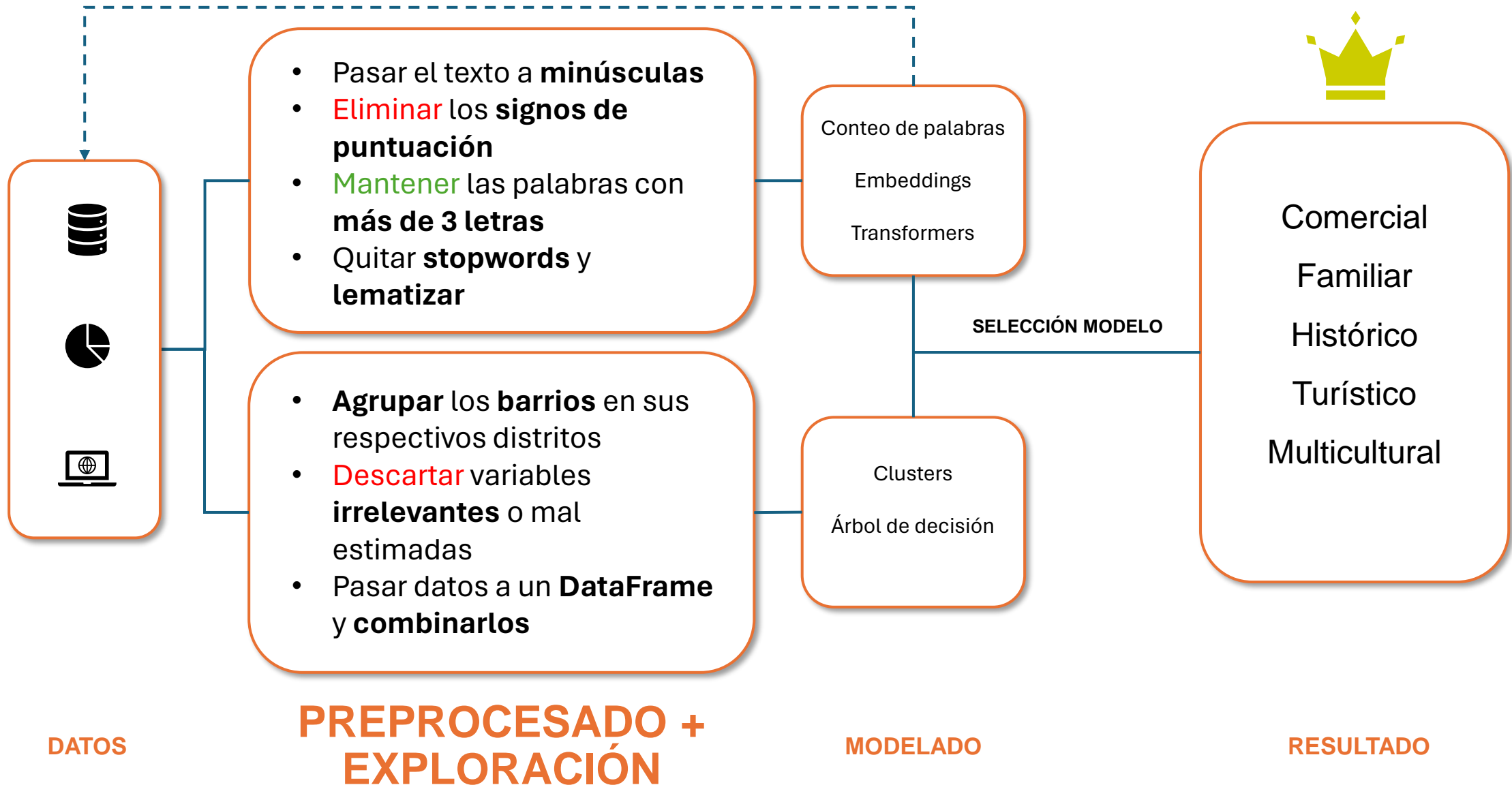
PREPROCESADO + EXPLORACIÓN



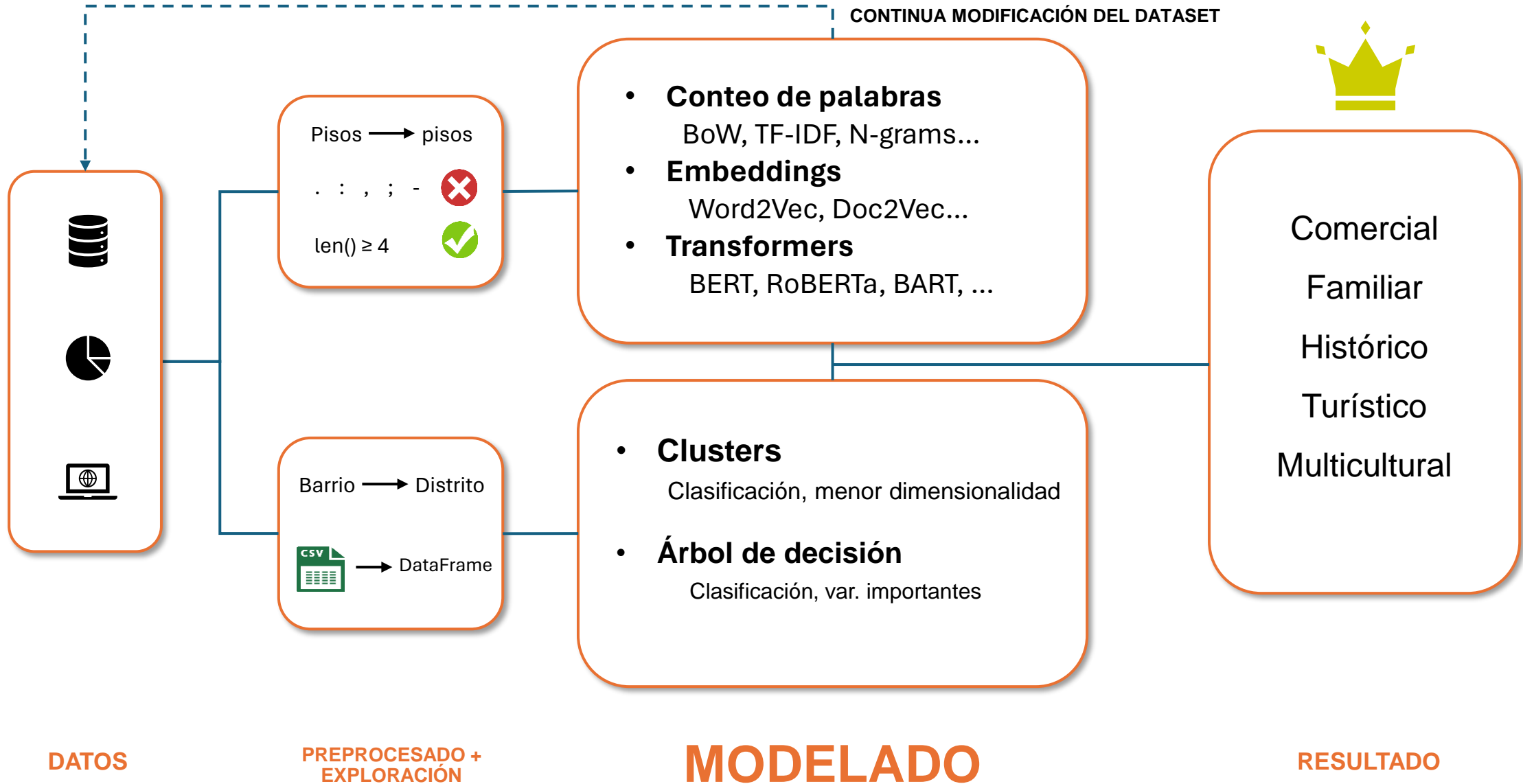
PREPROCESADO + EXPLORACIÓN



FLUJO DE TRABAJO



FLUJO DE TRABAJO



Bag of Words

- Se genera un **vocabulario** común para todos los documentos con todos los términos únicos.
- Se crea una matriz de **dimensiones $n \times m$** (n documentos, m términos).
- Cada **texto** se representa en una **fila** de la matriz, donde cada **columna** representa la **frecuencia de aparición** de un término del vocabulario en ese texto.
- CountVectorizer()



CONTEO

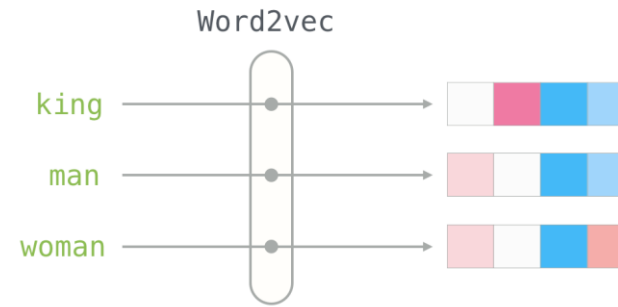
TF-IDF

- Representación sobre un vector BoW donde se aplica una **ponderación** para descontar los términos más comunes y promover los raros.
- Medida de **originalidad** de una palabra.
- TfidfVectorizer()



Embeddings

- **Word2vec**: Representación vectorial de los términos del vocabulario en un espacio vectorial continuo basado en la similitud semántica y de contexto.
- **Doc2vec**: Calcular el *embedding* del documento como la media de los *word embeddings* de sus términos.
- Spacy: `es_core_news_lg`



Transformers

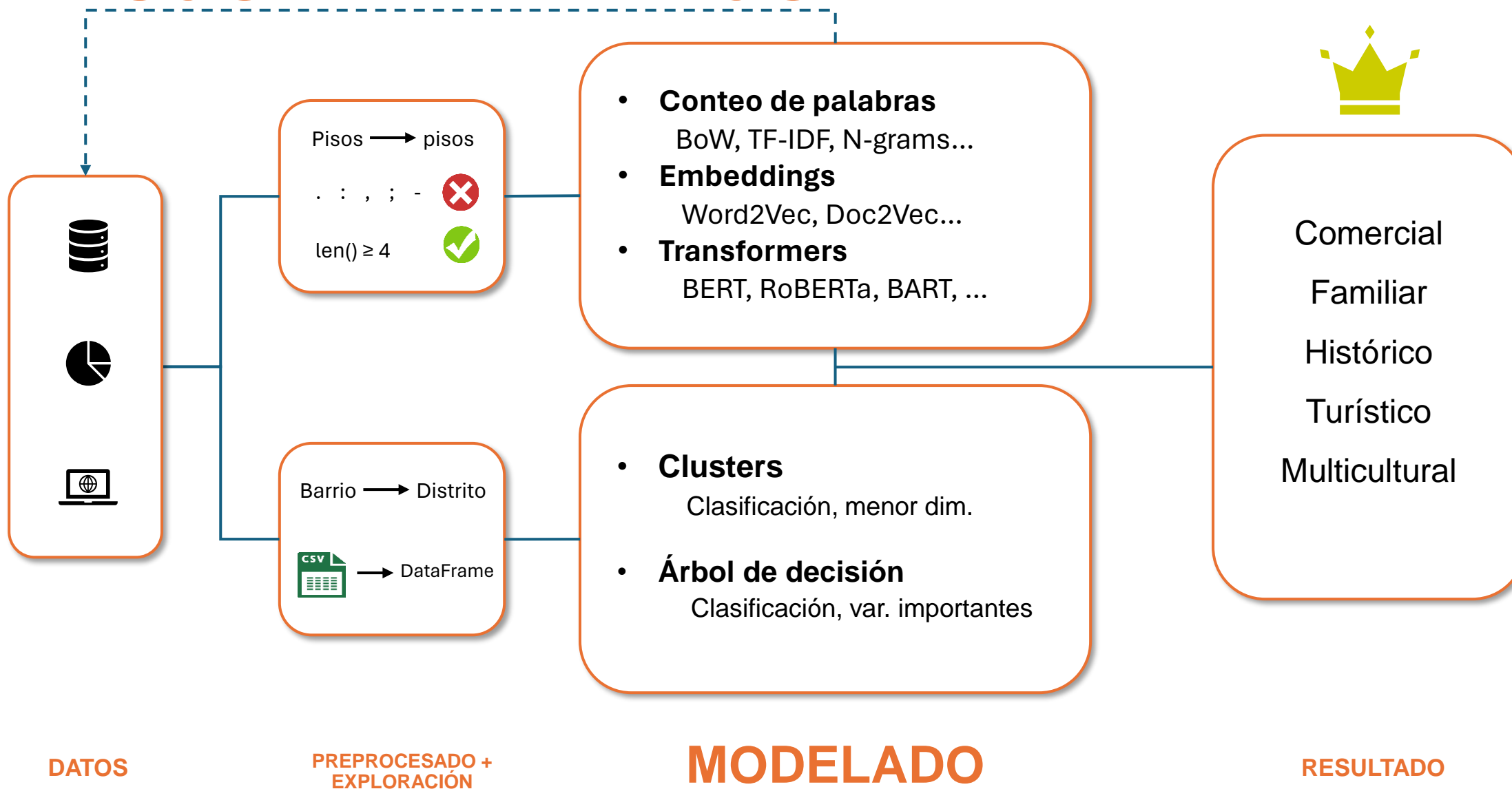
- Modelos de **aprendizaje profundo** que utilizan los *embeddings* de las palabras del texto como entrada.
- Hugging Face + transformers:
 - `bart-large-mnli` (BERT-like encoder + GPT-like decoder)
 - `mDeBERTa-v3-base-xnli-multilingual-nli-2mil7`
 - `bert-base-spanish-wwm-cased-xnli`



COMPARACIÓN ENTRE MODELOS

Característica	BoW	TF-IDF	Doc2Vec	Transformers
Representación	Matriz ocurrencia	Matriz ponderada por frecuencia	Vectores de documentos	Embeddings contextuales
Contexto	Lo ignora	Lo ignora	Captura contexto del documento	Contexto global y dependencias a largo plazo
Entrenamiento	Rápido y simple	Moderadamente rápido	Intensivo	Muy intensivo, muchos datos

FLUJO DE TRABAJO



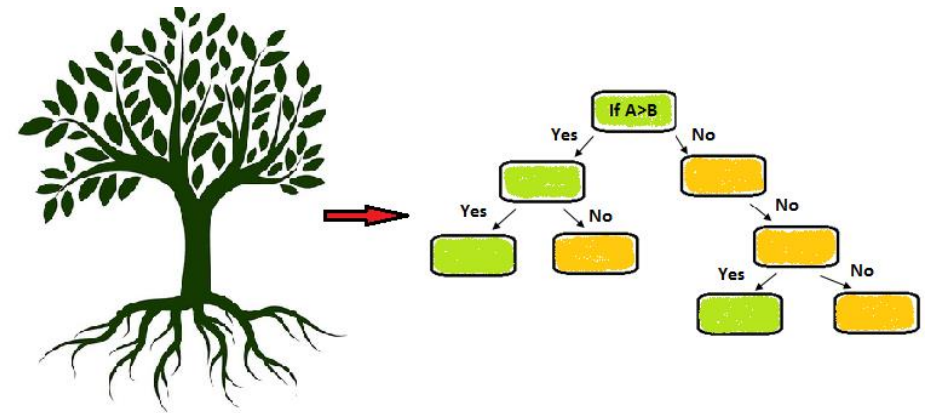
CLUSTERS

- **Uso de KMeans**, algoritmo básico de clustering.
- Permiten ver estructura subyacente en los datos estructurados que tenemos.
- Permitirán determinar el **número idóneo de etiquetas** a partir del *Silhouette score*.

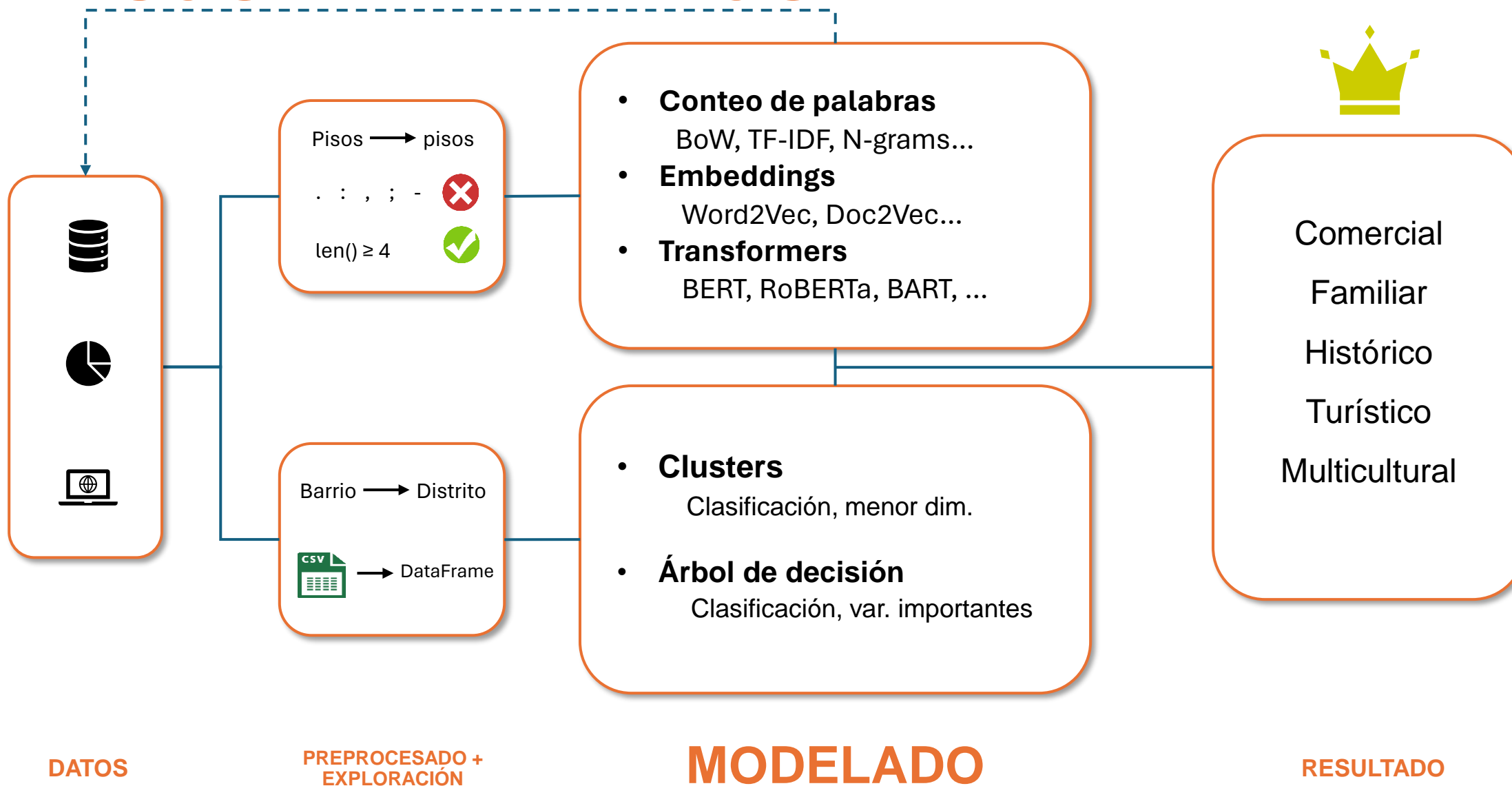


ÁRBOL DE DECISIÓN

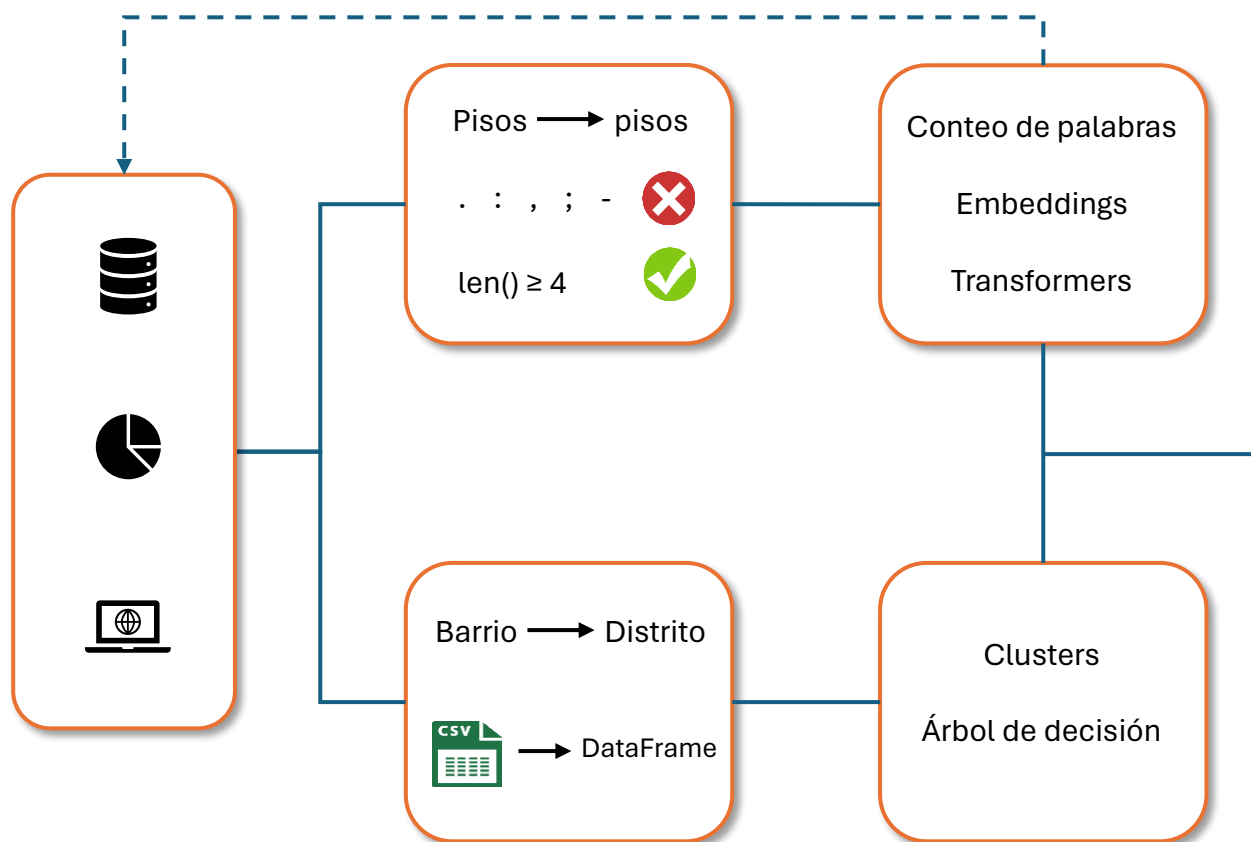
- Permite la creación de **reglas** basadas en separar los datos en cada nodo de la *mejor forma* según las **clases**.
- Con variables numéricas se utilizan *thresholds* que el algoritmo encuentra **óptimos** según la métrica usada.
- **Métricas:** Impureza de Gini, Entropía, ...
- `DecisionTreeClassifier()`



FLUJO DE TRABAJO



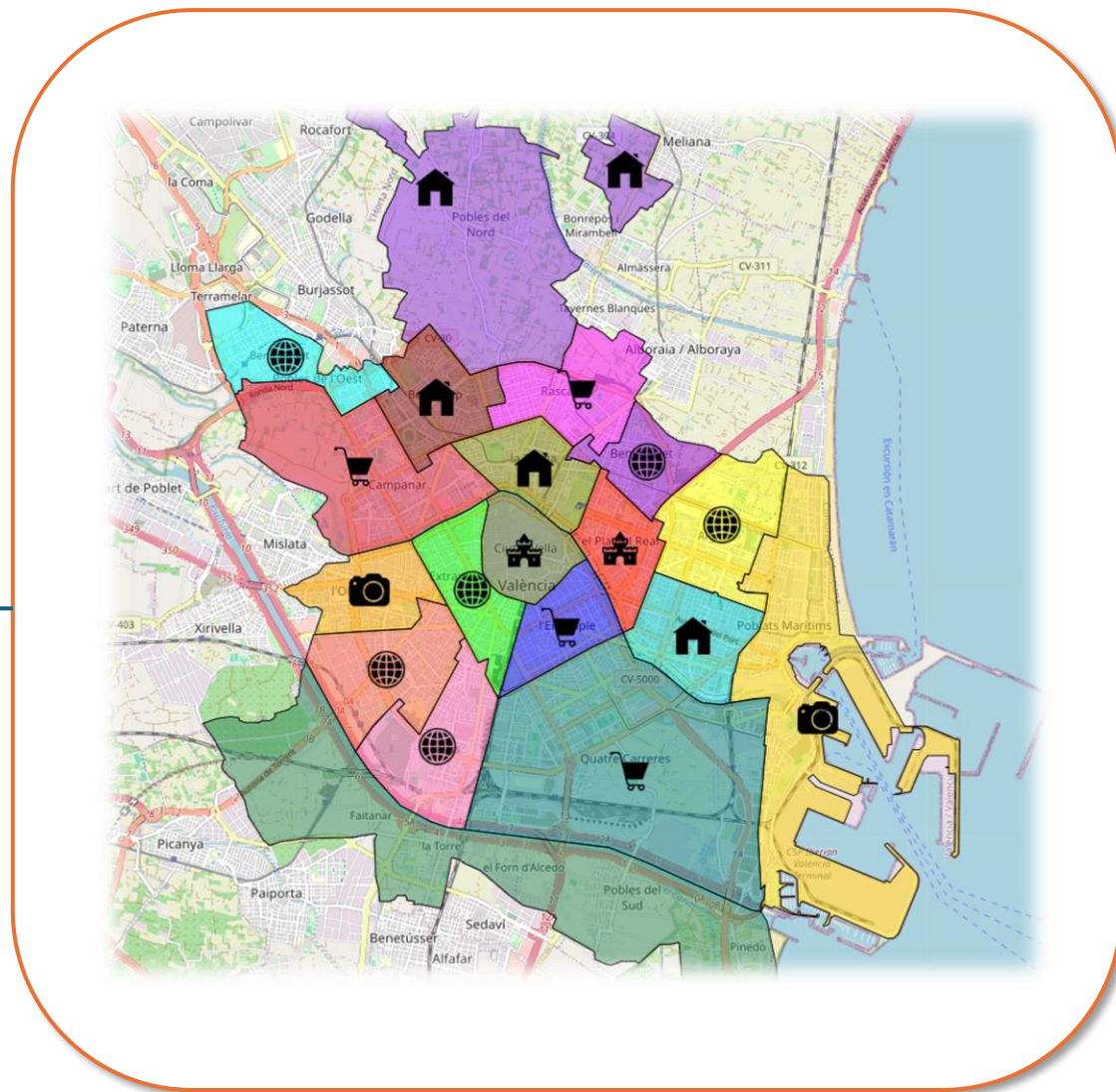
FLUJO DE TRABAJO



DATOS

PREPROCESADO +
EXPLORACIÓN

MODELADO



RESULTADO

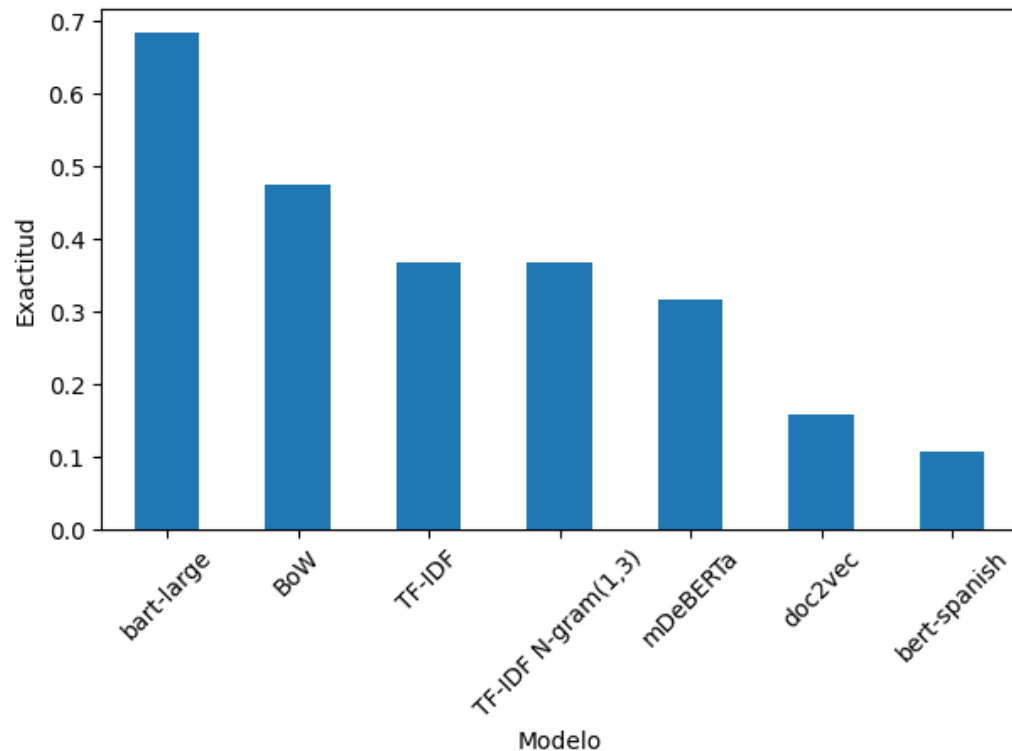
CLASIFICACIÓN

- **9 clases:** histórico, lujo, multicultural, familiar, universitario, turístico, alternativo, suburbano, comercial
- **7 modelos:** BoW, TF-IDF, TF-IDF N-gram, bart-large, mDeBERTa-v3, bert-base-spanish

Distrito	BoW	TF-IDF N-gram(1,3)	facebook/bart-large-mnli
Ciutat Vella	suburbano (0.27), histórico (0.22)	suburbano (0.04), histórico (0.03)	histórico (0.35), turístico (0.19)
L'Eixample	suburbano (0.32), comercial (0.27)	suburbano (0.06), universitario (0.05)	comercial (0.48), turístico (0.19)
Extramurs	turístico (0.20), comercial (0.19)	suburbano (0.04), multicultural (0.03)	multicultural (0.41), turístico (0.25)
Campanar	familiar (0.11), comercial (0.10)	comercial (0.01), suburbano (0.01)	comercial (0.36), turístico (0.28)
La Saldia	turístico (0.12), comercial (0.12)	comercial (0.01), turístico (0.01)	familiar (0.36), multicultural (0.32)
El Pla del Real	suburbano (0.19), familiar (0.10)	suburbano (0.02), familiar (0.02)	histórico (0.48), universitario (0.12)
L'Olivereta	comercial (0.10), suburbano (0.09)	comercial (0.01), lujo (0.01)	turístico (0.70), lujo (0.09)
Patraix	suburbano (0.31), familiar (0.23)	suburbano (0.05), familiar (0.04)	multicultural (0.67), turístico (0.07)
Jesus	turístico (0.12), comercial (0.12)	turístico (0.02), comercial (0.01)	multicultural (0.63), suburbano (0.08)

CLASIFICACIÓN

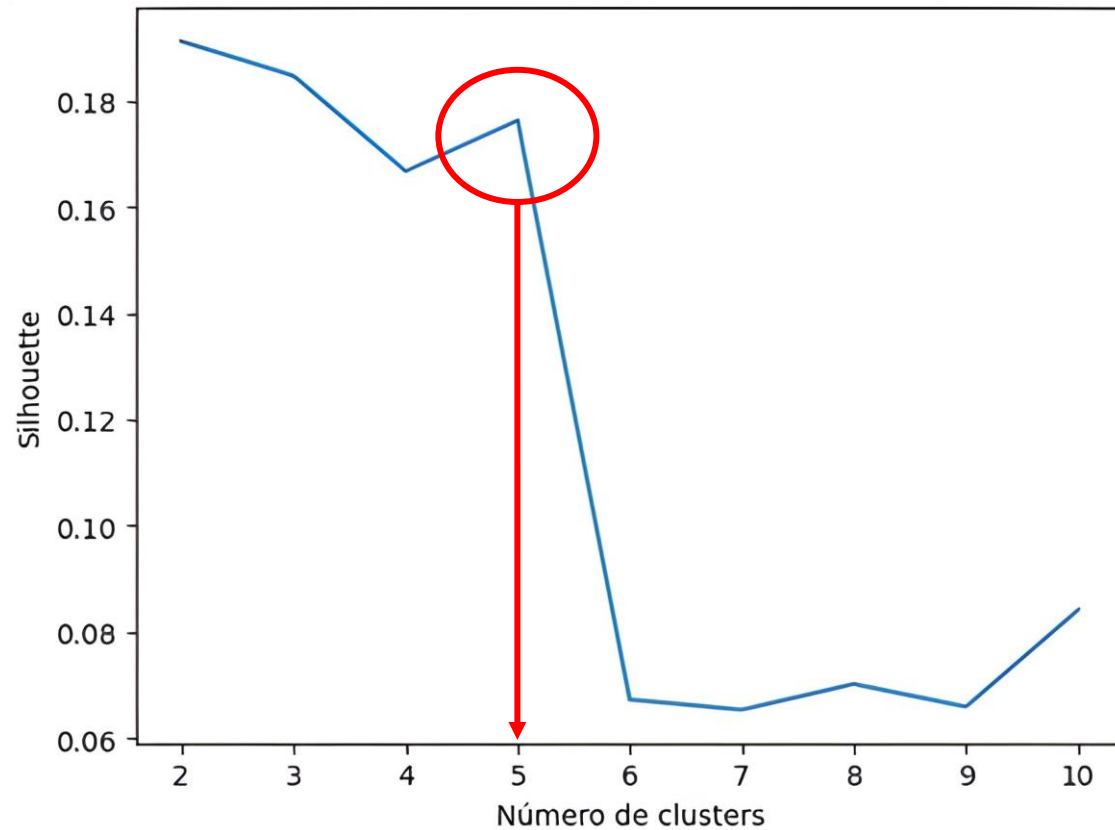
- Nueva clasificación con las **5 etiquetas**
- Comparación con la **opinión del experto**



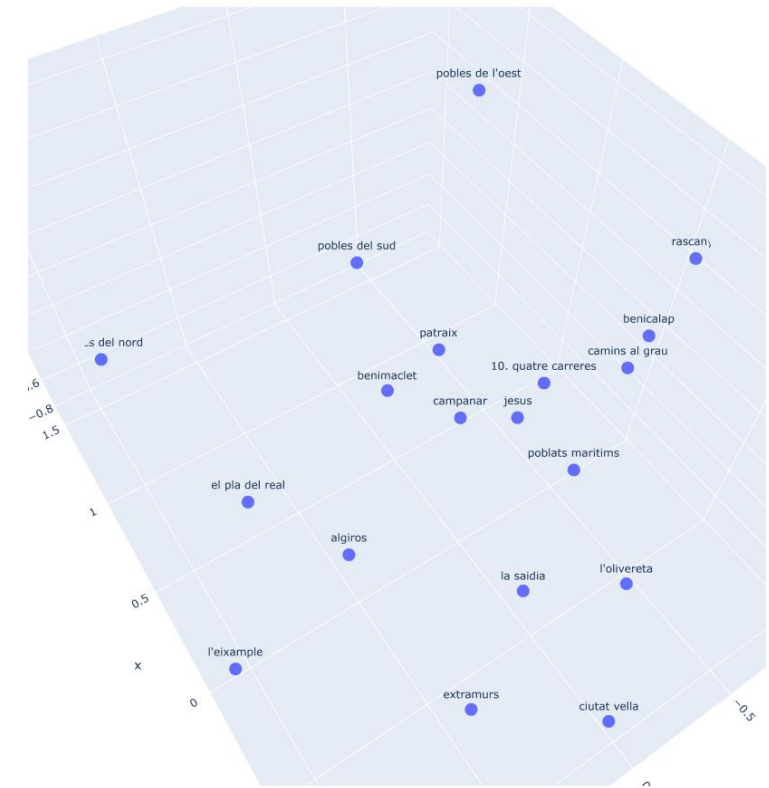
Distrito	Experto	bart-large
Ciutat Vella	histórico	histórico
L'Eixample	comercial	comercial
Extramurs	multicultural	multicultural
Campanar	familiar	comercial
La Saldia	familiar	familiar
El Pla del Real	histórico	histórico
L'Olivereta	multicultural	turístico
Patraix	multicultural	multicultural
Jesus	multicultural	multicultural
Quatre Carreres	turístico	comercial
Poblats Maritims	turístico	turístico
Camins al Grau	familiar	familiar
Algiros	multicultural	multicultural
Benimaclet	multicultural	multicultural
Rascanya	familiar	comercial
Benicalap	familiar	familiar
Poblats de L'oest	familiar	multicultural
Poblats del Nord	histórico	familiar
Poblats del Sud	familiar	familiar

CLUSTERING

SILHOUETTE

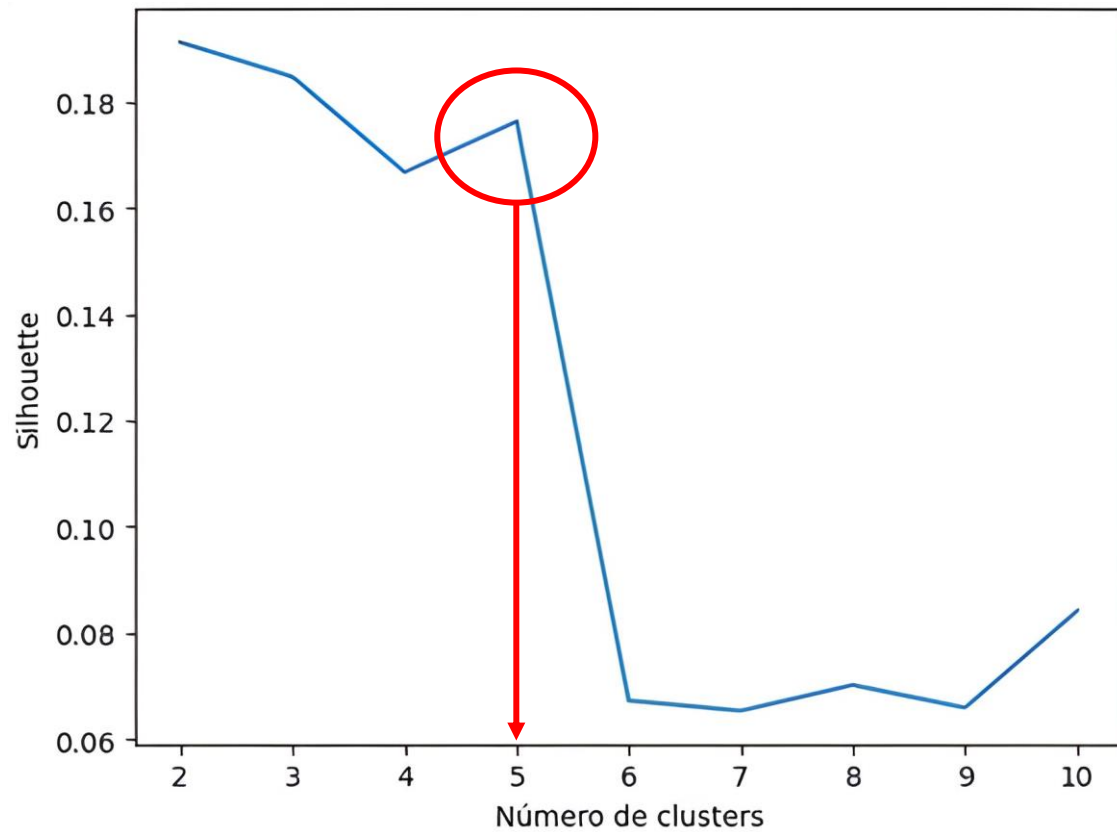


3-D CLUSTER

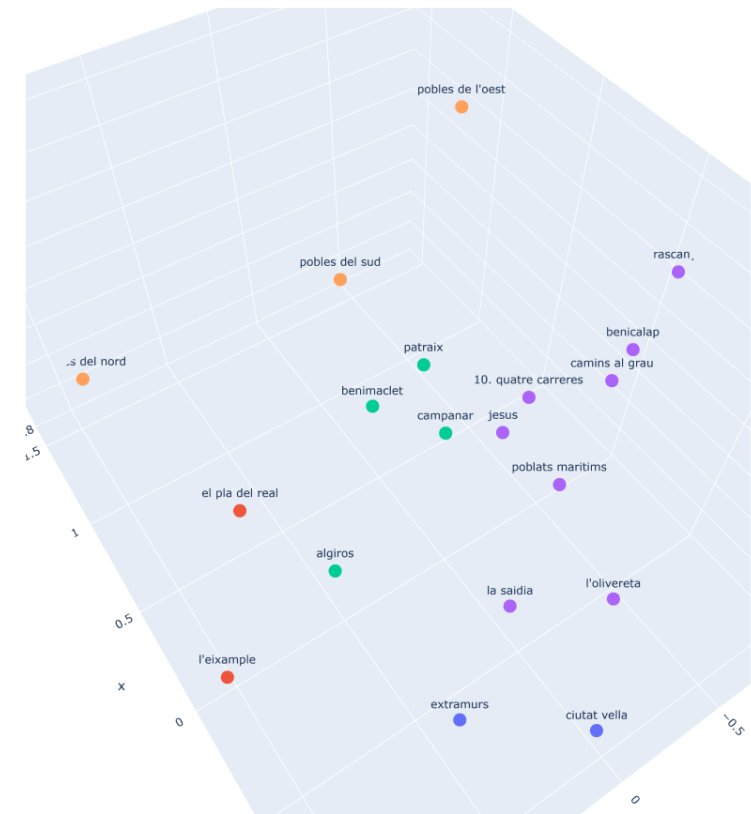


CLUSTERING

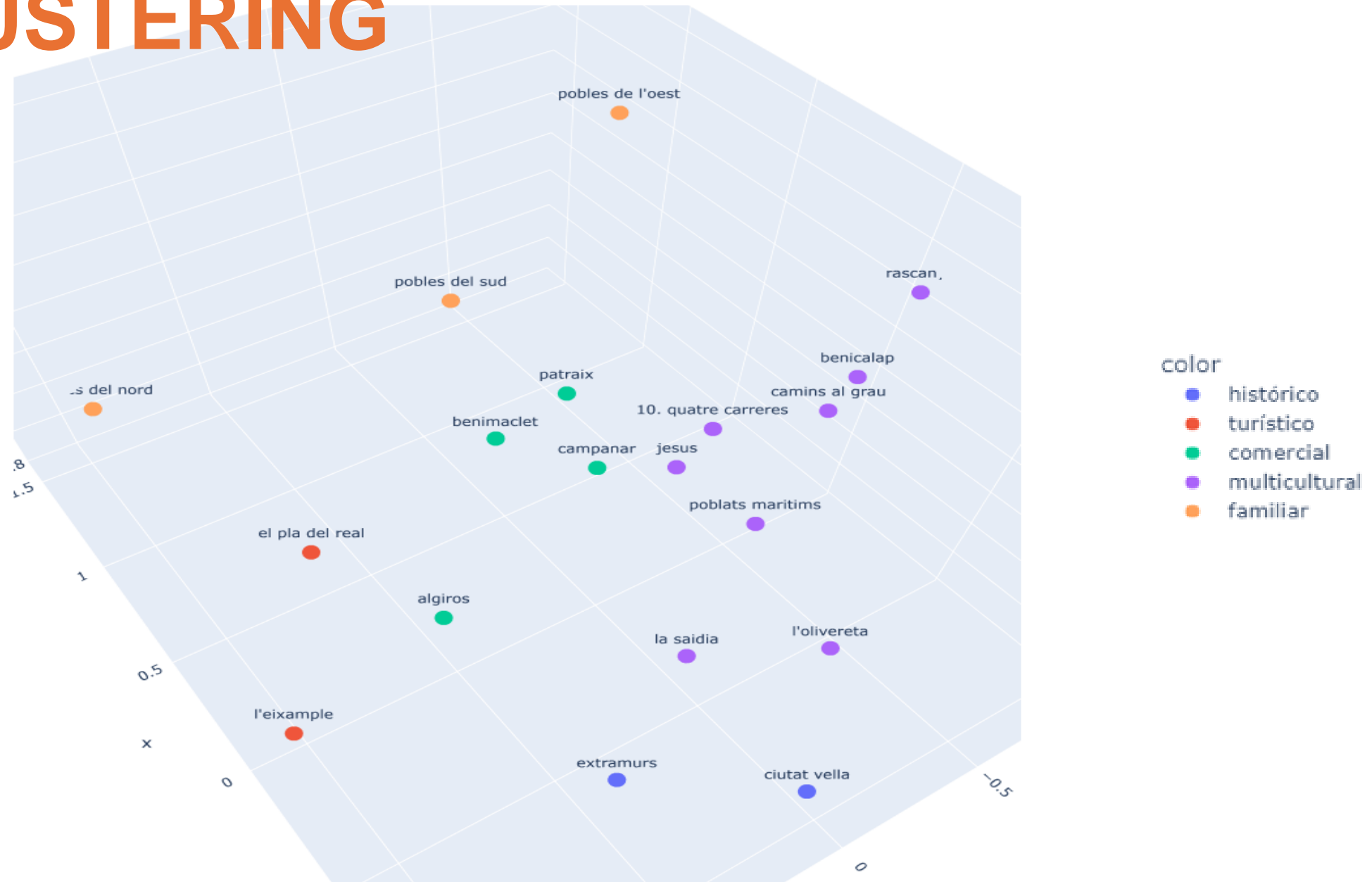
SILHOUETTE



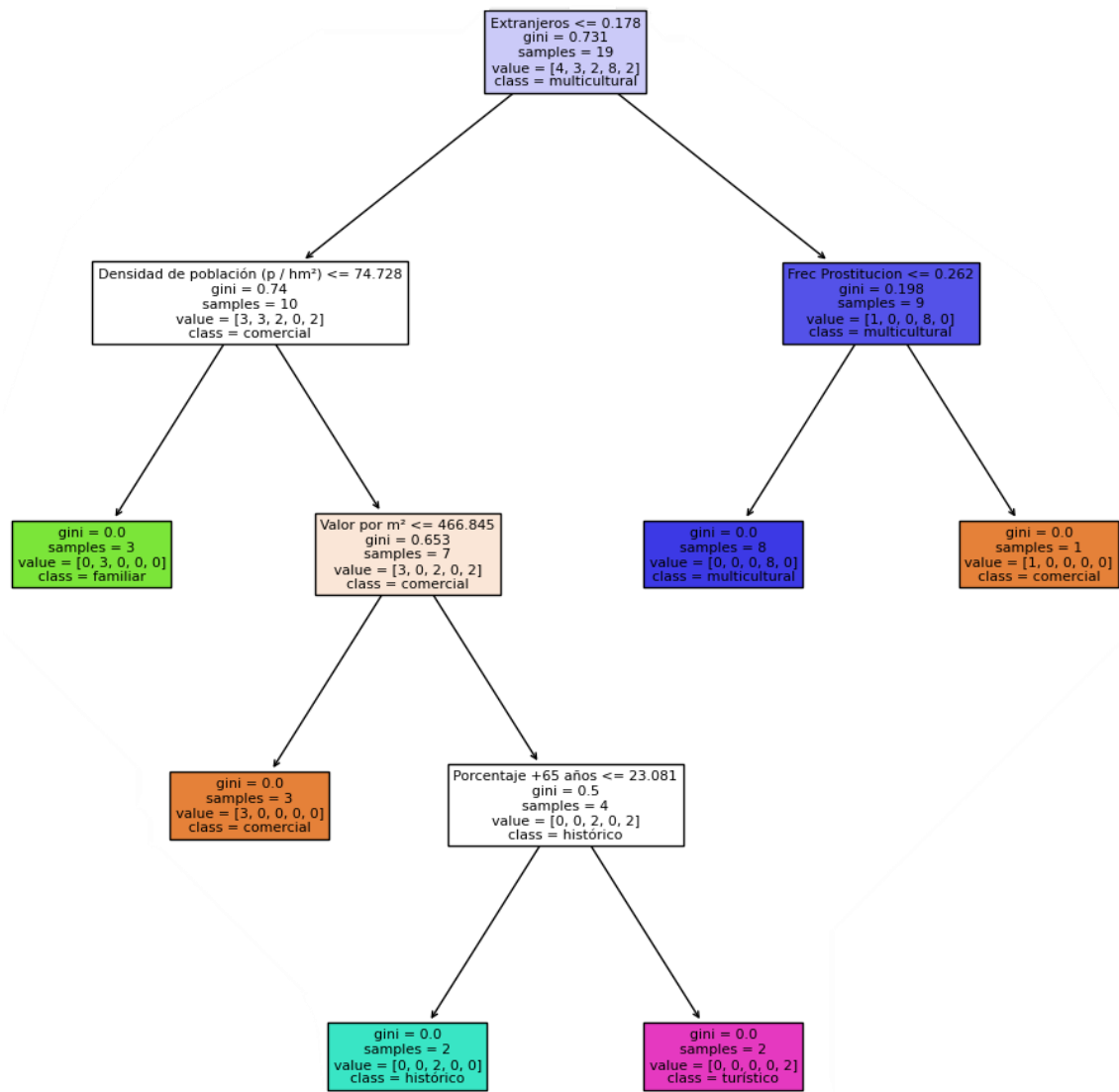
3-D CLUSTER



CLUSTERING



MODELO DE REGLAS

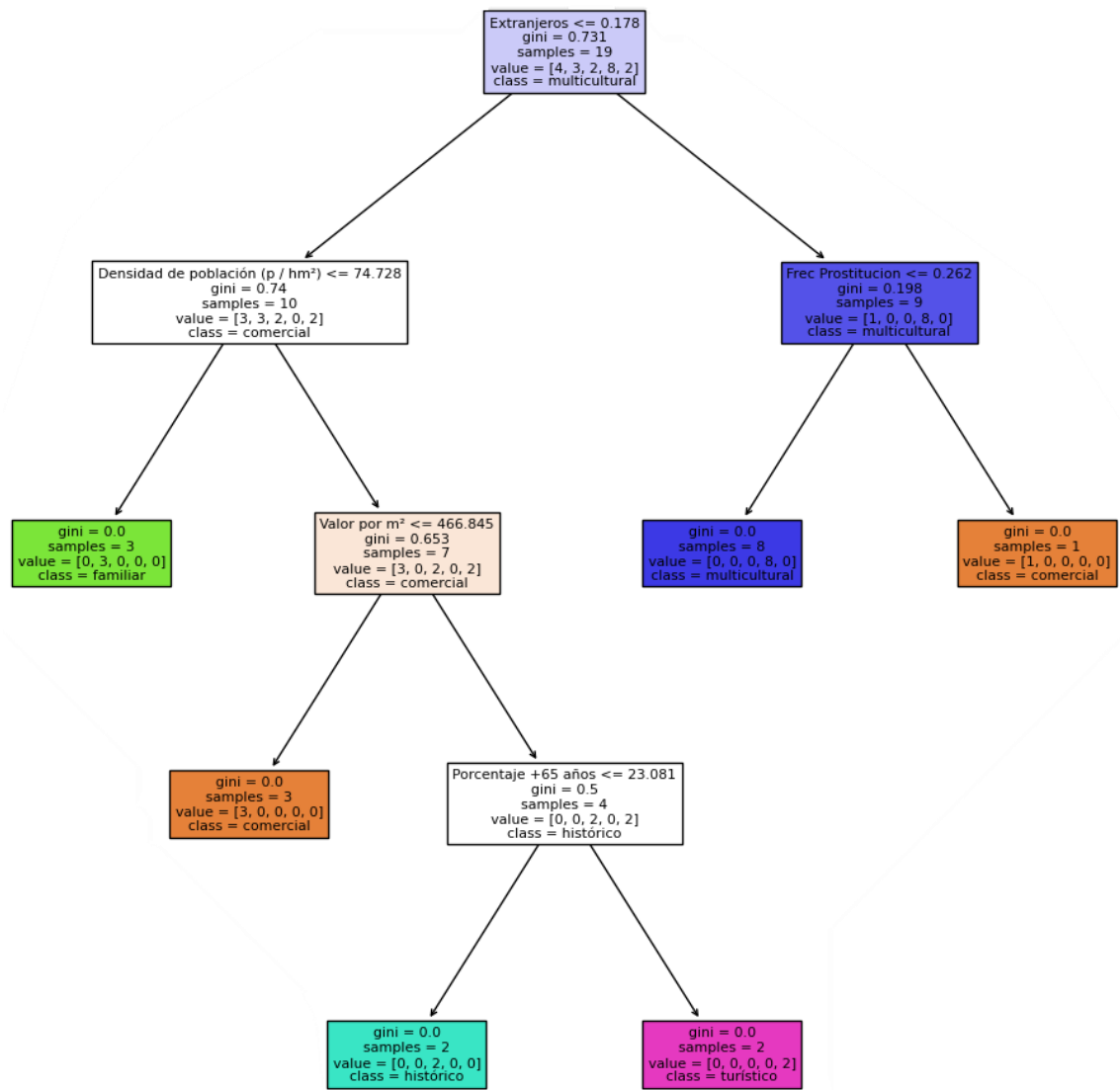


Previa clasificación con K-Means

- Agrupamos con 5 características

	Distrito	Extranjeros	Densidad de población (p / hm²)	Valor por m²	Porcentaje +65 años	Frec Prostitucion	Cluster_name
	ciutat vella	0.140351	170.210374	504.081523	20.165658	0.289474	histórico
	l'eixample	0.055556	251.357324	616.036472	23.970437	0.119048	turístico
	extramurs	0.152672	250.216819	491.626363	22.191162	0.083969	histórico
	campanar	0.177966	76.074215	442.062998	23.489702	0.093220	comercial
	la saidia	0.242188	244.347732	372.342184	23.198932	0.039062	multicultural
	el pla del real	0.000000	181.221057	638.134024	24.461370	0.247863	turístico
	l'olivereta	0.230769	249.316171	339.068317	21.063360	0.084615	multicultural
	patraix	0.308824	203.064080	412.463723	21.432930	0.397059	comercial
	jesus	0.218045	176.187174	362.409969	20.824373	0.060150	multicultural
	quatre carreres	0.190476	67.611932	385.316458	19.966008	0.095238	multicultural
	poblat marítims	0.186567	140.919175	335.106466	20.475484	0.126866	multicultural
	camins al grau	0.178571	279.844205	455.111580	19.226224	0.057143	multicultural
	algiros	0.059322	121.043258	434.574451	25.658769	0.127119	comercial
	benimaclet	0.140351	180.329748	421.983774	23.941020	0.008772	comercial
	rascanya	0.280303	207.725462	399.984546	17.049181	0.090909	multicultural
	benicalap	0.234375	219.806305	365.884949	18.918313	0.023438	multicultural
	pobles del nord	0.053763	5.219512	246.799319	22.920545	0.182796	familiar
	pobles de l'oest	0.153846	73.381631	344.474109	17.937072	0.028846	familiar
	pobles del sud	0.141593	3.251460	324.953979	20.847012	0.035398	familiar

MODELO DE REGLAS



¿Lo hemos hecho mal?

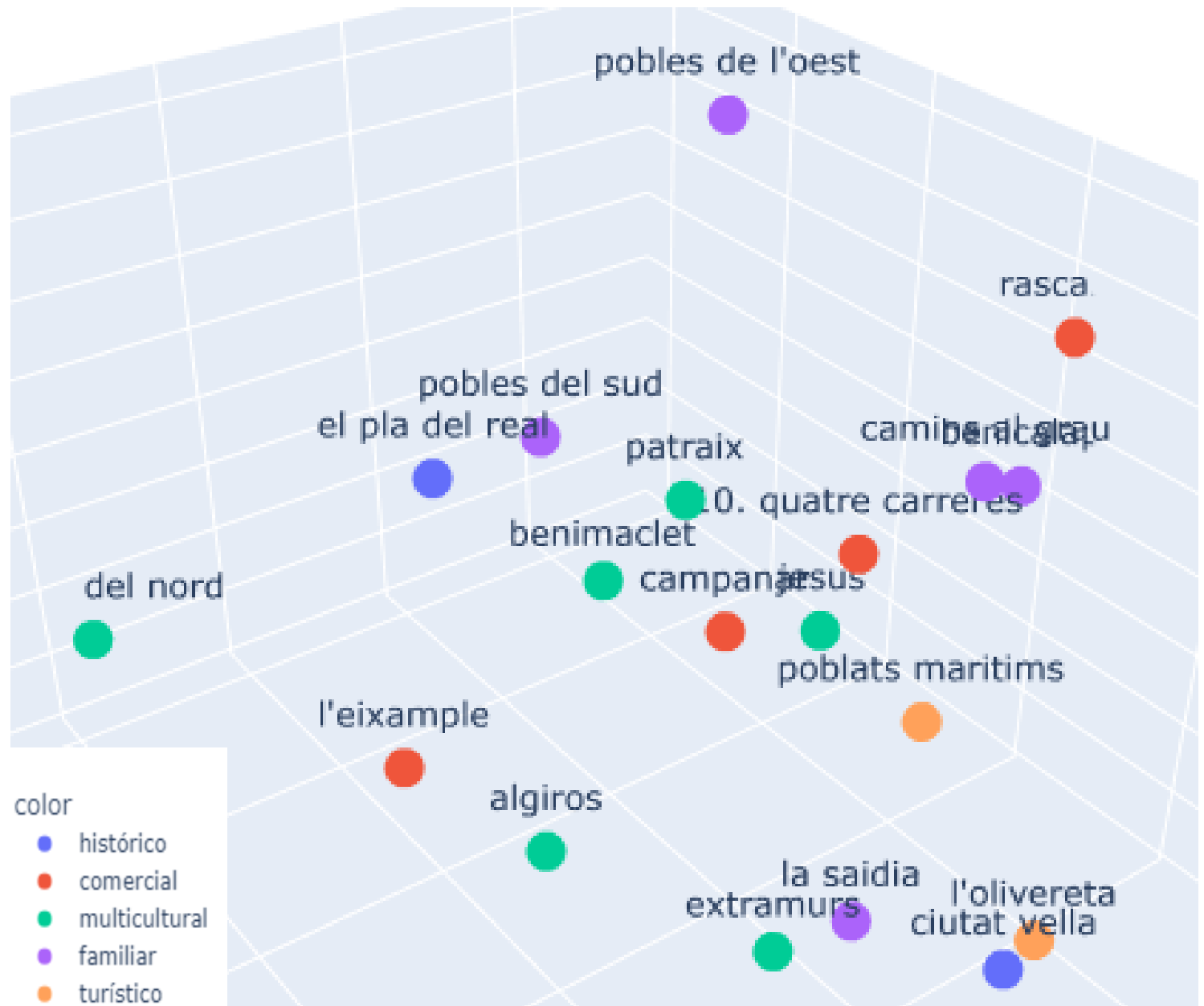
Distrito	Experto	Cluster/Reglas
Ciutat Vella	histórico	histórico
L'Eixample	comercial	turístico
Extramurs	multicultural	histórico
Campanar	familiar	comercial
La Saldia	familiar	multicultural
El Pla del Real	histórico	turístico
L'Olivereta	multicultural	multicultural
Patraix	multicultural	comercial
Jesus	multicultural	multicultural
Quatre Carreres	turístico	multicultural
Poblats Maritims	turístico	multicultural
Camins al Grau	familiar	multicultural
Algirós	multicultural	comercial
Benimaclet	multicultural	comercial
Rascanya	familiar	multicultural
Benicalap	familiar	multicultural
Poblats de L'oest	familiar	familiar
Poblats del Nord	histórico	familiar
Poblats del Sud	familiar	familiar

CLUSTERING

Coloreamos los distritos según las etiquetas del **experto** en el espacio de dimensionalidad reducida.

¡Los distritos del mismo tipo no están juntos!

No existe una estructura latente

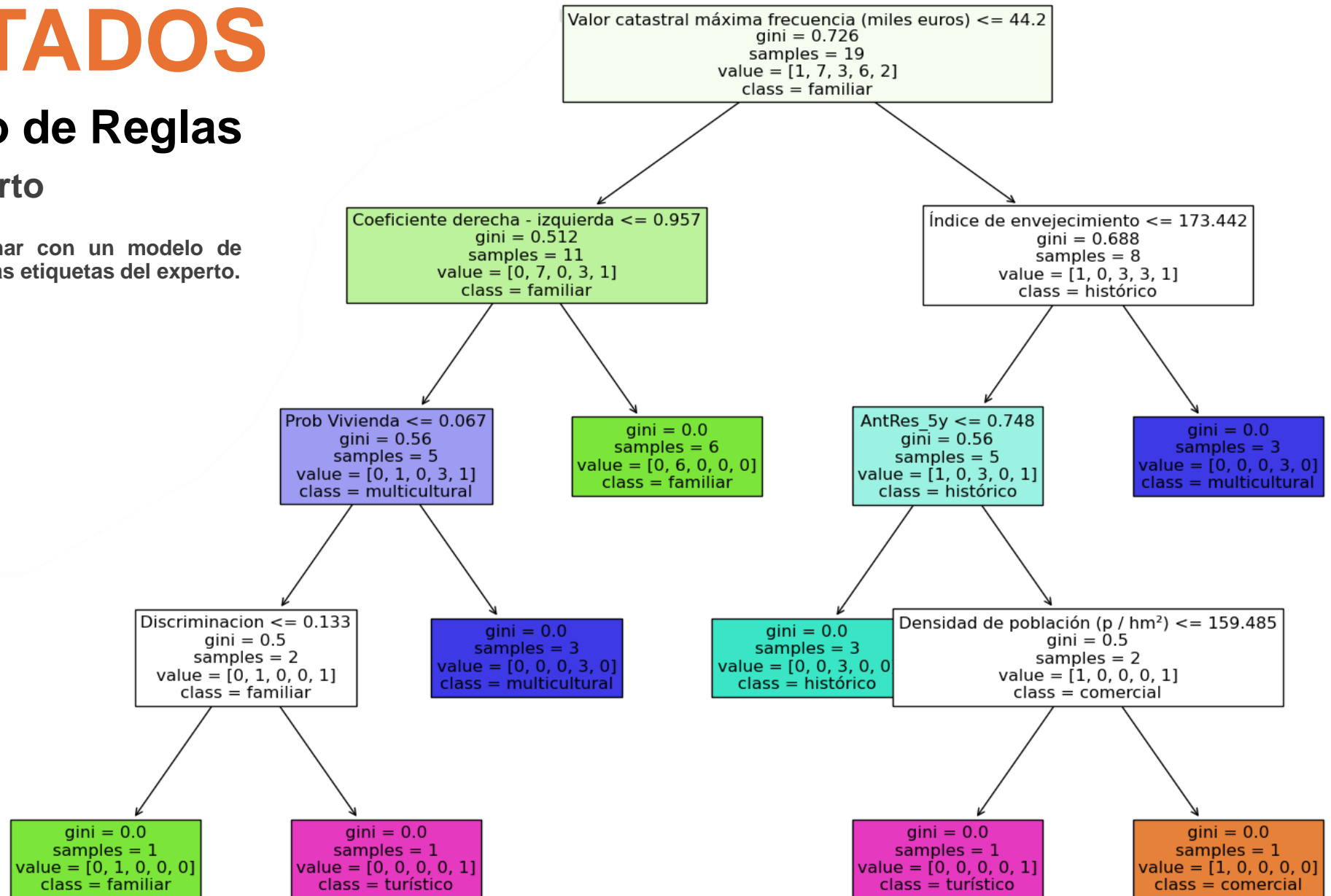


RESULTADOS

Nuevo Modelo de Reglas

Etiquetas del experto

Nos tendremos que conformar con un modelo de reglas construido a partir de las etiquetas del experto.



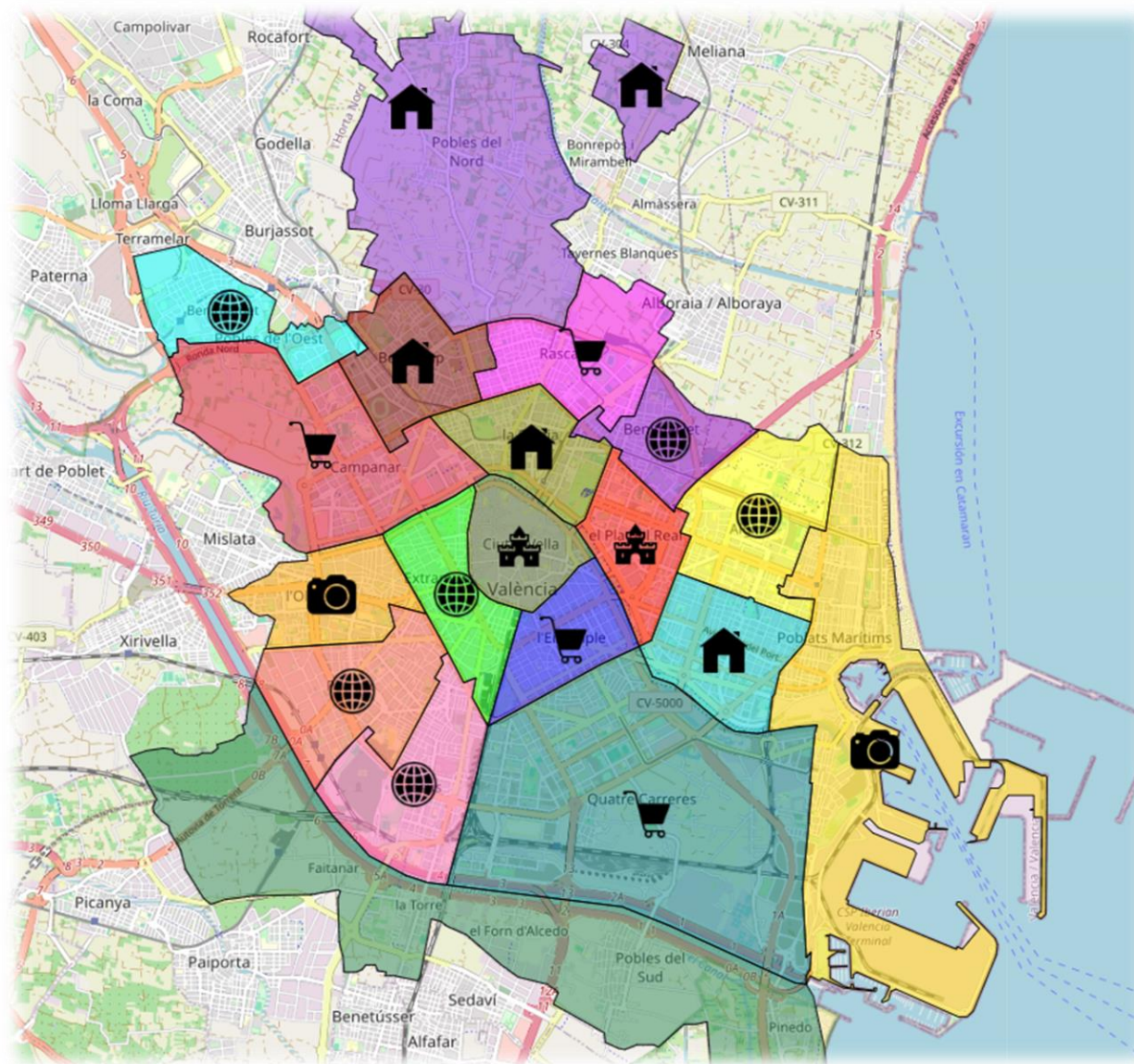
RESULTADOS



<https://qrco.de/mapavlc>







Folium



CONCLUSIONES

Veamos si hemos cumplido los **objetivos**:

- **Identificar y recopilar datos** utilizando técnicas automáticas 
- **Explorar** datos reales y hacer un **análisis** preliminar 
- Emplear técnicas de **procesamiento de lenguaje natural** y **Web Scraping** 
- Usar herramientas de control de versiones: **GitHub** 

CONCLUSIONES

Clasificar los distritos según su tipología y características.

- 7 modelos de Lenguaje ✓
- 2 modelos de Reglas ✓
- La clasificación de un **experto**

REUNIÓN DE EXPERTOS EN
DISTRITOS



CONCLUSIONES

Algunas cuestiones más:

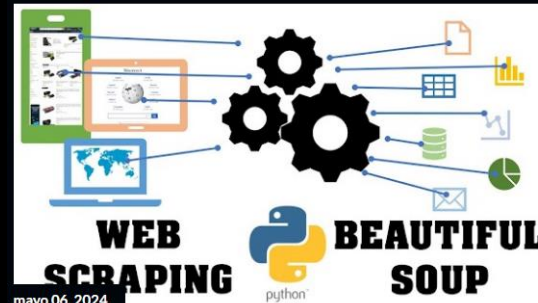
- Pocos datos realmente útiles
- Tarea mayoritariamente subjetiva
- Problema de clasificación y denominación:
no hay datos etiquetados
 - Elegir etiquetas
 - Repartir etiquetas
 - No hay forma de validar

Y POR ÚLTIMO...

¡No olvides
echarle un ojo a
nuestro blog!

En él encontrarás interesantes
artículos acerca del desarrollo
del proyecto, así como de los
resultados obtenidos

consultatubarrio.blogspot.com



mayo 06, 2024

DATOS NO ESTRUCTURADOS: USO DE
WEBSCRAPING

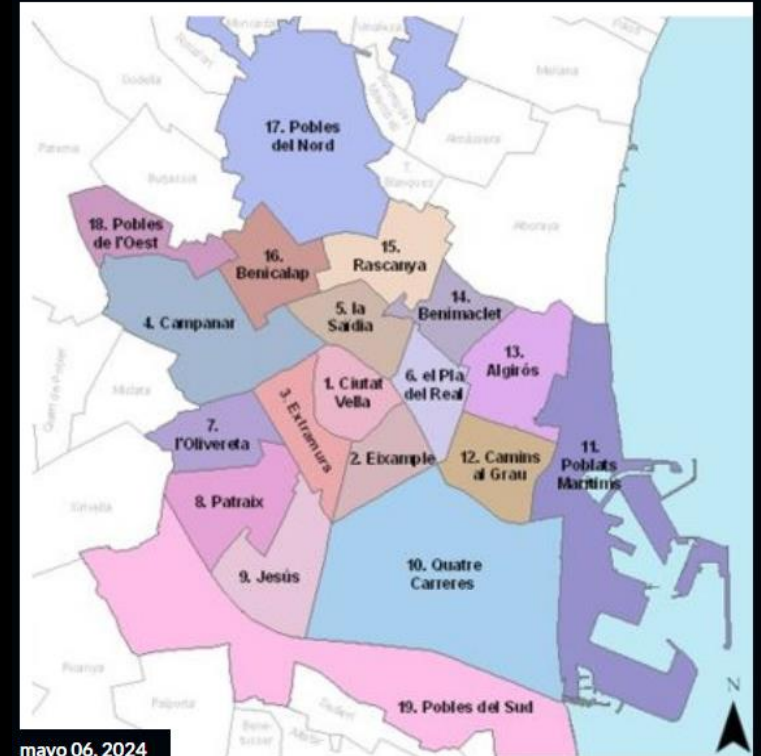
Compartir 5 comentarios



mayo 06, 2024

¡LOS DATOS SON IMPORTANTES PARA
UN ESTUDIO!

Compartir 5 comentarios



mayo 06, 2024

DATOS ESTRUCTURADOS:
AYUNTAMIENTO DE VALENCIA

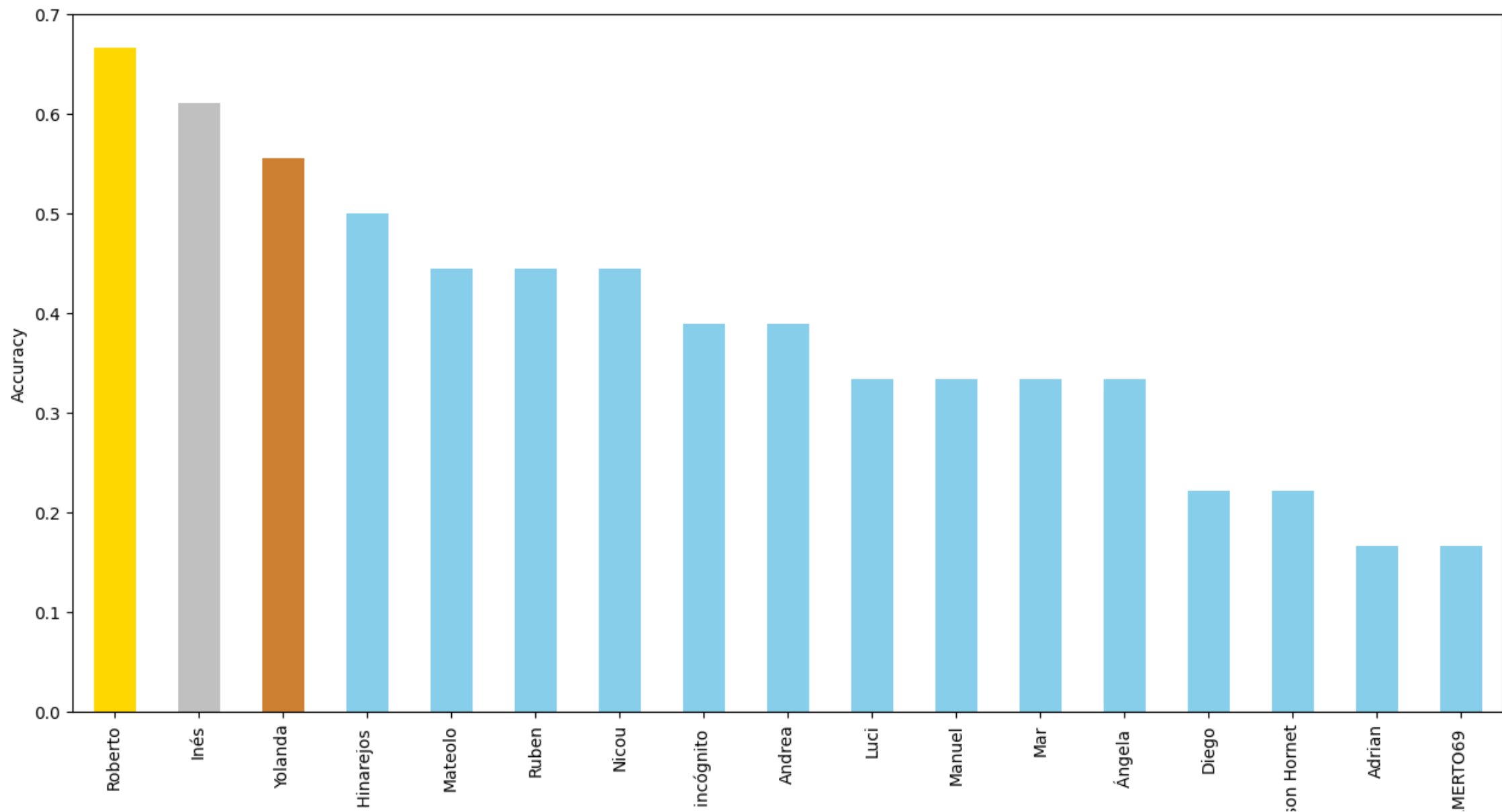
Compartir 5 comentarios

¡ÚNETE A NUESTRA COMUNIDAD!

Usuarios ▼ por ID del país ▼



PAÍS	USUARIOS	
Spain	37	↓ 62,2 %
United States	9	-
Argentina	1	-
Brazil	1	-
Canada	1	-
Türkiye	1	-
Colombia	0	↓ 100,...



Autores:

Isaac Martí

Jesús Martínez

Samuel Ortega

Carlos Sánchez

Pablo Vicente

Con la colaboración de:



CLASIFICACIÓN DE DISTRITOS DE VALENCIA