

Examen Desarrollo

Nombre: Washington Adrian Lumbi Sánchez

1,"Descargue los seguidores de este usuario de Instagram @juanperez a través de las técnicas estudiadas en clase, los datos a recolectar deben ser: seguidores, seguidos, biografía, name, username"

Se uso playwright para scrapear al usuario @luisandreluna01 se scrapeo exactamente 153 seguidores en los cuales los parametros fueron:

'username': username, username

'full_name': full_name, name

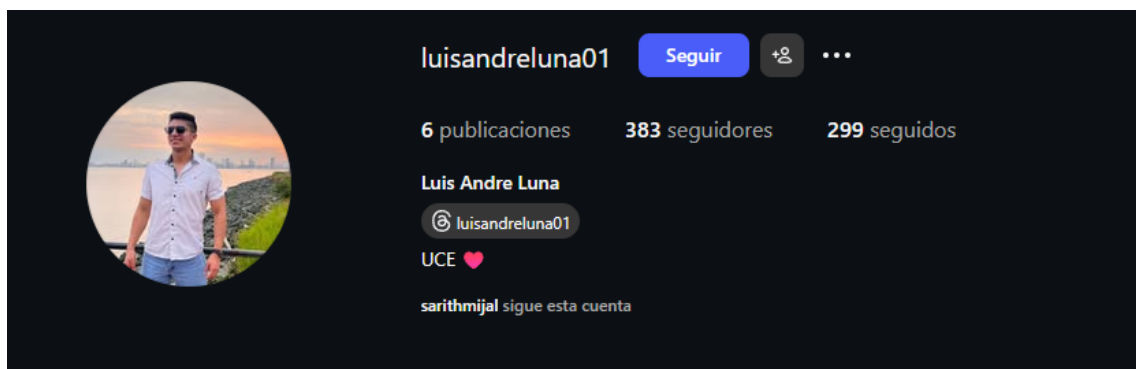
'biography': biography, biografía

'follower_count': followers, seguidores

'following_count': following, seguidos

'is_private': is_private, (no se pidió en la rubrica pero lo dejo porque ya tenia hecho)

'is_verified': is_verified, (no se pidió en la rubrica pero lo dejo porque ya tenia hecho)



Instagram

instagram.com/danny_fern/

danny_fern Seguir + ...

87 publicaciones 326 seguidores 572 seguidos

Danny Chicaiza
@danny_fern
Ingeniero en diseño industrial

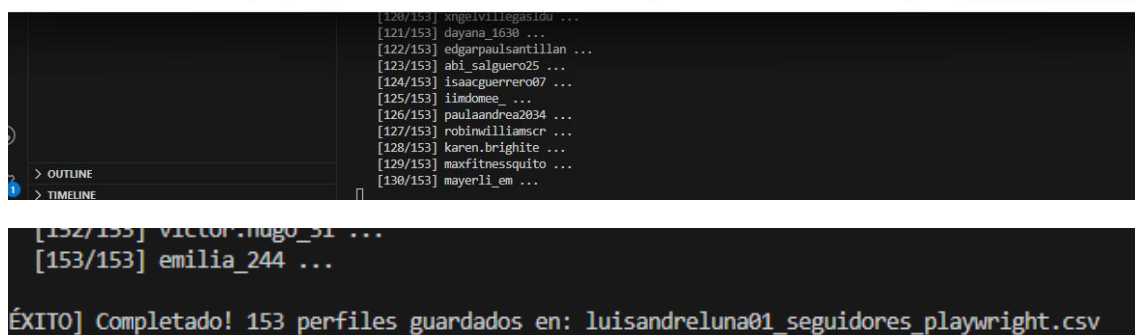
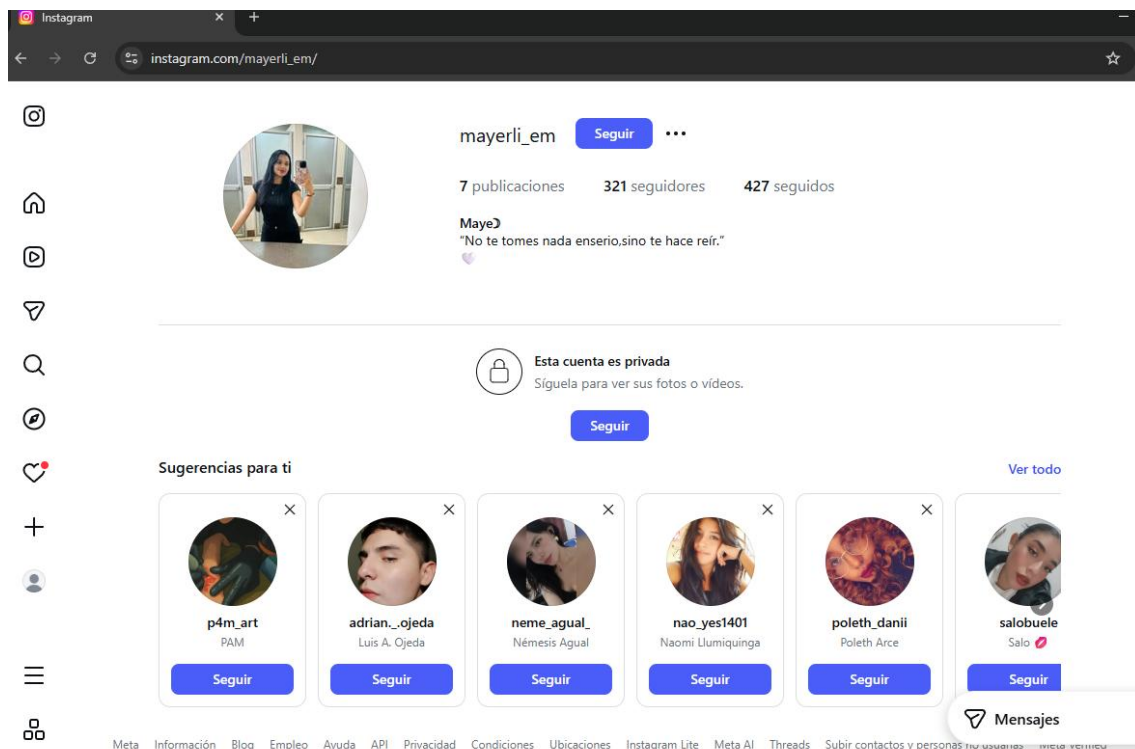
Destacada

Mensajes

- [86/153] misodayan ...
- [86/153] valeponce_g ...
- [87/153] alessandrorcm10 ...
- [88/153] tridydan ...
- [89/153] daniel.ale02 ...
- [90/153] victor_hugo263 ...
- [91/153] rodneyan3 ...
- [92/153] davila_sheyia ...
- [93/153] ders.e ...
- [94/153] mica.amifjos_12 ...
- [95/153] danny_fern ...

> OUTLINE

i > TIMELINE



Se uso cookies para el scrapeo

Application		Filter	
		Name	Value
Manifest		csrftoken	jlutSv0fxkuv
Service workers		datr	7bOPaYLi3C
Storage		ds_user_id	472810717
Storage		ig_did	8606DA5D-
Local storage		mid	aY-z7QALA
Session storage		rur	"LDC\05447
Extension storage		sessionid	472810717
IndexedDB		wd	252x948
Cookies			
https://www.instagram....			
https://www.facebook....			

Dichas cookies se almacenan en un .env


```
# Google Gemini API Key
# Obtén tu clave gratuita en: htt
GOOGLE_API_KEY=AIzaSyCL7ACa_Rzly9

OPENAI_API_KEY=sk-or-v1-4e89d7dcb

# Instagram Cookies (JSON en una
INSTAGRAM_COOKIES=[{"domain": ".in
```

Si usted lo desea le envío el .env o le dejo el .env.example para que pueda usar el código

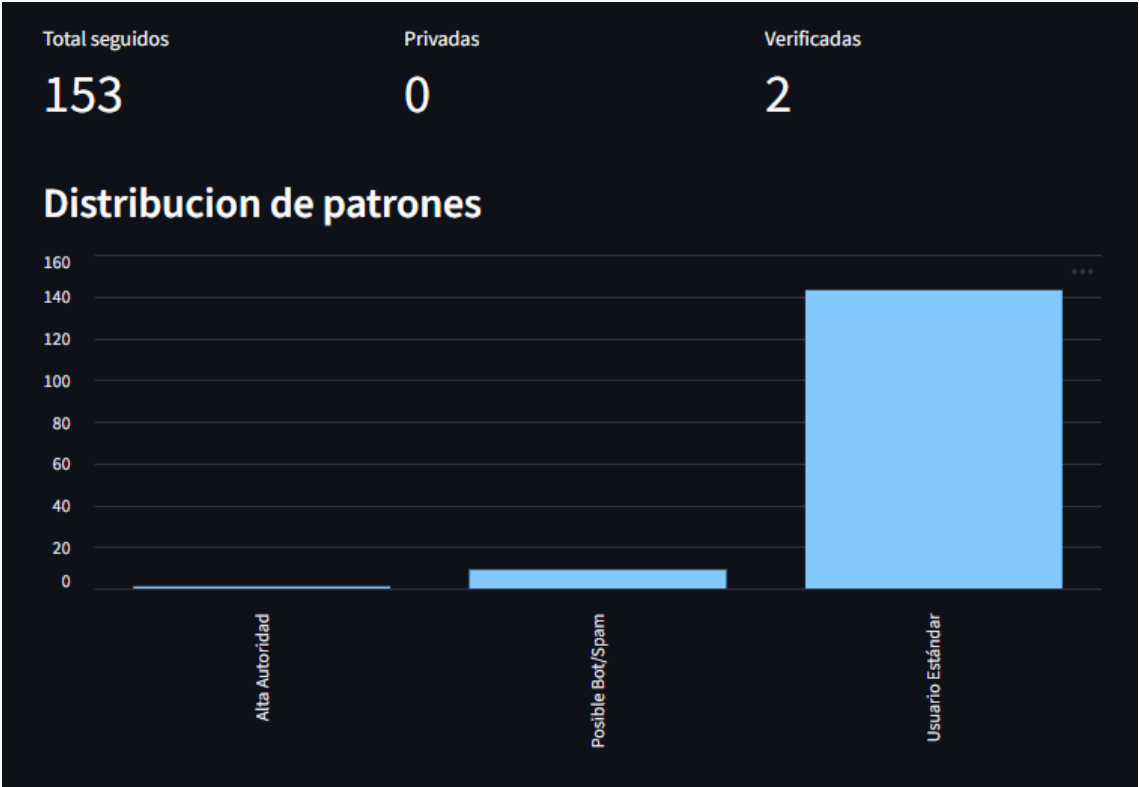
2,"Con los datos que ha descargado, suba el archivo al software solicitado en el deber que recibe archivos y da un resumen, ahora su funcionalidad es que reciba las biografías descargadas (por lo menos 30) y que su resultado le proporcione un listado de palabras y conceptos de lo que más hablan."

Para ello se usó el repositorio Chatpdf que nos proporciona el uno de expositores de las clases anteriores <https://github.com/BluerZu/ChatPDF>

Con la ayuda de Streamlit y ChromaDB se pudo realizar el análisis, algo que agregar que para la preparación de dicho examen ya había realizado con la ayuda de la IA ciertas funciones para analizar los chats de Instagram



Análisis estadístico



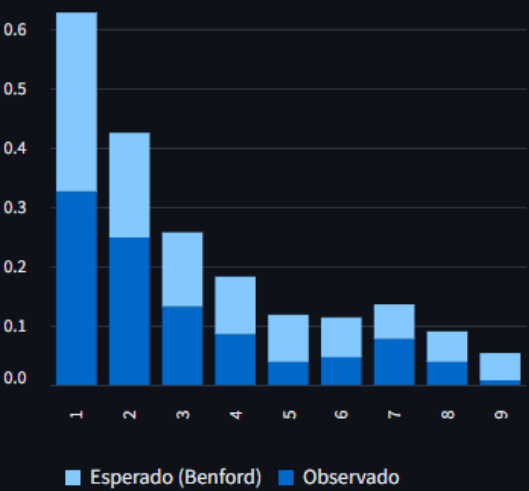
Ley de benford

Ley de Benford

Compara la distribucion del primer digito de los datos vs la distribucion teorica de Benford. Desviaciones grandes pueden indicar datos anomalos o manipulados.

Columna a analizar:

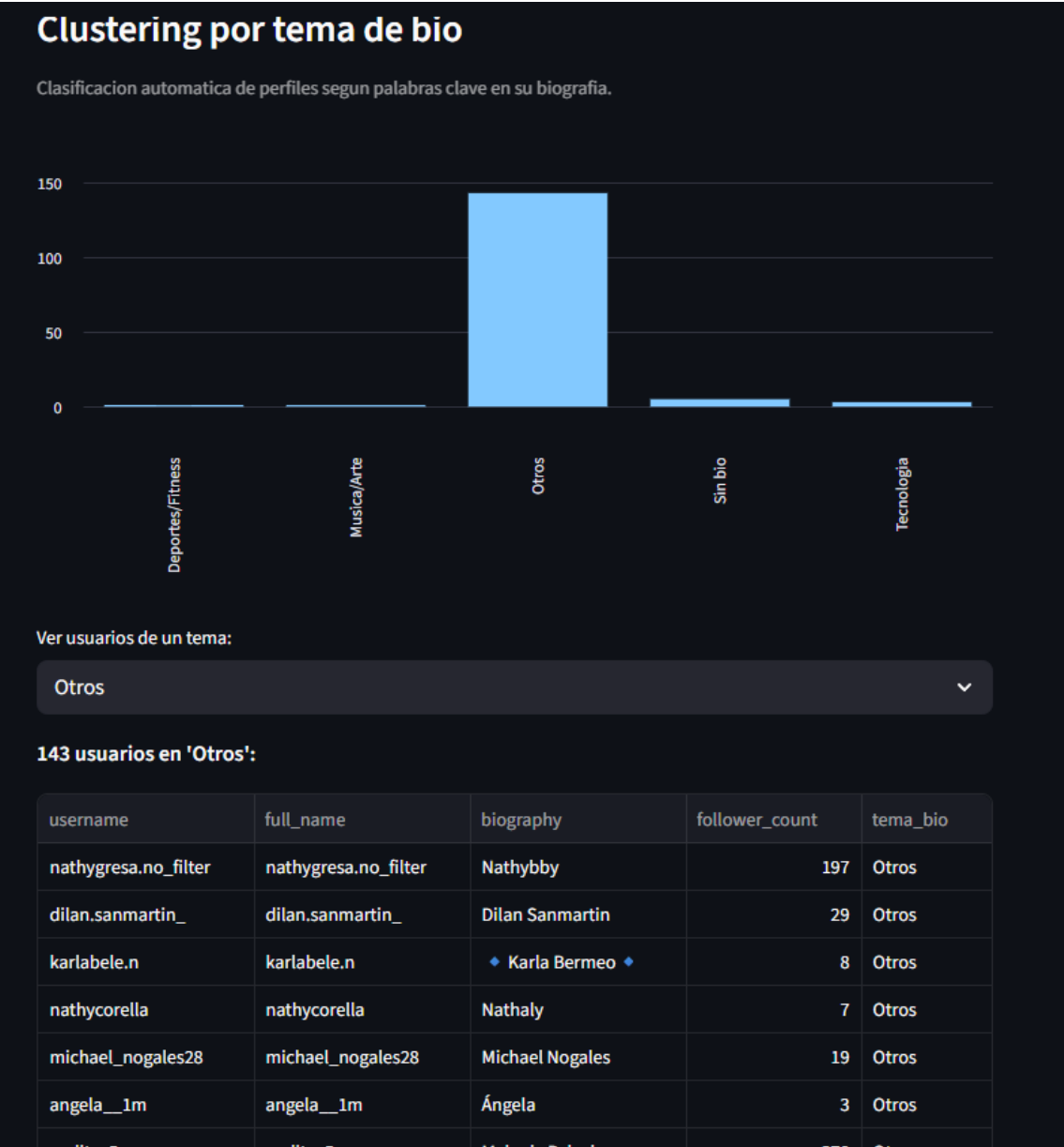
follower_count



Digito	Observado	Esperado (Benford)	Diferencia
1	0.3256	0.301	0.0246
2	0.2481	0.1761	0.072
3	0.1318	0.1249	0.0068
4	0.0853	0.0969	0.0116
5	0.0388	0.0792	0.0404
6	0.0465	0.0669	0.0204
7	0.0775	0.058	0.0195
8	0.0388	0.0512	0.0124
9	0.0078	0.0458	0.038

Desviacion moderada detectada (max: 7.2%)

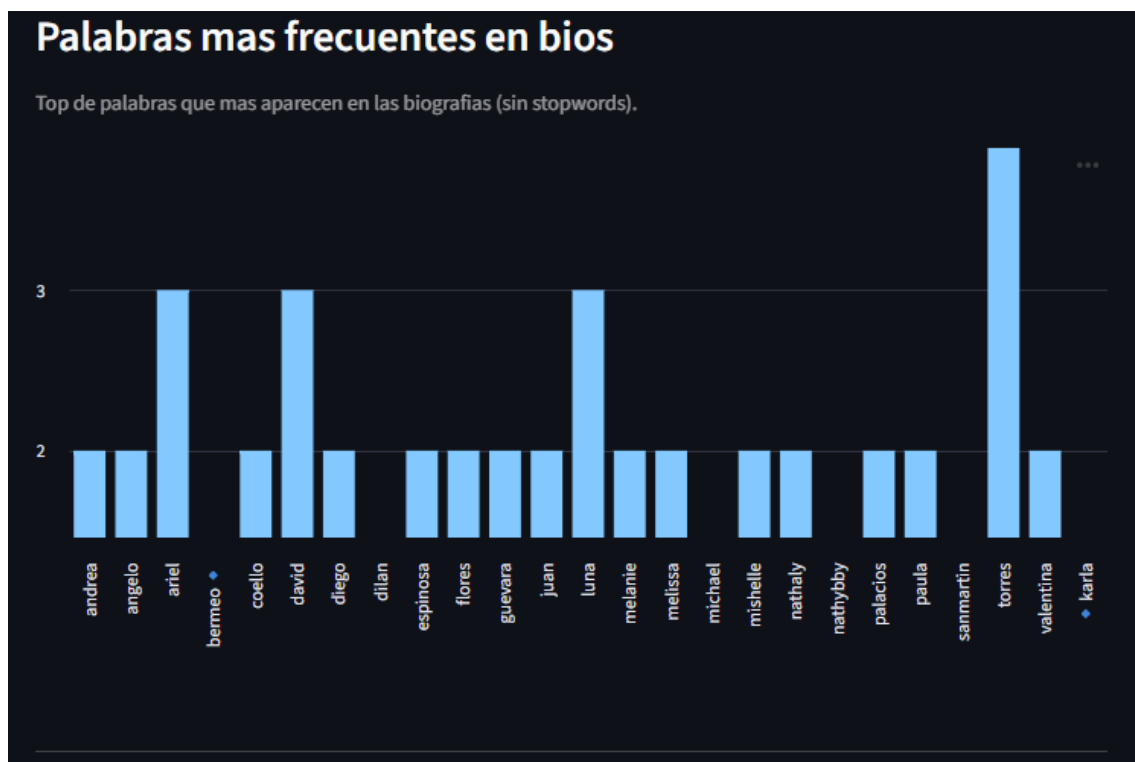
Clusterizacion



En este caso la IA me ayudo para el desarrollo de esta función y me propociono estos temas


```
THEME_KEYWORDS = {
    "Deportes/Fitness": ["futbol", "football", "soccer", "basket", "gym", "fitness", "crossfit",
                        "runner", "running", "athlete", "atleta", "entrenador", "coach", "deporte",
                        "sport", "tennis", "boxing", "boxeo", "yoga", "surf", "ciclismo", "cycling"],
    "Musica/Arte": ["musica", "music", "cantante", "singer", "rapper", "dj", "productor", "producer",
                   "artista", "artist", "arte", "pintor", "painter", "actor", "actriz", "actress",
                   "teatro", "band", "banda", "guitarra", "piano", "drummer"],
    "Belleza/Moda": ["makeup", "maquillaje", "beauty", "belleza", "fashion", "moda", "modelo", "model",
                    "skincare", "nails", "uñas", "hair", "cabello", "stylist", "estilista", "outfit"],
    "Tecnologia": ["developer", "desarrollador", "programador", "programmer", "software", "tech",
                  "codigo", "code", "data", "ai", "web", "devops", "hacker", "ciberseguridad",
                  "cybersecurity", "startup", "fintech", "blockchain", "crypto"],
    "Negocios/Emprendimiento": ["ceo", "founder", "fundador", "entrepreneur", "emprendedor", "empresa",
                                "business", "marketing", "ventas", "sales", "manager", "gerente",
                                "director", "consultor", "consultant", "coach empresarial", "negocio"],
    "Politica/Activismo": ["politica", "politics", "diputado", "senador", "congresista", "presidente",
                           "activista", "activist", "derechos", "rights", "justicia", "democracia",
                           "partido", "gobierno", "government", "feminista", "feminist"],
    "Educacion": ["profesor", "teacher", "educacion", "education", "universidad", "university",
                  "estudiante", "student", "docente", "academia", "investigador", "researcher",
                  "phd", "maestro", "escuela", "school"],
    "Fotografia/Video": ["fotografo", "photographer", "fotografia", "photography", "videographer",
                         "filmmaker", "cineasta", "video", "camera", "foto", "retrato", "portrait"],
    "Gastronomia": ["chef", "cocina", "cooking", "food", "comida", "restaurante", "restaurant",
                    "recetas", "recipes", "foodie", "gastronomia", "panadero", "baker", "barista"],
    "Viajes": ["travel", "viaje", "viajero", "traveler", "wanderlust", "mochilero", "backpacker",
              "turismo", "tourism", "aventura", "adventure", "explore", "nomad"],
    "Salud/Bienestar": ["doctor", "medico", "salud", "health", "psicologo", "psychologist",
                       "terapeuta", "therapist", "nutricion", "nutrition", "bienestar", "wellness",
                       "mental health", "enfermero", "nurse", "dentista"],
}
```

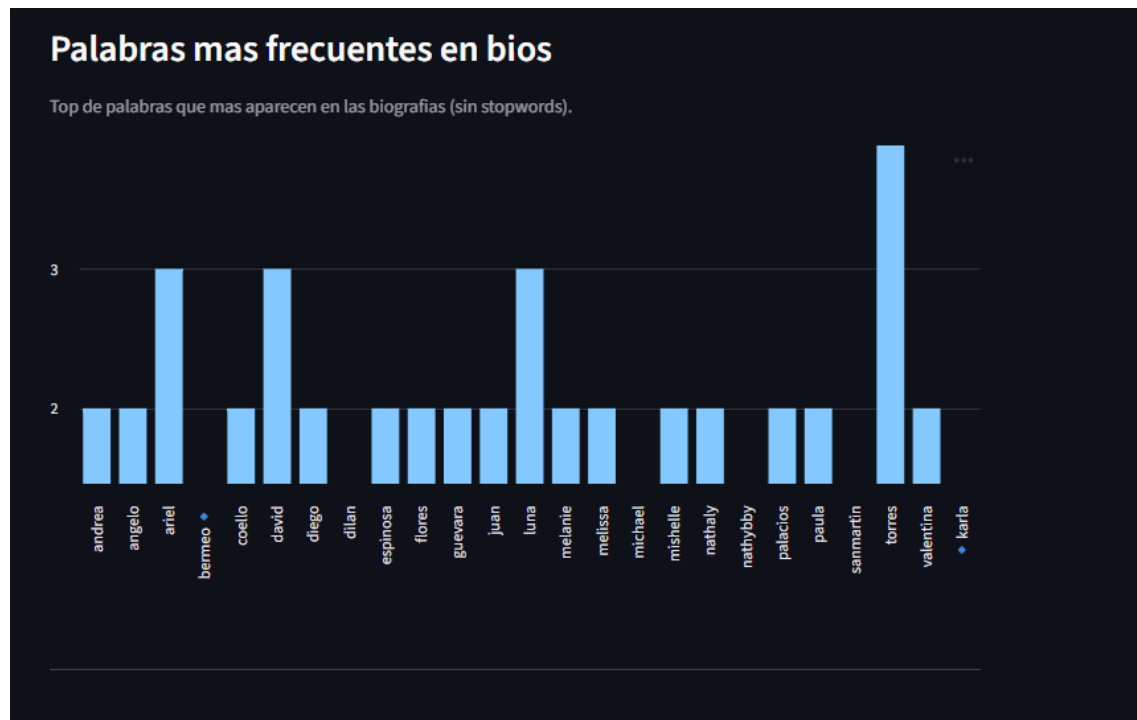
Palabras mas frecuentes



Me apoye de la ia para el desarrollo de esta función por eso el stopwords en el prompt le pedí que no me ponga conectores y palabras irrelevantes para no sesgar el diagrama de barras

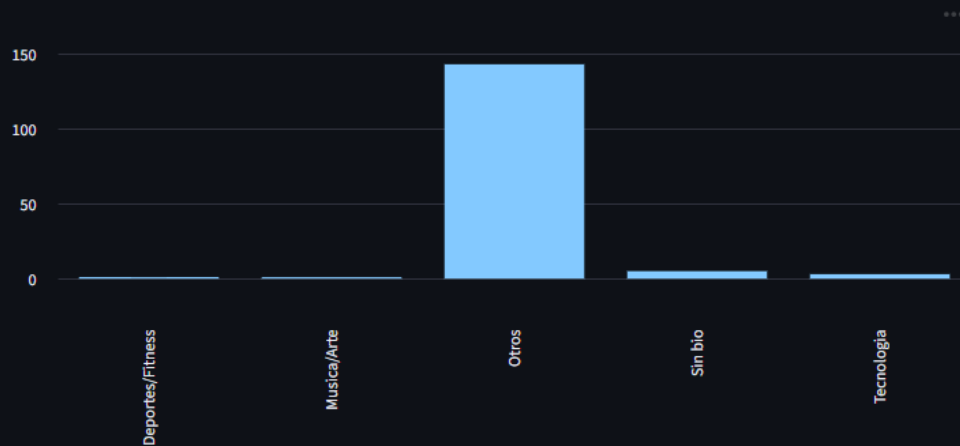

```
stopwords = {
    "de", "la", "el", "en", "y", "a", "los", "las", "del", "un", "una", "que",
    "es", "por", "con", "para", "al", "se", "lo", "no", "mi", "me", "te", "tu",
    "su", "nos", "le", "si", "ya", "o", "e", "i", "the", "and", "of", "to", "in",
    "is", "for", "on", "my", "it", "at", "we", "be", "do", "an", "or", "so", "if",
    "he", "she", "as", "up", "all", "but", "not", "you", "are", "was", "has", "had",
    "nan", "sin", "bio", "com", "www", "https", "http"
}
all_words = []
```

3,"Mostrar el listado de conceptos y palabras, en cualquier gráfico, nube de palabras, diagramas de barra, etc"



Clustering por tema de bio

Clasificación automática de perfiles según palabras clave en su biografía.



Ver usuarios de un tema:

Otros

143 usuarios en 'Otros':

username	full_name	biography	follower_count	tema_bio
nathygresano_filter	nathygresano_filter	Nathybby	197	Otros
dilan.sanmartin_	dilan.sanmartin_	Dilan Sanmartin	29	Otros
karlahelo_n	karlahelo_n	Karla Bermes	8	Otros

4,"Subir el deber a GIT que sea público, enviar al correo:

dfnoguera@uce.edu.ec con el asunto: Examen Desarrollo 2025-2026."

https://github.com/Wadri02/Examen_Desarrollo

Enviar

Para

DIEGO FRANS NOGUERA VASCONEZ

CC

Examen Desarrollo 2025-2026.

Borrador guardado a las 10:00

https://github.com/Wadri02/Examen_Desarrollo

envió el link de github