

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 scraeo 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)



Proceso de Scrapeo:

The screenshot shows an Instagram profile page for the user 'danny_fern'. The profile picture is a circular photo of a man with glasses and a blue shirt. The bio reads 'Danny Chicaiza' and 'Ingeniero en diseño industrial'. The stats show 87 publicaciones, 326 seguidores, and 572 seguidos. A 'Destacada' post is visible. The main feed shows three photos: one of a man looking at his phone, one of a forest, and one of a group of people taking a selfie. A message bubble icon with 'Mensajes' is overlaid on the right side of the feed. On the left, there's a sidebar with navigation icons and a list of 153 messages, starting with 'alissdayan ...' and ending with 'danny_fern ...'. At the bottom left, there are buttons for 'OUTLINE' and 'TIMELINE'.

- [85/153] alissdayan ...
- [86/153] valeponce g ...
- [87/153] alessandrorm10 ...
- [88/153] triidylan ...
- [89/153] daniel.ale02 ...
- [90/153] victor.hugo063 ...
- [91/153] rodneyand3 ...
- [92/153] davila_sheyla ...
- [93/153] ders.e ...
- [94/153] mica.arnijos_12 ...
- [95/153] danny_fern ...

mayerli_em Seguir ...

7 publicaciones 321 seguidores 427 seguidos

Maye
"No te tomes nada enserio,sino te hace reír."

Esta cuenta es privada
Síguela para ver sus fotos o videos.

Sugerencias para ti

p4m_art PAM	adrian...ojeda Luis A. Ojeda	neme_agual_ Nemesia Aguilar	nao_yes1401 Naomi Llumiquinga	poleth_daniil Poleth Arce	salobule Salo
Seguir	Seguir	Seguir	Seguir	Seguir	Seguir

Ver todo

Meta Información Blog Empleo Ayuda API Privacidad Condiciones Ubicaciones Instagram Lite Meta AI Threads Subir contactos y personas

[120/153] xngelvillegas1du ...
[121/153] dayana_1630 ...
[122/153] edgarpaulsantillan ...
[123/153] abi_salguero25 ...
[124/153] isaacguerrero07 ...
[125/153] imidomee ...
[126/153] paulandrea2034 ...
[127/153] robinwilliamscr ...
[128/153] karen.brighte ...
[129/153] maxfitnessquito ...
[130/153] mayerli_em ...

> OUTLINE > TIMELINE

[152/153] victor.hugo_51 ...
[153/153] emilia_244 ...

ÉXITO! Completado! 153 perfiles guardados en: luisandreluna01_seguidores_playwright.csv

Se uso cookies para el scraping

Application

- Manifest
- Service workers
- Storage
- Storage
 - Local storage
 - Session storage
 - Extension storage
- IndexedDB
- Cookies
 - https://www.instagram....
 - https://www.facebook....

Filter

Name	Value
csrf-token	jlutSv0fxkuv
datr	7bOPaYLi3C
ds_user_id	4728107171
ig_did	8606DA5D-
mid	aY-z7QALA
rur	"LDC\05447
sessionid	4728107171
wd	252x948

Dichas cookies se almacenan en un .env

```
# Google Gemini API Key
# Obtén tu clave gratuita en: http://www.googleapis.com/auth/gemini
GOOGLE_API_KEY=AIzaSyCL7ACa_Rzly9

OPENAI_API_KEY=sk-or-v1-4e89d7dcbae64f8

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

Si usted lo desea le envio 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 uso el repositorio Chatpdf que nos proporciono 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

Chat Instagram - Análisis de Seguidores

Selecciona un archivo CSV de seguidos

Drag and drop file here
Limit 200MB per file • CSV

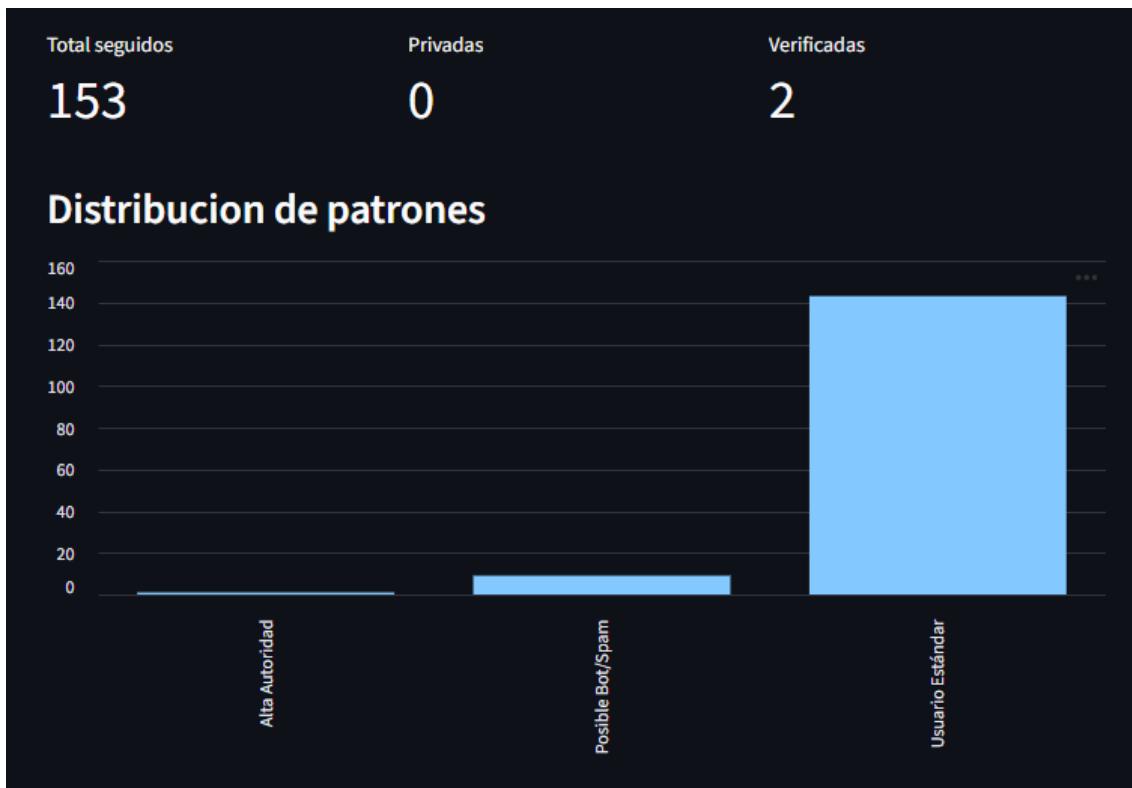
Browse files

luisandreluna01_seguidores_playwright.csv 11.9KB X

Cargar datos de Instagram

Listo! 153 perfiles cargados en 68 fragmentos.

Análisis estadístico



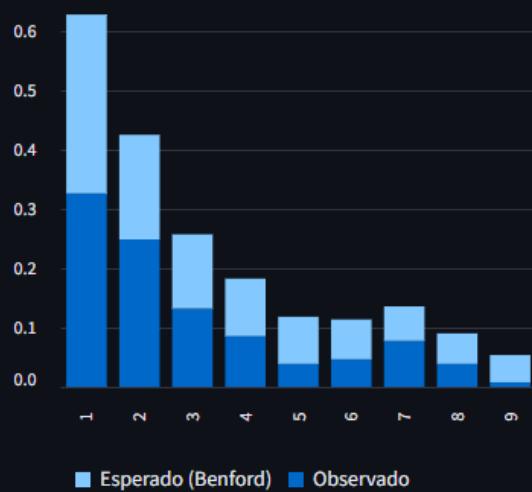
Ley de benford

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



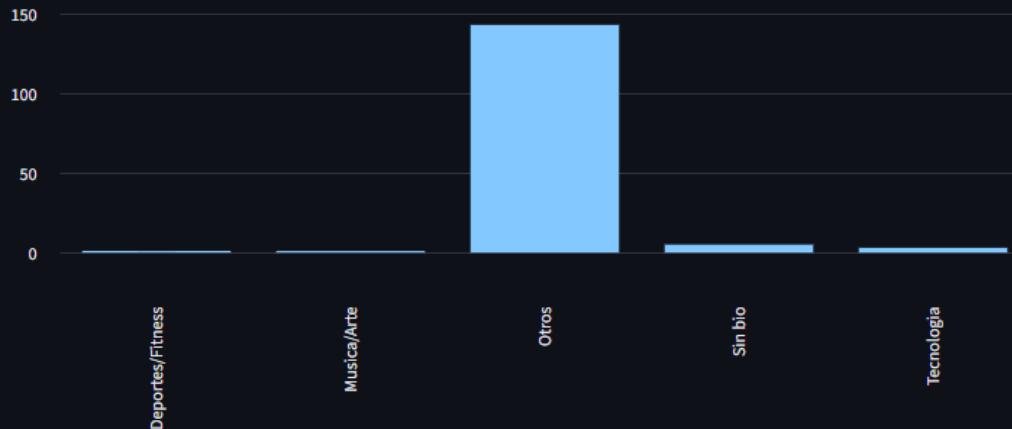
Dígito	Observado	Esperado (Benford)	Diferencia
1	0.3256	0.301	0.0246
2	0.2481	0.1761	0.072
3	0.1318	0.1249	0.0068
4	0.0853	0.0969	0.0116
5	0.0388	0.0792	0.0404
6	0.0465	0.0669	0.0204
7	0.0775	0.058	0.0195
8	0.0388	0.0512	0.0124
9	0.0078	0.0458	0.038

Desviación moderada detectada (max: 7.2%)

Clusterizacion

Clustering por tema de bio

Clasificacion automatica de perfiles segun palabras clave en su biografia.



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
karlabele.n	karlabele.n	♦ Karla Bermeo ♦	8	Otros
nathycorella	nathycorella	Nathaly	7	Otros
michael_nogales28	michael_nogales28	Michael Nogales	19	Otros
angela_1m	angela_1m	Ángela	3	Otros

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

```

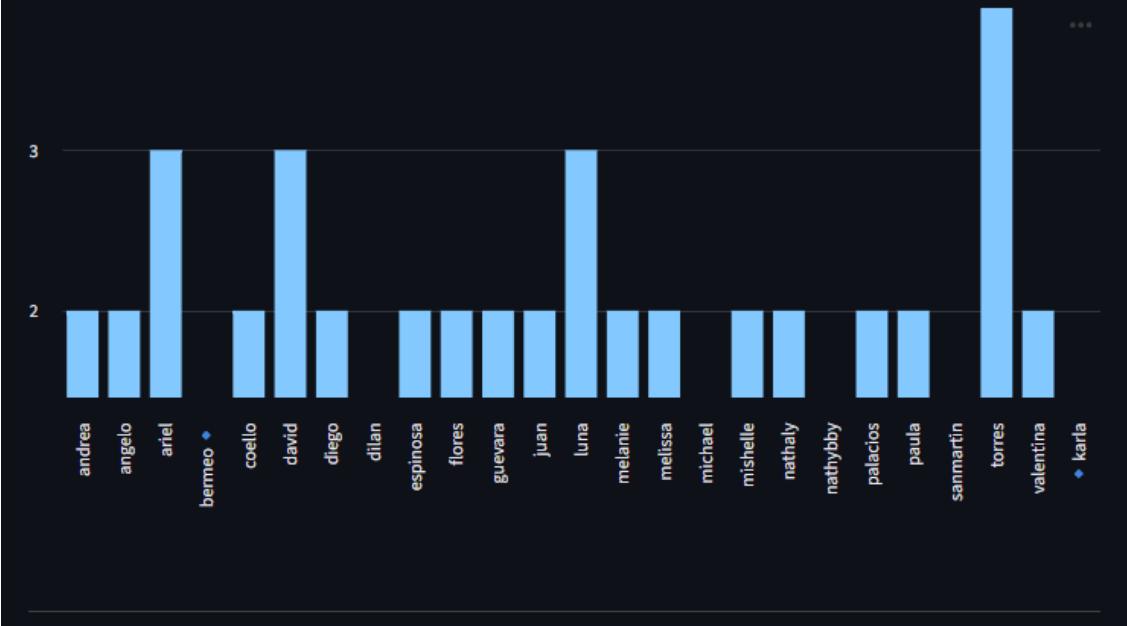
THEME_KEYWORDS = {
    "Deportes/Fitness": ["futbol", "football", "soccer", "basket", "gym", "fitness", "crossfit",
                         "runner", "running", "athlete", "atleta", "entrenador", "coach", "deporte",
                         "sport", "tennis", "boxing", "boxeo", "yoga", "surf", "ciclismo", "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

Palabras mas frecuentes en bios

Top de palabras que mas aparecen en las biografias (sin stopwords).



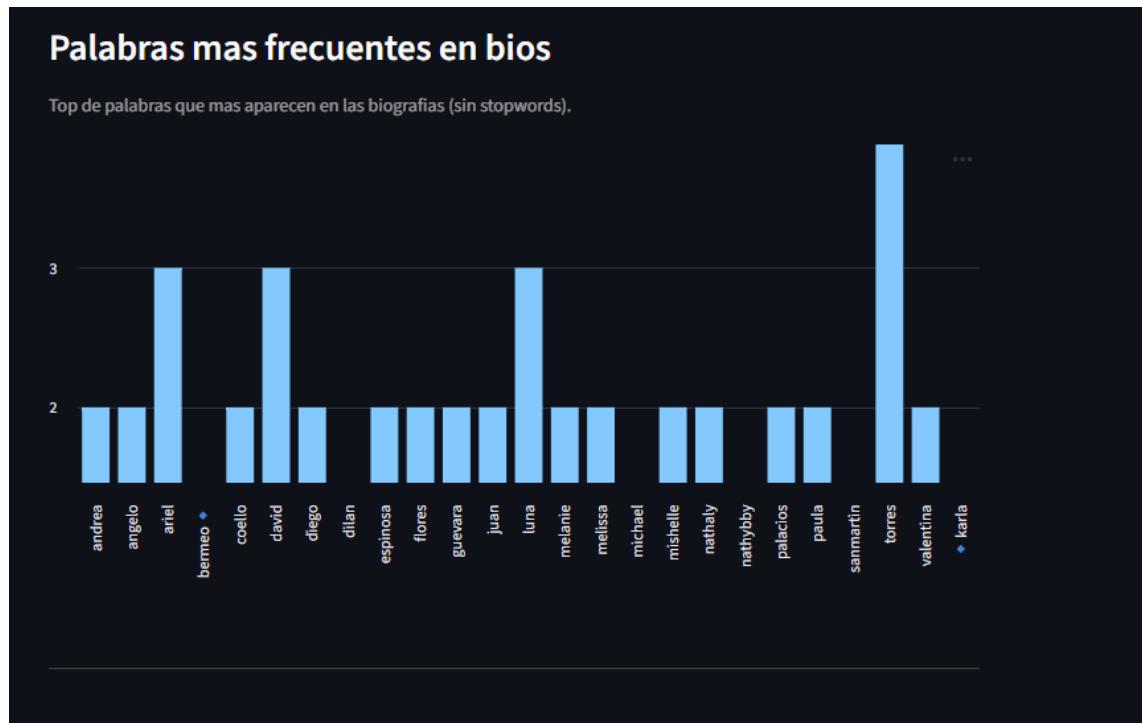
Me apoye de la ia para el desarrollo de esta función por eso el stopwords en el prompt le pedí que no me ponga conectores y palabras irrelevantes para no sesgar el diagrama de barras

```

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 = []

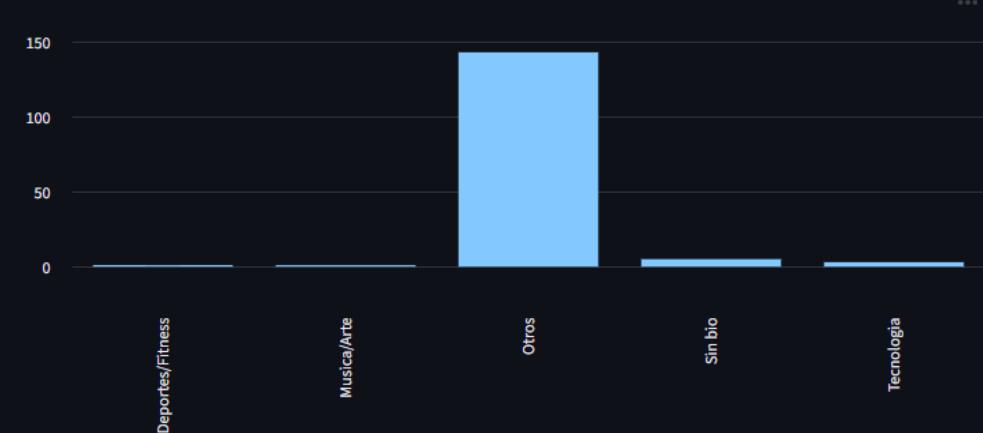
```

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
KarlaBermoo...	KarlaBermoo...	Karla Bermoo...	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 la nube

https://github.com/Wadri02/Examen_Desarrollo
envió el link de github