

Mediavida Extractor 1.0

Herramientas utilizadas:

- Python
 - Pandas
 - Matplotlib/Seaborn
 - BeautifulSoup
 - Nltk

- PowerBi




Introducción

Mediavida es uno de los foros más populares y antiguos de España con casi 25 años de antigüedad con miles de usuarios que participan diariamente.

Su temática principal es la de videojuegos pero abarca todo tipo de temas que se dividen en **subforos**













Breve explicación de la estructura del foro


 mediavida **Foros** ▾ Spy Top Grupos

Foros


















General

 OFF-Topic Mi novia me ha dejado por un reptiliano	 FEDA Foro de no-Billas
 Ofertas Shut up and take my money!	 Política Ni de derechas ni de izquierdas
 Streamers e Influencers Conectamos con Andorra	 Criptomonedas y finanzas Lambos, sí, pero a plazo fijo
 Compra-Venta Me lo quitan de las manos	 Estudios y trabajo La solución es 0,67
 Ciencia En este foro obedecemos las leyes de la termodinámica	 Música Este foro es más comercial ahora que cuando yo lo descubrí



 **Ciencia**

Activos Nuevos Por leer (137) Top Noticias **Nuevo hilo**

Tema	Resp.	Lect.	Usuarios	Activ.
 Todo sobre ciencia [Vídeos, blogs, webs, libros, canales, podcasts]  Ciencia	71	11.7K		May '23
 [HO] NASA/SpaceX - Visitando otros planetas 1K+  Ciencia	3.1K	133.4K		16h
 No sé dónde más preguntar: reproducción asistida  Salud	6	299		2d
 ¿Qué es la felicidad?	73	1.2K		6d
 ¿Por qué orbitan en vez de chocar directamente?  Astronomía	11	999		11d
 Epic Spaceman  Ciencia	2	694		17d
 Iniciación a la astronomía 163  Astronomía	192	24.8K		22d

Una vez accedamos a www.mediavida.com clickamos en la pestaña “Foros”

Dentro de cada subforo encontraremos hilos de esa temática donde los usuarios podrán abrir y comentar hilos

Francisco Ibáñez el autor de Mortadelo y Filemon acaba de fallecer



ChaRliFuM

Jul '23 #31

Buff qué recuerdos de la infancia. Yo creo que del 2000 hacia atrás tenía prácticamente TODOS. Tendré que mirar en casa de mis abuelos a ver...

DEP.



2

1 respuesta ▾



sagem9

Jul '23 #32

Se va un grande. DEP



Tintintopete

Jul '23 #33

El Sulfato Atómico, El profeta Jeremías, Valor y al toro, Clínicas Antibirria...joder si puedo tener como 100 mortadelos en casa. De crío dice mi madre que hablaba con los mortadelos, diciendo cáspita, rayos, sapristi...si en catequesis llamaron la atención a mis padres porque le puse a Jesucristo la nariz de mortadelo. Es como si se hubiera muerto una parte de mi, es difícil de explicar para alguien que no le gustasen sus tebeos.



15



nosoyuncion

Jul '23 #34

Aprendí a leer con un Mortadelo en las manos, primero mirando los dibujos y poco a poco leyendo las viñetas

Descanse en paz, leyenda



9

1 respuesta ▾

Dentro de cada hilo los comentarios se dividirán en bloques donde veremos el nombre del usuario, su avatar, fecha del comentario, número de likes recibidos, etc

Se crea una nueva página cada 30 comentarios

Es posible **filtrar** los mensajes de un usuario en un hilo haciendo click en su nick y clicando en “Filtrar X posts”



ChaRliFuM

🕒 hace 5 horas

📅 13 jul 02

✉ Mensaje

🚫 Ignorar aquí

🔍 Filtrar 1 posts



chuckneitor

NOTICIAS

2

HILOS

9

POSTS

4.6K

FIRMAS

79

Últimos posts



Presentación de PlayStation 5 Pro, 10 de septiembre a las 17 horas

9d



< Formula 1 2024 - Hilo Oficial >

30 Jul



Stream de IlloJuan | Boof

18 Jul



Dark Souls II #HO | Scholar of the First Sin

8 Jul



Conflicto Israel-Irán

14 Abr



Final Fantasy VII Rebirth #HG | Empieza la fiesta

6 Mar



Así son las nuevas 'Tarjetas monedero' para familias vulnerabl

24 Ene



Tiendas Game dejara de vender juegos de segunda mano en UK

17 Ene



Resultados de Mediavida Game of the Year 2023

15 Ene



¿cuál ha sido vuestra mayor nómina?

28 Dic

Firmas



tasslehoff34

May '23

Bro tienes mp



shemar

Jun '19

PAYASO



CaraLimon

May '18

Hola, sabes donde ver la de Vengadores online a buena calidad? gracias 😊



T-1000

Jul '17

<http://www.mediavida.com/foro/off-topic/desde-cuando-existe-espana-como-nacion-588512>

Haciendo click en el nombre de un usuario accederemos a su perfil donde podremos ver:

- Últimos hilos comentados
- Número de hilos abiertos
- Número de comentarios totales
- Firmas de otros usuarios

Propósitos del proyecto

Obtener un dataset con toda la información del usuario o hilo para poder realizar un análisis de este, como por ejemplo:

- Evolución de la actividad a lo largo del tiempo
- Comentarios con mayor número de likes
- Palabras más frecuentes
- Análisis de sentimientos
- etc

El proyecto se divide en dos partes diferentes:

- 1. Extracción de datos**
- 2. Análisis de datos**

Extracción de datos

Referente al archivo *Extractor de datos.ipynb*

Usaremos las funciones *analisis_usuario(user)* y *analisis_hilo(hilo)* según la información que queramos extraer. Al finalizar obtendremos un dataset en formato .csv como el ejemplo, en este caso, se trata de un usuario

	Usuario	Subforo	idHilo	Hilo	Etiqueta	NumComent	Comentario	Likes	Fecha
0	Calles	off-topic	680521	rusia-invade-ucrania-hilo-general-	Noticia	20418	\n#20410 Jajajajajaja por Valencia son fallas...	0	2022-02-27 21:33:00
1	Calles	juegos	534364	realidad-virtual-hg-	PC	9243	\nUno que se ha comprado también Contractors a...	0	2022-02-21 18:35:00
2	Calles	juegos	534364	realidad-virtual-hg-	PC	9250	\n#9247 Puedes compartir link o algo? Gracias!	0	2022-02-23 19:47:00
3	Calles	off-topic	656229	coronachat-ii-filarmonica-edition-	NaN	20526	\nEn Valencia ya tenemos toque de queda, desde...	0	2020-10-25 09:23:00
4	Calles	juegos	506260	juegos-indie-de-gestion-hg-	Hilo general	3493	\n#3491 Nadie sabe de ningún digno sucesor? (n...	0	2019-12-07 20:19:00

Análisis de datos

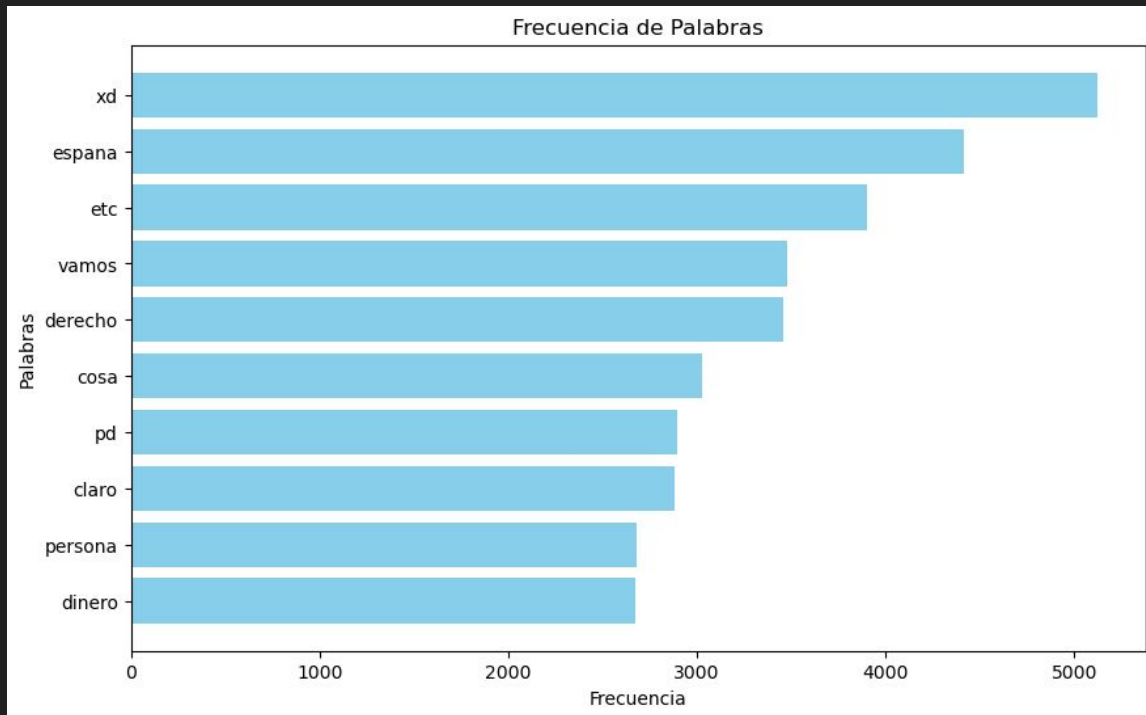
Referente al archivo *Análisis de datos.ipynb*

Este archivo servirá para cargar el dataset y consta de varias funciones:

1. **top_palabras(df, num_palabras)** Ordenará las palabras por su frecuencia de uso
2. **busca_palabras(df,palabra)** Encontrará los comentarios donde se ha dicho una palabra en específico
3. **top_manitas(df,n)**: Sacará los n comentarios con más likes obtenidos y creará la url para acceder a ellos

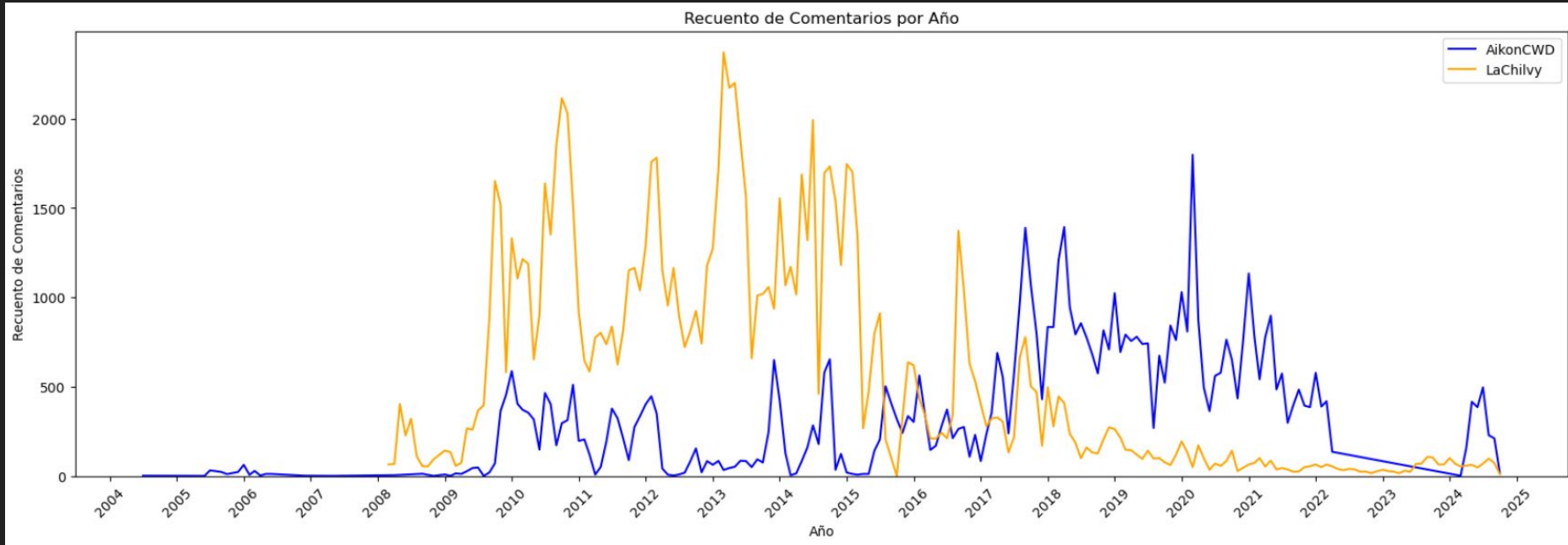
Para el historial de actividad se usará la herramienta PowerBi ya que nos permite hacer dashboards interactivos pero podríamos usar cualquiera que acepte .csv

Ejemplos



Palabras más frecuentes de un usuario

Ejemplos



Grado de actividad de dos usuarios diferentes a lo largo del tiempo

Fecha registro

11/08/2003 19:47:00

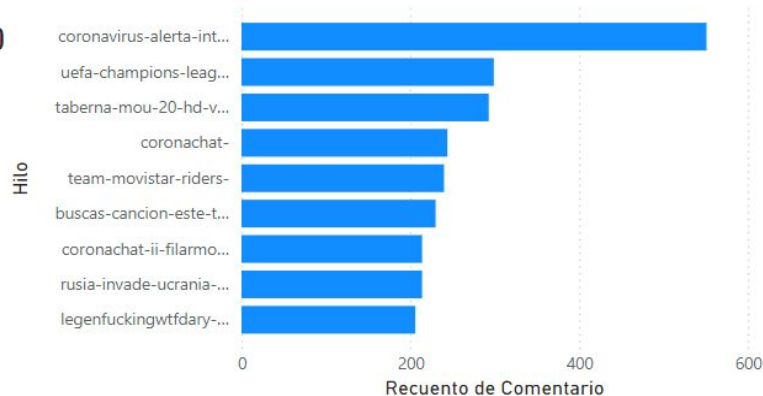
Comentarios totales

30393

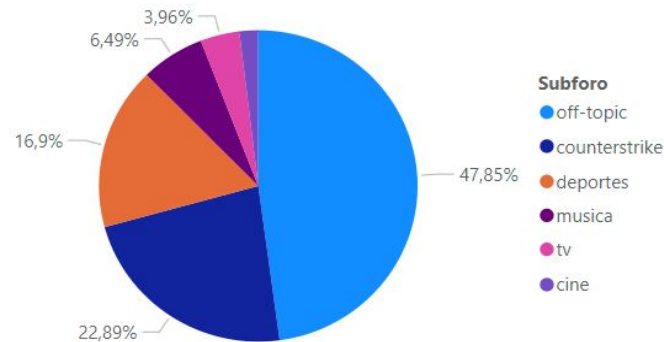
Likes totales

20051

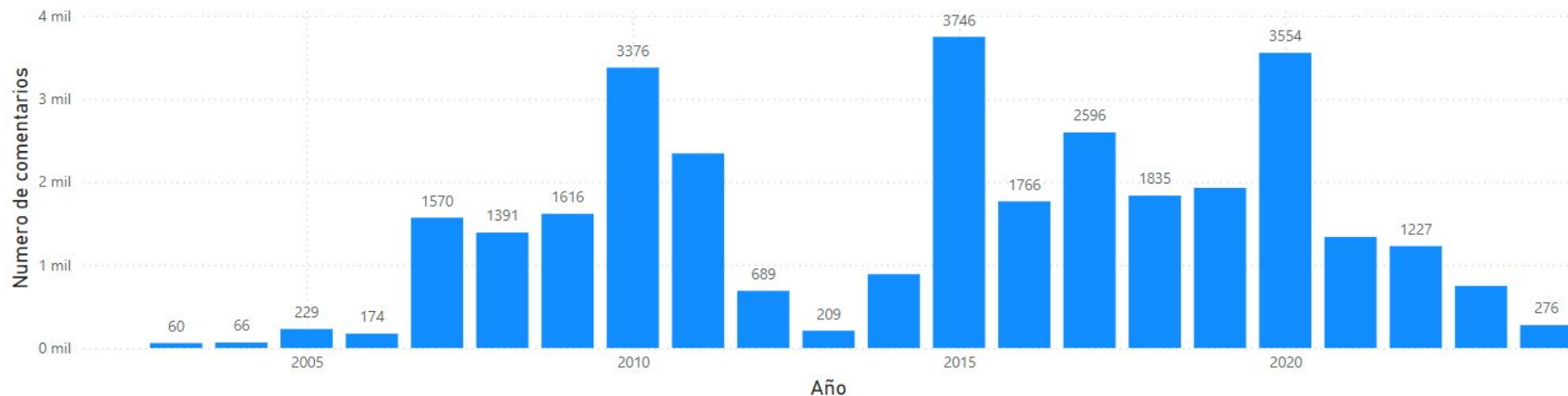
Hilos más comentados



Subforos más comentados



Actividad a lo largo de los años



Dashboard interactivo en PowerBi para un usuario