

Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología

Grado en Ingeniería Informática

**Análisis de datos en comunidades
virtuales: #claustrovirtual en Twitter**

Trabajo fin de estudio presentado por:	Alfonso Izquierdo Carrasco
Línea de investigación:	Ingeniería en la Web, servicios Web y seguridad en la Web
Director/a:	José Alberto Benítez Andrades
Fecha:	19/07/2022
Código fuente:	https://github.com/sitol/edu_

Resumen

#claustrvirtual es una comunidad educativa surgida de Twitter, en la cual se comparten contenidos, consejos y experiencias en materia educativa. Este trabajo busca, través de un análisis de datos en el espacio de dos meses, con las librerías Pandas y Matplotlib de Python, y una implementación iterativa y fundamentada en las metodologías ágiles, hacer un análisis de los tuits recientes publicados por los componentes de la comunidad que utilizan este hashtag. De esta manera, se busca conocer mejor las relaciones entre las diversas cuentas, los contenidos que comparten y los motivos por los que lo hacen. Se ha encontrado que esta comunidad es tal y como se define: colaborativa, solidaria y con un fin constructivo. Se comparten muchos contenidos relacionados con las nuevas tecnologías, además de experiencias en aulas, materiales didácticos y mensajes joviales y alentadores, entre otros. Se usan diversas redes y material audiovisual. Como es evidente, al ser un *hashtag* en español, predomina el castellano como lengua.

Palabras clave:

Análisis, datos, Twitter, educación

Abstract

#claustrovirtual is an educational community raised in Twitter, in which experiences, advice and contents related to education are shared. Through a two months length data analysis, using agile foundations and an iterative approach as methodologies, and using Pandas and matplotlib libraries, this work aims to perform an analysis of recent tweets published by the community members who use this hashtag. This way, the objective is to get a better understanding of the relationship between the different accounts, the content they share and the motivations that lie behind. It was found that the community is as it is stated to be by its members: very collaborative and with a positive mindset. A great part of the shared content is related to new technologies, classroom activities, learning materials and encouraging and leisure messages. Many different social networks and media are used, and Spanish dominates as the main language used.

Keywords: Analysis, data, twitter, education

Índice de contenidos

1.	Introducción.....	8
1.1.	Justificación del tema elegido	8
1.2.	Problema y finalidad del trabajo	9
1.3.	Objetivos del TFE.....	12
1.3.1.	Objetivo General	12
1.3.2.	Objetivos específicos	12
2.	Marco teórico	13
2.1.	Ser profesor, buscar y compartir: El aprendizaje informal en la formación del profesorado	13
2.2.	El análisis de datos	14
2.2.1.	El análisis de grandes cantidades de datos: Big data	14
2.2.2.	El análisis exploratorio de datos	15
3.	Contextualización.....	18
3.2.1.	Brandmentions	19
3.2.2	Tweetbinder.....	20
3.3.	Tecnologías a aplicar	21
3.3.1.	Jupyter Notebook	22
3.3.2.	Pandas.....	23
3.3.3.	Matplotlib	24
3.4.	Resumen de conclusiones.....	25
4.	Diseño de la propuesta	25
4.1.	Objetivos	25
4.1.1.	Objetivo General	25
4.1.2.	Objetivos específicos	26
4.1.2.1.	Aprender y profundizar sobre las maneras de compartir contenido entre el profesorado a través de la lectura y análisis de los mismos	26
4.1.2.2.	Crear un sistema que interactúe con la API de Twitter para que permita extraer los datos sobre la comunidad durante el tiempo de desarrollo del proyecto.....	26
4.1.2.3.	Crear un script en Python a través de Jupyter Notebooks en el cual se pueda volcar la información extraída de la API de Twitter.	26
4.1.2.4.	Utilizar la librería Pandas para modelar la información extraída y hacer un análisis a partir de los datos obtenidos.....	26
4.1.2.5.	Utilizar la librería Matplotlib para realizar gráficos, cuando sea necesario, que permitan visualizar de manera eficaz los datos analizados.	27
4.1.2.6.	Redactar y poner de relieve las conclusiones extraídas a través del análisis de la información y su visualización, a modo de resumen e informe.....	27
4.2.	Contenidos.....	27

4.2.1.	Obtener los datos: Obtención de los datos a través de la API Twitter.....	27
4.2.1.1.	La API de Twitter : Tuits recientes.....	27
4.2.1.2.	Script: Lambda.py	29
4.2.2.	Volcado de información.....	30
4.2.2.1.	Script de volcado de información.....	30
4.2.3.	Análisis de datos	30
4.2.3.1.	Muestra de los datos	30
4.2.3.2.	Tuits por día.....	30
4.2.3.3.	Ranking de tuits por persona	31
4.2.3.4.	Ranking de retuits por persona	32
4.2.3.5.	Ranking de <i>likes</i> por persona.....	33
4.2.3.6.	Nombre y número de <i>hashtags</i> relacionados.....	33
4.2.3.7.	Los tuits con más retuits.....	35
4.2.3.8.	Los tuits con más <i>likes</i>	36
4.2.3.9.	Los tuits con más respuestas.....	38
4.2.3.10.	Lista de sitios web más mencionados en los tuits.....	41
4.2.3.11.	Personas con más menciones	42
4.2.3.12.	Recursos multimedia más utilizados	43
4.2.3.13.	Lenguas utilizadas en los tuits	45
4.2.3.14.	Apps utilizadas para enviar los tuits.....	45
4.2.3.15.	Número de tuits con “contenido sensible”	46
4.2.3.16.	Número de tuits abiertos a la comunidad.....	48
4.3.	Metodología.....	48
4.3.1.	El marco de trabajo SCRUM.....	48
4.3.2.	Uso de SCRUM	49
4.3.2.1.	Roles dentro de SCRUM	49
	Desarrollador	49
	Product Owner	49
	Scrum Master	50
4.3.2.2.	Enfoque incremental e iterativo.....	50
4.4.	Evaluación	51
4.4.1.	Consistencia de datos (extracción manual vs automática)	52
4.4.1.1.	Obtención de datos	52
4.4.1.2.	Consistencia de los datos	53
4.4.2.	Efectividad de la herramienta en el análisis.....	54
5.	Conclusiones y trabajo futuro.....	57
5.1.	Conclusiones	57

5.1.1.	Resumen de las conclusiones obtenidas.....	57
5.1.1.1.	Volumen de publicaciones	57
5.1.1.2.	Actividad de la comunidad	57
5.1.1.3.	Temáticas de la comunidad.....	57
5.1.1.4.	Relaciones dentro de la comunidad	58
5.1.1.5.	Perfil de la comunidad.....	58
5.1.2.	El proceso de digitalización en las aulas	58
5.2.	Trabajo futuro	60
5.2.1.	Posibles mejoras en el estudio.....	60
5.2.2.	Más allá de #claustrvirtual	61

Índice de figuras

Figura 1.	Tuit en #claustrvirtual.....	10
Figura 2.	Tuits en #claustrvirtual	11
Figura 3.	Grafo de representación de grupos de influencers educativos según modularidad	14
Figura 4.	Cómo el análisis exploratorio de datos encaja en el proceso global de análisis.....	16
Figura 5:	Fases del análisis exploratorio de datos según la guía práctica de datos.gob.es	17
Figura 6.	Sección general del informe de Brandmentions sobre #claustrvirtual.....	19
Figura 7.	Nube de keywords, top influencers, hashtag relacionados y últimas menciones del informe de Brandmentions sobre #claustrvirtual	20
Figura 8.	Sección general del informe de Tweetbinder sobre #claustrvirtual	20
Figura 9.	Sección de información sobre usuarios (actividad, impacto, popularidad, retuits) del informe de Tweetbinder sobre #claustrvirtual	21
Figura 10.	Tecnologías basadas en Python para análisis de datos.....	22
Figura 11.	Ejemplo de Jupyter Notebook	23
Figura 12.	Tipos básicos de diagramas disponibles en Matplotlib.....	24
Figura 13:	Número de tuits por día	30
Figura 14:	Histograma sobre el número de tuits por persona	31
Figura 15:	Histograma sobre el número de retuits por persona.....	32
Figura 16:	El tuit con más retuits de #claustrvirtual	35
Figura 17:	El tuit con más likes de #claustrvirtual	37
Figura 18:	Número de likes por cada tuit.	38
Figura 19:	Tuit con más respuestas	39
Figura 20:	Encuesta con respuestas	40
Figura 21:	Tuit buscando visibilidad	42
Figura 22:	Tuit con gif animado	44

Figura 23: Gráfico de sectores con las apps más utilizadas.....	46
Figura 24: Tuit calificado por Twitter como “posible contenido sensible”	47
Figura 25: Tuit calificado por Twitter como “posible contenido sensible”	47
Figura 26: Archivos de datos .json sobre los tuits	53
Figura 27: Detalle de diferencias entre los .json obtenidos de manera manual y automática 54	
Figura 28: Tuit de apoyo a los miembros de #claustrvirtual	59

Índice de tablas

Tabla 1: Parámetros utilizados para obtener la información sobre los tuits #claustrvirtual.	28
Tabla 2: Ranking de tuits por persona	31
Tabla 3: Análisis estadístico de tuits por persona	32
Tabla 4: Ranking de <i>likes</i> recibidos por persona	33
Tabla 5: Ranking de apariciones de cada <i>hashtag</i>	34
Tabla 6: Ranking de tuits con más retuits	36
Tabla 7: Ranking de tuits con más likes	36
Tabla 8: Ranking de tuits con más respuestas.....	38
Tabla 9: Ranking de <i>URLs</i> compartidas en los tuits de #claustrvirtual.....	41
Tabla 10: Ranking de personas más mencionadas.....	43
Tabla 11: Ranking de recursos multimedia más utilizados	44
Tabla 12: Ranking de lenguas más utilizadas en los tuits.....	45
Tabla 13: Ranking de aplicaciones más utilizadas para enviar tuits.....	46
Tabla 14: Número de tuits con contenido sensible.....	47
Tabla 15: Tuits abiertos solo al público o a seguidores.....	48
Tabla 16: Tabla de evaluación de parámetros del análisis.....	55

1. Introducción

Hoy en día, hablar de redes sociales y de las oportunidades que ofrecen para compartir experiencias, historias y aprendizajes no es nada nuevo. De hecho, llevamos viviendo este fenómeno desde hace más de una década, cuando explotaron plataformas como Facebook o Twitter. Esta última, por su estructura más abierta en la que se unas personas siguen a otras sin necesidad de ser “amigos” (como ocurre en Facebook), ha dado pie a innumerables debates y charlas y conversaciones entre gente desconocida.

Esto no ha ocurrido solo a través de la herramienta del Retuit, sino también del uso de hashtags que cualquier usuario puede seguir, y que permiten categorizar el conocimiento y participar de él. Es esto lo que ha ocurrido con la creación del #claustrvirtual, una comunidad educativa con distintos perfiles de personas que comparte contenido relacionado con la educación en la plataforma de Twitter.

1.1. Justificación del tema elegido

El profesorado está viviendo, con mejor o peor suerte, una revolución en torno a lo digital en los últimos años, acrecentada sobre todo por la reciente pandemia de COVID-19 (Grinin et al., 2022). Como ocurre en estos casos, se ha pasado de obviar la tecnología en las aulas, a sentir curiosidad por ella y, finalmente a, de manera inevitable y con una pandemia mediante, verse abocados a utilizarla. No son pocos los casos de profesores que se han sentido perdidos o han llegado a sentir ansiedad al verse obligados a utilizar nuevas tecnologías que no comprendían.

Otro sector del gremio, normalmente más joven y abierto a nuevas experiencias, ha entrado de lleno a usar la tecnología en el aula, no solo a través de la utilización de nuevos dispositivos, sino también incluyéndola como parte de sus métodos de enseñanza. No estamos hablando de buscar nuevas actividades o ideas que llevar a los alumnos, sino de algo mucho más grande: nuevas maneras de evaluar, estrategias de enseñanza-aprendizaje, metodologías como el aprendizaje por proyectos o el aprendizaje cooperativo... Incluso existen mezclas de todos estos elementos en los que la tecnología juega un papel importante, como el Flipped Classroom, donde los alumnos, a través de vídeos y otros recursos acceden al conocimiento para compartirlo después en el aula.

Sin embargo, aún no hemos hablado de cómo los profesores llegan a esta información y posteriormente la comparten. Indudablemente, el boca a boca ayuda, pero, ¿qué hace un

profesor que busca inspiración y no la encuentra entre sus compañeros, que siguen haciendo lo mismo año tras año?

Durante muchos años, ya fuese unas horas, a media jornada, o incluso a tiempo completo, he completado mi profesión de ingeniero informático con la de profesor particular, ya sea a través de clases particulares o a pequeños grupos. Preparar clases solo basadas en un manual o libro de referencia se me hacía bastante pobre, así que siempre he buscado inspiración en Internet. Esta búsqueda no solo consiste en utilizar el buscador de Google, sino que muchas otras veces se utilizan otras herramientas, como son las redes sociales.

En Twitter esta práctica se ha extendido en los últimos años. Sobre todo, a partir del año 2014, un grupo de personas comenzó a usar el hashtag #claustrovirtual para pedir consejo, preguntar dudas, compartir recursos, y a veces fotos, escenas, o experiencias de su día a día en las aulas.

El presente trabajo busca analizar la red social Twitter, y en concreto la comunidad que ha surgido en torno al *hashtag* #claustrovirtual. Cada día se comparte mucha información de manera informal, y es interesante analizarla y ver no solo quién la comparte o cómo ha sido la evolución de la comunidad, sino también qué otros hashtags o temas se relacionan con esta comunidad, o cuáles son las páginas que se suelen compartir.

Esto permitirá dar un paso más, y pasar de tener solo información a saber cómo esta está organizada. De esta manera, una persona podría fácilmente ir descubriendo la gente que participa en la comunidad, y así tener acceso a los recursos que utilizan y los consejos que dan.

Para mí, que he utilizado Twitter alguna vez para conocer nuevos recursos y plataformas, este trabajo representa una aportación práctica y valiosa a la comunidad, para así predicar con el ejemplo y compartir este análisis con todos ellos, partícipes de querer convertir la educación en un entorno colaborativo y de calidad.

1.2. Problema y finalidad del trabajo

Si hablamos de la comunidad #claustrovirtual, nos referimos a una red de personas compartiendo distintos recursos educativos, formaciones, actividades o reflexiones, entre otros contenidos. Un ejemplo es este tuit:

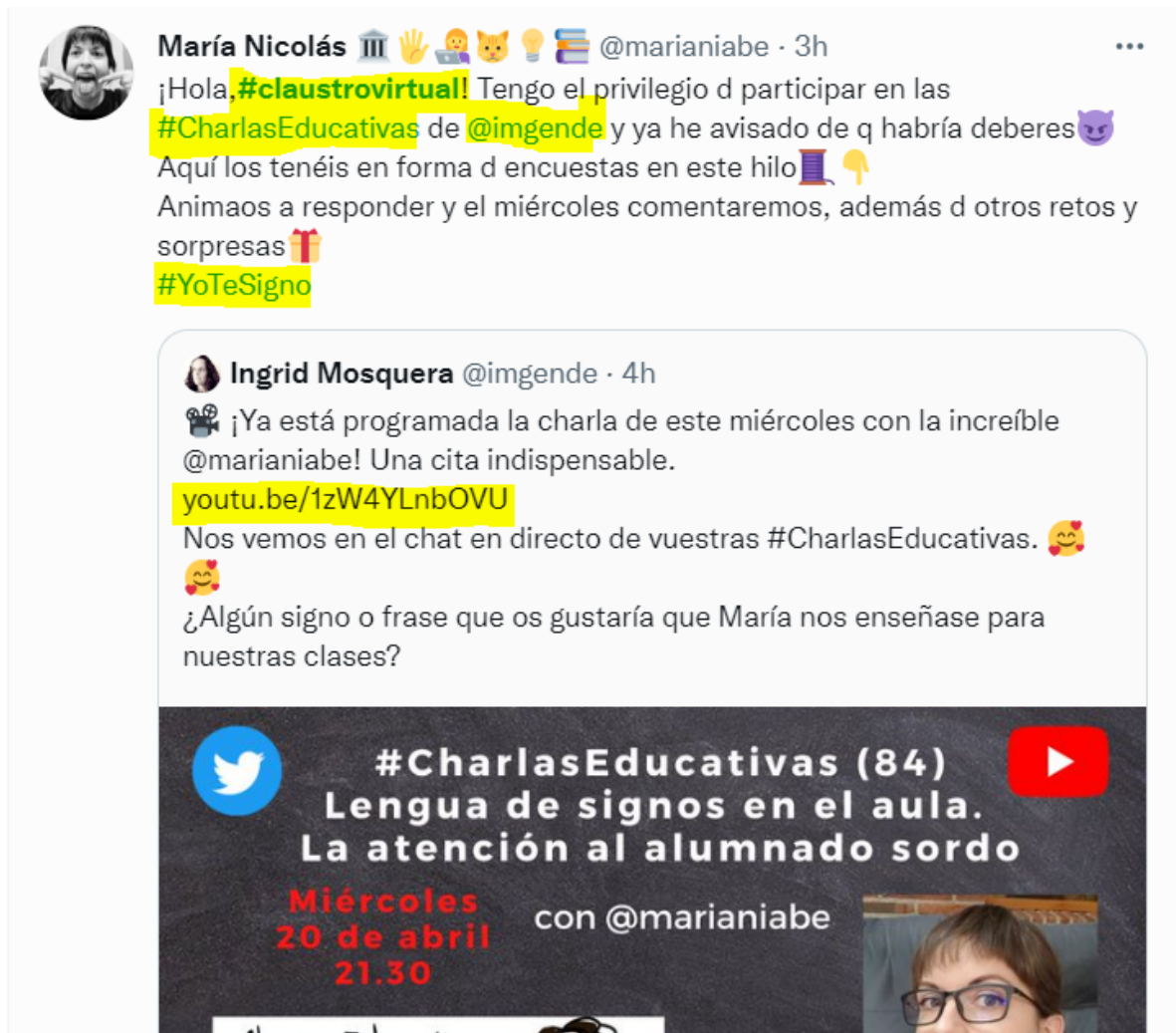


Figura 1. Tuit en #claustrvirtual

Fuente: <https://twitter.com/marianiabe/status/1516075330427142149>

En él, podemos ver que hay bastante información:

- Se nombran otros dos hashtags: #CharlasEducativas, una serie de vídeo-entrevistas en las que un ponente habla sobre distintas herramientas o enfoques educativos. También aparece #YoTeSigno, que tiene que ver con el lenguaje de signos.
- Se menciona a otra persona, @imgende, la persona que se dedica a hacer estas entrevistas.
- Vemos que hay un enlace a YouTube, la página en la que se hacen estas entrevistas.

Con este tuit podemos ver que hay personas interesadas en el lenguaje de signos en la educación, y que comparten su conocimiento utilizando otras plataformas externas para ello.

Ahora bien, a veces no buscamos informarnos sobre algo en concreto, sino más bien descubrir, explorar nuevas opciones, inspirarnos... En este caso, no es fácil navegar por la comunidad, pues la información es muy diversa y no está organizada.

Por ejemplo, si miramos los dos tuits anteriores al ejemplo que acabamos de comentar, vemos que las temáticas son totalmente distintas. En el primero se habla de ciencias naturales, y en el segundo sobre expresión oral:



Figura 2. Tuits en #claustrvirtual

Fuente: twitter.com

Así pues, es interesante poder hacer un análisis y ver cuáles son las relaciones entre diversas personas, y que así una persona interesada, por ejemplo, en las matemáticas, podría conocer la gente que está compartiendo recursos, prácticas y experiencias sobre las mismas.

También es una buena herramienta para ver de manera clara cuál está siendo la tendencia de la comunidad, si está creciendo o se mantiene estable, si hay solo algunas personas activas y una gran cantidad que solo la utilizan de manera puntual... Toda esta información puede servir para dar visibilidad y claridad a la comunidad.

De esta manera, podemos pasar de tener información en bruto y no organizada, a tener un análisis detallado a través del cual podemos ver tendencias y descubrir personas y temáticas que solo usando Twitter nos sería imposible descubrir, pues necesitaría hacer *scroll* durante horas anotando a mano la información.

1.3. Objetivos del TFE

1.3.1. Objetivo General

Diseñar un análisis sobre la comunidad educativa #claustrovirtual que permita descubrir nuevas personas y temáticas relacionadas con la misma, además de una idea general sobre el progreso de la comunidad (número de gente que participa, número de tuits, popularidad de los mismos e impacto...).

1.3.2. Objetivos específicos

1. Aprender y profundizar sobre las maneras de compartir contenido entre el profesorado a través de la lectura y análisis de los mismos
2. Crear un sistema que interactúe con la API de Twitter para que permita extraer los datos sobre la comunidad durante el tiempo de desarrollo del proyecto.
3. Crear un script en Python a través de Jupyter Notebooks en el cual se pueda volcar la información extraída de la API de Twitter.
4. Utilizar la librería Pandas para modelar la información extraída y hacer un análisis a partir de los datos obtenidos.
5. Utilizar la librería Matplotlib para realizar gráficos, cuando sea necesario, que permitan visualizar de manera eficaz los datos analizados.
6. Redactar y poner de relieve las conclusiones extraídas a través del análisis de la información y su visualización, a modo de resumen e informe.

2. Marco teórico

Este trabajo se basa en dos fuentes a nivel teórico. En primer lugar, existe un contexto en el marco educativo, centrado a nivel del profesorado y la manera de acercarse al aprendizaje informal. En segunda instancia, se explora el análisis de datos en un estudio más técnico, que permitirá evaluar los datos obtenidos de la comunidad educativa #claustrvirtual.

2.1. Ser profesor, buscar y compartir: El aprendizaje informal en la formación del profesorado

Cuando una persona se convierte en profesor o profesora, tiene frente a sí una tarea infinitamente amplia y comprometida en cuanto a retos y posibilidades se refiere. Por un lado, se enfrenta a aulas con, muchas veces, veinticinco o treinta alumnos, que viven situaciones muy distintas a nivel cultural, familiar y socioeconómico. Hoy día, cada vez vivimos una mayor segregación escolar, sobre todo en ciudades grandes, como Madrid, y reducirla es una tarea compleja (Ferrer & Gortázar, 2021). Por tanto, acoger e incluir toda la diversidad en las aulas, en un mundo globalizado, no es fácil.

Sin embargo, esta complejidad ha traído más alternativas. Poco a poco vemos cómo la clase magistral tradicional ha dado paso a nuevos estilos de enseñanza-aprendizaje o metodologías activas, como el Aprendizaje basado en Proyectos (ABP), el Aprendizaje-Servicio (ApS), o el Flipped Classroom o aula invertida (Thinkö, 2020). Estas, a su vez, se han integrado con las nuevas tecnologías o Tecnologías de la Información y Comunicación (TIC) para ofrecer experiencias muy distintas a las sucedidas en la década pasada, donde el libro, la explicación del profesor, y si acaso alguna presentación de diapositivas predominaba.

Todas estas variables han hecho que el sector del profesorado más curioso y motivado busque nuevos recursos y maneras de enseñar, más allá de la educación formal que reciben, que a veces no es suficiente por los distintos tipos y niveles de interés en el sistema educativo, y todo el peso político y burocrático que subyace detrás (Menor, 2021).

Así pues, encontramos que una parte del contenido que algunos profesores llevan a sus aulas procede del compartir que acontece en Internet, y más concretamente en redes sociales como Twitter. Es en esta red donde ha surgido #claustrvirtual, una comunidad sin jerarquía ni promotor, construida para apoyarse y ayudarse de manera solidaria (Mosquera, 2022). Muchas de estas personas han llegado a tener una influencia notable en la práctica docente

Análisis de datos en comunidades virtuales: #claustrovirtual en Twitter de muchos centros educativos, y se han erigido como *influencers* educativos en Twitter, si analizamos los datos existentes (Marcelo & Marcelo, 2021).

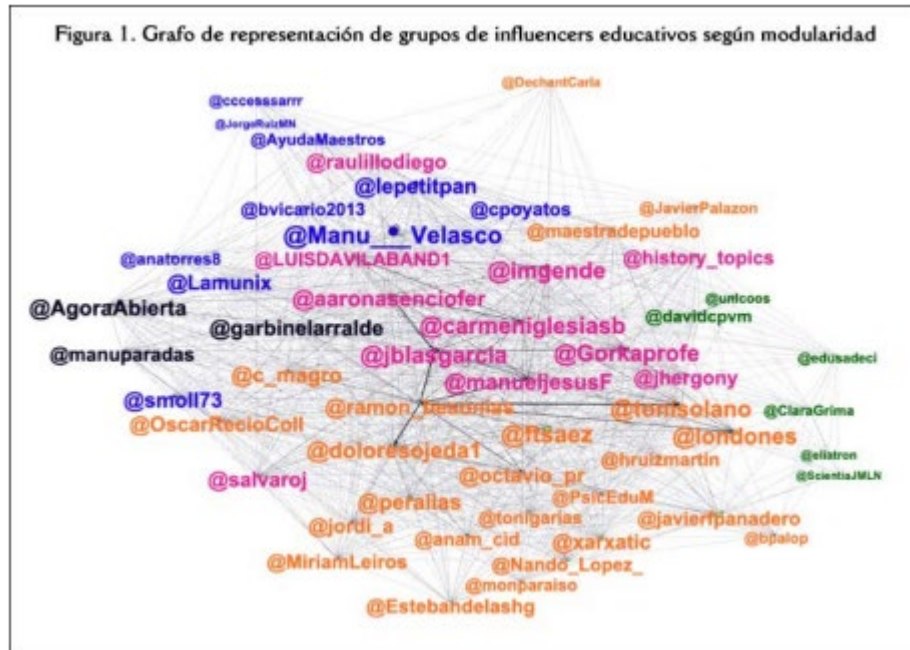


Figura 3. Grafo de representación de grupos de influencers educativos según modularidad

Fuente: Comunicar, 68, XXIX, 2021

2.2. El análisis de datos

2.2.1. El análisis de grandes cantidades de datos: Big data

Podría parecernos, a juzgar por la cantidad de términos nuevos que aparecen prácticamente a diario o cada mes, que el análisis de datos, u otros términos ampliamente utilizados como *data science* o *big data*, es algo que solo pertenece a nuestros días. Sin embargo, todo este arte de recopilar datos, analizarlos y mostrarlos visualmente ya se llevaba a cabo en muchos campos como la cartografía, la medicina y la estadística, cada uno en su tiempo, pero hace ya siglos (Friendly, 2008).

Ya situándonos en el día de hoy, con la gran cantidad de datos a nuestra disposición, se ha definido el término *big data* de distintas maneras. Un ejemplo sería :

Big Data is the amount of data beyond the ability of technology to store, manage and process efficiently (Big data: The next frontier for innovation, competition, and productivity | McKinsey, s. f.)

Esta definición nos habla de la gran cantidad de datos y del reto que supone analizarlos y sacar conclusiones a través de ellos.

Según algunos estudios, el *big data* tiene 5 características (Özköse et al., 2015):

- Volumen: El tamaño de los datos (comúnmente llamado *dataset*). Siempre se habla de una gran cantidad de datos.
- Variedad: El hecho de que los datos provengan de distintas fuentes, ya sean internas o externas al entorno de análisis.
- Velocidad: La generación de datos puede ser muy rápida. El aumento notable de datos significa que estos deberían de ser analizados con más celeridad. Cuanto más grande es el tamaño de los datos, más necesidad hay de datos.
- Veracidad: La precisión de los datos. La fuente de la cual los obtenemos ha de ser fidedigna y segura. Solo la gente autorizada podrá acceder a la misma.
- Valor: Se genera un resultado tras el tratamiento de los datos, el cual debe enriquecer el proceso.

Por lo tanto, ya concluyendo, hoy en día tenemos la oportunidad de hacer análisis sobre grandes cantidades de datos, que, si cumplen con las características ya mencionadas, podemos considerar *big data*.

2.2.2. El análisis exploratorio de datos

El análisis exploratorio de datos comenzó con el trabajo del estadístico John Tukey en la década de los 60 y los 70 (Wongsuphasawat et al., 2019). Desde entonces, son muchos los campos que han contribuido a hacer crecer a este, como son la estadística, la interacción persona-computadora, o la visualización de la información. Hay una amplia discusión sobre las fronteras de este tipo de análisis.

En el análisis exploratorio de datos, se busca complementar el análisis estadístico formal con este otro análisis, en cual requiere del investigador de una “actitud flexible” a la hora de examinar los datos, para luego complementarlos con un análisis estadístico formal (en palabras de Tukey, *confirmatory data analysis*) (Tukey, 1980).

El siguiente gráfico nos muestra las fases de un proceso global de análisis y cómo el análisis exploratorio de datos entraría dentro del mismo (Kandel et al., 2012). En itálica, las actividades de análisis exploratorio que se podrían introducir (Alspaugh et al., 2019).

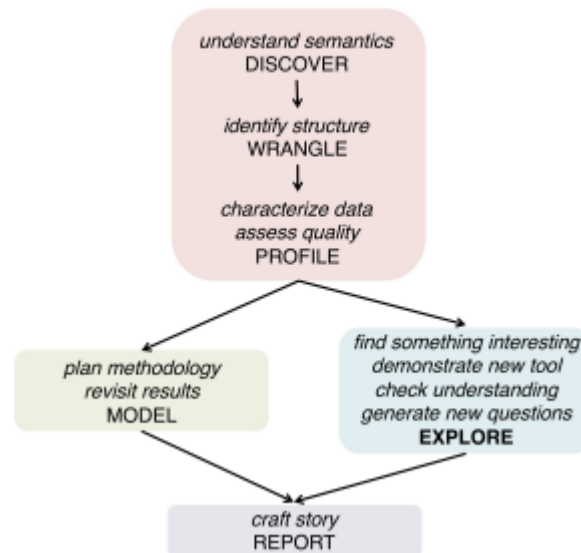


Figura 4. Cómo el análisis exploratorio de datos encaja en el proceso global de análisis.

Fuente: <https://people.ischool.berkeley.edu/~hearst/papers/vast2018.pdf>

Así pues, podemos entender que este análisis se encuentra en los dominios de aquel que pretende obtener datos y visualizarlos de una manera informal para obtener indicios que se confirmarán en un análisis estadístico formal. La plataforma datos.gob.es da una definición que resume esto:

El Análisis Exploratorio de Datos o EDA es el conjunto de técnicas estadísticas cuyo fin es explorar los datos de forma preliminar a la aplicación de cualquier proceso posterior como una investigación científica o una visualización interactiva de datos. Se trata de un proceso fundamental para el entendimiento básico de los datos y las relaciones que existen entre ellos.

En muchos de estos análisis se usan lenguajes como Python, R, o la librería Pandas de Python para hacer el análisis de datos, junto con herramientas de visualización como Matplotlib o Seaborn.

Una de las opciones para hacer este análisis exploratorio de datos según la guía de datos.gob.es ¹ es la siguiente:

¹ <https://datos.gob.es/es/>



Figura 5: Fases del análisis exploratorio de datos según la guía práctica de datos.gob.es

Fuente:

https://datos.gob.es/sites/default/files/doc/file/analisis_exploratorio_de_datos_2021_v6.pdf

f

En ella observamos 5 fases a la hora de hacer el análisis:

1. Realizar un análisis descriptivo de las variables, para obtener una idea que represente al conjunto de todos los datos.
2. Reajustar los tipos de las variables para que sean consistentes y no haya problema a la hora de realizar las operaciones de análisis de datos.
3. Detección y tratamiento de datos ausentes. Tratar con los datos que pudieran faltar, ya que no serían fiables a la hora de procesarlos y no se podría hacer el análisis con estas variables numéricas
4. Identificar y tratar los datos atípicos, ya que pueden dar problemas a la hora de realizar futuros análisis estadísticos.
5. Realizar un examen, ya sea numérico, gráfico o ambas opciones, sobre las relaciones entre las variables analizadas para determinar el grado de correlación entre ellas, de cara a ver qué relación existe entre las mismas, y las conclusiones que podemos obtener de su análisis.

3. Contextualización

3.1. Dominio de la aplicación. Necesidades detectadas

Aunque ha surgido de manera muy natural con el paso de los años, el caso de la comunidad #claustrovirtual no es muy común en tanto que no hay un claro grupo de personas que realmente velen por que este se mantenga y crezca, si bien es cierto que muchas personas contribuyen en su día a día y de manera desinteresada a ello.

Esto hace que, aunque haya parte de la comunidad que muchas veces se ha organizado para determinadas acciones, no hay un claro liderazgo que marque lo que se espera o lo que representa la comunidad, más allá de ser un espacio colaborativo y constructivo para compartir recursos y experiencias. Los tuits aparecen en la plataforma de manera desordenada y no organizada, y eso no ayuda a la hora de hacer un análisis general del contexto.

Por un lado, esto otorga una flexibilidad en la que toda iniciativa es acogida, pero a la vez no es fácil ver de manera clara los contenidos existentes, o hacerse a la idea de la evolución del grupo. Es en este punto en el que, un proyecto como este, cobra sentido. Se puede ver que hay un potencial para invitar a la reflexión, a pararse, observar los datos y dar oportunidad para tomar decisiones en torno a ellos. Estas conclusiones pueden convertirse en material para la concienciación, sensibilización y motivación de los docentes.

3.2. Análisis de hashtags: Herramientas existentes

A la hora de buscar herramientas que realicen análisis de datos por Twitter, da la impresión de que la mayoría van destinadas al mundo del marketing, ya que se insiste mucho en la idea del valor de la información aportada por la herramienta, a veces transformada en una estimación en dólares. Otro aspecto a reseñar es que la gran mayoría de herramientas encontradas siguen un modelo *freemium*, en el que se aporta un pequeño análisis del *hashtag* gratis, para ofrecerte una versión de pago en el que se accede a todo el potencial de la herramienta.

A continuación, se muestran un par de ejemplos de herramientas que hacen análisis de datos a través de *hashtag*.

3.2.1. Brandmentions

Brandmentions es una herramienta para hacer seguimiento de marcas y competidores. Permiten hacer una búsqueda simple por *hashtag* de manera gratuita, y hacen un informe en el que se pueden visualizar, para los últimos 7 días, las menciones, interacciones, alcance, las veces que se ha compartido el *hashtag* y los *likes*.

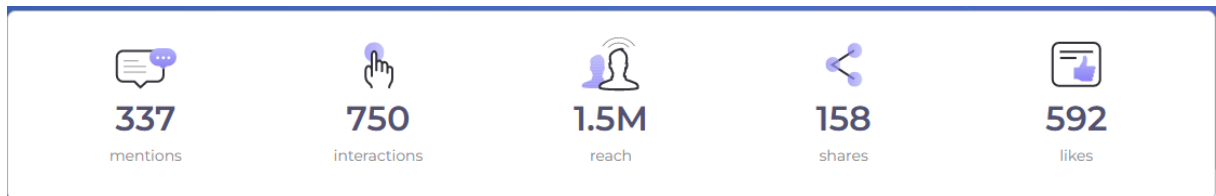


Figura 6. Sección general del informe de Brandmentions sobre #claustrvirtual

Fuente: <https://app.brandmentions.com/h/p/332484/%23claustrvirtual>

Además, se presenta una nube de *keywords*, los *top influencers* y los *hashtags* relacionados, además de las últimas menciones. No solo utilizan Twitter para el análisis, sino también Facebook. Accediendo a la suscripción de pago, puede verse el contenido de las últimas menciones (en la versión gratuita hay partes ocultas), y un análisis del sentimiento. También se necesita pagar para ver un análisis pormenorizado en cada categoría.

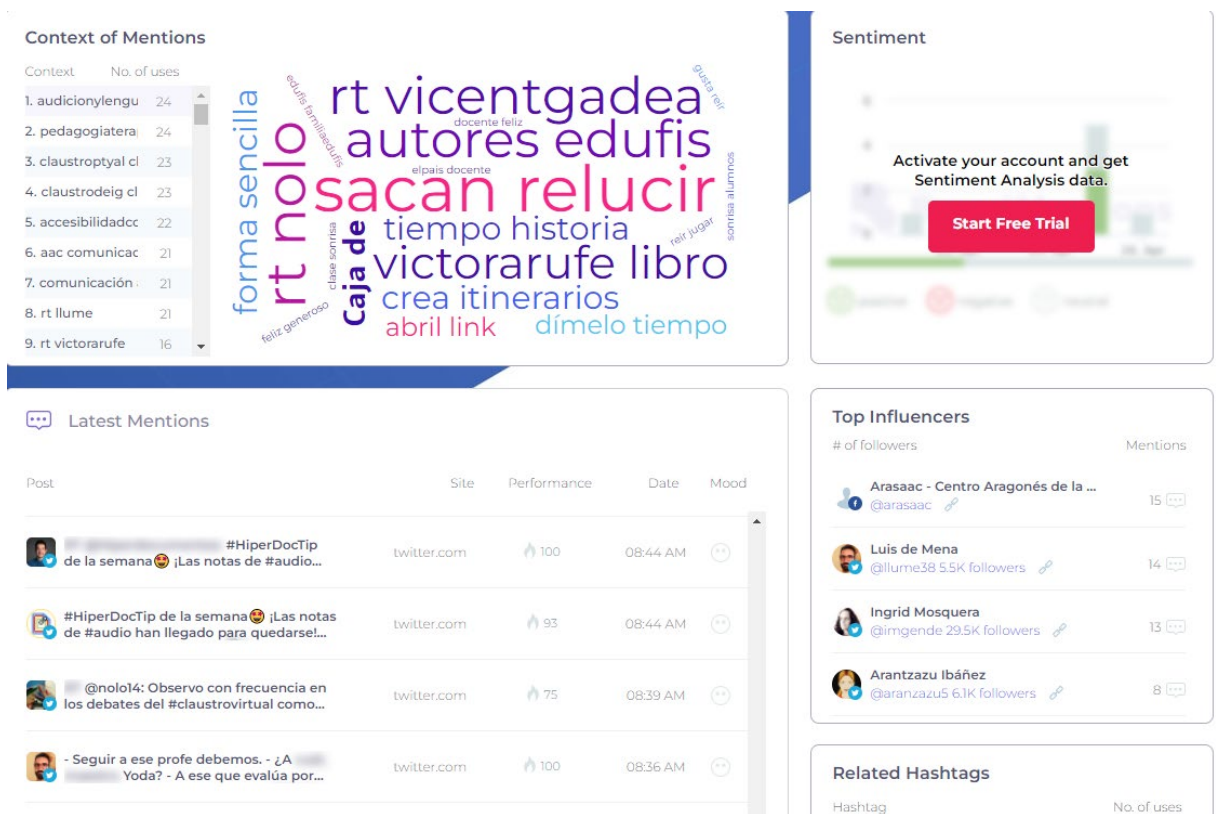


Figura 7. Nube de keywords, top influencers, hashtag relacionados y últimas menciones del informe de Brandmentions sobre #claustrvirtual

Fuente: <https://app.brandmentions.com/h/p/332484/%23claustrvirtual>

La versión gratuita permite hacerse una idea del alcance del *hashtag* sin poder profundizar mucho, y la nube de *keywords* es visual, pero no parece muy conseguida, pues las palabras clave no son relevantes respecto a la comunidad (no suelen aparecer palabras del ámbito de la educación).

3.2.2 Tweetbinder

Tweetbinder se presenta como una herramienta de análisis de Twitter, que permite crear informes de datos históricos (30 días o un año), y también sobre datos en tiempo real. También se pueden comprar informes de análisis de perfiles específicos. Para probar la herramienta, se ha utilizado la versión de prueba, que analiza los últimos 500 tuits en los últimos 7 días.

En esta versión, se pueden observar datos como el número de contribuciones, los nombres de los contribuidores y su alcance, o el número y porcentaje de links e imágenes por cada 500 tuits. Además, se hacen estimaciones a nivel económico, como los impactos potenciales a usuarios y su alcance, y el valor económico del *hashtag* en la red, que desconozco cómo son calculados.

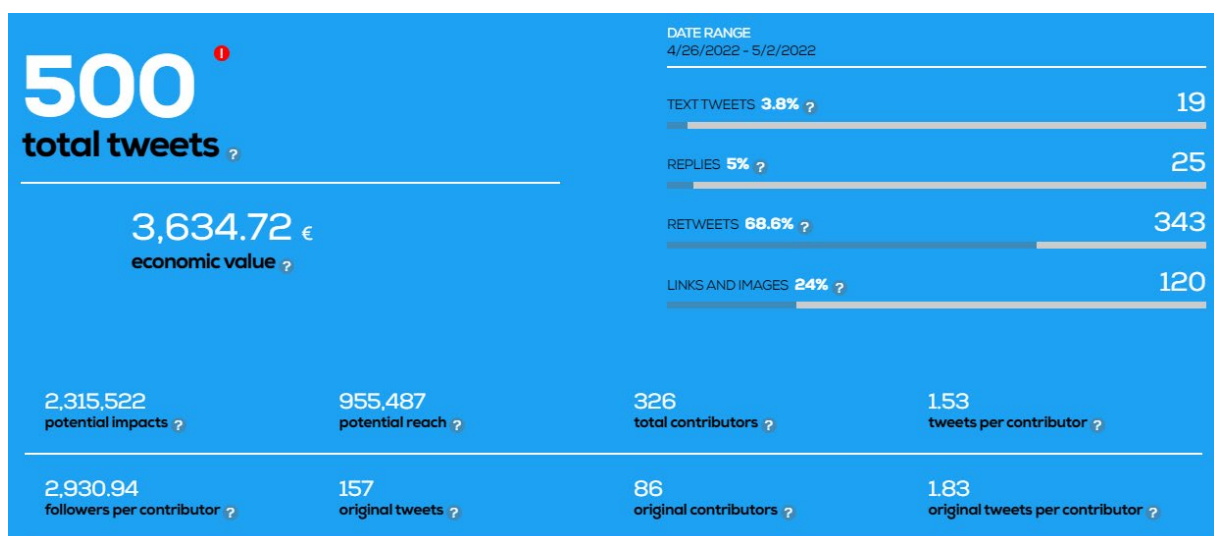


Figura 8. Sección general del informe de Tweetbinder sobre #claustrvirtual

Fuente: <https://dash.tweetbinder.com/report/free>

Por último, también se ofrece en esta versión gratuita la evolución en el tiempo del número de tuits posteados cada día durante la última semana y un ranking de contribuidores basado en distintos parámetros (los más populares, más activos, los que retuitean más, impacto más alto...).

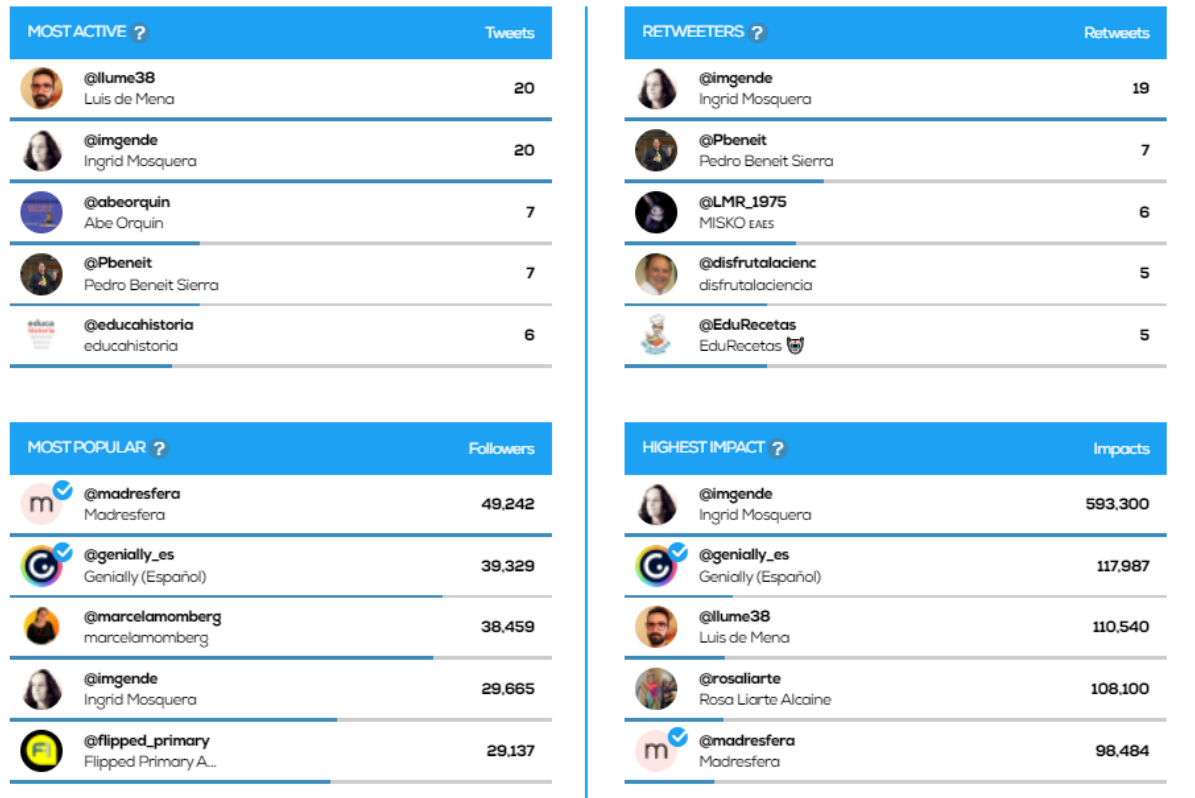


Figura 9. Sección de información sobre usuarios (actividad, impacto, popularidad, retuits) del informe de Tweetbinder sobre #claustrvirtual

Fuente: <https://dash.tweetbinder.com/report/free>

3.3. Tecnologías a aplicar

En el contexto actual de la analítica de datos, existen distintas tecnologías a aplicar en lo referente al análisis de datos. Una de las primeras decisiones a tomar es la elección del lenguaje de programación. Hoy en día, existen opciones como Python, R o Matlab. Por lo natural de su aprendizaje en un inicio (Ozgur et al., 2021) y las librerías que tiene, las cuales son muy populares y gozan de un buen soporte, se ha seleccionado Python como elección de lenguaje de programación. Además, la facilidad a la hora de entender el código permite que la curva de aprendizaje se haga más corta en relación a otros lenguajes, y tiene una amplia selección a la hora de crear gráficos para visualizar los datos.

A continuación, se describen las tecnologías que se utilizarán para la realización del proyecto:

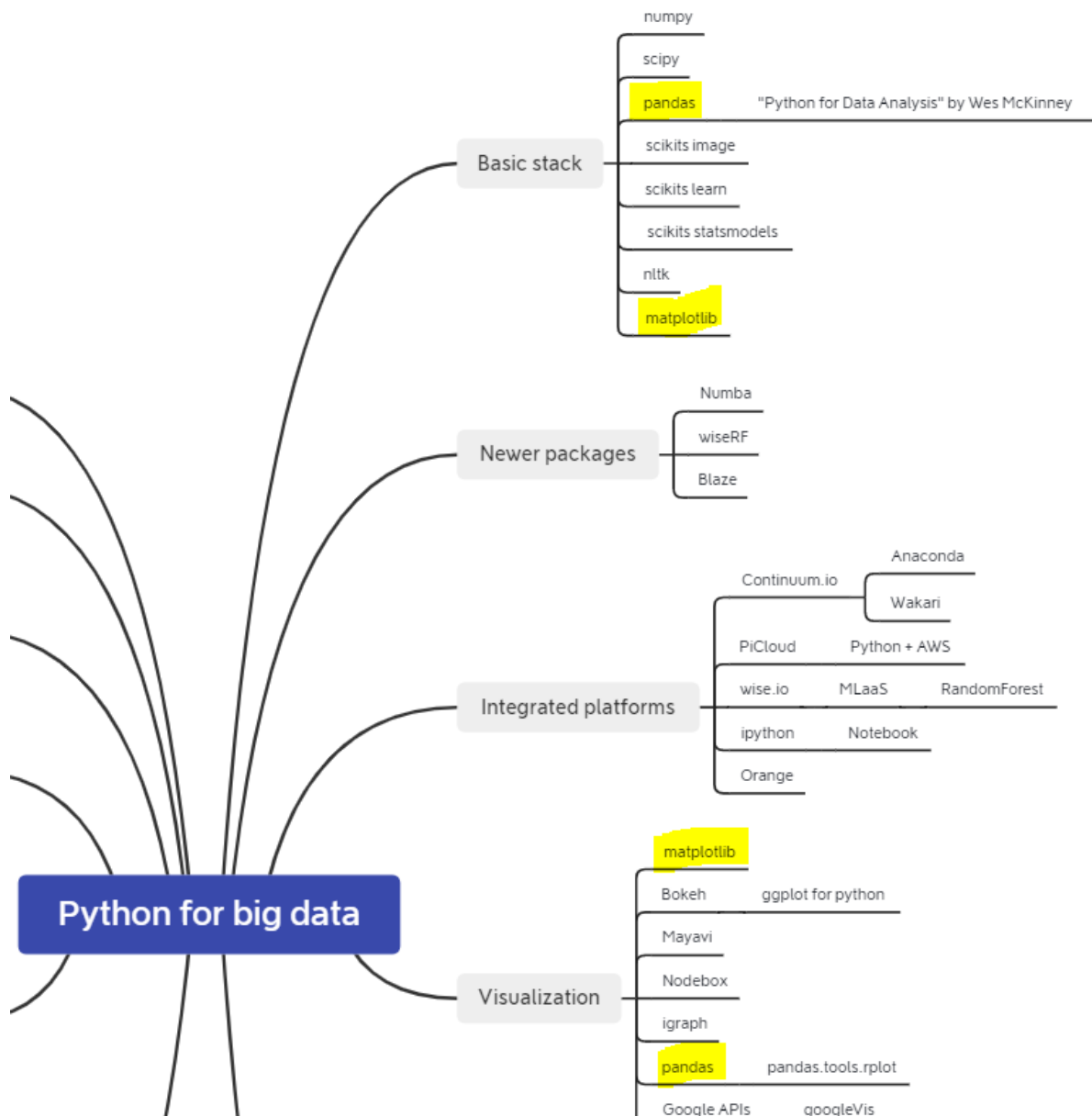


Figura 10. Tecnologías basadas en Python para análisis de datos

Fuente: <https://xmind.net/m/WvFC/#>

3.3.1. Jupyter Notebook

De igual manera que a la hora de desarrollar código solemos utilizar entornos de desarrollo integrado o IDEs (*Integrated Development Environment*), para hacer análisis de datos existe una herramienta que nos permite realizar de manera muy sencilla los scripts, y poder ejecutarlos y ver sus resultados en tiempo real. Esta herramienta son los Jupyter Notebooks, bajo el amparo de Project Jupyter, una organización sin ánimo de lucro y de código abierto nacida en 2014 a raíz del IPython Project.

En un Jupyter Notebook se puede crear un script e ir ejecutándolo paso a paso, pudiendo ver en cada uno de ellos los resultados, y siendo posible volver a ejecutarlos o cambiarlos para obtener el resultado esperado.

```
[1]: from matplotlib import pyplot as plt
import numpy as np

# Generate 100 random data points along 3 dimensions
x, y, scale = np.random.randn(3, 100)
fig, ax = plt.subplots()

# Map each onto a scatterplot we'll create with Matplotlib
ax.scatter(x=x, y=y, c=scale, s=np.abs(scale)*500)
ax.set(title="Some random data, created with JupyterLab!")
plt.show()
```

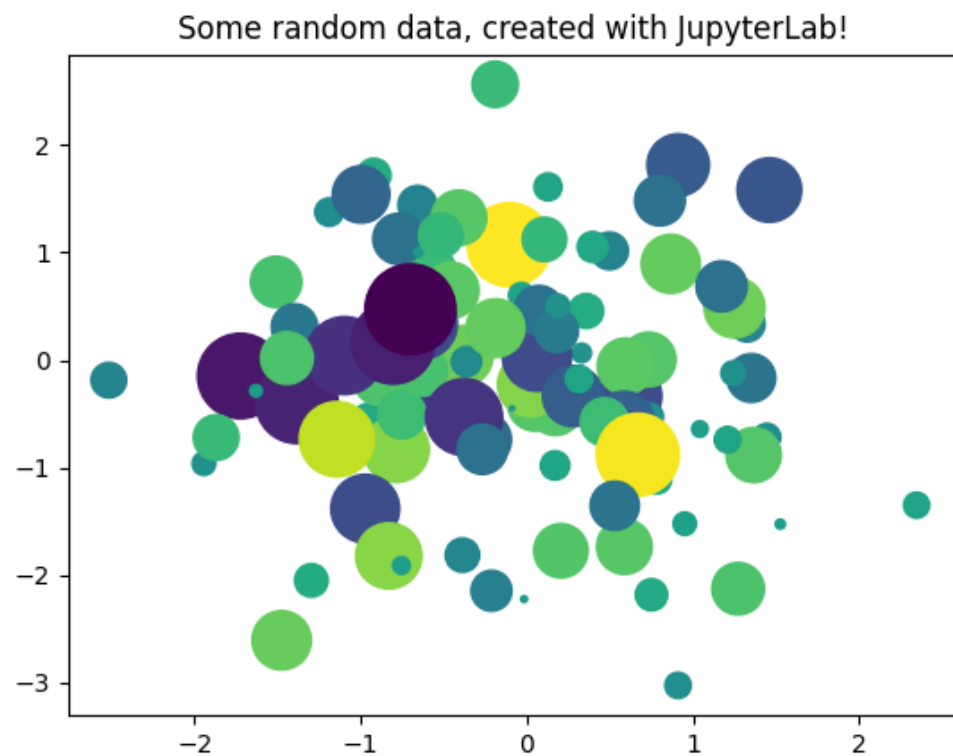


Figura 11. Ejemplo de Jupyter Notebook

Fuente: <https://jupyter.org/try-jupyter/retro/notebooks/?path=notebooks/Intro.ipynb>

3.3.2. Pandas

Pandas es una librería de código abierto de Python, que nos permite hacer operaciones de tratar y manipular los datos de los que disponemos. Por todo ello, con el tiempo se ha hecho muy popular en los ámbitos del aprendizaje automático, o *machine learning* y la ciencia de datos o *data science*.

Un aspecto a destacar son sus dos estructuras de datos principales: *serie* (un array unidimensional) y *dataframe* (array bidimensional de *series*). A través de ellos podemos cargar fácilmente archivos de distintos tipos (en el caso de este proyecto, .json, pero también .csv o .html) y manipularlos (fusionarlos, unirlos, etiquetarlos de distintas maneras...) para poder luego crear visualizaciones de los mismos a nuestra elección. Esta visualización está apoyada por otra librería que utiliza a su vez y que describiremos a continuación, Matplotlib.

3.3.3. Matplotlib

Matplotlib es una librería creada en Python para crear visualizaciones (gráficos). Estos pueden ser tanto estáticos como animados o incluso interactivos, y de diversos tipos, no solo diagramas de barras o histogramas, sino también diagramas de contorno o mapas de color.

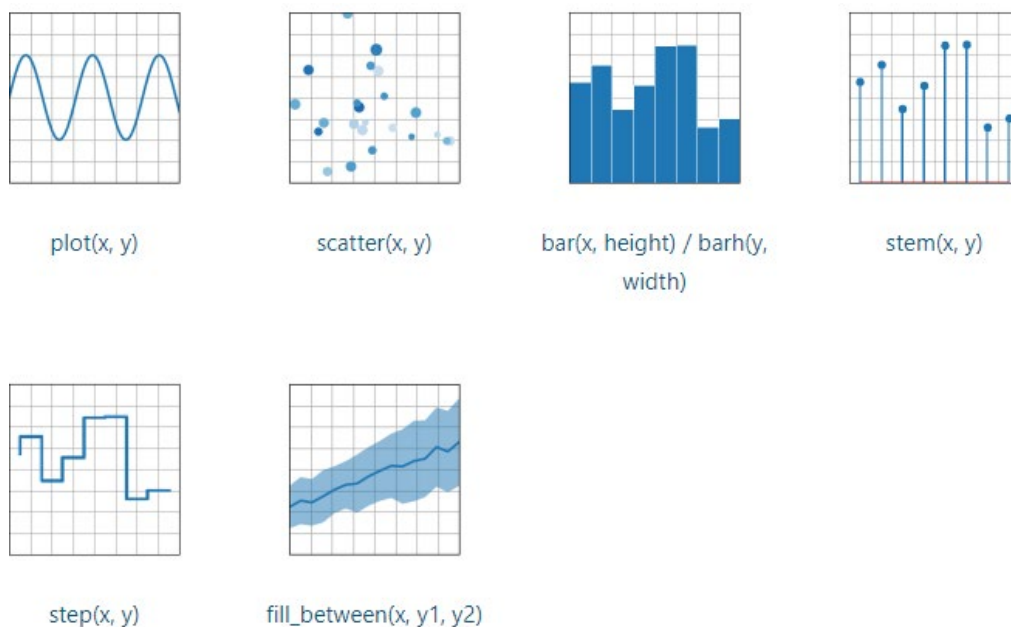


Figura 12. Tipos básicos de diagramas disponibles en Matplotlib

Fuente: https://matplotlib.org/stable/plot_types/index.html

Estos gráficos pueden customizarse de distintas maneras, y el usuario puede llegar a interactuar con ellos. Además, es posible exportarlos, una vez terminados, a distintos formatos.

3.4. Resumen de conclusiones

Así pues, y como resumen, podemos ver que, dada la horizontalidad y el liderazgo compartido de la comunidad educativa #claustrvirtual, es útil el recabar datos y analizarla de manera práctica. Las herramientas existentes no permiten hacerlo de una manera concreta y exhaustiva a día de hoy, pues solo permiten trabajar con datos de los 7 días anteriores o los últimos 500 tuits.

Estamos también ante la oportunidad de, en vez de usar una herramienta genérica, poder examinar primero los datos obtenidos de la API de Twitter y, después, ver cuáles son los aspectos más interesantes a analizar. De esta manera, se da igual peso a la herramienta y al contexto, y se pueden obtener conclusiones más contextualizadas conociendo la situación del docente en España.

Por último, es posible realizar este trabajo pues existen las herramientas adecuadas para llevarlo a cabo. El ecosistema de librerías alrededor de Python nos permite manejar los datos y realizar visualizaciones de los mismos.

4. Diseño de la propuesta

4.1. Objetivos

4.1.1. Objetivo General

Como ya se ha indicado en la introducción, el objetivo general consiste en diseñar un análisis sobre la comunidad educativa #claustrvirtual que permita descubrir nuevas personas y temáticas relacionadas con la misma, además de una idea general sobre el progreso de la comunidad (número de gente que participa, número de tuits, popularidad de los mismos e impacto...).

En este proceso se busca aprender de esta comunidad educativa a través de la tecnología. Precisamente, hoy en día existe un interés muy grande en indagar sobre cómo la tecnología puede ponerse al servicio de la pedagogía y la educación en general. Este es un ejemplo sobre cómo usarla para observar y sacar conclusiones sobre cómo el profesorado aprende a través de las redes sociales.

4.1.2. Objetivos específicos

A continuación, y a partir de los objetivos específicos definidos en la introducción, pasamos a desarrollar los mismos:

4.1.2.1. Aprender y profundizar sobre las maneras de compartir contenido entre el profesorado a través de la lectura y análisis de los mismos

Para hacernos una idea sobre los temas y datos que se manejan en la comunidad educativa, siempre ayuda, en primera instancia, el hacer una revisión de los mismos. Esta revisión se puede hacer de manera informal y no rigurosa, como antesala al análisis, para tener una idea de los datos que pueden ser interesantes a analizar. Además, nos darán pistas para saber los tipos de dinámicas a la hora de compartir contenidos existentes en la comunidad, lo cual nos puede ayudar a la hora de, más tarde, interpretar los resultados analizados.

4.1.2.2. Crear un sistema que interactúe con la API de Twitter para que permita extraer los datos sobre la comunidad durante el tiempo de desarrollo del proyecto.

Una de las partes cruciales del trabajo son los datos obtenidos a través de Twitter. Aunque estos datos podrían obtenerse de forma manual, esto requeriría de un trabajo diario para obtenerlos, el cual es a la vez tedioso y requiere de constancia. Es por eso que se ha optado por hacer la extracción a través de un script en Python, el cual se encarga de extraer los tuits con el *hashtag* #claustrovirtual con 7 días de anterioridad de manera automática.

4.1.2.3. Crear un script en Python a través de Jupyter Notebooks en el cual se pueda volcar la información extraída de la API de Twitter.

Una vez tengamos la información que necesitamos de Twitter, es el momento de analizarla. Sin embargo, primero es necesario establecer una manera de pasar multitud de archivos .json, el formato utilizado para guardar los datos, en el Jupyter Notebook que utilizamos para hacer el análisis. Para esto se ha creado un pequeño script en Python que hará la tarea.

4.1.2.4. Utilizar la librería Pandas para modelar la información extraída y hacer un análisis a partir de los datos obtenidos.

En este punto, se pasa a la cuestión clave del trabajo, en la que, utilizando la librería Pandas de Python, se hace un análisis de los datos obtenidos a través de Twitter. Haciendo un análisis de la información obtenida y los tuits ya estudiados previamente sobre la comunidad, se hace el análisis para obtener una serie de resultados que se presentarán de manera gráfica, y con ello llegar a las conclusiones y el aprendizaje que buscamos.

4.1.2.5. Utilizar la librería Matplotlib para realizar gráficos, cuando sea necesario, que permitan visualizar de manera eficaz los datos analizados.

Para representar los resultados del análisis se utiliza la librería Matplotlib (y pandas, que también utiliza Matplotlib a su vez). Con ello, podemos ver de una manera eficaz y precisa los resultados obtenidos a través del análisis.

4.1.2.6. Redactar y poner de relieve las conclusiones extraídas a través del análisis de la información y su visualización, a modo de resumen e informe.

Una vez terminado el proceso de análisis y representación, se llega al estadio final, en el que se analiza el resultado obtenido y se extraen las conclusiones derivadas del mismo.

4.2. Contenidos

4.2.1. Obtener los datos: Obtención de los datos a través de la API Twitter

4.2.1.1. La API de Twitter : Tuits recientes

En este caso, nuestro objetivo es obtener los últimos tuits que se han tuiteado con el *hashtag* #clastrovirtual. Para ello, y después de buscar en la documentación de la API abierta de Twitter, se ha decidido utilizar el *endpoint* de los tuits recientes, es decir, los últimos que se han publicado, para los últimos días.

En este sentido, existe una restricción de 7 días para extraer todos estos tuits. Así pues, el objetivo será, de una manera recurrente y durante la duración del estudio, recopilar día tras día los tuits que mencionen al #clastrovirtual para luego analizarlos.

El *endpoint* utilizado ha sido:

<https://api.twitter.com/2/tweets/search/recent?query=%23clastrovirtual+-is%3Aretweet+-is%3Aquote>

En lo referente a estos campos, se ha filtrado la búsqueda para que los tuits no sean directamente retuits ni citas (*quotes*). Estos tuits no aportan valor cualitativo al análisis, ya que su contenido es el mismo, y por lo tanto se han decidido descartar del estudio.

Por otro lado, para completar la búsqueda se han utilizado los siguientes parámetros adicionales:

Tabla 1: Parámetros utilizados para obtener la información sobre los tuits #claustrovirtual

Parámetro	Valor	Descripción
start_time	Timestamp (YYYY-MM-DDTHH:mm:ssZ)	El timestamp a partir de la cual empieza la búsqueda de tuits
end_time	Timestamp (YYYY-MM-DDTHH:mm:ssZ)	El timestamp hasta el que se ejecuta la búsqueda de tuits
max_results	100	El número máximo de tuits devueltos por la petición (el máximo es 100)
expansions	attachments.media_keys, attachments.poll_ids, referenced_tweets.id.author_id, geo.place_id, in_reply_to_user_id, referenced_tweets.id.author_id, entities.mentions.username	Lista de campos extra sobre el tuit, relacionados con ids referenciadas al mismo según distintos parámetros (adjuntos, autor, ids de tuits referenciados...)
tweet.fields	attachments, author_id, context_annotations, conversation_id, entities, created_at, geo, in_reply_to_user_id, id, lang, possibly_sensitive, public_metrics, referenced_tweets, reply_settings, source, withheld, text	Valores extra sobre el tuit, que van desde el texto del mismo (<i>text</i>), la lengua utilizada (<i>lang</i>), la aplicación desde la cual se ha utilizado (<i>source</i>), si el tuit está categorizado como sensible o no (<i>possibly_sensitive</i>), si está abierto o no a responder (<i>reply_settings</i>), o los <i>likes</i> , retuits y <i>quotes</i> del tuit (<i>public_metrics</i>)
media.fields	alt_text, duration_ms, height, media_key, public_metrics, width, type, url,	Campos relacionados con los datos sobre el contenido multimedia (<i>media</i>)

	preview_image_url	que los usuarios que han añadido en sus tuits en el rango de tiempo determinado
place.fields	contained_within, country, full_name, id, country_code, geo, name, place_type	Campos relacionados con los datos sobre los lugares (<i>places</i>) que los usuarios que han creado los tuits han añadido en el rango de tiempo determinado
poll.fields	voting_status, options, id, end_datetime, duration_minutes	Campos relacionados con los datos sobre las encuestas (<i>polls</i>) creadas en el rango de tiempo determinado

Como se puede apreciar, son muchos los campos utilizados, y evidentemente, finalmente no se han utilizado todos para el análisis de datos. Al no existir muchas encuestas y lugares (*polls* y *places*), estos datos no han sido utilizados por no considerarse relevantes. Los campos del contenido multimedia (*media*) no se han utilizado apenas, solo el campo *type*), para conocer los tipos de contenido multimedia utilizados dentro de la comunidad.

4.2.1.2. Script: Lambda.py

Para obtener de manera automática todos los tuits de la comunidad de #claustrvirtual, se ha escrito un *script* lambda.py en Python (el cual se puede ver en el Anexo A) usando AWS (Amazon Web Services), que permite hacer una llamada cada día a las 0:05 y guardar todos los tuits con el hashtag que se postearon el día X - 7 (siete días antes del día actual) a través de un *bucket* de S3, un servicio de almacenamiento de objetos en la nube de Amazon.

El *script* define la *url* deseada y configura el token (*Bearer token*) y los *headers* necesarios (*Accept, Authorization*) para que la llamada se pueda hacer con los permisos y configuración necesarias. Después, guarda el resultado en una variable, la cual vuelca en un fichero .json. Finalmente, se crea un *bucket* S3 de AWS y se almacena en el mismo.

Para este proyecto, se han descargado todos los ficheros .json y se han analizado. De esta manera, podemos trabajar de manera cómoda con los datos obtenidos de la API de Twitter de manera automática.

4.2.2. Volcado de información

4.2.2.1. Script de volcado de información

Para volcar la información obtenida en todos nuestros ficheros de tuits en formato .json, se ha escrito otro script (el cual se puede ver en el Anexo B).

Este script abre todos los ficheros ubicados en un directorio local (según el script del anexo, uno elegido a mi elección llamado "eduproject_data", donde están todos los ficheros del *bucket s3*), y uno a uno los abre. Cada vez que se abre un fichero, se obtiene la información deseada (los campos *data*, *users* y *media*), y se pasan a variables creadas con anterioridad para almacenar esta información (en este caso, *df_data_all*, *df_users_all* y *df_media_all*).

Toda esta información se fusiona en una única variable con la que se trabajará (*df_all*).

4.2.3. Análisis de datos

4.2.3.1. Muestra de los datos

Para este estudio, se han utilizado los datos obtenidos en un intervalo de tiempo de dos meses, desde el 09/04/2022 hasta el 09/06/2022.

4.2.3.2. Tuits por día

El número de tuits por día nos ayuda a entender cuándo la comunidad está más activa. En este caso, podemos observar que hay un número más o menos constante de tuits que oscila en una franja de entre 10 y 40 tuits diarios. El número máximo de tuits es 44, y el mínimo es 7. Por lo general, es entre semana cuando existe más actividad en la comunidad, y durante el fin de semana se publica mucho menos contenido. Los momentos de menor actividad son, de hecho, en los días festivos: el último fin de semana de Semana Santa (del 14 al 17 de abril, con 7 tuits el 15 de abril), el puente de Mayo (del fin de semana del 30 de abril y 1 de mayo, con 11 tuits en este día).

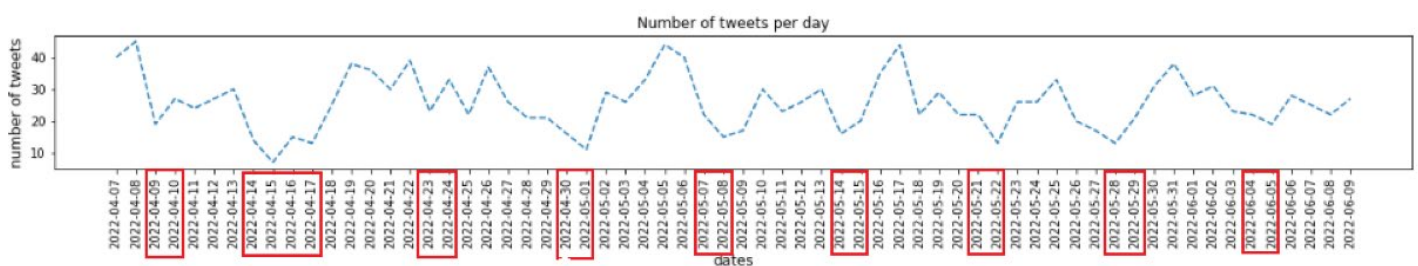


Figura 13: Número de tuits por día

4.2.3.3. Ranking de tuits por persona

En lo referente al número de tuits por persona, se ha hecho un ranking que muestra a las 30 personas más activas en la comunidad durante el lapso de tiempo analizado.

De todas estas personas, se observa que hay una pequeña parte que son muy activas. Las 28 personas con más tuits han publicado al menos 10 de ellos durante el tiempo analizado (recuadros rojos en la figura 12). A partir de este número, y como es normal, se observa un grupo más amplio que publican de manera menos habitual (recuadro azul en la figura 12). Vemos entonces, que hay un fuerte grupo implicado, que impulsa la comunidad, y otro que lo sigue.

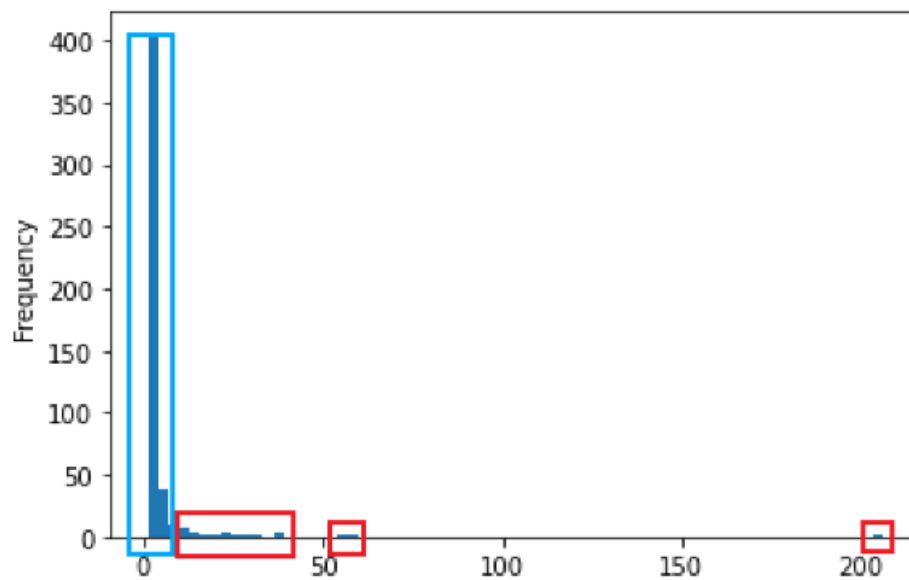


Figura 14: Histograma sobre el número de tuits por persona

Tabla 2: Ranking de tuits por persona

username	name	
llume38	Luis de Mena	206
ladeidiomas	Jimena ESGBDEFRITBRRU #TVTTagTeam 🇵🇷	59
srviruel	Sergio Ruiz Viruel	56
abeorquin	Abe Orquín	39
agr_psp_iplacex	Agrupación Estudiantes Psicopedagogía Online	38
edpuzzle_es	Edpuzzle (español)	38
123eraseunavez	Cuentitis Crónica	38
vicentgadea	Vicent Gadea	33
rosaliarte	Rosa Liarte Alcaine	31
area44cp	Área 44 Servicios Educativos	30
imgende	Ingrid Mosquera	26
EliaZapico	Elia	24
ProfeCarlos67	Carlos Espinoza R. #SoyProfesor - #claustrvirtual	23
manueljesusF	Manuel Jesús Fernández	22
gustavolorentel	gustavolorentel	19
asunflowers	Asun Flores	17
JUNIORResp_es	JUNIOR Esports	17
Hiperdocumentos	#Hiperdocumentos	14

La media de tuits es 3,43 por persona.

Tabla 3: Análisis estadístico de tuits por persona

count	480.00000
mean	3.43125
std	11.13044
min	1.00000
25%	1.00000
50%	1.00000
75%	2.00000
max	206.00000

4.2.3.4. Ranking de retuits por persona

En el caso de los retuits recibidos por persona, la dinámica es parecida a la de los tuits por persona. Aunque cambia el orden, se observa que son las 15 personas que más publican en la comunidad las cuales tienen el mayor número de retuits. El orden cambia por la popularidad de la persona que publica, pues hay ciertas personas que publican más, pero a la vez tienen menos seguidores. Por ejemplo, @llume38 publicó 206 tuits y tiene 269 retuits. @vicentgadea tiene 312, y sin embargo tan solo publicó 33. Esto se puede explicar porque el segundo tiene casi el doble de seguidores que el primero, y por lo tanto sus publicaciones son más susceptibles a ser más populares, y por tanto más retuiteadas.

Además, vemos una distribución parecida a la de los tuits por persona, en la que hay un grupo de personas muy activo y otros participantes que tienen un número muy pequeño (1 o 2 retuits por persona)

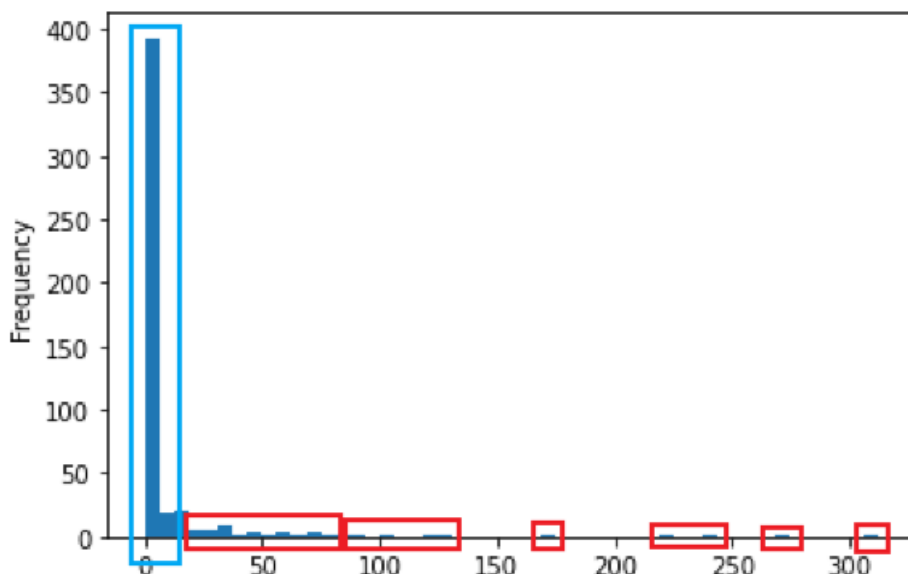


Figura 15: Histograma sobre el número de retuits por persona

4.2.3.5. Ranking de *likes* por persona

Como era de esperar, existe un número mucho mayor de *likes* por persona, que el de tuits o retuits. A pesar de que los números crecen, el histograma es similar a estos dos ya mencionados, por lo que se puede observar que se guarda una coherencia entre tuits, retuits y *likes*.

Tabla 4: Ranking de *likes* recibidos por persona

username	
llume38	1971
gustavolorente1	1700
vicentgadea	936
EliaZapico	861
edpuzzle_es	598
AzaharaTIC	504
ladeidiomas	400
rosaliarte	353
manueljesusF	332
larotesmeyer	304
doloresojeda1	284
lygobu	281
genially_es	271
srviruel	244
imgende	212
Hiperdocumentos	201
clasehistoria79	201

4.2.3.6. Nombre y número de *hashtags* relacionados

Como se puede ver en la Tabla 5, acompañando al *hashtag* #claustrvirtual hay toda una serie de *hashtags* relacionados con el mundo educativo que están a su lado. Muchos de ellos van juntos, como puede observarse con #psicopedagogia (97 menciones) y #pedagogia (91 menciones). Aunque no son lo mismo, se entiende que las personas interesadas en un campo también lo estarán por el otro.

Si analizamos los *hashtags* utilizamos, vemos que la comunidad está interesada por las nuevas tecnologías (#ensenandocontic, #podcast), por metodologías activas de enseñanza/aprendizaje (#abj, el aprendizaje basado por juegos, o #gamificacion), o por compartir recursos (#charlaseducativas o #edurecetas). Además, observamos el interés de algunas empresas por estar presentes en las publicaciones, o su éxito al ser mencionadas en la comunidad (#edpuzzle o #microsoftedu).

Tabla 5: Ranking de apariciones de cada *hashtag*

clastrovirtual	1651
educacion	228
psicopedagogia	97
pedagogia	91
docentes	75
edutwitter	56
mfltwitterati	55
profesores	54
edtech	52
maestros	52
edchats	48
teacher5oclockclub	48
maestro	48
docente	47
profesor	45
teachers	42
edufis	40
abj	37
ensenandocontic	37
edurecetas	36
hashtags	35
unitedteachers	34
librosrecomendados	34
edpuzzle	34
microsoftedu	31
charlaseducativas	31
podcast	29
inclusioneducativa	29
gamificacion	28
	27

4.2.3.7. Los tuits con más retuits

Cuando analizamos los tuits más populares entre la comunidad, se observa que hay un gran interés a la hora de aprender sobre nuevas herramientas, y más específicamente, los tuits más populares tienen relación con herramientas digitales. Si observamos los 10 tuits que más retuits han tenido, vemos que la totalidad de los mismos tienen esta finalidad.



Figura 16: El tuit con más retuits de #claustrvirtual

Fuente: <https://twitter.com/AzaharaTIC/status/1516291004109885443>

La comunidad entiende que este tipo de herramientas ayudan mucho en el día a día, y es por eso que son al final estos tuits los que se expanden con más fuerza entre la misma.

Tabla 6: Ranking de tuits con más retuits

	username	retweet_count	text
6246	AzaharaTIC	89	Herramienta digital para crear vídeos con voz realista a partir de texto. #VÍdeo #Audio #Claustrvirtual 🗣️🎧📺 https://t.co/51MCNIG086
676	larotesmeyer	84	Para animar a los opositores. Para levantar el ánimo del #claustrvirtual. Para hacer más llevadera la recta final. ¿Os hace un hilo de #portadasdiscosdocentes? ¡Venga, que luego lo pasamos bien! @envozalta_libro ¿te animas? Si os apetece, enganched en el hilo. ¡Empiezo yo:
3209	Mustriasm	69	Hace un mes pedí ayuda al #ClaustrVirtual con unas dudas sobre Genially. Gracias a @imgende @ProfeDe_ELE_DAF @fernando_marit7 @AlicitaTF y @FDuchement di forma a un taller sobre uso seguro de TIC en 5º y 6º de EP y aquí está para quien lo quiera usar 🙌
20284	vicentgadea	67	Existen multitud de herramientas 🧰 que nos permiten crear recursos digitales 📱 para nuestro alumnado. ¡Hoy os comparto mi "Top 10 apps para crear fichas y actividades interactivas" 📄 https://t.co/Xx3vsQV8q0 ! ¿Cuál es tu favorita? 🤔 https://t.co/DnW01odeKM #edtech
26008	AzaharaTIC	54	Esta divertida web te permite enviar audios de WhatsApp a tus amigos como si fueran Yoda, Homer, Vegeta y más. #Claustrvirtual https://t.co/GLYOY8emD a través de @genbeta
7160	gustavolorente1	50	Estoy de vuelta 🙌 "PONER EL RELOJ EN HORA" Trabajando percepción espacial+temporal+desplaza+colabor. (más grande) lanzamientos de saquitos Relacionado con mates #edufis #familiaedufis #physed #claustrvirtual #PODIUMEF @piEFcitosCaMiNo #school #educacion #educacionfísica https://t.co/4kWm6eLHBC
25598	llume38	49	Yo, intentando acabar el temario. https://t.co/XFGIPoYnDe #claustrvirtual
31778	lygobu	43	#claustrvirtual Recién salido del horno!!! Este año nos propusimos diseñar para las excursiones cuadernillos didácticos. Con ellos pretendemos que las visitas sean activas. La próxima excursión será al Parque de las Ciencias y éste es el resultado. https://t.co/Vy1uK28vOF
31447	gustavolorente1	39	BALLOON HOOP. Este es el último video de #edufis de este curso. Han sido más de 100 videos donde he mostrado ideas, propuestas y opciones que espero hayan podido ayudar a compañer@s. Deseo que así haya sido. #familiaedufis #physed #pe #claustrvirtual #PODIUMEF @piEFcitosCaMiNo https://t.co/cepTYFHJ4y
1496	Hiperdocumentos	38	📄 Hoy toca #HiperDocTic de una de las más recientes incorporaciones de @Google a @googledocs. los #ChipsInteligentes y #EnlacesInteligentes. ¿para qué sirve? Sigue leyendo 📄 https://t.co/YsOU644WhM #googleforedu #claustrvirtual #edtech

4.2.3.8. Los tuits con más likes

Curiosamente, al igual que los tuits que tienen más retuits tratan en su totalidad sobre trucos y herramientas a la hora de usar las nuevas tecnologías o TIC (Tecnologías de la Información y Comunicación), en el caso de los tuits con más likes, el tema a tratar es distinto.

Tabla 7: Ranking de tuits con más likes

	username	like_count	text
20701	llume38	312	Yo, intentando acabar el temario. https://t.co/XFGIPoYnDe #claustrvirtual
6200	gustavolorente1	310	BALLOON HOOP. Este es el último video de #edufis de este curso. Han sido más de 100 videos donde he mostrado ideas, propuestas y opciones que espero hayan podido ayudar a compañer@s. Deseo que así haya sido. #familiaedufis #physed #pe #claustrvirtual #PODIUMEF @piEFcitosCaMiNo https://t.co/cepTYFHJ4y
7831	larotesmeyer	301	Para animar a los opositores. Para levantar el ánimo del #claustrvirtual. Para hacer más llevadera la recta final. ¿Os hace un hilo de #portadasdiscosdocentes? ¡Venga, que luego lo pasamos bien! @envozalta_libro ¿te animas? Si os apetece, enganched en el hilo. ¡Empiezo yo:
28248	lygobu	239	#claustrvirtual Recién salido del horno!!! Este año nos propusimos diseñar para las excursiones cuadernillos didácticos. Con ellos pretendemos que las visitas sean activas. La próxima excursión será al Parque de las Ciencias y éste es el resultado. https://t.co/Vy1uK28vOF
16924	AzaharaTIC	220	Herramienta digital para crear vídeos con voz realista a partir de texto. #VÍdeo #Audio #Claustrvirtual 🗣️🎧📺 https://t.co/51MCNIG086
23027	AzaharaTIC	210	Esta divertida web te permite enviar audios de WhatsApp a tus amigos como si fueran Yoda, Homer, Vegeta y más. #Claustrvirtual https://t.co/GLYOY8emD a través de @genbeta
7060	gustavolorente1	196	WATER MARKER. Rotuladores de agua. Es una idea que ha salido muy chula. Sencilla, barata y reciclando-reutilizando. Útil para trabajar cualquier contenido de cualquier área. Atrayente para l@s niñ@s. ¡ideas #ideasdeclase #podiumef #claustrvirtual #educacion #escuela #school https://t.co/Cp3RFkcy9
13177	vicentgadea	192	Existen multitud de herramientas 🧰 que nos permiten crear recursos digitales 📱 para nuestro alumnado. ¡Hoy os comparto mi "Top 10 apps para crear fichas y actividades interactivas" 📄 https://t.co/Xx3vsQV8q0 ! ¿Cuál es tu favorita? 🤔 https://t.co/DnW01odeKM #edtech
2923	LAURADULCINEA	161	Lo último: Se van los chiquititos de excursión y prefiero que vayan 2 profes por si acaso. ¡nPues tiene que pagar uno! 😂🙄 ¡Queridos...ha llegado el siguiente nivel: ya no solo "vamos gratis" sino que tenemos que pagar! https://t.co/51MCNIG086 #claustrvirtual
27196	profdhistoria	159	Els docents sabem que s'acosta final de curs quan trobem correus de les famílies molt preocupades perquè els seus fills (que han estudiat molt) no han aprovat l'últim examen. Després que portem 8 mesos dient que el nen o la nena de torn no està totent res. #claustrvirtual

En este caso, y aunque también encontremos tuits de estas características, estamos hablando más bien de mensajes de ánimo, experiencias con los alumnos, o ideas para utilizar en clase que no tienen por qué ver con la tecnología. De hecho, el segundo tuit con más likes es una broma:



Figura 17: El tuit con más likes de #claustrovirtual

Fuente: <https://twitter.com/llume38/status/1528973120027213825>

Esto nos hace ver que, a la vez que la comunidad se preocupa por que se divulgue el conocimiento práctico sobre herramientas y recursos, estos no dejan de apoyar y gustar del contenido más ameno y entretenido. No estamos hablando de un grupo de personas que utilizan el *hashtag* de manera práctica, sino que se trata de una verdadera comunidad de gente que charlan también de manera distendida sobre lo que les divierte y también sobre lo que les preocupa.

Por último, el histograma de frecuencias está alineado con el de retuits.

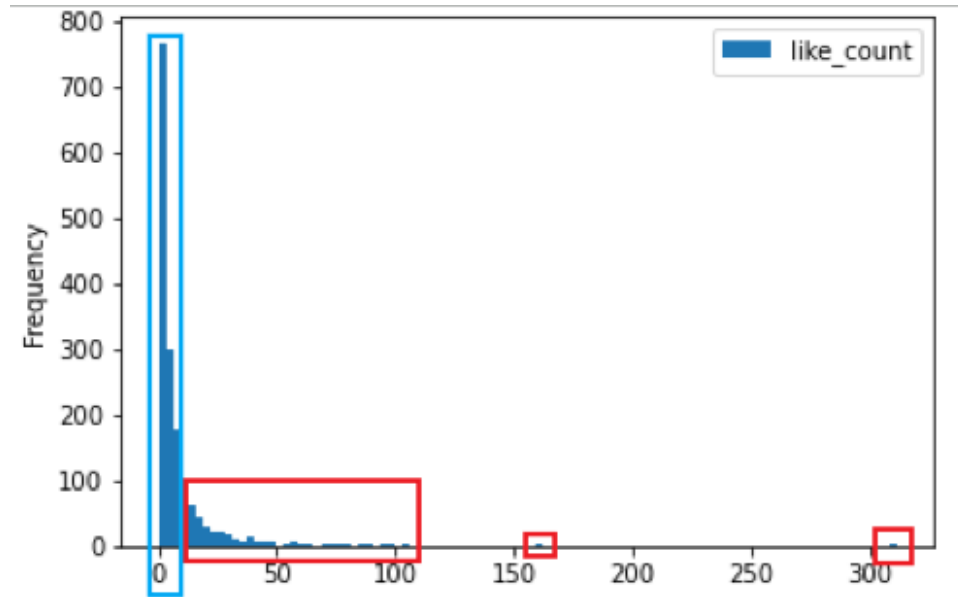


Figura 18: Número de likes por cada tuit. El eje X representa el número de tuits, y el eje Y la frecuencia de likes que ha tenido cada uno de los tuits. Hay muchos tuits que han recibido 1 o 2 likes y muchos menos que han recibido 20 o 50, por ejemplo.

4.2.3.9. Los tuits con más respuestas

Finalmente, los tuits con más respuestas van en la línea de los tuits con más likes.

Tabla 8: Ranking de tuits con más respuestas

	username	reply_count	text
676	larotesmeyer	110	Para animar a los opositores.\nPara levantar el ánimo del #clastrovirtual.\nPara hacer más llevadera la recta final.\n¿Os hace un hilo de #portadasdiscosdocentes?\n¡Venga, que luego lo pasamos bien!\n@envozalta_libro ¿te animas?\nSi os apetece, enganched en el hilo.\nEmpiezo yo:
29332	sruedapascual	37	A ver si me puede ayudar el #clastrovirtual o la magia de Twitter.\nHijo cumple 10 años y le encantan las matemáticas.\nMe pide para su cumpleaños un libro con problemas.\nYa tiene "Hasta el infinito y más allá" de mi querida @ClaraGrima\n¿Alguna sugerencia?
31778	lygobu	25	#clastrovirtual Recién salido del horno!!! Este año nos propusimos diseñar para las excursiones cuadernillos didácticos. Con ellos pretendemos que las visitas sean activas.\nLa próxima excursión será al Parque de las Ciencias y éste es el resultado. https://t.co/Vy1uK28vOF
30138	LAURADULCINEA	25	Lo último: Se van los chiquititos de excursión y prefiero que vayan 2 profes por si acaso.\nPues tiene que pagar uno! 😊😊\nQueridos... ha llegado el siguiente nivel: ya no solo "vamos gratis" sino que tenemos que pagar!\n#clastrovirtual
3209	Mustriasm	19	Hace un mes pedí ayuda al #ClastroVirtual con unas dudas sobre Genially, Gracias a @imgende @ProfeDe_ELE_DAF @fernando_marti7 @AlicitaTF y @PDuchement di forma a un taller sobre uso seguro de TIC en 5º y 6º de EP y aquí está para quien lo quiera usar! 📄
21797	isicollados	19	Pregunta importante #clastrovirtual\n¿Utilizáis horas lectivas con alumnos en clase para corregir exámenes de otras materias (mientras los alumnos realizan alguna otra tarea)?...me gustaría saber los motivos de vuestra elección. ¡Gracias! 😊\n\n@imgende (me ayudas 🙏)
21214	AlbaCandelaR	17	Pregunta para mi querido #clastrovirtual o quien pueda ayudar. ¿Puede una persona hacer bullying sin ser consciente de ello? ¿E, incluso, sin q los q están alrededor y/o lo sufren lo sean?\nLlevo unos meses con un runrún... pero no acabo de dar con la tecla. \n@imgende @PDuchement
14858	ladeidiomas	17	It was my birthday yesterday. Thank you for your messages #mftwitterati #clastrovirtual https://t.co/TzPLBs2qCP
29819	MateOpositora	17	Toca empezar con una de mis áreas menos favoritas: la estadística 😊 ¿Alguna idea para el tema en el escaso tiempo que nos queda? #clastrovirtual\n\nEmpiezo, por supuesto, revisando los hilos de @pbeltranp, pero cualquier cosa que podáis aportar es bienvenida.
31447	gustavolorente1	16	BALLOON HOOP. Este es el último video de #edufis de este curso. Han sido más de 100 videos donde he mostrado ideas, propuestas y opciones que espero hayan podido ayudar a compañer@s. Deseo que así haya sido. #familiaedufis #physed #pe #clastrovirtual #PODIUMEF @piEFcitosCaMiNo https://t.co/cepTYFHJ4y
25788	profedeoz	16	¡Nesesito ayuda! En Word TODA mi PD está en "Arial 11", pero al convertir a .pdf y ver sus propiedades aparece esto (cuadrado rojo) y no consigo identificar dónde está. ¿Qué puedo hacer? #clastrovirtual @imgende https://t.co/9JUvYmW9QL

Su propósito, más que práctico o de información, es el de crear comunidad, buen ambiente, y compartir sin mayores pretensiones. Así lo muestra el tuit analizado con más respuestas:



Figura 19: Tuit con más respuestas

Fuente: <https://twitter.com/larotesmeyer/status/1534554592003899393>

Otra de las modalidades de este tipo de tuits es la de pedir ayuda respecto a ciertos temas. Por ejemplo, se pide ayuda sobre aspectos técnicos a la hora de utilizar nuevas tecnologías (o no tan nuevas, como el paquete MS Office), recomendaciones de libros educativos, inspiración a la hora de tomar decisiones sobre la gestión del centro (por ejemplo, cómo se están implementando las metodologías activas, la eliminación de libros de texto...).

Finalmente, lo que se puede observar es que estos tuits comparten sobre todo el aspecto del debate, lo que genera muchas respuestas. Muchas veces son temas controvertidos en los que la comunidad participa como “termómetro” de lo que se vive en las aulas.

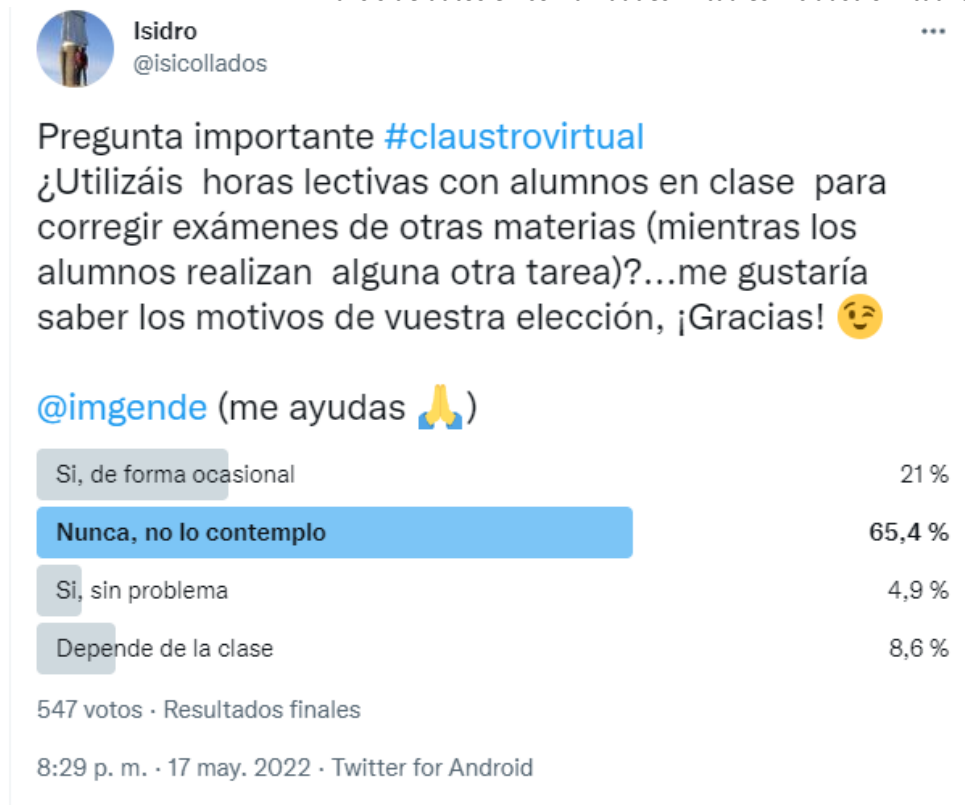


Figura 20: Encuesta con respuestas

Fuente: <https://twitter.com/isicollados/status/1526630962892881920>

Roser @inici_de_cantic · 18 may.
 En respuesta a @isicollados y @imgende
 Más q "no lo contemplo" es "no me da tiempo" (q rima y poco más). Cuando los alumnos están trabajando suelo pasear entre las mesas para apoyar o tomar nota de su aprendizaje.

1 comentario · 1 me gusta

Isidro @isicollados · 18 may.
 Genial, ¡Gracias!

Iván G. Lemos @Me_apunto · 18 may.
 En respuesta a @isicollados y @imgende
 La única vez que lo he hecho es si he tenido que hacer alguno suelto, porque el día D no pudo venir y lo hace unos días después. Si ese día están trabajando en silencio, a ese alumno se lo corrijo para que compruebe cómo le ha salido

1 comentario · 1 me gusta

4.2.3.10. Lista de sitios web más mencionados en los tuits

Un aspecto interesante a analizar son los sitios web utilizados para compartir recursos entre la comunidad educativa.

Para analizar y obtener estos sitios, que no aparecen directamente en la información proveniente de la API, se ha hecho una expresión regular (*regex*) con la cual se han extraído todas las *urls* de los tuits, para así analizarlas y clasificarlas:

Tabla 9: Ranking de URLs compartidas en los tuits de #claustrvirtual

twitter	1200
youtu	55
youtube	50
instagram	31
go.ivoox	30
bit	25
view.genial	18
tweetedtimes	17
vm.tiktok	16
claustrvirtual	12
anchor	10
lnkd	9
rosaliarte	8
l8r	7
paper	5
abeorquin	5
meet.google	5
blog.genial	5
magisnet	5
docs.google	5
t	4
drive.google	4
vicentgadea	4
gedu	4
valverdeando.blogspot	4
teachandbytes	4
genbeta	3
exitoseducativo	3
mdpi	3
open.spotify	3

Como podemos ver en la lista, la mayoría de enlaces son internos, es decir, provienen de la misma plataforma de Twitter. Lo realmente interesante es la variedad de formatos que le siguen. El primero es el vídeo, a través de [YouTube](#)²(youtube +youtu.be, más de 100 menciones). Después le sigue [Instagram](#)³(foto y vídeo, 31 menciones), y el formato audio/podcast ([iVoox](#)⁴, 30 menciones).

² www.youtube.com

³ www.instagram.com

⁴ www.ivoox.com

Otro recurso muy utilizado entre docentes es la plataforma de presentaciones genial.ly⁵, la cual se esfuerza mucho para estar en contacto con ellos. Le sigue la ya conocida [Tiktok](https://www.tiktok.com/es)⁶, que pese a ser utilizada por los más jóvenes, precisamente la usa el profesorado para llegar a ellos.

Otras webs y recursos a mencionar son:

- [Tweetedtimes](http://www.tweetedtimes.com)⁷: Curación de contenidos a través de Twitter
- [Claustrovirtual](http://www.claustrovirtual.com)⁸: Educlaustr, newsletter semanal sobre educación
- [magisnet](http://www.magisnet.com)⁹: Periódico sobre magisterio
- [rosaliarte](http://www.rosaliarte.com)¹⁰ o [abeorquin](http://www.abeorquin.com)¹¹: Webs personales

4.2.3.11. Personas con más menciones

Las personas con más menciones vienen a ser, de alguna manera, las personas más influyentes en la comunidad, pues muchas veces los miembros de la misma los mencionan para que sus tuits tengan más visibilidad y lleguen a más gente.



Figura 21: Tuit buscando visibilidad

Fuente: <https://twitter.com/Patrysa10/status/1534517016677302273>

⁵ www.genial.ly

⁶ www.tiktok.com/es

⁷ www.tweetedtimes.com

⁸ www.claustrovirtual.com

⁹ www.magisnet.com

¹⁰ www.rosaliarte.com

¹¹ www.abeorquin.com

Esta es la lista de los perfiles más mencionados:

Tabla 10: Ranking de personas más mencionadas

imgende	200
fernando_marti7	48
ProfeCarlos67	42
isaacguerrerop	32
daidsantos_a	30
serendipium	29
edpuzzle_es	28
GEGSpain	27
flipped_primary	22
genially_es	21
kikeguerrerot	20
educacyl	19
piEFcitosCaMiNo	18
rosaliarte	18
Mercurio_Juegos	18
EABE2020	17
BeatrizCerdan	16

Destaca en esta lista el perfil de @imgende, Ingrid Mosquera, la creadora de las llamadas #charlaseducativas, que han reunido a muchos y distintos perfiles educativos en su canal para charlar sobre educación. Además, es una persona muy mencionada ya que comparte de manera proactiva muchos de los contenidos de #clastrovirtual, y tiene mucha visibilidad en la red social (a día de hoy, más de 30.000 seguidores).

Otros perfiles mencionados son:

- Profesores influyentes: @fernando_marti7 o @profecarlos67
- Instituciones/Otras comunidades: @educacyl o @GEGSpain
- Empresas / APPs: @genially_es o @flipped_primary

4.2.3.12. Recursos multimedia más utilizados

En la comunidad #clastrovirtual, los tipos de archivos multimedia que se han publicado han sido de tres tipos:

- Foto (961)
- Vídeo (141)
- Gif animado (73)

Tabla 11: Ranking de recursos multimedia más utilizados

photo	961
video	141
animated_gif	73

Si tenemos en cuenta que hemos analizado una muestra de 1647 tuits, esto quiere decir que más de dos tercios de los tuits tienen contenido multimedia. Esto es así porque hoy en día prácticamente todos los tuits van enriquecidos con imagen y vídeo. Los gifs animados se utilizan normalmente para aportar amenidad y diversión, en un ambiente de compartir más distendido.



Figura 22: Tuit con gif animado

Fuente:

https://twitter.com/para_profes/status/1394336041050550272

4.2.3.13. Lenguas utilizadas en los tuits

Como se podía prever, la lengua más utilizada a la hora de publicar dentro de la comunidad es el español.

Tabla 12: Ranking de lenguas más utilizadas en los tuits

es	1458
en	84
und	80
ca	13
pt	8
qme	1
fi	1
tl	1
ht	1

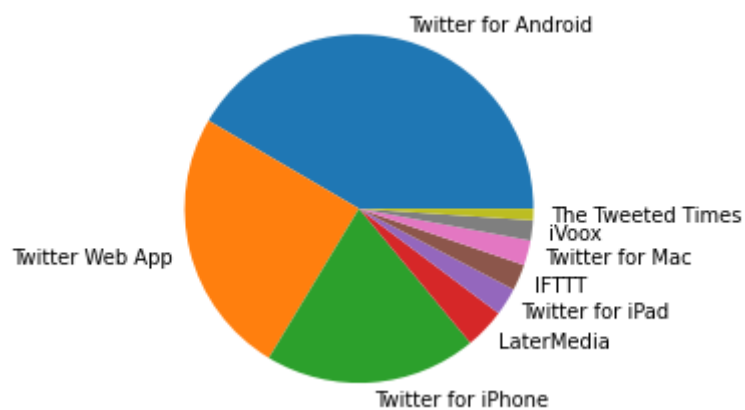
La práctica totalidad de los tuits provienen del perfil @ladeidiomas. Solo 8 personas han postado tuits en inglés, lo cual es un número bajo de personas. Respecto al resto de idiomas, salvando el catalán (ca), del que solo hay 13 tuits, el resto son tuits con hashtags que a nivel de lengua no tienen relevancia, y que sin embargo han sido tratados como provenientes de otras lenguas.

4.2.3.14. Apps utilizadas para enviar los tuits

Observando los datos, podemos ver que más de la mitad de los tuits se publican desde el móvil (un 60%). De ellos, dos tercias partes son publicadas a través de dispositivos Android, y el resto desde iPhone. Un 24% se publica directamente desde la web (PC). El resto de aplicaciones y dispositivos usados son aplicaciones como IFTTT (publica tuits de manera automatizada), iVoox (de podcast), o Hootsuite (gestor de redes sociales).

Tabla 13: Ranking de aplicaciones más utilizadas para enviar tuits

Twitter for Android	666
Twitter Web App	397
Twitter for iPhone	315
LaterMedia	60
Twitter for iPad	41
IFTTT	39
Twitter for Mac	37
iVoox	30
The Tweeted Times	17
Hootsuite Inc.	11
Metricool	10
Twitter Media Studio	7
Paper.li	6
LinkedIn	4
TweetDeck	3
Tweetbot for iOS	2
Twitter for Advertisers	1
Sprinklr Publishing	1

*Figura 23: Gráfico de sectores con las apps más utilizadas*

4.2.3.15. Número de tuits con “contenido sensible”

Twitter tiene un sistema interno que determina si los tuits pueden tener contenido sensible a través del campo booleano “*possibly_sensitive*”. Este campo pasa a activarse como TRUE si el tuit tiene un link que Twitter considera que puede dañar la sensibilidad de la persona que lo lee.

En el contexto de nuestro análisis, solo encontramos dos tuits potencialmente sensibles (*possibly_sensitive* = TRUE)

Tabla 14: Número de tuits con contenido sensible

False	1645
True	2

Analizando los dos tuits en cuestión, vemos que, aunque Twitter los considere potencialmente sensibles, no lo son en realidad. Uno de ellos es un correo en el que una persona le agradece su trabajo al profesor.

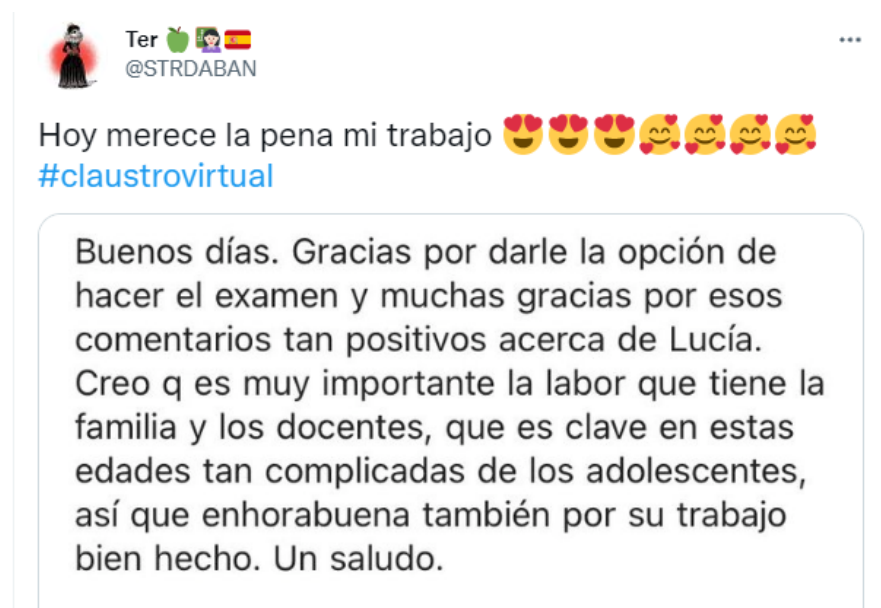


Figura 24: Tuit calificado por Twitter como “posible contenido sensible”

Fuente: <https://twitter.com/STRDABAN/status/1523667169023692800>

El otro tiene que ver con un evento en La Palma:



Figura 25: Tuit calificado por Twitter como “posible contenido sensible”

Fuente: <https://twitter.com/jarmapd/status/1534529868309872643>

De todas maneras, no parece que una comunidad como la de #claustrovirtual, la cual es en positivo y sobre el entorno educativo, pueda contener material ofensivo para el público.

4.2.3.16. Número de tuits abiertos a la comunidad

El campo *reply_settings* dentro de la información del tuit nos habla de las personas que pueden responder al tuit en cuestión. Hay tres posibilidades: *everyone* (todo el mundo), *mentionedUsers* (los perfiles mencionados en el tuit) y *following* (las personas que siguen a la persona que ha publicado el tuit). Se entiende que, si la gente publica usando el *hashtag* #claustrovirtual, normalmente estará usando este campo para que todo el mundo pueda responder (*reply_settings* = *everyone*).

Analizando los datos, podemos confirmar que esto es así. Solo un 0,46% de los tuits se publican para que solo los seguidores respondan (en total, 7).

Tabla 15: Tuits abiertos solo al público o a seguidores

<i>everyone</i>	1639
<i>following</i>	8

De hecho, estos 7 tuits provienen de una usuaria (@rocioqnr) y la persona que la contestó (@imgende), así que podemos confirmar que la comunidad se comporta de una manera abierta.

4.3. Metodología

4.3.1. El marco de trabajo SCRUM

A la hora de definir la metodología a utilizar, he optado por seguir el marco de procesos (*framework*) SCRUM, ya que estoy habituado al mismo y me ha dado buenos resultados en mi carrera profesional. El primer paso es, entonces, y aunque ya sea popular en el entorno del desarrollo de software, definir SCRUM y sus posibilidades.

SCRUM se define como un marco de procesos que se usa para organizar el desarrollo del producto y otras obras que reporten conocimiento. (Alliance, 2017). SCRUM es empírico en tanto que, cuando se sigue de manera correcta, provee los medios para los equipos para que establezcan hipótesis sobre cómo piensan que algo funciona, para después probarlo, reflexionar sobre la experiencia y hacer, en base a ello, los ajustes apropiados para mejorar la manera de trabajar en el proyecto.

SCRUM está estructurado de una manera que permite a los equipos incorporar prácticas de otros marcos de procesos (*frameworks*) que tengan sentido en el contexto del equipo.

En este sentido, y en relación con el párrafo anterior, se considera que SCRUM es apto, por su flexibilidad, a un trabajo de estas características, pues se adapta a equipos de distintas características y número de participantes.

4.3.2. Uso de SCRUM

4.3.2.1. Roles dentro de SCRUM

Desarrollador

Los desarrolladores son las personas dentro del equipo de SCRUM que crean la aplicación y cumplen en cada *sprint* (cada periodo iterativo, de tiempo determinado, en el que se realiza el trabajo) con las tareas que se han marcado.

Normalmente, los desarrolladores, además de hacer su trabajo de creación del producto, participan en ciertas reuniones:

- *Sprint Planning*, reunión en la que se crean las tareas para cada *sprint*, creando el *Sprint Backlog*.
- *Daily Scrum*, una reunión diaria en la que informan de sus avances y de las situaciones que les impiden avanzar.
- *Sprint Review*, en el que hacen una demostración de lo conseguido en el *sprint* al Product Owner y otras personas encargadas de la parte comercial o de producto.
- *Retrospective*, en la que discuten con el resto del equipo sobre cómo ha ido el *sprint*, qué ha funcionado, y qué se puede mejorar en la siguiente iteración.

Product Owner

El *Product Owner* (PO) se encarga de que haya un alineamiento entre el equipo de desarrolladores y los equipos comerciales, destinados a la parte de negocio y producto. Esto quiere decir, a fin de cuentas, que todo lo que se construya esté alineado con los objetivos estratégicos del proyecto.

El gran objetivo del *Product Owner* es mantener un buen *backlog* (la lista de próximas tareas a realizar) priorizado, es decir, que las siguientes tareas que se vayan a hacer sean las más importantes y necesarias de hacer en el proyecto.

Por lo tanto, sus tareas más importantes consistirán en estar en contacto continuo con los distintos actores dentro del proyecto y definir los objetivos del mismo, y las prioridades existentes en cada momento. También es importante que todos estos actores (desarrolladores, equipo de negocio, equipo de producto) entiendan el *backlog* y los objetivos de cada *sprint*, así como que las tareas a realizar permitirán cumplir con los objetivos proyectados.

Scrum Master

El *Scrum Master* es el guía y el apoyo del resto del equipo dentro del proceso SCRUM. De manera general, su objetivo principal es el de hacer avanzar el proyecto, de manera que cuando alguno de los miembros del equipo tenga algún contratiempo y se le impida avanzar, él estará allí para ayudar a la persona y permitirle continuar su trabajo. Es un verdadero líder de equipo.

Muchas veces en los proyectos hay interrupciones que no nos permiten avanzar: una persona externa que te pide ayuda en ciertas tareas, una tarea que no es del todo clara, o una reunión que se hace muy pesada para el equipo. El *Scrum Master* se dedica a guiar las reuniones para que sean fluidas y efectivas, a interceder por los miembros del equipo para que se puedan centrar en sus tareas, o a aportar claridad a la hora de realizar las tareas dentro del *sprint*. Estos son algunos ejemplos, pero el *Scrum Master* tiene que encargarse de muchos otros aspectos que impactan en el equipo para que este no pierda su guía y pueda llegar al máximo de su eficiencia.

4.3.2.2. Enfoque incremental e iterativo

Hay que tener en cuenta que SCRUM es una práctica a implementar en equipos. Hacer este trabajo de manera individual es algo distinto, ya que no existen tantos actores como, por ejemplo, en el departamento de desarrollo de una empresa, en la que hay distintos equipos en los que cohabitan distintos *sprints*.

Por todo esto, para este proyecto no tenía sentido decir que se ha hecho SCRUM, y a la vez definirme como desarrollador, *Product Owner* y *Scrum Master*. Es cierto que el tutor aporta cierta visión sobre las prioridades del proyecto y el cómo ir desarrollando las tareas, pero no considero que esto haya tenido suficiente impacto, ya que su rol no implica el volcarse en el proyecto, sino más bien asesorar sobre él para que la persona que lo desarrolle pueda llevarlo a buen puerto.

Lo que sí que he hecho, a pesar de todo esto, ha sido guiarme por las prácticas anteriormente descritas, para así moldear un modo de trabajar que me permitiera sentirme cómodo y ser eficaz y eficiente a la hora de trabajar. Esto se ha definido en las siguientes prácticas:

- He definido mi tiempo de trabajo en bloques de dos semanas, en los que iba poniendo las tareas que había esbozado en un principio. Antes de comenzar cada bloque, definía con claridad lo que iba a hacer. Por ejemplo, antes de comenzar el análisis de datos, definí todas las métricas y consulté con mi tutor si tenían sentido dentro del encuadre del proyecto.
- También, antes de comenzar cada *sprint* o bloque de trabajo, contrastaba con el tutor en una reunión las tareas que iba a realizar, y cuándo sería la fecha en la que estarían listas.
- Una vez terminaba el *sprint*, compartía con el tutor todo el trabajo, para que lo revisara y pudiera opinar sobre él.
- Cada *sprint* se trabajaba a partir del anterior, es decir, de manera iterativa e incremental. Una vez tuve un entorno de trabajo, creé los scripts para poder tener todos los datos, y ya con los datos de los primeros días fui creando versiones primitivas del análisis que me permitieron ir viendo qué podría analizar y qué no, en vez de esperar a tener todos los datos para realizar el análisis.

4.4. Evaluación

A la hora de evaluar este proyecto, vamos a analizar si la herramienta desarrollada nos ha permitido conseguir nuestro objetivo general, que es el de conocer las personas de la comunidad #claustrovirtual, así como sus motivaciones, y cómo la comunidad está estructurada de manera general.

Para ello, será importante ver primero si hemos podido obtener de manera correcta los datos (en el contexto de este proyecto, una muestra de 1647 tuits). Al haberse hecho esto de manera automática, comprobaremos el resultado comparándolo con una muestra tomada de manera manual, y así podremos ver si los datos han cambiado o no de un método a otro.

Por otro lado, también se va a valorar si la herramienta ha sido efectiva a la hora de valorar las variables, si se han podido sacar conclusiones claras, y si estos hallazgos nos han permitido conocer mejor a la comunidad.




















4.4.1. Consistencia de datos (extracción manual vs automática)

Cuando trabajamos en un proyecto sobre análisis de datos, como es el caso, hemos de evaluar los datos extraídos, ya que, sin unos datos correctos y consistentes, las conclusiones extraídas de los mismos corren el riesgo de no ser reales.

Además, en nuestro caso se ha desarrollado un script en Python que ha permitido obtener estos datos de manera automática. Se va a valorar que hemos podido obtener los datos (no faltan días, no ha fallado el script) y que los datos obtenidos de manera automática son similares que de manera manual.

4.4.1.1. Obtención de datos

Analizando los datos obtenidos a través del script lambda.py (Anexo A), podemos observar que hemos conseguido recabar datos sobre todos los días en los que hemos analizado la comunidad #claustrvirtual para nuestra muestra.

Specified objects					
<input type="text" value="Find objects by name"/>					
Name	Type	Last modified	Size		
 2022-04-07.json	json	April 14, 2022, 00:52:30 (UTC+02:00)	180.5 KB		
 2022-04-08.json	json	April 14, 2022, 02:05:21 (UTC+02:00)	215.5 KB		
 2022-04-09.json	json	April 15, 2022, 02:05:21 (UTC+02:00)	87.9 KB		
 2022-04-10.json	json	April 16, 2022, 02:05:21 (UTC+02:00)	120.6 KB		
 2022-04-11.json	json	April 17, 2022, 02:05:21 (UTC+02:00)	119.3 KB		
 2022-04-12.json	json	April 18, 2022, 02:05:21 (UTC+02:00)	114.5 KB		
 2022-04-13.json	json	April 19, 2022, 02:05:21 (UTC+02:00)	97.8 KB		
 2022-04-14.json	json	April 20, 2022, 02:05:21 (UTC+02:00)	72.6 KB		
 2022-04-15.json	json	April 21, 2022, 02:05:21 (UTC+02:00)	30.5 KB		
 2022-04-16.json	json	April 22, 2022, 02:05:21 (UTC+02:00)	45.3 KB		
 2022-04-17.json	json	April 23, 2022, 02:05:21 (UTC+02:00)	48.3 KB		
 2022-04-18.json	json	April 24, 2022, 02:05:21 (UTC+02:00)	98.1 KB		
 2022-04-19.json	json	April 25, 2022, 02:05:21 (UTC+02:00)	160.9 KB		
 2022-04-20.json	json	April 26, 2022, 02:05:21 (UTC+02:00)	128.5 KB		
 2022-04-21.json	json	April 27, 2022, 02:05:21 (UTC+02:00)	129.6 KB		
 2022-04-22.json	json	April 28, 2022, 02:05:21 (UTC+02:00)	152.1 KB		
 2022-04-23.json	json	April 29, 2022, 02:05:21 (UTC+02:00)	100.4 KB		
 2022-04-24.json	json	April 30, 2022, 02:05:21 (UTC+02:00)	167.8 KB		
 2022-04-25.json	json	May 1, 2022, 02:05:21 (UTC+02:00)	93.1 KB		

Análisis de datos en comunidades virtuales: #claustrovirtual en Twitter

2022-04-25.json	json	May 1, 2022, 02:05:21 (UTC+02:00)	93.1 KB
2022-04-26.json	json	May 2, 2022, 02:05:21 (UTC+02:00)	148.7 KB
2022-04-27.json	json	May 3, 2022, 02:05:21 (UTC+02:00)	106.1 KB
2022-04-28.json	json	May 4, 2022, 02:05:21 (UTC+02:00)	104.3 KB
2022-04-29.json	json	May 5, 2022, 02:05:21 (UTC+02:00)	88.9 KB
2022-04-30.json	json	May 6, 2022, 02:05:21 (UTC+02:00)	60.2 KB
2022-05-01.json	json	May 7, 2022, 02:05:21 (UTC+02:00)	60.1 KB
2022-05-02.json	json	May 8, 2022, 02:05:21 (UTC+02:00)	128.3 KB
2022-05-03.json	json	May 9, 2022, 02:05:21 (UTC+02:00)	106.9 KB
2022-05-04.json	json	May 10, 2022, 02:05:21 (UTC+02:00)	146.9 KB
2022-05-05.json	json	May 11, 2022, 02:05:22 (UTC+02:00)	217.8 KB
2022-05-06.json	json	May 12, 2022, 02:05:21 (UTC+02:00)	195.3 KB
2022-05-07.json	json	May 13, 2022, 02:05:21 (UTC+02:00)	94.1 KB
2022-05-08.json	json	May 14, 2022, 02:05:21 (UTC+02:00)	66.9 KB
2022-05-09.json	json	May 15, 2022, 02:05:21 (UTC+02:00)	69.9 KB
2022-05-10.json	json	May 16, 2022, 02:05:21 (UTC+02:00)	130.3 KB
2022-05-11.json	json	May 17, 2022, 02:05:21 (UTC+02:00)	97.5 KB
2022-05-12.json	json	May 18, 2022, 02:05:21 (UTC+02:00)	98.3 KB
2022-05-13.json	json	May 19, 2022, 02:05:21 (UTC+02:00)	105.4 KB
2022-05-14.json	json	May 20, 2022, 02:05:21 (UTC+02:00)	57.6 KB
2022-05-15.json	json	May 21, 2022, 02:05:21 (UTC+02:00)	82.3 KB
2022-05-16.json	json	May 22, 2022, 02:05:21 (UTC+02:00)	129.4 KB
2022-05-17.json	json	May 23, 2022, 02:05:21 (UTC+02:00)	184.9 KB
2022-05-18.json	json	May 24, 2022, 02:05:21 (UTC+02:00)	81.6 KB
2022-05-17.json	json	May 23, 2022, 02:05:21 (UTC+02:00)	184.9 KB
2022-05-18.json	json	May 24, 2022, 02:05:21 (UTC+02:00)	81.6 KB
2022-05-19.json	json	May 25, 2022, 02:05:21 (UTC+02:00)	128.6 KB
2022-05-20.json	json	May 26, 2022, 02:05:21 (UTC+02:00)	73.4 KB
2022-05-21.json	json	May 27, 2022, 02:05:21 (UTC+02:00)	103.1 KB
2022-05-22.json	json	May 28, 2022, 02:05:21 (UTC+02:00)	54.9 KB
2022-05-23.json	json	May 29, 2022, 02:05:21 (UTC+02:00)	93.2 KB
2022-05-24.json	json	May 30, 2022, 02:05:21 (UTC+02:00)	93.5 KB
2022-05-25.json	json	May 31, 2022, 02:05:21 (UTC+02:00)	131.9 KB
2022-05-26.json	json	June 1, 2022, 02:05:21 (UTC+02:00)	97.0 KB
2022-05-27.json	json	June 2, 2022, 02:05:21 (UTC+02:00)	63.7 KB
2022-05-28.json	json	June 3, 2022, 02:05:21 (UTC+02:00)	61.4 KB
2022-05-29.json	json	June 4, 2022, 02:05:21 (UTC+02:00)	84.4 KB
2022-05-30.json	json	June 5, 2022, 02:05:21 (UTC+02:00)	121.0 KB
2022-05-31.json	json	June 6, 2022, 02:05:21 (UTC+02:00)	145.6 KB
2022-06-01.json	json	June 7, 2022, 02:05:21 (UTC+02:00)	127.1 KB
2022-06-02.json	json	June 8, 2022, 02:05:21 (UTC+02:00)	140.7 KB
2022-06-03.json	json	June 9, 2022, 02:05:21 (UTC+02:00)	101.3 KB
2022-06-04.json	json	June 10, 2022, 02:05:22 (UTC+02:00)	79.0 KB
2022-06-05.json	json	June 11, 2022, 02:05:21 (UTC+02:00)	92.7 KB
2022-06-06.json	json	June 12, 2022, 02:05:21 (UTC+02:00)	97.1 KB
2022-06-07.json	json	June 13, 2022, 02:05:22 (UTC+02:00)	82.7 KB

Figura 26: Archivos de datos .json sobre los tuits

4.4.1.2. Consistencia de los datos

La segunda parte consiste en comprobar que estos datos obtenidos de manera automática no difieren de los obtenidos de manera manual. Para ello, hemos comparado dos archivos json: uno obtenido de manera automática y otro manual, y los hemos comparado a través de una herramienta online, [JSON Diff](http://www.jsondiff.com)¹².

Los resultados son prácticamente iguales, con mínimas diferencias. Cogiendo una de las últimas referencias, un fichero del 06/06/2022, vemos que han cambiado 16 líneas de 2600.

¹² <http://www.jsondiff.com>

La razón de estos cambios resulta de haber obtenido el fichero de manera manual el día después (07/06/2022) en vez de 7 días antes, como hace el script (05/06/2022). En el caso del script, al haber pasado más días, esto hace que ciertos usuarios lleguen a estos tuits, y, por ejemplo, le den al *like* o al retuit, lo que hace cambiar la información.

1482.	"public_metrics": {	1482.	"public_metrics": {
1483.	"like_count": 311,	1483.	"like_count": 310,
1484.	"quote_count": 4,	1484.	"quote_count": 4,
1485.	"reply_count": 16,	1485.	"reply_count": 16,
1486.	"retweet_count": 39	1486.	"retweet_count": 38
1487.	}	1487.	}

Figura 27: Detalle de diferencias entre los .json obtenidos de manera manual y automática

Por todo esto, se entiende que la mejor opción es la de obtener los datos como se hizo con el script automático, una semana después de la publicación de los tuits, ya que por las restricciones de la API es la fecha máxima para obtenerlos, y a la vez van a estar sujetos a menos cambios en el futuro, pues al haber pasado una semana tras su publicación, es menos probable que los usuarios los vean y hagan *like*, retuit, o *quote*.

4.4.2. Efectividad de la herramienta en el análisis

Una vez sabemos que los datos que hemos obtenido son correctos, es necesario verificar que los resultados obtenidos a través del análisis nos ayudan a conocer mejor nuestra comunidad educativa objetivo. En este sentido, hemos explorado distintas facetas de esta comunidad:

- La progresión de sus publicaciones en el tiempo: durante el periodo de tiempo analizado, ¿vemos si hay una evolución y si se publican más o menos tuits? ¿Cuáles son los días en los que las publicaciones crecen o decrecen?
- La actividad de la comunidad: ¿Cuántas personas publican en la comunidad? ¿Cuántos tuits han publicado durante el tiempo analizado? ¿Hay un grupo que "lidera" la comunidad respecto al resto?
- Temáticas dentro de la comunidad: ¿Cuáles son los *hashtags* que acompañan a #claustrvirtual? ¿Qué temáticas se tratan? ¿De qué maneras se comparte el contenido? ¿Cuáles son los contenidos que gozan de mayor popularidad? ¿Desde qué sitios web se comparte el contenido? ¿Qué tipo de contenido multimedia se comparte?
- Relaciones dentro de la comunidad: ¿Hay un grupo de gente que sobresalen en la comunidad? ¿Actúan como influencia para el resto de la misma? ¿Quiénes reciben más menciones? ¿Aquellos que publican más, son igualmente populares?

- Perfil de la comunidad: ¿Qué lenguas se utilizan a la hora de publicar los tuits? ¿Desde qué aplicaciones publican los usuarios? ¿Se publica de manera abierta al resto de Twitter? ¿Se publica contenido sensible?

La respuesta a estas preguntas se puede visualizar en la siguiente tabla:

Tabla 16: Tabla de evaluación de parámetros del análisis

Parámetro	¿Qué resuelve?	Análisis
Volumen de publicaciones	<p>¿Vemos si hay una evolución y si se publican más o menos tuits?</p> <p>¿Cuáles son los días en los que las publicaciones crecen o decrecen?</p>	4.2.3.1. Análisis por día
Actividad de la comunidad	<p>¿Cuántas personas publican en la comunidad?</p> <p>¿Cuántos tuits han publicado durante el tiempo analizado?</p> <p>¿Hay un grupo que “lidera” la comunidad respecto al resto?</p>	<p>4.2.3.2. Ranking de tuits por persona</p> <p>4.2.3.3. Ranking de retuits por persona</p> <p>4.2.3.4. Ranking de likes por persona</p>
Temáticas de la comunidad	<p>¿Cuáles son los hashtags que acompañan a #claustrvirtual?</p> <p>¿Qué temáticas se tratan?</p> <p>¿De qué maneras se comparte el contenido?</p> <p>¿Cuáles son los contenidos que gozan de mayor popularidad?</p>	<p>4.2.3.5. Nombre y número de <i>hashtags</i> relacionados</p> <p>4.2.3.6. Los tuits con más retuits</p> <p>4.2.3.7. Los tuits con más likes</p> <p>4.2.3.8. Los tuits con más respuestas</p> <p>4.2.3.9. Webs más mencionadas en los tuits</p>

	<p>¿Desde qué sitios web se comparte el contenido?</p> <p>¿Qué tipo de contenido multimedia se comparte?</p>	4.2.3.11. Recursos multimedia más utilizados
Relaciones dentro de la comunidad	<p>¿Hay un grupo de gente que sobresalen en la comunidad?</p> <p>¿Actúan como influencia para el resto de la misma?</p> <p>¿Quiénes reciben más menciones?</p> <p>¿Aquellos que publican más, son igualmente populares?</p>	4.2.3.10. Personas con más menciones
Perfil de la comunidad	<p>¿Qué lenguas se utilizan a la hora de publicar los tuits?</p> <p>¿Desde qué aplicaciones publican los usuarios?</p> <p>¿Se publica de manera abierta al resto de Twitter?</p> <p>¿Se publica contenido sensible?</p>	<p>4.2.3.12. Lenguas utilizadas en los tuits</p> <p>4.2.3.13. Apps utilizadas para enviar los tuits</p> <p>4.2.3.14. Número de tuits con “contenido sensible”</p> <p>4.2.3.15. Número de tuits abiertos a la comunidad</p>

5. Conclusiones y trabajo futuro

5.1. Conclusiones

5.1.1. Resumen de las conclusiones obtenidas

Desde sus inicios, la comunidad #claustrvirtual se ha definido como una comunidad abierta y en positivo, en tanto que sus fines están asentados en el compartir, ayudarse, mejorar y aprender en el difícil campo de la educación.

Esta comunidad ha ido surgiendo poco a poco y de manera espontánea a través de los años. ¿Cuál es su status quo?

A raíz del análisis realizado, podemos obtener distintas conclusiones:

5.1.1.1. Volumen de publicaciones

A día de hoy y según los datos analizados, el flujo de publicaciones se mantiene constante, manteniéndose en alza los días lectivos y decreciendo los fines de semana y días festivos, momentos en los que los docentes y otros miembros de la comunidad aprovechan para descansar.

5.1.1.2. Actividad de la comunidad

En nuestro análisis, que se ha realizado desde el 09/04/2022 al 09/06/2022, en un tramo de dos meses, se han publicado 1647 tuits, por parte de 480 perfiles distintos.

Se ha observado que hay un grupo de personas con un nivel alto de publicaciones, seguido por otro grupo, mucho más amplio, que solo han publicado de manera muy puntual. Llegando al 50% de los perfiles, estos solo habrían publicado 1 tuit en este periodo de tiempo, y en el 75% solo 2 tuits. Por tanto, dentro de la comunidad hay una pequeña parte muy activa y con mucha influencia, y otra parte grande con mucha menos participación.

5.1.1.3. Temáticas de la comunidad

Dentro de la comunidad, muchas de las temáticas tienen que ver con la pedagogía y psicopedagogía, las nuevas tecnologías, las metodologías activas y por los eventos educativos en los que se comparten herramientas digitales, o se discuten sobre distintos temas. También las empresas relacionadas con la educación tienen su espacio.

Las nuevas tecnologías, con el uso de las TIC dentro del aula, están en auge. La mayoría de los tuits con más retuits tienen que ver con esta temática. Los tuits con más likes tienen también

que ver con experiencias en el aula, o incluso contenido más lúdico, con bromas y chistes. Los tuits con más respuestas suelen ser debates o encuestas relacionadas con la educación.

Por último, a la hora de compartir contenido, se utiliza sobre todo el mismo Twitter, seguido de plataformas como YouTube y Tiktok (vídeo), iVoox (audio, podcast), presentaciones (genially) y webs personales y newsletters (claustrovirtual, rosaliarte).

5.1.1.4. Relaciones dentro de la comunidad

Hay un grupo de personas que sobresalen dentro de la comunidad. En este caso, coincide que por un lado son de los perfiles más activos y que publican tuits con el hashtag #claustrovirtual, y que además gozan de muchos seguidores y likes y retuits en sus contenidos.

Dentro de este grupo, hay personas que publican más tuits, pero a la vez, a causa de tener menos seguidores, gozan de menos popularidad, con lo cual se observa que el número de seguidores es clave dentro de la comunidad a la hora de que los tuits tengan más likes y retuits.

Por último, muchas de las personas con más seguidores actúan de altavoz para la comunidad, pues son constantemente mencionadas en los tuits de usuarios con más popularidad para que su mensaje sea compartido a más gente, y así ocurre muchas veces. Por tanto, se observa cierta relación entre el hecho de tener muchos seguidores y, a la vez, muchas menciones.

5.1.1.5. Perfil de la comunidad

Dentro de la comunidad, la práctica totalidad de los tuits son publicados en español. Los dispositivos utilizados son el móvil, con preferencia de Android sobre iOS, seguida de la aplicación web. Se publica de una manera abierta en Twitter, dejando que cualquier persona responda a los tuits, y el contenido no es nunca sensible. Esto demuestra que la declaración de intenciones de ser una comunidad “abierta, solidaria y constructiva” se cumple.

5.1.2. El proceso de digitalización en las aulas

Dentro de este estudio, y en el ámbito de la ingeniería informática, es importante resaltar el papel de las TIC y de los recursos digitales dentro de la educación, y más aún después de una pandemia como la surgida en marzo de 2020, en la cual todo el profesorado tuvo que vivir un intenso proceso de digitalización.

Como se ha mencionado en el anterior apartado de conclusiones, se está utilizando el entorno abierto y digital de Twitter para compartir multitud de recursos para hacer la vida más fácil y efectiva a la hora de enseñar. Muchas de las prácticas de enseñanza/aprendizaje están

Análisis de datos en comunidades virtuales: #claustrovirtual en Twitter empapadas por las nuevas tecnologías, que se han puesto al servicio del ABP (Aprendizaje Basado en Proyectos), ABJ (Aprendizaje Basado en Juegos), la gamificación del aprendizaje, o el aprendizaje cooperativo.

Además, lo digital ha ayudado mucho en este proceso de compartir. A día de hoy, un docente puede crear sus materiales en su ordenador para imprimirlos y usarlos en el aula. Sin embargo, después puede subirlos a un Google Drive en la nube y compartirlos por Twitter, o compartir su presentación interactiva hecha en Genial.ly. Para terminar, hay muchas de estas figuras activas de la comunidad, además de empresas, que organizan charlas y eventos para conocer más sobre estas nuevas tecnologías. Estos son ejemplos reales de lo que ocurre en la comunidad #claustrovirtual cada mes.

Para terminar, es importante señalar que este apoyo entre miembros de la comunidad educativa (y no solo docentes, sino también orientadores, pedagogos, o padres y madres) se realiza en un entorno virtual como Twitter, donde no solo lo práctico, como compartir una herramienta es posible, sino que también hay un amplio espacio para el debate, el recreo y diversión, y el apoyo emocional.

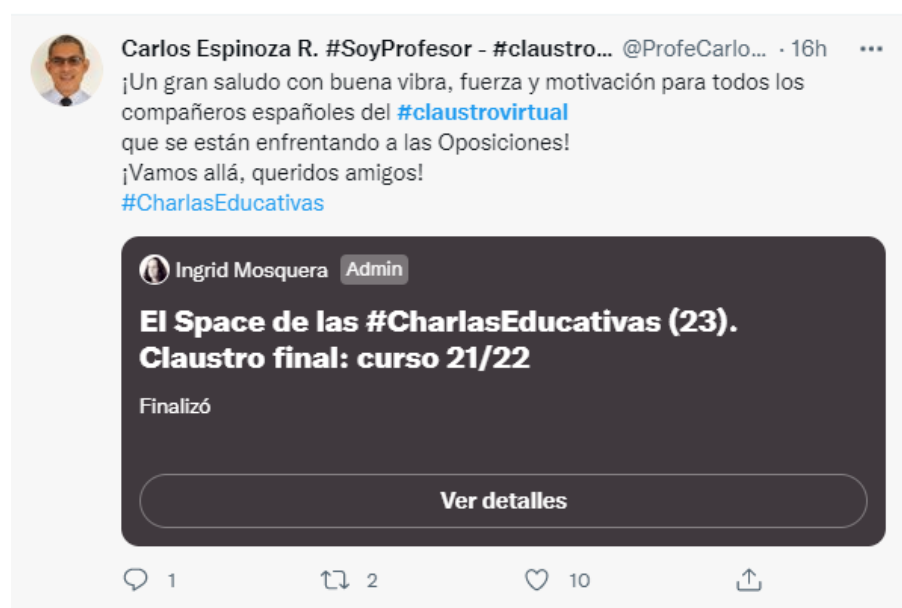


Figura 28: Tuit de apoyo a los miembros de #claustrovirtual

Fuente: <https://twitter.com/ProfeCarlos67/status/1536032647738171392>

5.2. Trabajo futuro

5.2.1. Posibles mejoras en el estudio

Como todo análisis de datos, hay muchas variables que podrían hacer dicho análisis más rico y dotarlo de nuevas conclusiones. Estas son algunas:

- Tiempo de análisis y volumen de datos: Para hacer este análisis se han utilizado datos en un intervalo de tiempo de dos meses, con 1647 tuits analizados. Ampliar estos números ayudaría a verificar las conclusiones obtenidas en este estudio, además de tener más datos para sacar más conclusiones.
- Variables no estudiadas: Aunque se han utilizado una gran parte de las variables en el estudio, hay algunas que no se han usado finalmente. En algunos casos, como la información sobre lugares (*places*) o encuestas (*polls*) porque no existían datos suficientes para sacar conclusiones en los dos meses de datos. Tampoco se han utilizado datos como el número de seguidores, o el recuento de visualizaciones de archivos multimedia, porque habrían hecho el estudio demasiado amplio.
- Nuevas relaciones entre variables: Aunque se han estudiado algunas variables por sí solas, como el número de *likes*, *retuits* y *quotes* recibidos, o las menciones recibidas, sería interesante estudiar con rigor en un futuro trabajo la relación existente entre estas variables. Un ejemplo sería estudiar si hay una relación directa a través de los *retuits* y menciones recibidas respecto a los seguidores de un perfil. Esto podría ser visualizado a través de un diagrama de dispersión.
- Análisis de Redes Sociales (ARS): Además de estudiar las relaciones entre variables, sería interesante estudiar a fondo las relaciones entre cada uno de los perfiles de la comunidad, sobre todo entre aquellos que publican más y tienen, por lo tanto, más peso dentro de la comunidad. Esto se puede hacer con un Análisis de Redes Sociales (ARS), en el que a través de un grafo se puede observar quién menciona a quién. Hay ciertos parámetros, como la centralidad de intermediación (*betweenness centrality* en inglés), que podrían ser estudiados para enriquecer el análisis y conocer mejor cómo se relacionan entre sí los miembros de la comunidad.

Por todo esto, se imagina un trabajo futuro más amplio, con un rango mayor de tiempo y más tuits analizados, en el que variables como el número de seguidores, los lugares, encuestas y contenido multimedia se analicen en profundidad. Además, se podrían estudiar ciertas

Análisis de datos en comunidades virtuales: #claustrovirtual en Twitter relaciones entre variables, como el número de retuits y menciones recibidas comparado con el número de seguidores, y hacer un análisis más visual de las relaciones entre los integrantes de la comunidad con un análisis de redes sociales (ARS).

5.2.2. Más allá de #claustrovirtual

Otra de las maneras de continuar este trabajo sería expandirlo a otras comunidades educativas y de aprendizaje. Una de ellas podría ser #TwitterELE (Español como Lengua Extranjera), donde se comparte contenido sobre esta materia. Sería interesante hacer el mismo estudio y observar si obtenemos las mismas conclusiones que en #claustrovirtual. Otra opción sería estudiar otras comunidades alrededor de un *hashtag* que estén fuera del ámbito de la educación, o que estén dentro, pero en otra geografía, o utilizando distintas lenguas.

Referencias bibliográficas

- Agile Alliance. (2017, Abril 7). What is Scrum? *Agile Alliance* |. <https://www.agilealliance.org/glossary/scrum/>
- Big data: The next frontier for innovation, competition, and productivity* | McKinsey. (s. f.). Recuperado 29 de junio de 2022, de <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- Ferrer, Á., & Gortázar, L. (2021). *Diversidad y libertad: Reducir la segregación escolar respetando la capacidad de elección de centro. EsadeEcPol Insight*(29), 23.
- Friendly, M. (2008). A Brief History of Data Visualization. En C. Chen, W. Härdle, & A. Unwin (Eds.), *Handbook of Data Visualization* (pp. 15-56). Springer. https://doi.org/10.1007/978-3-540-33037-0_2
- Grinin, L., Grinin, A., & Korotayev, A. (2022). COVID-19 pandemic as a trigger for the acceleration of the cybernetic revolution, transition from e-government to e-state, and change in social relations. *Technological Forecasting and Social Change*, 175, 121348. <https://doi.org/10.1016/j.techfore.2021.121348>
- Marcelo, C., & Marcelo, P. (2021). Influencers educativos en Twitter. Análisis de hashtags y estructura relacional. *Comunicar: Revista Científica de Comunicación y Educación*, 29(68), 73-83. <https://doi.org/10.3916/C68-2021-06>
- Menor, M. (s. f.). La formación del profesorado escolar desde los años sesenta: Una historia interminable. *Colecciones informes*, 96(160). <https://fe.ccoo.es/d6de493ed0a1d9bbbc09431f94111b81000063.pdf>
- Mosquera, I. (2022, marzo 24). Cómo los docentes pueden aprovechar el aprendizaje informal en Twitter. *The Conversation*. <https://theconversation.com/como-los-docentes-pueden-aprovechar-el-aprendizaje-informal-en-twitter-178730>
- Ozgun, C., Colliau, T., Rogers, G., & Hughes, Z. (2021). MatLab vs. Python vs. R. *Journal of Data Science*, 15(3), 355-372. [https://doi.org/10.6339/JDS.201707_15\(3\).0001](https://doi.org/10.6339/JDS.201707_15(3).0001)
- Özköse, H., Arı, E. S., & Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data. *Procedia - Social and Behavioral Sciences*, 195, 1042-1050. <https://doi.org/10.1016/j.sbspro.2015.06.147>

- Thinkö. (2020, junio 15). Qué son las metodologías activas y cómo aplicarlas en el aula. *Thinkö education*. <https://thinkoeducation.com/metodologias-activas>
- Alspaugh, S., Zokaei, N., Liu, A., Jin, C., & Hearst, M. A. (2019). Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 22-31. <https://doi.org/10.1109/TVCG.2018.2865040>
- Kandel, S., Paepcke, A., Hellerstein, J. M., & Heer, J. (2012). Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2917-2926. <https://doi.org/10.1109/TVCG.2012.219>
- Tukey, J. W. (1980). We Need Both Exploratory and Confirmatory. *The American Statistician*, 34(1), 23-25. <https://doi.org/10.2307/2682991>
- Wongsuphasawat, K., Liu, Y., & Heer, J. (2019). *Goals, Process, and Challenges of Exploratory Data Analysis: An Interview Study* (arXiv:1911.00568). arXiv. <http://arxiv.org/abs/1911.00568>

Anexo A. Script lambda.py

```
import json
import os
import requests
from requests.structures import CaseInsensitiveDict
import boto3
import datetime

# Env var: token from twitter dev api
# Trigger: setup to run daily at 0:05 and record the tweets of the day
before
# Tweets are collected with the requests library and stored in S3 as
json files (each file corresponds to a single day)

def _format_date(date):
    return datetime.datetime.strftime(date, "%Y-%m-%d")

def lambda_handler(event, context):

    # initial url
    url =
    "https://api.twitter.com/2/tweets/search/recent?query=%23claustrvirtual
+-is%3Aretweet+-
is%3Aquote&tweet.fields=attachments%2Cauthor_id%2Ccontext_annotations%2C
conversation_id%2Centities%2Ccreated_at%2Cgeo%2Cin_reply_to_user_id%2Cid
%2Clang%2Cpossibly_sensitive%2Cpublic_metrics%2Creferenced_tweets%2Crepl
y_settings%2Csource%2Cwithheld%2Ctext&media.fields=alt_text%2Cduration_m
s%2Cheight%2Cmedia_key%2Cpublic_metrics%2Cwidth%2Ctype%2Curl%2Cpreview_i
mage_url&place.fields=contained_within%2Ccountry%2Cfull_name%2Cid%2Ccoun
try_code%2Cgeo%2Cname%2Cplace_type&poll.fields=voting_status%2Coptions%2
Cid%2Cend_datetime%2Cduration_minutes&expansions=attachments.media_keys%
2Cattachments.poll_ids%2Centities.mentions.username%2Cgeo.place_id%2Cin_
reply_to_user_id%2Creferenced_tweets.id.author_id%2Creferenced_tweets.id
%2Cauthor_id&max_results=100" # add the restriction of time
dynamically (gather the tweets from yesterday only)
    today = datetime.datetime.now()
    start_day = today + datetime.timedelta(days=-6)
    end_day = start_day + datetime.timedelta(days=1)
    start_day = _format_date(start_day)
    end_day = _format_date(end_day)
    url +=
    "&start_time={}T00:00:00Z&end_time={}T00:00:00Z".format(start_day,
end_day)

    # authentication for get request
    token = os.environ['TOKEN']
    headers = CaseInsensitiveDict()
    headers["Accept"] = "application/json"
```



```
headers["Authorization"] = "Bearer {}".format(token)

resp = requests.get(url, headers=headers)

assert resp.status_code == 200

# store response in a file
resp_json = json.dumps(resp.json(), indent = 4)
fname = '/tmp/result.json'
with open(fname, 'w') as fh:
    fh.write(resp_json)

# upload file to s3
bucket = 'tweets-claustro-virtual'
s3 = boto3.client('s3')
s3.upload_file(fname, bucket, '{}.json'.format(start_day))
```

Anexo B. Script de volcado de información

```
import pandas as pd
import json
import matplotlib.pyplot as plt
import os
import warnings
import numpy as np
warnings.filterwarnings('ignore')
pd.set_option('display.max_colwidth', None)

directory = os.fsencode("eduproject_data")
df_data_list = []
df_users_list = []
df_media_list = []

for file in os.listdir(directory):
    filename = os.fsdecode(file)
    if filename.endswith(".json"):
        json_file = open(os.path.join(directory.decode("utf-8"),
str(filename)), encoding="utf8")
        data = json.load(json_file)
        df_data = pd.DataFrame(data['data'])
        df_users = pd.DataFrame(data['includes']['users'])
        df_media = pd.DataFrame(data['includes']['media'])
        df_data_list.append(df_data)
        df_users_list.append(df_users)
        df_media_list.append(df_media)

df_data_all = pd.concat(df_data_list, ignore_index=True)
df_users_all = pd.concat(df_users_list, ignore_index=True)
df_media_all = pd.concat(df_media_list, ignore_index=True)

df_all = pd.merge(df_data_all,df_users_all, left_on='author_id',
right_on='id', how='left')
```