

# **Proyecto: Tipificación de usuarios de PlayStation**

**Social Web Behaviour**

**Alumnos:** *Sergi Albiach Caro, Daniel Constantín Birdici & Stéphane Díaz-Alejo León*

# ÍNDICE

<b>1.- INTRODUCCIÓN</b>	<b>3</b>
<b>2.- OBJETIVOS</b>	<b>4</b>
<b>3.- METODOLOGÍA</b>	<b>5</b>
3.1.- CLUSTERING	5
3.2.- ANÁLISIS DE SENTIMIENTOS	5
<b>4.- RESULTADOS</b>	<b>7</b>
4.1.- CLUSTERING	7
4.1.1. - ANÁLISIS DE LOS RESULTADOS	7
4.1.2. - ANÁLISIS DE K-MEANS	10
4.2.- ANÁLISIS DE SENTIMIENTOS	13
<b>5.- CONCLUSIONES</b>	<b>15</b>
<b>6.-DISCUSIÓN</b>	<b>17</b>
<b>ANEXO A</b>	<b>18</b>
<b>ANEXO B</b>	<b>19</b>

# 1.- INTRODUCCIÓN

Los usuarios de la Web 2.0 y sus aplicaciones comparten información constantemente. Las redes sociales son usadas para mostrar sus aficiones y/o gustos, reivindicar o sumarse a una causa, debatir sobre un tema de actualidad, ver lo que publican otros o simplemente realizar lo que comúnmente se conoce ahora como “shitposting”, que es nada más y nada menos que publicar contenido absurdo o de baja calidad.

Todas estas acciones que se realizan dentro de la web nos ayudan a comprender y entender cómo es la persona detrás de la pantalla de cada usuario y también cómo se sentía en cada momento específico. Esta tipificación es sumamente valiosa y útil por diversas razones, cómo pueden ser recomendaciones de contenido, marketing o análisis de consumidores, entre otros.

En la tipificación de este trabajo hemos decidido abogar por la comunidad de usuarios relacionada con PlayStation, la serie de consolas de videojuegos creadas por Sony Interactive Entertainment.

La videoconsola creada por la compañía japonesa está en el mercado desde 1994 y ha conseguido más de 400 millones de ventas en los sucesivos modelos que se han ido sacando desde su nacimiento. La consola nipona es la más popular entre el sector que ocupa, contando actualmente con más de 100 millones de usuarios activos, lo cual convierte a esta comunidad en un grupo social interesante a analizar.

Se tomó la decisión de analizar este segmento de la sociedad porque además de contar con una gran cantidad de usuarios activos, esperábamos encontrar un amplio espectro de perfiles con opiniones muy dispares.

## 2.- OBJETIVOS

Mediante este análisis de Twitter buscamos averiguar un poco más sobre la comunidad de PlayStation. Pretendemos encontrar patrones de comportamiento, ver qué tipo de usuarios existen, qué grado de actividad tienen y qué opiniones ponen de manifiesto respecto a PlayStation y sus diferentes productos. Sería interesante descubrir si esta comunidad está tan polarizada como se cree comúnmente que son las comunidades de videojuegos.

Para la primera parte, el análisis de clusters, nos hemos propuesto los siguientes objetivos:

- Descubrir cuántos grupos podemos obtener y qué características separan unos grupos de otros.
- Determinar si existen correlaciones entre las características de los tweets.
- Ver qué características son más importantes a la hora de explicar la población de cada grupos.

A continuación, para el análisis de sentimientos, nos gustaría averiguar información sobre los siguientes puntos:

- Averiguar la opinión de los miembros de la comunidad sobre los productos de Sony.
- Descubrir el grado de polarización de la comunidad; es decir, si los usuarios suelen tener opiniones muy tajantes o si, por el contrario, no se inclinan fuertemente en ninguna dirección particular.

## 3.- METODOLOGÍA

### 3.1.- CLUSTERING

La primera metodología usada es clustering, en concreto, la técnica no-supervisada llamada *K-means clustering* que realiza la herramienta *XLSTAT*. Se trata de un algoritmo que, dado un número de grupos o clases, intenta minimizar la suma de cuadrados (la diferencia entre un punto y el promedio de la clase al cuadrado) dentro del mismo grupo. El vector de características utilizado en este algoritmo consta de cinco campos:

- *Retweets*: las veces que los tweets de este usuario se han vuelto a publicar por otros usuarios.
- *Favorites*: las veces que los tweets de este usuario han sido añadidos a favoritos.
- *Followers*: los seguidores que tiene la cuenta.
- *Follows*: el número de cuentas seguidas por el usuario.
- *Listed*: el número de listas personalizadas a las que la cuenta de este usuario fue añadido.

Nuestro equipo decidió descargar la información que ha utilizado este algoritmo mediante *Twitter Archiver*, utilizando la siguiente regla:

“#PlayStation OR #PS4 OR #PS5 OR #PSVR OR #PS4share OR #PS5share”

Esta regla simplemente busca aquellos tweets en los que aparezca uno de los *hashtags* mencionados. Se obtuvieron 55.664 tweets pertenecientes al periodo siguiente: 02/23/2021 19:32:18 - 02/24/2021 16:51:51, extrayendo alrededor de 2000 tweets por hora.

### 3.2.- ANÁLISIS DE SENTIMIENTOS

La segunda metodología que hemos utilizado es el análisis de sentimientos mediante la herramienta *Mozdeh*.

Teniendo en mente que nos interesa analizar el texto de los tweets, se ha decidido eliminar de la query original los hashtags #PS4share y #PS5share, que se utilizan al compartir fotos o vídeos de las sesiones de juego. Nos hemos dado cuenta de que en la mayoría de casos no tiene ningún texto asociado, por lo que se incrementa mucho la cantidad de tweets y realmente no podíamos extraer ninguna información de ellos.

Además, como los usuarios genéricos escriben PlayStation de diferentes maneras, también hemos tenido en cuenta esto a la hora de realizar la búsqueda, incluyendo las principales variaciones del nombre.

Finalmente la lengua elegida ha sido inglés ya que la mayoría de publicaciones son en este idioma. Con estos cambios la query final ha sido:

“PlayStation OR playstation OR Playstation OR #Playstation OR #PS4 OR #PS5 OR #PSVR”

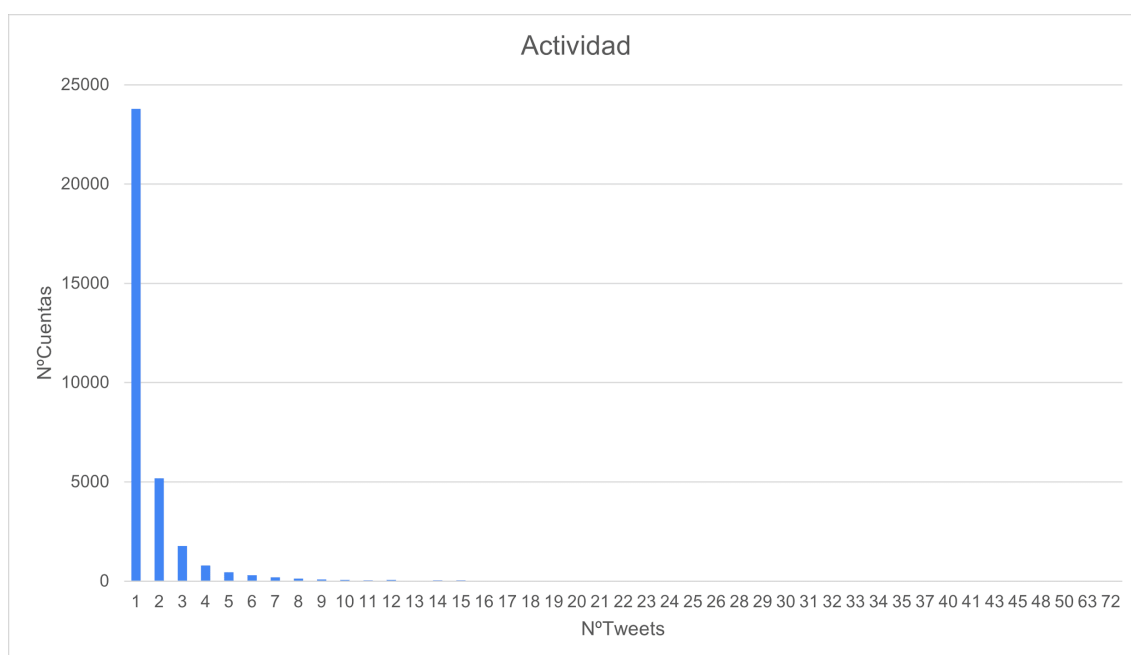
Una vez realizada la consulta se han obtenido unos 8500 tweets pertenecientes al espacio temporal comprendido entre el 10 y el 18 de marzo de 2021. Es un número significativamente menor al obtenido en el apartado anterior, ya que la herramienta utilizada no ha sido la misma y el tiempo de recolección ha sido muy inferior.

Creemos que, a pesar de haber cambiado la consulta y haber recogido menos muestras, los datos obtenidos siguen siendo representativos del conjunto de usuarios de la comunidad, ofreciendo resultados estadísticamente similares al primer dataset.

## 4.- RESULTADOS

### 4.1.- CLUSTERING

Del conjunto de 55.664 tweets que se mencionó en el apartado 3.1 hemos podido analizar la actividad de las 33.017 cuentas únicas dentro del mismo. Los resultados nos arrojan que la gran mayoría de cuentas sólo habían publicado un tweet en el período de extracción, 23.779 cuentas (72% respecto del total), mientras que conforme el número de tweets aumenta, el número de cuentas decrece exponencialmente, cosa que se puede ver en la figura 1. Esto parece indicar que la gran mayoría de usuarios de nuestro conjunto de datos presenta una gran inactividad mientras que son pocas las cuentas que son realmente activas.



**Fig. 1:** *número de cuentas por número de tweets publicados*

#### 4.1.1. - ANÁLISIS DE LOS RESULTADOS

Tras seguir la metodología indicada en el apartado anterior, hemos obtenido la información estadística de las variables que se han seleccionado, figura 1 del anexo A.

Esta información nos puede dar una idea de la distribución de población de cada característica. Por ejemplo, la población de la variable followers, a pesar de tener una cantidad máxima de 2.754.074, debe de tener una enorme cantidad de muestras con valores muy bajos, debido a que la media es tan solo de 1.246,328, menos de la

mitad, y unas pocas muestras con valores muy altos, cómo nos muestra la desviación típica. Por norma general, se puede observar que las poblaciones presentan unas pocas muestras con valores muy altos y el resto bajos, ya que las medias tienden a ser bajas y la desviación típica un orden de magnitud o varios por encima.

Para analizar la naturaleza de los datos hemos realizado un test de correlación entre las variables para ver si existen dependencias entre ellas. Los resultados se pueden observar en la figura 2.

Correlation matrix (Pearson):

Variables	Retweets	Favorites	Followers	Follows	Listed
Retweets	<b>1</b>	<b>0,822</b>	<b>0,309</b>	<b>0,050</b>	<b>0,237</b>
Favorites	<b>0,822</b>	<b>1</b>	<b>0,307</b>	<b>0,054</b>	<b>0,226</b>
Followers	<b>0,309</b>	<b>0,307</b>	<b>1</b>	<b>0,154</b>	<b>0,819</b>
Follows	<b>0,050</b>	<b>0,054</b>	<b>0,154</b>	<b>1</b>	<b>0,179</b>
Listed	<b>0,237</b>	<b>0,226</b>	<b>0,819</b>	<b>0,179</b>	<b>1</b>

Values in bold are different from 0 with a significance level alpha=0,05

**Fig. 2:** correlación entre variables

Los valores obtenidos siguen lo que la lógica nos puede decir en un primer momento. Existe una fuerte relación entre *retweets* y *favorites*, ya que a más *retweets*, el mensaje llega más lejos y es más probable que la gente le de favorite. Por otra parte, existe una fuerte relación entre los followers que se tienen y la pertenencia a listas, lo cual es lógico ya que los perfiles con más followers suelen tratarse de medios de comunicación, divulgación o perfiles verificados de la propia PlayStation, por lo que tienden a estar en listas. Existe también una mínima relación entre los *retweets*, *favorites* y *followers*, también una correlación lógica y esperable, ya que a más followers más probabilidad de que den retweet y favorite.

Con el objetivo de ver qué variables son más significativas, hemos realizado un análisis PCA (*Principal Component Analysis*), en el que intentamos eliminar variables del conjunto original de modo que obtengamos un conjunto más reducido pero manteniendo la mayor parte de la información del grupo de variables original.

Como podemos ver en la gráfica de la figura 3 y la tabla de la figura 4 resultantes de este análisis, con el primer factor podemos explicar hasta el 50% de la variabilidad de los datos. Este factor está compuesto mayoritariamente por las variables *retweets*, *favorites*, *followers* y *listed*, como nos indica la tabla.



Añadiendo el segundo factor, compuesto por las variables *retweets*, *favorites* y *listed*, podemos llegar hasta un 75% de variabilidad más o menos. De este modo, podemos comprobar como la variable *follows*, que representa el número de cuentas seguidas por un usuario concreto, no es muy representativa. Siguiendo una lógica intuitiva, esto tiene sentido ya que el número de cuentas que un usuario sigue no afecta a la popularidad de su cuenta; es decir, no tiene gran impacto en el número de *retweets*, followers o favorites que pueda obtener.

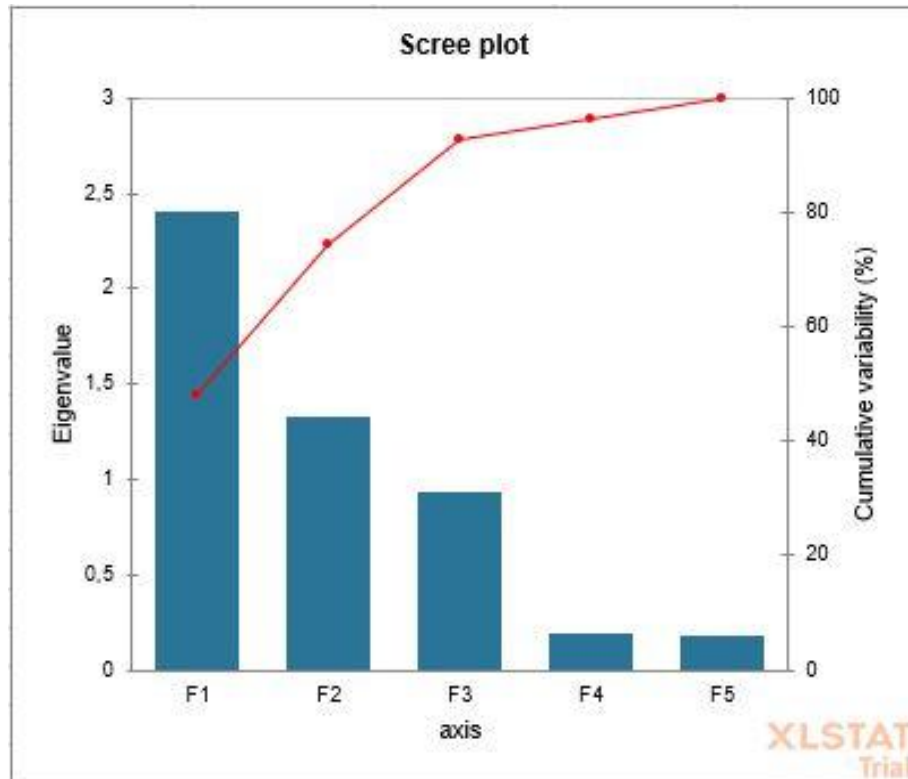


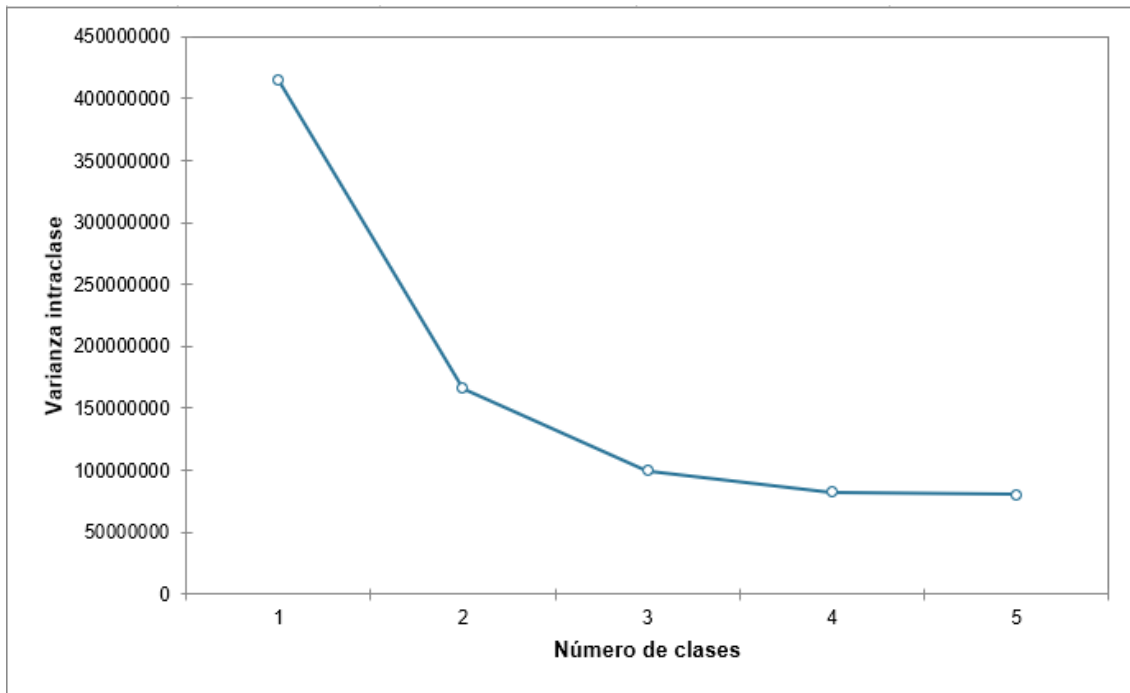
Fig. 3: gráfico con los resultados del análisis PCA

Contribution of the variables (%):					
	F1	F2	F3	F4	F5
Retweets	23,849	25,256	0,597	30,266	20,032
Favorites	23,598	25,691	0,745	24,914	25,053
Followers	26,666	16,691	5,406	23,292	27,945
Follows	2,416	9,051	88,484	0,012	0,038
Listed	23,471	23,312	4,768	21,516	26,933

Fig. 4: tabla con los resultados del análisis de PCA

#### 4.1.2. - ANÁLISIS DE K-MEANS

En los resultados proporcionados por *K-means clustering* se nos presenta una gráfica con la varianza intraclase en base al número de clusters, la cual podemos observar en la figura 5.



**Fig. 5:** número de clases y varianza de cada una de ellas

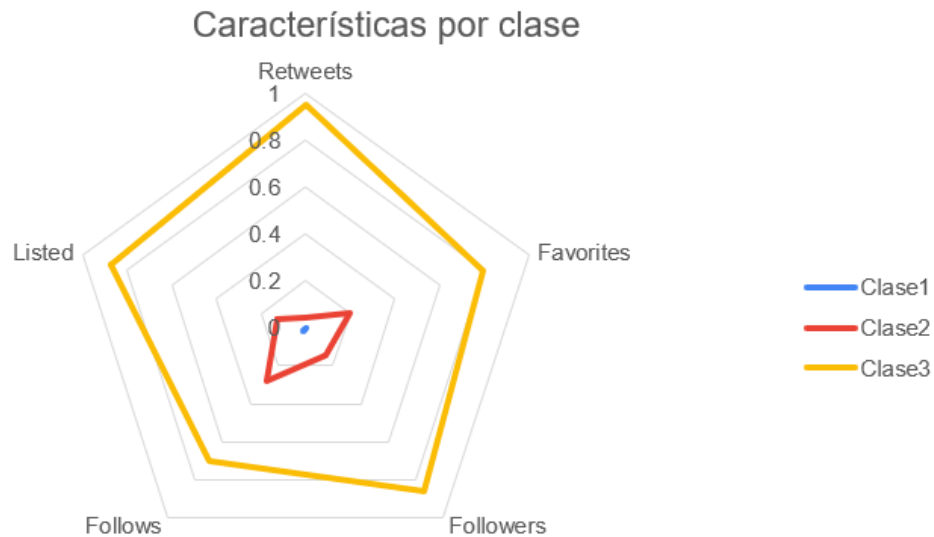
Siguiendo la “regla del codo”, podemos concluir que con estos datos la mejor agrupación serían tres clusters, ya que es el punto en el que la disminución de la varianza ya no es significativa; es decir, no vamos a conseguir una clasificación significativamente mejor utilizando más grupos.

En la tabla de la figura 6 podemos ver los resultados de los tres clusters o clases:

Clase	1	2	3
Objetos	55564	92	9
Suma de los pesos	55564	92	9
Varianza intraclase	18140491.237	11716557031.673	434484813253.139
Distancia mínima al centroide	10.164	5482.366	164646.114
Distancia media al centroide	1242.312	82778.183	487842.228
Distancia máxima al centroide	119417.653	446397.196	1516077.368

**Fig. 6:** tabla con los resultados del clustering

Se puede apreciar una reducción significativa de muestras conforme cambiamos de clase, siendo la primera clase la que más muestras posee y la tercera la que menos. Para poder razonar a qué se debe esto hemos realizado la gráfica de la figura 7 en la que se puede observar los valores normalizados del centroide de cada clase:



**Fig. 7:** representación gráfica de los resultados del clustering

El grupo 1 es el usuario mayoritario. Se trata de cuentas con pocos *retweets*, *favorites*, *followers*, *follows* y *listed*, en comparación con cuentas más grandes. Los usuarios de la clase 2 son usuarios medianos con una contribución moderada sobre el total, tratándose principalmente de cuentas de empresas que se dedican al mundo de los videojuegos, como *Best Buy Canada*. Por último, la clase 3 comprende las cuentas con un mayor número de *followers*, *follows*, *retweets* y *favorites*, es decir, sobre todo, las cuentas oficiales de *PlayStation*, y algunas revistas como *CNET* o *famitsu*.

La tabla de la figura 8 muestra los centroides de cada clase normalizados por característica:

Clase	Retweets	Favorites	Followers	Follows	Listed
1	0.00102653	0.00411926	0.00049287	0.01955169	0.0007631
2	0.04449321	0.19956296	0.14116623	0.28081515	0.12373949
3	0.95448026	0.79631778	0.8583409	0.69963316	0.87549741

**Fig. 8:** tabla con los centroides normalizados de las clases

La característica principal de la clase 1 son los *follows*. Los usuarios de este grupo siguen a muchas cuentas en comparación con sus *followers*. Puesto que no son

cuentas populares, tampoco tienen demasiados *retweets*, *favorites* ni aparecen en listas personalizadas.

Si intentamos clasificarlos por su tipología y en función de su actividad en la red social, una parte de estos usuarios serían mayoritariamente *espectadores* que consumen el contenido que crean los otros dos grupos de esta clasificación. Si bien es cierto que la mayoría de tweets son de este grupo, muchos de ellos son simplemente segmentos de vídeo de las sesiones de juego de los usuarios. Este tipo de contenidos son poco populares y atractivos según las estadísticas que hemos obtenido, lo que nos lleva a pensar que sus creadores no han invertido mucho esfuerzo en intentar que este contenido llegue a una amplia audiencia. Consecuentemente, no podemos clasificarlos como creadores.

La característica representativa del grupo 2 son los *follows* también, seguido de los *favorites* y *followers*. Estas cuentas suelen ser de tiendas de videojuegos, revistas y otros tipos de empresas que tienen más presencia en la red social, por lo que tiene sentido que tengan más *followers* y más *favorites*.

Este grupo es el que está formado verdaderamente por creadores; es decir, empresas y revistas que valoran su presencia en las redes y las utilizan como herramientas para la difusión de contenidos que promuevan sus productos y sus marcas. Es por ello que esta priorización de contenido se ve reflejada en los *ratings* de sus tweets, ya que tienen más *followers* y la gente interactúa de manera activa con ellos.

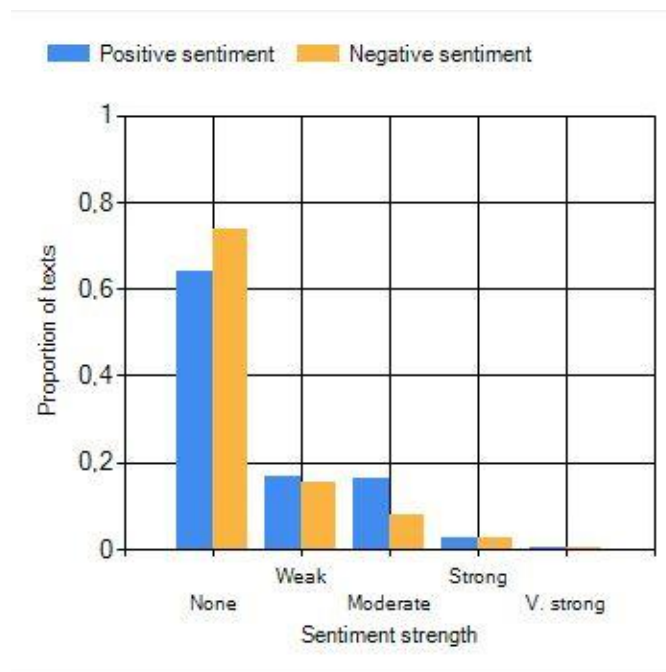
La característica principal del grupo 3 son los *retweets* seguido de los *followers*. Un dato que está en concordancia con lo dicho anteriormente, ya que se trata de cuentas de grandes empresas con una gran cantidad de *followers* que proporcionan muchos *retweets*. También se pueden calificar como creadores, con la diferencia de que los contenidos que publican son mucho más populares y llegan mucho más lejos que aquellos del grupo 2.

Finalmente, un dato insólito es que el grupo 2 y 3, a pesar de estar compuestos por cuentas grandes, también cuentan con la mayor cantidad de *follows* o usuarios seguidos. Si analizamos las cuentas pertenecientes a estos grupos, podremos observar que algunas de ellas disponen de bots que siguen automáticamente a una gran cantidad de cuentas relacionadas con el sector de los videojuegos. Como ejemplo, solo la revista japonesa *famitsu* sigue a alrededor de 45 mil otras cuentas. Por este hecho, podríamos clasificarlas como *moscardones*, al menos en parte, ya que si bien estas cuentas pueden estar manejadas por un equipo de marketing, es muy

probable que dispongan de programas adicionales que les permitan seguir a muchas cuentas de manera rápida y automatizada. Siguiendo en la línea de las tipologías de usuario específicamente para Twitter, también podemos añadir que estas cuentas pertenecen a la clasificación de *massmedia*, debido al elevado número de *followers* y al hecho de que son medios de comunicación dedicados a los videojuegos. El resto de cuentas que se dedican a promocionar la venta de videojuegos serían cuentas *empresariales*, que buscan cumplir sus objetivos de negocios mediante el uso de las redes sociales.

## 4.2.- ANÁLISIS DE SENTIMIENTOS

Con la herramienta Mozdeh, como ya se ha comentado anteriormente, se han analizado aproximadamente unos 8500 tweets desde el día 10 de marzo hasta el 18 de marzo, y se ha obtenido la gráfica de la figura 9 asociada al análisis de sentimientos.



**Fig. 9:** clasificación de la potencia de los sentimientos positivos y negativos

En este primer gráfico observamos cómo la mayoría de los tweets son clasificados con una fuerza muy baja, el 65% de los positivos y el 74% de los negativos. Van seguidos de otros con una potencia baja o moderada, el 32% de los positivos y el 23% de los negativos. Aquellos tweets con la clasificación de sentimiento fuerte o muy fuerte son minoritarios, un 3% de los positivos y 3% de los negativos.

De aquí se puede concluir que a la gran mayoría de publicaciones no se les puede asociar un sentimiento claro, probablemente porque son tweets poco relevantes, noticias que normalmente no incluyen una valoración en el tweet o comentarios muy genéricos sobre un juego que están jugando. Aquellos tweets que realmente están valorando enérgicamente un juego o la consola son más bien pocos, y vemos que los que tienen un sentimiento asociado fuerte son tanto positivos como negativos por igual, mientras que aquellos que tienen un sentimiento asociado muy fuerte, son mayoritariamente positivos. Cabe destacar que hay una cantidad significativa de publicaciones con un sentimiento moderado que también son principalmente positivos. Con esto podemos suponer que, al tratarse de una videoconsola, solamente la gente que se lo pasa realmente bien jugando es la que decide publicar un comentario sobre ello, mientras que, si el juego no le ha gustado mucho, el usuario parece que no está dispuesto a invertir su tiempo en comentarlo en redes. Además, como el usuario sabe el tipo de juegos que le gustan, solo compra estos. Al hacer comentarios sobre los títulos que está jugando van a ser mayoritariamente positivos, pues ya han pasado un filtro y ha descartado aquellos juegos que no le gustan, no le gustan a sus amigos o las reviews son malas, y sobre estos no comenta.

En otra de las gráficas que ofrece Mozdeh y que se encuentra en el anexo (B.1), se representa el número de tweets por la puntuación otorgada en cada sentimiento. Esta representación pone de manifiesto que cuanto más polarizado esté un mensaje, tanto positivamente como negativamente, menos lo estará del sentimiento opuesto, ya que para que se mezclaran los dos sentimientos en la gráfica tendríamos que encontrar una diagonal con cuadrados mucho más grandes.

Aunque no se han podido recuperar los tweets originales de la primera parte del proyecto debido al cambio de herramienta, sí que se han obtenido nuevas publicaciones con la misma query para ver si existía alguna diferencia en el análisis de los sentimientos. Los resultados se encuentran en el anexo B bajo el nombre B.2.

Cabe destacar que el porcentaje de tweets positivos es mayor en todas las categorías en las que se puede detectar alguna potencia de sentimiento. Esto puede deberse a que aquellas personas que comparten el contenido con los hashtags #PS4Share o #PS5Share, lo hacen porque les está gustando mucho la partida y por eso lo comparten, teniendo en cuenta que los pocos que añaden un texto a la publicación lo hacen con un sentido mayoritariamente positivo.

## 5.- CONCLUSIONES

Para afianzar los descubrimientos del comportamiento de los usuarios en relación a la comunidad de PlayStation, presentaremos las principales conclusiones a las que hemos llegado después de realizar el estudio estadístico sobre el conjunto de usuarios y un análisis de sentimientos sobre los tweets que se han publicado.

Uno de los objetivos que nos propusimos al comienzo del análisis era averiguar cuantos subgrupos había en nuestra comunidad. Los resultados obtenidos nos indican que el mejor número de clusters son 3, ya que a partir de este punto la varianza no sigue disminuyendo significativamente, a pesar de añadir más clases. Para poder entender mejor estos grupos de usuarios hemos generado una serie de gráficos que nos han ayudado a ver que uno de estos grupos representa al usuario medio con cuentas pequeñas; es decir, pocos seguidores, y que comparten vídeos de sus sesiones de juego, o son meramente espectadores. Son gente corriente sin mucha influencia y eso explica sus características principales. Otra agrupación modela a cuentas medianas con bastantes seguidores y algunos *retweets*, de tiendas de venta al por menor o similares. El último contiene un número muy reducido de grandes cuentas con alta popularidad, como revistas muy importantes o cuentas oficiales de PlayStation, con un gran número de seguidores y *retweets*. Los últimos dos grupos contienen cuentas empresariales y de massmedia, que se centran en la creación de contenidos y en publicitar sus productos, y es por ello que cuentan con este número de seguidores y *retweets*. Esto nos lleva a cumplir con el objetivo inicial de tipificar las características más relevantes de los grupos.

A continuación, con el análisis PCA hemos observado que la componente del número de cuentas seguidas posee poca relevancia a la hora de tipificar los usuarios mientras que el resto de características sí que ofrecen un peso significativo. Además, también se ha podido constatar la fuerte relación que hay entre los *retweets* y los *favorites*, algo que era bastante lógico pues cuantos más *retweets* tenga una publicación a más gente le va a llegar y es más probable que la gente le de favorito, cumpliendo así con el objetivo de determinar correlaciones entre características.

Respecto al análisis de sentimientos, a la mayoría de publicaciones no se les puede otorgar una potencia de sentimiento mínimamente elevada, ya que la mayoría de estos tweets son sorteos o anuncios y aquellas publicaciones que hacen referencia a estos, es decir, son textos que no contienen ninguna opinión. Por otra parte, debido al filtro que existe a la hora de comprar un juego (lo adquirimos si pensamos que nos va a

gustar o si le ha gustado a algún amigo), cuando un usuario realiza un comentario respecto a este, es mucho más probable que sea positivo y es por ello que las publicaciones con una potencia  $> 1$  son mayoritariamente positivas. Estos descubrimientos han desmentido nuestras hipótesis iniciales, en las que suponíamos que la polarización sería considerablemente mayor.

Atendiendo a la tipología de usuario mayoritario de cada grupo, los descubrimientos del análisis de sentimientos corroboran las suposiciones realizadas en el apartado de clustering. Los usuarios espectadores no suelen manifestar sus opiniones, y además las cuentas corporativas prefieren evitar las opiniones controvertidas y fuertes dado que estas repercutirían negativamente en sus objetivos de ventas. Generalmente, las opiniones sobre los productos de Sony tienden a ser ligeramente positivas, aunque no estamos seguros si esto es debido a las publicaciones de los usuarios o si, por el contrario, han sido las grandes cuentas las que hayan influenciado más este resultado.



## 6.-DISCUSIÓN

Aunque el trabajo realizado y las conclusiones obtenidas son interesantes, pensamos que ha habido ciertos factores que han limitado la presentación de resultados más complejos.

En primer lugar, las propias limitaciones que impone la API que usamos para recoger los datos Twitter, como que se obtienen solamente 2000 tweets cada 60 minutos, lo que condiciona un análisis por hora al poder estar recibiendo unos datos sesgados.

Otro factor bastante importante que ha condicionado directamente la obtención de conclusiones ha sido la imposibilidad de utilizar los mismos datos del apartado de clustering en el apartado de análisis de sentimientos. Si bien esto nos ofrece una mayor visión del conjunto de datos, ya que están tomados con una semana de diferencia, por otra parte nos reduce la capacidad de inferir conclusiones sobre el mismo dataset.

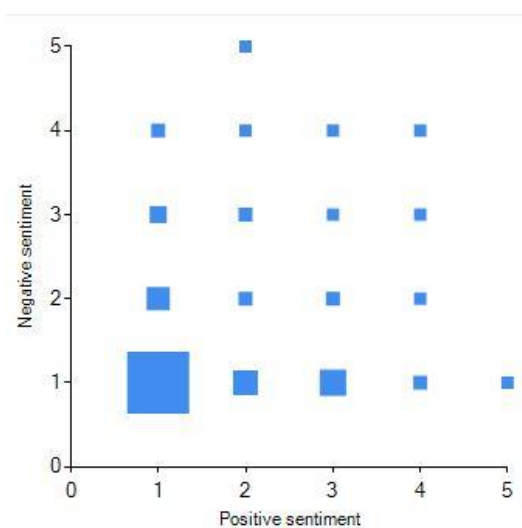
Por otra parte, con un periodo más extenso de tiempo, nos hubiera gustado centrar más nuestra atención en el análisis de sentimientos, realizando un análisis por cluster para ver si diferían los sentimientos entre los grupos o haber estudiado emociones concretas.

# ANEXO A

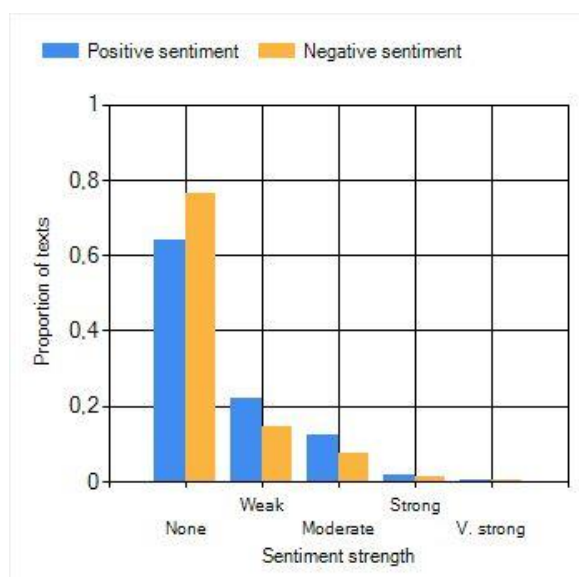
Variable	Observaciones	Obs. con datos perdidos	Obs. sin datos perdidos	Mínimo	Máximo	Media	Desv. típica
Retweets	55665	0	55665	0.000	579.000	0.130	3.480
Favorites	55665	0	55665	0.000	991.000	0.811	7.167
Followers	55665	0	55665	0.000	2754074.000	1246.328	20302.640
Follows	55665	0	55665	0.000	180379.000	453.575	1702.642
Listed	55665	0	55665	0.000	26332.000	12.947	186.337

**Fig. A.1:** *datos estadísticos sobre las características de los tweets*

## ANEXO B



**Fig. B.1:** *correlación sentimientos positivos y negativos*



**Fig. B.2:** *clasificación de la potencia de los sentimientos positivos y negativos con la query antigua*