

# Informe RecSys T01

Sistemas Recomendadores IIC3633 | 2021-2

Recomendación no personalizada, basada en feedback implícito y basada en contenido.

[Anime Recommendation Database - Kaggle](#)

Alumnos

- Diego Iruretagoyena - 14619164
- Víctor Tirreau - 17637171

💡 Recomendamos fuertemente ver este informe en su [versión online](#) en Notion. 💡

---

## [AC01 | Exploración de datos](#)

[Interacciones por usuarios y usuarios más activos](#)

[Interacciones por animé y animes más vistos](#)

[Indicadores estadísticos del dataset](#)

## [AC02 | Recomendación no personalizada](#)

[Most popular](#)

[Random](#)

## [AC03 | Recomendación feedback implícito](#)

[Análisis de sensibilidad - rendimiento contra factores latentes](#)

[Análisis de sensibilidad - tiempos de entrenamiento](#)

## [AC04 | Recomendación basada en contenido](#)

[Análisis de sensibilidad en base a factores latentes](#)

## [AC05 | Ejemplos de recomendación de animé](#)

[Análisis y discusión de resultados](#)

## [AC06 | Ejemplos de recomendación de animé](#)

[Recomendaciones](#)

[Recomendación usuario perfil Warm Start](#)

[Recomendación usuario con participación media](#)

[Recomendación usuario con alta participación](#)

[Historial perfiles de usuario](#)

[Historial usuario Warm Start](#)

[Géneros más frecuentes de usuario Warm Start](#)

[Géneros más frecuentes de usuario usuario con participación media](#)

[Historial usuario con alta participación](#)

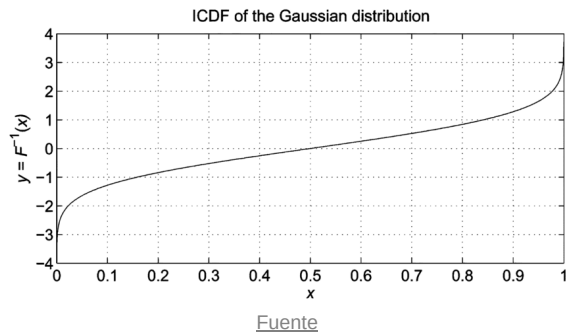
[Géneros más frecuentes de usuario usuario con participación alta](#)

[Experimentos futuros](#)

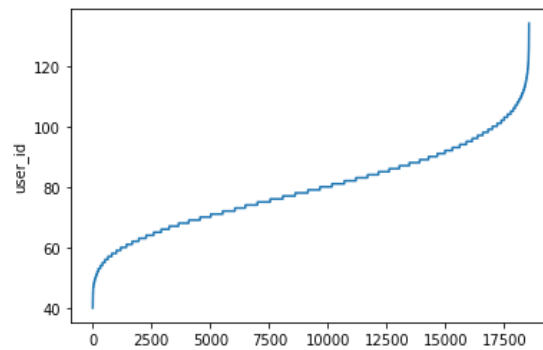
## AC01 | Exploración de datos

### Interacciones por usuarios y usuarios más activos

Gran parte de los usuarios ha registrado entre 60 y 100 recomendaciones de anime, con una distribución similar a una gaussiana. De hecho, la función inversa de la CDF de una gaussiana se ve la siguiente forma:



Mientras que la distribución de los ratings en el dataset se ve así:

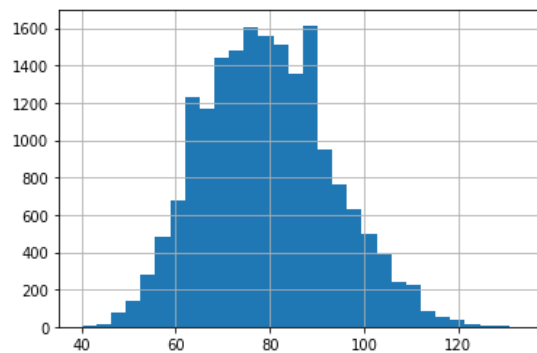


Distribución acumulada de animes vistos (eje Y) según cantidad de usuarios (eje X).

Nótese la similitud entre la forma de las distribuciones acumuladas.

Con respecto al histograma, se confirma la sospecha de la forma Gaussiana en la distribución, con una pequeña anomalía en torno a 85 - 90 ratings.

Además, se observa una leve asimetría a la izquierda:



Histograma de cantidad de animes vistos (eje X) según cuántas personas han visto dicha cantidad (eje Y)

A continuación, podemos observar que los usuarios que más animes han evaluado tienen entre 128 y 134 ítems distintos.

ID	Count
297988	134
469	131
6088	129
214484	129
303684	128

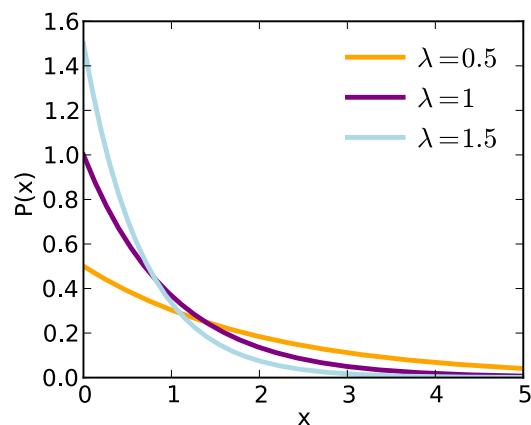
La cantidad de usuarios únicos en el dataset asciende a 18.552. Los 5 usuarios más activos contribuyen un 0,044% de las reviews totales, lo que **no es relevante** con respecto a el resto de los usuarios.

## Interacciones por animé y animes más vistos

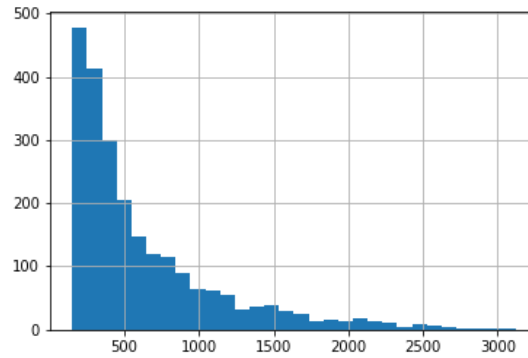
Los 5 animes más vistos se detallan en la siguiente tabla:

anime_id	Rating Count ↓	Name
16498	3119	No Game No Life
11757	3108	Sword Art Online
6547	2923	Angel Beats!
1535	2876	Death Note
19815	2843	No Game No Life

Los 5 animes más vistos acaparan el 1% del total de ratings, siendo esta distribución significativamente más asimétrica que la de ratings por usuario. En efecto, la distribución de ratings por ítem se asemeja bastante a una distribución exponencial:

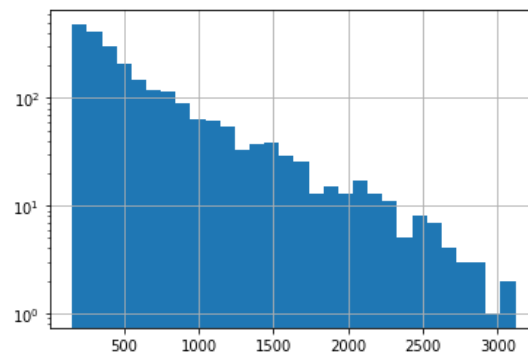


Fuente. Ejemplo de PDF distribución exponencial



Histograma para la distribución de ratings por animé. En el eje X se presenta la cantidad de ratings y en el eje Y la cantidad de animes que han recibido dicha cantidad de ratings.

Para comprobar que la distribución sea efectivamente exponencial, se puede graficar un histograma con eje Y logarítmico, tras lo cual debería observarse una distribución triangular derecha.



Histograma para la distribución de ratings por animé, escala de eje Y logarítmica.

Se observa que la distribución de ratings por animé tiene forma exponencial.

## Indicadores estadísticos del dataset

Estadístico	Valor
#Usuarios únicos	18552
#Animes únicos	2326
Animes/usuario AVG	79.51
Animes/usuario STD	13.70
Usuarios/animé AVG	634.16
Usuarios/animé STD	513.60
Sparsity	96.58%

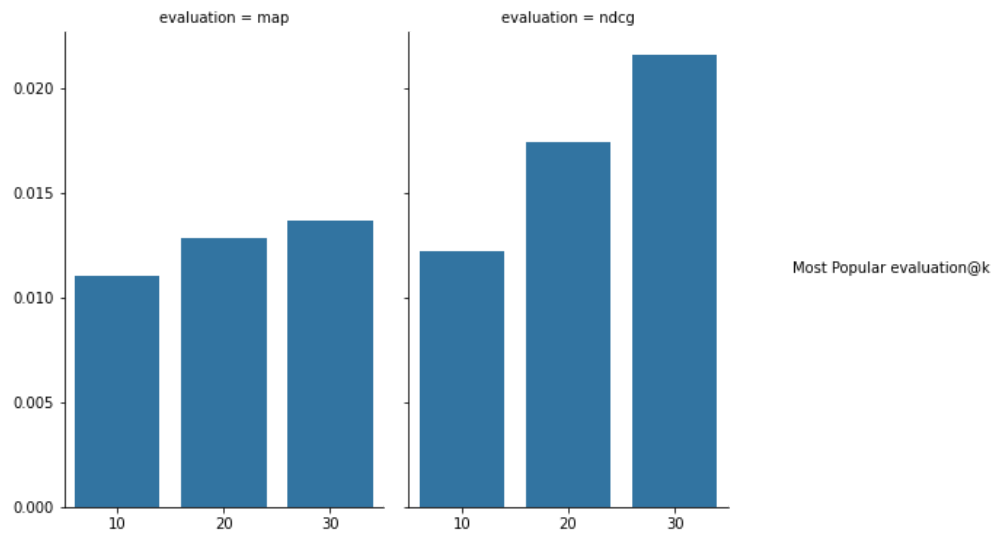
## AC02 | Recomendación no personalizada

### Most popular

Utilizamos la librería `pyreclab` para realizar las recomendaciones más populares. Recomendamos 30 ítems a cada usuario del set de validación y computamos  $nDCG@k$  y  $mAP@k$ , para  $k = 10, 20, 30$ . A continuación se tabulan y

grafican los resultados obtenidos por este método *baseline*.

k	nDCG@k	mAP@k
10	0.0122	0.0110
20	0.0174	0.0128
30	0.0216	0.0136



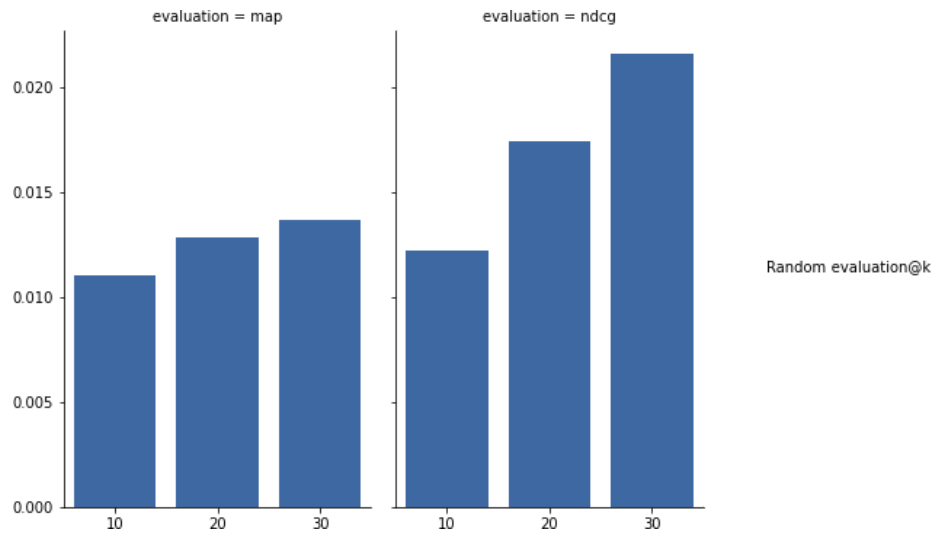
Rendimiento Most Popular para mAP@k y nDCG@k. En el eje x se ubican los distintos valores de k.

## Random

Realizamos 30 recomendaciones aleatorias a cada usuario del set de validación, descontando de los animes candidatos a recomendar aquellos que ya hubiesen sido vistos por los usuarios.

Los resultados se reportan en la tabla y gráfico a continuación:

k	nDCG@k	mAP@k
10	0.0070	0.0020
20	0.0141	0.0024
30	0.0224	0.0028

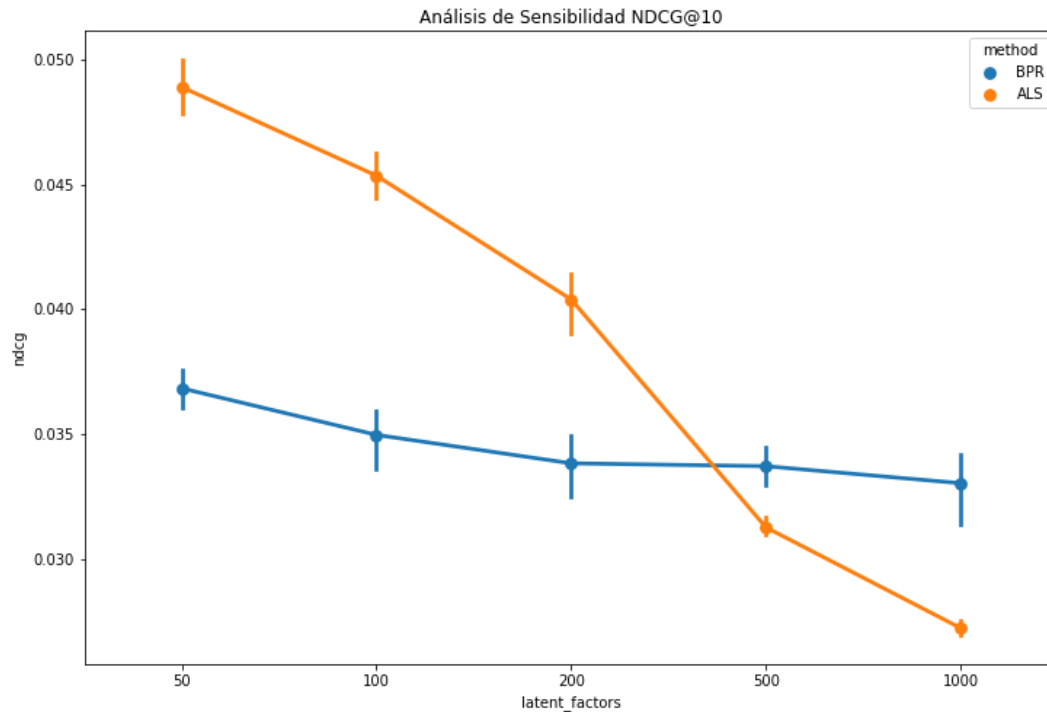


A modo de análisis, si bien tanto Random como Most Popular son resultados *baseline*, es interesante que al cortar las primeras  $k = 20$  posiciones, Most Popular es mejor recomendador en términos de  $\text{nDCG}@k$  que Random, pero es superado por este último cuando  $k = 30$ . Si bien el incremento no es significativo, da nociones que recomendar ítems populares funciona mejor en las primeras posiciones, pero al ser una recomendación prácticamente idéntica para todos, no puede capturar diversos tipos de usuario que Random podría acertar por azar. No obstante, en términos de  $\text{mAP}@k$ , Most Popular es una alternativa sustancialmente más robusta.

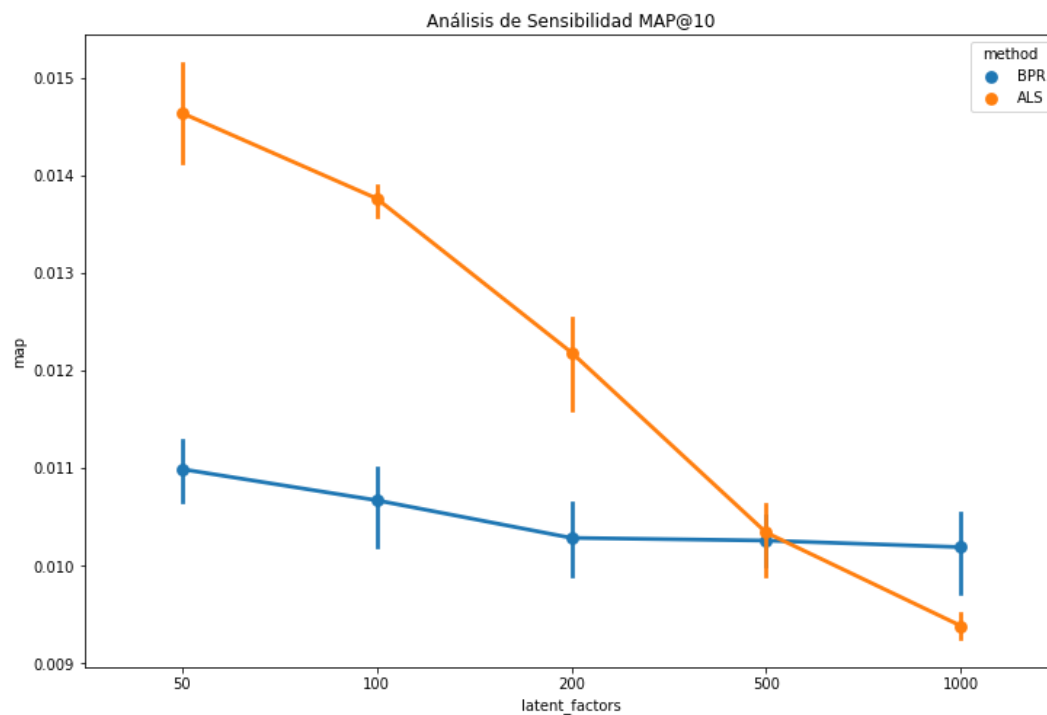
## AC03 | Recomendación feedback implícito

### Análisis de sensibilidad - rendimiento contra factores latentes

Utilizamos la librería `implicit` para entrenar modelos de factorización matricial basados en feedback implícito. En particular, entrenamos los modelos ALS (Alternating Least Squares) y BPR (Bayesian Personalized Ranking). A continuación, se presentan los resultados de ambos métodos en cuanto a  $\text{nDCG}@10$ ,  $\text{mAP}@10$  y tiempo de entrenamiento.



Resultados nDCG@10 de los modelos ALS y BPR (eje Y), para distintas configuraciones de factores latentes (eje X).



Resultados map@10 de los modelos ALS y BPR (eje Y), para distintas configuraciones de factores latentes (eje X).

Se puede observar una tendencia fuerte a la baja en el rendimiento para ambas métricas en ALS. Para BPR, existe también una leve baja en el rendimiento. Esto se puede explicar por un posible *overfitting* en los modelos con mayor

cantidad de factores latentes, hipótesis que comprobamos al graficar el rendimiento en función de la cantidad de iteraciones de entrenamiento en la actividad 5.

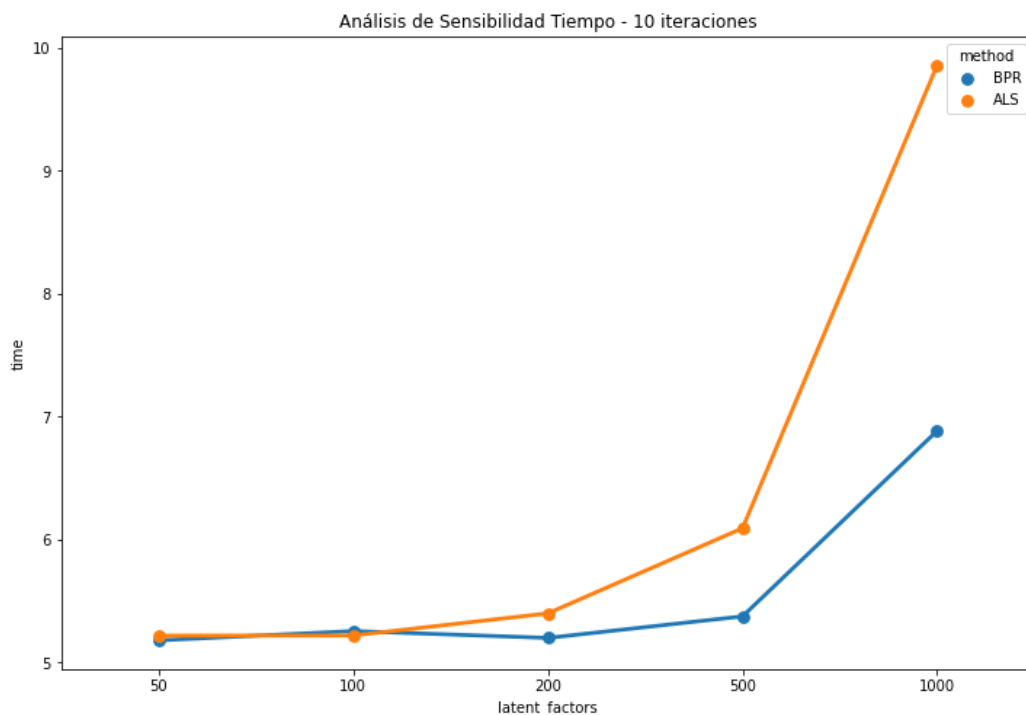
Dado que durante el entrenamiento, la métrica `train AUC` (utilizada por la librería para reportar el estado de entrenamiento) alcanzaba valores cercanos a 100% y, a la vez, el rendimiento en validación disminuía fuertemente (sobre todo en BPR), lo más probable es que el algoritmo esté memorizando los datos de entrenamiento en desmedro del rendimiento y capacidad de generalización en validación.

## Análisis de sensibilidad - tiempos de entrenamiento

Del análisis experimental del rendimiento en función de la cantidad de iteraciones de entrenamiento, dedujimos que para la convergencia de ALS requerimos menos de 20 iteraciones, mientras que BPR requiere alrededor de 100.

Fijándonos en la misma cantidad de iteraciones (en lugar de esperar hasta la convergencia), observamos que el proceso de entrenamiento en función de la cantidad de factores latentes es considerablemente más rápido en BPR que ALS.

Cabe mencionar que todos los tiempos reportados corresponden a segundos y que los algoritmos de entrenamiento fueron acelerados mediante una GPU Nvidia V100 - 16GB.



Tiempo de entrenamiento para 10 iteraciones según cantidad de factores latentes - ALS y BPR

Evidentemente, una estrategia de *early-stopping* puede amortizar el costo de entrenar un algoritmo ALS contrastado con un algoritmo BPR, dado que no se observan incrementos en el rendimiento incluso tras solo 10 iteraciones de ALS (graficado en la sección 5).

## AC04 | Recomendación basada en contenido

Las recomendaciones basadas en contenido se caracterizan por basarse en las descripciones de los ítems presentes y en perfiles de usuario.



En este ejercicio, hemos utilizado **Universal Sentence Encodings**, codificaciones de texto en vectores de alta dimensionalidad (512), los que podemos utilizar para hacer clasificación de texto, similitud semántica y otras operaciones de Procesamiento de Lenguaje Natural, aprovechando que tenemos una base de datos con información útil sobre cada item, en la forma de descripciones sobre cada animé, tales como su título, género, sinopsis y descripción de contenido.

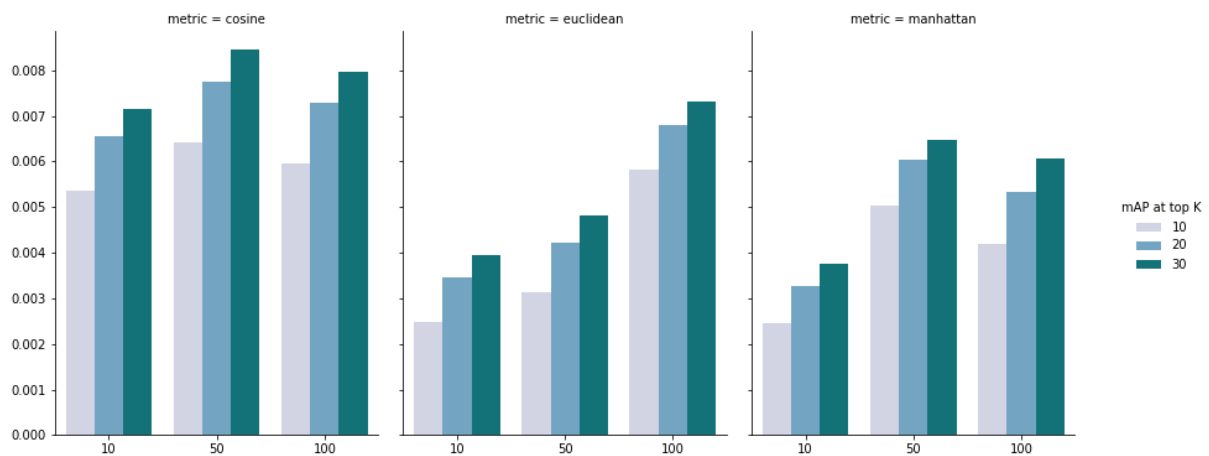
Ahora bien, la dimensionalidad de estos vectores es muy alta para poder utilizar la totalidad de la información. Es por esto que utilizamos **Principal Component Analysis** para reducirla, utilizando solo aquellos componentes con mayor representatividad sobre los embeddings.

Hemos reducido los embeddings a representaciones de **10, 50 y 100** componentes, para luego producir recomendaciones usando operaciones vectoriales. Calculamos las **distancias coseno, euclideana y manhattan**, para luego obtener los K más cercanos, con K entre **10, 20 y 30**.

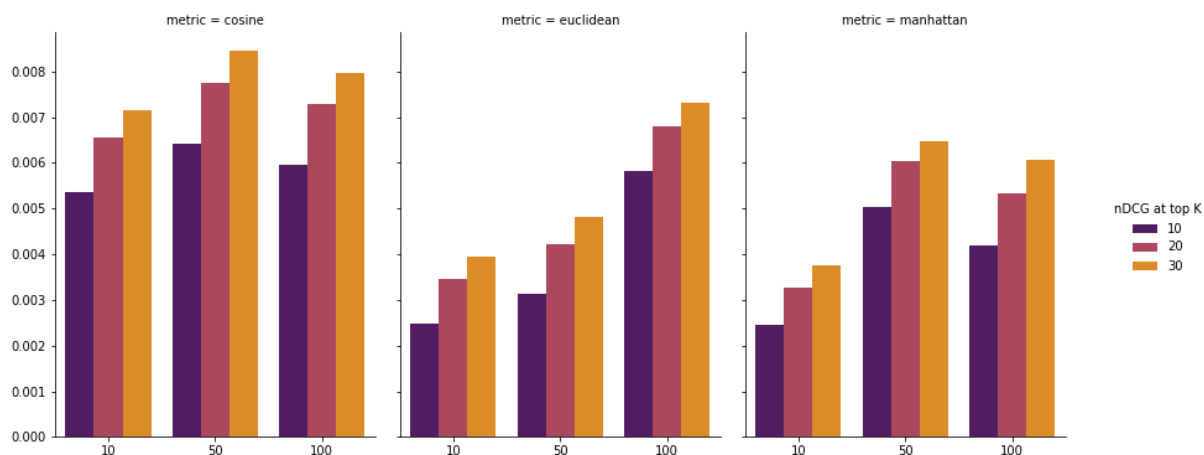
## Análisis de sensibilidad en base a factores latentes

En los siguientes `sns.FacetGrid.barplots`, tenemos el comportamiento de los hiperparámetros.

### Análisis MAP@K



### Análisis ncdg@K



Es interesante observar que para toda métrica de distancia, el resultado mejora al utilizar mayor cantidad de vectores vecinos cercanos. Los mejores resultados tanto para **mAP** como **ndcg** fueron usando **distancia coseno**, **50 componentes principales** y **30 vectores vecinos**. Todos los resultados están dentro del rango de resultado **0.005 - 0.008** en **ndcg** y **0.0025 - 0.009 mAP**, obteniendo resultados menores que el resto de los modelos.

## AC05 | Ejemplos de recomendación de animé

Una vez completados nuestros cinco métodos, hicimos diversos experimentos con el fin de buscar los mejores hiperparámetros para cada modelo. Para todos se utilizó un threshold de relevancia de **7+ reviews** por ítem. Se presentan los mejores resultados para cada método.

**Tabla comparativa de resultados - nDCG@30**

Método	Resultado
ALS	<b>0.133144</b>
BPR	0.102840
PCA	0.061013
Random	0.022371
Most Popular	0.021628

**Tabla comparativa de resultados - mAP@30**

Método	Resultado
ALS	<b>0.019522</b>
BPR	0.015077
Most Popular	0.013672
PCA	0.008609
Random	0.002817

**Detalle hiperparámetros nDCG@30**

Método	Resultado	Factores Latentes	Iteraciones	# Componentes	Métrica Distancia
ALS	0.133144	50	40	-	-
BPR	0.102840	50	110	-	-
PCA	0.061013	-	-	50	Coseno
Random	0.022371	-	-	-	-
Most Popular	0.021628	-	-	-	-

#### Detalle hiperparámetros MAP@30

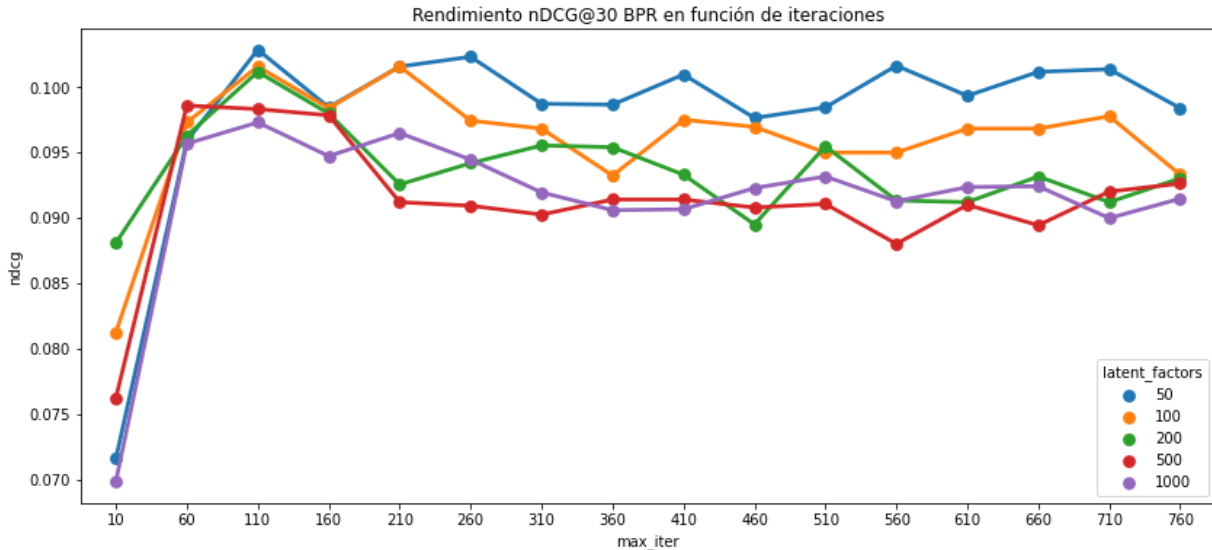
Método	Resultado	Factores Latentes	Iteraciones	# Componentes	Métrica Distancia
ALS	0.019522	50	50	-	-
BPR	0.015077	50	610	-	-
Most Popular	0.013672	-	-	-	-
PCA	0.008609	-	-	50	Coseno
Random	0.002817	-	-	-	-

## Análisis y discusión de resultados

En primer lugar, sobre los métodos de factorización matricial, se observa que la convergencia de ALS se logra mucho antes que BPR, sin embargo, el comportamiento asintótico del tiempo de ejecución del primero es mucho peor que el segundo cuando se modifica la cantidad de factores latentes. Esto es particularmente evidente para ALS con 1000 factores latentes, que tarda en torno a 10 segundos para entrenamiento y evaluación, de los cuales 5 segundos son atribuibles a la evaluación. En contraste, BPR tarda 7 segundos para entrenar y evaluar bajo la misma cantidad de factores latentes, con igualmente 5 segundos destinados a evaluación. En la misma línea, se observó empíricamente un descenso en el rendimiento al aumentar la cantidad de factores latentes. Como el máximo rendimiento se observó en el comienzo de la escala probada, queda abierta la posibilidad de probar con aún menos factores latentes, lo que no se ha reportado en este informe.

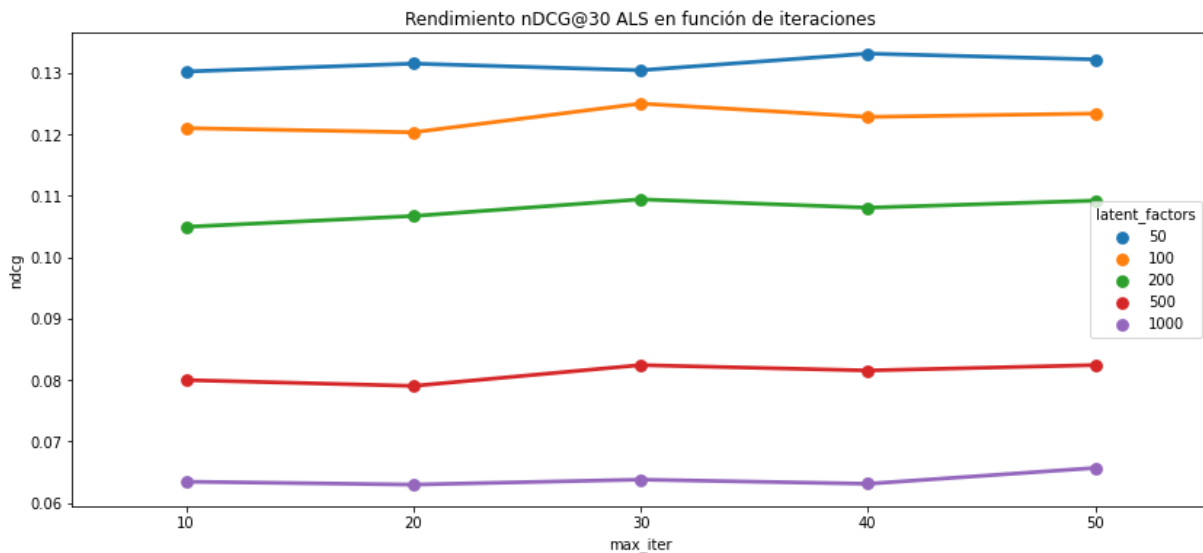
Como segundo punto, la técnica de content-based recommendation tiene un resultado mediocre comparado a las técnicas de feedback implícito. Esto se puede deber a que modela las percepciones como un promedio entre los ítems relevantes, y suele sufrir de problemas como *echo-chamber* o *information-bubble*. Así, la calidad de recomendaciones se ve perjudicada por la localidad de los embeddings, y rara vez logra recomendar ítems novedosos o con serendipia. De las métricas de distancia, se observó que la mejor resulta ser cosine-distance. Si bien puede tener similitud con la técnica euclideana, la primera solo rankea en función del ángulo entre dos vectores, permitiendo que aquellos colineales o que apunten a la misma región del espacio k-dimensioal sean mejor recomendados que bajo el esquema euclideano. El método Manhattan resultó ser el peor, probablemente porque la métrica L1 ha sido probada como útil en regímenes de mucho más alta dimensionalidad.

Finalmente, retomando el argumento de Random contra Most Popular, es entendible que el segundo presente un mAP@30 rotundamente mejor que Random, sin embargo, resalta que haya tenido un peor rendimiento en nDCG@30. Esto se puede explicar por el incremento en diversidad que puede aportar Random, contrastando con la recomendación prácticamente constante de Most Popular.



Rendimiento nDCG@30 para el método BPR en función de iteraciones de entrenamiento

Se observa en el gráfico anterior una clara tendencia a la baja a medida que aumentan las iteraciones máximas y también los factores latentes, particularmente evidente para los casos con 500 y 1000 factores. Se observa un *peak* de rendimiento en datos de validación en torno a las 110 iteraciones, decreciendo a medida que progresa el entrenamiento. En contraste, cuando los factores latentes son bajos, no solo se observa un mejor rendimiento, sino que también consistencia en su valor a lo largo del entrenamiento.



Rendimiento nDCG@30 para el método ALS en función de iteraciones de entrenamiento

Para ALS la convergencia es sustancialmente más rápida, no pudiendo evidenciarse cambios en el rendimiento tras solo 20 iteraciones. Además, es evidente la superioridad en el rendimiento para una menor cantidad de factores latentes al modelar las interacciones usuario-item.

Hemos entregado un archivo `/test_predictions.json` que contiene diez recomendaciones a cada usuario presente en el set de testing. [Link a archivo en Github.](#)

## AC06 | Ejemplos de recomendación de animé

Con el fin de analizar cualitativamente nuestras recomendaciones, tomamos tres perfiles de usuarios y estudiamos las recomendaciones que les hicimos. Nuestros perfiles fueron usuarios **Warm Start**, **Participación Media** y **Alta participación**. Para hacer esto, ordenamos la lista de usuarios según la cantidad de reviews que tenían, y obtuvimos los índices de usuarios que se encontraban en las posiciones representativas dentro del dataset de testing.

ID Usuario	Perfil	# Reviews
29392	Warm Start	40
297988	Participación media	79
46569	Alta participación	134

### Recomendaciones

A continuación se presenta tabla de diez recomendaciones por usuario, indicando detalle de su Nombre, Género y Sinopsis. Se comenta también acerca de los géneros recomendados, comparando con los items que el usuario ya había consumido.

#### Recomendación usuario perfil Warm Start

Name	Genres	Synopsis	anime_id
Clannad: After Story	Slice of Life, Comedy, Supernatural, Drama, Ro...	Clannad: After Story , the sequel to the criti...	4181
Suzumiya Haruhi no Yuuutsu	Comedy, Mystery, Parody, School, Sci-Fi, Slice...	Kyon, your typical high school student, has lo...	849
Shokugeki no Souma	Ecchi, School, Shounen	Ever since he was a child, fifteen-year-old So...	28171
Zankyou no Terror	Mystery, Psychological, Thriller	Painted in red, the word "VON" is all that is ...	23283
Shingeki no Kyojin Season 2	Action, Military, Mystery, Super Power, Drama,...	For centuries, humanity has been hunted by gia...	25777
Gekkan Shoujo Nozaki-kun	Comedy, Romance, School	Chiyo Sakura is a cheerful high school girl wh...	23289
Yahari Ore no Seishun Love Comedy wa Machigatt...	Slice of Life, Comedy, Drama, Romance, School	Hachiman Hikigaya is an apathetic high school ...	14813
Fate/stay night	Action, Supernatural, Magic, Romance, Fantasy	fter a mysterious inferno kills his family, Sh...	356
Clannad: After Story - Mou Hitotsu no Sekai, K...	Drama, Romance, School	Included in the 8th and final DVD of Clannad ~...	6351
Nanatsu no Taizai	Action, Adventure, Ecchi, Fantasy, Magic, Shou...	In a world similar to the European Middle Ages...	23755

Género	Cantidad
Romance	5
School	5
Comedy	4
Drama	4
Slice of Life	3
Supernatural	3
Mystery	3
Shounen	3
Action	3
Fantasy	3

### Recomendación usuario con participación media

Name	Genres	Synopsis	anime_id
Hyouka	Mystery, School, Slice of Life	Energy-conservative high school student Houtar...	12189
Akame ga Kill!	Action, Adventure, Drama, Fantasy, Shounen	Night Raid is the covert assassination branch ...	22199
Yahari Ore no Seishun Love Comedy wa Machigatt...	Slice of Life, Comedy, Drama, Romance, School	Yahari Ore no Seishun Love Comedy wa Machigatt...	23847
Boku wa Tomodachi ga Sukunai	Ecchi, Slice of Life, Comedy, Harem, Romance, ...	hen Kodaka Hasegawa finds out that he will be ...	10719
Nisemonogatari	Mystery, Comedy, Supernatural, Ecchi	Surviving a vampire attack, meeting several gi...	11597
Toaru Majutsu no Index II	Action, Magic, Sci-Fi, Super Power	s tensions between the world of magic and Acad...	8937
Fate/stay night: Unlimited Blade Works	Action, Fantasy, Magic, Supernatural	The Holy Grail War is a battle royale among se...	22297
Durarara!!	Action, Mystery, Supernatural	In Tokyo's downtown district of Ikebukuro, ami...	6746
Overlord II	Action, Game, Adventure, Supernatural, Magic, ...	z Ooal Gown, the undead sorcerer formerly know...	35073
Fate/stay night	Action, Supernatural, Magic, Romance, Fantasy	fter a mysterious inferno kills his family, Sh...	356

Género	Cantidad
Action	6
Supernatural	5
Fantasy	4
Magic	4
Mystery	3
School	3
Slice of Life	3
Comedy	3
Romance	3
Adventure	2

### Recomendación usuario con alta participación

Name	Genres	Synopsis	anime_id
Kyoukai no Kanata	Slice of Life, Supernatural, Fantasy	ai Kuriyama is the sole survivor of a clan of ...	18153
Mahou Shoujo Madoka★Magica	Psychological, Drama, Magic, Thriller	adoka Kaname and Sayaka Miki are regular middl...	9756
Mahouka Koukou no Rettousei	Action, Sci-Fi, Supernatural, Magic, Romance, ...	In the dawn of the 21st century, magic, long t...	20785
Psycho-Pass	Action, Sci-Fi, Police, Psychological	Justice, and the enforcement of it, has change...	13601
Death Parade	Game, Mystery, Psychological, Drama, Thriller	fter death, there is no heaven or hell, only a...	28223
Fullmetal Alchemist	Action, Adventure, Comedy, Drama, Fantasy, Mag...	Edward Elric, a young, brilliant alchemist, ha...	121
Ore no Imouto ga Konnani Kawaii Wake ga Nai	Slice of Life, Comedy	Kirino Kousaka embodies the ideal student with...	8769
Gekkan Shoujo Nozaki-kun	Comedy, Romance, School	Chiyo Sakura is a cheerful high school girl wh...	23289
Aldnoah.Zero	Action, Military, Sci-Fi, Mecha	The discovery of a hypergate on the Moon once ...	22729
Gate: Jieitai Kanochi nite, Kaku Tatakaeri	Action, Adventure, Fantasy, Military	Off-duty Japan Self-Defense Forces (JSDF) offi...	28907

Género	Cantidad
Action	5
Fantasy	3
Psychological	3
Drama	3
Magic	3
Sci-Fi	3
Comedy	3
Military	3
Slice of Life	2
Supernatural	2

## Historial perfiles de usuario

Qué habían visto estos usuarios ? Con esta información podremos comparar las recomendaciones versus los items consumidos históricamente.

## Historial usuario Warm Start

Name	Genres	Synopsis	anime_id
Toki wo Kakeru Shoujo	Adventure, Drama, Romance, Sci-Fi	akoto Konno is in her last year of high school...	2236
Grisaia no Kajitsu Specials	Ecchi	Short specials added to Blu-ray/DVD volumes.	29101
Hoshi no Koe	Sci-Fi, Space, Drama, Romance, Mecha	It is the year 2046, Noboru Terao and Mikako N...	256
Kokoro Connect: Michi Random	Comedy, Drama, Romance, School, Slice of Life,...	Not long after putting the previous supernatur...	16001
Owari no Seraph	Action, Military, Supernatural, Drama, Vampire...	h the appearance of a mysterious virus that ki...	26243
Dragon Ball	Adventure, Comedy, Fantasy, Martial Arts, Shou...	Gokuu Son is a young boy who lives in the wood...	223
Chuunibyou demo Koi ga Shitai!: Kirameki no.....	Comedy, Drama, Romance, School, Slice of Life	lthough Yuuta Togashi and Rikka Takanashi have...	16934
Akatsuki no Yona	Action, Adventure, Comedy, Fantasy, Romance, S...	Princess Yona lives a life of luxury and ease,...	25013
Aiura	Comedy, School, Shounen, Slice of Life	The story centers around Amaya, Iwasawa, and U...	17082
Teekyu 2	Comedy, School, Shounen, Sports	Second season of Teekyu series.	18121

## Géneros más frecuentes de usuario Warm Start

Género	Cantidad
Drama	18
Comedy	18
Romance	13
School	13
Supernatural	12
Action	12
Shounen	12
Slice of Life	10
Adventure	7
Sci-Fi	7

Analizando el historial y recomendaciones del usuario 29392, el cual corresponde a un perfil Warm Start, vemos que los top cuatro géneros recomendados Romance, School, Comedy, Drama, y en su historial vemos que es Drama, Comedy, Romance, School. Existe una fuerte correlación entre los géneros recomendados e históricos. De esto podemos concluir que el modelo "se va a la segura" al recomendar items con cualidades parecidas a lo que ya había visto anteriormente, dado que al tener menos información del usuario, no tiene cómo perfilar correctamente al individuo.

## Historial usuario con participación media

Name	Genres	Synopsis	anime_id
Eizouken ni wa Te wo Dasu na!	Adventure, Comedy, School, Seinen	dori Asakusa sees the world a bit differently....	39792
Love Live! The School Idol Movie	Music, School, Slice of Life	Hot on the heels of the third year students' g...	24997
Working!!!: Lord of the Takanashi	Comedy, Romance, Seinen, Slice of Life	The light-hearted Working!! franchise comes to...	31715
Kore wa Zombie Desu ka? of the Dead: Hai, Mino...	Action, Harem, Comedy, Supernatural, Magic, Ecchi	unaired OVA episode of Kore wa Zombie Desu ka?...	15437
Kore wa Zombie Desu ka? of the Dead	Action, Harem, Comedy, Supernatural, Magic, Ecchi	kawa Ayumu was revived as a zombie by the cute...	10790
High School DxD	Comedy, Demons, Ecchi, Harem, Romance, School	High school student Issei Hyoudou is your run-...	11617
Kono Subarashii Sekai ni Shukufuku wo! 2	Adventure, Comedy, Parody, Supernatural, Magic...	hen Kazuma Satou died, he was given two choice...	32937
Kami nomi zo Shiru Sekai	Comedy, Harem, Romance, Shounen, Supernatural	Keima Katsuragi, known online as the legendary...	8525
Strike the Blood II	Action, Harem, Supernatural, Ecchi, Vampire, F...	The second season of Strike the Blood which ad...	33286
Clannad: After Story	Slice of Life, Comedy, Supernatural, Drama, Ro...	Clannad: After Story , the sequel to the criti...	4181

## Géneros más frecuentes de usuario usuario con participación media



Género	Cantidad
Comedy	66
Action	50
Romance	47
School	46
Supernatural	35
Fantasy	34
Shounen	31
Drama	31
Slice of Life	28
Sci-Fi	27

Para el usuario con perfil de participación media, vemos que los principales géneros que hemos recomendado son Action, Supernatural, Fantasy, Magic, mientras que sus géneros históricos son Comedy, Action, Romance, School. Llama la atención que Comedy no sea predominante en sus recomendaciones, aún cuando es su tipo preferido. Si podemos ver items como Action, repetirse. De esto podemos inferir que el modelo ya logra discernir ciertos perfiles entre usuarios con ~70 reviews, recomendado cosas fuera de lo que ya han visto.

## Historial usuario con alta participación

Name	Genres	Synopsis	anime_id
Inu to Hasami wa Tsukaiyou	Comedy, Supernatural	onsense comical mystery. Harumi Kazuhito is a ...	17831
Mouretsu Pirates	Sci-Fi, Space	The story centers around a spirited high schoo...	8917
Captain Earth	Action, Mecha, Romance, Sci-Fi, Space	One night, right before summer vacation, Manat...	21677
Angel Beats!	Action, Comedy, Drama, School, Supernatural	Otonashi awakens only to learn he is dead. A r...	6547
Persona 4 the Animation	Sci-Fi, Adventure, Mystery, Super Power, Super...	Yuu Narukami moves to Inaba, a seemingly quiet...	10588
Durarara!!x2 Ketsu	Action, Mystery, Supernatural	s Mikado Ryuugamine continues to purge the Dol...	27833
Medarot	Adventure, Comedy, Sci-Fi, Shounen	dabots—powerful robots granted artificial inte...	1138
Hanasaku Iroha	Slice of Life, Comedy, Drama	Ohana Matsumae is an energetic and wild teenag...	9289
Interstella5555: The Story of The 5cret 5tar ...	Adventure, Drama, Music, Sci-Fi	This music video is the visual realization of ...	731
Durarara!!x2 Ten	Action, Mystery, Supernatural	In Ikebukuro, the lives of its citizens contin...	27831

## Géneros más frecuentes de usuario usuario con participación alta

Género	Cantidad
Action	51
Sci-Fi	29
Comedy	27
Supernatural	24
Fantasy	23
Adventure	19
Mecha	17
Shounen	16
Drama	15
School	15

Para el usuario de perfil más activo, hemos recomendado principalmente géneros como Action, Fantasy, Psychological y Drama, mientras que sus items más consumidos eran Action, Sci-Fi, Comedy, Supernatural. Vemos que Action toma una posición predominante, pero el sistema recomendador es capaz de entregar nuevos tipos de items, distintos a los que ya había visto.

## Experimentos futuros

De la misma forma que hemos aprendido en clases, es posible generar un ensamble de modelos llamado VBPR, el cual utiliza features visuales junto con feedback implícito para recomendar items. En vez de usar features visuales, podríamos combinar embeddings de texto como un TBPR.

En los modelos normales de Matrix Factorization, cuando nos pidan recomendar usuarios que tienen muy pocos reviews, por ejemplo menos de 15, podemos hacer fallback a Most Popular, generando un ensamble de modelos más completo.

Otro approach sería involucrar técnicas de Deep Learning y generar redes neuronales que manejen los embeddings para usuarios e items, para luego hacer búsquedas por similitud de distancia coseno para obtener los vectores vecinos a nuestra representación vectorial.