

# Memoria EDA: "Mi primer anime"

domingo, 16 de abril de 2023 12:10

## 1. Objetivo del análisis y formulación de hipótesis

Como consumidora habitual de este tipo de formato, el análisis se centrará en confirmar o desmentir ideas personales preconcebidas sobre el anime. Además, con las conclusiones podremos determinar qué características debe tener un anime para considerarse un éxito. Esos datos se podrían utilizar para recomendar un "primer anime ideal" a alguien que nunca haya visto uno.

### 1.1 Formulación de hipótesis

- Los 100 animes mejor valorados son mayoritariamente Shounen de acción. Si bien hay animes para todo tipo de público y de temáticas diversas, creo que el top 100 está dominado por animes de acción dirigidos a público masculino de entre 12 y 18 años.
- La puntuación está relacionada de forma inversa con la duración en minutos (a mayor duración, menor puntuación). Basándome en la percepción personal de que actualmente se valora mucho más lo inmediato, o lo que menos tiempo nos ocupe, considero que cuanto más dure un anime, peor puntuación tendrá.
- Los animes cuya emisión empezó antes de los 2000 tienen peor puntuación que los animes que empezaron a emitirse después del 2000. Aunque el anime ha sido conocido internacionalmente desde los años 60-70 (Heidi, Mazinger Z), no fue hasta la década de los 90 cuando los animes llegaron a la televisión para quedarse (Dragon Ball, OLiver y Benji). Sin embargo, en las últimas décadas el aspecto artístico y narrativo del anime ha evolucionado de forma notable. Esto unido a que cada vez más gente joven consume anime, considero que los animes de este siglo estarán mejor valorados que los del siglo pasado.
- Los animes de género comedia tienen una puntuación media superior que el resto de los géneros. Desde un punto de vista personal, el ocio es sinónimo de pasárselo bien y lograr estar de mejor humor. Por eso considero que el género de anime con mayor valoración media es la comedia.
- La puntuación media de las películas de anime es inferior a las series, y ha sido así históricamente. Aunque desde principio de los 2000 las películas de anime han ido adquiriendo más popularidad (gracias, en parte, a las películas de Estudio Ghibli), esto no es suficiente para que estén mejor valoradas que las series.
- Los animes de suspense tienen menos usuarios que puntúan en comparación a comedia, pero la puntuación media es mayor. Partiendo de la misma base que una de las hipótesis anteriores, el poder de evasión de la comedia consigue que más gente la vea. Sin embargo, el hecho de que tenga más votos influye negativamente en su media, en comparación a suspense.
- La cantidad de usuarios que puntúa un anime influye en su puntuación.

## 2. DataSet original

El DataSet original se obtiene de Kaggle, bajo el nombre Anime 2022. El contenido se basa en web MyAnimeList.com. Originalmente tiene 21460 filas y 28 columnas. Tras la limpieza de datos, su tamaño pasa a ser 2548 filas y 16 columnas.

Enlace al dataset: <https://www.kaggle.com/datasets/harits/anime-database-2022>

Explicación de las columnas originales:

Column	Description
ID	ID del anime en MyAnimeList.net
Title	Título original del anime
Synonyms	Sinónimo del nombre del anime
Japanese	Título en japonés
English	Título en inglés
Synopsis	Breve descripción del anime
Type	Tipo de formato ( serie, película, etc)
Episodes	Número de episodios
Status	Status de emisión
Start_Aired	Fecha de comienzo de emisión
End_Aired	Fecha de final de emisión
Premiered	Temporada del estreno
Broadcast	Día y hora de emisión
Producers	Productores
Licensors	Licencias
Studios	Estudios
Source	Fuente del anime
Genres	Géneros
Themes	Temas
Demographics	Público objetivo
Duration_Minutes	Duración total en minutos de cada episodio
Rating	Edad recomendada para ver el anime
Score	Puntuación media de los usuarios de MyAnimeList
Scored_Users	Número de usuarios que puntuaron el anime
Ranked	Ranking basado en puntuación
Popularity	Ranking basado en la popularidad. Viene dado por cuántos usuarios añaden el anime a su lista
Members	Número de usuarios que añaden el anime a su lista
Favorites	Número de usuarios que añaden el anime como favorito

### 3. Limpieza de datos

- Se eliminan las columnas "ID", "Synonyms", "Japanese", "Synopsis", "End\_Aired", "Producers", "Licensors", "Popularity", "Members", "Favourite", "Premiered" y "Broadcast" por no aportar datos relevantes o por exceso de "Unknowns".
- Se plantea eliminar la columna "Title" y renombrar la columna "English" por "Title". Sin embargo, aunque eso lo haría más comprensible para el público general, se comprueba que la columna "English" demasiados valores "Unknown". Se decide dejar ambas columnas
- Tras limpiar todo el dataframe, se decide que hay que quitar todos los Unknown. A la hora de graficar los resultados, destacan demasiado y no tienen valor.
- En las columnas "Demographics", "Themes" y "Genres" se reemplazan los valores que tienen más de una palabra y se añaden a la categoría de la primera palabra. Por ejemplo, en "Demographics", "Kids, Shounen" se reemplaza por "Kids". Al ser una categoría ya

existente, se suman los valores.

- En la categoría "Start\_Aired" se unifica el formato para que muestre sólo años y se cambia el tipo de la columna a int64.
- La columna "Ranked" se ordena para que tenga una secuencia numérica lógica y se pone como índice.

#### 4. Valores faltantes

- Tras hacer dos pruebas donde, por un lado, se reemplazan los NaN por la media de los valores en las columnas "Score" y "Scored\_User" y, por otro lado, se eliminan las filas de esas columnas cuando hay un NaN, se decide mantener las filas y rellenarlas con el valor medio de esas columnas

- Se decide hacer lo mismo con las columnas "Episodes" y "Duration\_Minutes", tras comprobar que los NaN aparecen tanto en series, como en pelis.

- En "Ranked" no tiene sentido rellenar los NaN con la media, porque es un ranking del 1 al \*\*. Tras verificar la columna Ranked, se decide eliminar las filas donde "Ranked" == NaN. Se comprueba además que eliminando los Unknown de "Ranked" desaparecen los "Status" == "Not yet aired", dejando datos que tienen más valor para nuestras hipótesis

#### 5. Enlace GitHub

#### 6. Bibliografía:

<https://www.kaggle.com/datasets/harits/anime-database-2022>

<https://es.wikipedia.org/wiki/Anime>

<https://www.menzig.es/a/historia-anime-espana-origenes/>

# Resumen tema

lunes, 17 de abril de 2023 10:09

La animación japonesa, también conocida como anime, es un formato audiovisual originario de Japón y potencialmente dirigido a todos los públicos. Es un gran medio de expansión en Japón, siendo al tiempo un producto de entretenimiento cultural y comercial, lo que ha ocasionado un fenómeno cultural en de masas y una forma de arte tecnológico.

El anime es un medio diverso que combina arte gráfico, caracterización, cinematografía y otras formas de técnicas imaginativas e individualistas. Desde la década de 1980, el anime ha incrementado su éxito también de forma internacional debido sobre todo al auge de la programación subtitulada y doblada en el extranjero y su creciente distribución a través de servicios de transmisión. En 2002 fue la primera vez que una película de anime recibió un Oscar y, en 2016, la animación japonesa representó el 60% de los programas de televisión animados del mundo.

# Objetivo del análisis

lunes, 17 de abril de 2023 11:59

Como consumidora habitual de este tipo de formato, el análisis se centrará en confirmar o desmentir ideas personales preconcebidas sobre el anime.