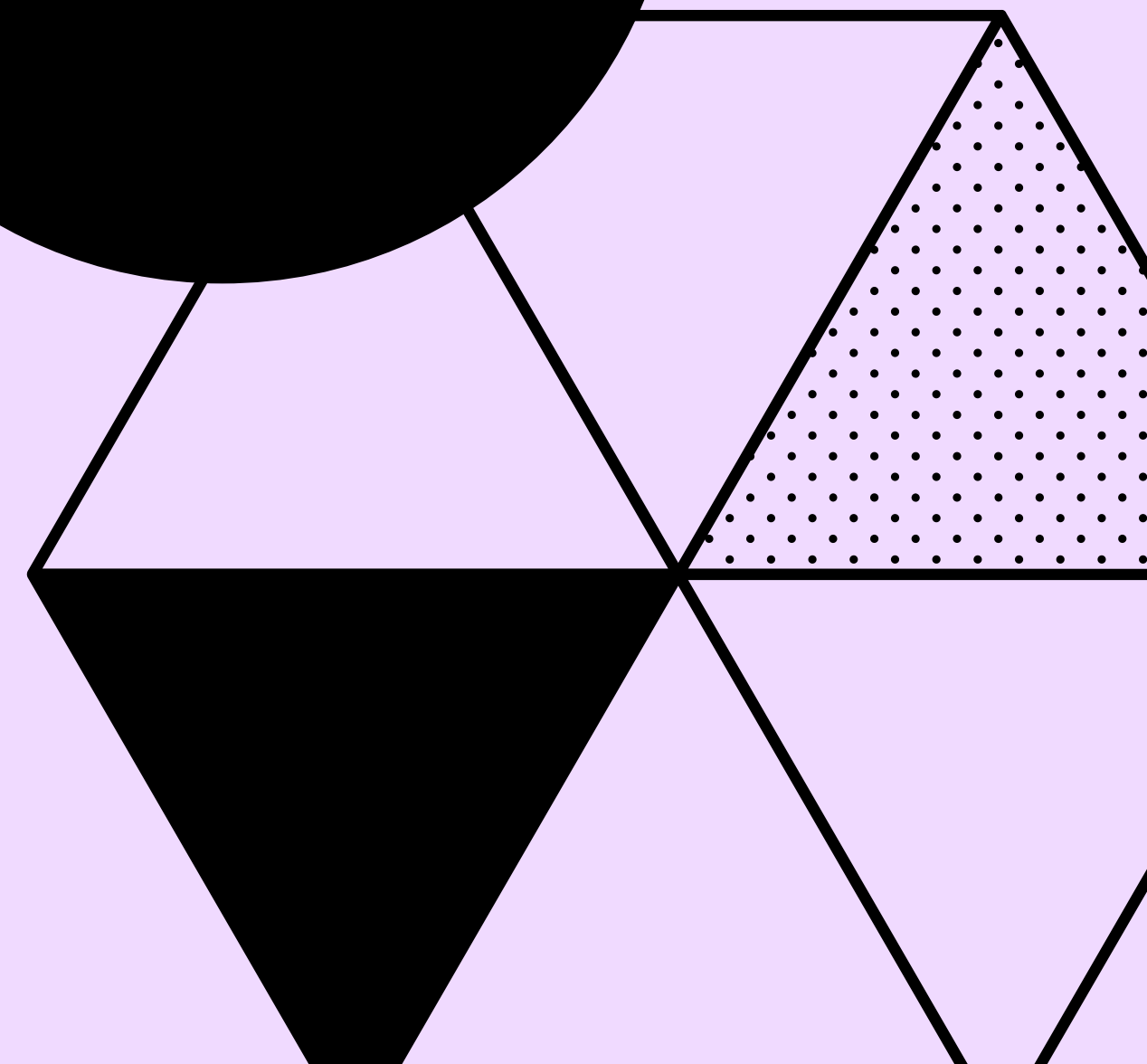
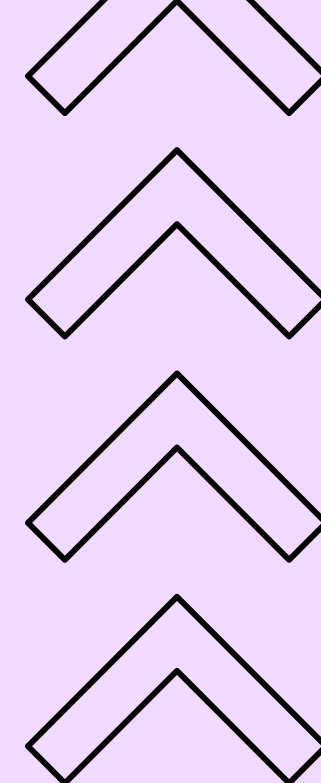
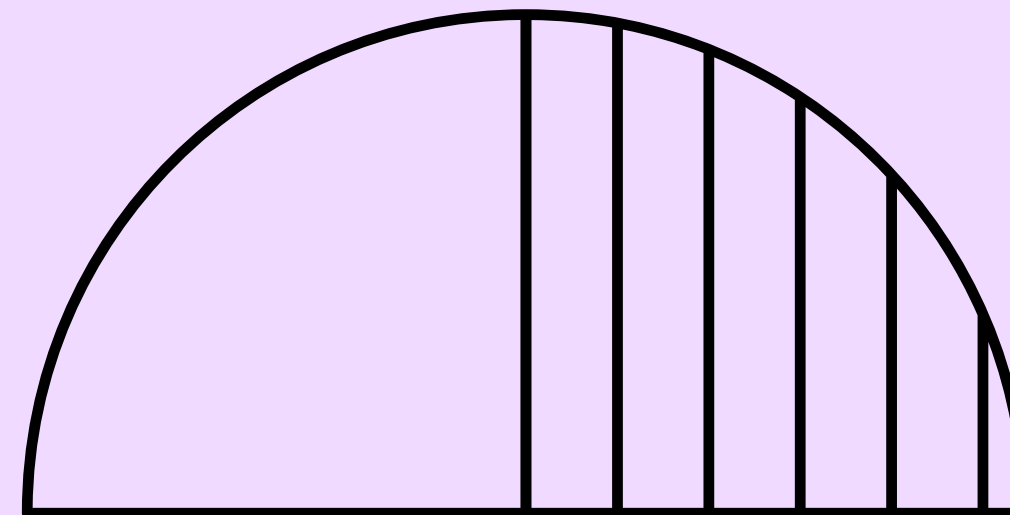
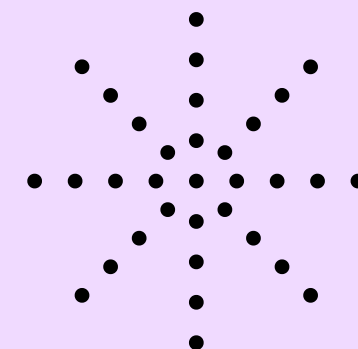
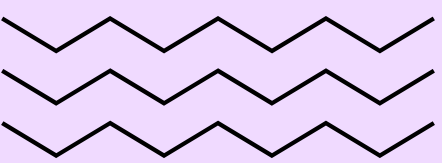
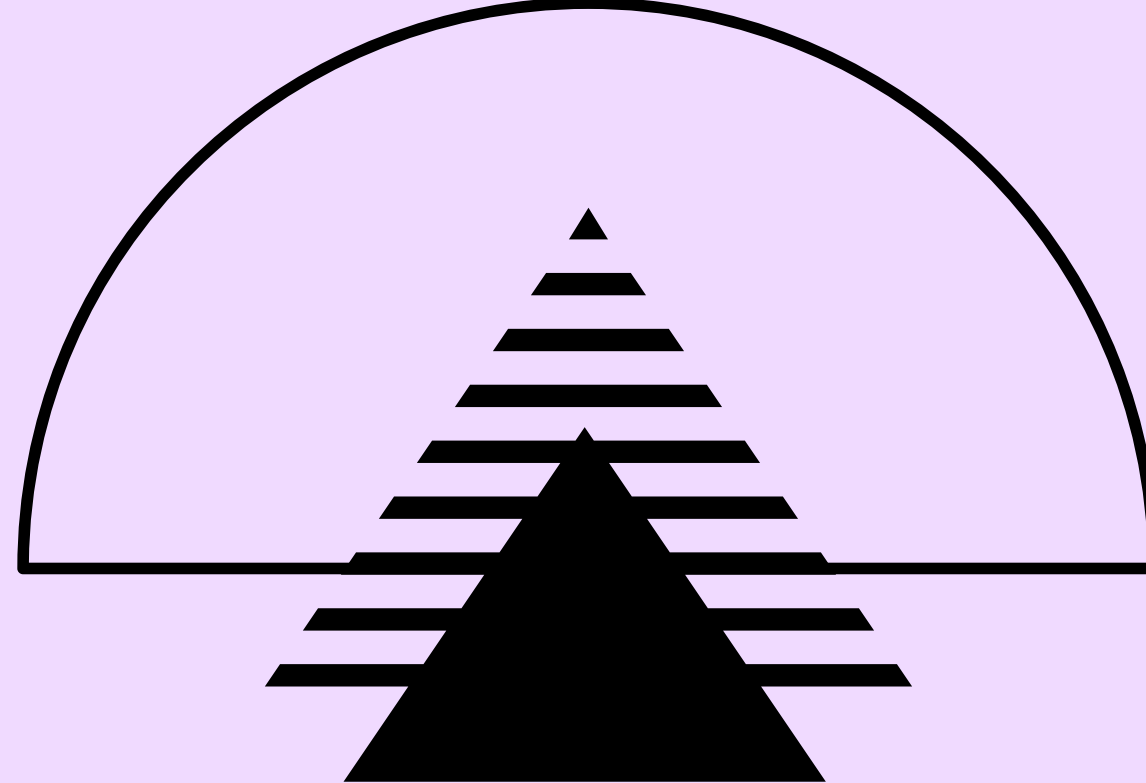
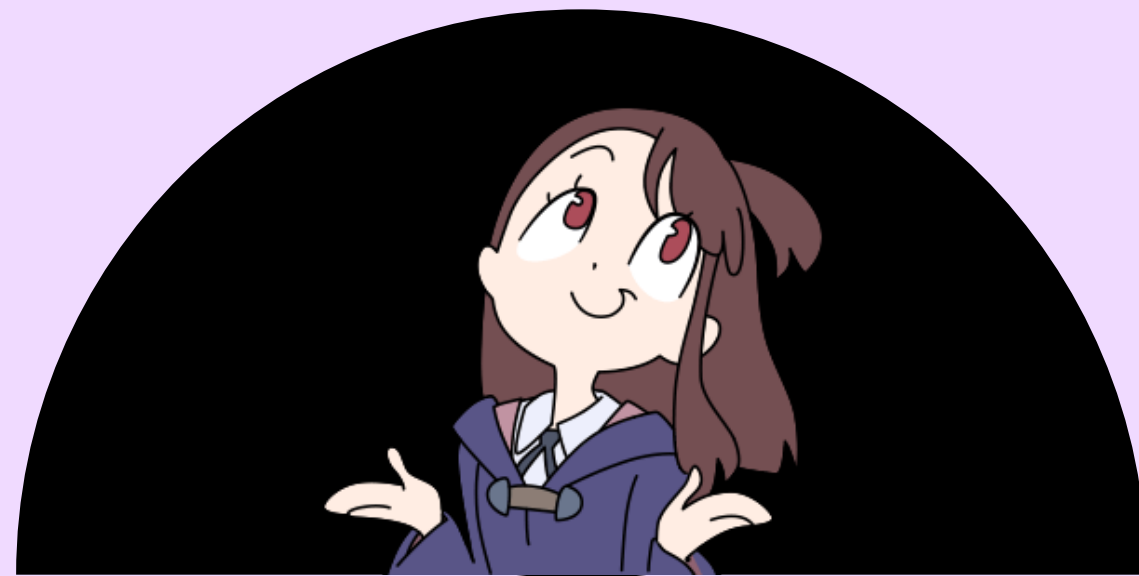




ANIMELIST DATASET

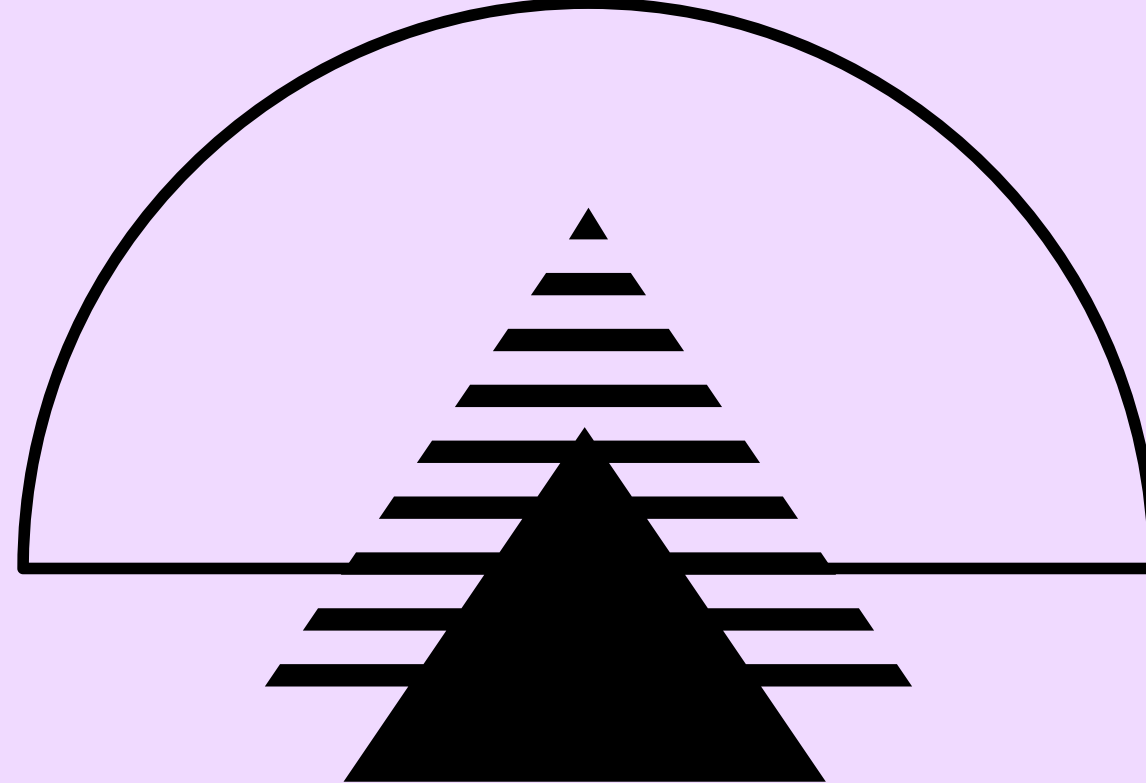
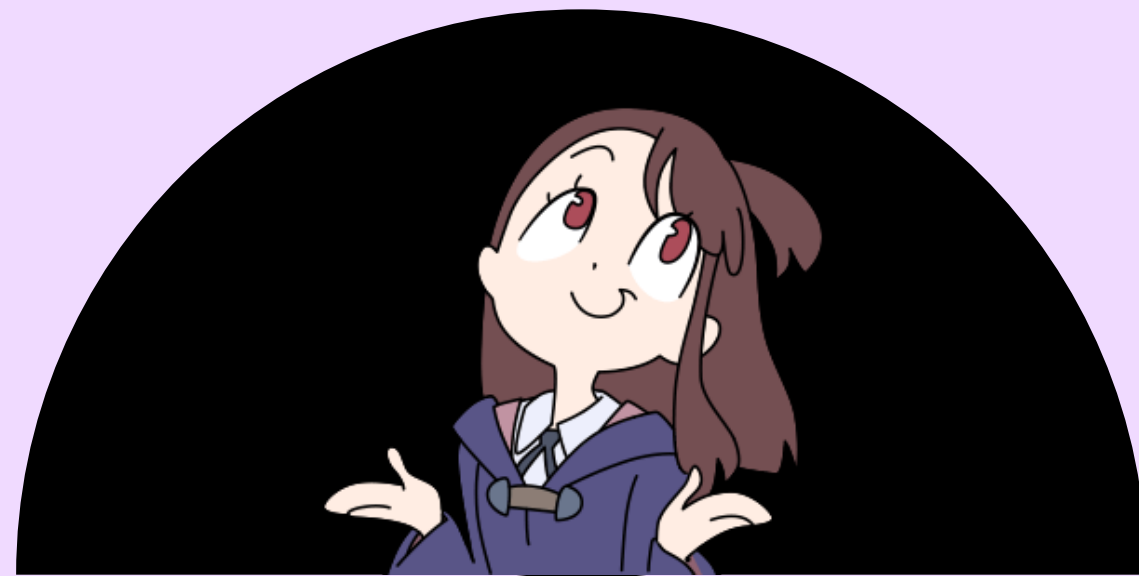
Integrantes: Amaro Zurita
Javier Facondi
Marco Martinez
Mario Fuentes
Pedro Escobar





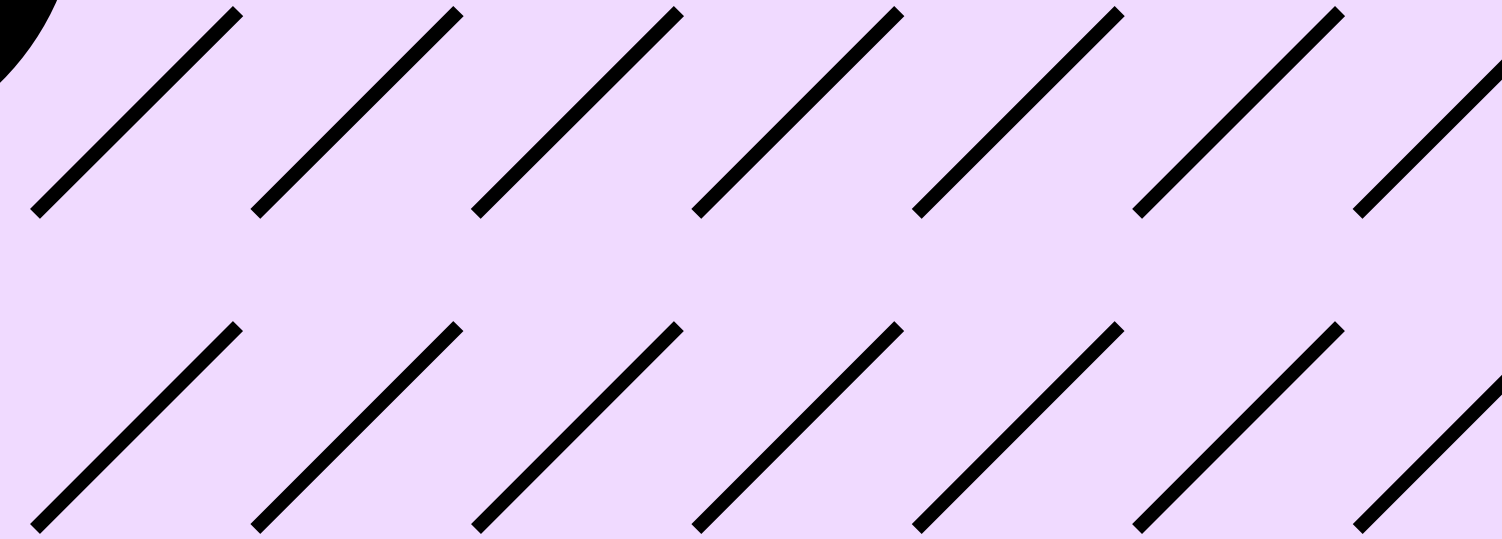
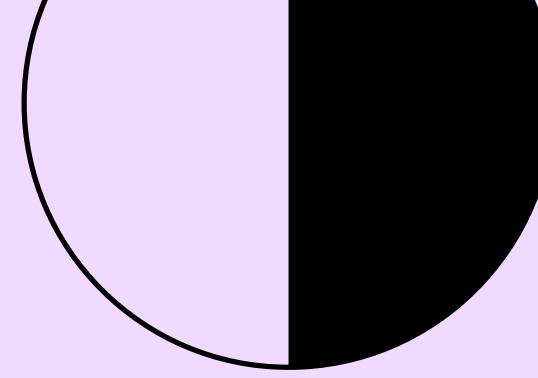
Motivacion

¿Cuál es el contexto general del tema/problema/datos de estudio?



Objetivos

¿Por qué podría ser
interesante estudiar estos
datos?



Preguntas y Problemas.

01

¿Qué características de un anime (género, estudio de animación, número de episodios) son más predictivas de una alta puntuación (mayor a 8) en MyAnimeList?

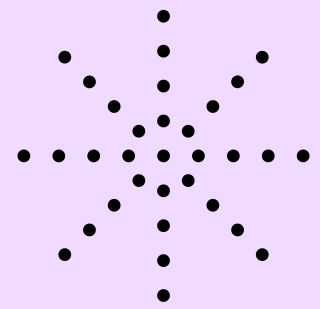
02

¿Existirá la posibilidad de predecir el estudio de animación con ciertas características de un anime?

03

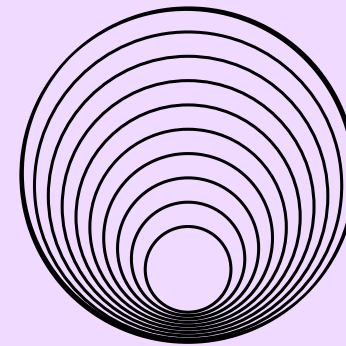
¿Existen patrones temporales (meses o años) en los que se lancen más animés de alta calidad (puntuación > 8)?

Exploracion de Datos

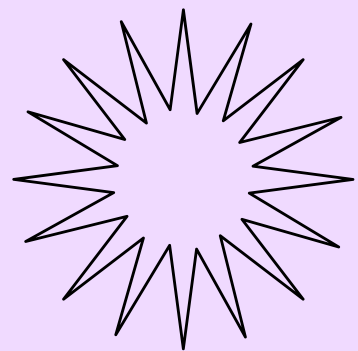


Filtracion de Generos

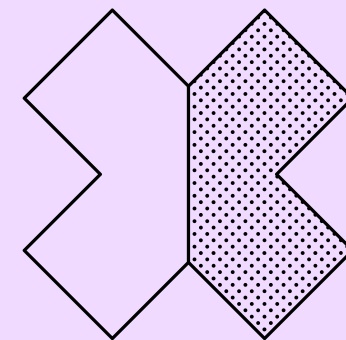
Filtracion de animes
“hentai”



Filtracion de OVA`s y
peliculas

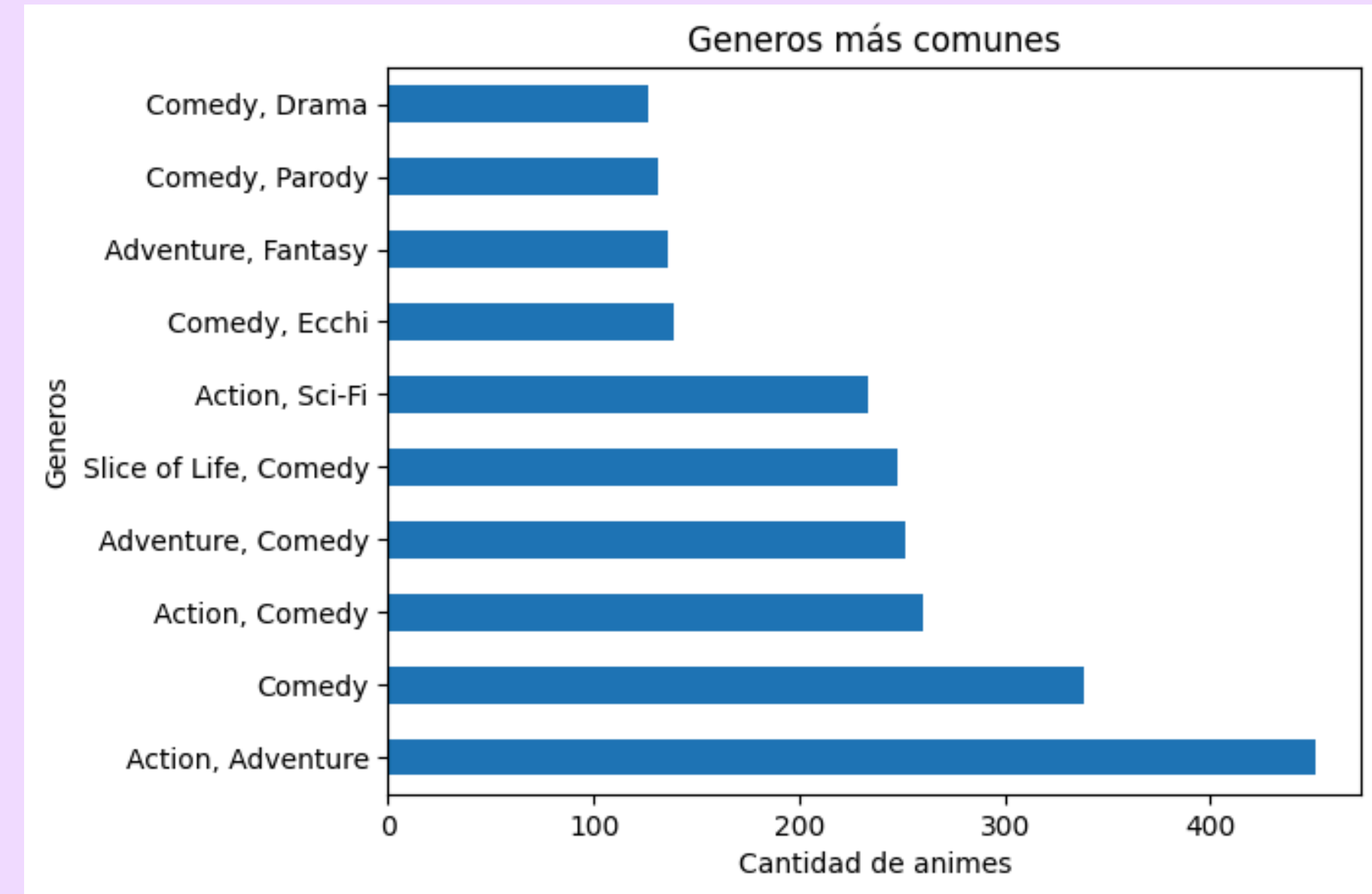
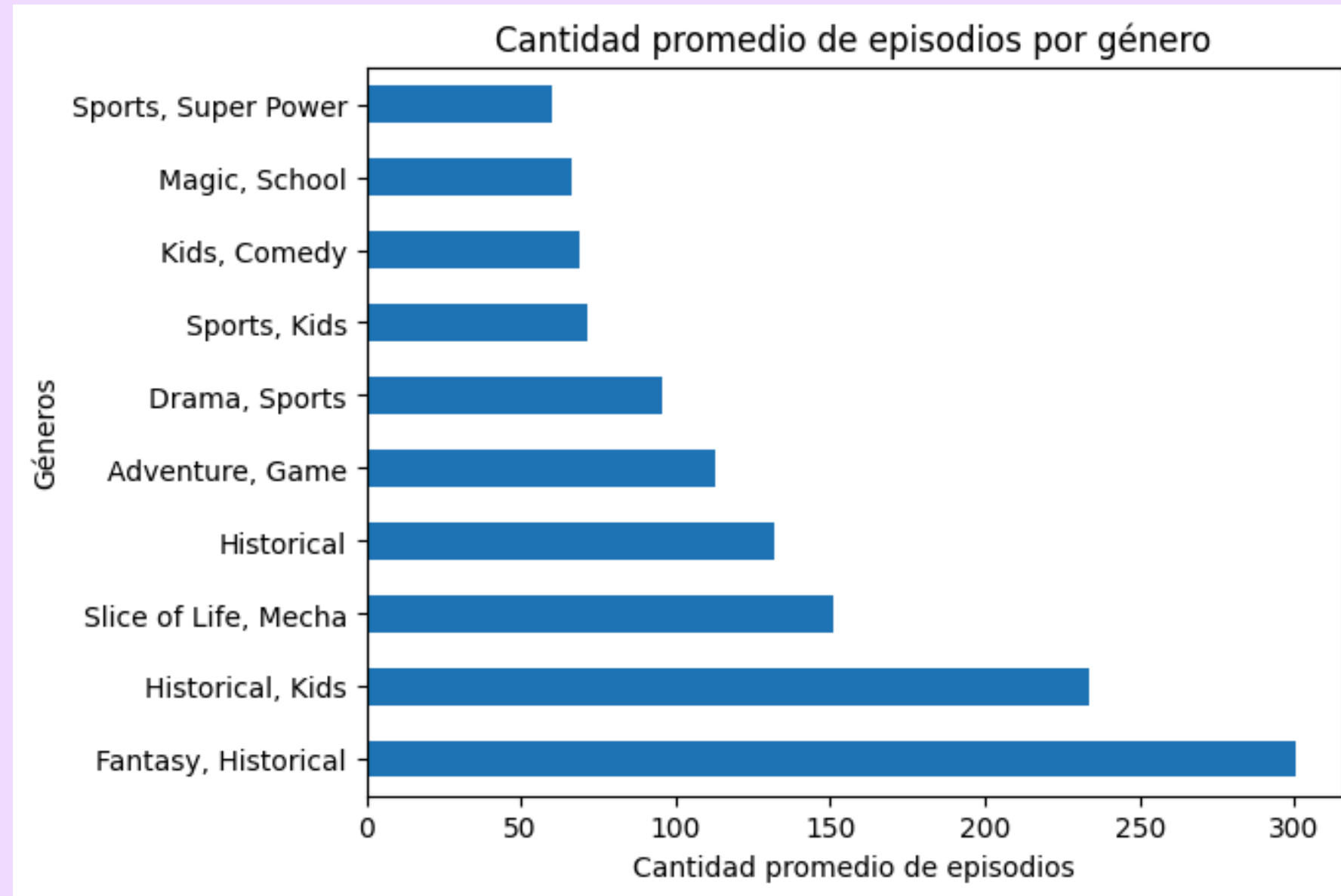


Filtración de Columnas

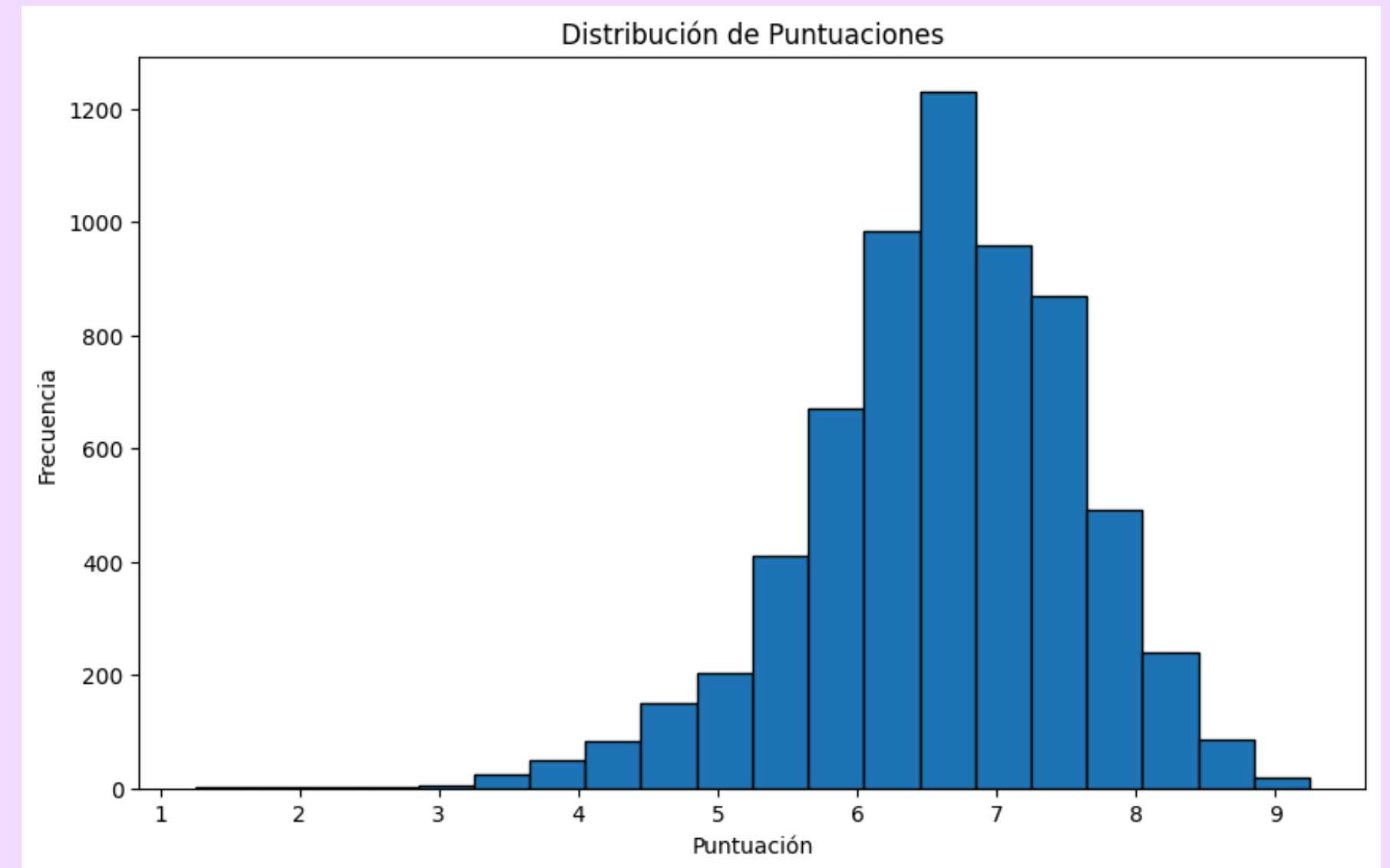
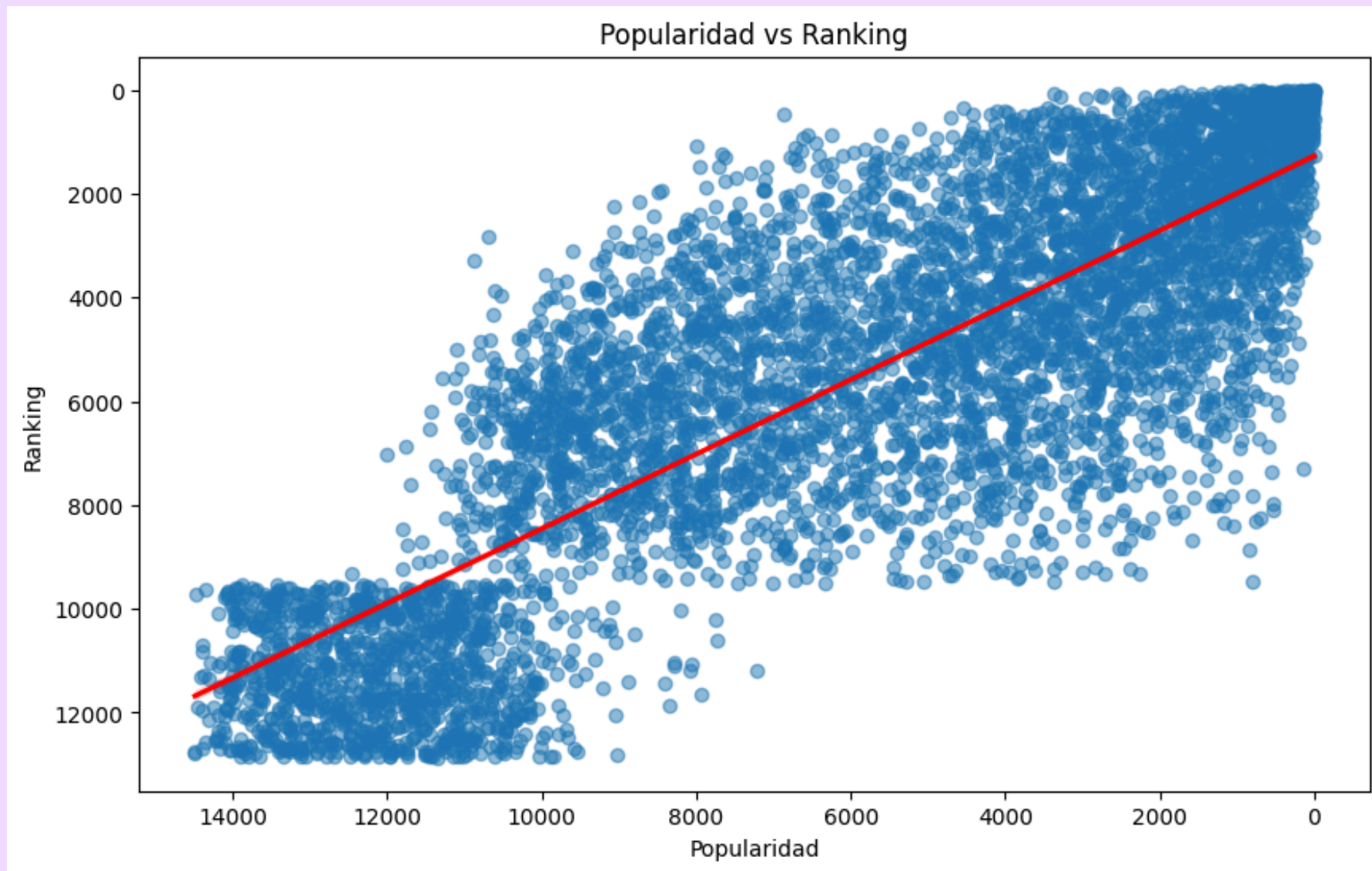


Filtración de animes que
aun no finalizan

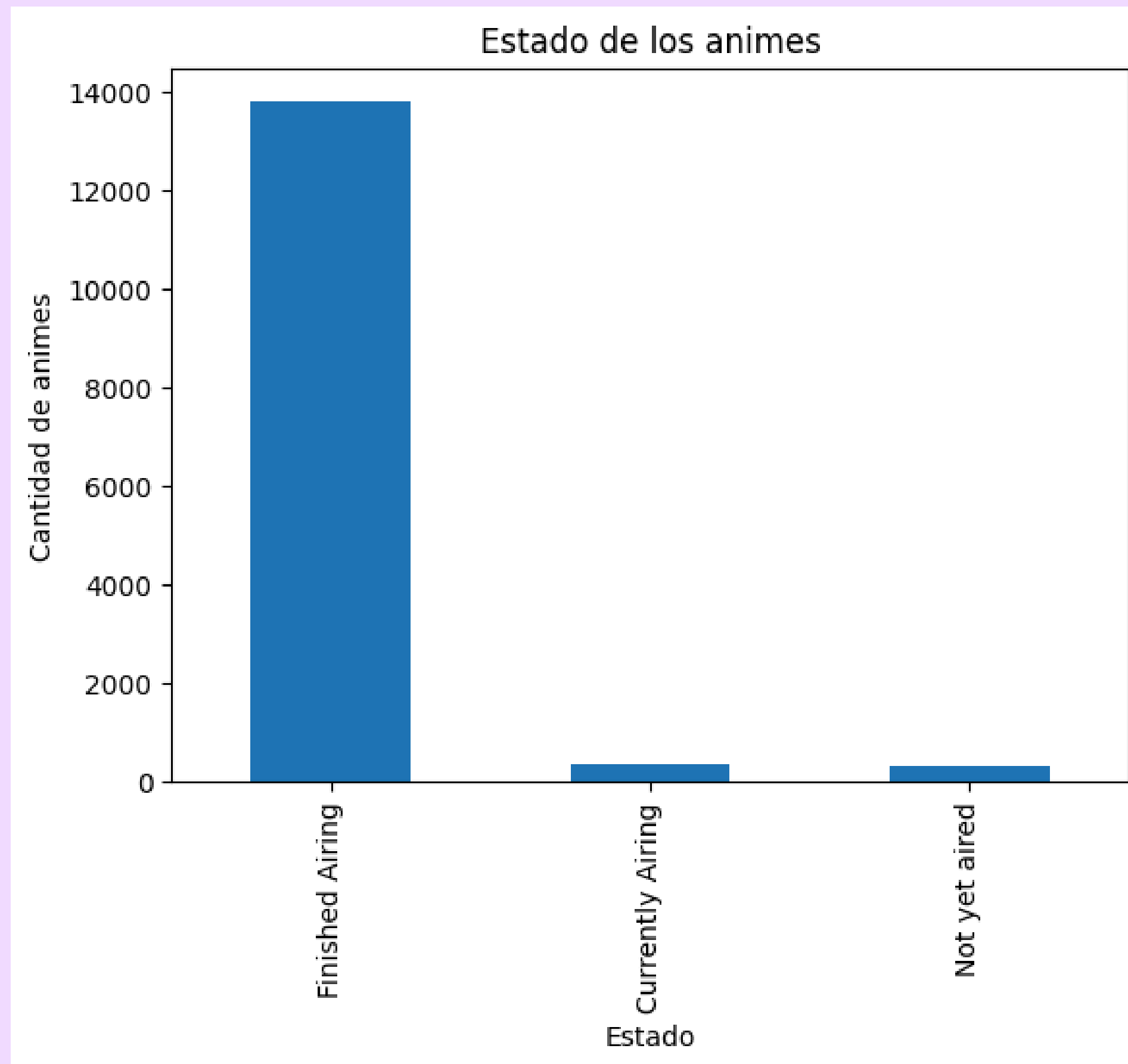
Exploracion de Datos

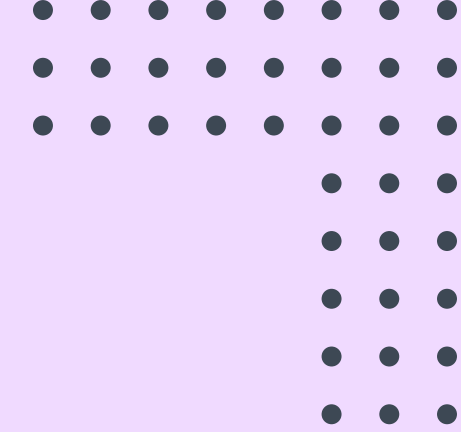
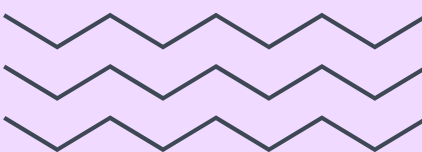


Exploracion de Datos



Exploracion de Datos





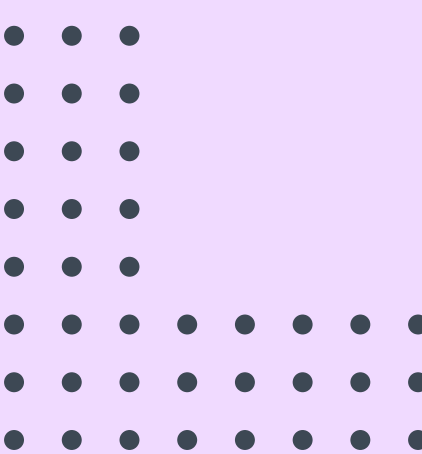
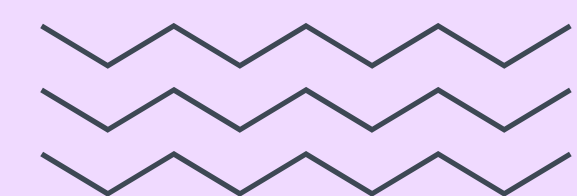
¿QUÉ CARACTERÍSTICAS DE UN ANIME (GÉNERO, ESTUDIO DE ANIMACIÓN, NÚMERO DE EPISODIOS) SON MÁS PREDICTIVAS DE UNA ALTA PUNTUACIÓN (MAYOR A 8) EN MYANIMELIST?

Metodología

Modelos a entrenar

- Decision Tree
- Random Forest
- Gradient Boosting

Atributos a usar

- Genero
 - Estudio de animación
 - Cantidad de episodios
- 
- 

¿QUÉ CARACTERÍSTICAS DE UN ANIME (GÉNERO, ESTUDIO DE ANIMACIÓN, NÚMERO DE EPISODIOS) SON MÁS PREDICTIVAS DE UNA ALTA PUNTUACIÓN (MAYOR A 8) EN MYANIMELIST?

¿Cuál es la importancia de los atributos?

Decision Tree

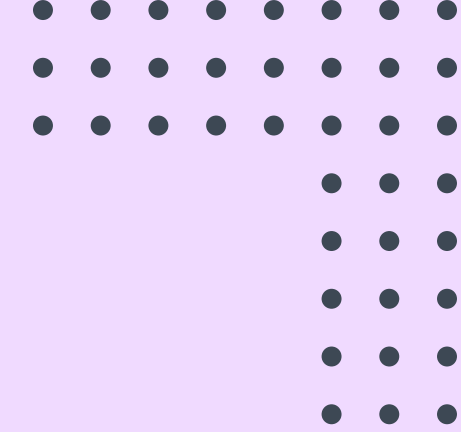
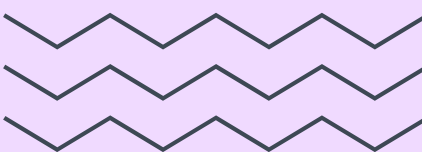
Característica	Importancia
genre_encoded	0.5667
studio_encoded	0.2846
normalized_episodes	0.1486

Random Forest

Característica	Importancia
genre_encoded	0.4986
studio_encoded	0.3204
episodes	0.1811

Gradient Boosting

Característica	Importancia
genre_encoded	0.4613
studio_encoded	0.3093
episodes	0.2294

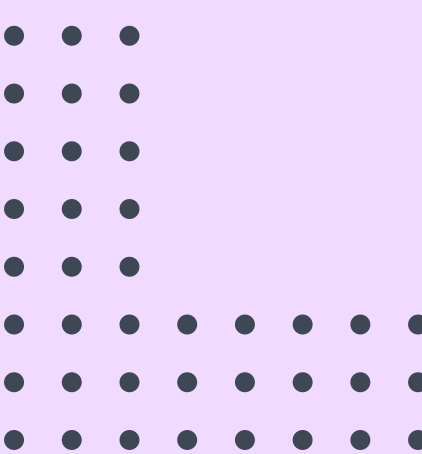
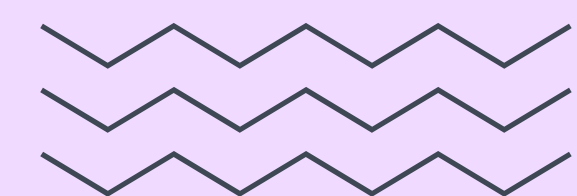


¿QUÉ CARACTERÍSTICAS DE UN ANIME (GÉNERO, ESTUDIO DE ANIMACIÓN, NÚMERO DE EPISODIOS) SON MÁS PREDICTIVAS DE UNA ALTA PUNTUACIÓN (MAYOR A 8) EN MYANIMELIST?

Intento de mejorar los resultados

Técnica SMOTE: ayuda a equilibrar las clases en los datos de entrenamiento.
Mejora mínima en identificación de animés con altas puntuaciones,
y una reducción en la precisión general de las predicciones.

¿Resultados?

1. Genero
 2. Estudio de animación
 3. Cantidad de episodios
- 
- 



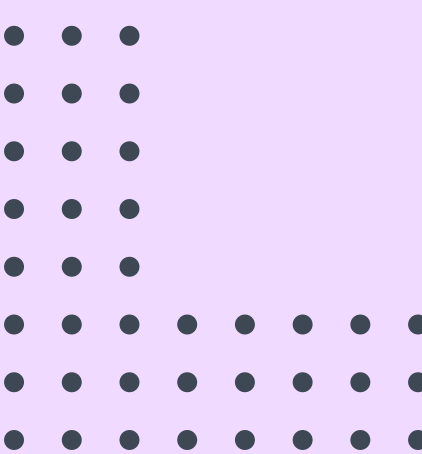
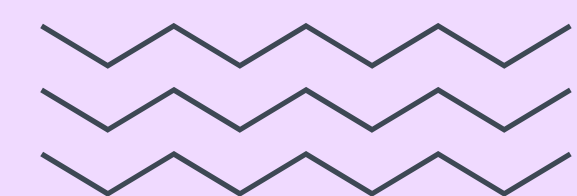
¿EXISTIRÁ LA POSIBILIDAD DE PREDECIR EL ESTUDIO DE ANIMACIÓN CON CIERTAS CARACTERÍSTICAS DE UN ANIME?:

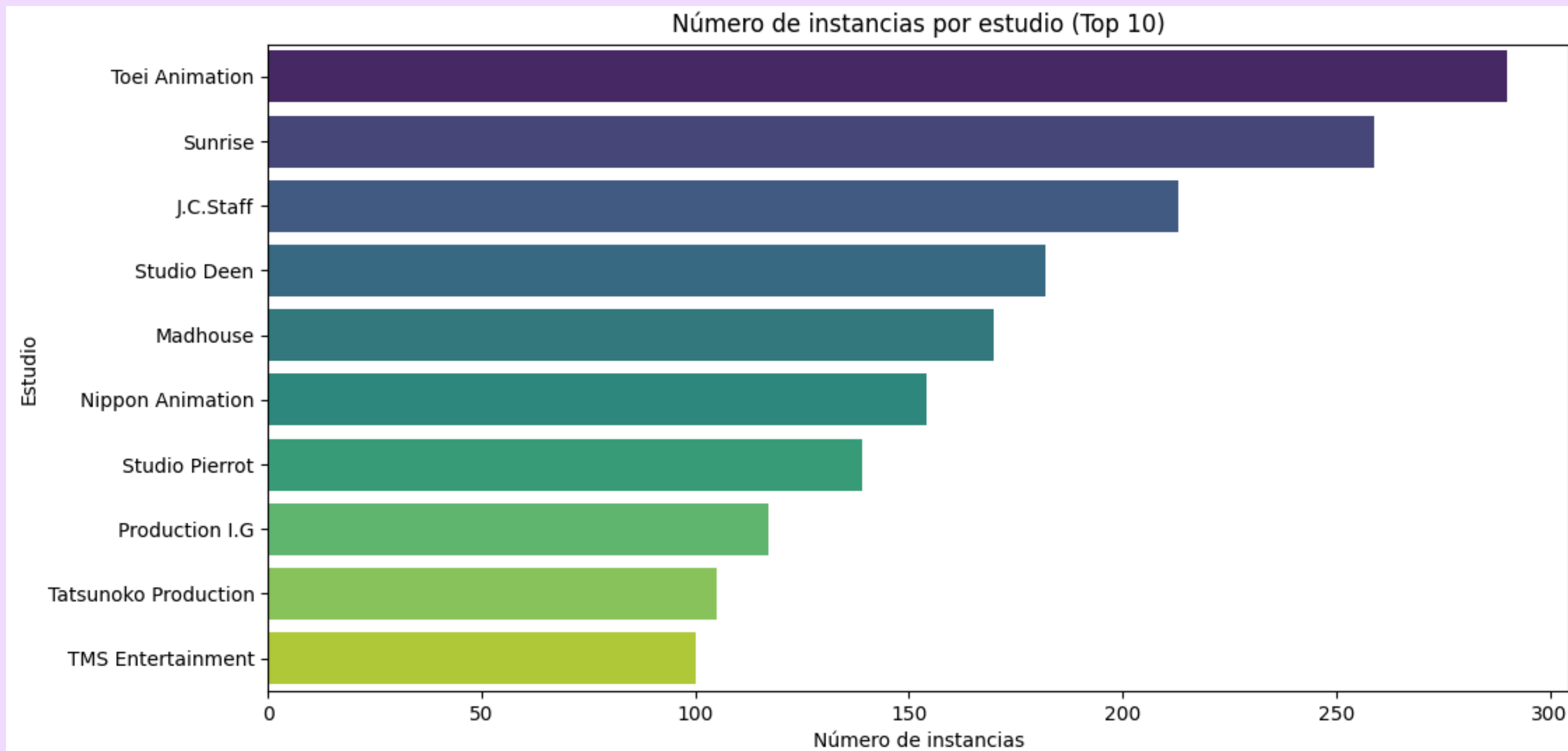
Metodología

Modelos a entrenar

- Decision Tree
- KNeighbors
- GaussianNB
- Support Vector Machine

Atributos clave a usar

- Genero
 - Puntuación
 - Duración de episodios
 - Ranking de animes
 - Popularidad
- 
- 



¿EXISTIRÁ LA POSIBILIDAD DE PREDECIR EL ESTUDIO DE ANIMACIÓN CON CIERTAS CARACTERÍSTICAS DE UN ANIME?:

Resultados Decision Tree

Top 10

	precision	recall	f1-score
J.C.Staff	0.28	0.24	0.26
Madhouse	0.10	0.09	0.09
Nippon Animation	0.25	0.29	0.27
Production I.G	0.14	0.13	0.14
Studio Deen	0.09	0.08	0.08
Studio Pierrot	0.07	0.07	0.07
Sunrise	0.30	0.37	0.33
TMS Entertainment	0.16	0.15	0.15
Tatsunoko Production	0.06	0.05	0.05
Toei Animation	0.28	0.28	0.28
accuracy			0.20
macro avg	0.17	0.17	0.17
weighted avg	0.20	0.20	0.20

Top 5

Estudio	Precision	Recall	F1-Score
J.C.Staff	0.40	0.43	0.41
Madhouse	0.21	0.18	0.19
Studio Deen	0.28	0.28	0.28
Sunrise	0.39	0.42	0.41
Toei Animation	0.48	0.47	0.47
accuracy			0.37
macro avg	0.35	0.35	0.35
weighted avg	0.37	0.37	0.37

¿EXISTIRÁ LA POSIBILIDAD DE PREDECIR EL ESTUDIO DE ANIMACIÓN CON CIERTAS CARACTERÍSTICAS DE UN ANIME?:

Resultados KNeighbors

Top 10

	precision	recall	f1-score
J.C.Staff	0.24	0.48	0.32
Madhouse	0.09	0.12	0.10
Nippon Animation	0.14	0.19	0.16
Production I.G	0.16	0.17	0.17
Studio Deen	0.28	0.22	0.25
Studio Pierrot	0.15	0.07	0.10
Sunrise	0.14	0.13	0.14
TMS Entertainment	0.00	0.00	0.00
Tatsunoko Production	0.09	0.05	0.06
Toei Animation	0.17	0.14	0.15
accuracy			0.17
macro avg	0.15	0.16	0.15
weighted avg	0.16	0.17	0.16

Top 5

Estudio	Precision	Recall	F1-Score
J.C.Staff	0.33	0.48	0.39
Madhouse	0.10	0.09	0.10
Studio Deen	0.42	0.36	0.39
Sunrise	0.33	0.35	0.34
Toei Animation	0.45	0.36	0.40
accuracy			0.34
macro avg	0.33	0.33	0.32
weighted avg	0.34	0.34	0.34

¿EXISTIRÁ LA POSIBILIDAD DE PREDECIR EL ESTUDIO DE ANIMACIÓN CON CIERTAS CARACTERÍSTICAS DE UN ANIME?:

Resultados GaussianNB

Top 10

	precision	recall	f1-score
J.C.Staff	0.17	0.19	0.18
Madhouse	0.24	0.12	0.16
Nippon Animation	0.12	0.74	0.21
Production I.G	0.11	0.04	0.06
Studio Deen	0.08	0.03	0.04
Studio Pierrot	0.00	0.00	0.00
Sunrise	0.25	0.02	0.04
TMS Entertainment	0.00	0.00	0.00
Tatsunoko Production	0.06	0.05	0.05
Toei Animation	0.12	0.09	0.10
accuracy			0.13
macro avg	0.11	0.13	0.08
weighted avg	0.13	0.13	0.09

Top 5

Estudio	Precision	Recall	F1-Score
J.C.Staff	0.32	0.21	0.26
Madhouse	0.24	0.12	0.16
Studio Deen	0.06	0.03	0.04
Sunrise	0.35	0.15	0.21
Toei Animation	0.35	0.81	0.48
accuracy			0.31
macro avg	0.26	0.26	0.23
weighted avg	0.28	0.31	0.26

¿EXISTIRÁ LA POSIBILIDAD DE PREDECIR EL ESTUDIO DE ANIMACIÓN CON CIERTAS CARACTERÍSTICAS DE UN ANIME?:

Resultados Support Vector Machine

Top 10

	precision	recall	f1-score
J.C.Staff	0.24	0.40	0.30
Madhouse	0.00	0.00	0.00
Nippon Animation	0.00	0.00	0.00
Production I.G	0.00	0.00	0.00
Studio Deen	0.00	0.00	0.00
Studio Pierrot	0.00	0.00	0.00
Sunrise	0.31	0.08	0.12
TMS Entertainment	0.00	0.00	0.00
Tatsunoko Production	0.00	0.00	0.00
Toei Animation	0.19	0.86	0.31
accuracy			0.21
macro avg	0.07	0.13	0.07
weighted avg	0.11	0.21	0.11

Top 5

Estudio	Precision	Recall	F1-Score
J.C.Staff	0.27	0.40	0.33
Madhouse	0.00	0.00	0.00
Studio Deen	0.00	0.00	0.00
Sunrise	0.00	0.00	0.00
Toei Animation	0.31	0.86	0.46
accuracy			0.30
macro avg	0.12	0.25	0.16
weighted avg	0.13	0.30	0.18



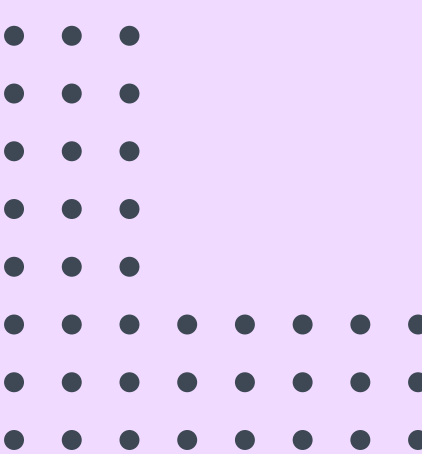
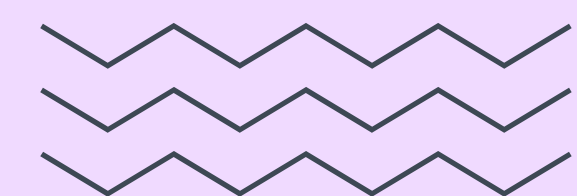
¿EXISTEN PATRONES TEMPORALES (MESES O AÑOS) EN LOS QUE SE LANCEN MÁS ANIMÉS DE ALTA CALIDAD (PUNTUACIÓN > 8)?

Metodología

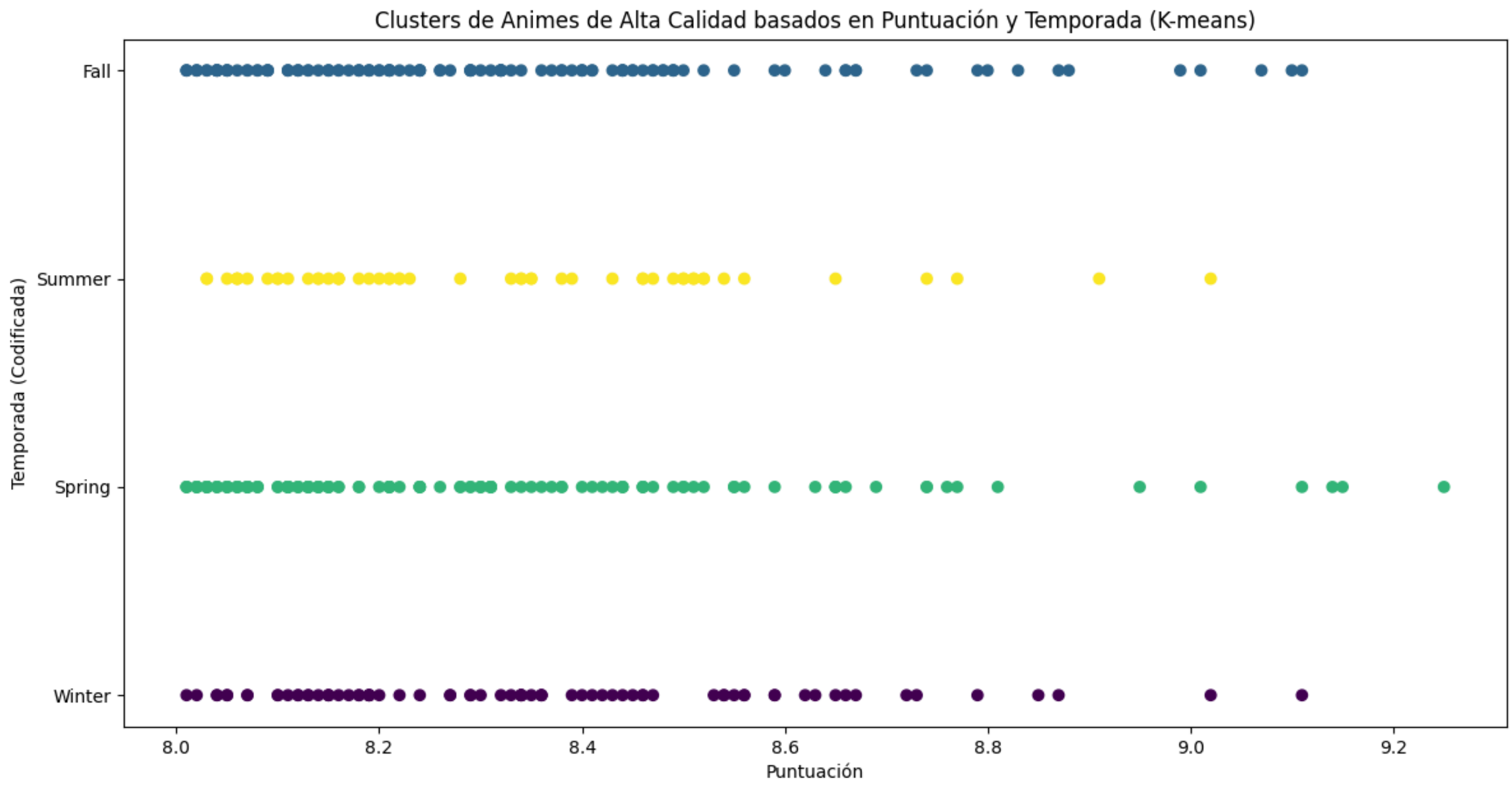
Métodos de Clustering

- K-Means
- Agglomerative
- DBSCAN

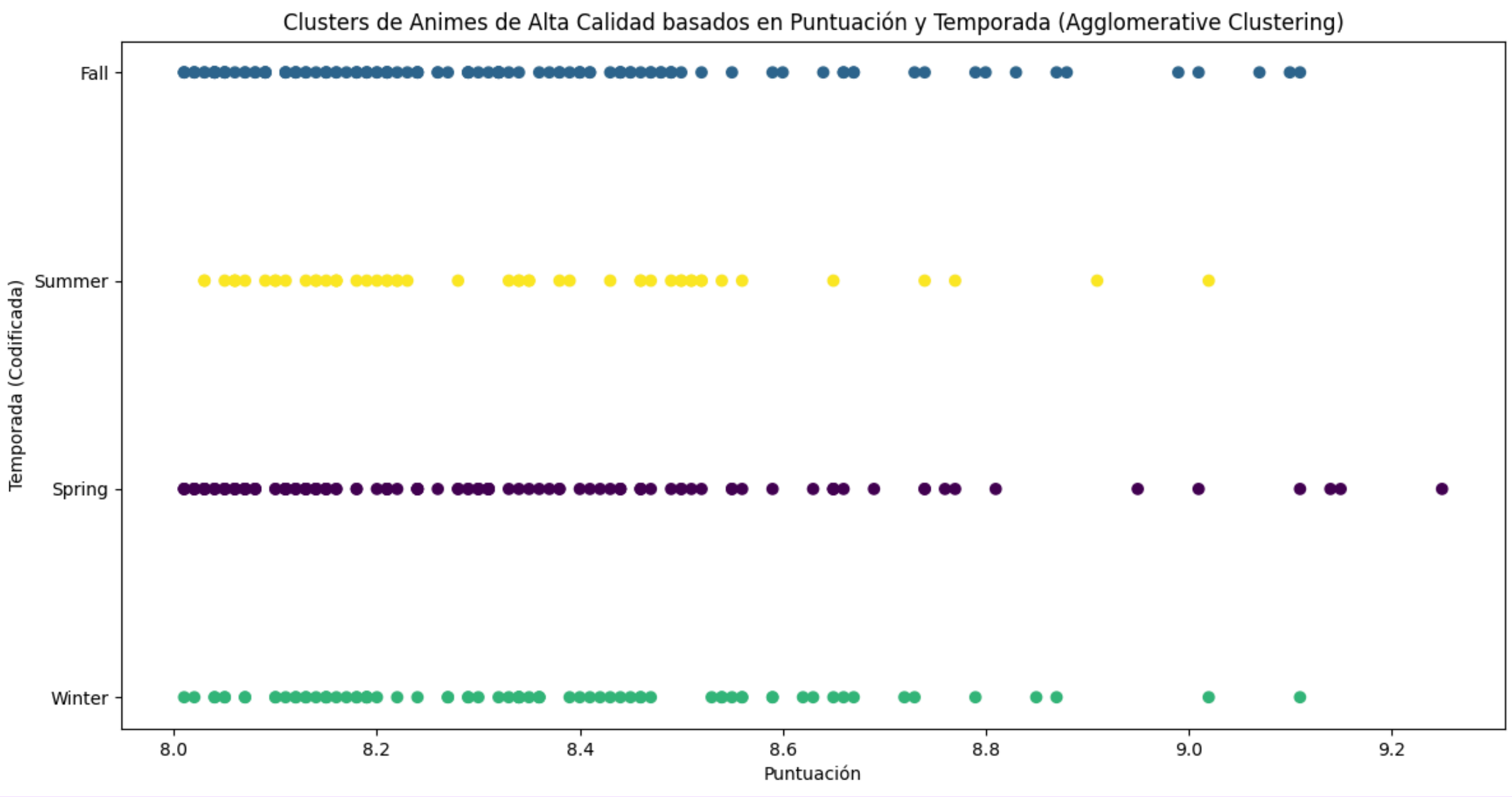
Atributos clave a usar

- Fechas de lanzamiento
 - Puntuación
 - Temporada
- 
- 

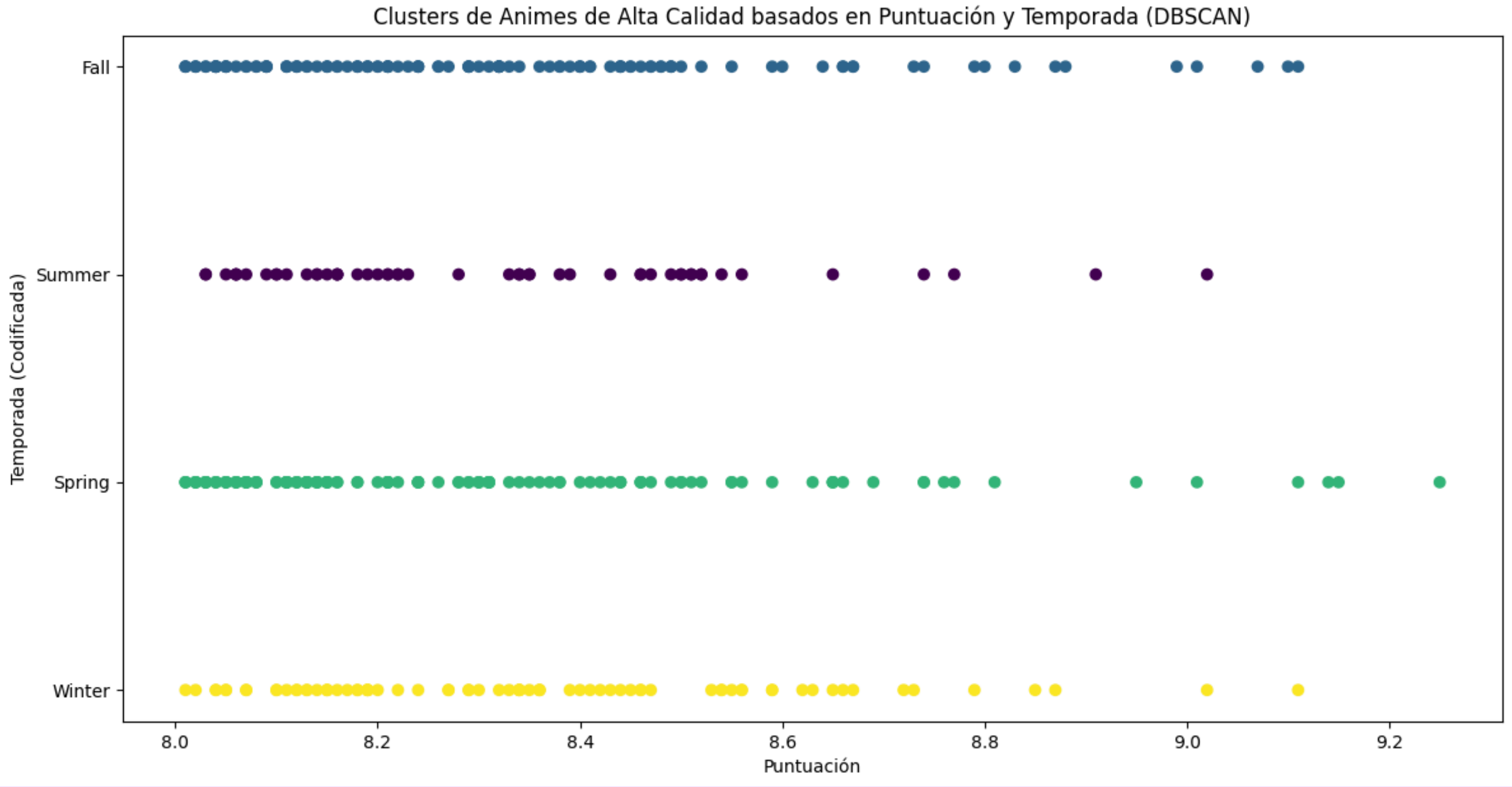
PUNTUACIÓN VS TEMPORADA



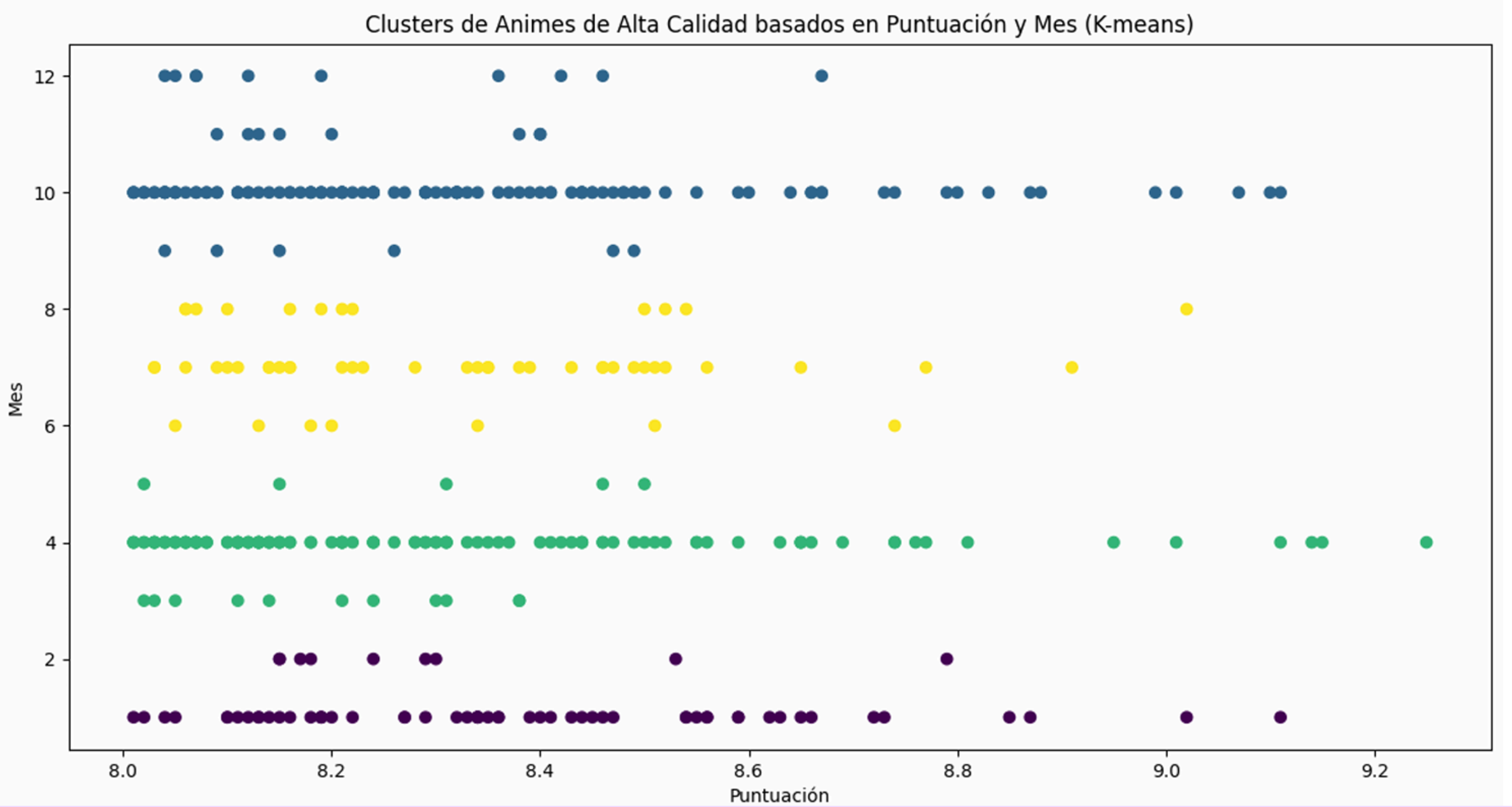
PUNTUACIÓN VS TEMPORADA



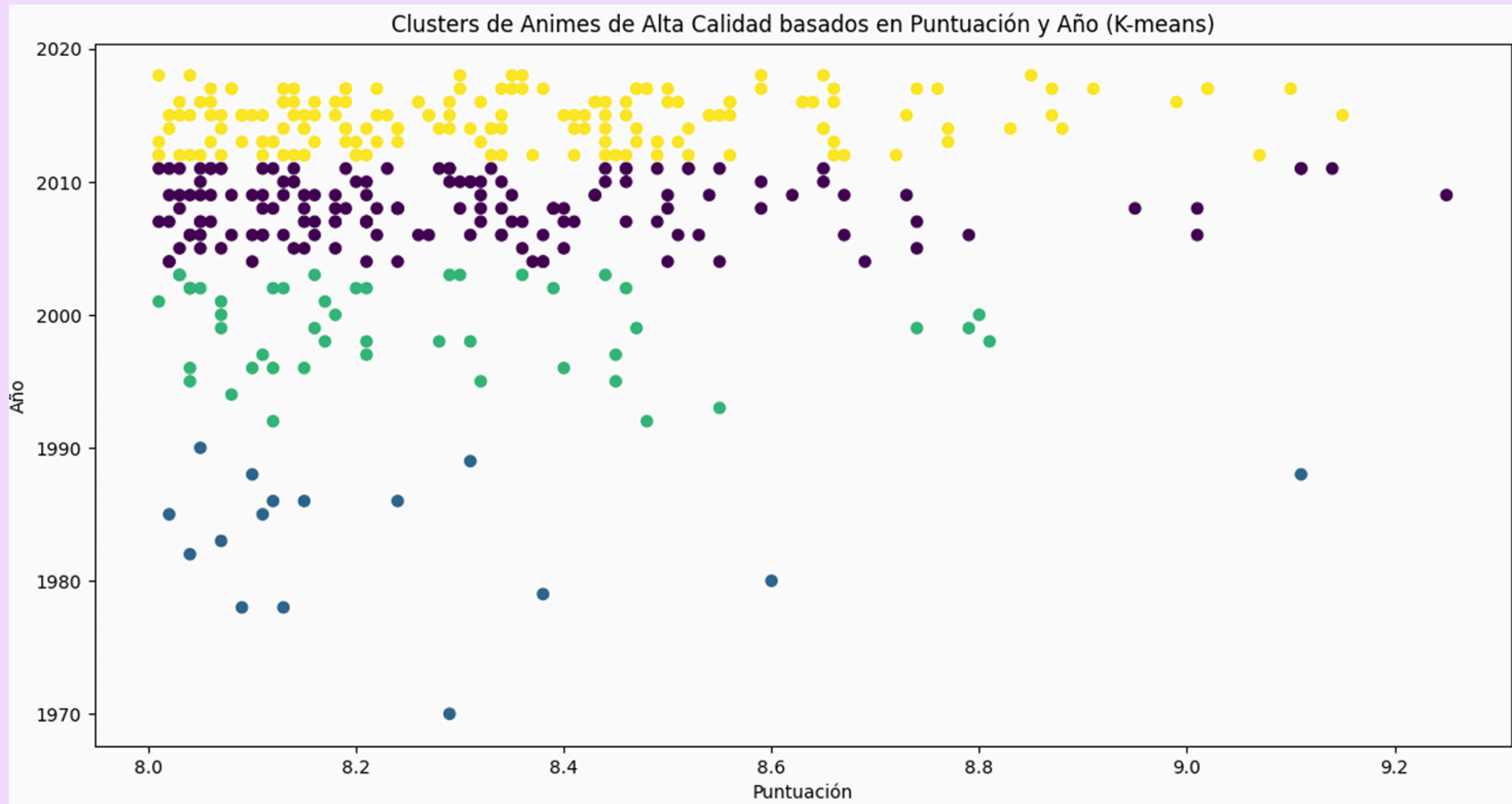
PUNTUACIÓN VS TEMPORADA



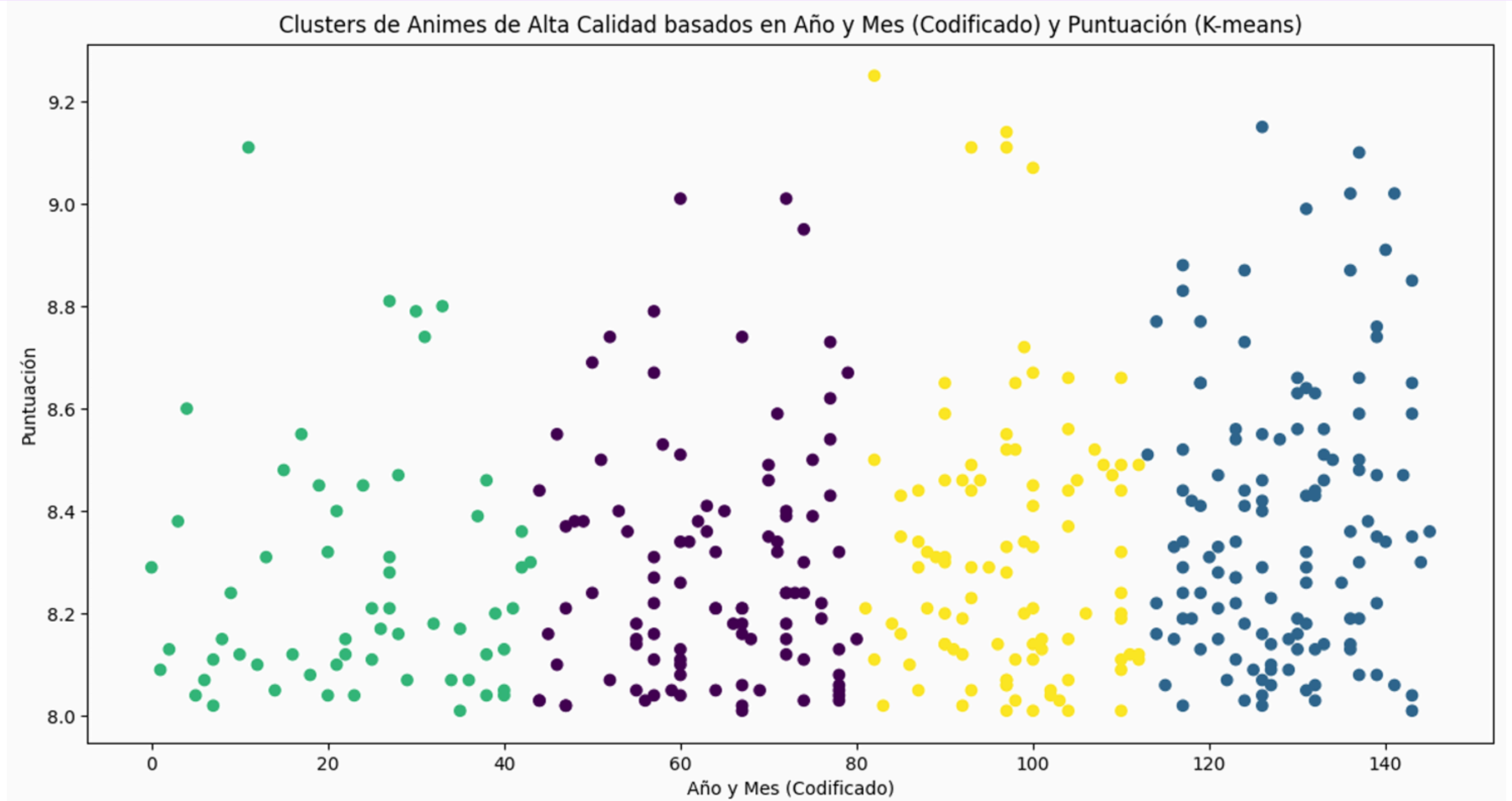
PUNTUACIÓN VS MES



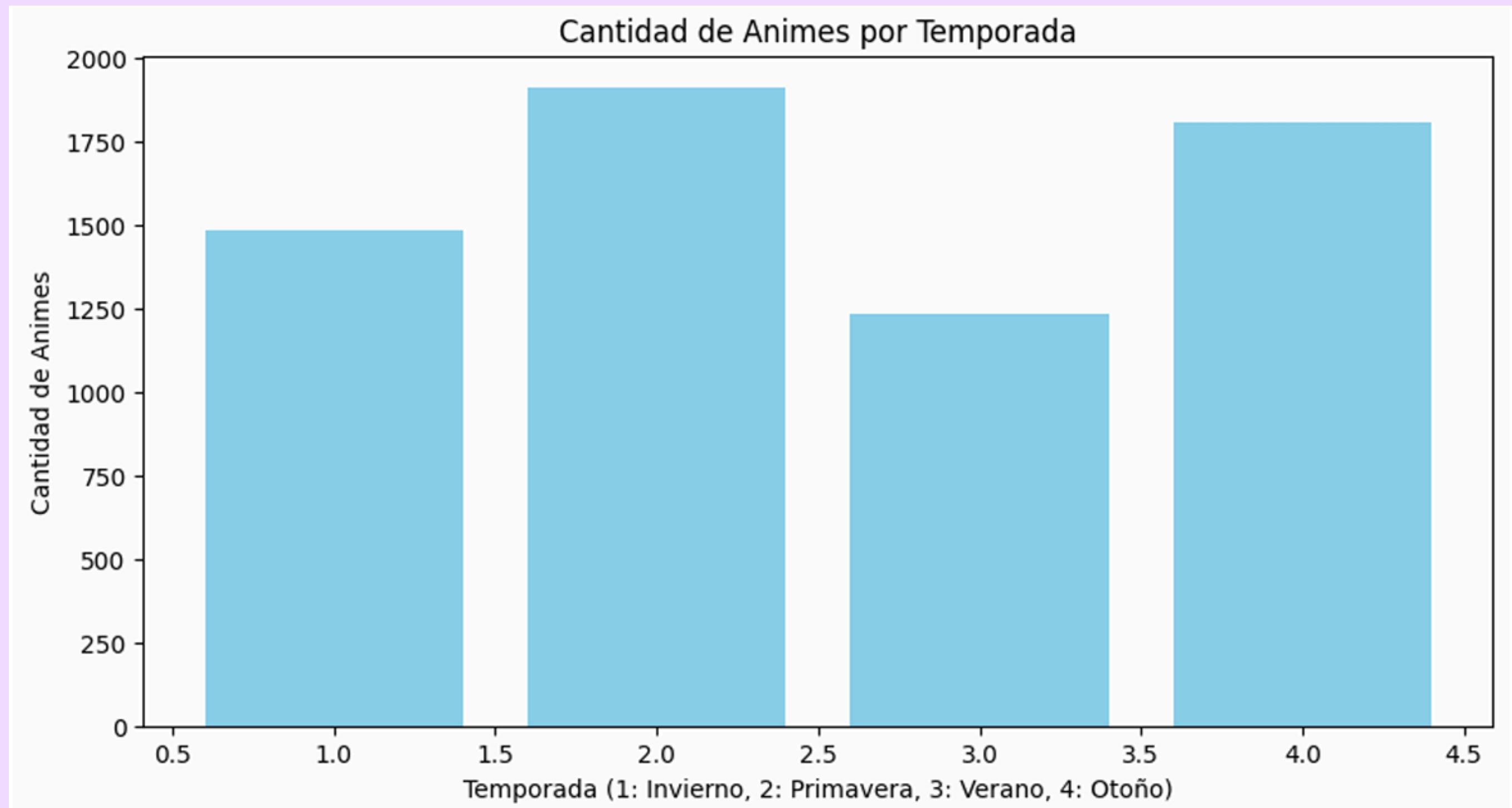
PUNTUACIÓN VS AÑO



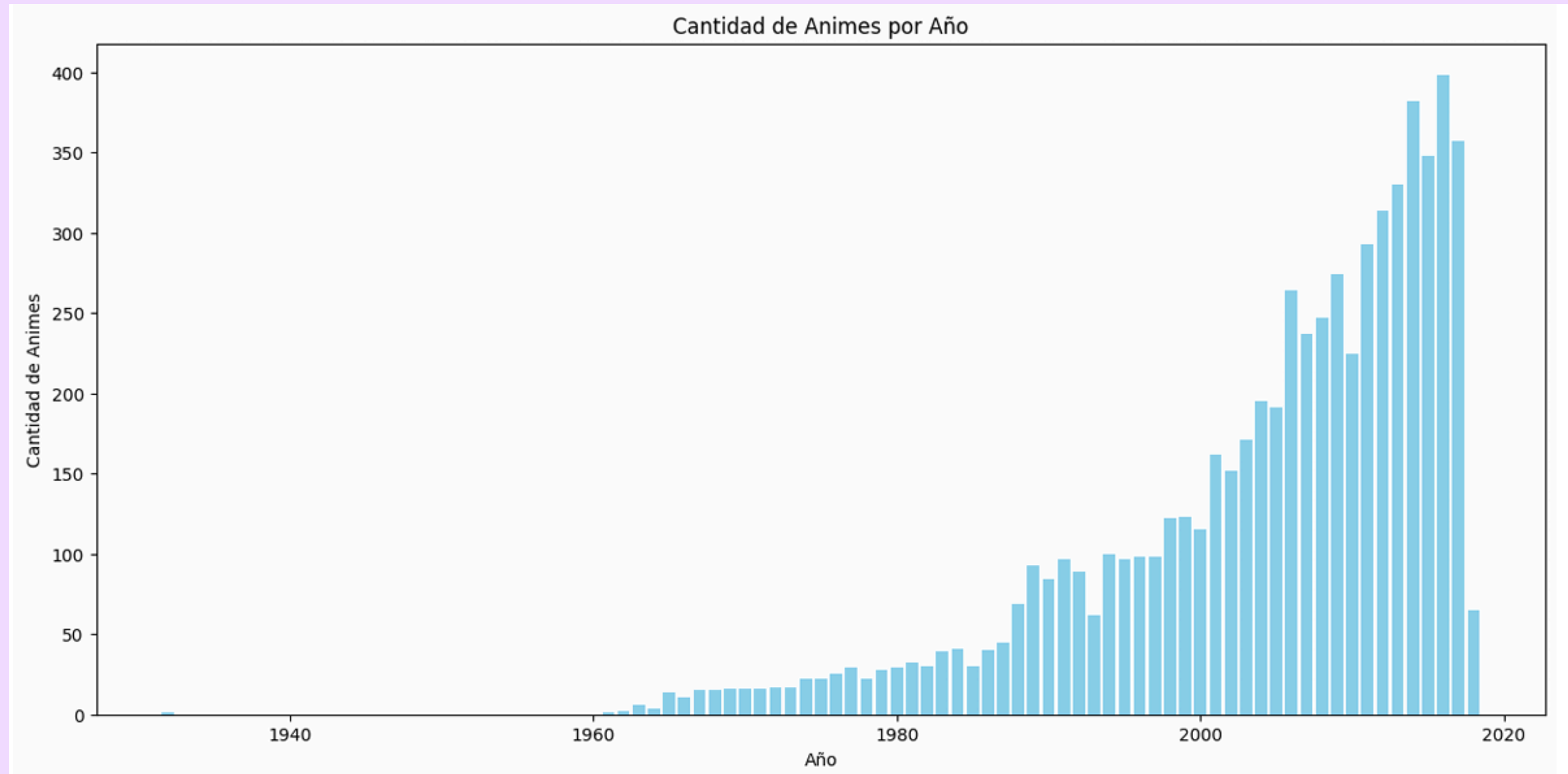
PUNTUACIÓN VS AÑO Y MES



CANTIDAD DE ANIMES POR TEMPORADA



CANTIDAD DE ANIMES POR AÑO

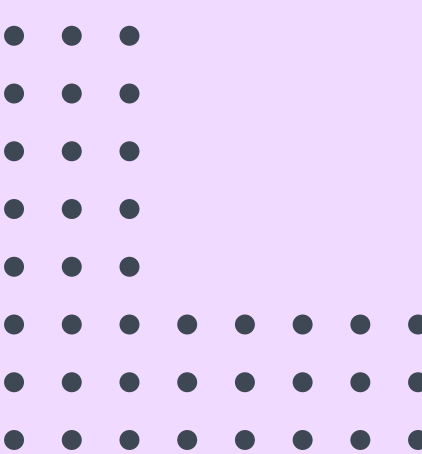
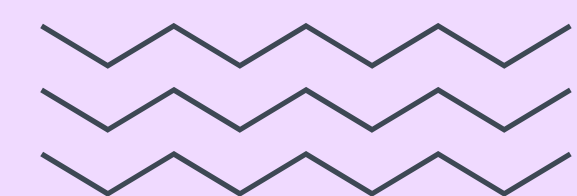




¿EXISTEN PATRONES TEMPORALES (MESES O AÑOS) EN LOS QUE SE LANCEN MÁS ANIMÉS DE ALTA CALIDAD (PUNTUACIÓN > 8)?

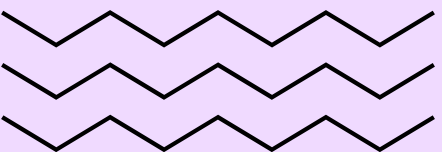
Resultados

Silhouette Score

- K-Means: 0.452
 - Agglomerative: 0.403
 - DBSCAN: No encontró suficientes clusters
- 
- 



ANIMELIST DATASET



Integrantes: Amaro Zurita
Javier Facondi
Marco Martinez
Mario Fuentes
Pedro Escobar

