



# Steam Games

# Analysis



Grupo 6 - CC5205 Integrantes:

Cristóbal Braga, Isabel Marx, Sofía Medina,  
Manuel Ortiz e Iván Salas



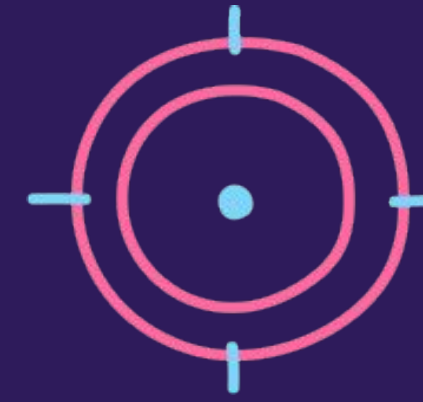
START

# Motivación

En un mundo donde la tecnología y la innovación están en constante evolución, la **industria de los videojuegos** emerge como uno de los sectores más **dinámicos y prometedores**.

Según el informe Gaming & Esports Report 2023 de YouGov, el **31% de la población mundial** juega activamente en diversas plataformas semanalmente





# Objetivo de Estudio

La **expansión de los videojuegos** genera la interrogante de ¿qué hace que un juego sea exitoso en un mercado tan diverso y competitivo?

El estudio del comportamiento de plataformas como Steam proporciona información sobre las **tendencias y las estrategias** que pueden llevar a un juego al éxito en este entorno digital en constante expansión.



# Preguntas y problemas abordados

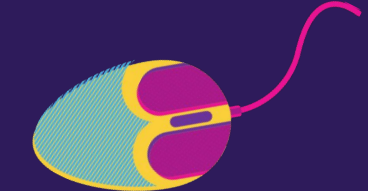
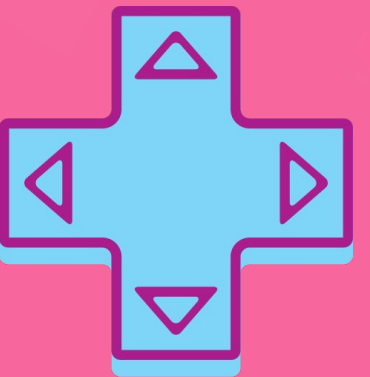
¿Qué es lo que hace que  
un juego sea bueno?

¿Cuál es el atributo más  
importante que define a  
un juego bueno?

¿Hay algún patrón  
que identifique a los  
mejores juegos?

¿Hay algún género que  
sea más exitoso que  
otros, hay una trama  
en común en los  
juegos exitosos?

¿Hay algún/os atributo/s  
que permita/n predecir si  
un juego será exitoso o  
no?



# Información de los Datos



## Indie

Género de juego más popular



## English

98,11% de los juegos del dataset están en inglés



## Windows

99,98% de los juegos corren en Windows

## Single Player

Categoría más popular con 25.678 menciones



## RPG

Género con mejor rating promedio



## Massively Multiplayer

Género con peor rating promedio

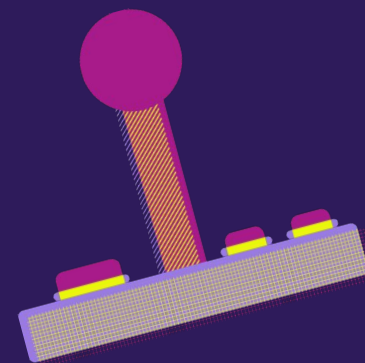


# Experimentos Realizados

**- Experimento Preliminar Hito 2: Clustering**

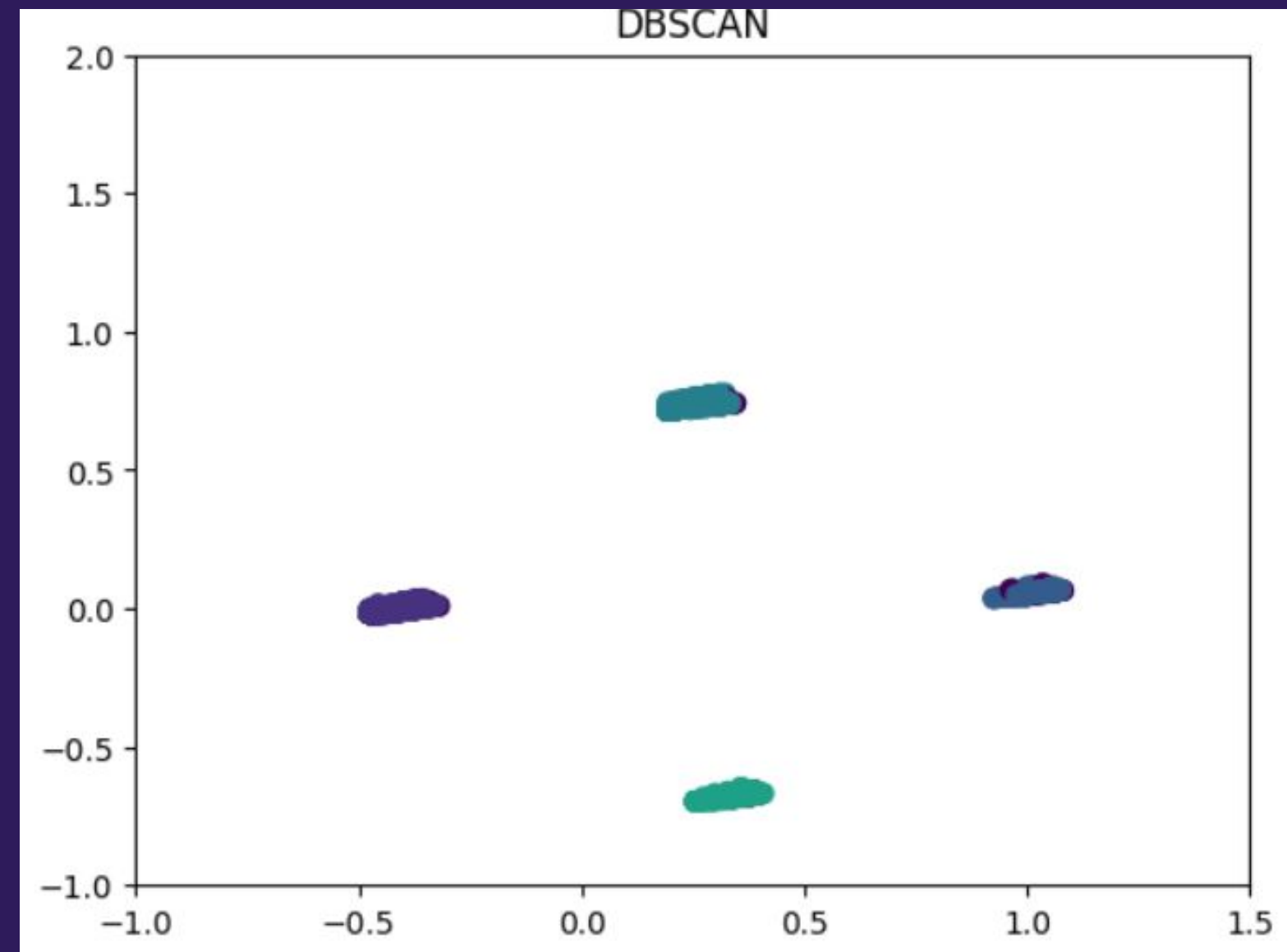
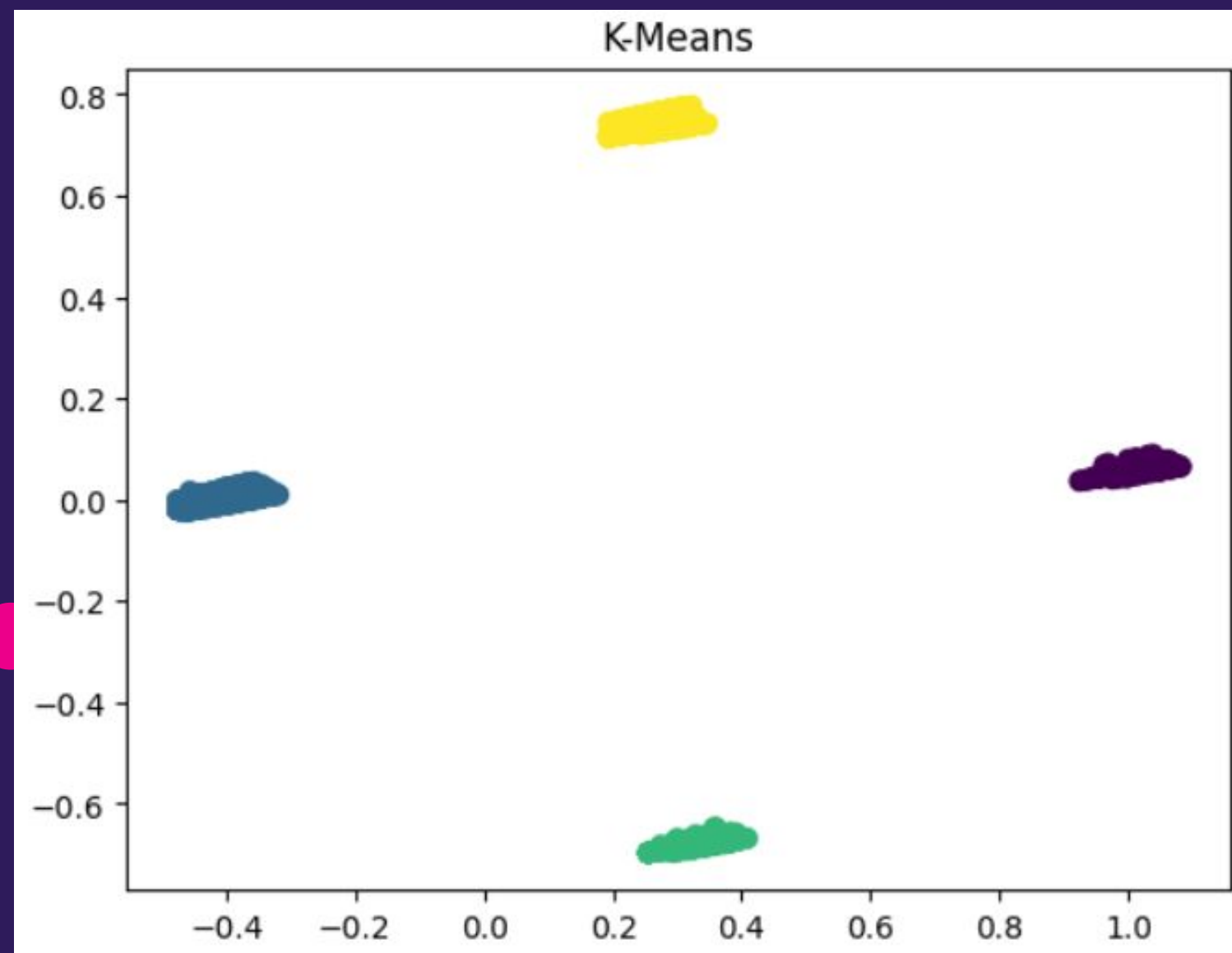
**- Experimentos Hito 3:**

- 1. Exploracion de Clusters**
- 2. Clasificación: Labels manuales**
- 3. Regresión y Clasificación del Rating en base a descripciones o Keyphrases de los juegos**



# Resultados

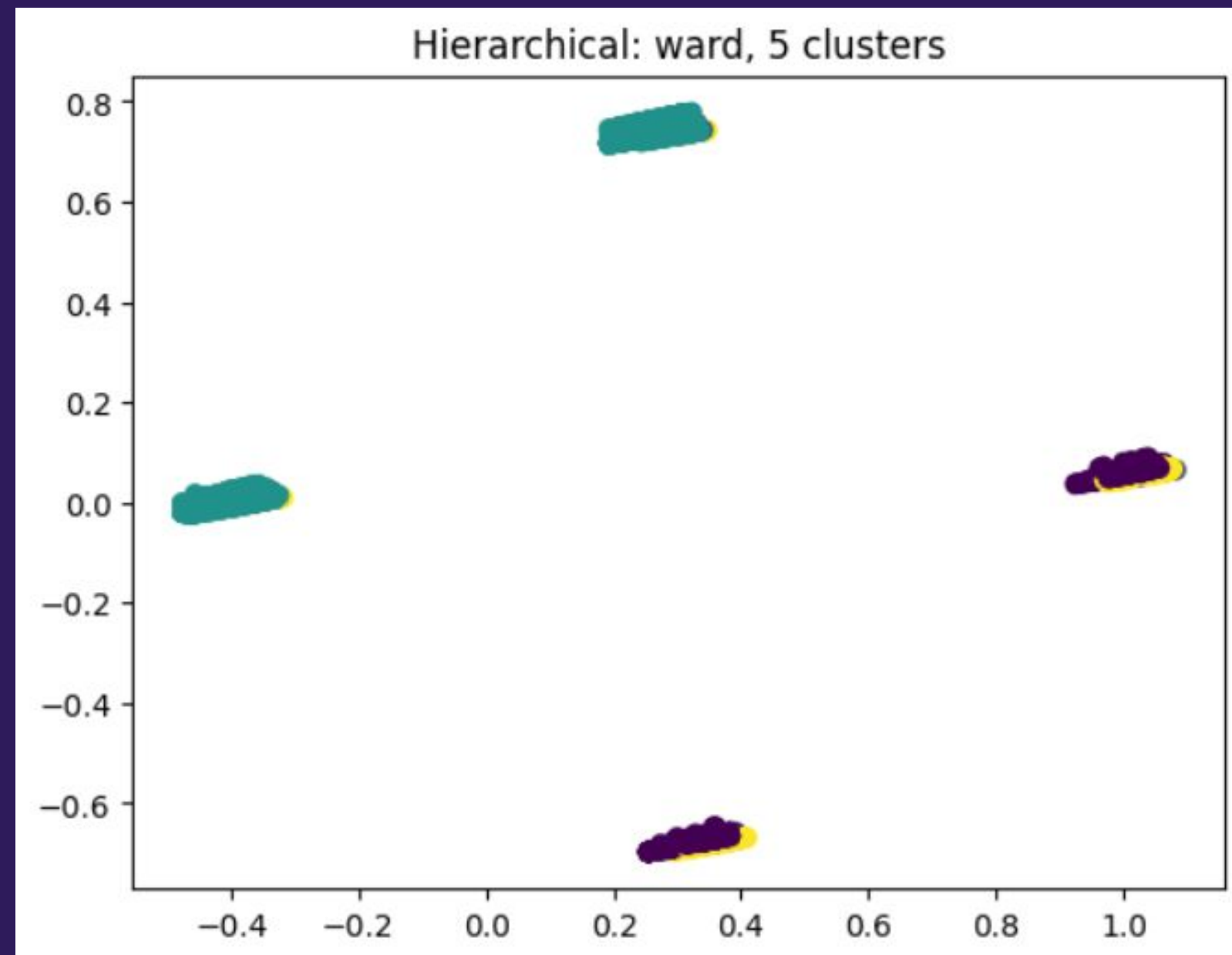
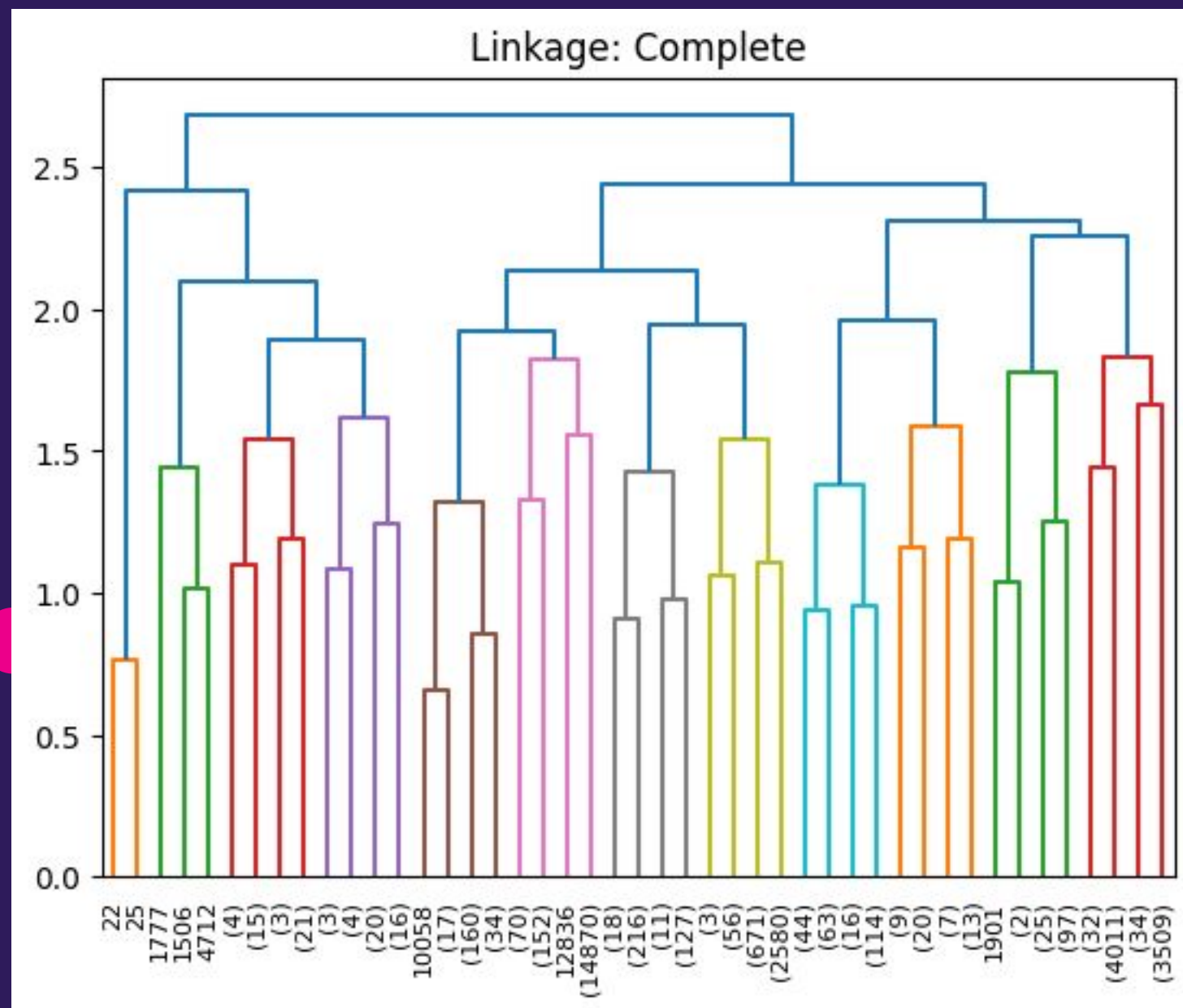
## Experimento Preliminar





# Resultados

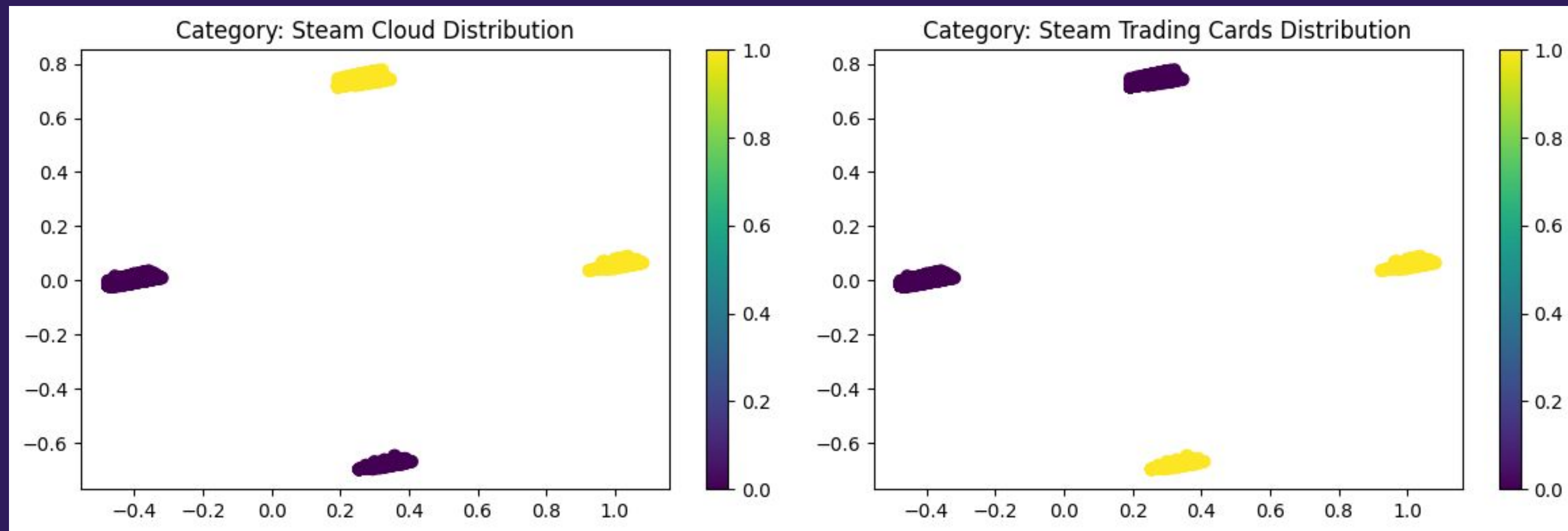
## Experimento Preliminar





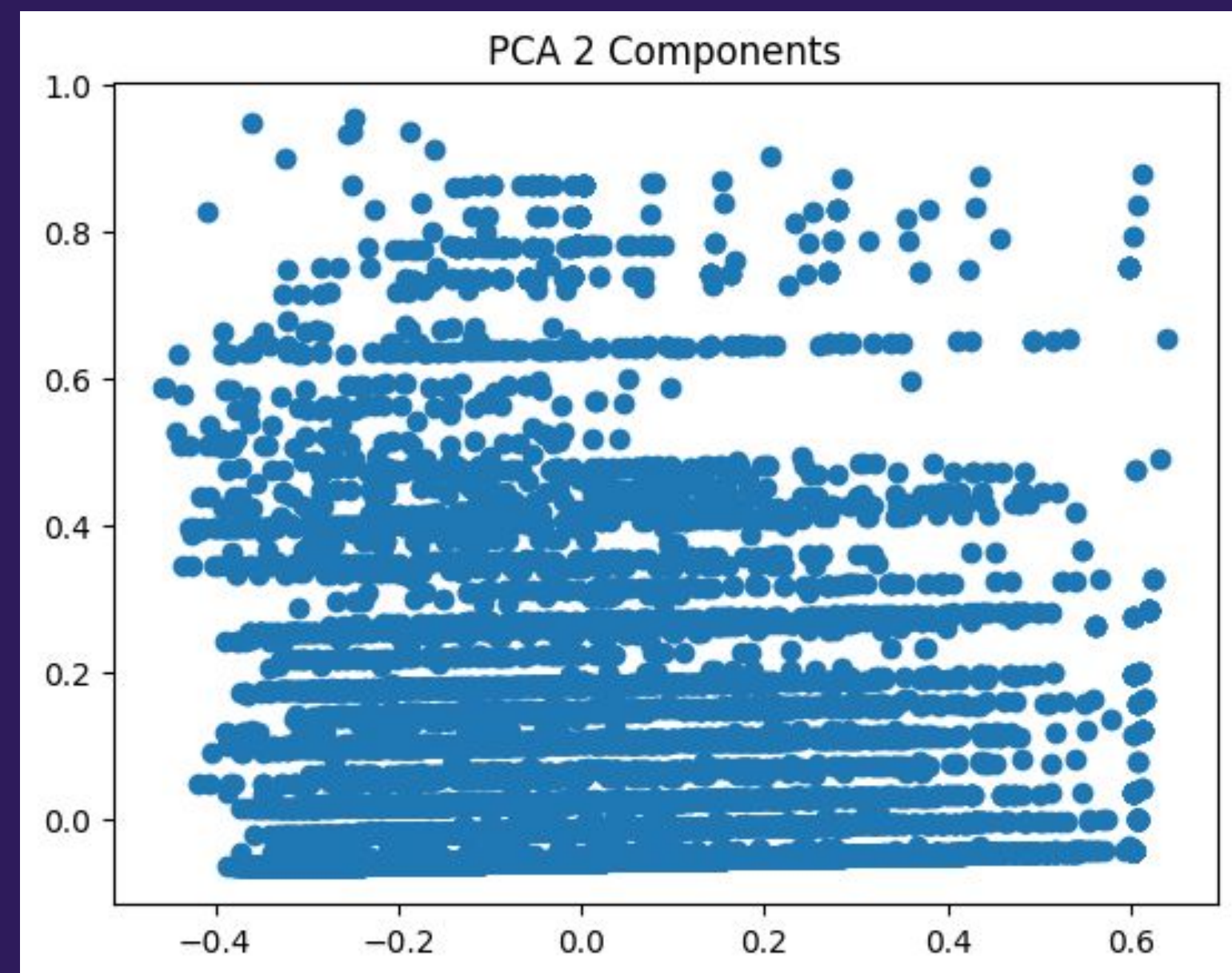
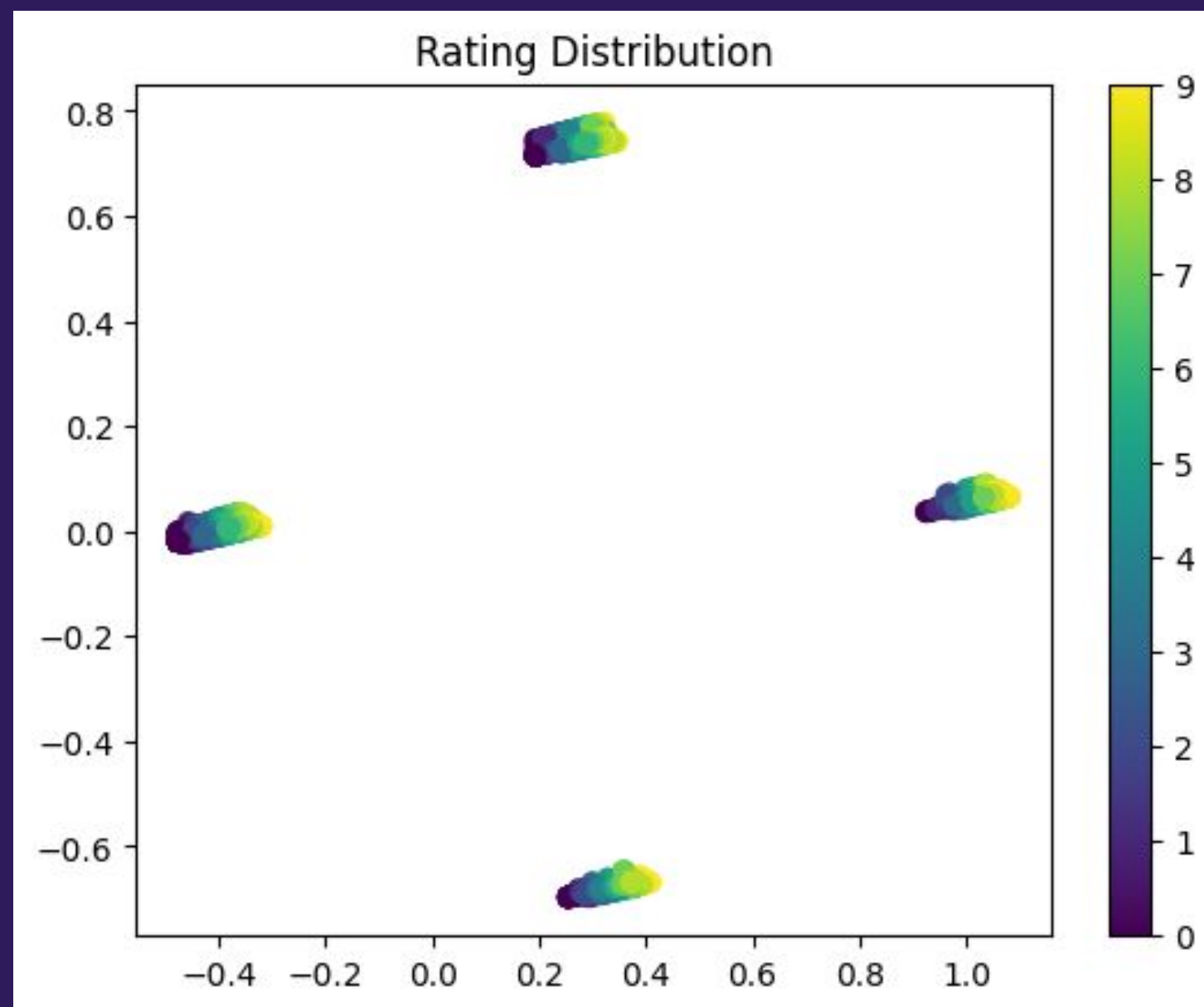
# Resultados

## Experimento Preliminar



# Resultados

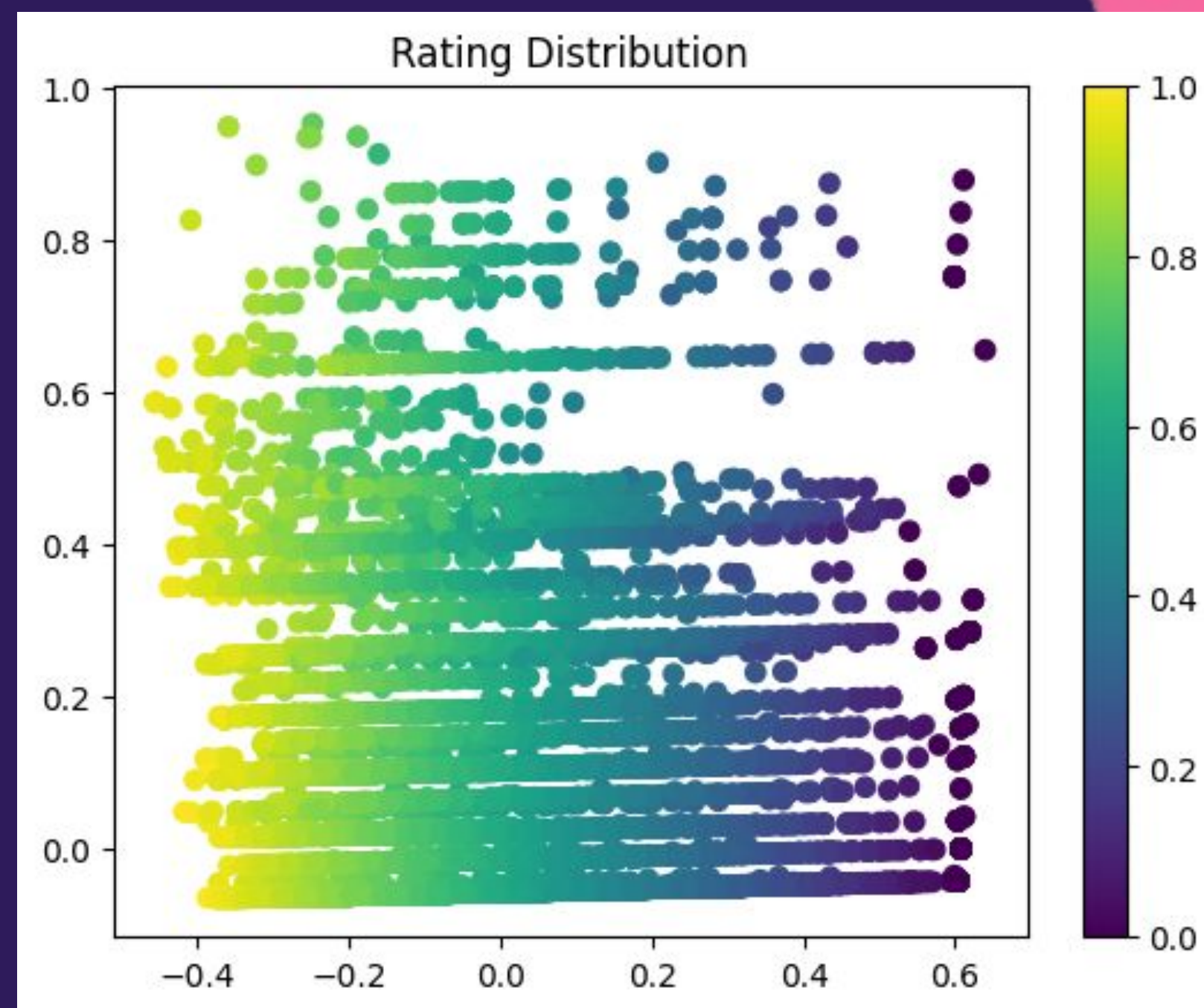
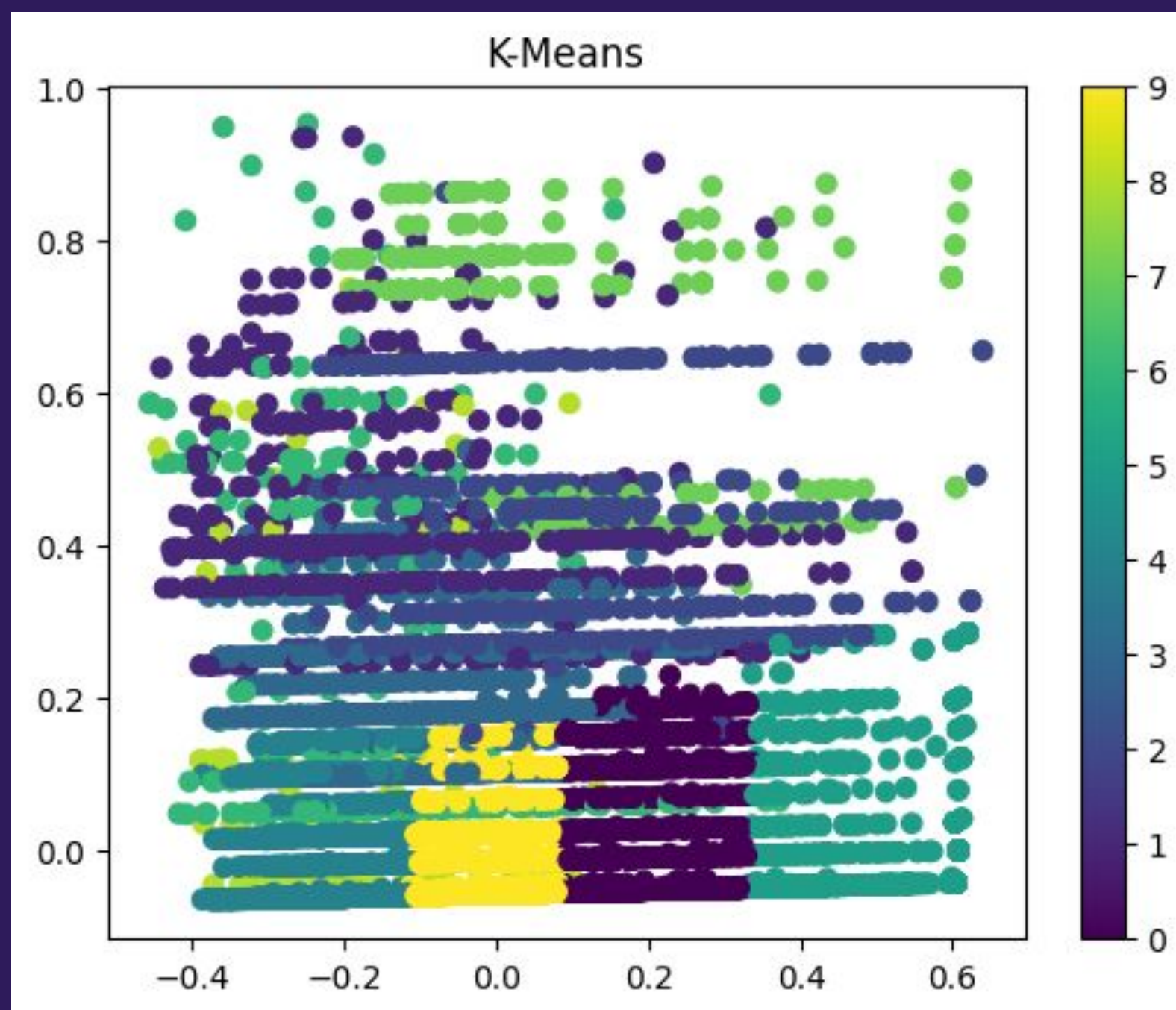
## Exploración de Clusters





# Resultados

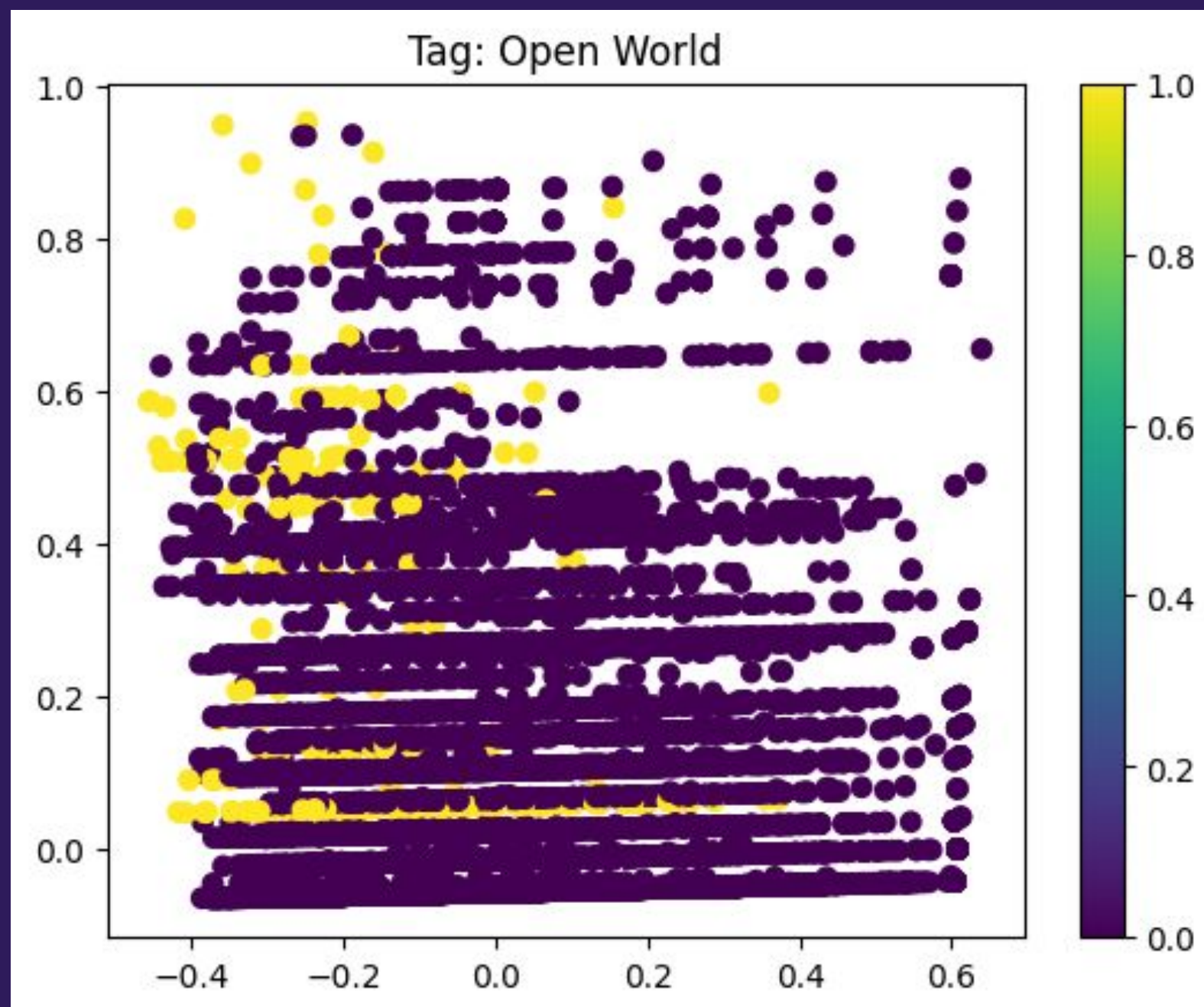
## Exploración de Clusters





# Resultados

## Exploración de Clusters

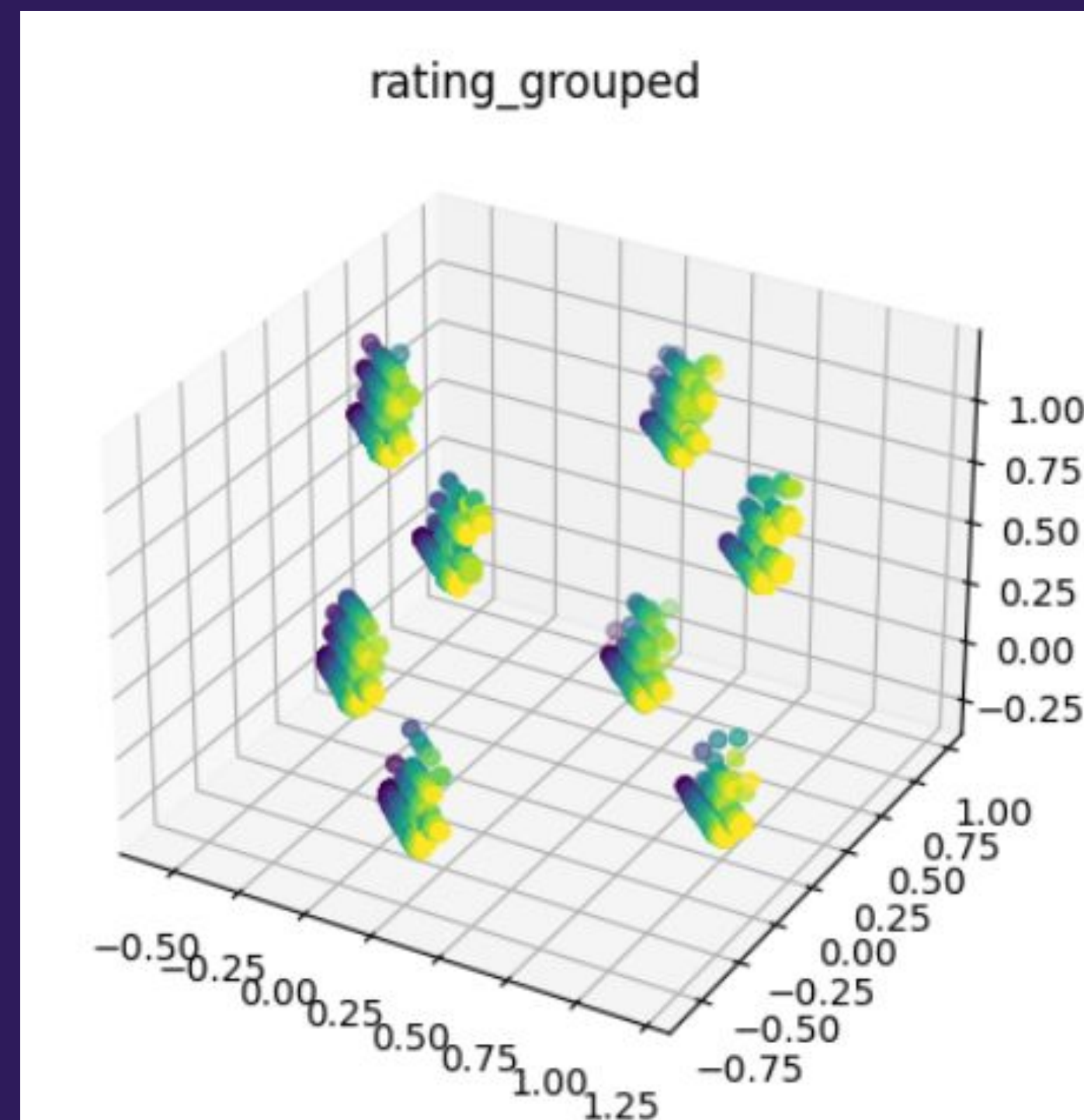
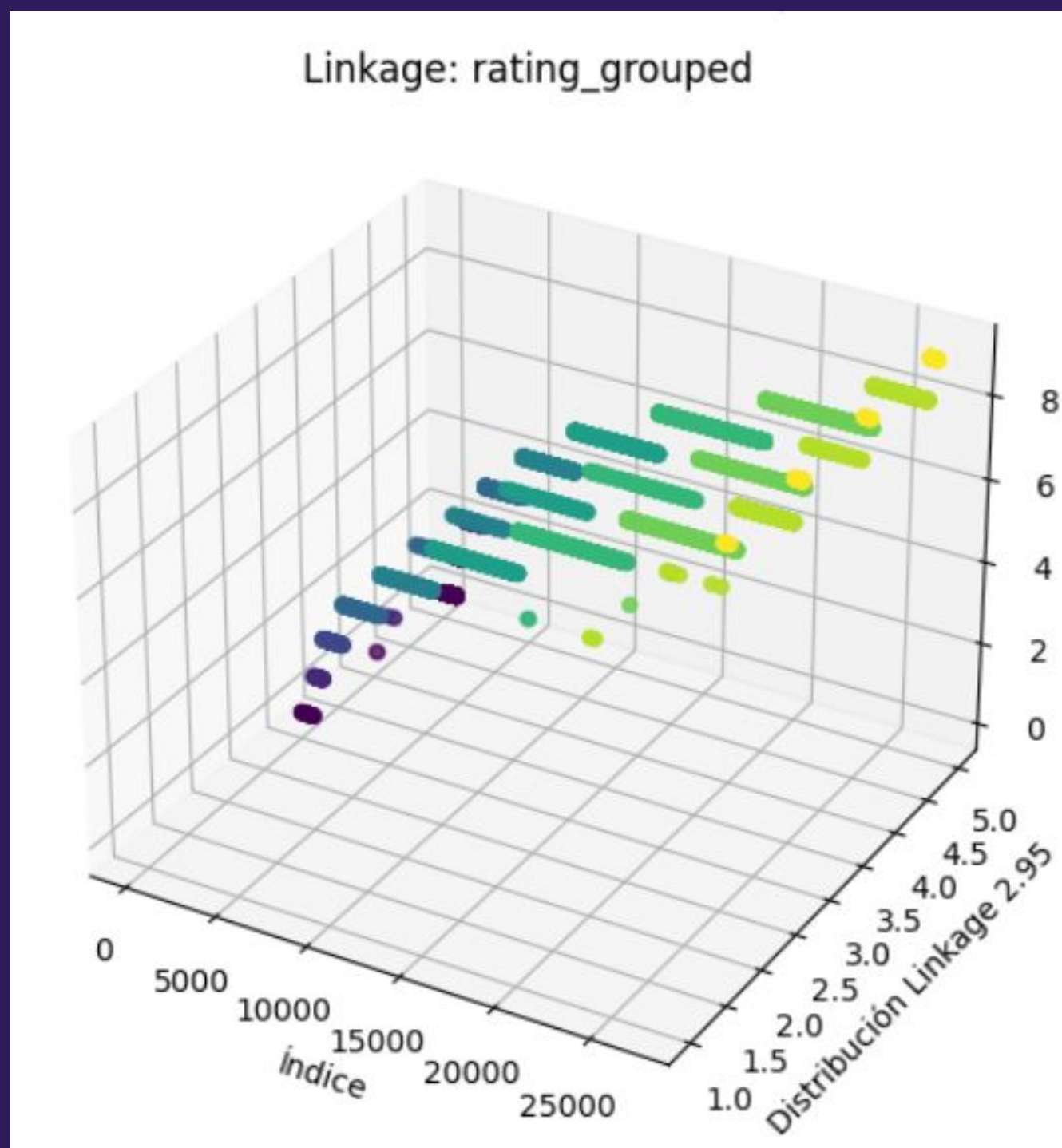


**¿Juegos Open World  
tienen más rating?**



# Resultados

## Exploración de Clusters



# Resultados

## Regresiones:

El objetivo (output) era obtener el rating en base a las descripciones o Keyphrases de un videojuego (input).

Se experimentó con diversos modelos, diversas formas de vectorizar los textos y variando el input (keyphrases con KeyBert o descripción completa).

Finalmente los mejores resultados se obtuvieron con el modelo Ridge Regression, con TF-IDF y usando la detailed\_description como input.





# Resultados



## Regresiones:

TF-IDF Vectorization with detailed_description			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	241.65	12.54	0.1551
LR	295.50	13.62	-0.0330
RF	280.80	13.42	0.0183
GB	261.07	13.29	0.0872
SVR	281.04	13.30	0.0174

TF-IDF Vectorization with filter_description			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	243.15	12.50	0.1499
Lasso	286.36	14	-0.0011
LR	300.34	13.72	-0.0500
RF	286.52	13.66	-0.0016
GB	262.82	13.28	0.0811
SVR	282.83	13.36	0.0112

TF-IDF Vectorization with top_keyphrases			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	298.75	13.78	-0.0444
Lasso	286.36	14	-0.0011
LR	647.86	20.25	-1.2649
RF	321.23	13.88	-0.1230
GB	285.30	13.88	0.0025
SVR	291.08	13.53	-0.0176

Glove2Vec Vectorization with detailed_description			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	263.21	13.20	0.0797
Lasso	275.78	13.7	0.0358
LR	265.23	13.23	0.0727
RF	289.01	13.78	-0.0100
GB	265.41	13.14	0.0720
SVR	280.55	13.34	0.0191

Glove2Vec Vectorization with filter_description			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	265.46	13.23	0.0719
Lasso	277.21	13.72	0.0308
LR	268.07	13.26	0.0628
RF	282.24	13.61	0.0132
GB	268.33	13.25	0.0619
SVR	279.40	13.30	0.0231

Glove2Vec Vectorization with top_keyphrases			
Model	MSE	MAE	R2
Dummy	286.36	14.00	-0.0011
Ridge	284.58	13.76	0.0051
Lasso	282.21	13.84	0.0133
LR	284.77	13.76	0.0044
RF	313.07	14.47	-0.0945
GB	291.85	14.01	-0.0203
SVR	286.38	13.40	-0.0012

# Resultados

## Clasificadores:

El objetivo (output) era obtener clases del rating en base a las descripciones o Keyphrases de un videojuego (input).

Se usan clasificadores para ver si se puede "predecir" el rating de una forma distinta que sea más efectiva a los modelos de regresión.

Se crearon labels en base al rating (malo, bueno, regular, etc) y se entrenan distintos clasificadores.

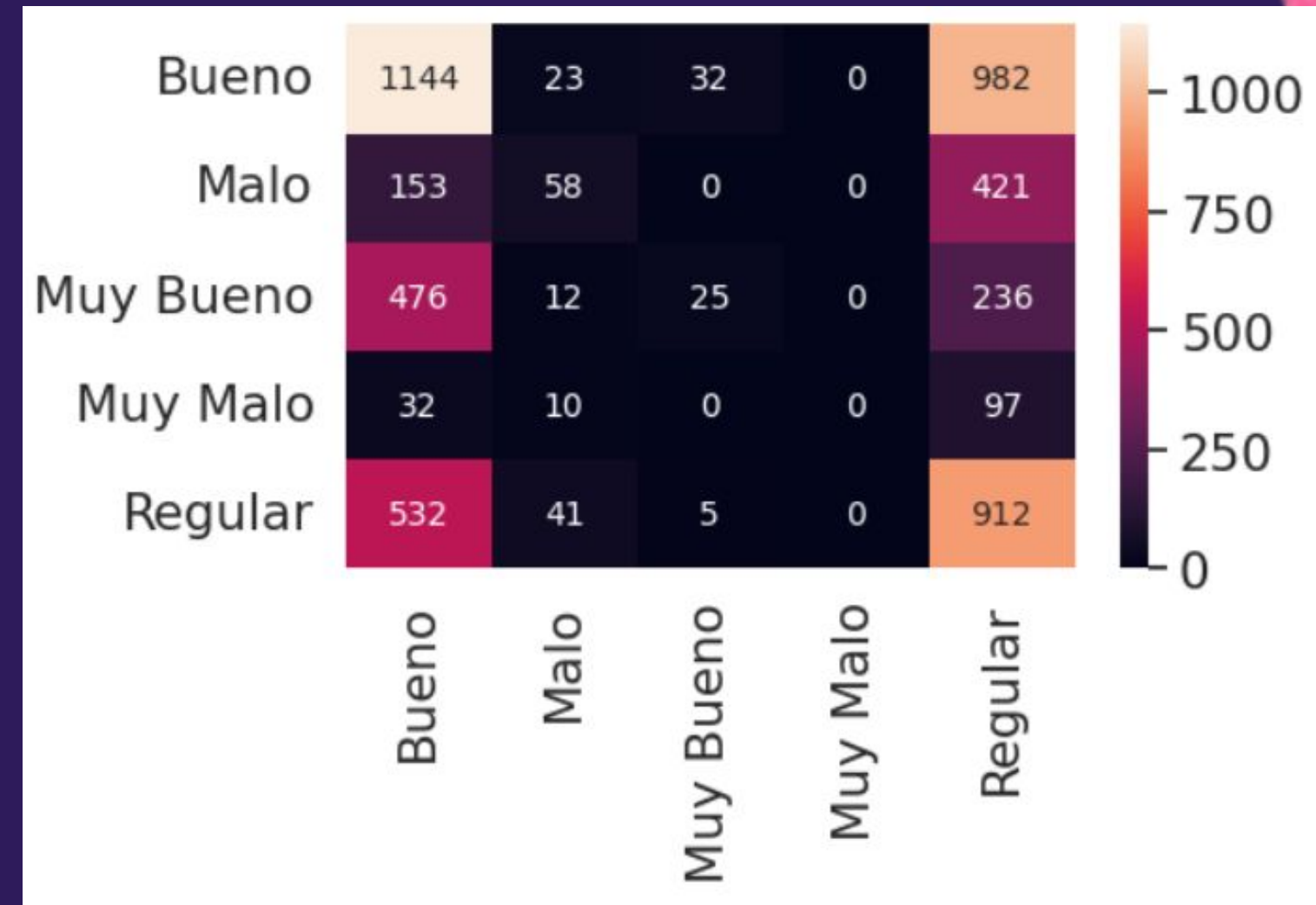




# Resultados

## Clasificación: Ridge Regression Classifier

	precision	recall	f1-score	support
Bueno	0.49	0.52	0.51	2181
Malo	0.40	0.09	0.15	632
Muy Bueno	0.40	0.03	0.06	749
Muy Malo	0.00	0.00	0.00	139
Regular	0.34	0.61	0.44	1490
accuracy			0.41	5191
macro avg	0.33	0.25	0.23	5191
weighted avg	0.41	0.41	0.37	5191

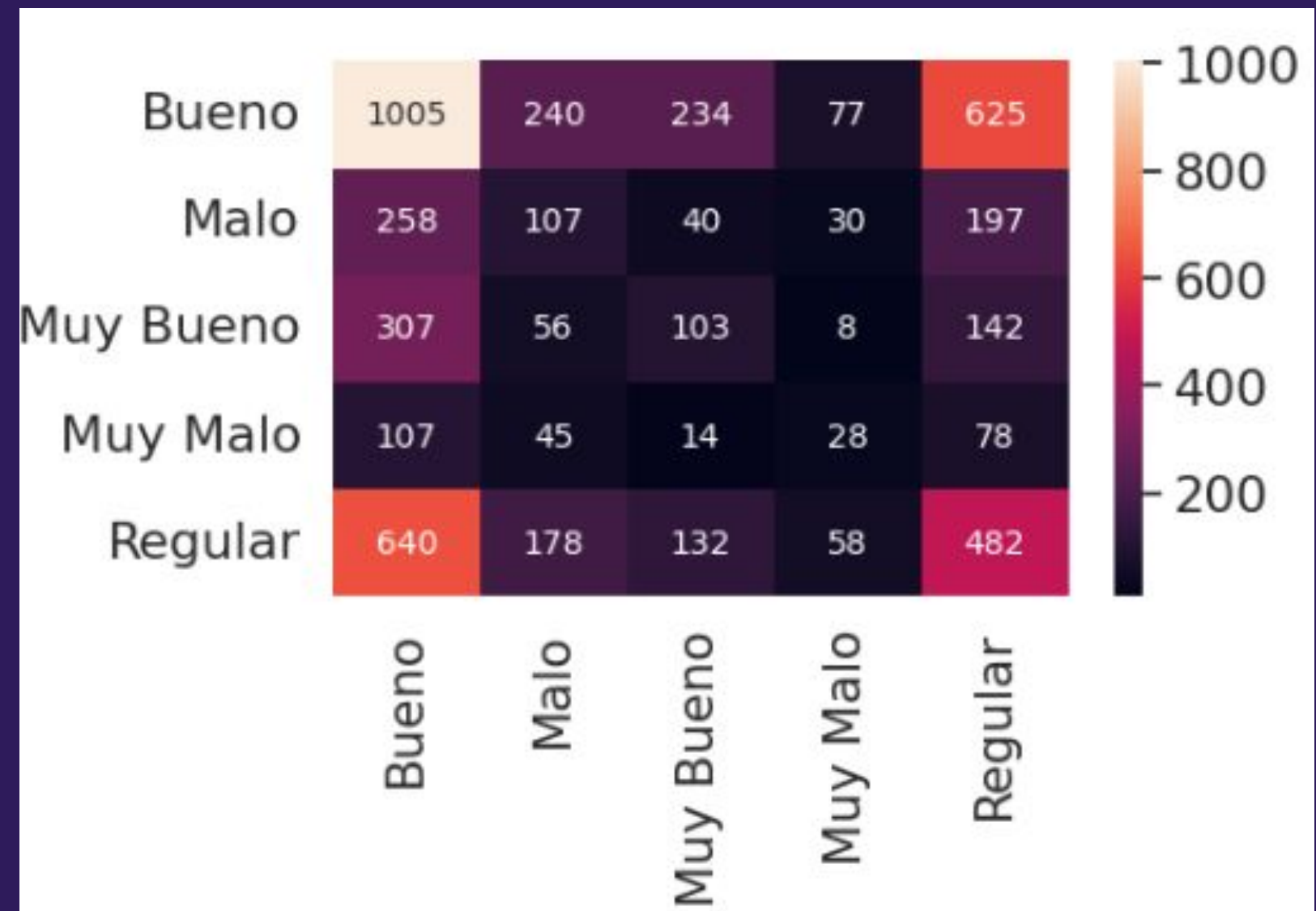


# Resultados

## Clasificación: Decision Tree

Accuracy of Decision Tree Classifier: 0.3323059140820651  
Classification Report:

	precision	recall	f1-score	support
Bueno	0.43	0.46	0.45	2181
Malo	0.17	0.17	0.17	632
Muy Bueno	0.20	0.17	0.18	616
Muy Malo	0.14	0.10	0.12	272
Regular	0.32	0.32	0.32	1490
accuracy			0.33	5191
macro avg	0.25	0.24	0.25	5191
weighted avg	0.32	0.33	0.33	5191



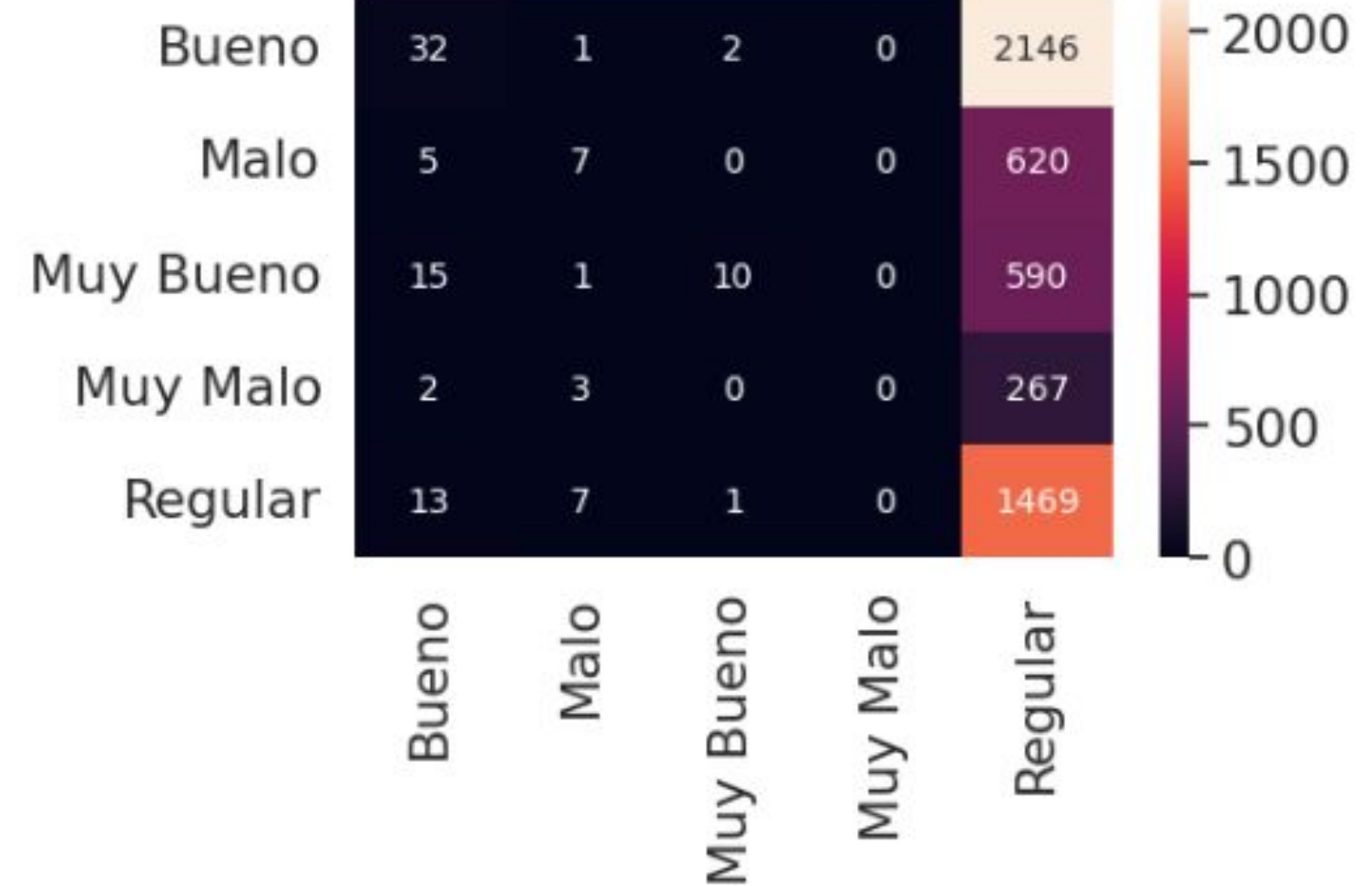


# Resultados

## Clasificación: KNN

Accuracy of k-NN Classifier: 0.2924292043922173  
Classification Report for k-NN:

	precision	recall	f1-score	support
Bueno	0.48	0.01	0.03	2181
Malo	0.37	0.01	0.02	632
Muy Bueno	0.77	0.02	0.03	616
Muy Malo	0.00	0.00	0.00	272
Regular	0.29	0.99	0.45	1490
accuracy			0.29	5191
macro avg	0.38	0.21	0.11	5191
weighted avg	0.42	0.29	0.15	5191



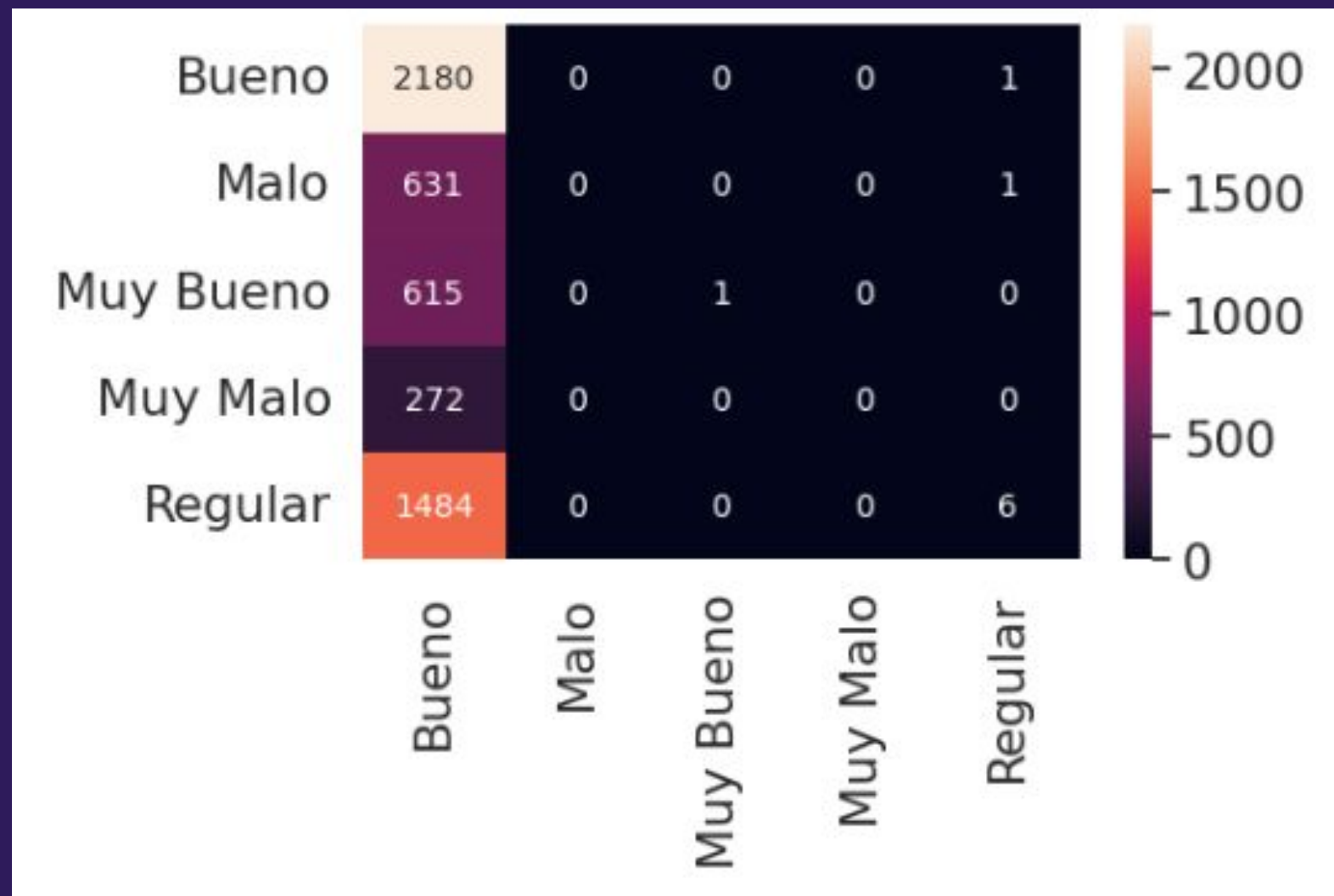
# Resultados

## Clasificación: Naive Bayes

Accuracy of Naive Bayes Classifier: 0.4213061067231747

Classification Report for Naive Bayes:

	precision	recall	f1-score	support
Bueno	0.42	1.00	0.59	2181
Malo	0.00	0.00	0.00	632
Muy Bueno	1.00	0.00	0.00	616
Muy Malo	0.00	0.00	0.00	272
Regular	0.75	0.00	0.01	1490
accuracy			0.42	5191
macro avg	0.43	0.20	0.12	5191
weighted avg	0.51	0.42	0.25	5191





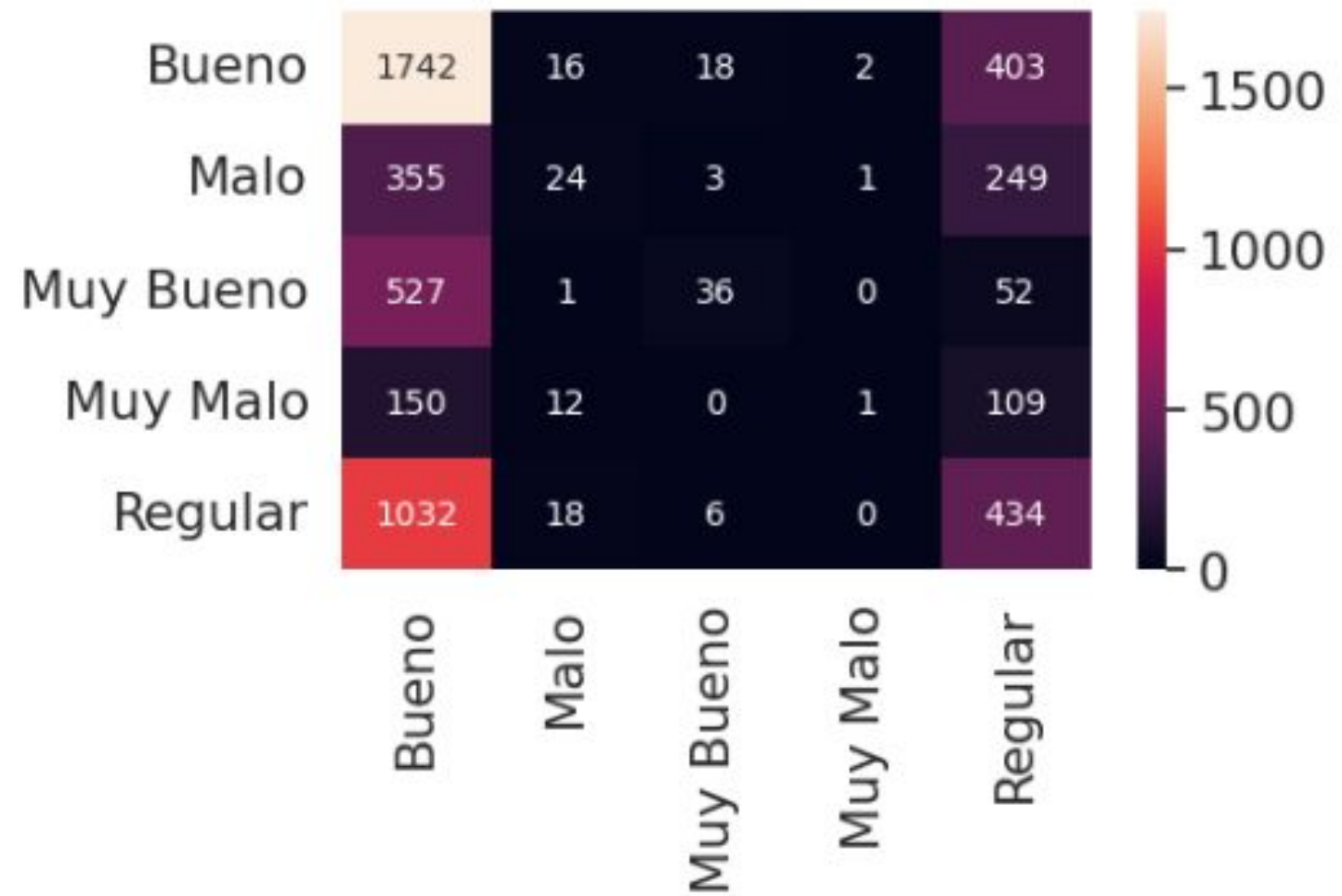
# Resultados

## Clasificación: SVM

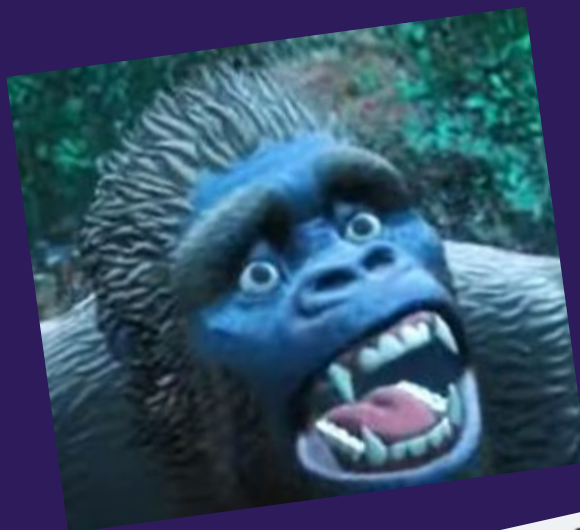
Accuracy of SVM Classifier: 0.4309381622038143

Classification Report for SVM:

	precision	recall	f1-score	support
Bueno	0.46	0.80	0.58	2181
Malo	0.34	0.04	0.07	632
Muy Bueno	0.57	0.06	0.11	616
Muy Malo	0.25	0.00	0.01	272
Regular	0.35	0.29	0.32	1490
accuracy			0.43	5191
macro avg	0.39	0.24	0.22	5191
weighted avg	0.41	0.43	0.36	5191



# Ejemplo



"Una completa estafa", Skull Island Rise of Kong es aclamado como el peor juego del 2023 y fue desarrollado en Chile



? 39.49%  
↑36 ↓64

1  
In-Game



Rating Prediction: 58.75415083780705  
Real Rating Value: 39.49

Most negative coefficients:

Coefficient	Feature Label (Words)
-21.1363	various
-18.575	person
-16.5263	history
-15.4461	jungle
-14.6015	king
-12.2693	waves
-10.8859	fury
-10.817	terrain
-10.281	heights
-9.40571	potential





# Futuras Direcciones

## 🌟 Tendencias temporales

Impacto de eventos relevantes (lanzamientos de consolas, plataformas, etc.)

## 🌟 Regresiones

Buscar formas adicionales de optimizar el modelo de regresión (Ej: Regularizaciones)

## 🌟 Análisis geográfico

Estudiar tendencias en distintas regiones geográficas: países o continentes.

## 🌟 Clasificación

Abordar el desafío del desequilibrio entre las clases mediante la implementación de técnicas de balanceo.

## 🌟 Clustering<sub>c</sub>

Variar preprocesamiento para reducir dimensionalidad, estudiar más los atributos en los clusters encontrados.

## 🌟 Clasificación<sub>c</sub>

Con clusters mejor construidos nombrar según rating y usar esas labels para clasificar.



# Conclusiones

- No se obtienen clusters claros separados por rating Se encuentra un indicio de un tipo de juego relacionado con alto rating.
- Regresiones obtienen predicciones aproximadas.
- Clasificación funciona aproximadamente.
- Las preguntas planteadas se contestan con baja certeza, más análisis es necesario.



# ¡Muchas Gracias!

Pd: si quieren hacer un juego exitoso no  
hagan un juego de un gorila en una isla.

