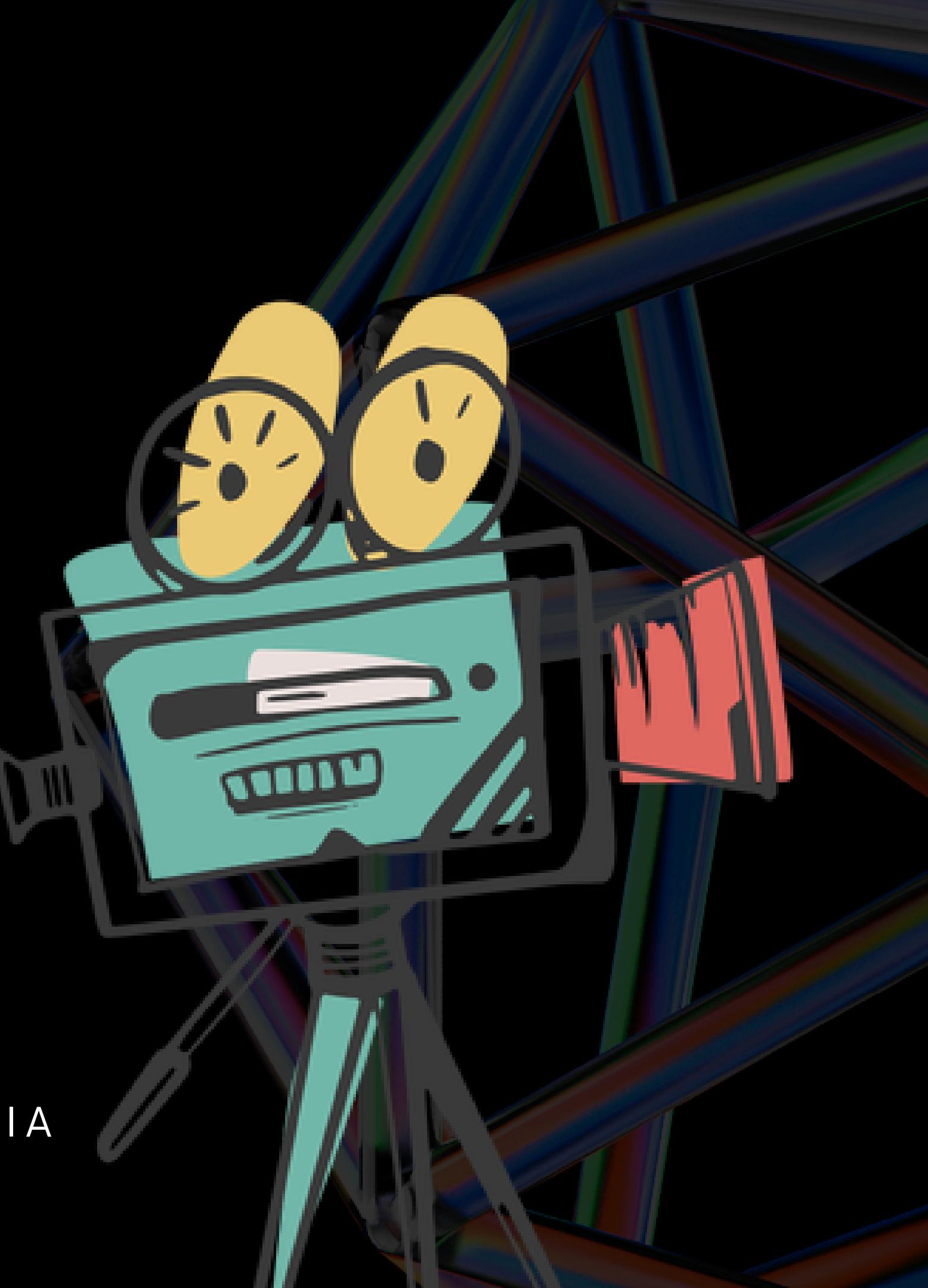


PROYECTO PREDICTOR DE PELÍCULAS TAQUILLERAS

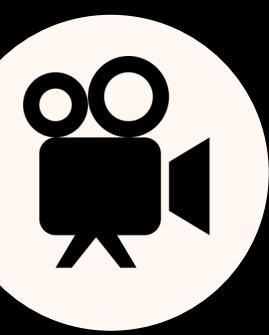
2190092 - HERNÁN SANTIAGO VALLEJO GARCIA
2191795 - JUAN DIEGO ÁLVAREZ SÁNCHEZ



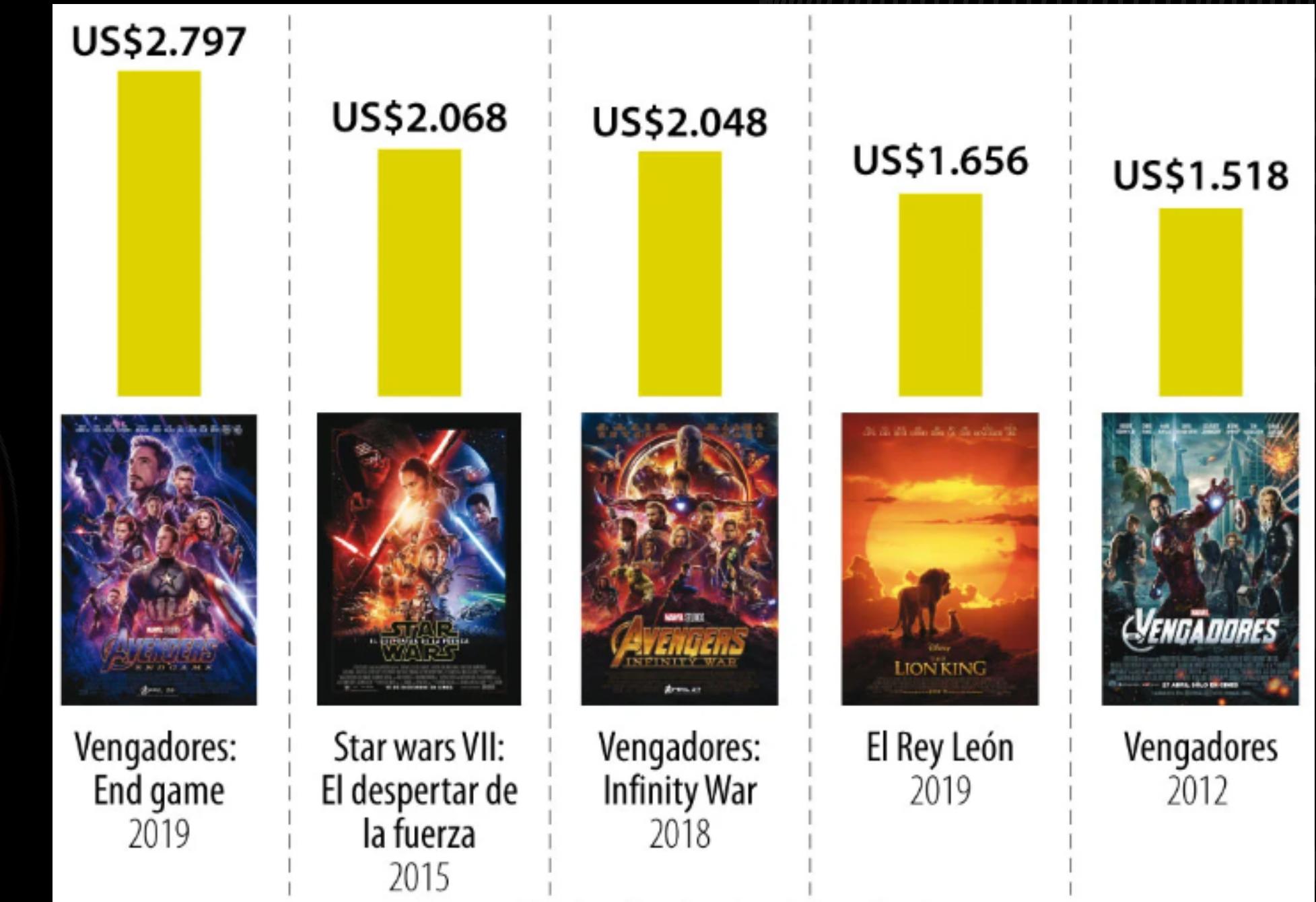
A cartoon illustration of two characters from the TV show Rick and Morty. On the left, Morty is shown with a shocked expression, wearing his signature green shirt and white headphones. On the right, Rick is shown with a more neutral or slightly smug expression, wearing his white lab coat and blue spiky hair. They are set against a dark background with some radial lines.

INDICE:

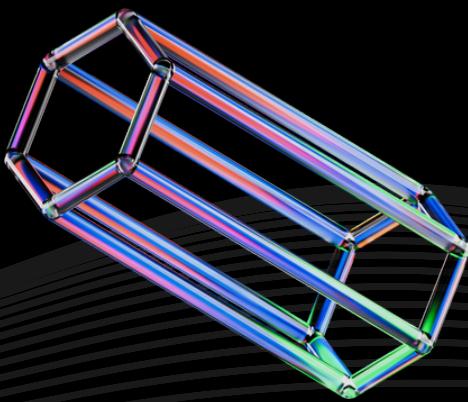
- 01 **RESUMEN**
- 02 **INTRODUCCIÓN**
- 03 **PROCESO Y MÉTODO**
- 04 **RESULTADOS**
- 05 **CONCLUSIONES**



RESUMEN

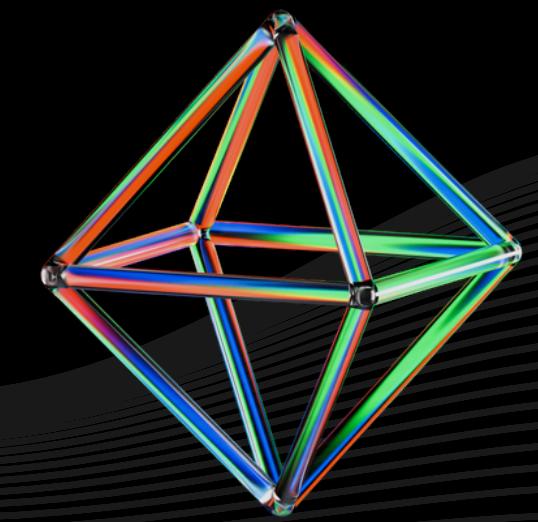


INTRODUCCIÓN



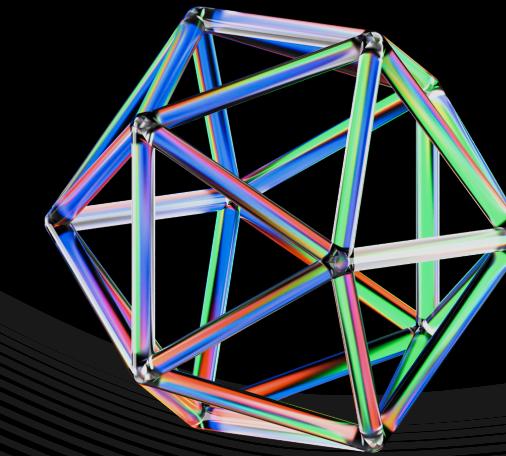
PELICULAS

¿Cuales son?



TAQUILLERAS

¿Cuanto generaron?



ÉXITO

¿Por que esas películas?

PROCESO Y MÉTODO

Generos

```
[489] testData.head()
      id budget popularity runtime log_budget genre_Count Has_En num_production_companies is_use_production number_of_spoken_languages size_of_cast
0 3001 0 3.851534 90.0 0.000000 4 0 3.0 1 2.0 7.0
1 3002 88000 3.559789 65.0 11.385103 2 0 1.0 1 1.0 10.0
2 3003 0 8.085194 100.0 0.000000 2 0 3.0 1 1.0 9.0
3 3004 6800000 8.596012 130.0 15.732433 3 0 3.0 0 3.0 23.0
4 3005 2000000 3.217680 92.0 14.508658 2 0 3.0 1 1.0 4.0

trainData.head()
      id budget popularity runtime revenue log_budget genre_Count Has_En num_production_companies is_use_production number_of_spoken_languages size_of_cast
0 1 14000000 6.575393 93.0 12314651 16.454568 1 1 3.0 1 1.0 24.0
1 2 40000000 8.248895 113.0 95149435 17.504390 4 1 1.0 1 1.0 20.0
2 3 3300000 64.299990 105.0 13092000 15.009433 1 1 3.0 1 1.0 51.0
3 4 1200000 3.174936 122.0 16000000 13.997833 2 1 3.0 0 2.0 7.0
5 6 8000000 0.743274 83.0 3261638 15.894952 2 1 2.0 1 1.0 14.0

mainData['genre_Count'] = 0
~ i in trainData['genres']:
d = get_dict(i)
if d != {}:
    trainData['genre_Count'][j] = len(d)
else:
    trainData['genre_Count'][j] = 0
j += 1

testData['genre_Count'] = 0
~ i in testData['genres']:
d = get_dict(i)
if d != {}:
    testData['genre_Count'][j] = len(d)
else:
    testData['genre_Count'][j] = 0
j += 1

trainData.drop(['genres'], axis=1, inplace=True)
testData.drop(['genres'], axis=1, inplace=True)
trainData.head()
```



trainData.isnull().sum()

id	0
belongs_to_collection	2396
budget	0
genres	7
homepage	2054
imdb_id	0
original_language	0
original_title	0
overview	8
popularity	0
poster_path	1
production_companies	156
production_countries	55
release_date	0
runtime	2
spoken_languages	20
status	0
tagline	597
title	0
Keywords	276
cast	13
crew	16
revenue	0
dtype: int64	

RESULTADOS

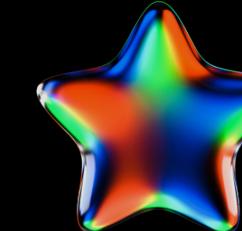
- INGRESOS
- E.E.U.U
- IDIOMA
- PREDICCIÓN



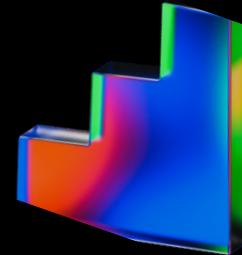
CON CLUSI ONES



Se acertaron algunas predicciones



Estados Unidos es el sector más fuerte



EstArboles de decisión fácilmente escalables



Resultados desfavorables



GRACIAS