

JUAN BAYÓN FERNÁNDEZ

PROYECTO EDA

VALORACIONES EN PÁGINAS ESPECIALIZADAS DE CINE



Proyecto EDA

TEMAS A DEBATIR

Presentación

Procedencia de los datos

Planteamiento

Datos recogidos

Análisis de los datos

Conclusiones

PRESENTACIÓN PERSONAL



ARQUITECTURA

Licenciado en arquitectura por la Universidad de la Coruña



CONSTRUCCIÓN

Trabajos previos en el mundo de la Construcción en obras como el centro comercial Las Arenas en Barcelona



DATA SCIENCE

Cambio profesional en el campo de la ciencia de datos.

OBTENCIÓN DE DATOS

Para el proyecto se han usado tres dataset de Kaagle:

- Movies. con información general de las películas como director, actores, género o título.
- IMDb. se recogen los ingresos además de las valoraciones según sexo, edad, o procedencia.
- TMDb. Con detalle de las valoraciones de TMDb y Metacritic.

Estos dataframe proceden del scraping de las respectivas páginas web.

PLANTEAMIENTO

**¿REPERCUTEN LAS NOTAS DE IMDB
EN LA RECAUDACIÓN TOTAL DE
LAS PELÍCULAS?**



TRABAJO PREVIO

LIMPIEZA Y PREPARACIÓN DE LOS DATOS

Se han limpiado los datos para que no haya valores inexistentes y estén todos en el mismo formato

ANÁLISIS

COMPROBACIÓN DE LOS TIPOS DE DATOS

Se analizan primero los datos obtenidos para saber sus características

DATOS RECOGIDOS

Director
Géneros
Idioma
Año
País
Actores

Valoraciones de TMDb
Valoraciones de IMDb
Valoraciones Metacritic

Votos por edad
Votos por sexo
Votos por edad y por sexo
Votos de EEUU resto del mundo

LOS DATOS SE SEPARAN EN TRES GRUPOS
PARA COMPROBAR LA RELACIÓN CON LA
RECAUDACIÓN

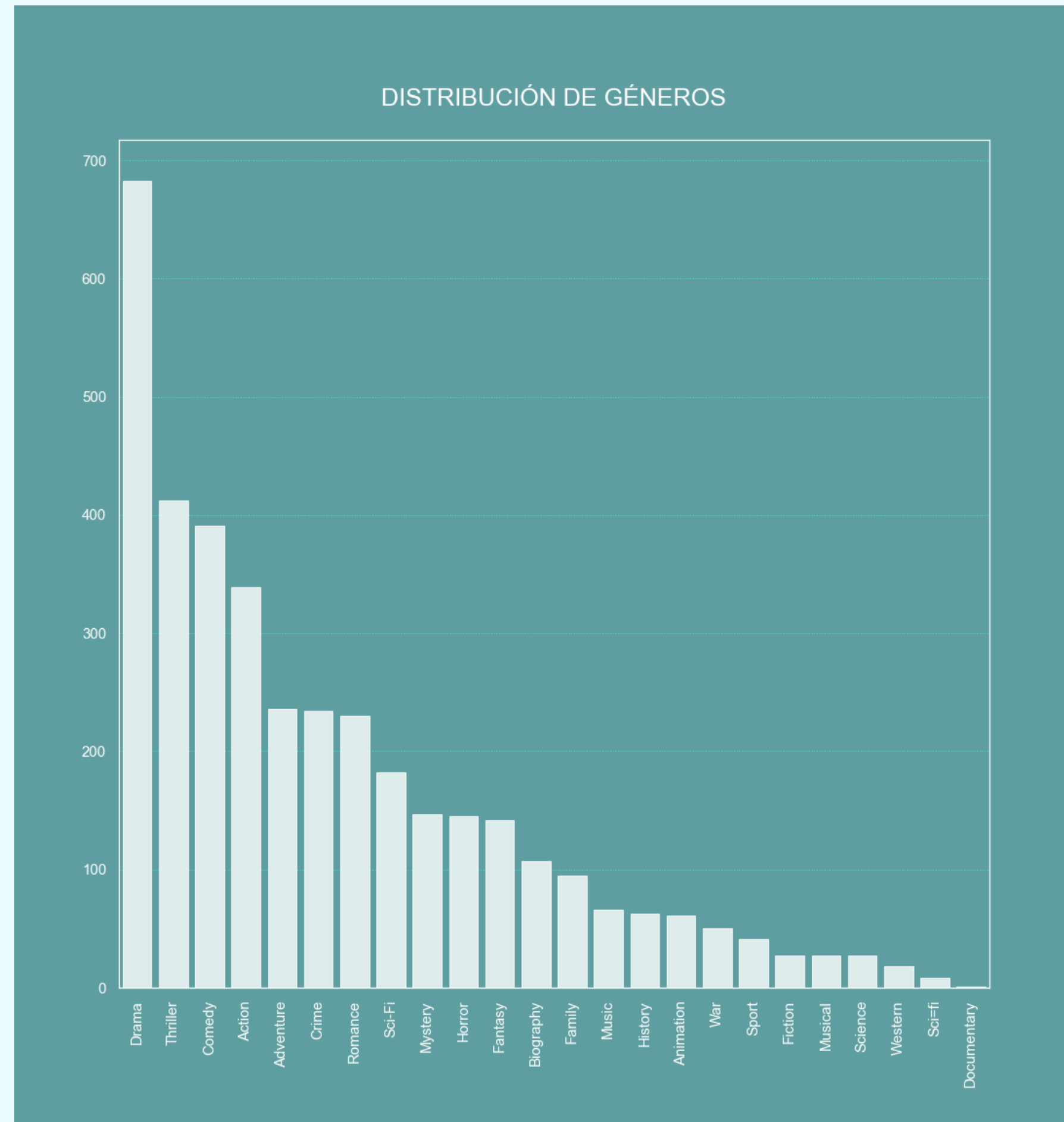
Recaudación mundial
Recaudación en EEUU

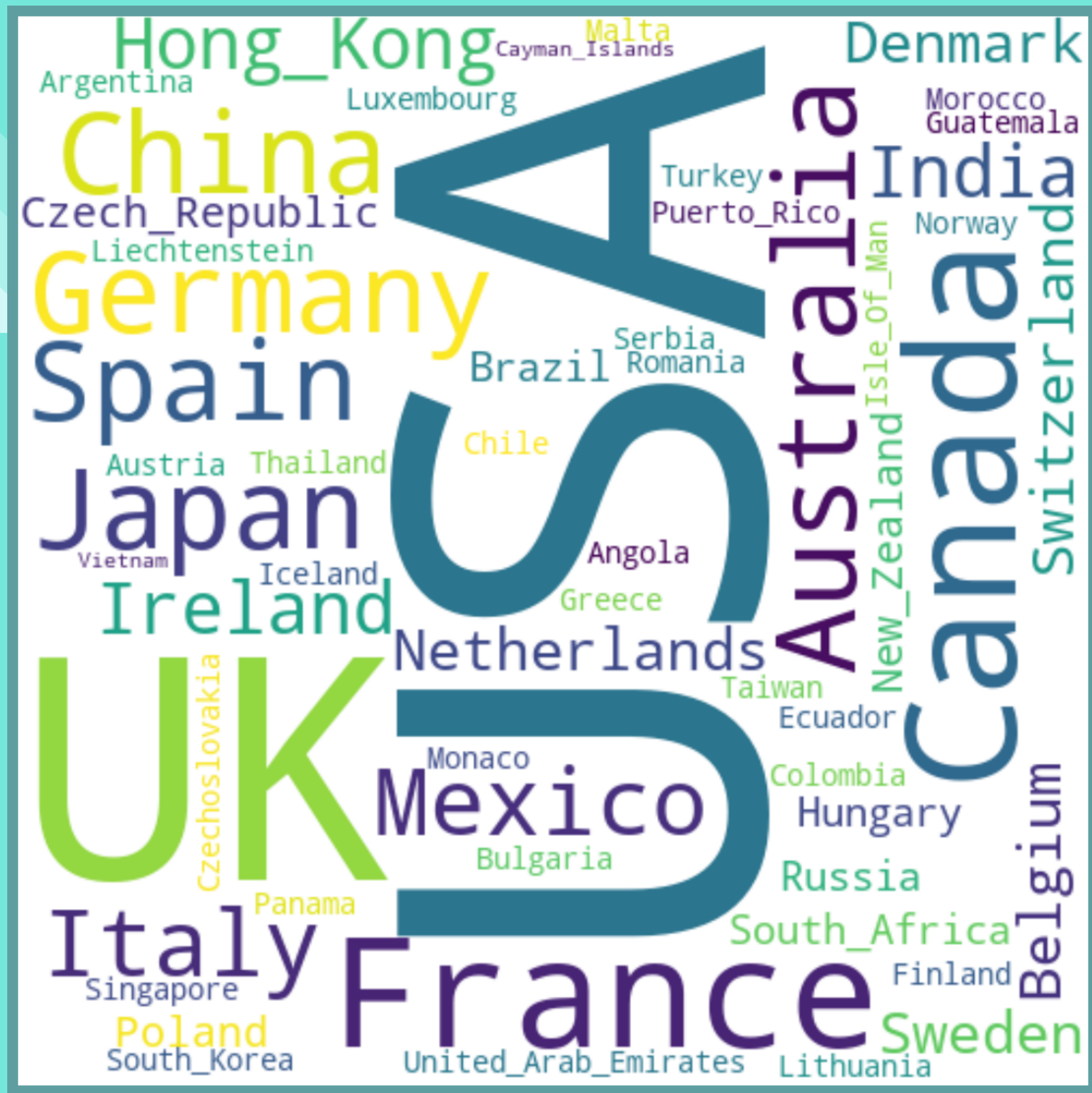
DISTRIBUCIÓN POR GÉNEROS

EL DRAMA ES EL GÉNERO
MÁS HABITUAL

THRILLER, COMEDIA O ACCIÓN SON LOS
SIGUIENTES.

SE REPRESENTA LAS PÉLICULAS MÁS
HABITUALES





PAÍS

ESTADOS UNIDOS ES
LA PROCEDENCIA
PREDOMINANTE EN
LA MUESTRA

Son 58 países en total

Reino Unido es la segunda
y luego el resto de valores
tienen mucha menos
representación

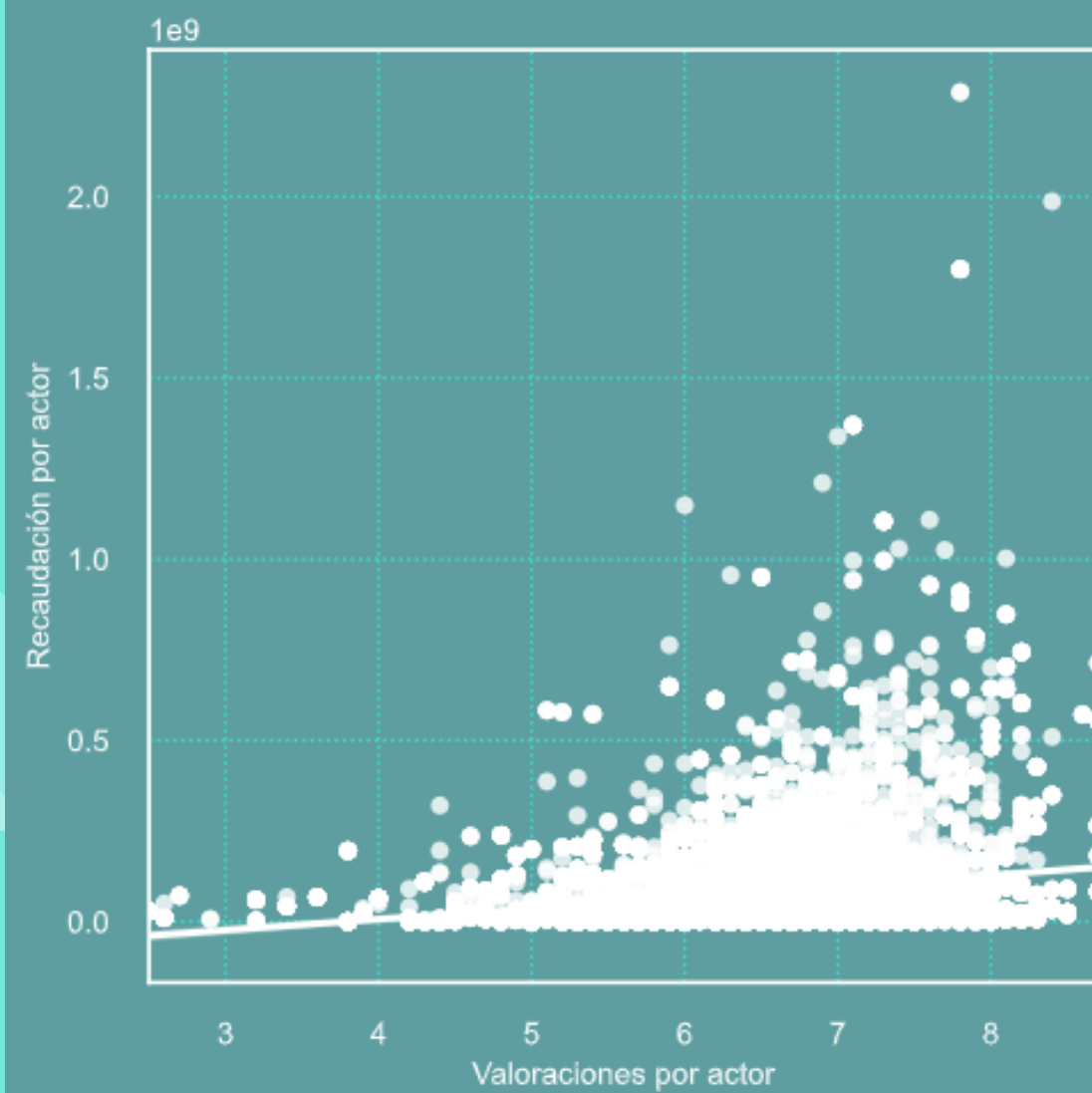
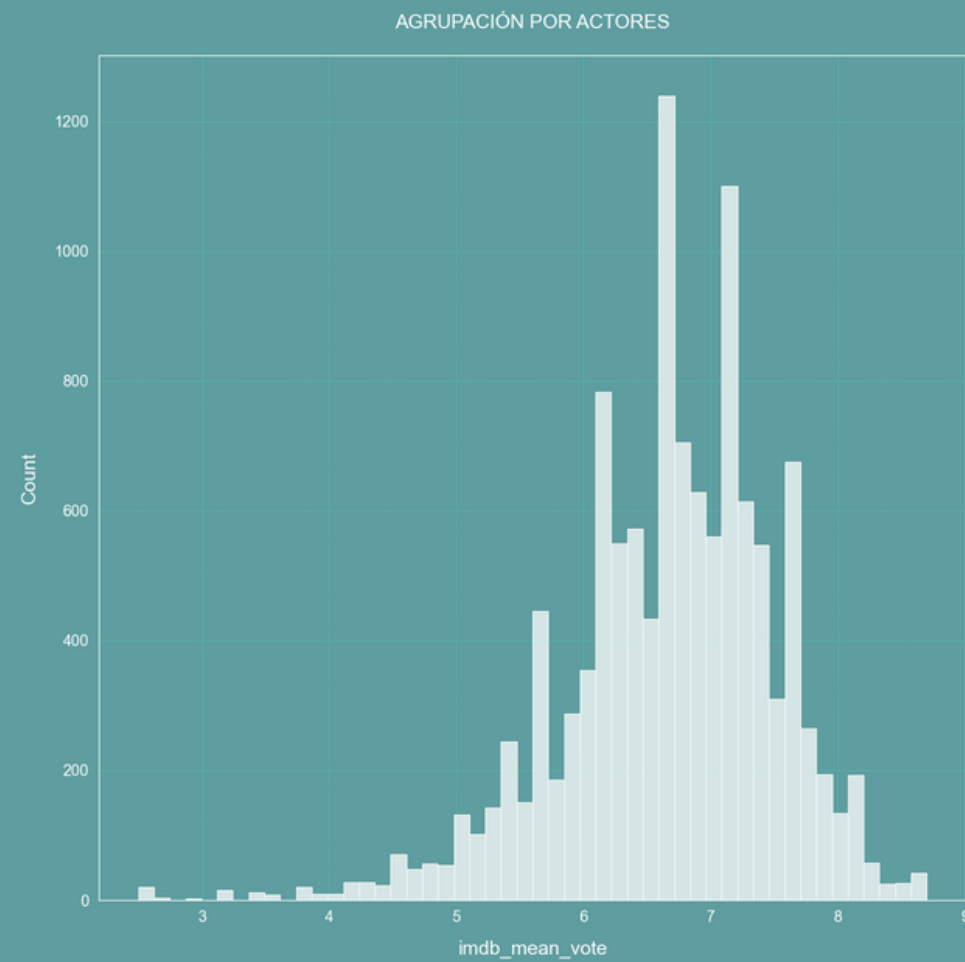
EL IDIOMA DE CASI
TODOS LOS TÍTULOS
ES EL INGLÉS

Distribución de películas por año de creación

HAY UN PUNTO DE INFLEXIÓN EN EL AÑO 1990

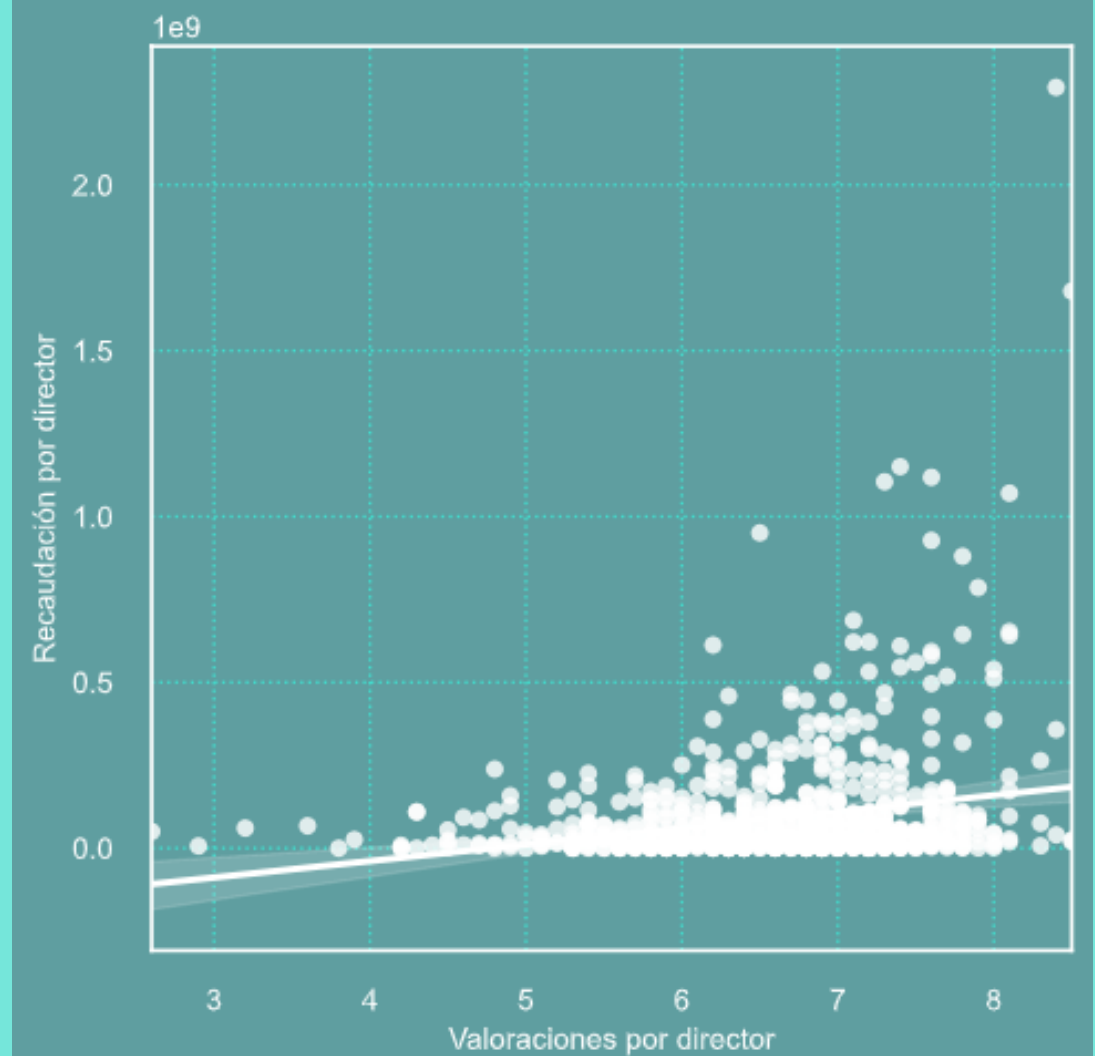
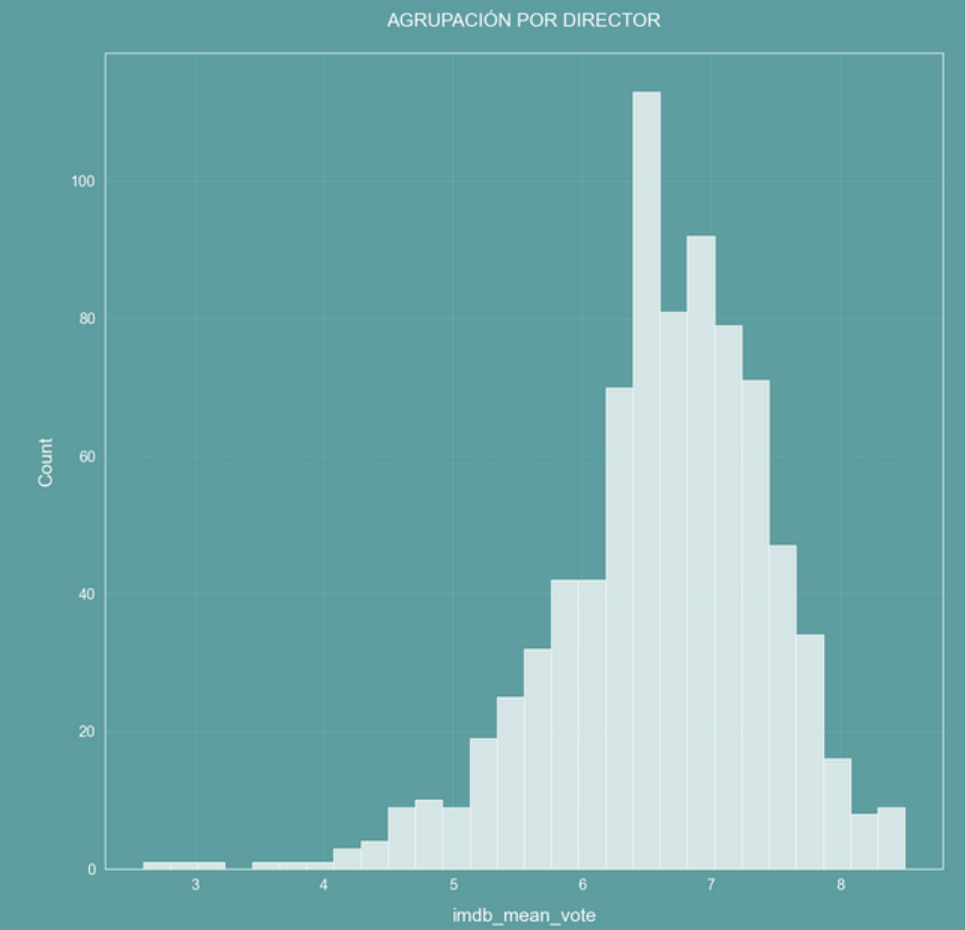


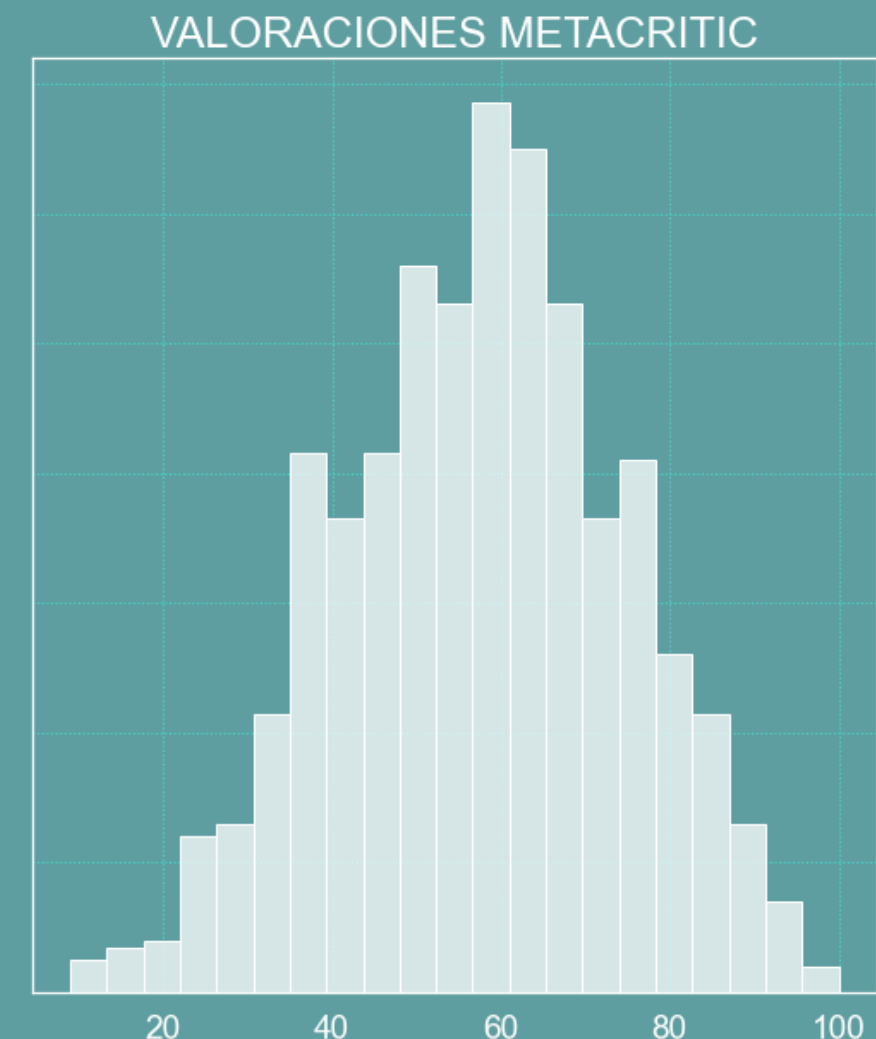
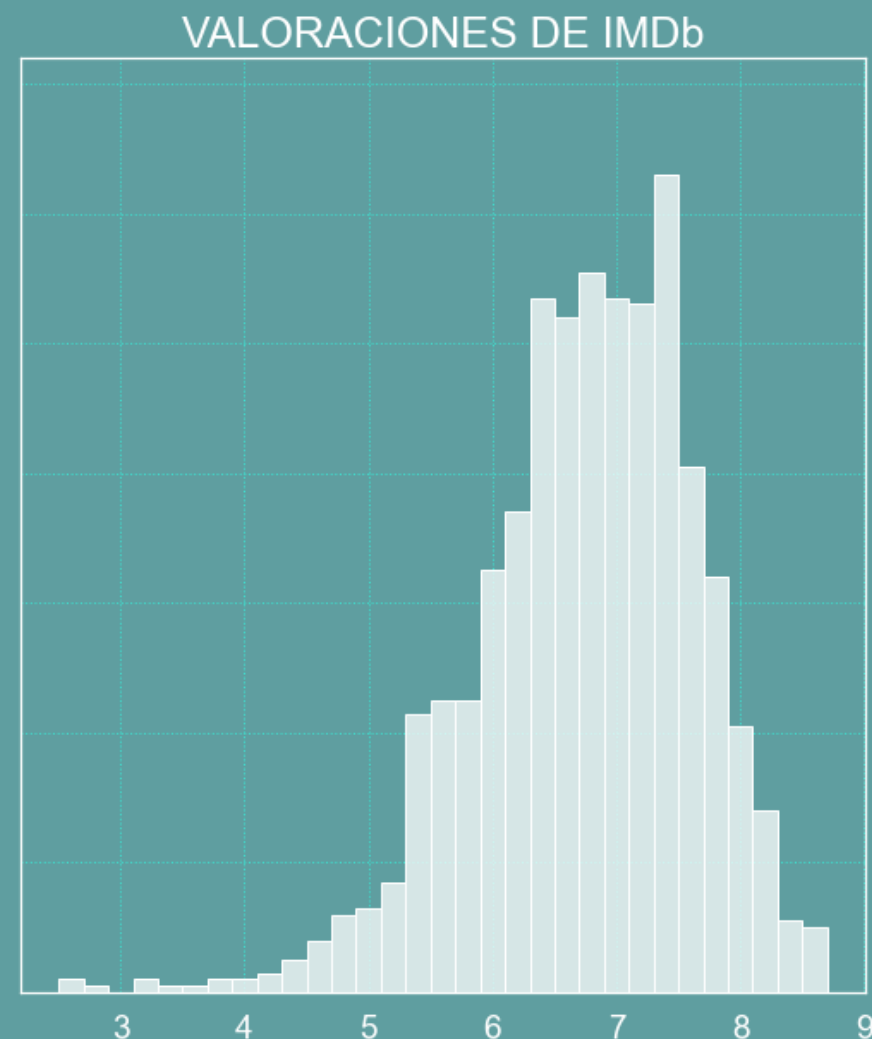
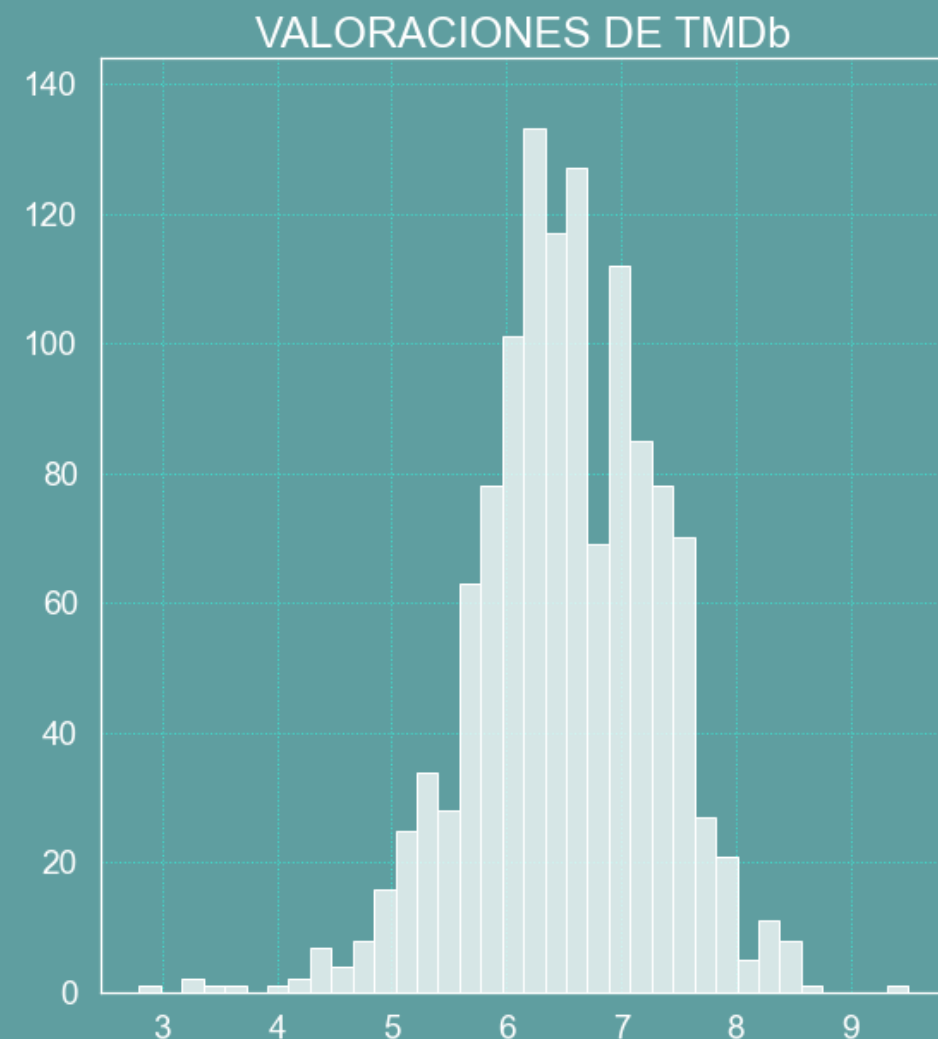
En el año 1990 se creó la página web IMDb por lo que la mayoría de películas recogidas son las creadas a partir de esta fecha



AGRUPANDO LOS VOTOS POR ACTOR Y DIRECTOR

Tanto para los actores como para los directores la distribución de votos agrupado es la similar. No existe la relación entre esos votos y la recaudación





Valoraciones de las 3 páginas

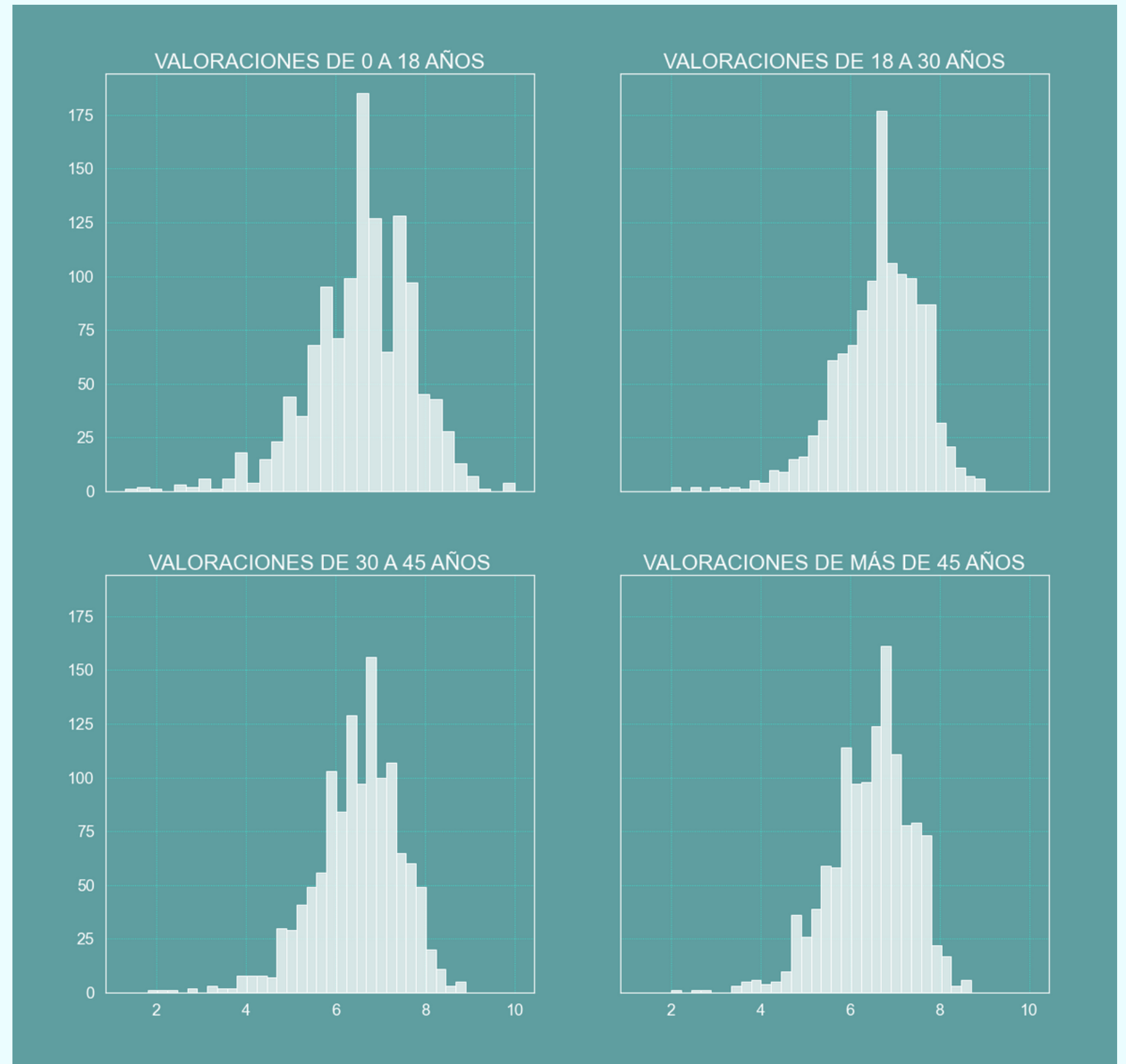
TMDB, IMDb Y METACRITIC

Las valoraciones de Metacritic están un poco más abiertas porque van del 1 al 100 y más centrada y uniforme porque es una suma de las valoraciones del resto de páginas. Sin embargo las gráficas son muy similares, con la mayoría de notas entre el 6 y el 7.

Valoraciones por edad

DISTRIBUCIÓN POR EDAD EN CUATRO GRUPOS

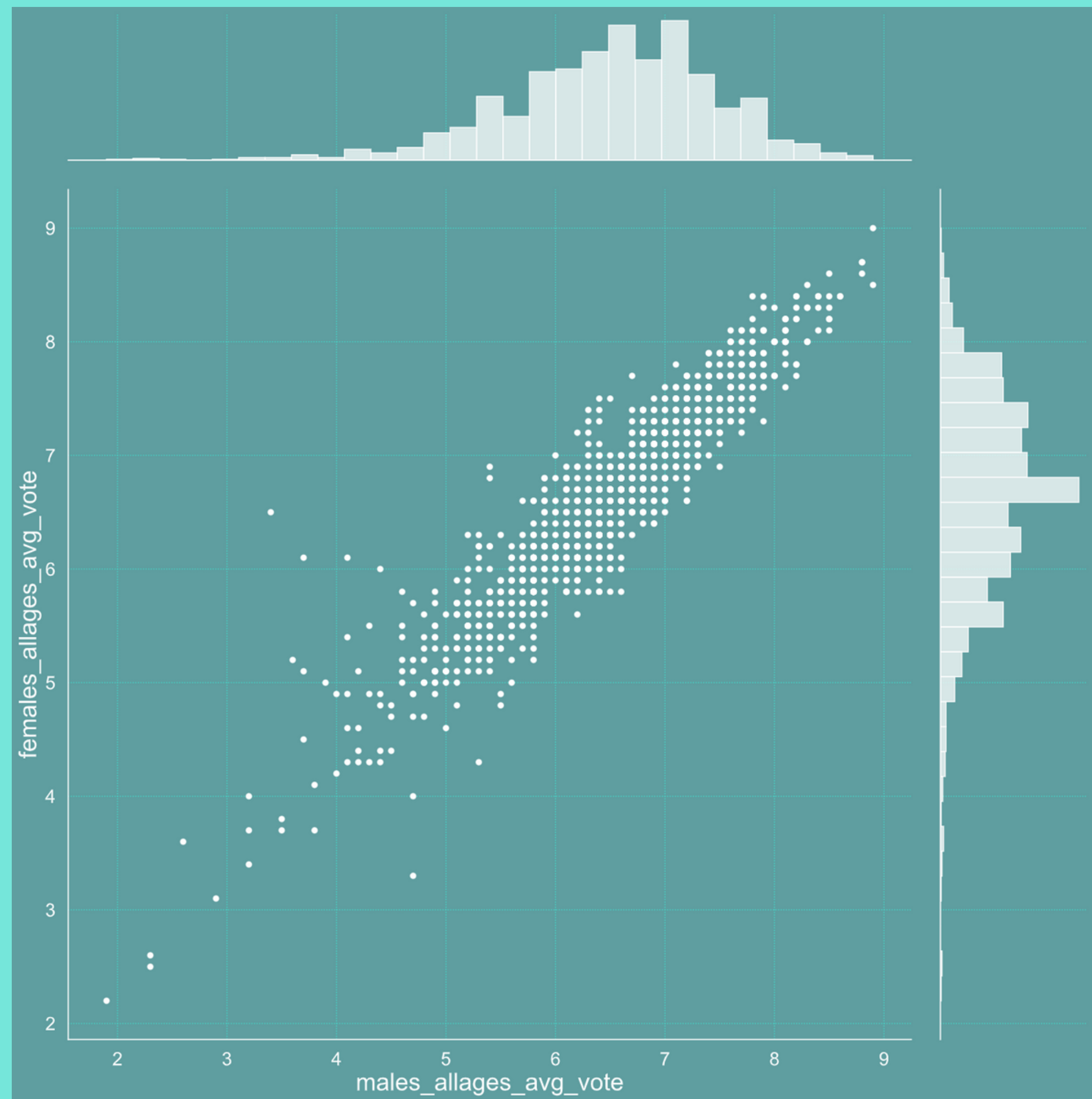
Se comprueba que la distribución es prácticamente igual. Con notas medias de entre 6 y 7



RELACIÓN DE LAS VALORACIONES DE HOMBRES Y MUJERES

ESTAN TOTALMENTE ALINEADOS

Las votaciones son muy similares independientemente del sexo.
Tanto hombres como mujeres votan igual.

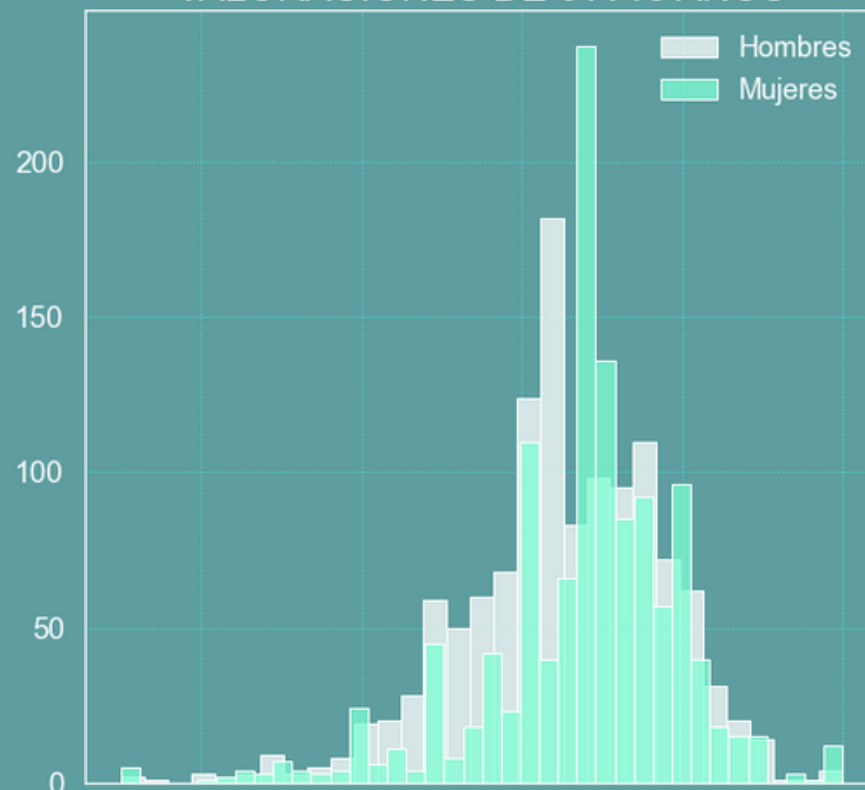


Valoraciones por sexo y tramos de edad

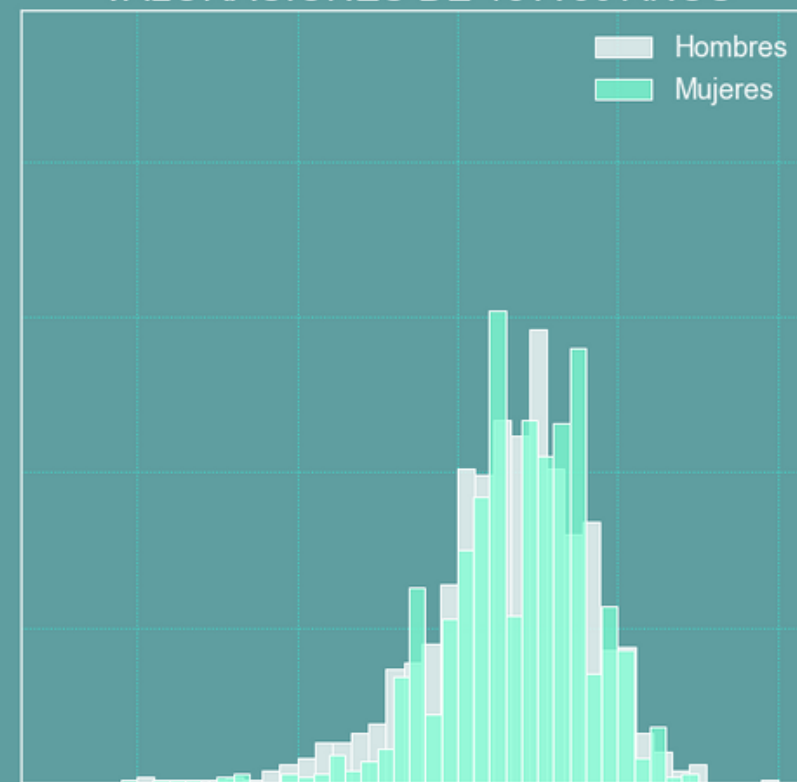
AL SOLAPAR LAS GRÁFICAS SE OBSERVA QUE SON MUY SIMILARES

Tanto por tramos de edad como por sexo las votaciones se mantienen con una distribución constante. La media es prácticamente constante para el grupo que elijamos

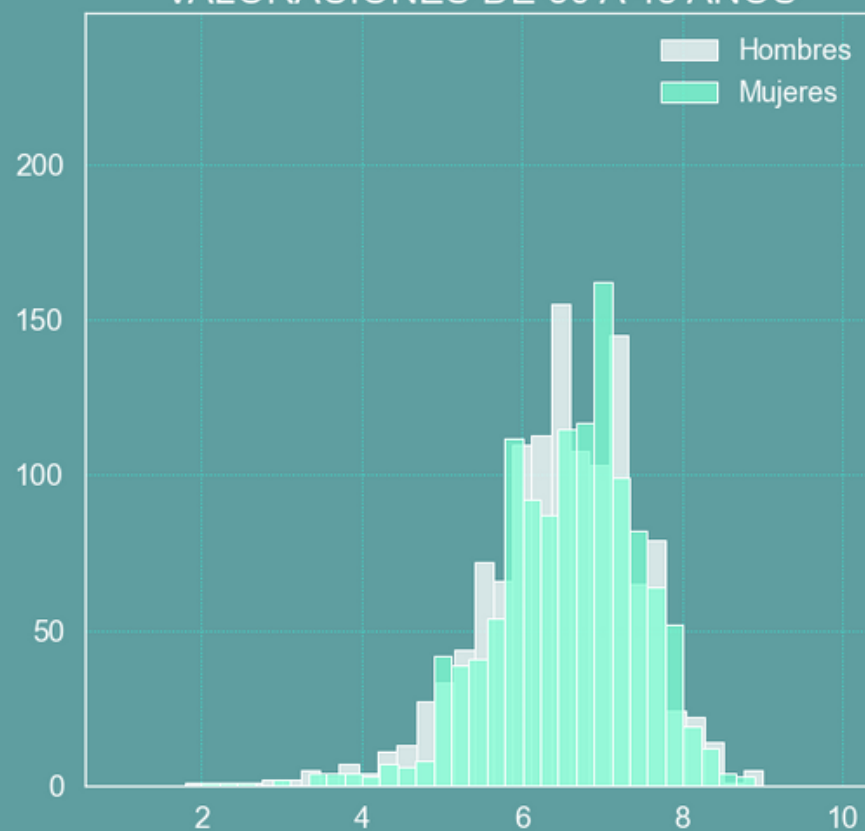
VALORACIONES DE 0 A 18 AÑOS



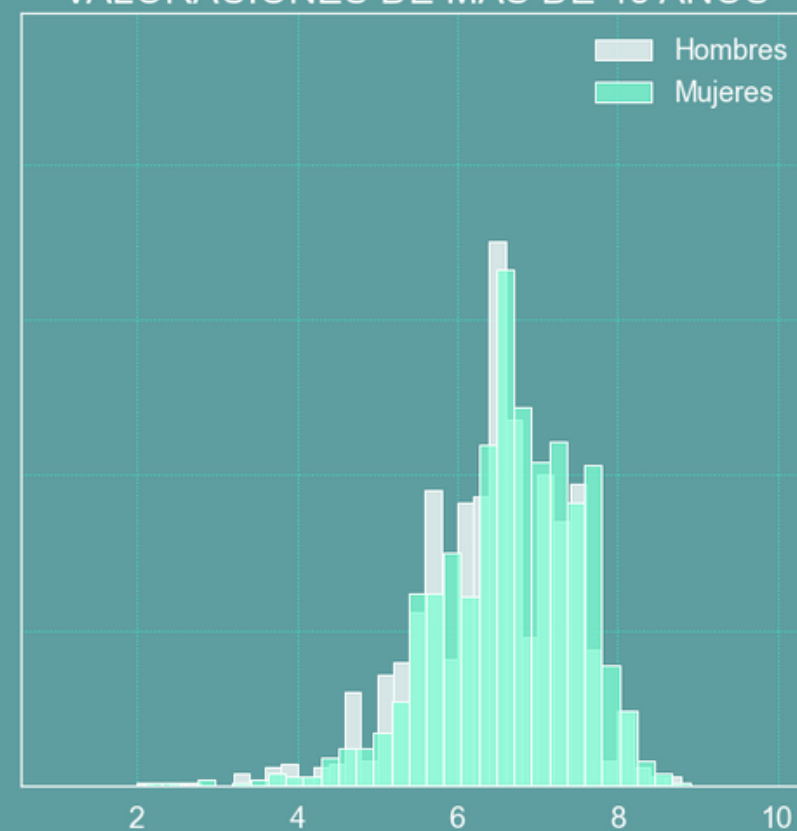
VALORACIONES DE 18 A 30 AÑOS



VALORACIONES DE 30 A 45 AÑOS

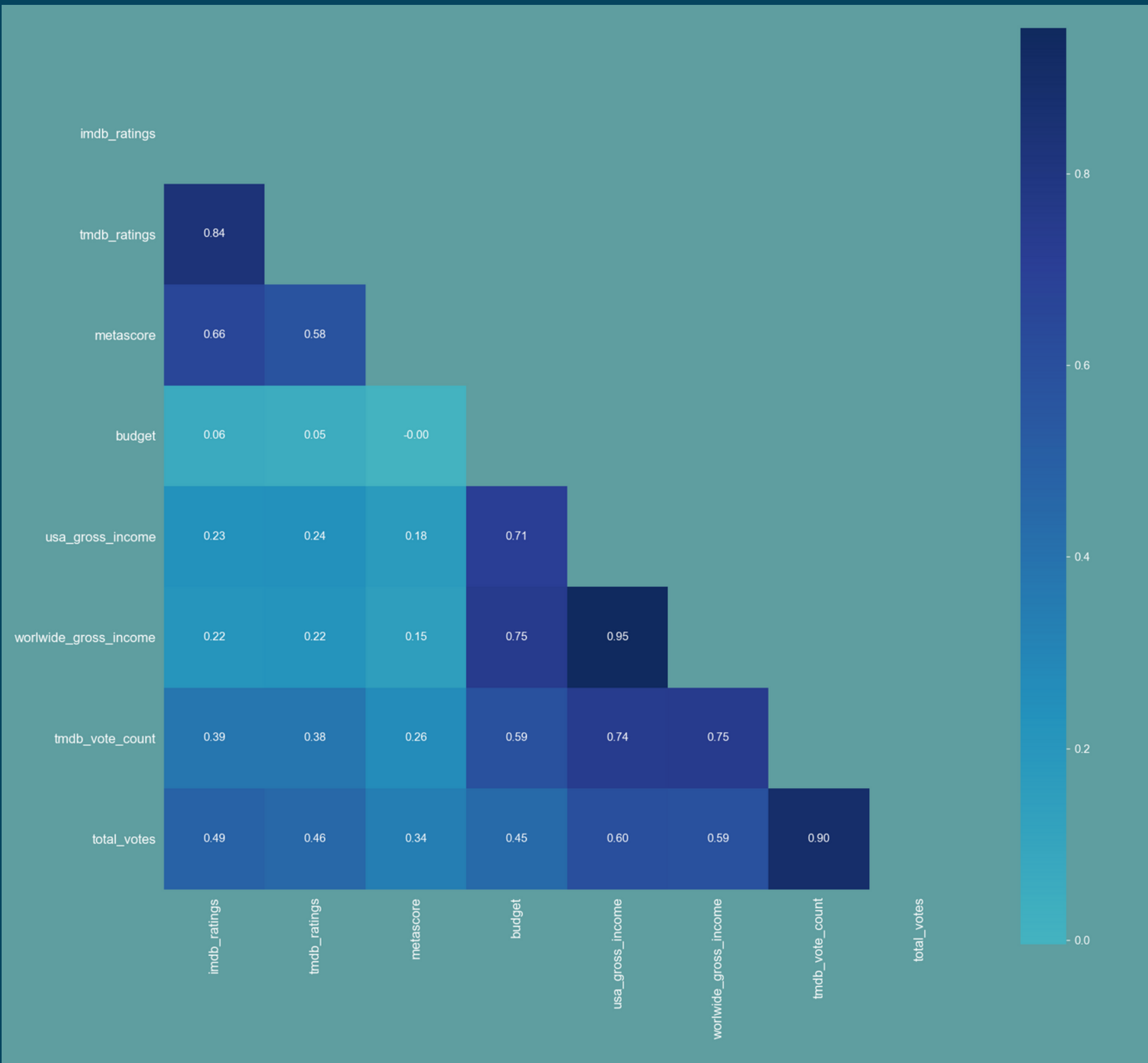


VALORACIONES DE MÁS DE 45 AÑOS



RELACIÓN ENTRE VARIABLES

- Valoraciones
- Presupuesto
- Cantidad de votos
- Recaudación



Las mayores relaciones son entre conceptos similares. Las valoraciones no afectan a la recaudación, pero el presupuesto y la cantidad de votos sí.

CONCLUSIONES

LA RESPUESTA A LA HIPÓTESIS INICIAL

- Las valoraciones de IMDb no afectan a las recaudaciones

ADEMÁS HEMOS CONSEGUIDO SABER

- El presupuesto afecta a las recaudaciones directamente. La publicidad detrás de ese presupuesto será posiblemente la razón
- Las votaciones se reparten igual independientemente del sexo, edad o procedencia.
- La cantidad de votos afecta ligeramente a las recaudaciones, aunque no de manera determinante.

A portrait of Salvador Dalí with a teal overlay. The overlay consists of diagonal bands of different shades of teal. The quote is in a light teal color, and the closing quote mark is white.

SALVADOR DALÍ

**Hablen bien o mal de mí,
lo importante es que
hablen de mí.**

”

MUCHAS GRACIAS

github_linkedin_correo:
Juan_Bayon
Juan Bayon Fernandez
juanbayonfernandez@gmail.com