

## **Práctico 2**

*Visualización de datos*



### **Asignatura:**

Análisis de Datos

### **Integrantes:**

Luciana Barberousse

Alfonso Sobral

### **Fecha:**

21/09/2025

## **Índice**

Índice.....	2
Objetivos.....	3
Alcance.....	3
Explicación de variables.....	3
Desarrollo.....	4
1. ¿Cómo ha cambiado la proporción entre películas y series a lo largo de los años?.....	4
2. ¿Cómo evolucionó la cantidad de estrenos a lo largo del tiempo para cada tipo de contenido?...6	
3. ¿Cuáles son los países con más producciones? ¿Se destacan más por películas o series?.....	7
4. ¿Qué tipo de contenido es más común para cada rating?.....	8
5. ¿Qué países producen más contenido para audiencias adultas e infantiles? ¿Se les ocurre otra separación además de estas dos?.....	9
6. ¿Se identifica alguna estacionalidad en los estrenos según la categoría (listed in)? ¿Qué meses concentran más lanzamientos?.....	10
7. ¿Qué directores tienen más títulos? ¿Se concentran en algún tipo de contenido o en una audiencia específica?.....	14
8. ¿Cuáles son los actores más populares?.....	15
9. ¿Cuál es la distribución en duración en series y películas?.....	16
10. ¿Hay palabras que se utilicen más que otras en títulos y descripciones?.....	17
Conclusiones.....	18
Referencias.....	18
Anexo.....	19
Prompts de Chat GPT.....	32

## Objetivos

Analizar el catálogo de películas y series de Netflix a través de visualizaciones de datos, identificando tendencias en el tiempo, diferencias entre países, géneros, ratings y protagonistas. El objetivo es aplicar las técnicas vistas en clase de análisis y visualización de datos, con énfasis en la claridad de las gráficas y la interpretación de resultados.

## Alcance

El informe se centra exclusivamente en el dataset provisto sobre Netflix. Se abordarán las 10 consignas propuestas en el práctico, utilizando Python y librerías de análisis de datos y visualización (pandas, matplotlib, etc.). El análisis incluirá gráficos, explicaciones y conclusiones, sin extenderse a otros aspectos.

## Explicación de variables

Para el análisis de este conjunto de datos se realizó un Data Profiling, con el objetivo de caracterizar cada una de las variables presentes y evaluar la calidad de los datos. A continuación, se detallan:

- **show\_id**: Identificador único de cada registro del dataset (7.787 en total). Es de tipo *text*, compuesto por la letra “s” seguida de un número natural.
- **type**: Variable categórica que define si el contenido corresponde a una *Movie* o un *TV Show*. Resulta especialmente relevante para varias de las visualizaciones planteadas en este trabajo.
- **title**: Título de la producción. Se trata de un valor único, dado que no pueden coexistir dos títulos idénticos en el catálogo debido a restricciones de derechos de autor. Es de tipo *text*.
- **director**: Nombre del director o directora de la obra. Es de tipo *text* y presenta un 30,7% de valores faltantes, posiblemente atribuibles a la ausencia de información disponible o errores en el registro.
- **cast**: Lista de los actores principales que participan en la producción. Es de tipo *text* y presenta un 9,2% de valores faltantes, cifra considerablemente menor a la observada en la variable *director*.
- **country**: País de origen de la producción. Es de tipo *text* y presenta un 6,5% de valores faltantes.
- **date\_added**: Fecha en la que el contenido fue incorporado al catálogo de Netflix. Es de tipo *date* y prácticamente no presenta valores faltantes.
- **release\_year**: Año en el que la producción fue creada. Es de tipo *real number* y no presenta valores faltantes.
- **rating**: Clasificación etaria de cada contenido (ej.: PG, R, TV-MA). Es de tipo *categorical* y contempla un total de 14 categorías diferentes.
- **duration**: Indica la duración de la producción. En el caso de películas se expresa en minutos, mientras que en series se registra en número de temporadas. Por esta dualidad, se considera una variable de tipo *text*.
- **listed\_in**: Géneros asociados a la película o serie, pudiendo asignarse uno o varios por registro. Es de tipo *text* y no presenta valores faltantes.
- **description**: Breve sinopsis que resume el contenido de la producción. Es de tipo *text* y no presenta valores faltantes.

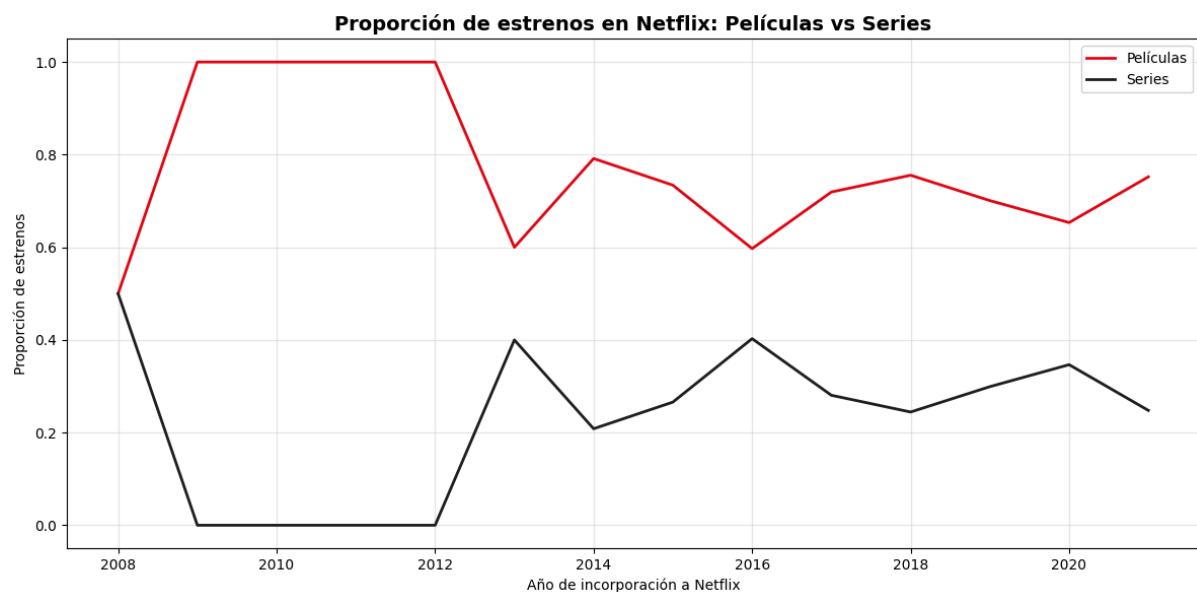
## Desarrollo

A continuación vamos a responder las preguntas que se plantean debajo con gráficos y texto que explique cada consigna.

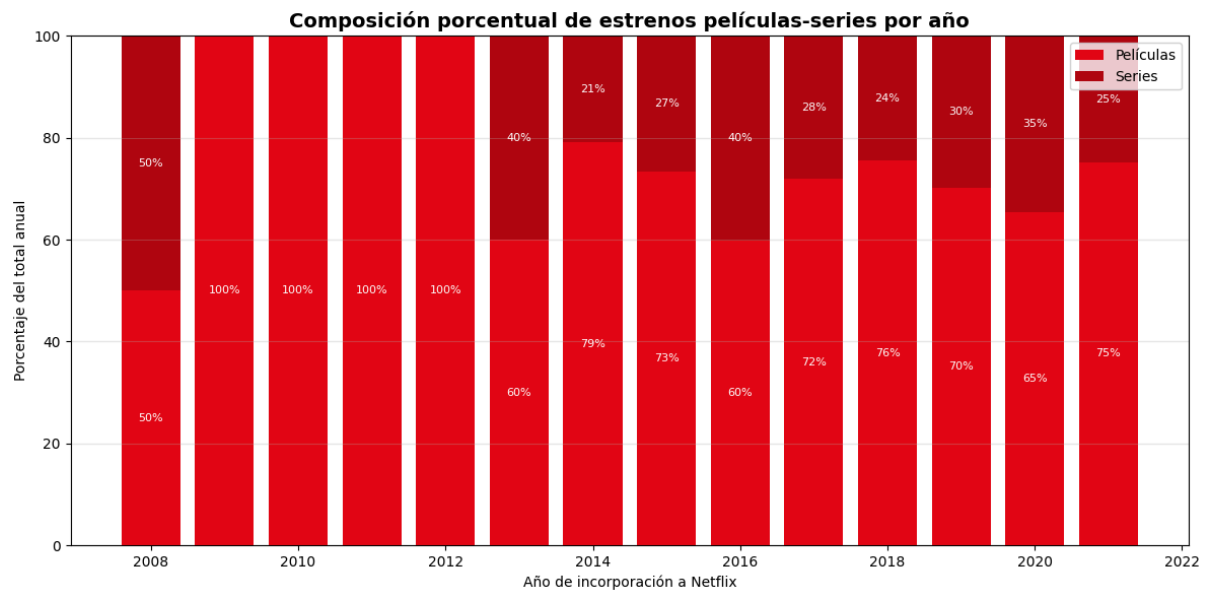
### 1. *¿Cómo ha cambiado la proporción entre películas y series a lo largo de los años?*

El código para esto carga el dataset de Netflix y transforma la columna `date_added` en fecha para extraer el año en que cada título ingresó a la plataforma. Luego agrupa los datos por año y por tipo de contenido (película o serie), calcula la cantidad de estrenos de cada uno y finalmente obtiene la proporción que representan dentro del total de estrenos anuales. Con esos valores se genera un gráfico de líneas que compara la evolución de películas y series a lo largo del tiempo, así como un gráfico de barras que muestra los porcentajes de estrenos de series y películas.

La gráfica muestra cómo, en los primeros años, Netflix incorporaba casi exclusivamente películas, mientras que a partir de 2013 comienzan a ganar espacio las series. Si bien las películas siguen siendo mayoría, se observa un crecimiento sostenido de las series que hace que la oferta se vuelva más equilibrada entre ambos formatos hacia los años más recientes.



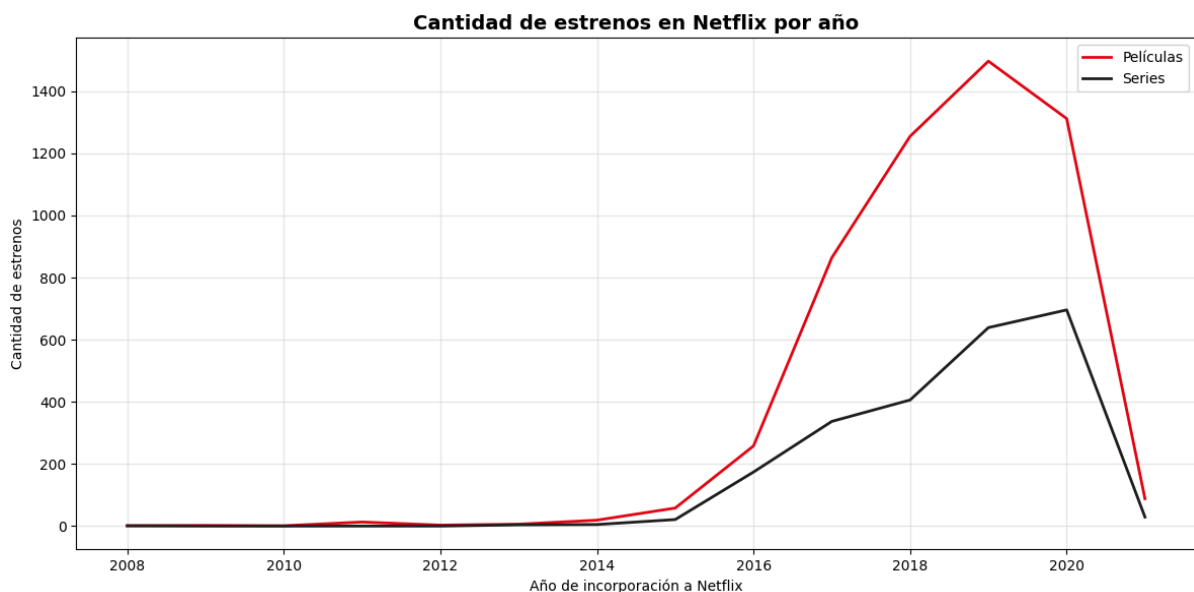
También podríamos representar la gráfica de esta otra forma. Decidimos añadirla para que sea más exacto cual es la proporción exacta.



## 2. ¿Cómo evolucionó la cantidad de estrenos a lo largo del tiempo para cada tipo de contenido?

El código carga el dataset de Netflix y convierte la columna *date\_added* en formato de fecha para obtener el año en que cada título fue incorporado. Luego agrupa los datos por año y tipo de contenido, contando cuántas películas y series se añadieron en cada período. Con esta información se genera un gráfico de líneas en el que cada curva representa la cantidad anual de estrenos de películas o series.

La gráfica muestra la evolución del número de títulos incorporados a Netflix a lo largo del tiempo. Durante los primeros años la cantidad de estrenos fue baja y con poca diferencia entre películas y series. A partir de 2015 se observa un crecimiento acelerado en ambos formatos, con un fuerte predominio de las películas. El pico máximo se alcanzó en 2019, donde se sumaron más de 1400 películas y alrededor de 650 series. En los años posteriores se nota una caída en el volumen de estrenos, aunque las películas continúan siendo el formato más numeroso. Esto refleja el crecimiento explosivo de la plataforma y su posterior estabilización.

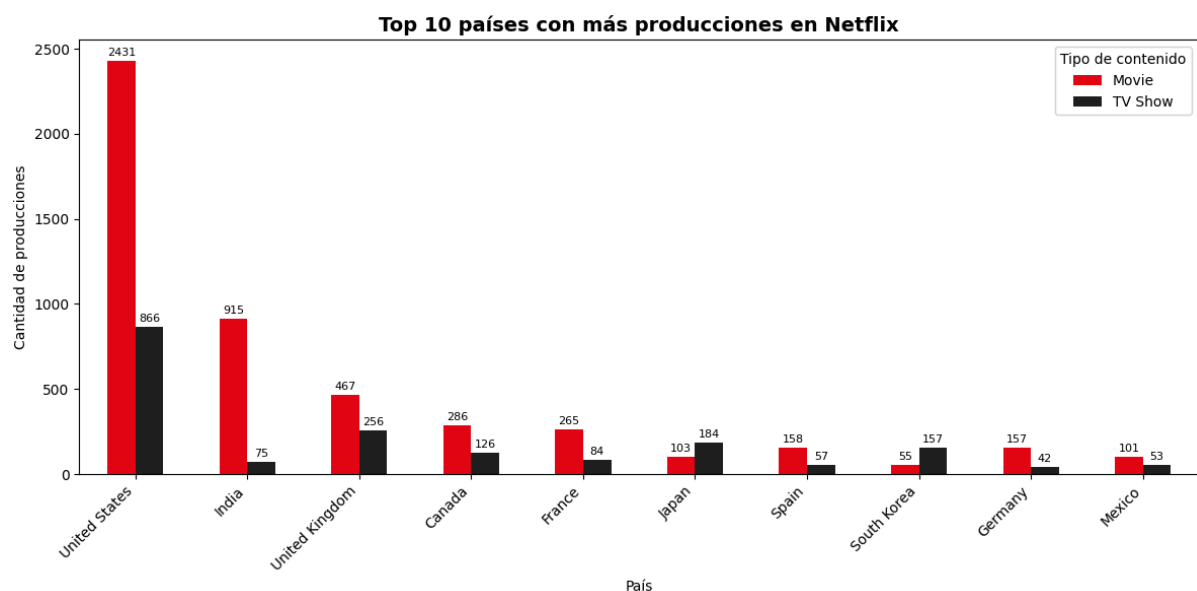


### 3. ¿Cuáles son los países con más producciones? ¿Se destacan más por películas o series?

El código comienza cargando el dataset de Netflix y limpiando la columna *country*, ya que en muchos casos un título puede pertenecer a más de un país. Para manejar esa situación, la columna se separa en varias filas, de modo que cada país cuente por separado. Después, los datos se agrupan por país y tipo de contenido (Movie o TV Show), calculando cuántas producciones corresponden a cada combinación. A continuación, se identifican los diez países con mayor número total de producciones y se reorganizan los datos para construir una tabla que permita graficar. Finalmente, se genera un gráfico de barras donde cada país aparece con dos barras: una roja para películas y otra negra para series, y se añaden etiquetas numéricas encima de cada barra para facilitar la lectura. Además los 10 países se encuentran ordenados en orden descendente en el total de producciones, ya sea series o películas.

La gráfica resultante muestra que Estados Unidos domina ampliamente el catálogo de Netflix, con más de 2400 películas y más de 800 series, muy por encima del resto de los países. Le sigue India, con un fuerte predominio de películas, y el Reino Unido, que combina una cantidad importante de ambos tipos de contenido. Otros países como Canadá, Francia, Japón, México, Corea del Sur, España y Alemania aparecen en el ranking, aunque con cifras mucho menores. En general, se observa que en la mayoría de los países predominan las películas, aunque en algunos casos, como Japón, Corea del Sur o el Reino Unido, las series tienen una presencia más significativa. Esto refleja tanto el peso de la industria cinematográfica en algunos mercados como la expansión global de la producción de series en la plataforma.

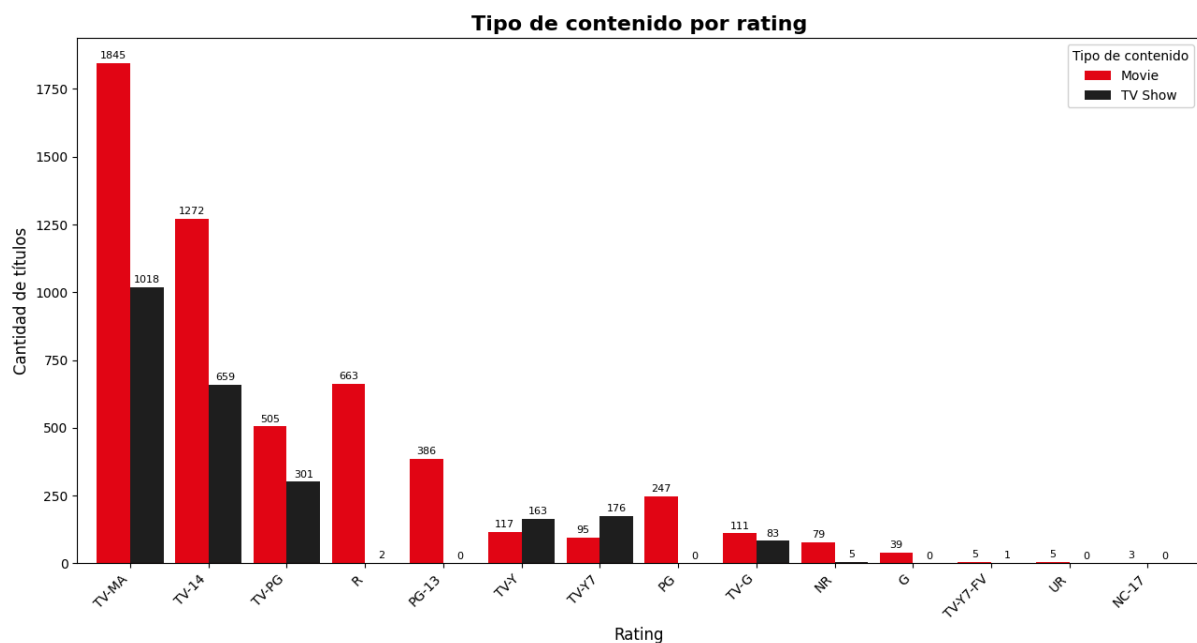
Cabe destacar que, dentro de los países del top 10, predominan las producciones de películas, con excepción de Japón y Corea del Sur, donde las series tienen un peso más relevante. Esto puede explicarse porque las series suelen requerir un mayor seguimiento y tiempos de grabación más prolongados, lo que implica costos adicionales en actores, escenarios, dirección y producción en general.



#### 4. ¿Qué tipo de contenido es más común para cada rating?

El código carga el dataset de Netflix y se centra en las columnas *type* y *rating*, eliminando los valores nulos y limpiando espacios en blanco. Luego agrupa la información por tipo de contenido (películas o series) y por clasificación (rating), contabilizando cuántos títulos corresponden a cada combinación. Para ordenar los resultados, calcula el total de títulos por rating y organiza las categorías de mayor a menor. Después convierte la tabla en un formato adecuado para graficar y genera un gráfico de barras agrupadas, donde cada clasificación aparece en el eje X y se comparan las cantidades de películas (en rojo) y series (en negro). Finalmente, se ajusta el tamaño de la figura en función del número de categorías y se agregan etiquetas con los valores exactos encima de cada barra.

La gráfica resultante muestra qué tipo de contenido predomina en cada clasificación. Se observa que las categorías más frecuentes en Netflix son TV-MA y TV-14, que concentran la mayor cantidad de títulos, tanto películas como series, seguidas por TV-PG y R. En la mayoría de los ratings, las películas superan en número a las series, aunque en ciertos casos como TV-Y o TV-Y7 las series tienen un peso relativamente mayor. También aparecen clasificaciones menos comunes como UR y NC-17, con muy pocos títulos. En conjunto, la gráfica refleja que el catálogo de Netflix se enfoca principalmente en contenidos para audiencias adolescentes y adultas.

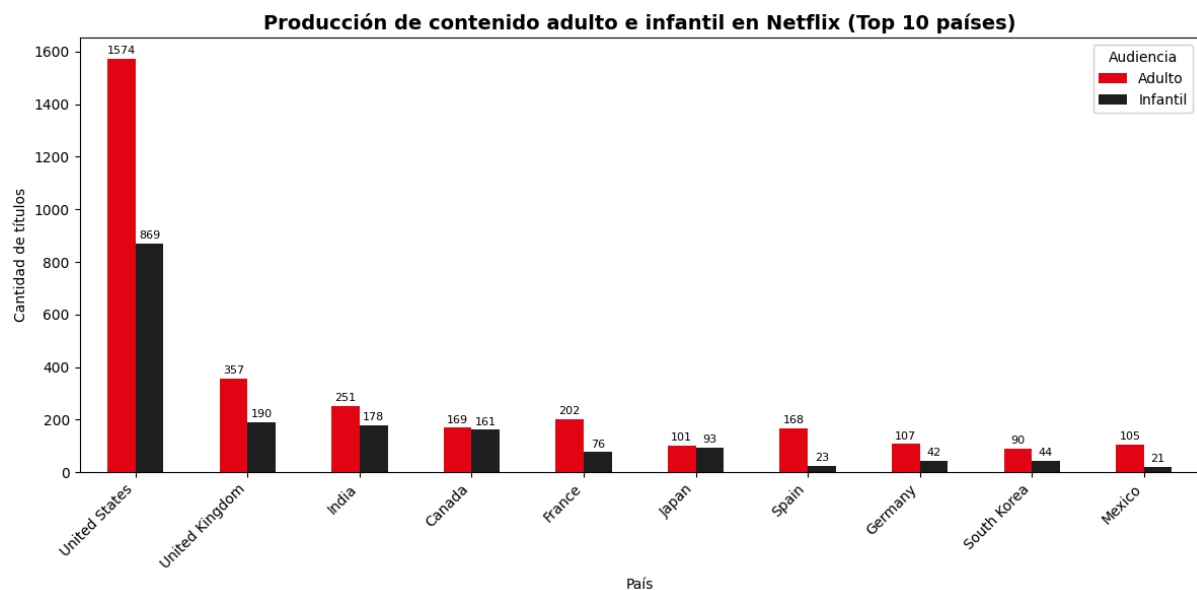




5. ¿Qué países producen más contenido para audiencias adultas e infantiles? ¿Se les ocurre otra separación además de estas dos?

El código carga el dataset de Netflix y se queda con *country* y *rating*. Primero clasifica cada título en una categoría de audiencia a partir del rating: Adulto (TV-MA, R, NC-17), Infantil (TV-Y, TV-Y7, TV-G, PG, TV-PG, G), (además contempla Adolescente y Otro, aunque en el gráfico no se usan). Luego limpia la columna *country* para separar títulos con múltiples países en filas individuales, cuenta cuántos títulos de cada audiencia tiene cada país, selecciona los 10 países con más títulos y arma una tabla pivote con las columnas Adulto e Infantil. Por último, ordena los países de mayor a menor según la suma de ambas audiencias y dibuja un gráfico de barras agrupadas, coloreando Adulto en rojo e Infantil en negro y agregando etiquetas numéricas sobre cada barra.

La gráfica compara, para los 10 países líderes, cuántos títulos son de audiencia adulta versus infantil. Se observa que Estados Unidos domina ampliamente en ambas categorías, seguido por Reino Unido e India; el resto (Canadá, Francia, Japón, España, Alemania, Corea del Sur y México) aparece bastante por detrás. En la mayoría de los países predomina el contenido adulto sobre el infantil, aunque algunos (por ejemplo Canadá o Japón) muestran una brecha menor entre ambas audiencias. En conjunto, el gráfico deja ver tanto el ranking de producción por país como su sesgo relativo hacia contenidos para adultos o para públicos infantiles.



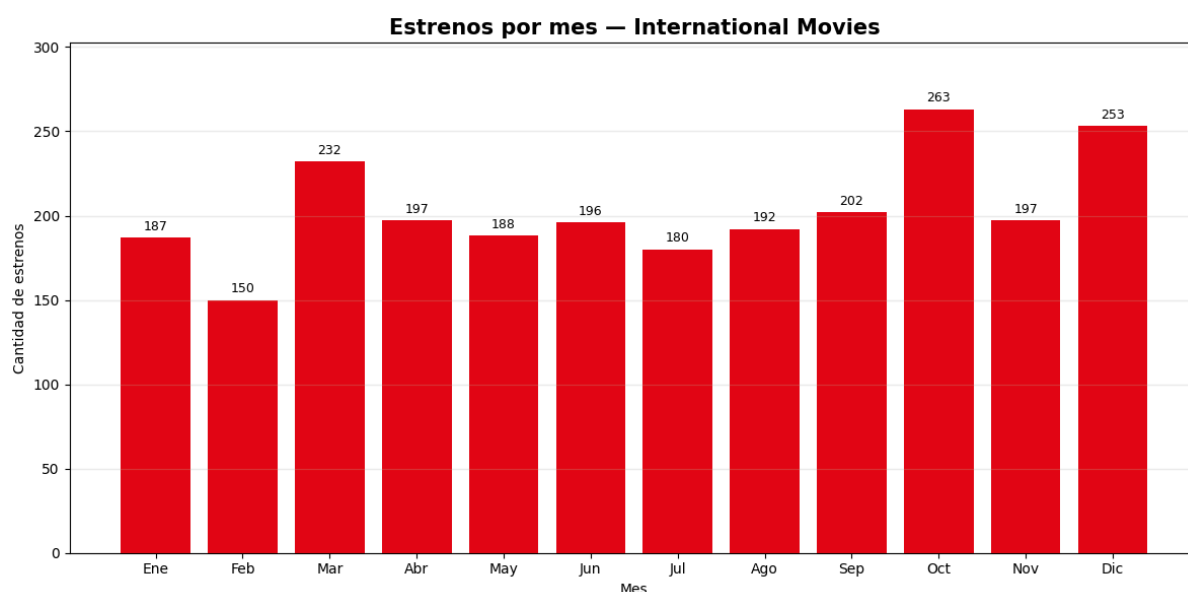
Otra categoría que podría incorporarse es la de Adolescentes, como un nivel intermedio entre Infantil y Adulto, que incluiría clasificaciones como PG-13 y TV-14. Asimismo, podría contemplarse una categoría de no definidos u “otros”, destinada a títulos con ratings poco claros o sin clasificación específica (por ejemplo, NR, UR, etc.).

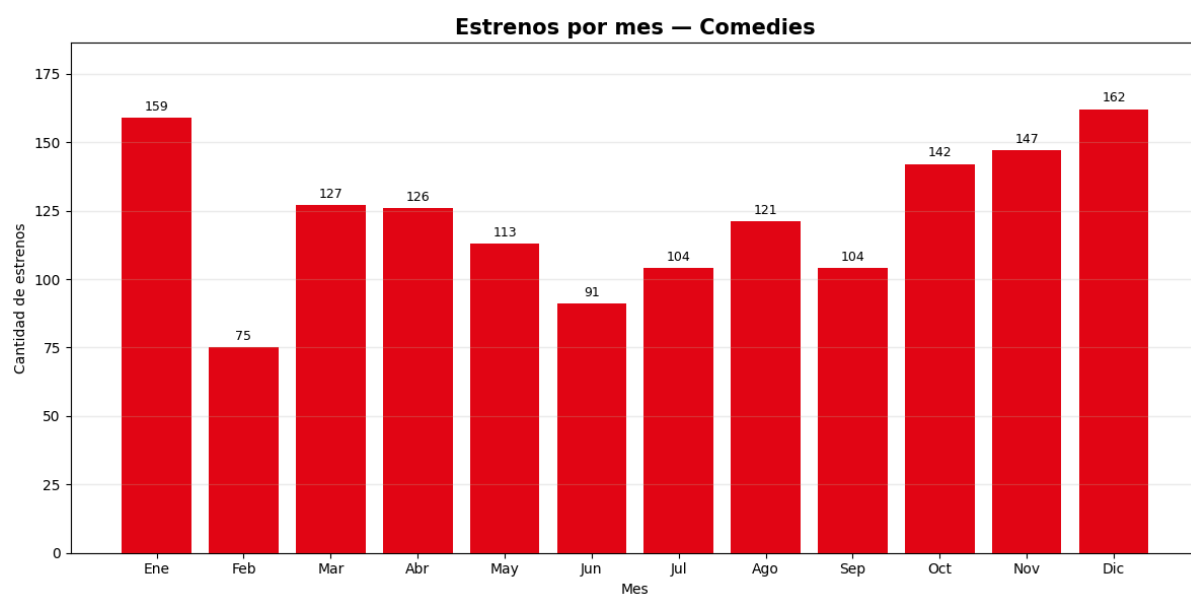
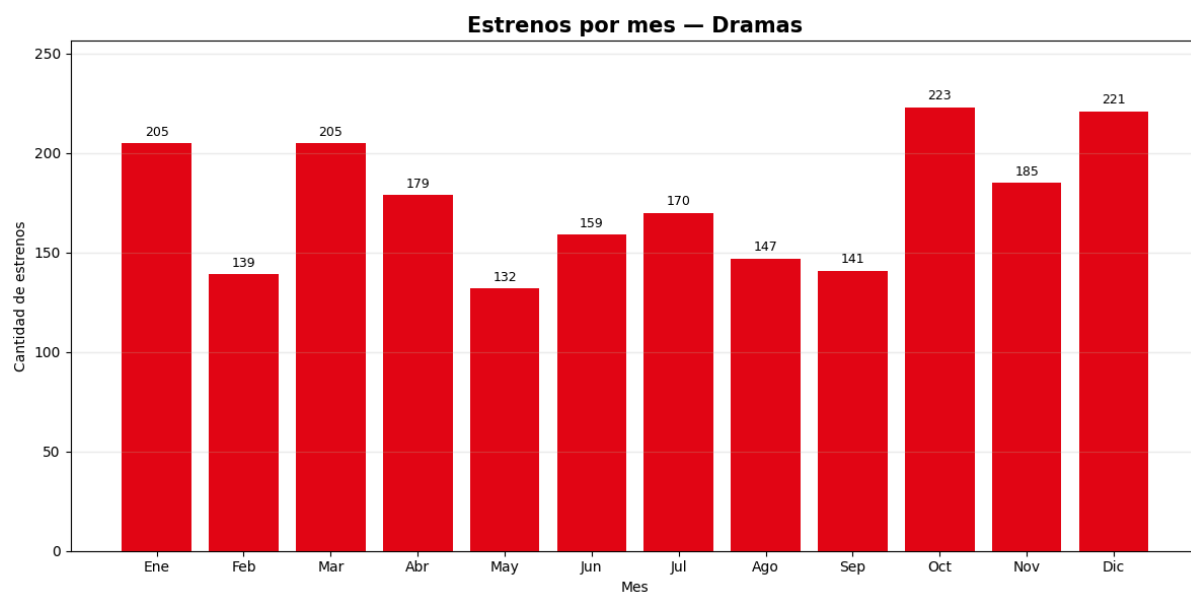
6. ¿Se identifica alguna estacionalidad en los estrenos según la categoría (*listed in*)? ¿Qué meses concentran más lanzamientos?

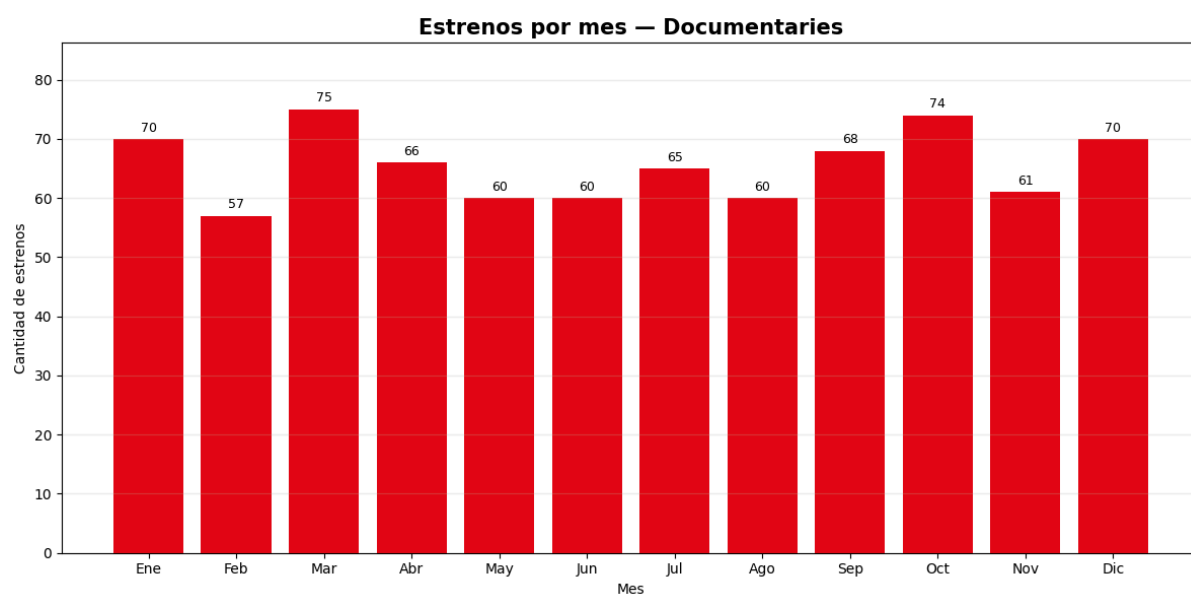
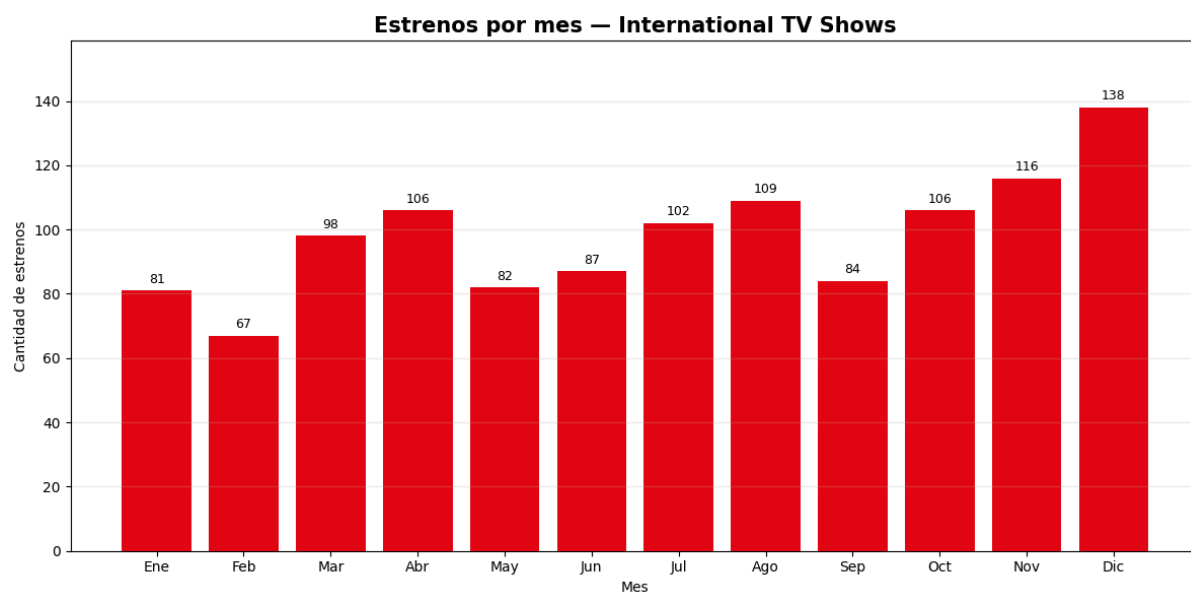
Primero realizamos un código que analiza los estrenos de Netflix por categorías y meses. Primero carga el dataset, convierte la columna *date\_added* en formato de fecha y extrae el mes en que cada título ingresó a la plataforma. Luego descompone la columna *listed\_in*, ya que un mismo título puede pertenecer a varias categorías, y reorganiza los datos para contar cuántos estrenos hubo por mes dentro de cada categoría. Con esos conteos genera un gráfico de barras individual para cada categoría, donde el eje X representa los meses del año y el eje Y la cantidad de estrenos, incluyendo además las cifras exactas encima de cada barra.

Estas gráficas permiten identificar si en ciertas categorías existe estacionalidad, es decir, meses en los que se concentran más lanzamientos. Por ejemplo, se podría observar que en categorías como Comedias o Dramas los estrenos se distribuyen más parejos durante el año, mientras que en otras, como Documentales o Navidad, pueden aparecer picos marcados en determinados meses (como diciembre). En conjunto, estas visualizaciones ayudan a entender la estrategia de estrenos de Netflix diferenciada por tipo de contenido.

Para mantener la claridad del informe y evitar una sobrecarga de información, dado que existen 42 categorías diferentes entre películas y series, se presentarán aquí únicamente las cinco con mayor cantidad de producciones, mientras que el detalle completo se incluirá en el anexo.

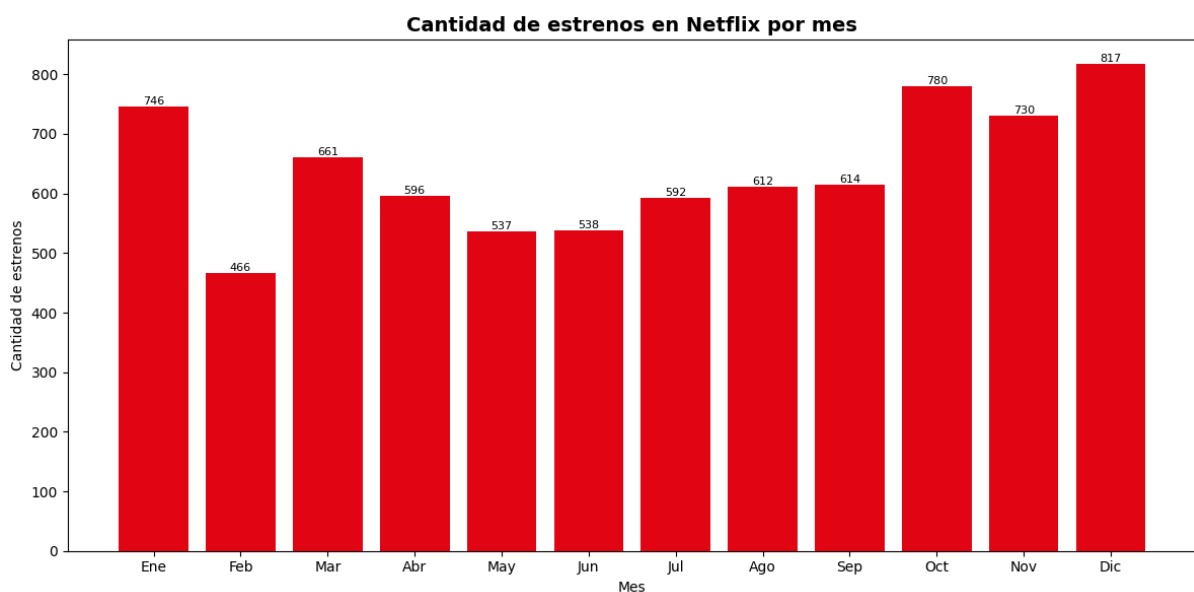






Por otro lado, realizamos un código para ver en qué mes se concentraban más lanzamientos, que carga el dataset de Netflix y convierte la columna *date\_added* en formato de fecha para poder extraer el mes en que cada título fue incorporado a la plataforma. A partir de allí, se agrupan todos los registros según el mes de incorporación y se cuentan los estrenos correspondientes a cada uno. Finalmente, con esos datos se genera un gráfico de barras donde el eje X representa los meses del año y el eje Y la cantidad total de títulos agregados.

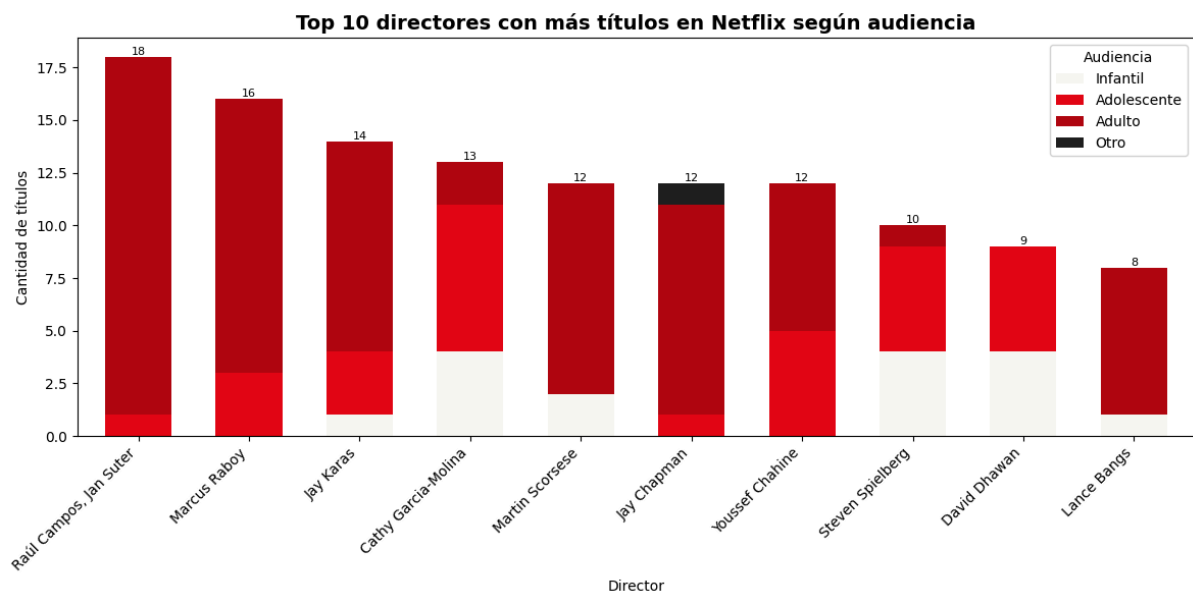
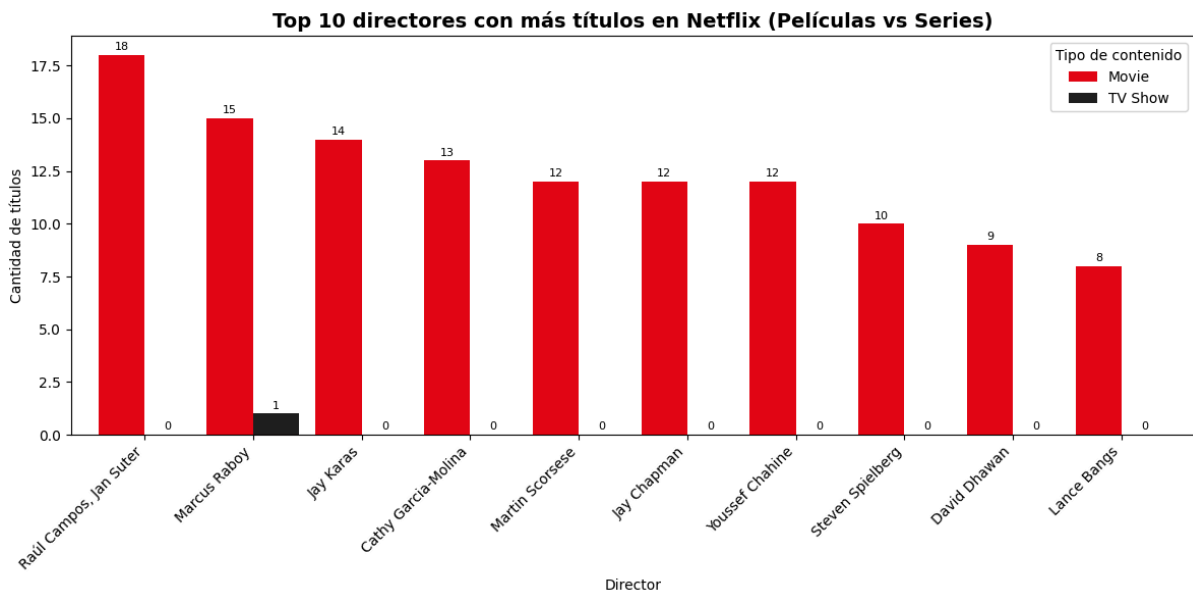
La gráfica resultante muestra la distribución de estrenos de Netflix a lo largo de los meses. Se observa que diciembre concentra la mayor cantidad de lanzamientos (817), seguido de cerca por octubre (780) y enero (746), lo que sugiere una estrategia vinculada a las vacaciones y a la mayor demanda de entretenimiento en esas fechas. En contraste, febrero aparece como el mes con menos estrenos. El resto de los meses mantiene valores más equilibrados, lo que indica que, aunque existe cierta estacionalidad, Netflix distribuye sus lanzamientos de forma relativamente uniforme a lo largo del año.



7. ¿Qué directores tienen más títulos? ¿Se concentran en algún tipo de contenido o en una audiencia específica?

El código para esta pregunta analiza los directores con mayor presencia en el catálogo de Netflix. Primero selecciona a los 10 directores con más títulos y genera un gráfico de barras comparando cuántas de sus producciones corresponden a películas y cuántas a series. Luego, clasifica los títulos según la audiencia a la que están dirigidos (infantil, adolescente, adulto u otros) y construye un gráfico de barras apiladas que muestra la distribución de estas categorías, agregando el total de títulos en la parte superior de cada barra.

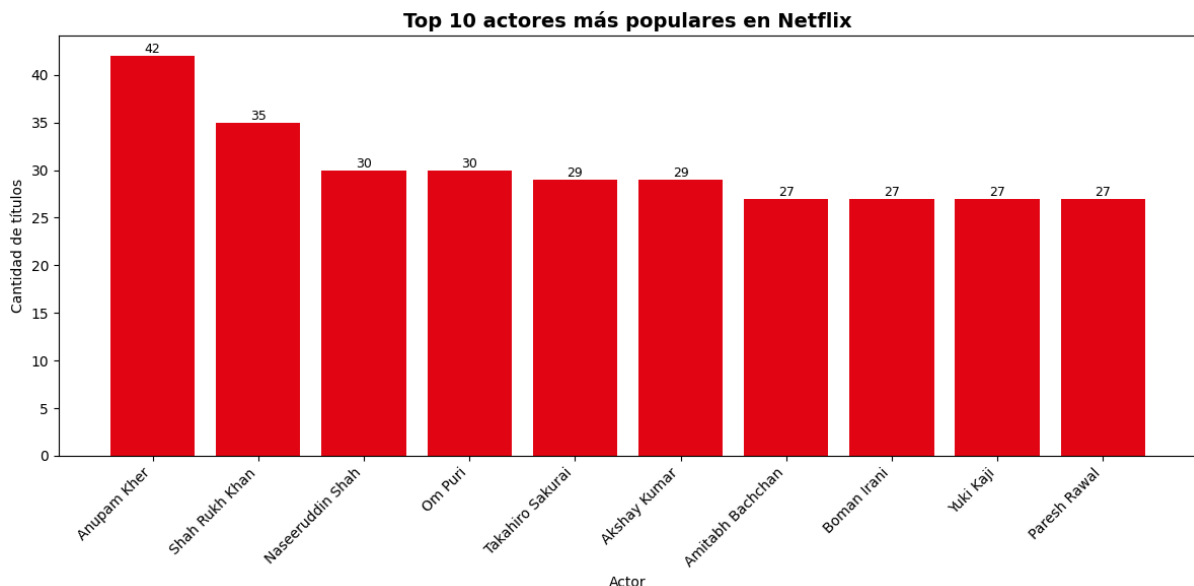
En las gráficas se observa que la gran mayoría de los directores del top 10 concentran sus producciones en películas, con escasa o nula participación en series. Asimismo, en el análisis por audiencias, predomina claramente el contenido para adultos, seguido en menor medida por títulos adolescentes e infantiles. Esto refleja tanto la orientación de Netflix hacia directores vinculados al cine como la preferencia por producciones dirigidas a un público adulto dentro de su catálogo más destacado.



## 8. ¿Cuáles son los actores más populares?

En esta pregunta utilizamos un código que analiza los actores con mayor presencia en el catálogo de Netflix. Primero carga el dataset, elimina los registros sin información en la columna *cast*, y separa en filas distintas a los actores cuando un título incluye varios nombres. Luego cuenta la cantidad de veces que aparece cada actor y selecciona a los 10 más frecuentes. Finalmente, construye un gráfico de barras donde cada barra representa a un actor y su altura indica la cantidad de títulos en los que participa, con el número exacto mostrado encima de cada barra.

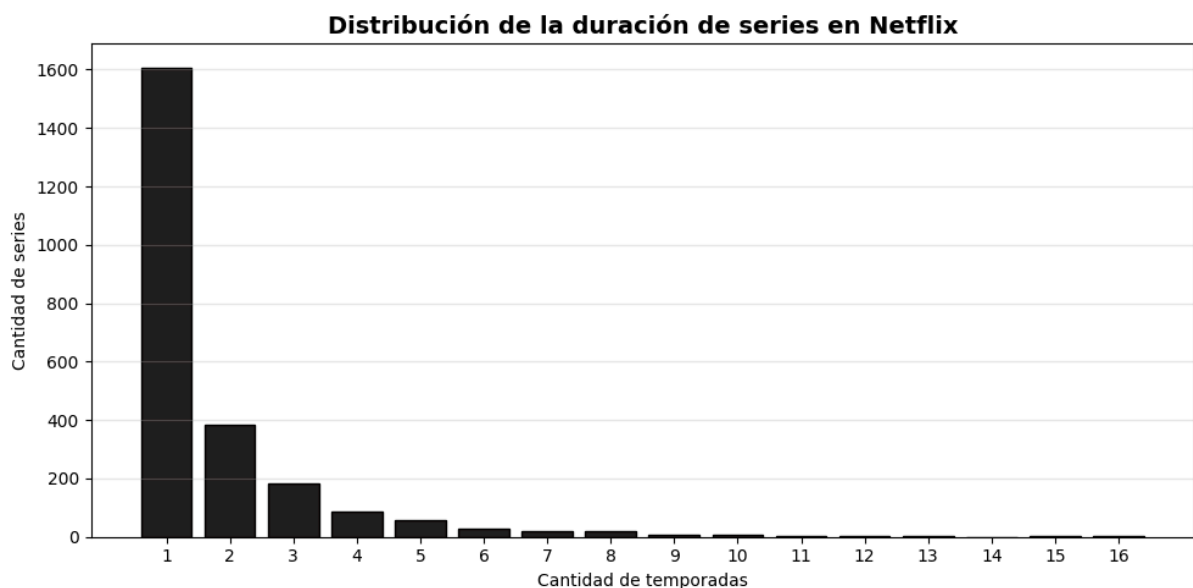
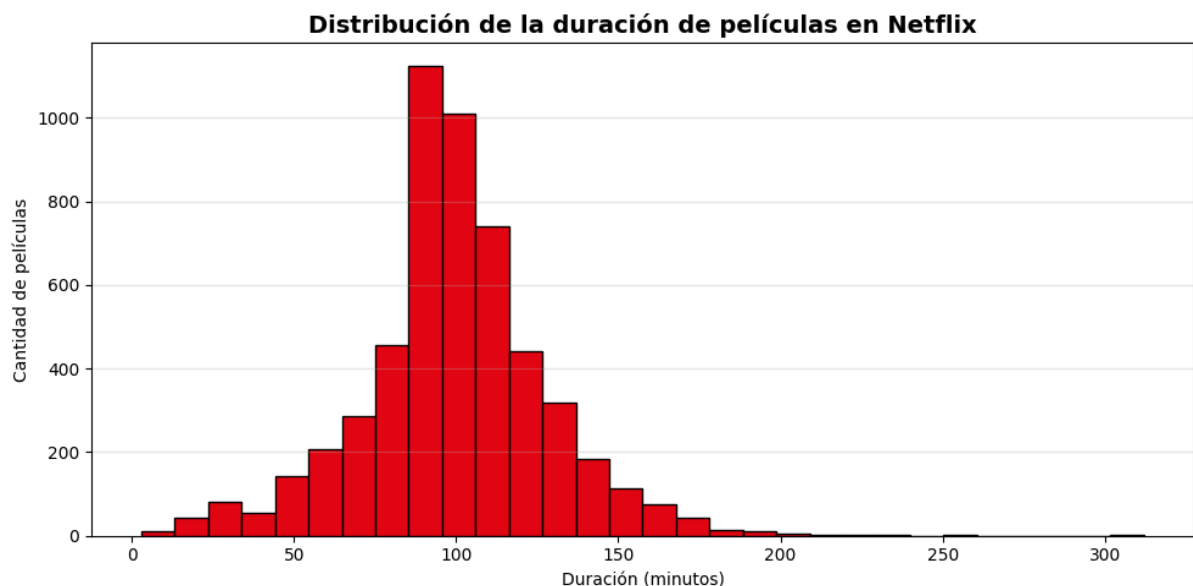
La gráfica resultante muestra que los actores con más apariciones en Netflix son principalmente figuras del cine y televisión de la India y Asia, destacándose Anupam Kher y Shah Rukh Khan como los más populares, con más de 35 títulos cada uno. Este análisis refleja cómo ciertos actores concentran una gran parte del contenido disponible, evidenciando la fuerte presencia del cine indio en el catálogo de la plataforma.



### 9. ¿Cuál es la distribución en duración en series y películas?

Este código estudia la distribución de la duración de las películas y series en Netflix. Para las películas, extrae de la columna *duration* la duración expresada en minutos y genera un histograma que muestra cuántos títulos caen en cada rango de tiempo. Para las series, en cambio, interpreta la duración como la cantidad de temporadas y construye otro histograma que indica cuántas series existen con 1, 2, 3 o más temporadas.

La gráfica de películas refleja una fuerte concentración alrededor de los 90 a 120 minutos, lo que confirma que la mayoría de los largometrajes de la plataforma se ajustan al estándar de cine tradicional. También se observa una caída progresiva hacia películas de mayor duración, con muy pocos títulos por encima de las 2 horas y media. En el caso de las series (según el segundo gráfico), la distribución suele mostrar que la mayoría cuentan con pocas temporadas, predominando aquellas con 1 o 2 temporadas, lo que sugiere que Netflix ofrece muchas producciones limitadas o de corta extensión.





El código genera una nube de palabras (WordCloud) con las palabras más frecuentes en los títulos del catálogo de Netflix. Para lograrlo, primero carga el dataset y procesa la columna *title*, convirtiendo todos los textos a minúsculas y eliminando valores nulos. Luego separa los títulos en palabras, descartando aquellas muy comunes o poco informativas (stopwords como the, de, la, and), y crea una lista con todos los términos relevantes.

La nube resultante muestra que palabras como “love”, “girl”, “world”, “story” y “life” destacan por su alta recurrencia, lo que refleja las temáticas más comunes de las producciones en la plataforma: historias románticas, personajes femeninos, relatos de vida y tramas de alcance global.



## **Conclusiones**

El análisis realizado sobre el catálogo de Netflix permitió identificar una serie de patrones y tendencias relevantes en la evolución y características de su oferta de contenidos.

En primer lugar, se observó que en los primeros años la plataforma se enfoca casi exclusivamente en películas, mientras que a partir de 2013 las series comenzaron a ganar protagonismo. Aunque las películas siguen siendo mayoría, las series muestran un crecimiento sostenido que ha equilibrado progresivamente la oferta. Asimismo, el volumen de estrenos alcanzó su pico en 2019, con una posterior disminución en años recientes, lo que refleja una etapa de expansión acelerada seguida de cierta estabilización.

En cuanto al origen geográfico de los contenidos, Estados Unidos domina ampliamente, seguido a distancia por India y el Reino Unido. La mayoría de los países producen principalmente películas, salvo excepciones como Japón y Corea del Sur, donde las series tienen un peso más destacado. Además, se confirmó que el catálogo se orienta en gran medida hacia audiencias adultas y adolescentes, mientras que el contenido infantil ocupa un lugar menos central.

Respecto a las clasificaciones por rating, predominan las categorías TV-MA y TV-14, lo que evidencia la importancia del contenido dirigido a un público maduro. De forma complementaria, los análisis de duración mostraron que las películas suelen ajustarse al estándar de 90 a 120 minutos, mientras que la mayoría de las series cuentan con 1 o 2 temporadas, lo que indica una preferencia por producciones cortas o limitadas.

En términos de protagonistas y creadores, los directores y actores más frecuentes provienen mayoritariamente de la industria cinematográfica india, lo que refleja la fuerte presencia de Bollywood en la plataforma. Entre los directores se observa una clara concentración en películas y, en cuanto a audiencias, un predominio del contenido para adultos. Entre los actores, destacan nombres como Anupam Kher y Shah Rukh Khan con una gran cantidad de apariciones.

Finalmente, el análisis de estacionalidad reveló que los meses con mayor concentración de estrenos son diciembre, octubre y enero, lo que sugiere una estrategia vinculada a temporadas festivas y de mayor consumo audiovisual. Por último, la nube de palabras mostró que términos como love, girl, world, story y life son los más recurrentes en los títulos, evidenciando el peso de temáticas universales como el romance, la vida cotidiana y las historias con alcance global.

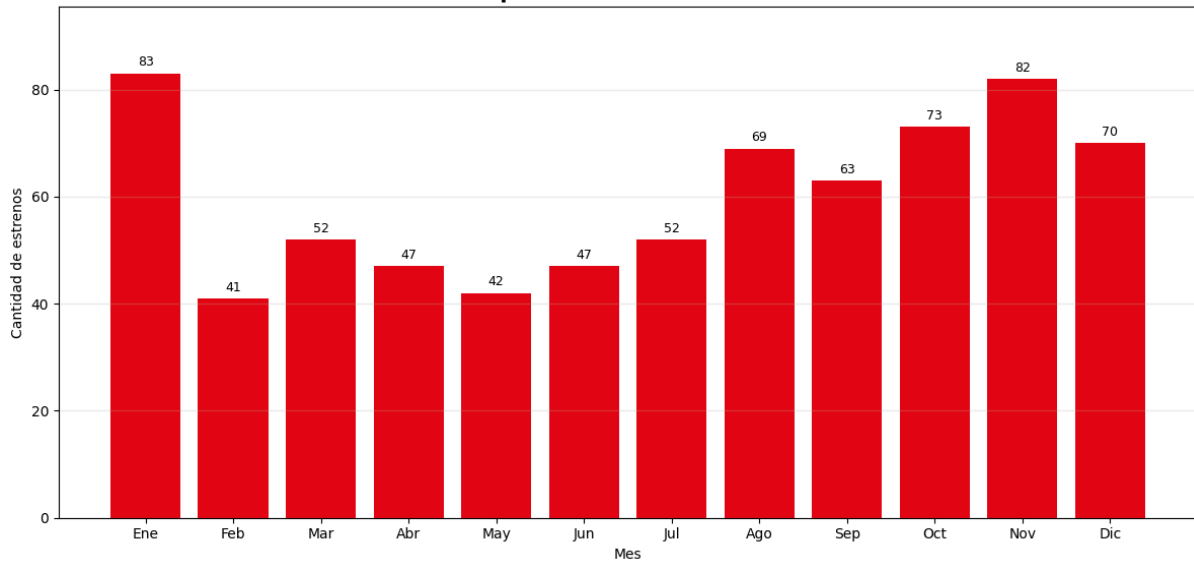
En conjunto, el trabajo permitió aplicar técnicas de análisis y visualización de datos para comprender mejor cómo Netflix organiza y distribuye sus contenidos, confirmando tendencias de mercado y mostrando la utilidad de estas herramientas para interpretar grandes volúmenes de información de forma clara y visual.

## **Referencias**

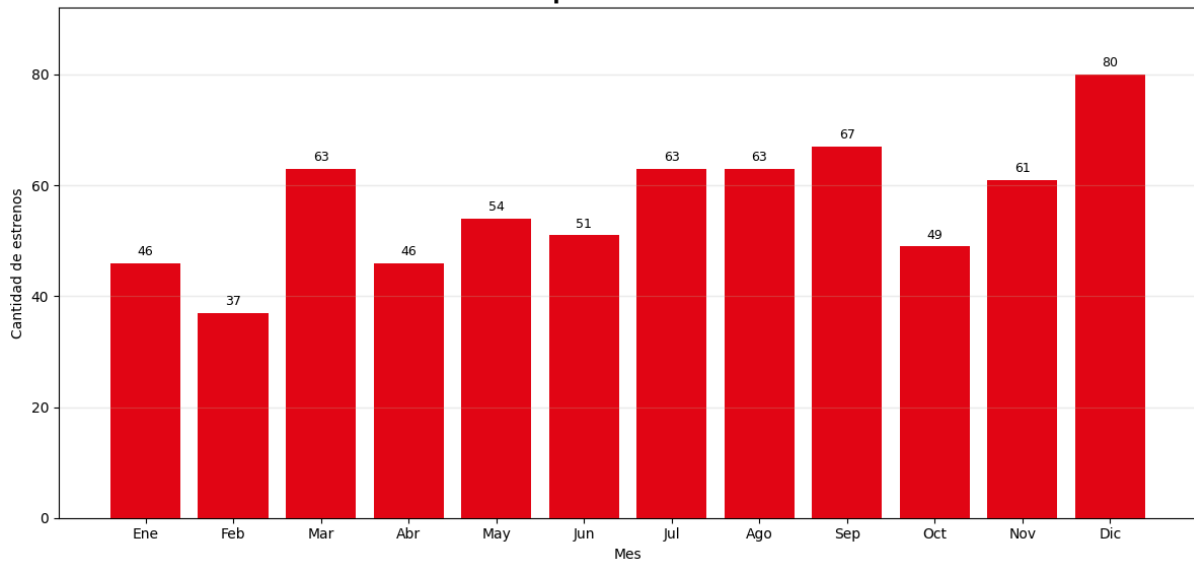
Para realizar el trabajo utilizamos principalmente los conceptos dados en clase y algunas herramientas de IA para aprender un poco más acerca de pandas y de python en general. A su vez, se utilizó para ayudar a mejorar la redacción.

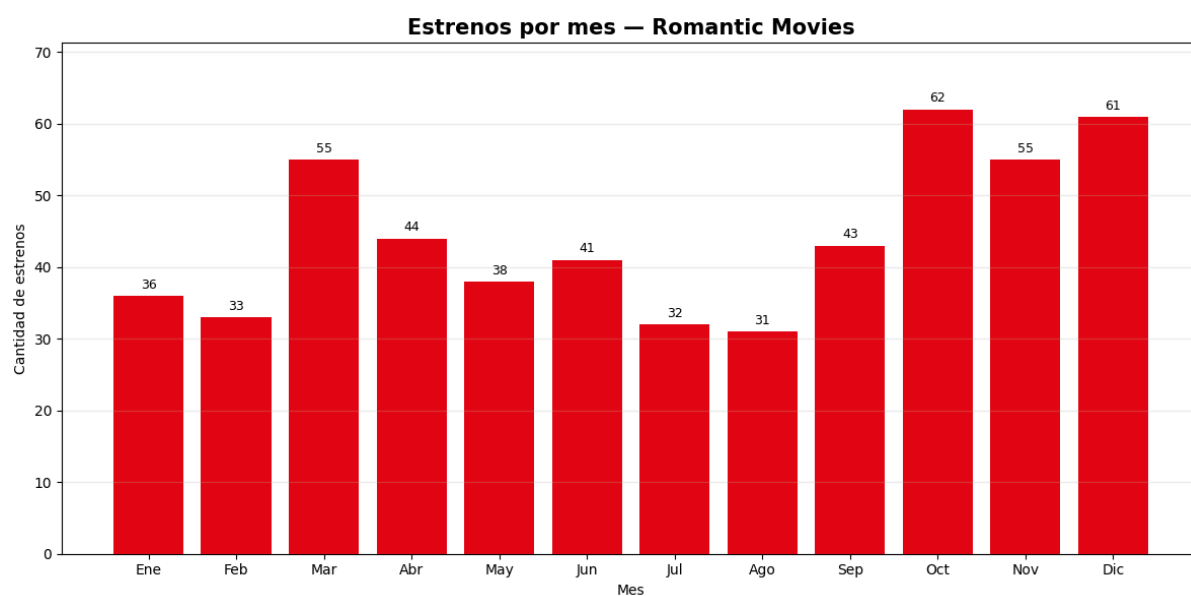
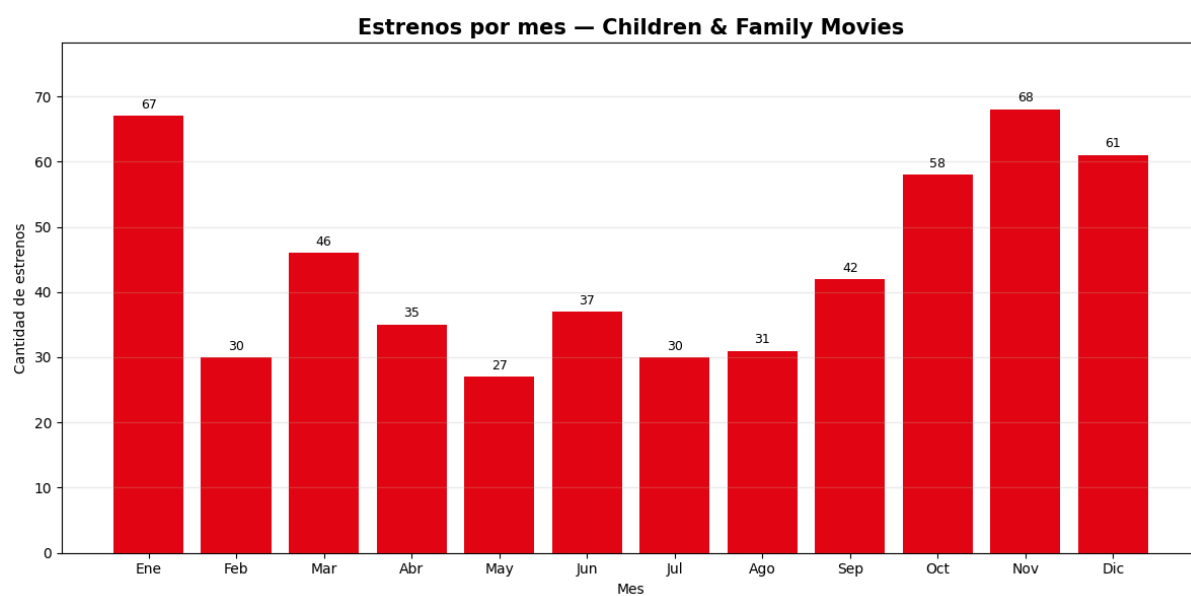
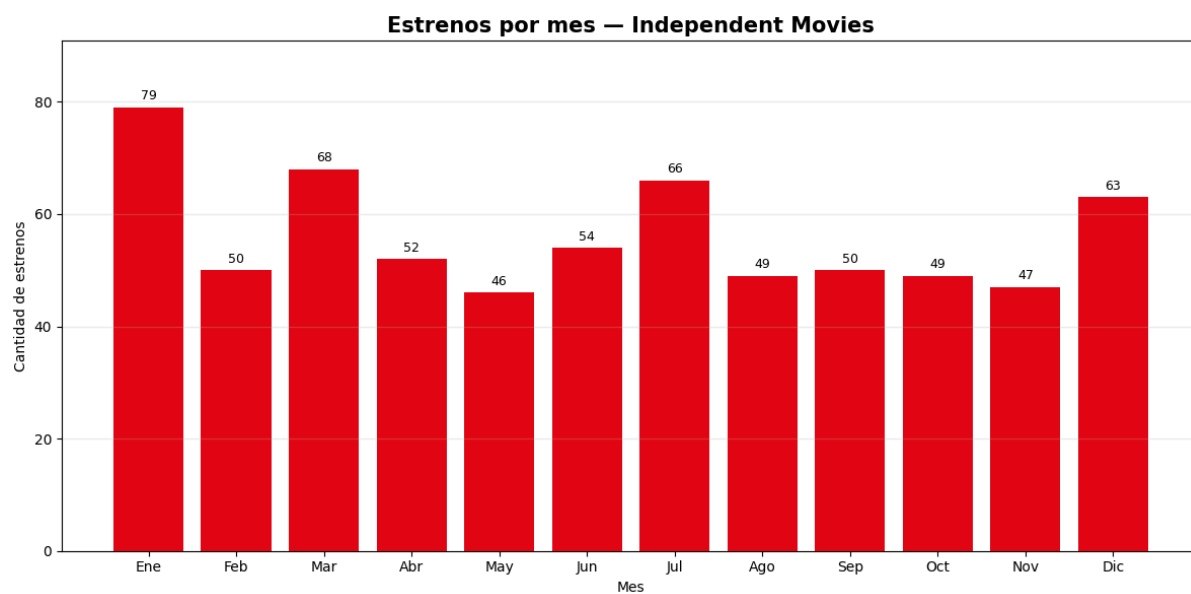
## Anexo

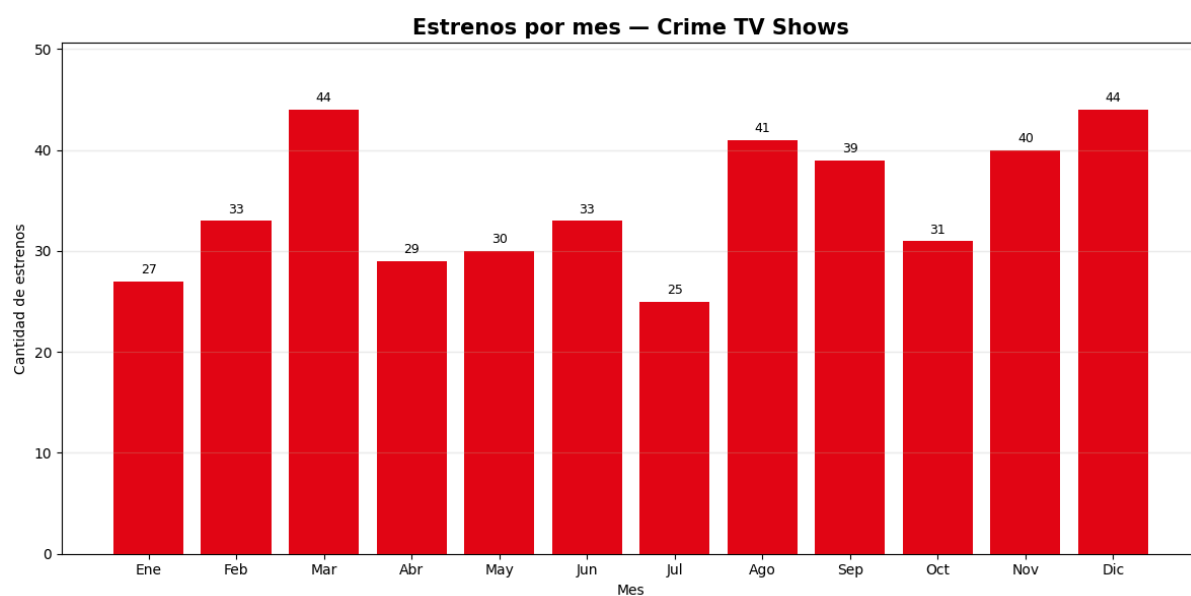
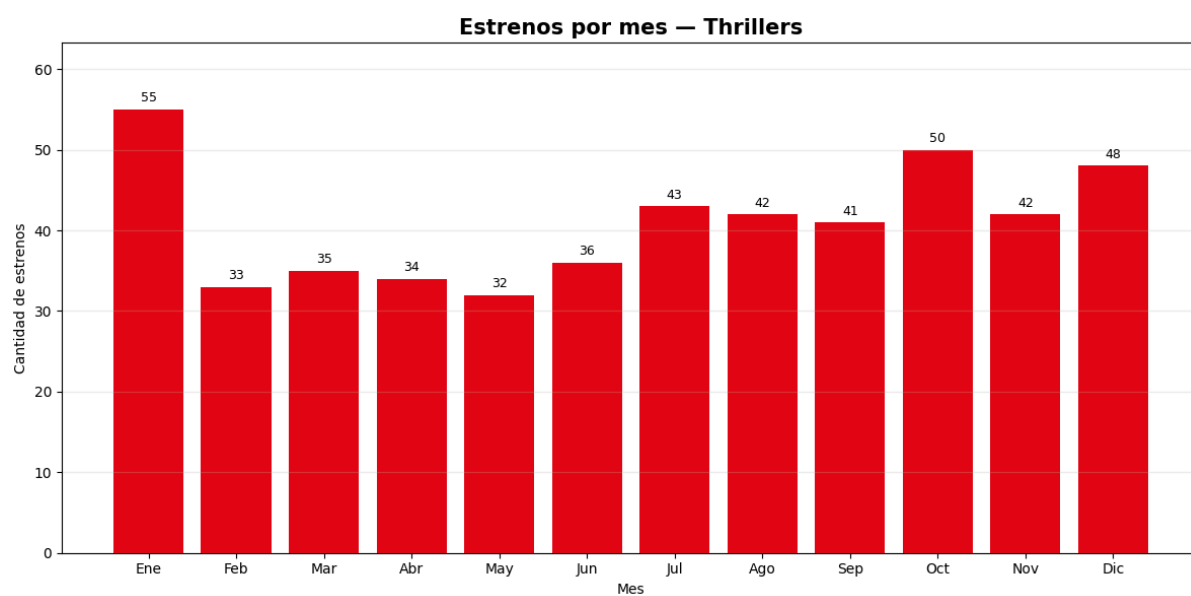
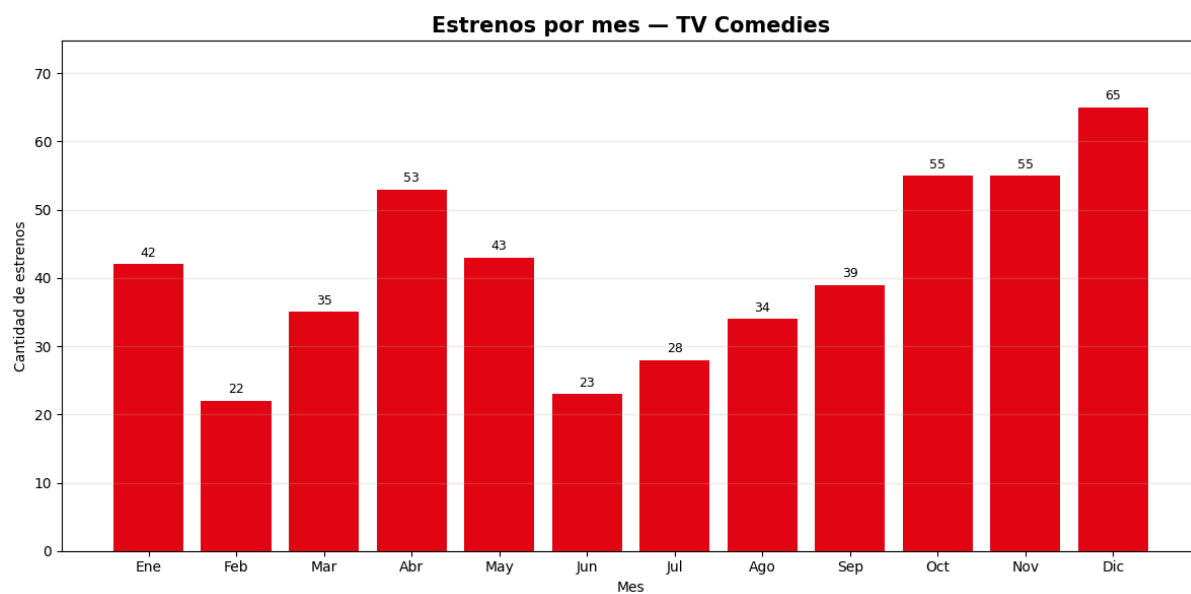
**Estrenos por mes — Action & Adventure**

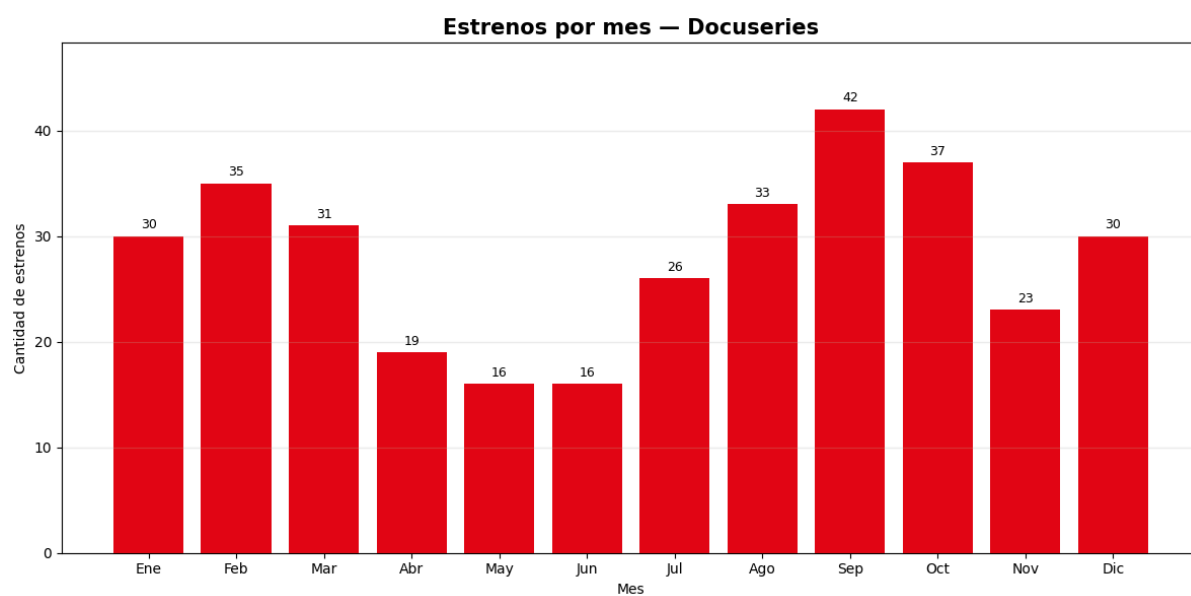
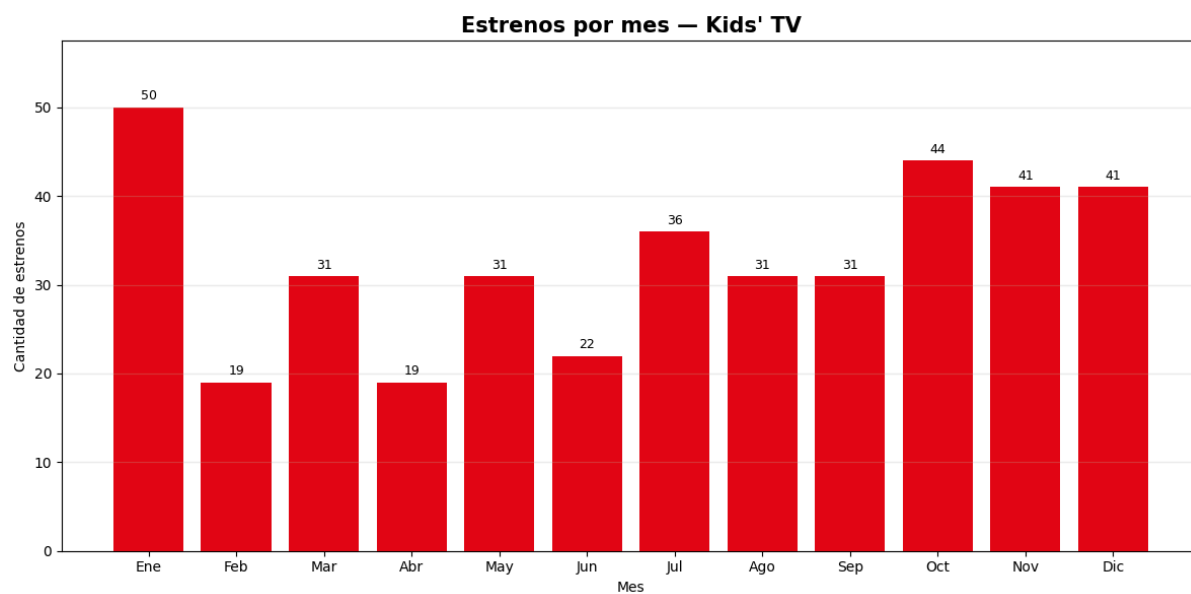


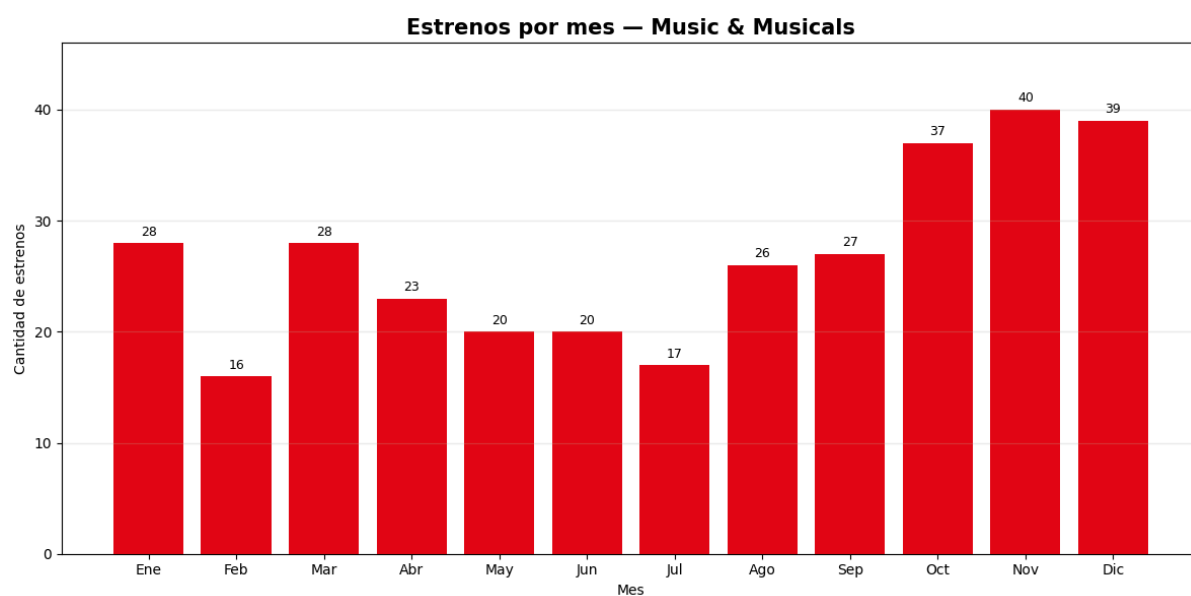
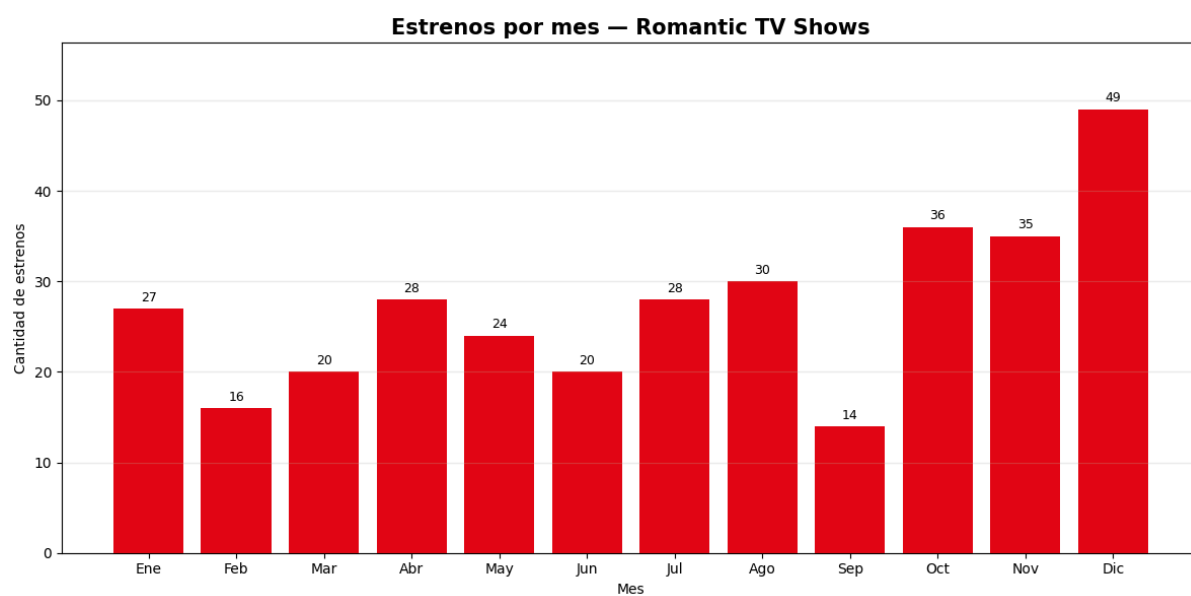
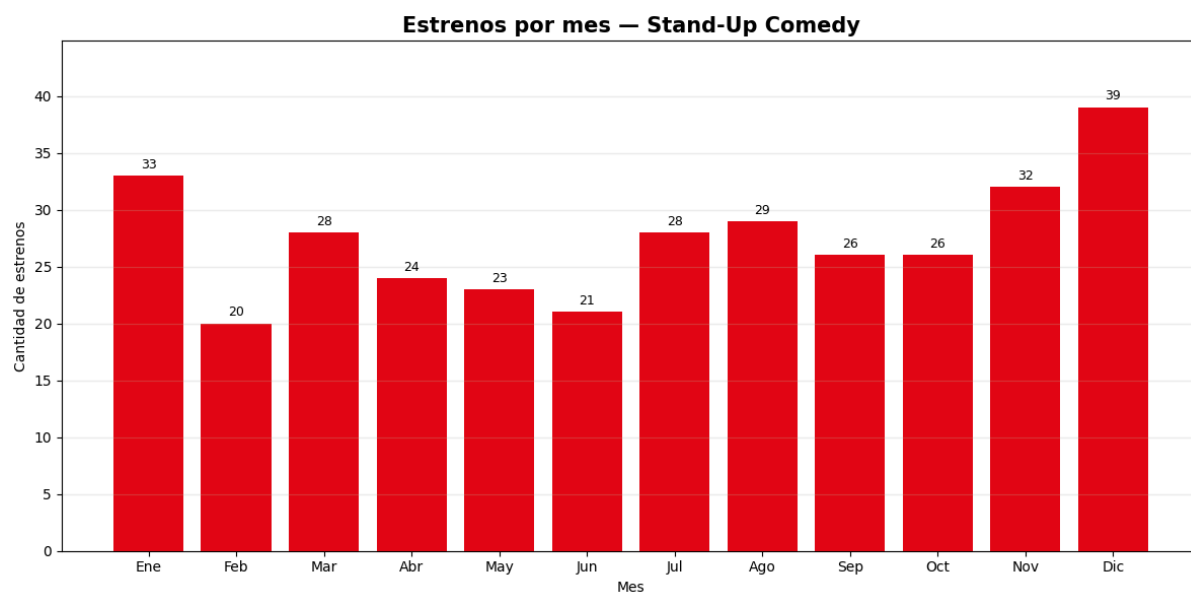
**Estrenos por mes — TV Dramas**

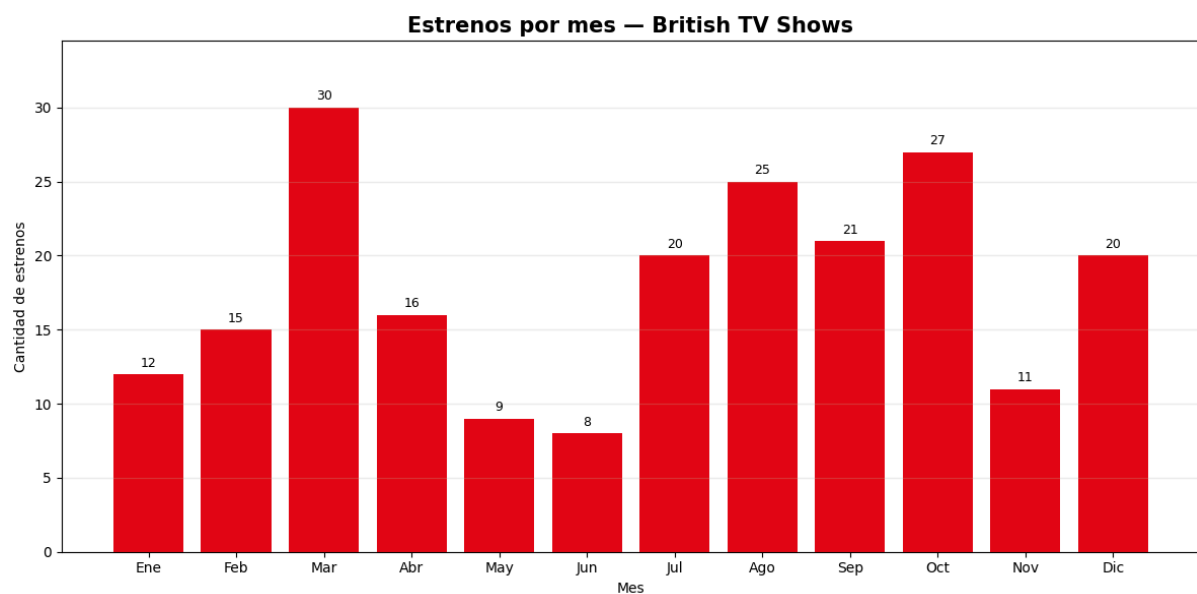
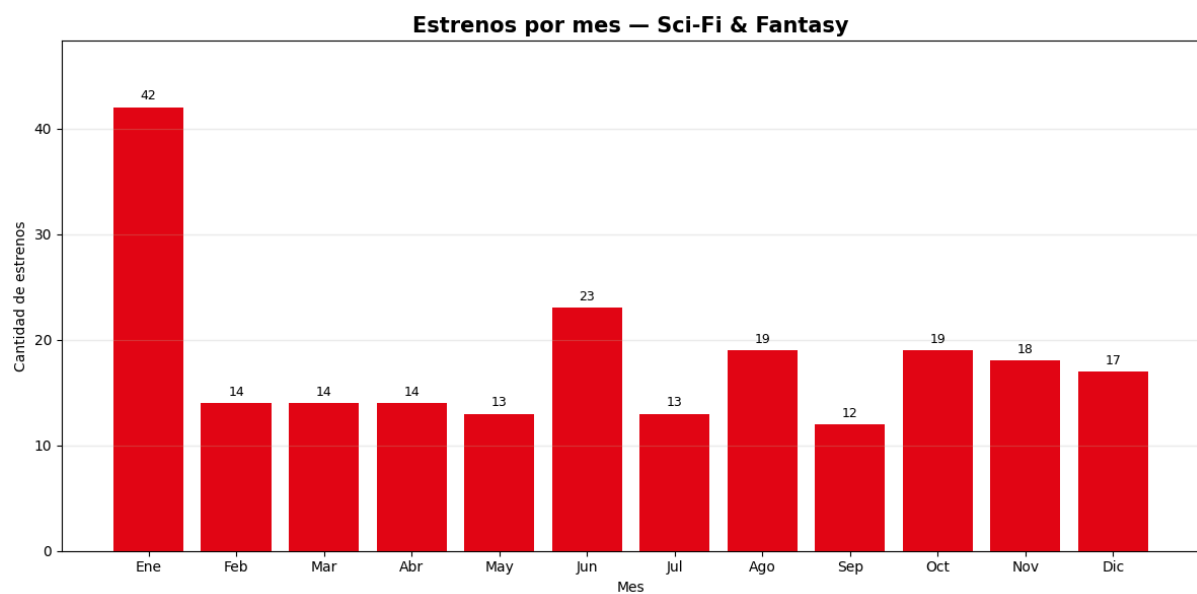
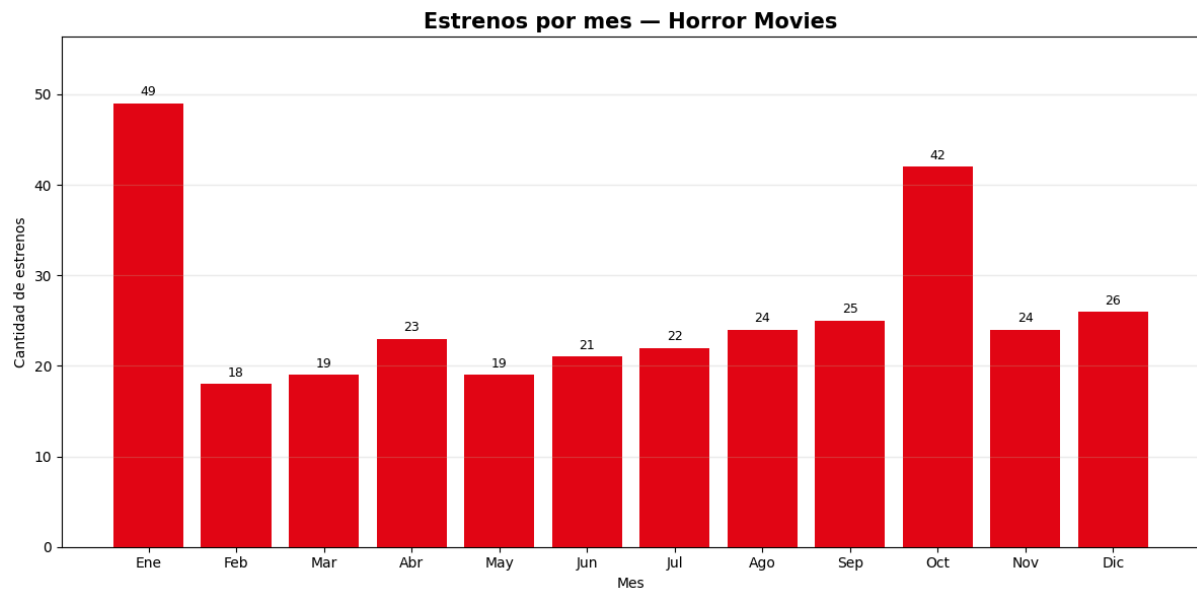




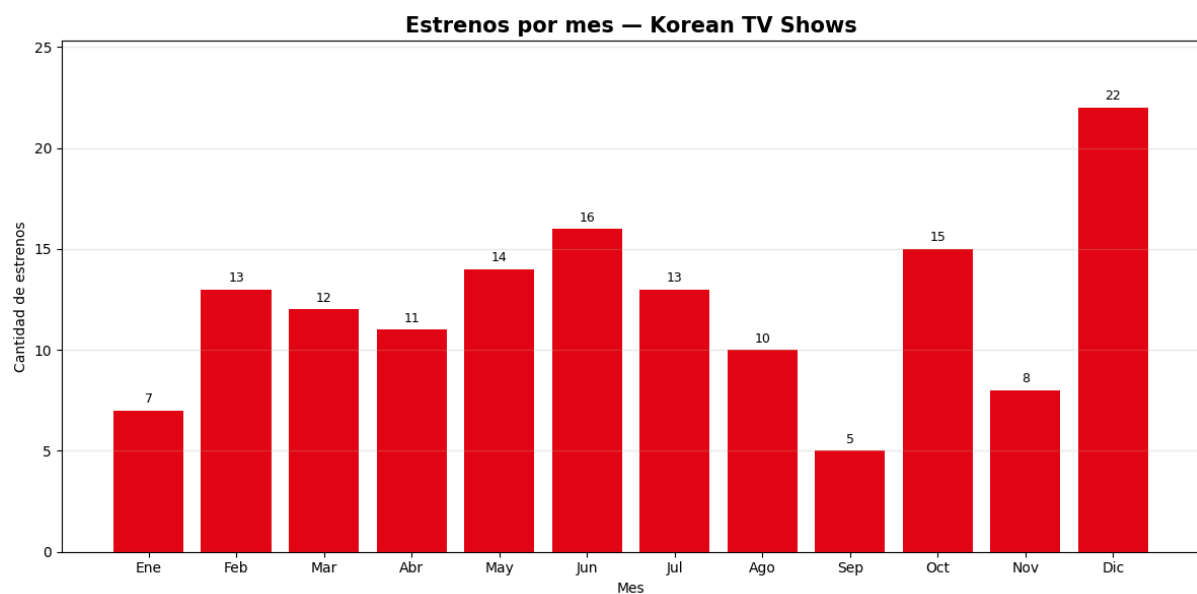
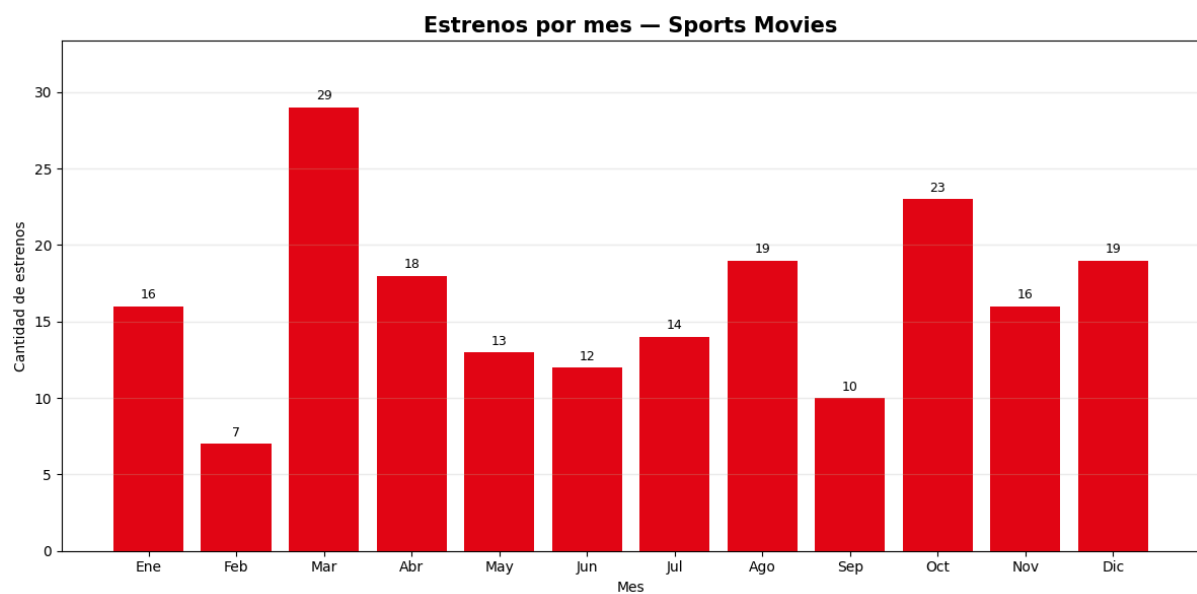
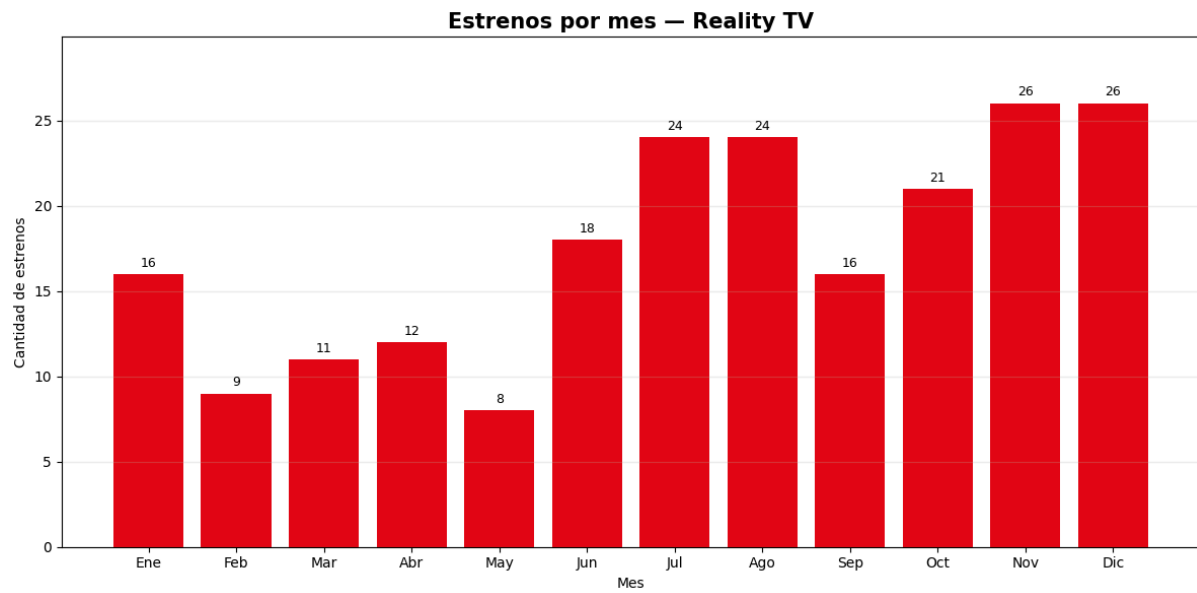


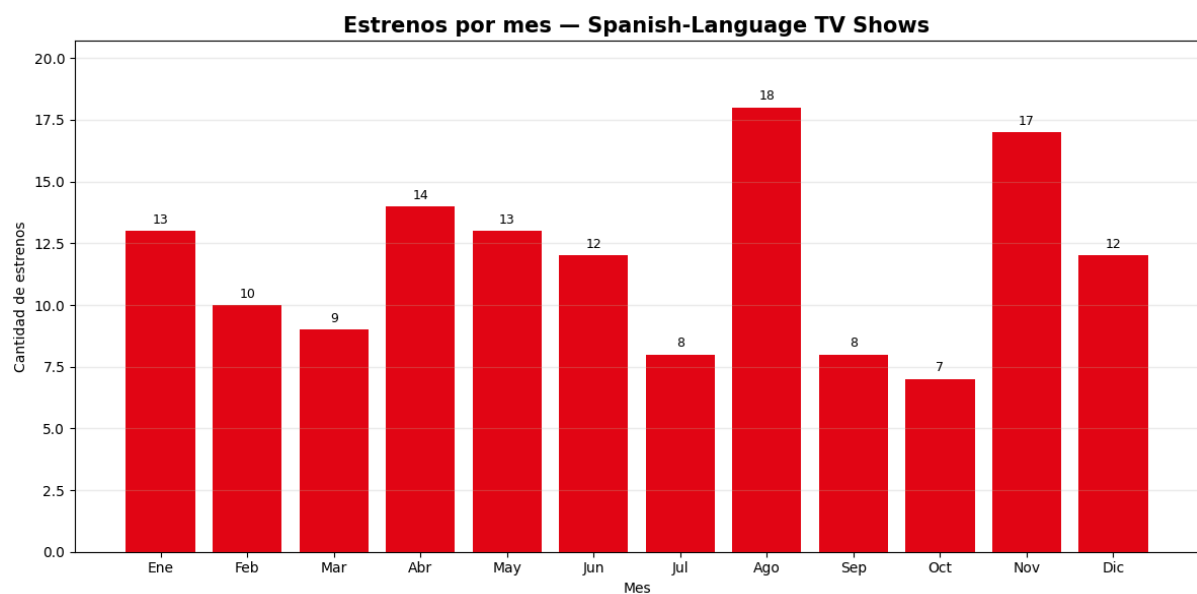
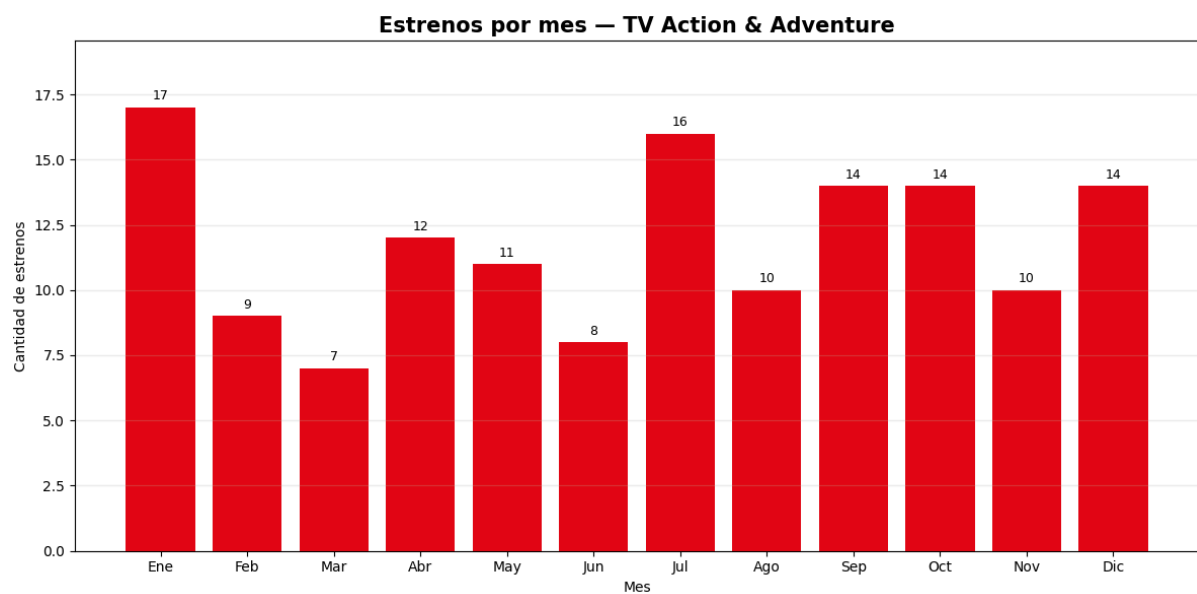
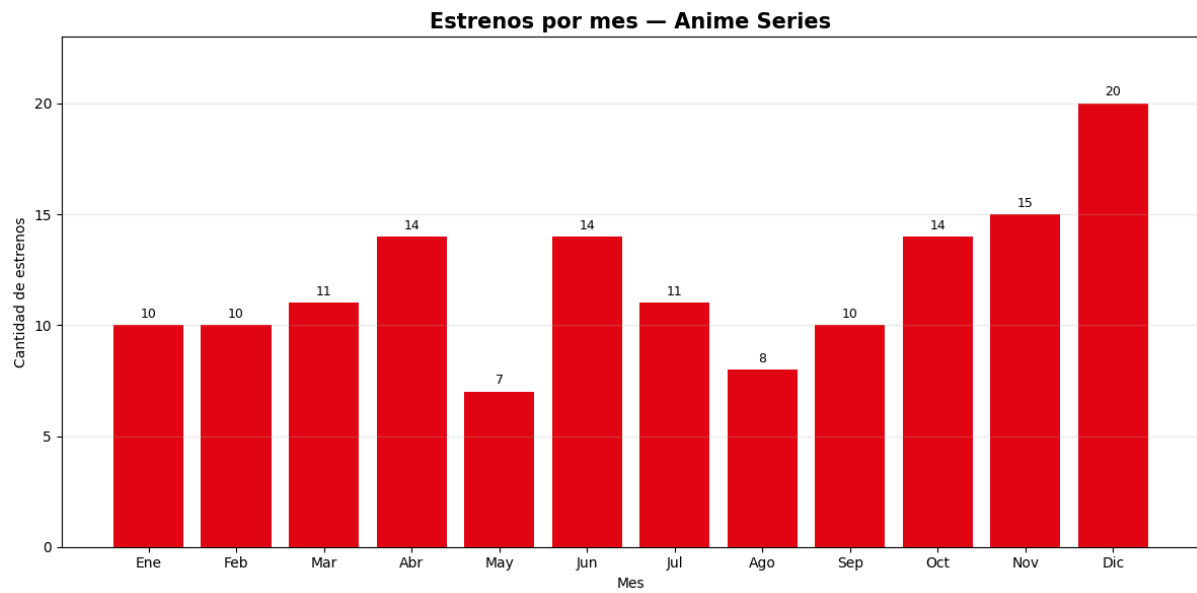


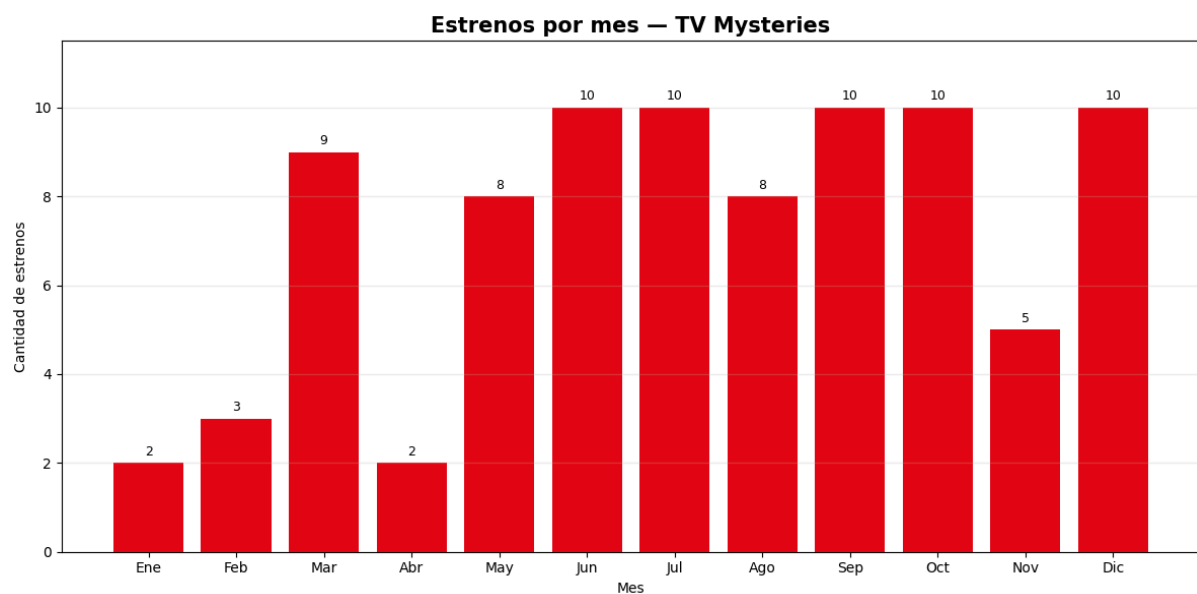
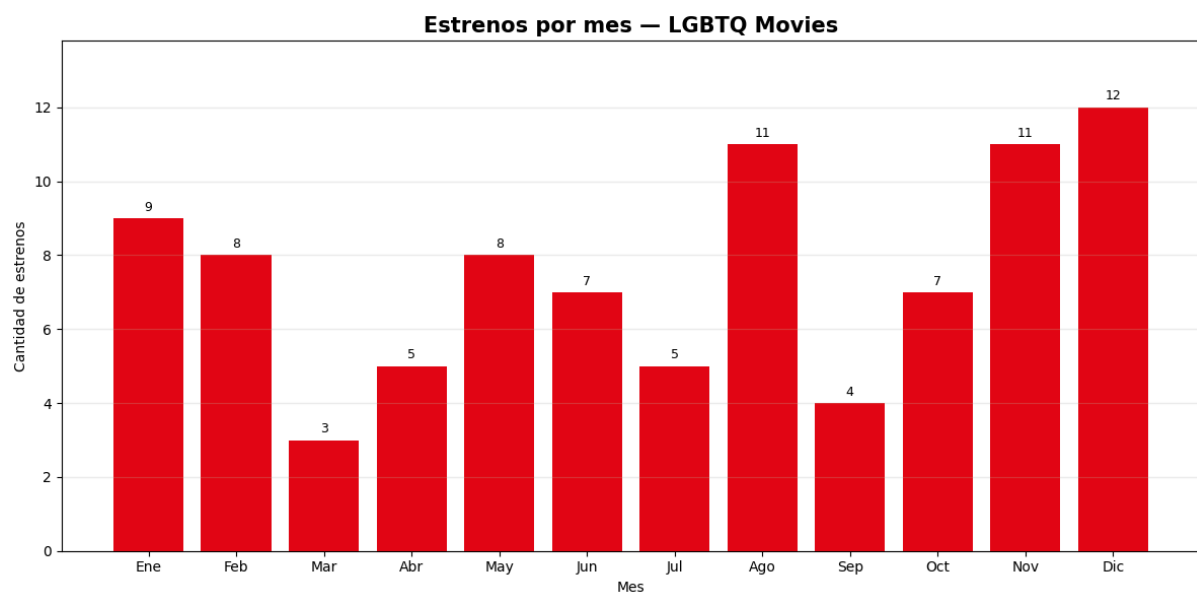
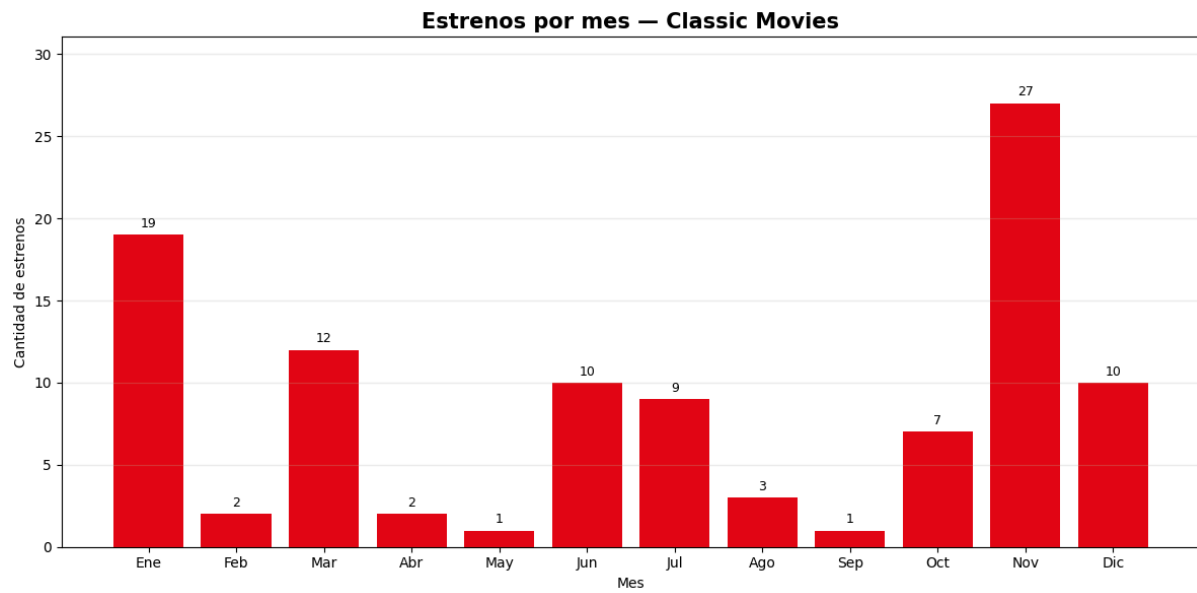


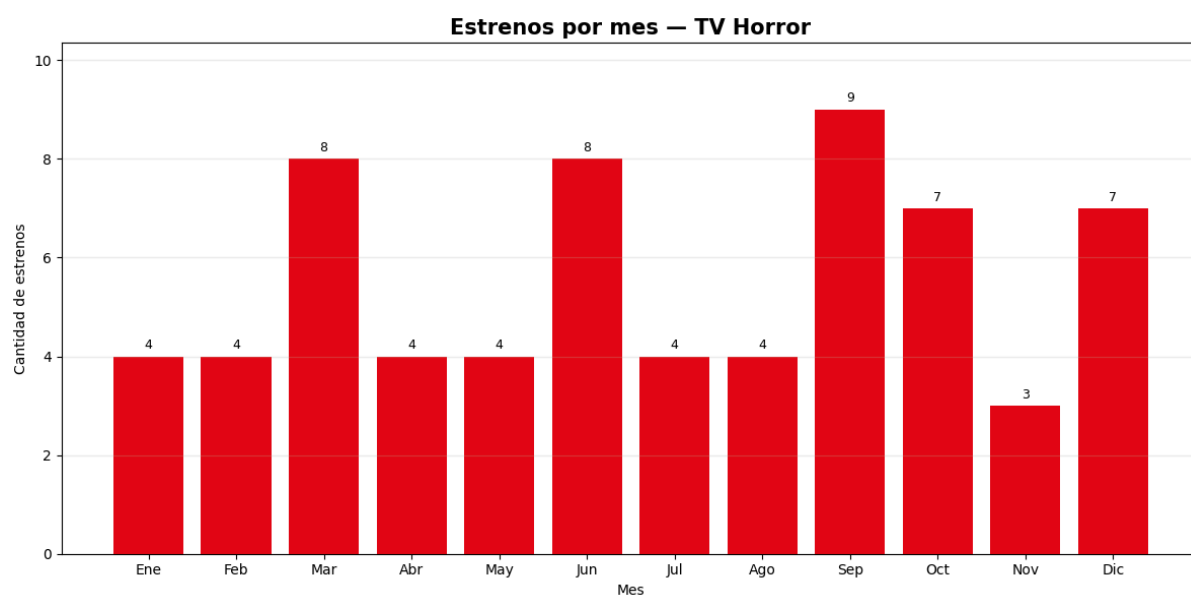
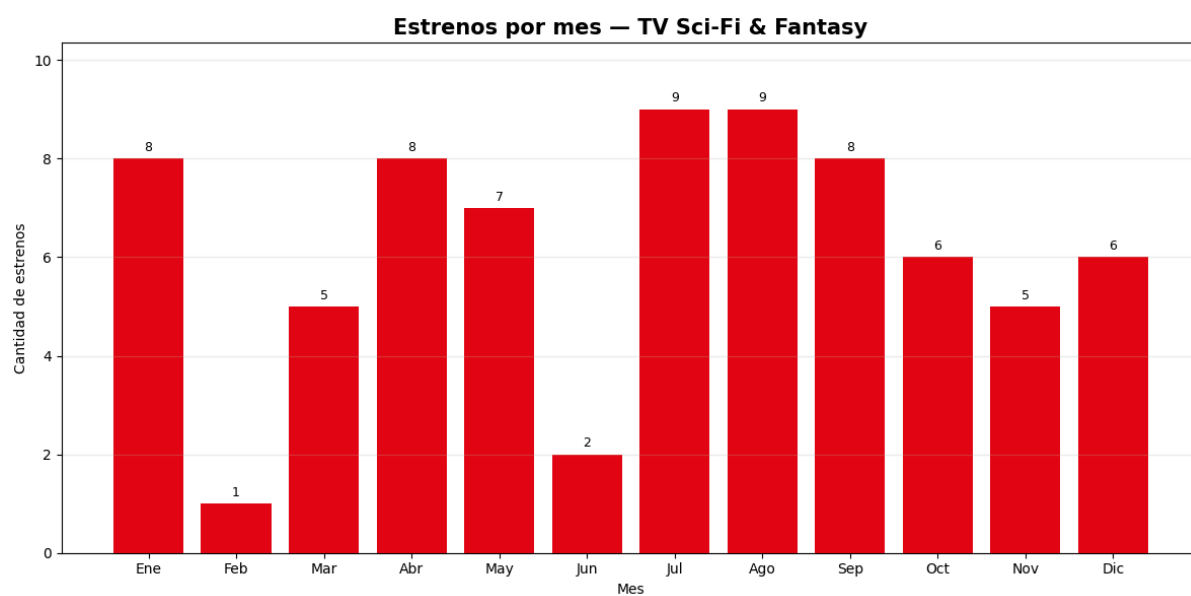
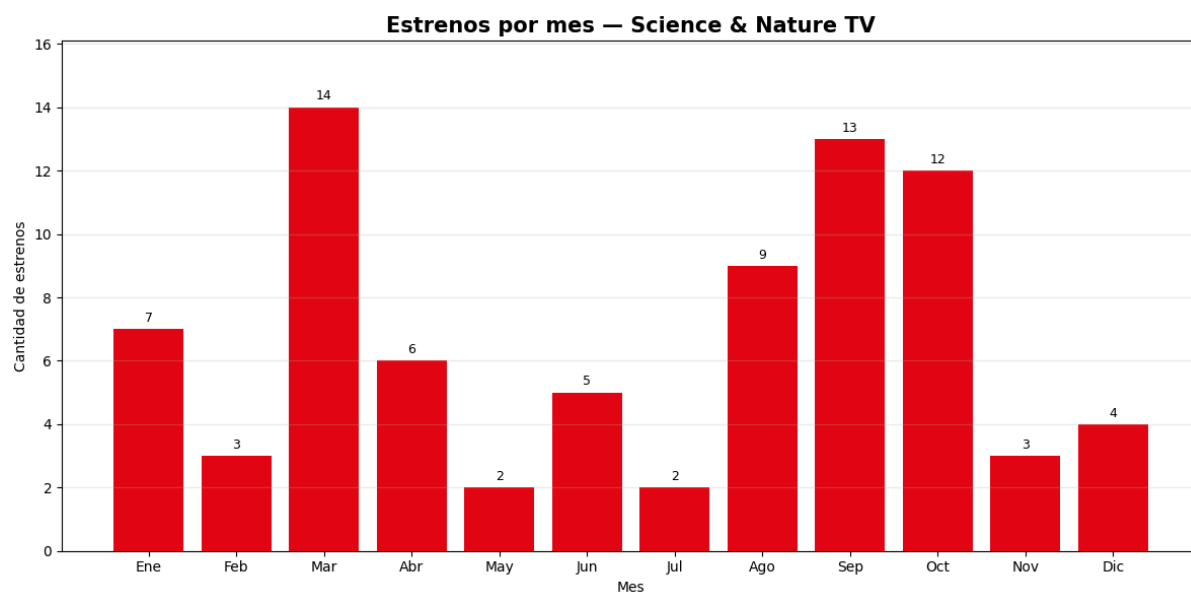


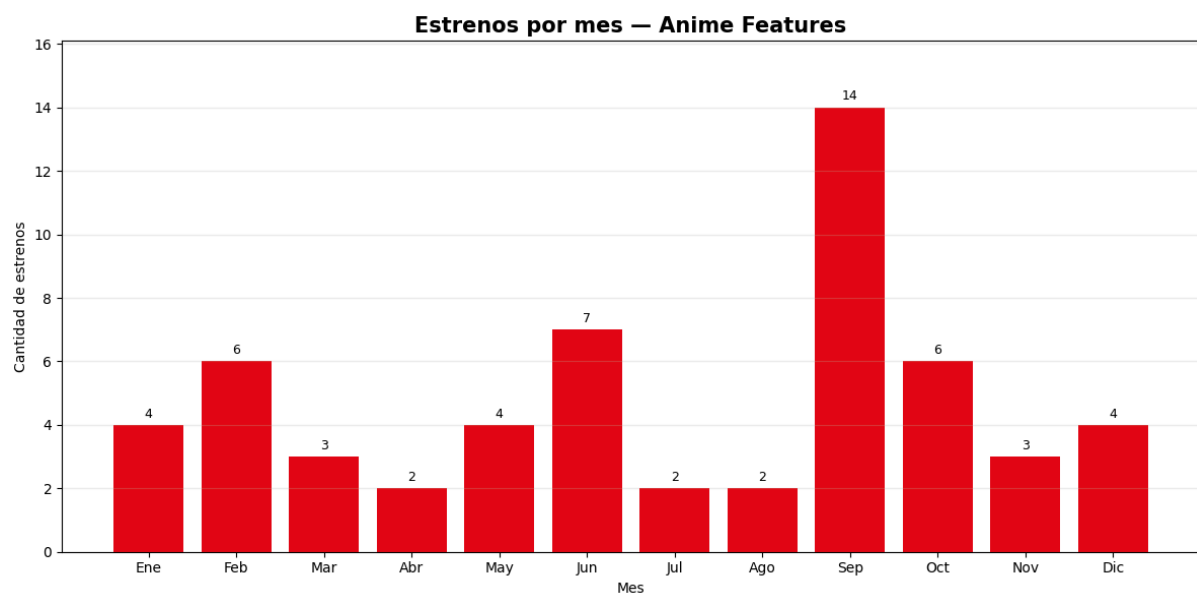
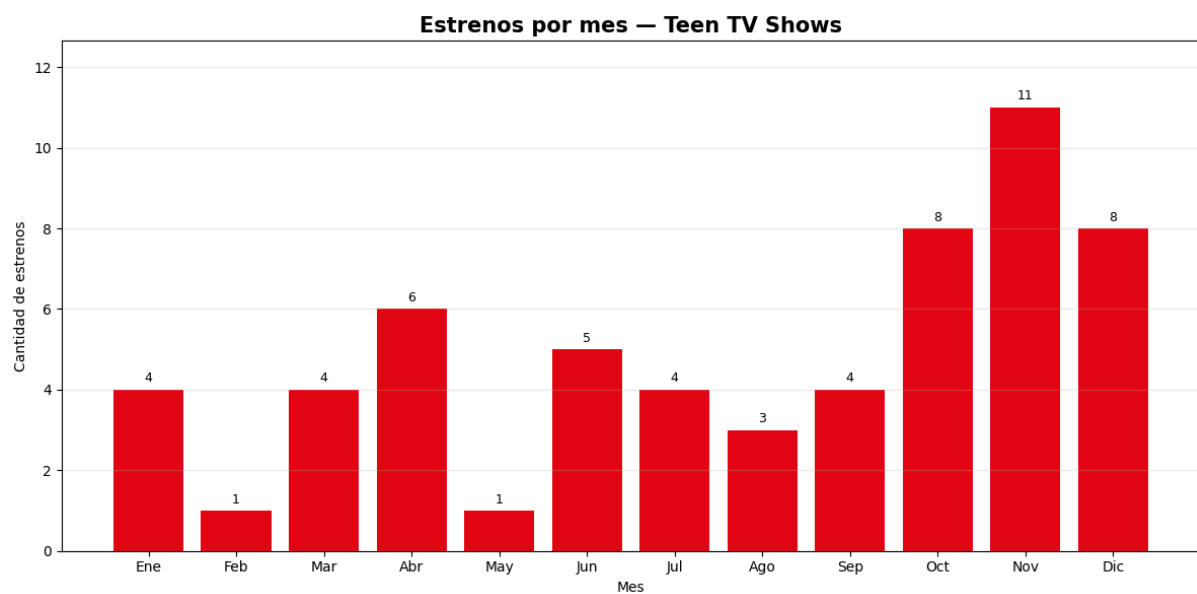
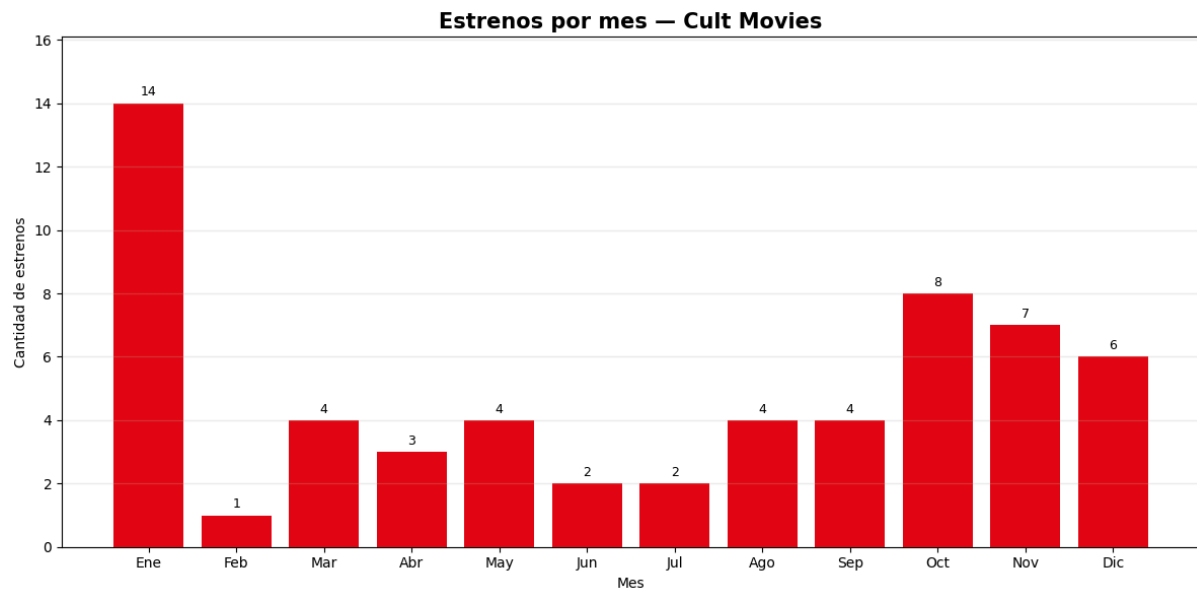


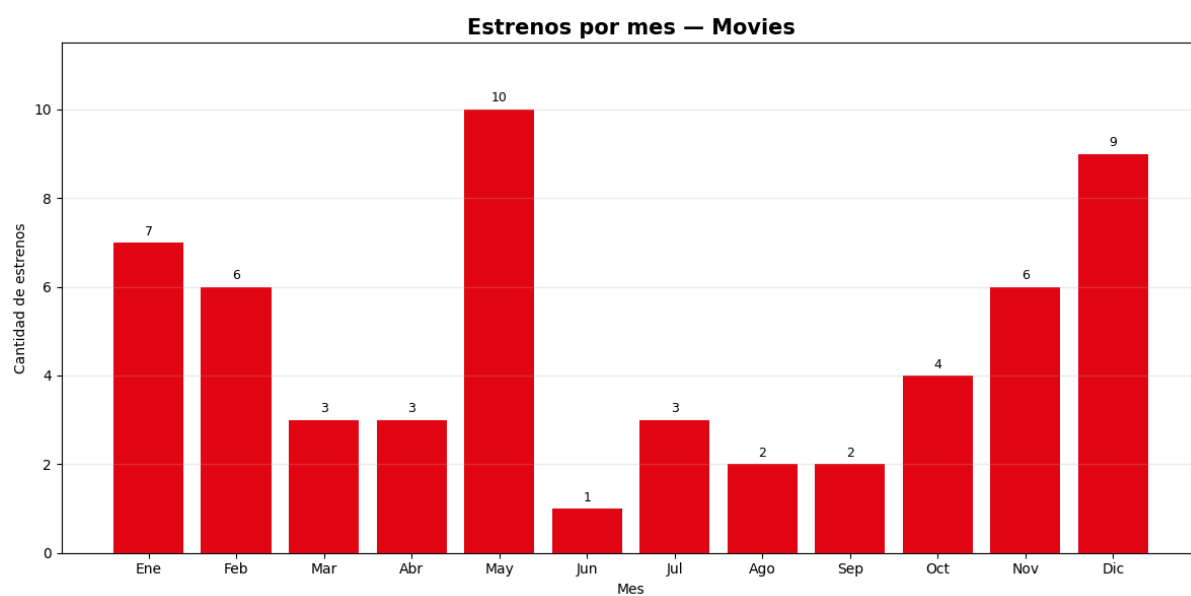
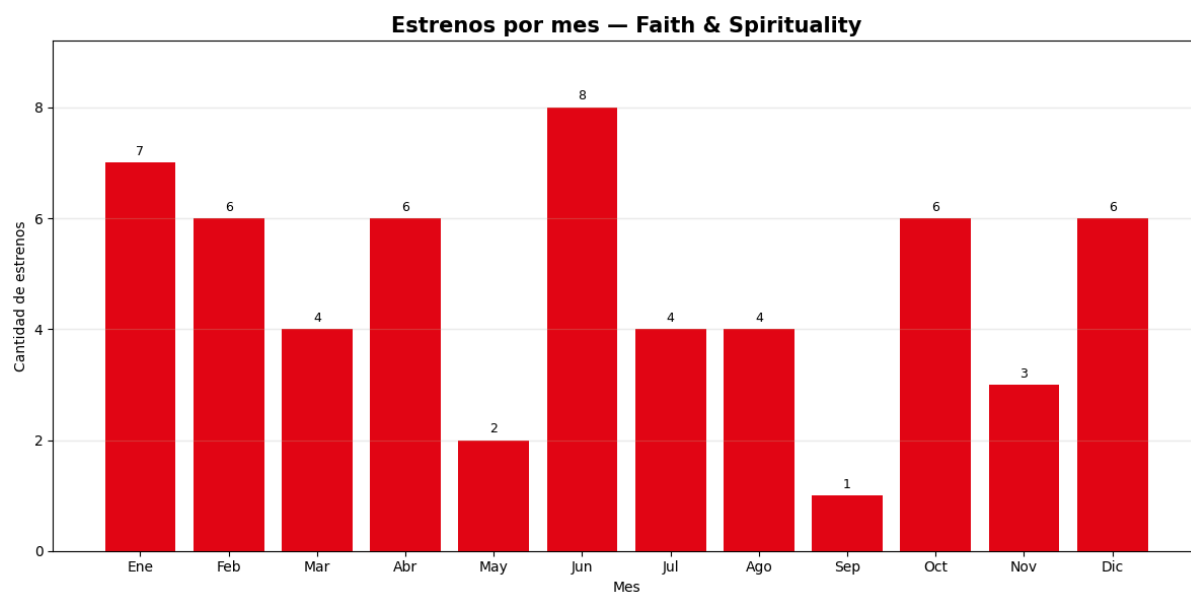


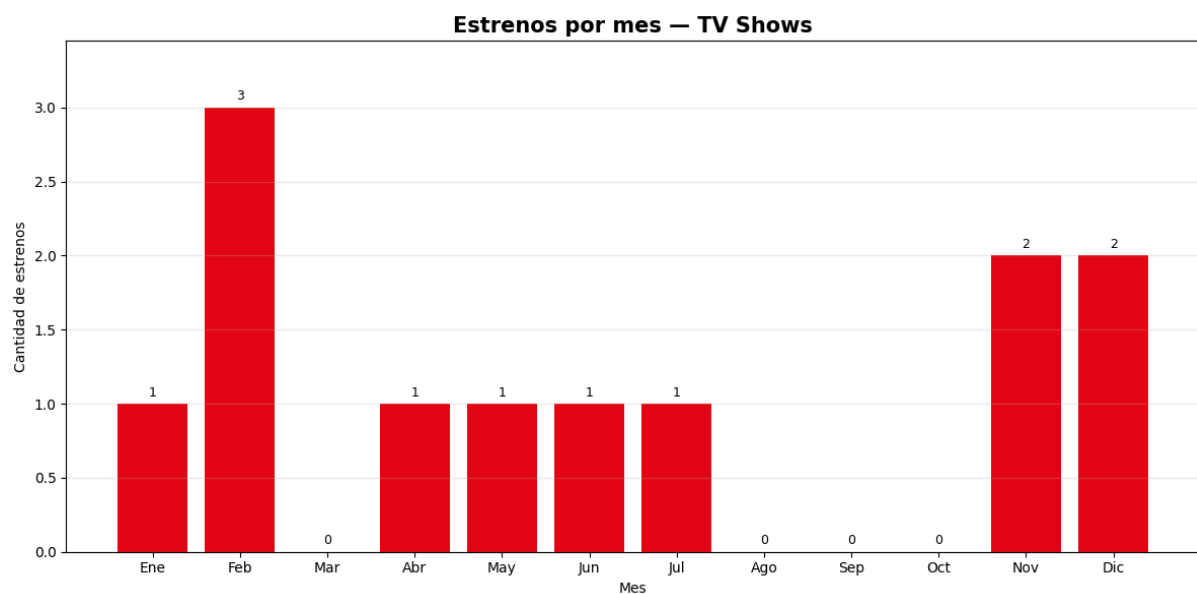
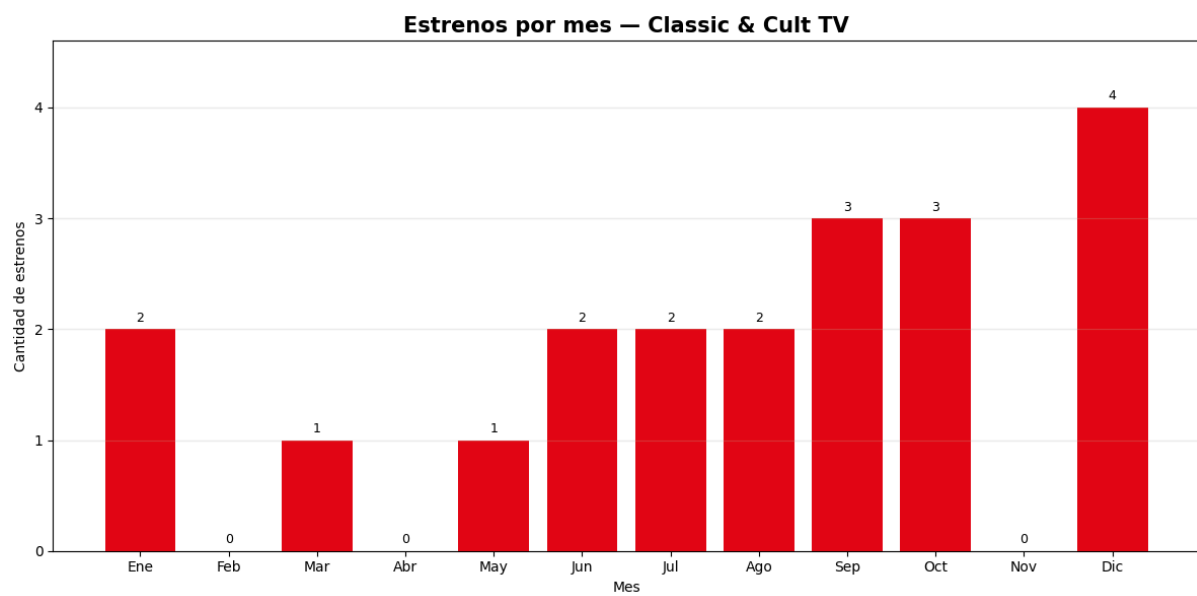
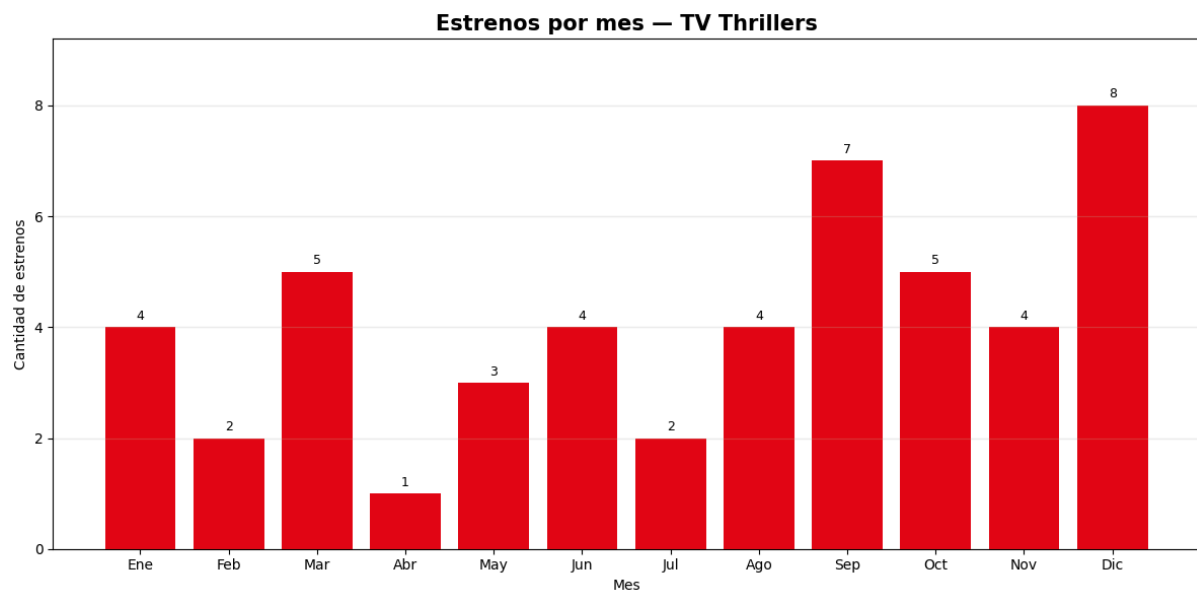












## Prompts de Chat GPT

### 1. Mejora redacción de explicación de variables

“ Para realizar la explicación de las variables utilizamos el Data Profiling.

- show\_id: es único, lo que significa que va a ser el identificador de cada una de las 7787 líneas que tiene el data set. Es del tipo text, pues está compuesto por la letra “s” y un número natural.
- type: define si es una Movie o un TV Show, dato que será relevante para varias de las visualizaciones solicitadas en este practico. Es del tipo categorical.
- title: es único, pues no puede haber dos nombres de películas o programas en general debido a derechos de autor. Es del tipo text.
- director: indica quien dirigió la película o serie y es del tipo text. A su vez, tiene un alto porcentaje (30.7%) de valores faltantes lo cual se puede deber a falta de conocimiento o a despistes a la hora de rellenar el formulario.
- cast: son los actores principales que participan en la producción. Es del tipo text y si bien tiene un porcentaje (9.2%) de valores faltantes, no es tan alto como el de director.
- country: indica el país de origen de la producción. Es del tipo text y cuenta con un porcentaje de 6.5% de valores faltantes.
- date\_added: fecha en que fue agregada a Netflix. Es del tipo date y cuenta con casi 0% de valores faltantes.
- release\_year: fecha en que se creó la película o serie. Es del tipo real number y tiene un 0% de valores faltantes.
- rating: es la clasificación en la que entra cada película o serie. Es del tipo categorical y cuenta con 14 clasificaciones diferentes.
- duration: es la duración de la serie en temporadas o de las películas en minutos. Debido a esta diferencia es que se trata de una variables del tipo text.
- listed\_in: indica el género de la película o serie pudiendo ser uno o varios. Es del tipo texto y tiene 0% de valores faltantes.
- description: es una breve descripción de lo que se trata la serie o la película. Es del tipo text y tiene un 0% de valores faltantes.

Podrías mejorar esta redacción para un informe académico ”

### 2. Mejora de redacción de las 10 preguntas

#### “ Pregunta 1

Para resolver esta pregunta primero cargamos el dataset de Netflix y transformamos la columna date\_added a formato fecha, lo que nos permitió extraer el año en que cada título fue agregado a la plataforma. Después agrupamos los datos por año y por tipo de contenido (película o serie), calculamos cuántos estrenos hubo de cada uno y finalmente obtuvimos la proporción que representan dentro del total anual. Con esos valores construimos un gráfico de líneas que compara la evolución de películas y series a lo largo del tiempo.

En el gráfico se ve que, en los primeros años, Netflix incorporaba casi únicamente películas, pero a partir de 2013 las series empezaron a ganar protagonismo. Aunque las películas siguen siendo mayoría, las series muestran un crecimiento sostenido que hace que la oferta se vuelva más equilibrada en los últimos años.

Quiero que me ayudes a mejorar esta redacción para utilizarlo en un informe y no quede tan informal. ”

#### “ Pregunta 2

Para responder esta pregunta trabajamos con la columna date\_added, que convertimos a formato fecha para poder identificar el año en que cada título fue incorporado a Netflix. Después agrupamos los datos por año y por tipo de contenido, contando cuántas películas y series se agregaron en cada periodo. Con esos resultados armamos un gráfico de líneas donde cada curva muestra la evolución anual de estrenos de películas o series.

El gráfico deja ver que en los primeros años la cantidad de estrenos era baja y las diferencias entre películas y series casi no se notaban. A partir de 2015 la situación cambia: los dos formatos empiezan a crecer con fuerza, aunque las películas siguen teniendo una presencia mucho mayor. El punto más alto se da en 2019, con más de 1400 películas y unas 650 series



agregadas. En los años siguientes se observa una caída en el número total de estrenos, pero las películas siguen siendo el formato más predominante. En general, la tendencia refleja el crecimiento explosivo de Netflix y cómo, después de unos años de expansión, la oferta comenzó a estabilizarse.

Ahora, para estos otros párrafos quiero que apliques un estilo similar al anterior, manteniendo la formalidad. ”

### “ Pregunta 3

Para responder esta pregunta primero limpiamos la columna country, ya que muchos títulos aparecen asociados a más de un país. Lo que hicimos fue separar esos casos en varias filas para que cada país sumara de manera independiente. Después agrupamos los datos por país y tipo de contenido (Movie o TV Show) y contamos cuántas producciones tenía cada combinación. Con eso armamos el top 10 de países con más producciones y lo pasamos a un gráfico de barras, donde cada país aparece con dos barras: una roja para películas y otra negra para series. Además, ordenamos los países de mayor a menor según el total de producciones y le agregamos etiquetas arriba de cada barra para que se vea más claro.

En el gráfico se nota enseguida que Estados Unidos está muy por encima del resto, con más de 2400 películas y más de 800 series. Bastante más abajo aparece India, donde predominan claramente las películas, y después el Reino Unido, que tiene una mezcla importante de los dos formatos. Otros países que completan el ranking son Canadá, Francia, Japón, México, Corea del Sur, España y Alemania, aunque todos con cifras mucho más bajas.

En general, en casi todos los países las películas son mayoría, pero hay excepciones como Japón, Corea del Sur o el Reino Unido, donde las series tienen un peso mucho más fuerte. Esto muestra cómo algunos mercados están más volcados al cine tradicional, mientras que otros vienen apostando cada vez más por las series dentro de Netflix.

Ahora ayúdame a emprolijar estos párrafos. ”

### “ Pregunta 4

Para analizar esta parte usamos las columnas type y rating. Primero limpiamos los datos, eliminando valores nulos y sacando los espacios de más. Después agrupamos por tipo de contenido (películas o series) y por clasificación, contando cuántos títulos había en cada combinación. Para que se entienda mejor, ordenamos los ratings según la cantidad total de títulos, de mayor a menor, y armamos un gráfico de barras agrupadas: en el eje X están las clasificaciones y en cada una se comparan películas (en rojo) y series (en negro). También ajustamos el tamaño del gráfico al número de categorías y le pusimos etiquetas arriba de cada barra con los valores exactos.

El gráfico deja claro qué clasificaciones son las más comunes en Netflix. Las más frecuentes son TV-MA y TV-14, que concentran gran parte del catálogo, tanto en películas como en series. Después siguen TV-PG y R. En la mayoría de las clasificaciones las películas superan a las series, aunque en ratings como TV-Y o TV-Y7 se nota que las series tienen un peso mayor. También aparecen algunas categorías poco comunes como UR y NC-17, con apenas unos pocos títulos. En general, los resultados muestran que Netflix se enfoca sobre todo en contenido pensado para adolescentes y adultos.

A estos párrafos también aplica el mismo estilo. ”

### “ Pregunta 5

En este caso trabajamos con las columnas country y rating. Primero armamos categorías de audiencia a partir del rating: Adulto (TV-MA, R, NC-17) e Infantil (TV-Y, TV-Y7, TV-G, PG, TV-PG, G). También definimos Adolescente y Otros, pero esas no las usamos en el gráfico. Después limpiamos la columna country para separar los títulos que tienen más de un país y que cada país sume por separado. Luego contamos cuántos títulos de cada audiencia tenía cada país, nos quedamos con el top 10 y armamos una tabla pivote con las columnas Adulto e Infantil. Por último, ordenamos los países de mayor a menor y generamos un gráfico de barras agrupadas, con rojo para adulto y negro para infantil, agregando etiquetas arriba de cada barra.

El gráfico muestra bien claro que Estados Unidos está lejísimos del resto, tanto en contenido adulto como infantil. Le siguen Reino Unido e India, y más atrás aparecen Canadá, Francia, Japón, España, Alemania, Corea del Sur y México. En la mayoría de los países se nota que predomina el contenido adulto, aunque en algunos como Canadá o Japón la diferencia con el infantil es bastante menor. En resumen, el gráfico sirve para ver no solo el ranking de producción por país, sino también hacia qué tipo de audiencia se inclina cada uno.

Como mejora, también se podría agregar una categoría Adolescente (con ratings como PG-13 y TV-14) para diferenciar un poco más los públicos, y otra categoría de Otros para clasificaciones raras o sin definir (NR, UR, etc.).

Mejorame esta redacción ”

#### “ Pregunta 6

Para este análisis primero cargamos el dataset, convertimos la columna date\_added a fecha y sacamos de ahí el mes en que cada título fue agregado. Como la columna listed\_in puede tener más de una categoría por título, la separamos y reordenamos los datos para poder contar cuántos estrenos hubo por mes en cada categoría. Con eso armamos gráficos de barras, uno por categoría, donde en el eje X están los meses, en el Y la cantidad de estrenos y además pusimos el número exacto arriba de cada barra.

Estos gráficos nos ayudan a ver si hay meses donde se concentran más lanzamientos en alguna categoría. Por ejemplo, en Comedias o Dramas los estrenos se ven bastante repartidos a lo largo del año, pero en Documentales o en Navidad aparecen picos más marcados, como en diciembre. Para no llenar el informe con las 42 categorías que hay, mostramos solo las cinco más importantes y dejamos el detalle completo en el anexo.

Por otro lado, también quisimos ver en qué meses en general se concentran más lanzamientos, sin importar la categoría. Para eso agrupamos los estrenos por mes y armamos otro gráfico de barras. Ahí se nota que diciembre es el mes con más títulos (817), seguido de octubre (780) y enero (746). En cambio, febrero es el mes con menos estrenos. El resto de los meses se mantienen más equilibrados, aunque queda claro que Netflix apuesta fuerte a las fechas de vacaciones y fiestas.

Podrías ayudarme a mejorar esta redacción ”

#### “ Pregunta 7

Para esta parte nos enfocamos en los directores. Primero sacamos los 10 con más títulos en el catálogo y armamos un gráfico de barras para ver cuántos de esos eran películas y cuántos series. Después clasificamos sus producciones según la audiencia (infantil, adolescente, adulto u otros) y generamos un gráfico de barras apiladas para ver cómo se distribuyen.

En los resultados se ve clarísimo que la mayoría de estos directores tienen casi todo concentrado en películas, con muy poca o nula participación en series. Además, cuando miramos las audiencias, la mayor parte del contenido es para adultos, seguido por adolescentes y en menor medida por infantil. Esto muestra que Netflix le da bastante peso a directores de cine y sobre todo a producciones pensadas para un público adulto.

Quiero mejorar esta redacción como para un informe académico ”

#### “ Pregunta 8

Acá miramos los actores que más aparecen en el catálogo. Primero filtramos los títulos sin información de elenco y después separamos en filas cuando un título tenía varios actores. Con eso contamos cuántas veces aparecía cada uno y nos quedamos con el top 10. Para visualizarlo hicimos un gráfico de barras donde cada barra es un actor y su altura muestra cuántos títulos tiene, con el número exacto arriba.

El gráfico muestra que los actores más presentes en Netflix son sobre todo del cine y la televisión de la India y Asia. Se destacan Anupam Kher y Shah Rukh Khan, con más de 35 títulos cada uno. Esto deja en evidencia la gran presencia del cine indio dentro del catálogo de la plataforma.

Mejorar mi redacción para que quede más formal y prolija ”

#### “ Pregunta 9

En este caso analizamos la duración según el tipo de contenido. Para las películas usamos la columna duration para extraer la cantidad de minutos y generamos un histograma con los rangos de tiempo. Para las series, interpretamos esa misma columna como la cantidad de temporadas y armamos otro histograma con la distribución.

Los resultados muestran que la mayoría de las películas duran entre 90 y 120 minutos, lo que coincide con el estándar del cine tradicional. A medida que aumenta la duración, el número de títulos va bajando y hay muy pocas películas que pasen las 2 horas y media. En cuanto a las series, la mayoría tiene 1 o 2 temporadas, lo que indica que Netflix apuesta mucho a producciones limitadas o de corta extensión.

Quiero que esta redacción quede formal y prolija como para ponerla en un informe académico ”

#### “ Pregunta 10

Para esta parte generamos una nube de palabras (WordCloud) con los términos más frecuentes en los títulos. Lo que hicimos fue procesar la columna title, pasar todo a minúsculas, eliminar los valores nulos y separar las palabras. Después filtramos las stopwords (palabras muy comunes como the, de, la, and), y con el resto armamos la nube.

En el WordCloud se ve que palabras como love, girl, world, story y life son las más repetidas, lo que refleja las temáticas más comunes en las producciones: romances, personajes femeninos, relatos de vida y tramas con alcance global. Además, usamos una paleta de colores que mantiene coherencia visual con el resto de los gráficos (rojo, negro y blanco, como la identidad de Netflix).

Podrías mejorar mi redacción con el estilo de las anteriores “

### 3. Redacción de la conclusión

“

- Al principio Netflix casi solo subía películas, pero a partir de 2013 las series empezaron a ganar terreno. Hoy las películas siguen siendo mayoría, pero la oferta está mucho más equilibrada.
- El año con más estrenos fue 2019, y después bajó un poco el ritmo. Eso muestra una etapa de crecimiento rápido y luego una estabilización.
- Estados Unidos lidera el catálogo por lejos, seguido por India y Reino Unido. En la mayoría de los países predominan las películas, salvo casos como Japón o Corea del Sur, donde las series tienen más peso.
- En cuanto a ratings, las categorías que más aparecen son TV-MA y TV-14, o sea, contenidos pensados para adolescentes y adultos. El material infantil es bastante menor.
- La duración también tiene patrones: la mayoría de las películas duran entre 90 y 120 minutos, mientras que la mayoría de las series tienen solo 1 o 2 temporadas (muchas son cortas o limitadas).
- Entre directores y actores, aparecen sobre todo nombres de la industria india. Por ejemplo, Anupam Kher y Shah Rukh Khan son los actores más recurrentes, y los directores se concentran mayormente en películas.
- En el análisis de estacionalidad, los meses con más estrenos son diciembre, octubre y enero, lo que parece estar relacionado con las fiestas y las vacaciones.
- En la nube de palabras, destacan términos como love, girl, world, story y life, lo que refleja que las temáticas más comunes giran en torno al romance, las historias de vida y los relatos de alcance global.

Podrías ayudarme a redactar este punteo considerando que es la conclusión para un informe académico ”