Francis Aguilar #22243

Angela Garcia #22869

# Hoja de Trabajo 01 – Análisis Exploratorio
# Proyecto 01

## INTRODUCCIÓN:

El estudio recién inaugurado, **"CineVision Studios"**, está especializado en la producción y distribución de películas a nivel global. Este, busca mejorar sus procesos de toma de decisiones basados en datos para maximizar la rentabilidad de sus producciones y satisfacer las expectativas del público.

En un mercado altamente competitivo, **CineVision Studios** enfrenta múltiples desafíos, como identificar las tendencias que generan mayor popularidad, optimizar presupuestos para producciones exitosas y atraer al mejor talento tanto en el elenco como en la dirección. El análisis de datos les permite comprender mejor el rendimiento histórico de las películas y anticipar el impacto de futuras decisiones.

Desde la perspectiva de **CineVision Studios**, este análisis tiene los siguientes propósitos:

1. **Optimizar la selección de directores y elenco**:
    - Evaluar el impacto de la popularidad de los actores y actrices en el éxito de las películas.
    - Determinar qué directores generan mayores ingresos y mejores calificaciones.
2. **Expandir mercados y audiencias**:
    - Detectar patrones en los países de producción que contribuyen al éxito financiero.
    - Analizar las tendencias de idioma y su impacto en la popularidad global.
    - Analizar los géneros populares entre el público.
3. **Tomar decisiones estratégicas en marketing**:
    - Evaluar la relación entre la existencia de videos promocionales y la popularidad de las películas.
4. **Fomentar la diversidad en las producciones**:
    - Analizar la representación de género en los elencos y cómo influye en la recepción del público.
    - Identificar las oportunidades para producir películas con enfoques inclusivos.

**Preguntas Clave desde la Perspectiva del Estudio**

1. ¿Cuáles son las 10 películas que contaron con más presupuesto?
2. ¿Cuáles son las 10 películas que más ingresos tuvieron?
3. ¿Cuál es la película que más votos tuvo?
4. ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?
5. ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras
6. ¿Cuál es el **género principal** de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Represéntelo usando un gráfico.
    1. ¿A qué género principal pertenecen las películas más largas?
7. ¿Las películas de qué genero principal obtuvieron mayores ganancias?

8. ¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?
9. ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?
10. ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?
11. ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión
12. ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?
13. ¿En qué meses se han visto los lanzamientos con mejores ingresos?¿cuantas películas, en promedio, se han lanzado por mes?
14. ¿Cómo se correlacionan las calificaciones con el éxito comercial?
15. ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?
16. ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?

**Aplicaciones del Análisis**

1. **Planificación de Producciones**: Con los insights obtenidos, **CineVision Studios** podrá enfocar recursos en géneros, directores y elencos que históricamente han mostrado ser rentables y populares.
2. **Optimización del Presupuesto**: Identificar el rango de presupuestos más eficiente para maximizar el rendimiento financiero.
3. **Estrategias de Distribución**: Adaptar películas para mercados globales considerando idiomas y países de producción con mejor desempeño.
4. **Diversificación de Portafolio**: Fomentar proyectos con mayor representación de género e inclusividad para satisfacer a audiencias modernas.

**Resultados Esperados**

Al aplicar este análisis, **CineVision Studios** puede fortalecer su posición competitiva en el mercado cinematográfico, reducir riesgos financieros y garantizar que sus producciones estén alineadas con las tendencias y preferencias del público global.

**Descripción de la consultoría:**

Han contratado a su grupo para que lleve a cabo los análisis que le sirvan al estudio para cumplir con sus objetivos estratégicos y le ha proporcionado un conjunto de datos. Sin embargo, algunos datos del conjunto no están disponibles para todos los años, como es el caso de la recaudación.

**Presentación de resultados**

La compañía espera **un informe** con todos los hallazgos que arrojaron las respuestas a las preguntas que planteó, estos deben estar bien explicados y se deben apoyar de gráficos que sostengan estas explicaciones.

También le han pedido que entregue **el código** utilizado pues están pensando en contratar un analista de datos de planta que continúe con el análisis a lo largo del tiempo.

## DESCRIPCIÓN DEL DATASET

El dataset contiene datos de 10000 películas obtenidos de la plataforma "The Movie DB".

**Variables:**

- **Id**: Id de la película
- **popularity**: Índice de popularidad de la película calculado semanalmente
- **budget**: El presupuesto para la película.
- **revenue**: El ingreso de la película.
- **original_title**: El título original de la película, en su idioma original.
- **originalLanguage**: Idioma original en que se encuentra la película
- **title**: El título de la película traducido al inglés
- **homePage**: La página de inicio de la película
- **video**: Si tiene videos promocionales o no
- **director**: Director de la película
- **runtime**: La duración de la película.
- **genres**: El género de la película.
- **genresAmount**: Cantidad de géneros que representan la película
- **productionCompany**: Las compañías productoras de la película.
- **productionCoAmount**: Cantidad de compañías productoras que participaron en la película
- **productionCompanyCountry**: Países de las compañías productoras de la película
- **productionCountry**: Países en los que se llevó a cabo la producción de la película
- **productionCountriesAmount**: Cantidad de países en los que se rodó la película
- **releaseDate**: Fecha de lanzamiento de la película
- **voteCount**: El número de votos en la plataforma para la película.
- **voteAvg**: El promedio de los votos en la plataforma para la película
- **actors**: Actores que participan en la película (Elenco)
- **actorsPopularity**: Índice de popularidad del elenco de la película.
- **actorsCharacter**: Personaje que interpreta cada actor en la película
- **actorsAmount**: Cantidad de personas que actúan en la película
- **castWomenAmount**: Cantidad de actrices en el elenco de la película
- **castMenAmount**: Cantidad de actores en el elenco de la película.

## EJERCICIOS

1. **(3 puntos)** Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.

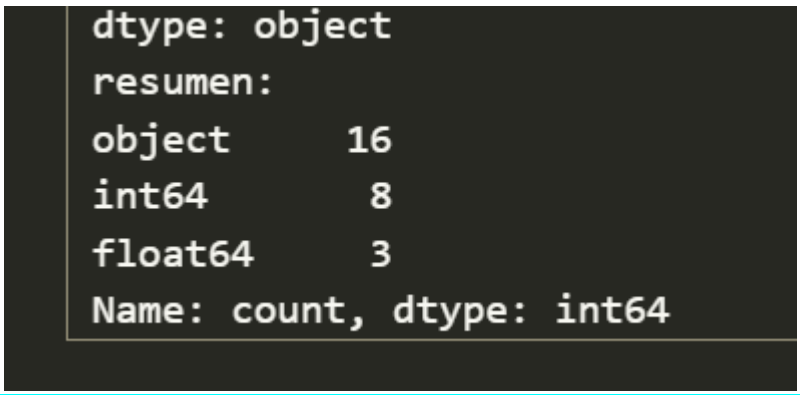
```
#exploracion de los datos
#mostrar las columanas
print("\n--Columnas:---")
print(list(df.columns))

#resumen del set de datos
print("\n--Resumen del set de datos:--")
print(df.describe())
```

[86] Python

```
--Columnas:---
['id', 'budget', 'genres', 'homePage', 'productionCompany', 'productionCompanyCountry', 'productionCountry', 'revenue', 'runtime', 'video', 'director', 'actors', 'actorsPopularity',

--Resumen del set de datos:--
                id        budget       revenue       runtime    popularity  \
count   10000.000000  1.000000e+04  1.000000e+04  10000.000000  10000.000000
mean   249876.829300  1.855163e+07  5.673793e+07    100.268100     51.393907
std    257380.109004  3.662669e+07  1.495854e+08     27.777829    216.729552
min         5.000000  0.000000e+00  0.000000e+00      0.000000      4.258000
25%     12286.500000  0.000000e+00  0.000000e+00     90.000000     14.577750
50%    152558.000000  5.000000e+05  1.631245e+05    100.000000     21.905500
75%    452021.750000  2.000000e+07  4.479661e+07    113.000000     40.654000
max    922260.000000  3.800000e+08  2.847246e+09    750.000000  11474.647000

           voteAvg     voteCount  genresAmount  productionCoAmount  \
count  10000.000000  10000.000000  10000.000000        10000.000000
mean       6.483490   1342.381800      2.596500            3.171400
std        0.984274   2564.196637      1.154565            2.539738
min        1.300000      1.000000      0.000000            0.000000
25%        5.900000    120.000000      2.000000            2.000000
50%        6.500000    415.000000      3.000000            3.000000
75%        7.200000   1316.000000      3.000000            4.000000
max       10.000000  30788.000000     16.000000           89.000000

      productionCountriesAmount   actorsAmount
count               10000.000000  10000.000000
mean                    1.751000   2147.666600
std                     3.012093  37200.075802
```

```python
#obtener el tipo de datos
print("\n----Tipo de datos:---")
print(df.dtypes)
tipos_de_datos = df.dtypes.value_counts()
print('resumen:')
print(tipos_de_datos)
```

```
----Tipo de datos:---
id                            int64
budget                        int64
genres                       object
homePage                     object
productionCompany            object
productionCompanyCountry     object
productionCountry            object
revenue                     float64
runtime                       int64
video                        object
director                     object
actors                       object
actorsPopularity             object
actorsCharacter              object
originalTitle                object
title                        object
originalLanguage             object
popularity                  float64
releaseDate                  object
voteAvg                     float64
voteCount                     int64
genresAmount                  int64
productionCoAmount            int64
productionCountriesAmount     int64
actorsAmount                  int64
castWomenAmount              object
castMenAmount                object
dtype: object
```




2. **(5 puntos)** Diga el tipo de cada una de las variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta)

**Tipo de Variables:**

**Variables Cualitativas:**

**Ordinales:**

- No hay

**Nominales:**

- original_title
- originalLanguage
- title
- homePage
- director
- genres
- productionCompany
- productionCompanyCountry
- productionCountry
- video
- actors
- actorsCharacter

**Variables Cuantitativas:**

**Continuas:**

- popularity

- releaseDate
- revenue
- voteAvg

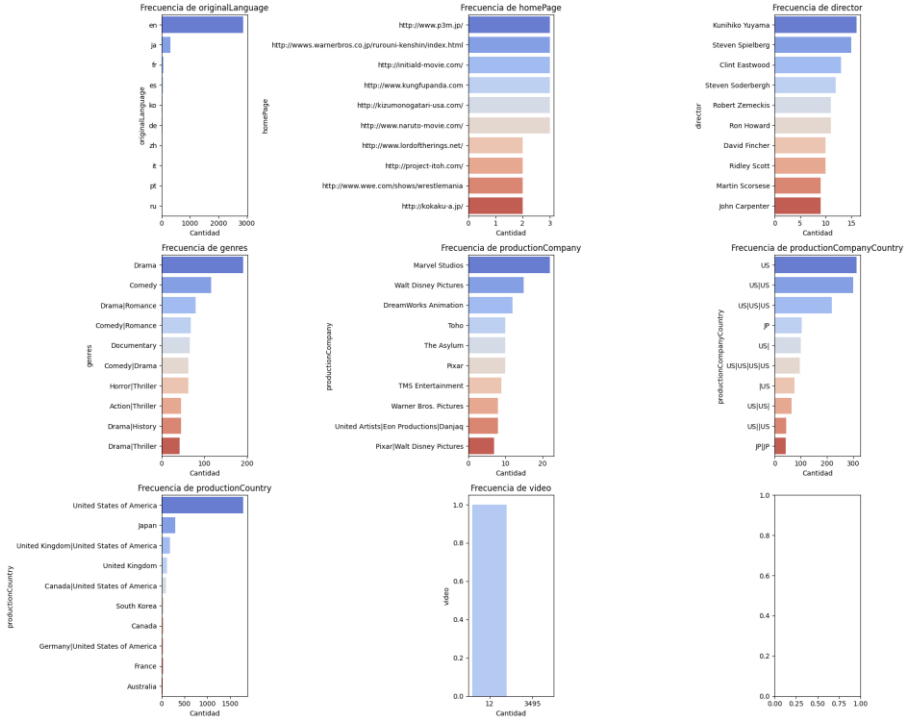
**Discretas:**

- Id
- budget
- runtime
- genresAmount
- productionCoAmount
- productionCountriesAmount
- voteCount
- actorsPopularity
- actorsAmount
- castWomenAmount
- castMenAmount

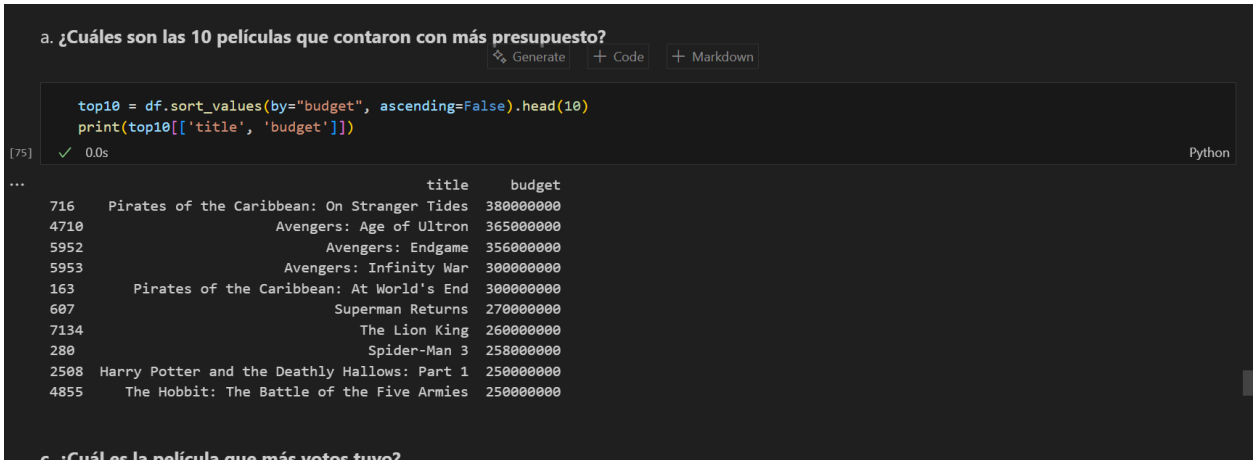
3. **(6 puntos)** Investigue si las variables cuantitativas siguen una distribución normal y haga una tabla de frecuencias de las variables cualitativas. **Explique todos los resultados.**

Investigando y haciendo pruebas estadísticas, logramos observar que ninguna de las variables cuantitativas tiene una distribución normal. Esto se puede dar a que existe un sesgo, ya sea a la derecha o a la izquierda, o que los datos se encuentran muy dispersos.

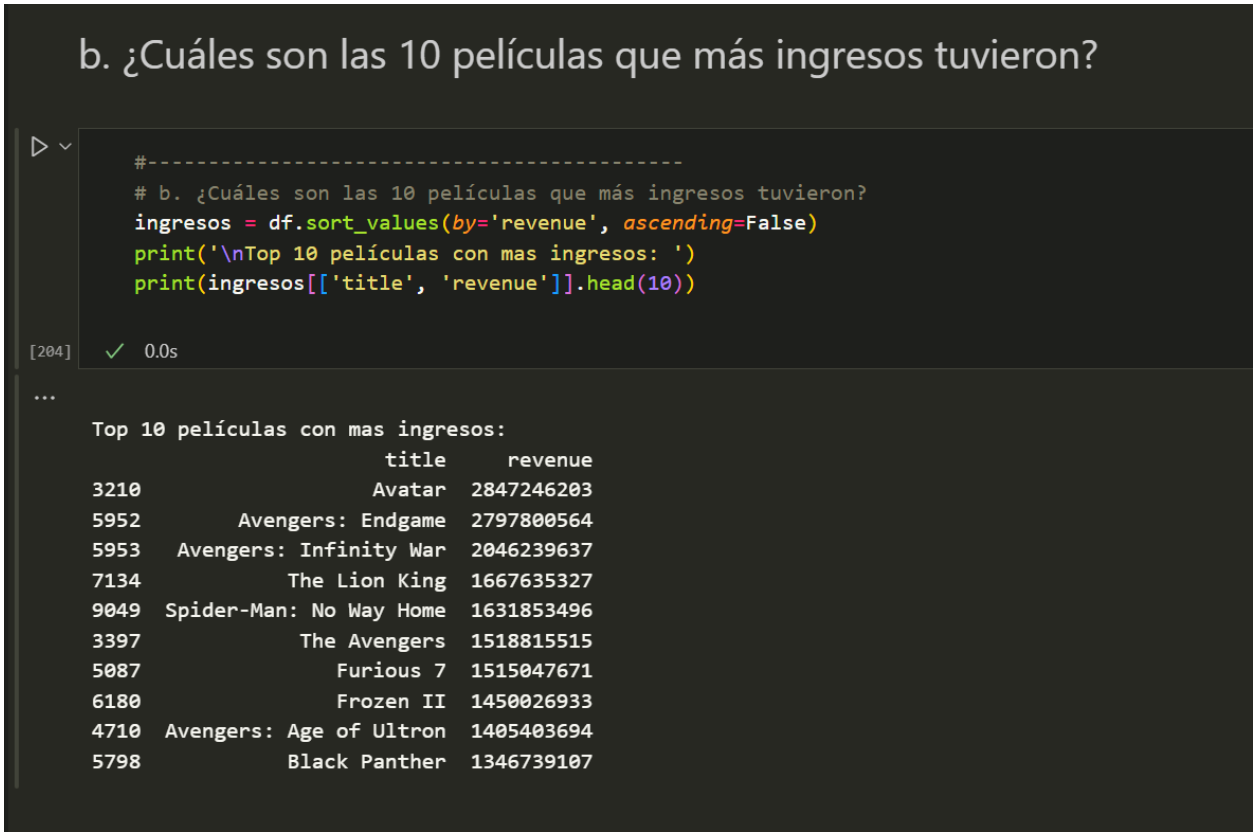
Ahora con los resultados de las variables cualitativas, podemos observar los diagramas de barra que se muestran a continuación. Aquí podemos ver con qué frecuencia aparecen cada una de las categorías. Podemos ver que algunas son de clasificación y existen otras que no se incluyeron en las gráficas debido a que no tendría sentirlo analizarlas, como lo es, el id, los actores (por la forma en la que está estructurada la información).


Frecuencia de originalLanguage
originalLanguage
en
ja
fr
es
ko
de
zh
it
pt
ru
0
1000
2000
3000
Cantidad
Frecuencia de homePage
homePage
http://www.p3m.jp/
http://wwws.warnerbros.co.jp/rurouni-kenshin/index.html
http://initiald-movie.com/
http://www.kungfupanda.com
http://kizumonogatari-usa.com/
http://www.naruto-movie.com/
http://www.lordoftherings.net/
http://project-itoh.com/
http://www.wwe.com/shows/wrestlemania
http://kokaku-a.jp/
0
1
2
3
Cantidad
Frecuencia de director
director
Kunihiko Yuyama
Steven Spielberg
Clint Eastwood
Steven Soderbergh
Robert Zemeckis
Ron Howard
David Fincher
Ridley Scott
Martin Scorsese
John Carpenter
0
5
10
15
Cantidad
Frecuencia de genres
genres
Drama
Comedy
Drama|Romance
Comedy|Romance
Documentary
Comedy|Drama
Horror|Thriller
Action|Thriller
Drama|History
Drama|Thriller
0
100
200
Cantidad
Frecuencia de productionCompany
productionCompany
Marvel Studios
Walt Disney Pictures
DreamWorks Animation
Toho
The Asylum
Pixar
TMS Entertainment
Warner Bros. Pictures
United Artists|Eon Productions|Danjaq
Pixar|Walt Disney Pictures
0
10
20
Cantidad
Frecuencia de productionCompanyCountry
productionCompanyCountry
US
US|US
US|US|US
JP
US|
US|US|US|US
|US
US|US|
US|US
JP|JP
0
100
200
300
Cantidad
Frecuencia de productionCountry
productionCountry
United States of America
Japan
United Kingdom|United States of America
United Kingdom
Canada|United States of America
South Korea
Canada
Germany|United States of America
France
Australia
0
500
1000
1500
Cantidad
Frecuencia de video
video
0.0
0.2
0.4
0.6
0.8
1.0
12
3495
Cantidad
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00

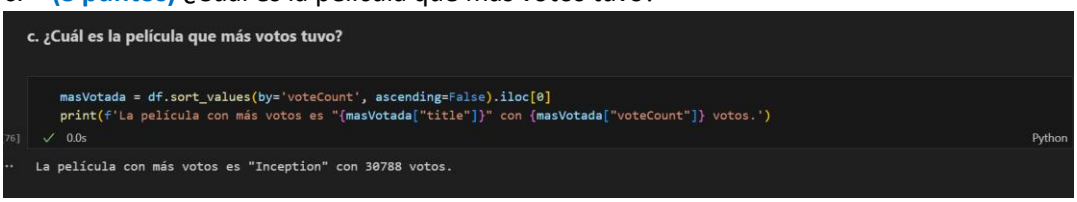
4. Responda las siguientes preguntas:
    a. **(3 puntos)** ¿Cuáles son las 10 películas que contaron con más presupuesto?




    b. **(3 puntos)** ¿Cuáles son las 10 películas que más ingresos tuvieron?




    c. **(3 puntos)** ¿Cuál es la película que más votos tuvo?




    d. **(3 puntos)** ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?

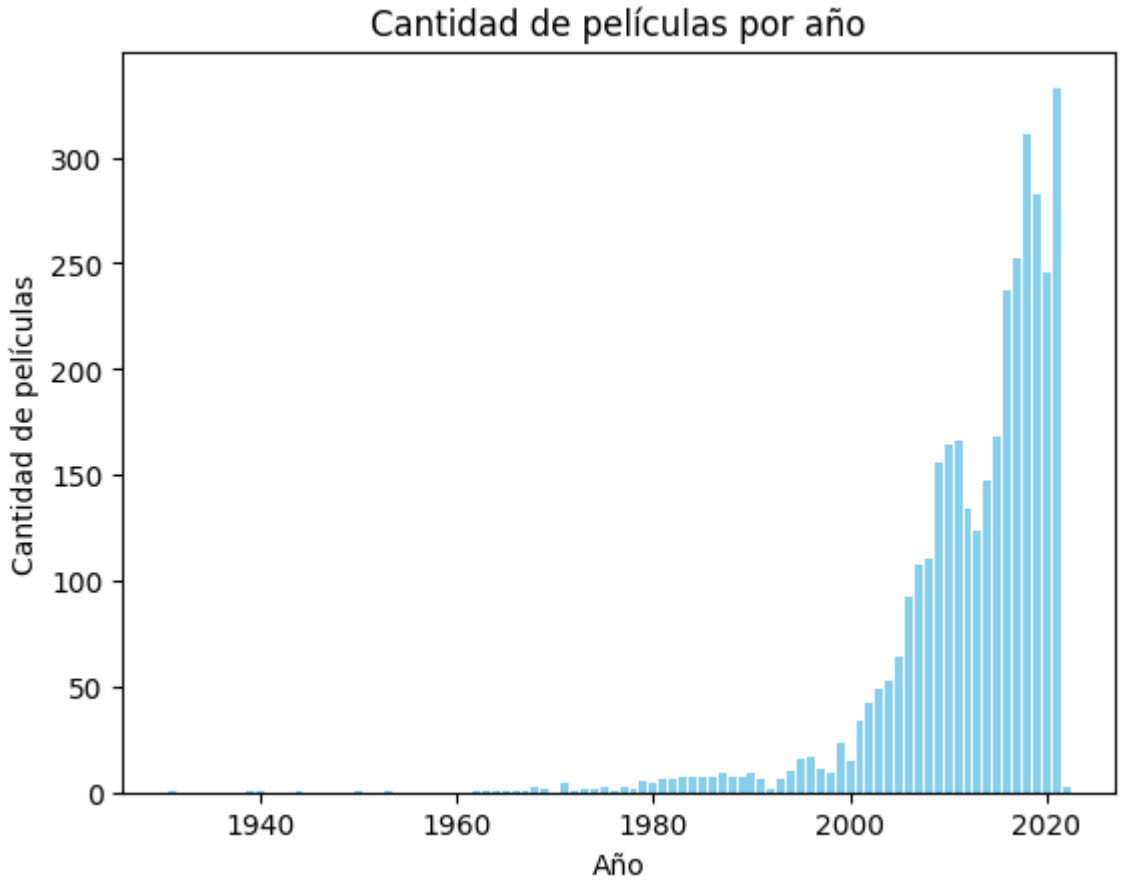
d. ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?

```
# d. ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?
peor = df.sort_values(by='voteAvg', ascending=True)
print('\nPeor pelicula segun los votos de todos los usuarios es: ')
print(peor[['title', 'voteAvg']].head(1))
```

[206] ✓ 0.0s Python

```
Peor pelicula segun los votos de todos los usuarios es:
                                                  title  voteAvg
9786  DAKAICHI -I'm Being Harassed by the Sexiest Ma...        1
```

e. **(8 puntos)** ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras



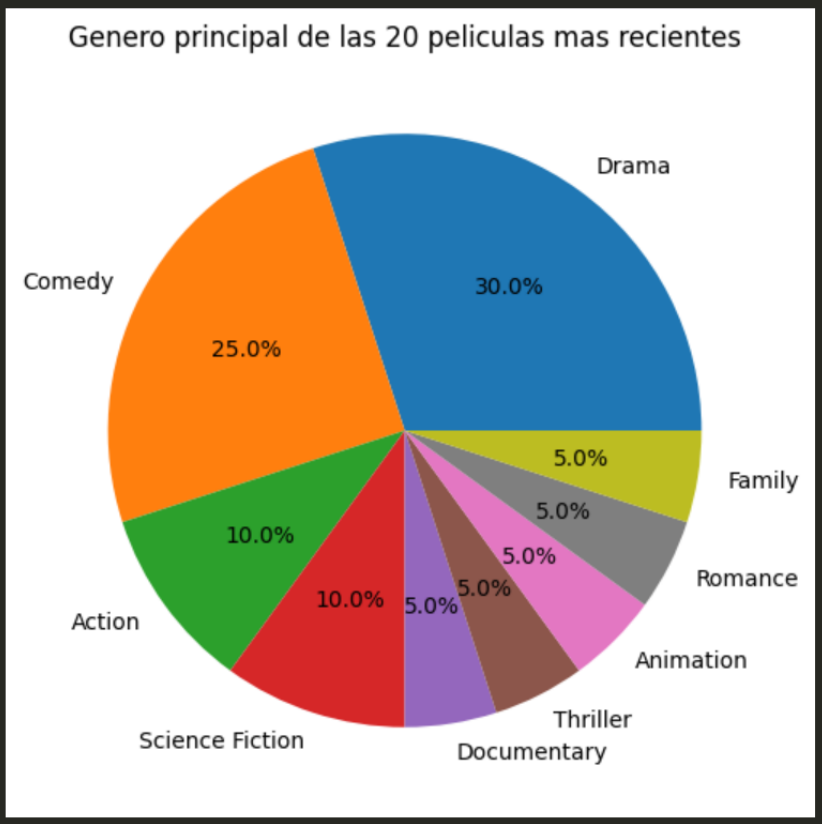

El año con más películas fue 2021 con 333 películas.

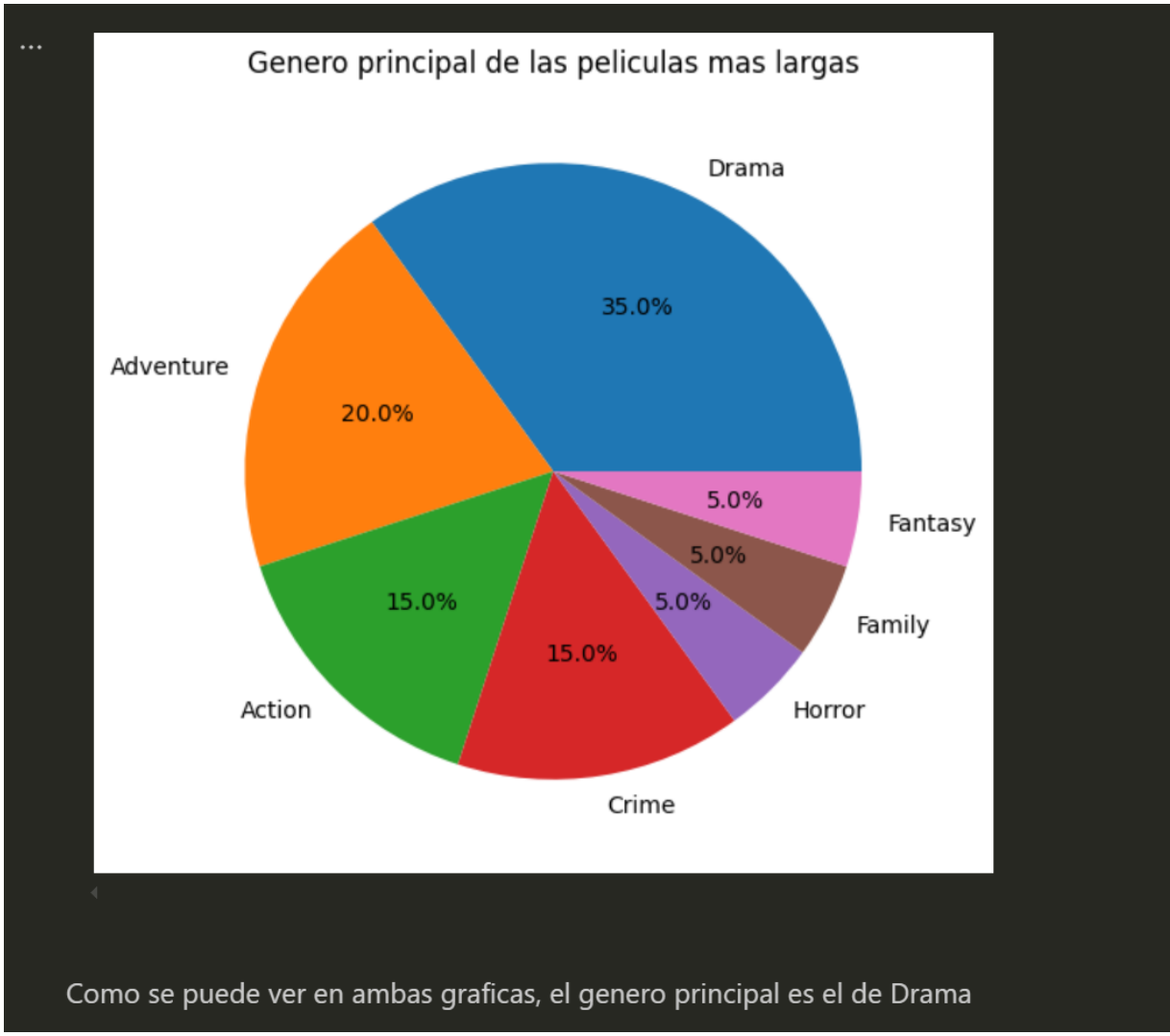
f. **(9 puntos)** ¿Cuál es el **género principal** de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Represéntelo usando un gráfico. ¿A qué género principal pertenecen las películas más largas?

f. ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Represéntelo usando un gráfico. ¿A qué género principal pertenecen las películas más largas?

```
#--------------------------------
'''
f.
¿Cuál es el género principal de las 20 películas más recientes?
¿Cuál es el género principal que predomina en el conjunto de datos?
 Represéntelo usando un gráfico.
 ¿A qué género principal pertenecen las películas más largas?
'''
#¿Cuál es el género principal de las 20 películas más recientes?
peliculas_recientes = df.sort_values(by='releaseDate', ascending=False)
top_20 = peliculas_recientes.head(20)
generos_top20= top_20['genres'].str.split('|').str[0]
generos_top20= generos_top20.value_counts()
print("Top generos:")
print(generos_top20)
nombres_generos= generos_top20.index
# print(nombres_generos)

#grafica
plt.figure(figsize=(10, 6))
plt.pie(generos_top20, labels=nombres_generos, autopct='%1.1f%%')
plt.title('Genero principal de las 20 peliculas mas recientes')
plt.show()

#¿A qué género principal pertenecen las películas más largas?
peliculas_mas_largas= df.sort_values(by='runtime', ascending=False)
top_20_pl = peliculas_mas_largas.head(20)
generos_pl = top_20_pl['genres'].str.split('|').str[0]
generos_pl = generos_pl.value_counts()
nombres_generos_pl = generos_pl.index

#grafica
plt.figure(figsize=(10, 6))
plt.pie(generos_pl, labels=nombres_generos_pl, autopct='%1.1f%%')
plt.title('Genero principal de las peliculas mas largas')
plt.show()
```

```
Top generos:
genres
Drama              6
Comedy             5
Action             2
Science Fiction    2
Documentary        1
Thriller           1
Animation          1
Romance            1
Family             1
Name: count, dtype: int64
```






g. **(8 puntos)** ¿Las películas de qué genero principal obtuvieron mayores ganancias?


h. **(3 puntos)** ¿La cantidad de actores influye en los ingresos de las películas? ¿Se han hecho películas con más actores en los últimos años?

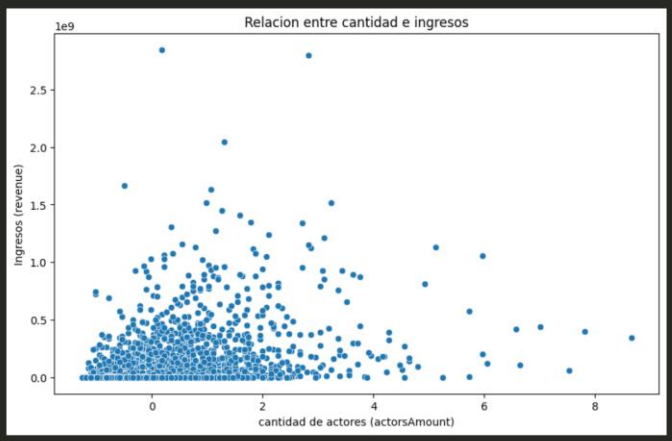
h. ¿La cantidad de actores influye en los ingresos de las películas? ¿Se han hecho películas con más actores en los últimos años?

```
'''h.
¿La cantidad de actores influye en los ingresos de las películas?
¿Se han hecho películas con más actores en los últimos años?
'''
#cantidad de actores influye en los ingresos de las películas
#correlacion de
correlacion_actores_ingresos = df[['actorsAmount', 'revenue']].corr()
print(correlacion_actores_ingresos)

plt.figure(figsize=(10, 6))
sns.scatterplot(x='actorsAmount', y='revenue', data=df)
plt.xlabel('cantidad de actores (actorsAmount)')
plt.ylabel('Ingresos (revenue)')
plt.title('Relacion entre cantidad e ingresos')
plt.show()
print('La grafica nos da una aproximacion de la relacion de la cantidad de actores y el ingreso, pero se necesita un analisis mas profundo y tomar en cuenta otras variables para determinar si la cantida


#anio mas reciente
anio_mas_reciente = df['releaseDate'].dt.year.max()
#filtrar los ultimos 10 a;os porque son los mas recientes
peliculas_5_ultimos_anios = df[df['releaseDate'].dt.year >= anio_mas_reciente - 10]

print(peliculas_5_ultimos_anios)

anios = []
cantidad_actores_anio = []
for anio, grupo in peliculas_5_ultimos_anios.groupby(peliculas_5_ultimos_anios['releaseDate'].dt.year):
    # print(anio)
    # print(grupo)
    #canidad de actores por a;o
    cantidad_actores = grupo['actorsAmount'] # grupo['castWomenAmount'] + grupo['castMenAmount']
    anios.append(anio)
    cantidad_actores_anio.append(cantidad_actores.mean())
    #print("Año:" + str(anio)+ " Cantidad de actores: " + str(sum(cantidad_actores)))
print(anios)
print(cantidad_actores_anio)

#grafica
plt.figure(figsize=(10, 6))
plt.plot(anios, cantidad_actores_anio, marker='o')
plt.title('Cantidad de actores en las peliculas por año')
```

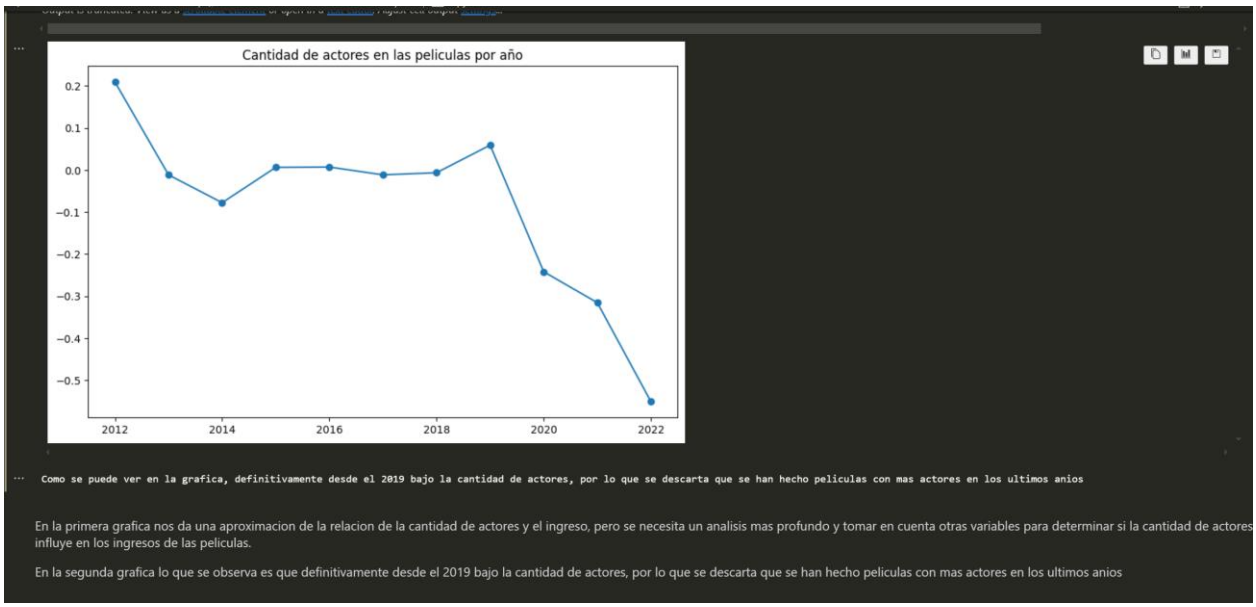
```
print(cantidad_actores_anio)

#grafica
plt.figure(figsize=(10, 6))
plt.plot(anios, cantidad_actores_anio, marker='o')
plt.title('Cantidad de actores en las peliculas por año')
plt.show()
print("Como se puede ver en la grafica, definitivamente desde el 2019 bajo la cantidad de actores, por lo que se descarta que se han hecho películas con mas actores en los ultimos anios")
```

✓ 0.1s

```
              actorsAmount   revenue
actorsAmount      1.000000  0.391464
revenue           0.391464  1.000000
```




```
La grafica nos da una aproximacion de la relacion de la cantidad de actores y el ingreso, pero se necesita un analisis mas profundo y tomar en cuenta otras variables para determinar si la cantidad de actores
          id     budget                         genres  \
110      189   65000000          Crime|Action|Thriller
739     1930  215000000        Action|Adventure|Fantasy
1874   10317   28000000                  Comedy|Drama
2007   10679    7500000  Action|Comedy|Science Fiction
2820   14564   25000000                         Horror
...      ...        ...                            ...
9955  885110     250000             War|History|Drama
9968  892342          0               Romance|Drama
9979  896633          0           Music|Documentary
9982  899082          0                 Documentary
9988  911068          0                      Comedy

                                          homePage  \
110                          http://sincity-2.com/
739                http://www.theamazingspiderman.com
1874             http://www.ourbrandisscrisismovie.com/
2007                         http://www.ironsky.net/
2820                       http://www.cingsmovie.com/
...                                              ...
9955           https://www.netflix.com/title/81450071
9968           https://www.netflix.com/title/81387787
9979    https://www.cbs.com/shows/adele-one-night-only/
9982  https://www.hbomax.com/feature/urn:hbo:feature...
...
```




i. **(3 puntos)** ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?

**i. ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?**

```
# Asegurar que las columnas sean numéricas
df["castWomenAmount"] = pd.to_numeric(df["castWomenAmount"], errors="coerce")
df["castMenAmount"] = pd.to_numeric(df["castMenAmount"], errors="coerce")
df["popularity"] = pd.to_numeric(df["popularity"], errors="coerce")
df["revenue"] = pd.to_numeric(df["revenue"], errors="coerce")

# Calcular correlaciones
correlations = df[["castWomenAmount", "castMenAmount", "popularity", "revenue"]].corr()
print(correlations)
```

[79] ✓ 0.0s Python

```
                  castWomenAmount  castMenAmount  popularity   revenue
castWomenAmount          1.000000       0.507684    0.048948  0.242131
castMenAmount            0.507684       1.000000    0.070101  0.440083
popularity               0.048948       0.070101    1.000000  0.173926
revenue                  0.242131       0.440083    0.173926  1.000000
```

Se puede observar que hay una correlación un poco más alta para los ingresos para los hombres que con las mujeres, entonces puede que aqui haya un sesgo, no es una gran diferencia, pero no deja de ser significativo. Se usó esta método para ver la correlación entre las variables

j. **(8 puntos)** ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

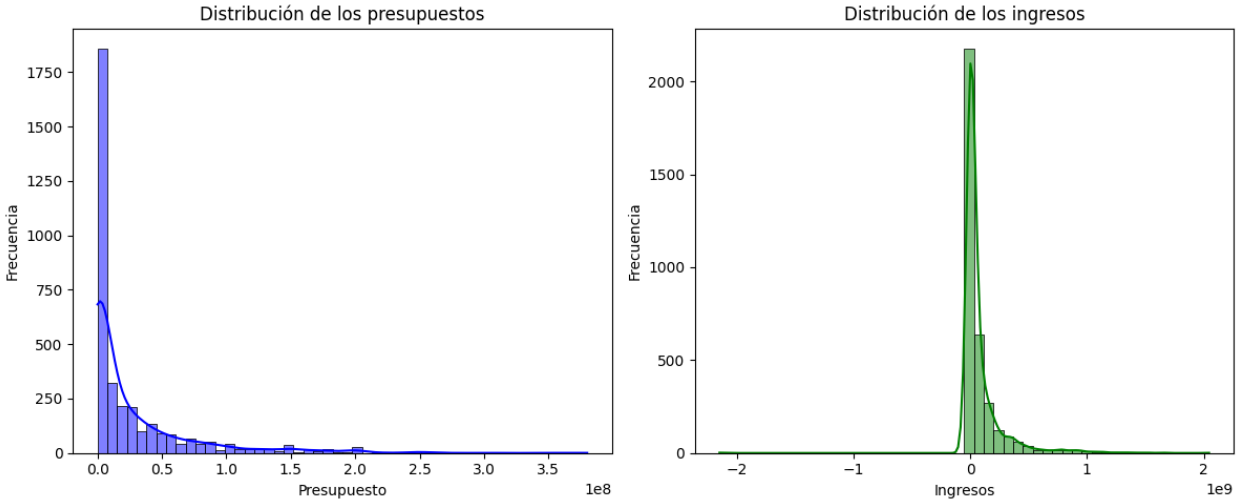
j. ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

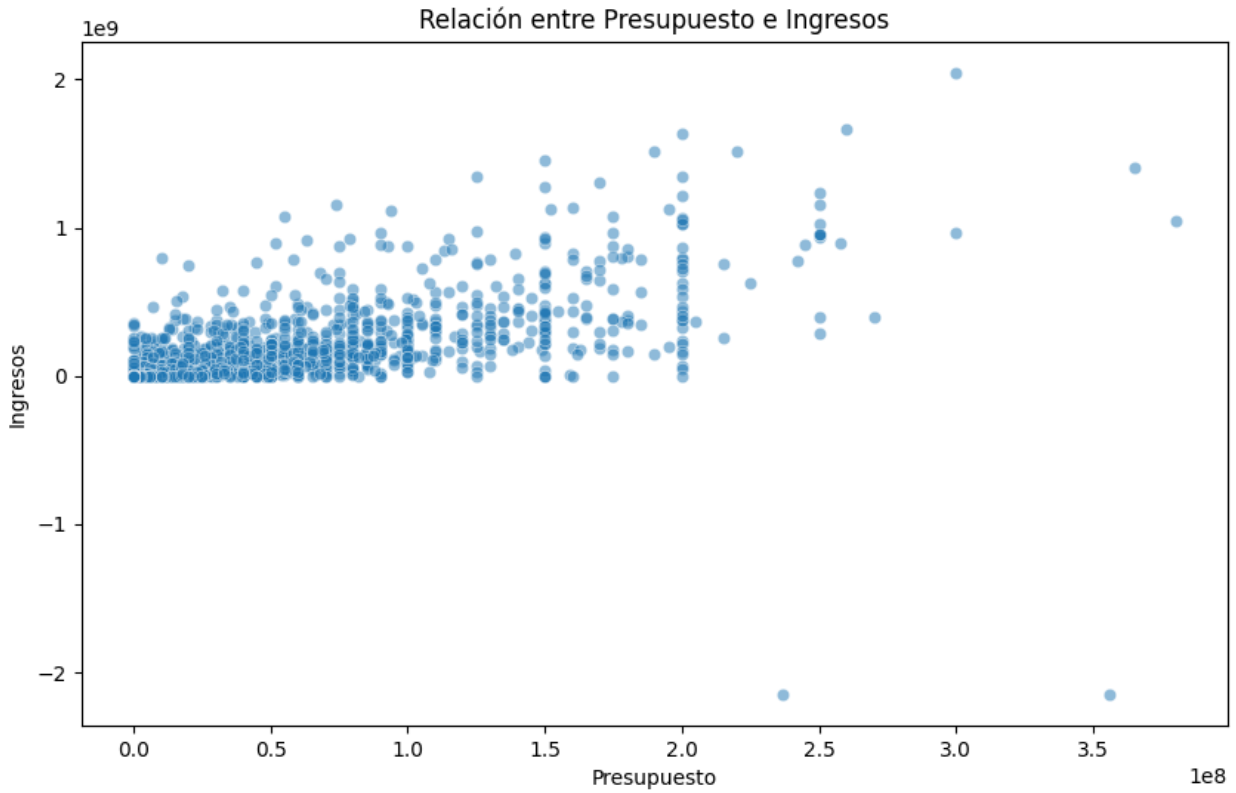
```
'''j.
¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?

'''
print("El listado de directores y con las 20 peliculas mejor calificadas es:\n")
peliculas_mejores_calificadas = df.sort_values(by='voteAvg', ascending=False).head(20)
for i, row in peliculas_mejores_calificadas.iterrows():
    print('director : '+str(row['director']) + '    | pelicula : '+str(row['title']))
```

✓ 0.0s

```
El listado de directores y con las 20 peliculas mejor calificadas es:

director : Park Jun-soo    | pelicula : Bring the Soul: The Movie
director : Park Jun-soo    | pelicula : Break the Silence: The Movie
director : Lin-Manuel Miranda    | pelicula : tick, tick… BOOM!
director : Martin Scorsese    | pelicula : GoodFellas
director : Roger Allers|Rob Minkoff    | pelicula : The Lion King
director : Eran Creevy|Giorgio Testi|Joe Pearlman|Casey Patterson    | pelicula : Harry Potter 20th Anniversary: Return to Hogwarts
director : Destin Daniel Cretton    | pelicula : Just Mercy
director : Paul Dugdale    | pelicula : ariana grande: excuse me, i love you
director : Mamoru Hosoda    | pelicula : Belle
director : Robert Rodriguez|Patrick Osborne    | pelicula : Happier Than Ever: A Love Letter to Los Angeles
director : Thor Freudenthal    | pelicula : Words on Bathroom Walls
director : Irvin Kershner    | pelicula : The Empire Strikes Back
director : Joe Wright    | pelicula : Pride & Prejudice
director : Takayuki Hamana    | pelicula : The Seven Deadly Sins: Cursed by Light
director : Carolina María García Fernández    | pelicula : Soy Luna: The Last Concert
director : Kenji Nagasaki    | pelicula : My Hero Academia: Heroes Rising
director : Makoto Shinkai    | pelicula : Weathering with You
director : Tosca Musk    | pelicula : Gabriel's Inferno
director : Alfonso Cuarón    | pelicula : Harry Potter and the Prisoner of Azkaban
director : Tsutomu Hanabusa    | pelicula : Tokyo Revengers
```

k. **(8 puntos)** ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión





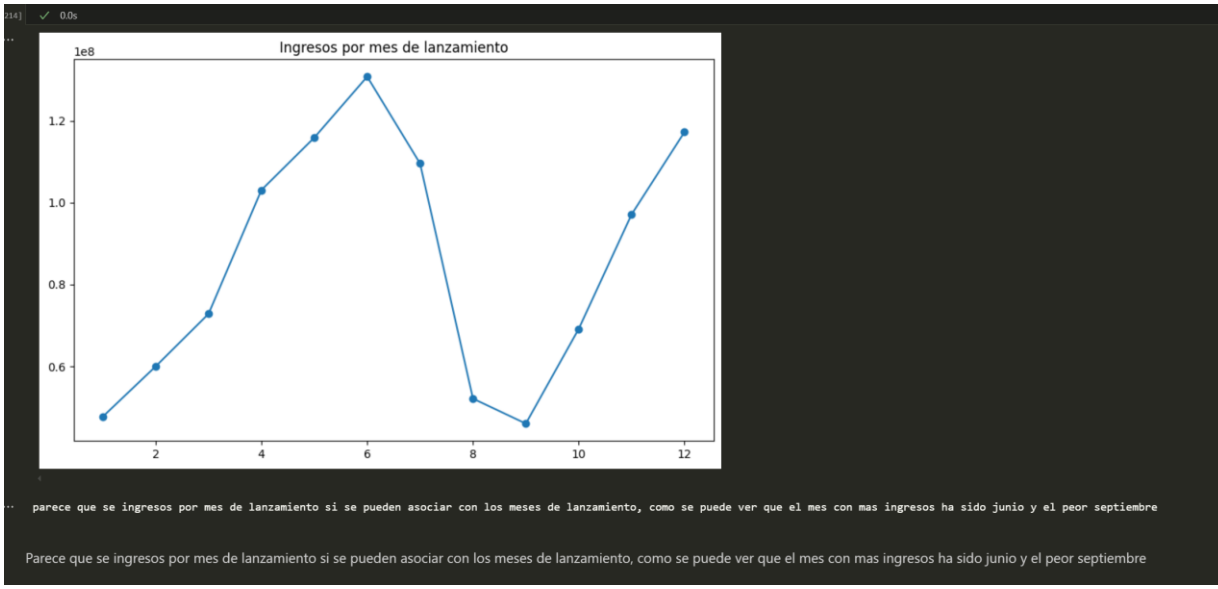

Se puede concluir que si hay una relación en la mayoría de casos, pero hay datos atípicos que sesgan la correlación entre las variables. Como la correlación no es mayor a 0.7, no se puede asegurar que en todos los casos mientras mas presupuesto haya, más ganancias habrá, pero si hay casos en los que esto se cumplirá.

l. **(5 puntos)** ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

l. ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

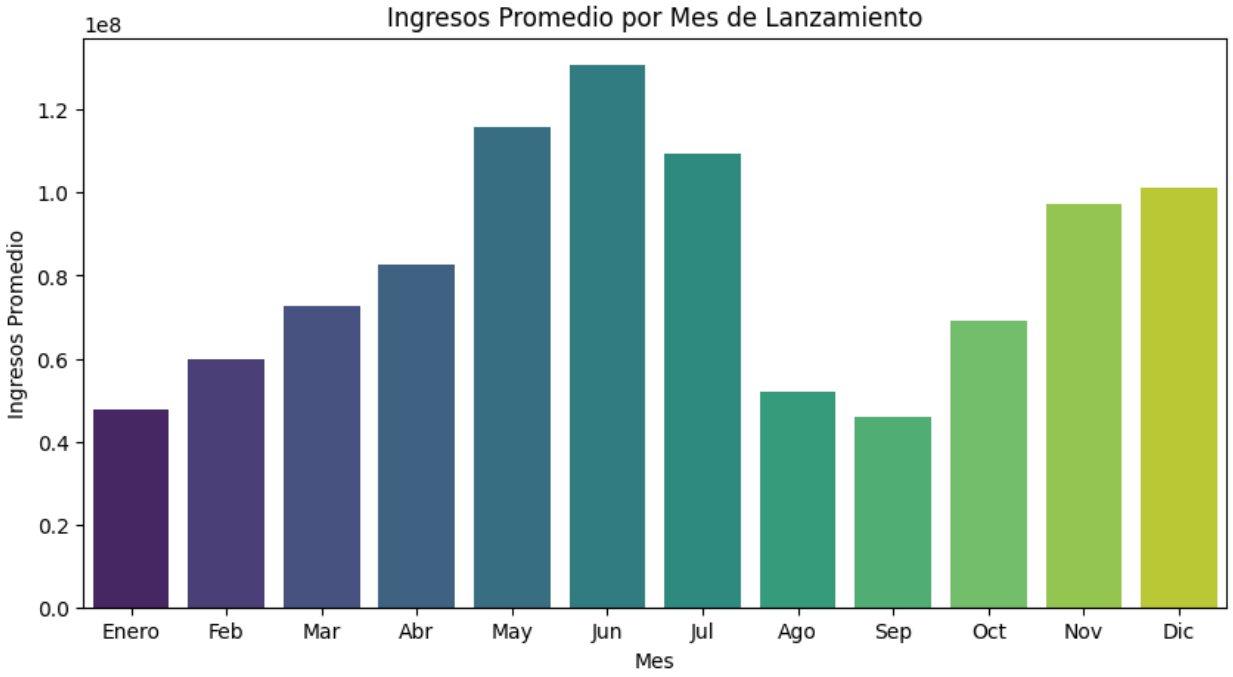
```
'''1.
¿Se asocian ciertos meses de lanzamiento con mejores ingresos?

'''
#meses de lanzamiento
meses_de_lanzamiento = df['releaseDate'].dt.month

#agrupar por mes y calcular promedio de ingresos
ingresos_por_mes = df.groupby(meses_de_lanzamiento)['revenue'].mean()

#grafica
plt.figure(figsize=(10, 6))
plt.plot(ingresos_por_mes.index, ingresos_por_mes, marker='o')
plt.title('Ingresos por mes de lanzamiento')
plt.show()
print('parece que se ingresos por mes de lanzamiento si se pueden asociar con los meses de lanzamiento, como se puede ver que el mes con mas ingresos ha sido junio y el peor septiembre')
```

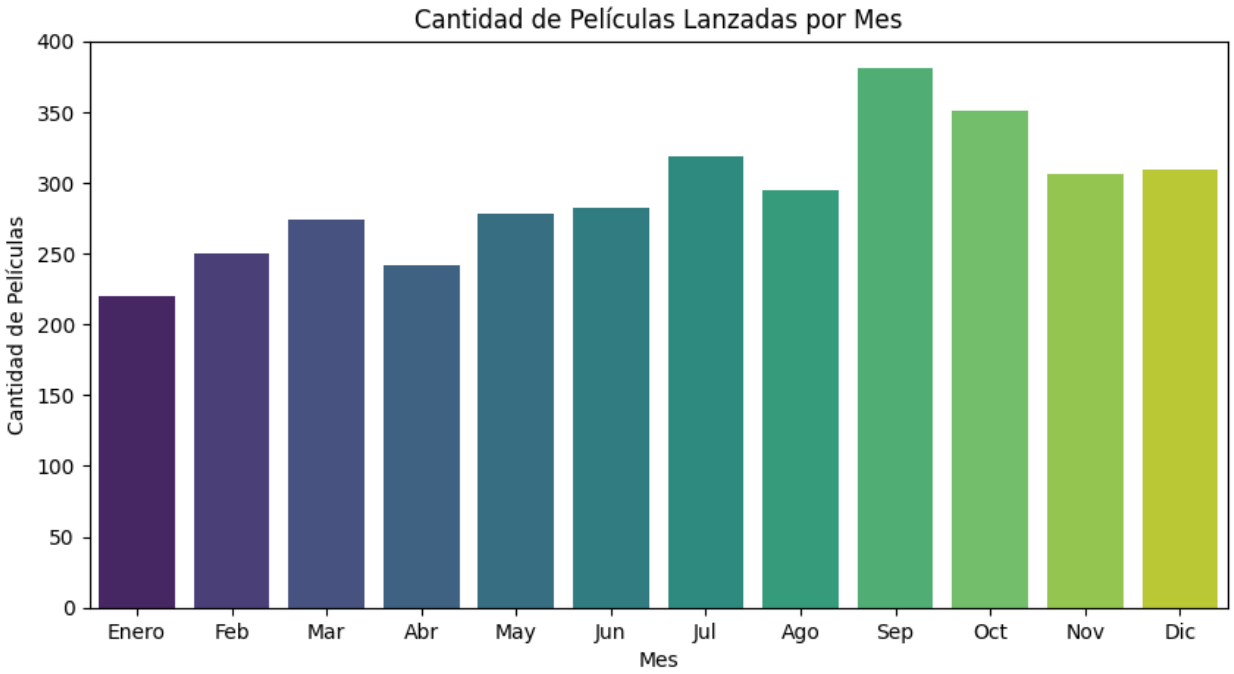
[214] ✓ 0.0s



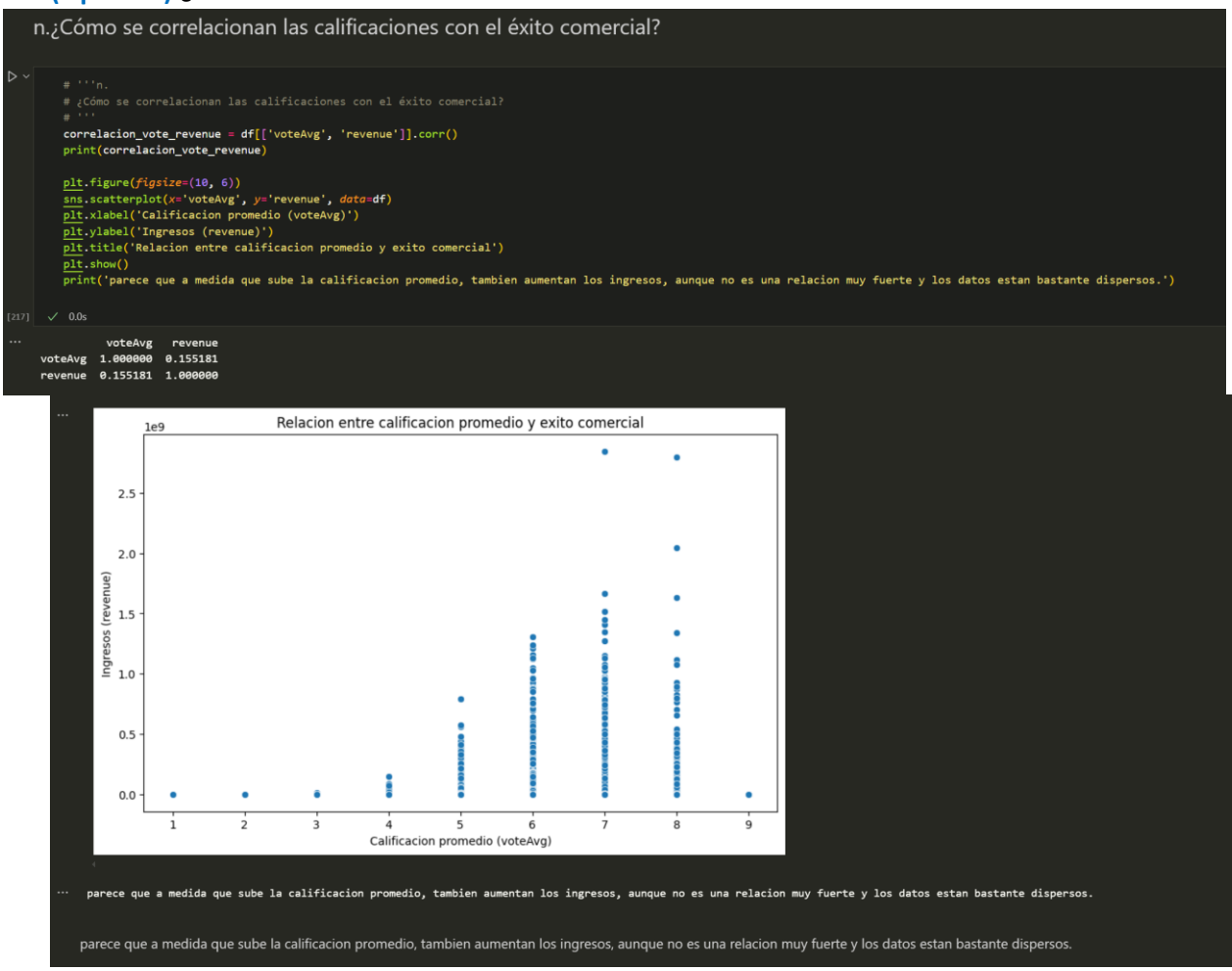

m. **(6 puntos)** ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿Cuántas películas, en promedio, se han lanzado por mes?

En los meses de mayo, junio y julio
Se han lanzado en promedio 300 peliculas





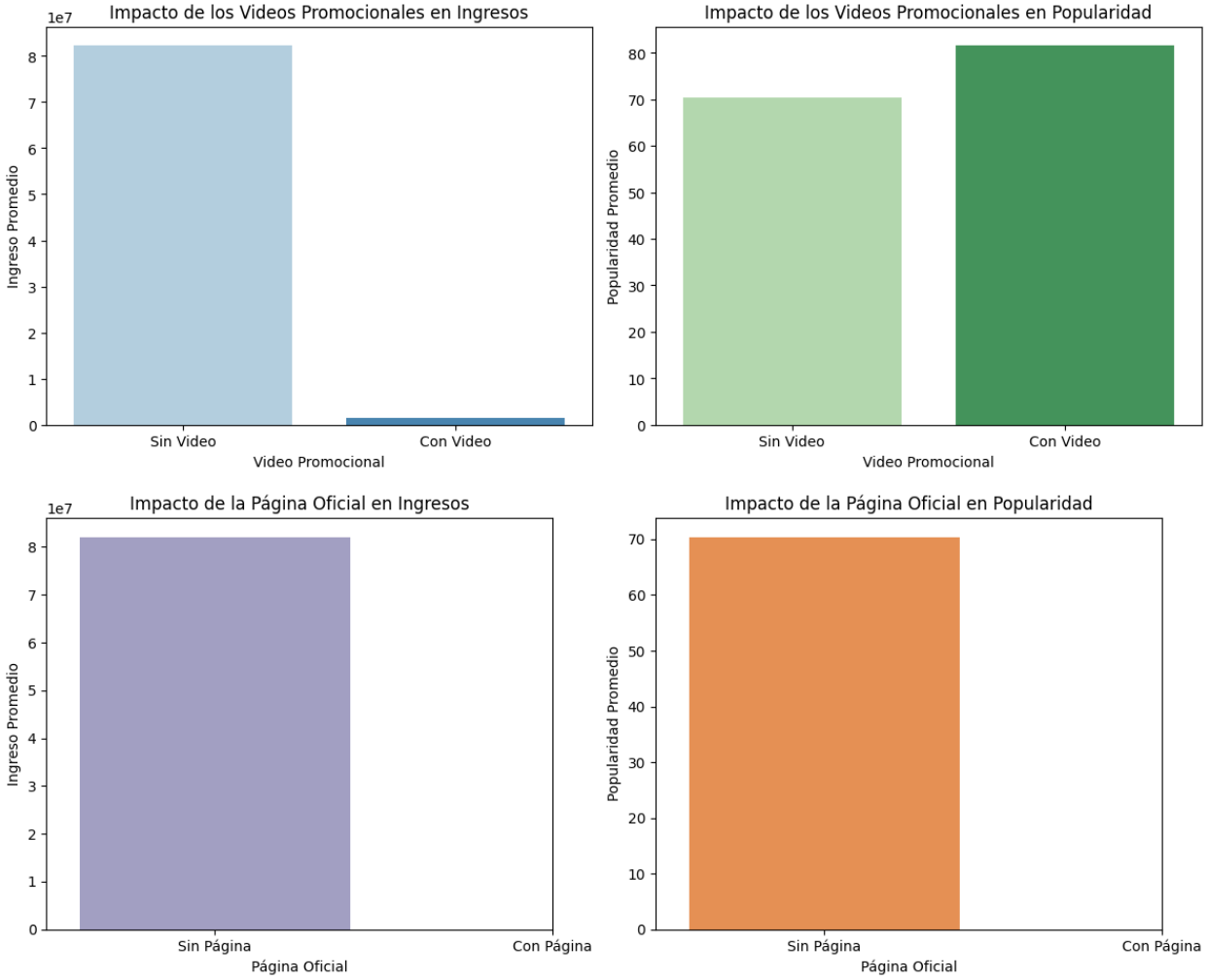

n. **(7 puntos)** ¿Cómo se correlacionan las calificaciones con el éxito comercial?



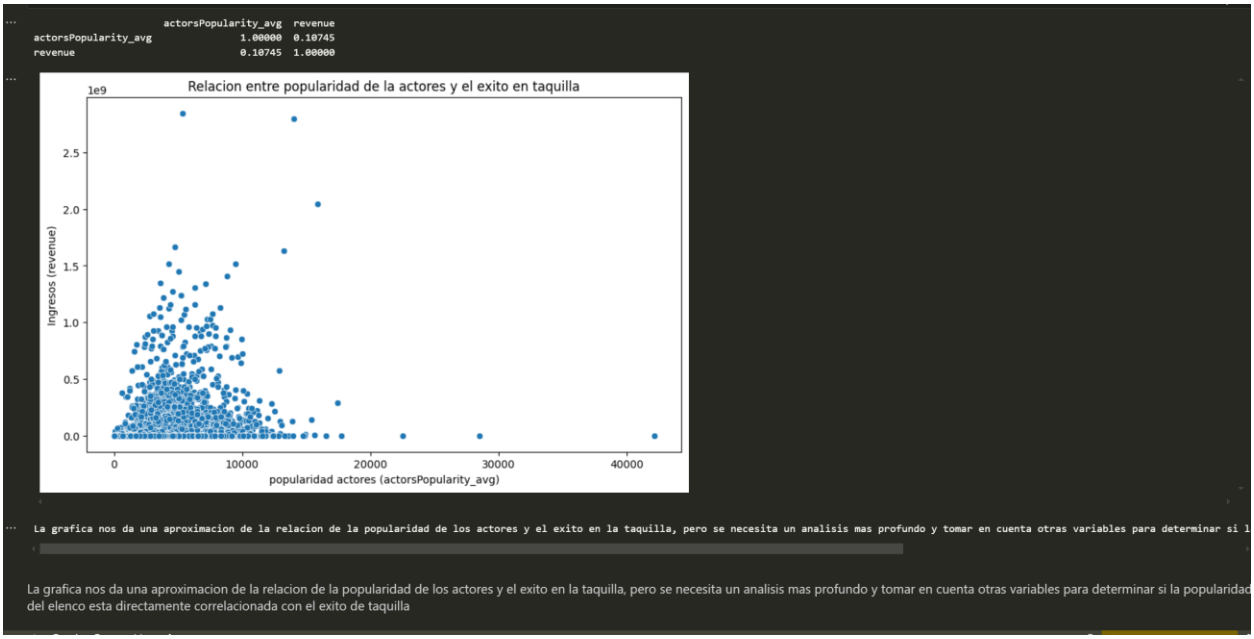

o. **(5 puntos)** ¿Qué estrategias de marketing, como videos promocionales o páginas

oficiales, generan mejores resultados?

Según las gráficas podemos concluir que los videos tienen más influencia en la popularidad




p. **(4 puntos)** ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?




## MATERIAL A ENTREGAR

- Enlace de **Sharepoint (Microsoft Word)** con el informe de análisis exploratorio. Se debe poder verificar el **historial de cambios**. Este informe debe incluir:
  - Enunciado de la pregunta que se está respondiendo.
  - Respuesta con su respectiva explicación.
  - Gráfico, si aplica, de acuerdo con la pregunta.

  Proyecto 01 - HT01 - Análisis Exploratorio-1.docx

- Enlace del **repositorio de github** donde se tendrán en cuenta los aportes de todos los integrantes del grupo. Este debe de estar **público**.
  angelargd8/HDT01-MDD