

Autores: Fabiola Contreras, 22787 María José Villafuerte, 22129	Docente: Mario Barrientos Laboratorio 1
Sección: 20	Fecha: 02/02/2025

Hoja de Trabajo 01 – Análisis Exploratorio Proyecto 01

Índice:

I. Exploración de datos:	2
II. Clasificación de variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta).....	2
III. Variables cualitativas – Distribución normal	3
IV. Preguntas Clave.....	4
1. ¿Cuáles son las 10 películas que contaron con más presupuesto?	4
2. ¿Cuáles son las 10 películas que más ingresos tuvieron?(en orden de mayor a menor)	4
3. ¿Cuál es la película que más votos tuvo?	4
4. ¿Cuál es la peor película de acuerdo con los votos de todos los usuarios?	4
5. ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras.....	4
6. ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Representelo usando un gráfico.....	5
7. ¿Las películas de qué genero principal obtuvieron mayores ganancias?	6
8. ¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?	6
9. ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?	7
10. ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?	7
11. ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión....	8
12. ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?.....	9
13. ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuántas películas, en promedio, se han lanzado por mes?	9
14. ¿Cómo se correlacionan las calificaciones con el éxito comercial?	10
15. ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?..	10
16. ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?.....	10
V. Repositorio de github:	11

I. Exploración de datos:

El análisis inicial del conjunto de datos reveló una estructura de 27 columnas. Mediante el uso de estadísticos descriptivos, obtuvimos una primera vista de los datos, incluyendo promedios y otras métricas básicas. Al examinar los tipos de datos, pudimos identificar la naturaleza de cada variable. Cabe mencionar que detectamos algunos valores faltantes en determinadas columnas, aspecto que consideraremos en nuestro análisis.

II. Clasificación de variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta)

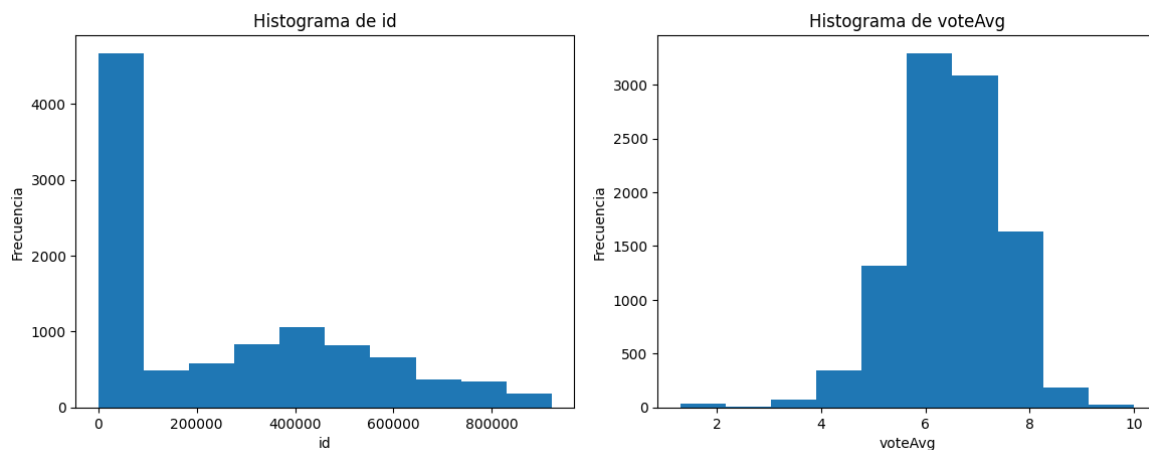
<i>variable</i>	<i>descripción</i>	<i>tipo</i>
<i>Id</i>	ID de la película	Cualitativa ordinal
<i>Popularity</i>	Índice de popularidad de la película calculado semanalmente	Cuantitativas continuas
<i>Budget</i>	El presupuesto para la película.	Cuantitativas continuas
<i>Revenue</i>	El ingreso de la película.	Cuantitativas continuas
<i>Original_title</i>	El título original de la película, en su idioma original.	Cualitativa nominal
<i>originalLanguage</i>	Idioma original en que se encuentra la película	Cualitativa nominal
<i>Title</i>	El título de la película traducido al inglés	Cualitativa nominal
<i>homePage</i>	La página de inicio de la película	Cualitativa nominal
<i>Video</i>	Si tiene videos promocionales o no	Cualitativa nominal
<i>Director</i>	Director de la película	Cualitativa nominal
<i>Runtime</i>	La duración de la película	Cuantitativas continuas
<i>Genres</i>	El género de la película	Cualitativa nominal
<i>genresAmount</i>	Cantidad de géneros que representan la película	Cuantitativas discretas
<i>productionCompany</i>	Las compañías productoras de la película	Cualitativa nominal
<i>productionCoAmount</i>	Cantidad de compañías productoras que participaron en la película	Cuantitativas discretas
<i>productionCompanyCountry</i>	Países de las compañías productoras de la película	Cualitativa nominal

<i>productionCountry</i>	Países en los que se llevó a cabo la producción de la película	Cualitativa nominal
<i>productionCountriesAmount</i>	Cantidad de países en los que se rodó la película	Cuantitativas discretas
<i>releaseDate</i>	Fecha de lanzamiento de la película	Cualitativa ordinal
<i>voteCount</i>	El número de votos en la plataforma para la película	Cuantitativas discretas
<i>voteAvg</i>	El promedio de los votos en la plataforma para la película	Cuantitativas continuas
<i>Actors</i>	Actores que participan en la película (Elenco)	Cualitativa nominal
<i>actorsPopularity</i>	Índice de popularidad del elenco de la película	Cuantitativas continuas
<i>actorsCharacter</i>	Personaje que interpreta cada actor en la película	Cualitativa nominal
<i>actorsAmount</i>	Cantidad de personas que actúan en la película	Cuantitativas discretas
<i>castWomenAmount</i>	Cantidad de actrices en el elenco de la película	Cuantitativas discretas
<i>castMenAmount</i>	Cantidad de actores en el elenco de la película	Cuantitativas discretas

- El análisis de las variables está presente en el jupyter notebook adjunto en el repositorio de github

III. Variables cualitativas – Distribución normal

El análisis de las variables cuantitativas revela que ninguna sigue una distribución normal. Si bien algunas variables como Id y VoteAvg exhiben visualmente patrones que se asemejan a una distribución normal, las pruebas estadísticas de normalidad demuestran que no cumplen con los criterios requeridos para ser consideradas normalmente distribuidas.



IV. Preguntas Clave

1. *¿Cuáles son las 10 películas que contaron con más presupuesto?*
 - a. Pirates of the Caribbean: On Stranger Tides
 - b. Avengers: Age of Ultron
 - c. Avengers: Endgame
 - d. Pirates of the Caribbean: At World's End
 - e. Justice League
 - f. Avengers: Infinity War
 - g. Superman Returns
 - h. Tangled
 - i. The Lion King
 - j. Spider-Man 3

2. *¿Cuáles son las 10 películas que más ingresos tuvieron?(en orden de mayor a menor)*
 - a. Avatar
 - b. Avengers: Endgame
 - c. Titanic
 - d. Star Wars: The Force Awakens
 - e. Avengers: Infinity War
 - f. Jurassic World
 - g. The Lion King
 - h. Spider-Man: No Way Home
 - i. Avengers
 - j. Furious 7

3. *¿Cuál es la película que más votos tuvo?*

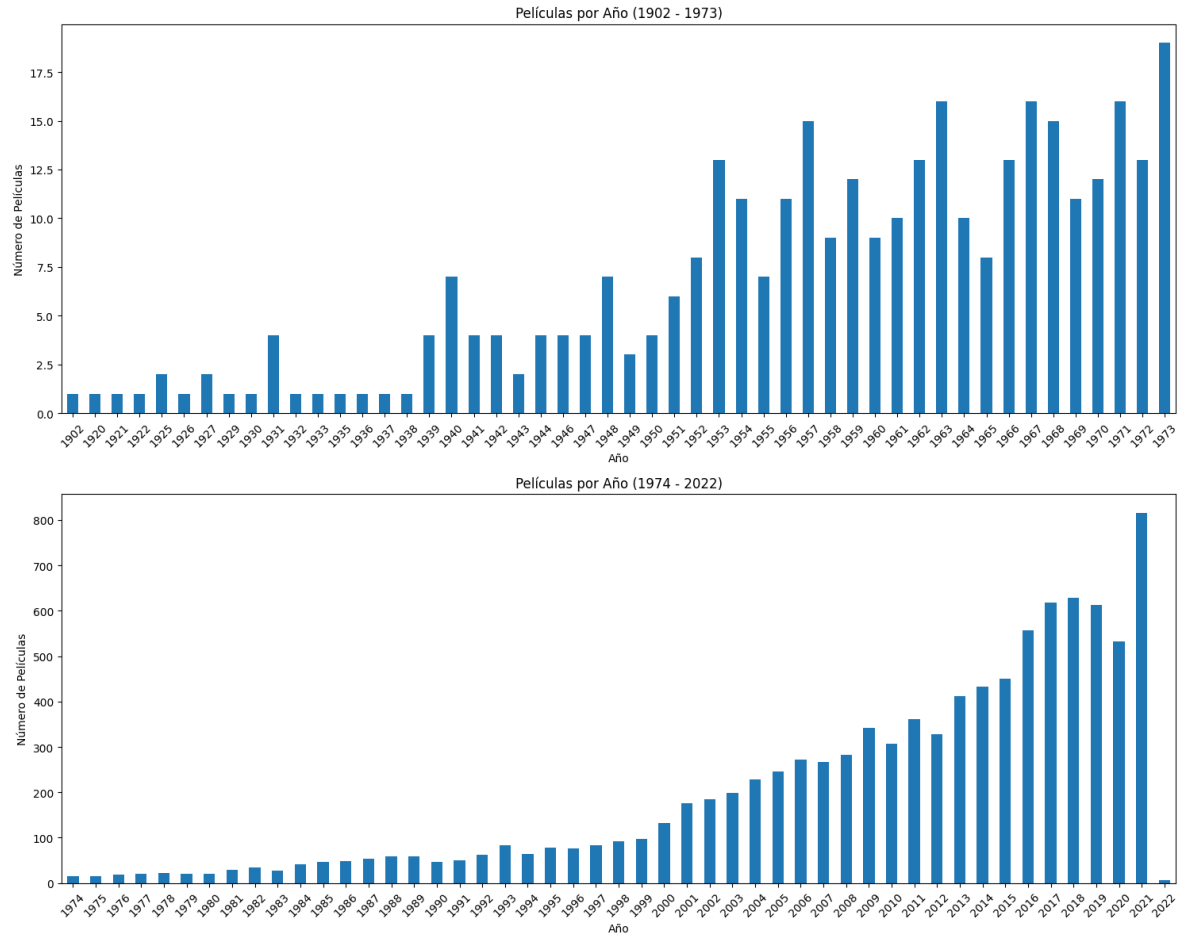
Inception 30788

4. *¿Cuál es la peor película de acuerdo con los votos de todos los usuarios?*

DAKAICHI -I'm Being Harassed by the Sexiest Man of the Year- The Movie: In Spain

Esta es una película japonesa, el promedio de votos fue de 1.3

5. *¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras*



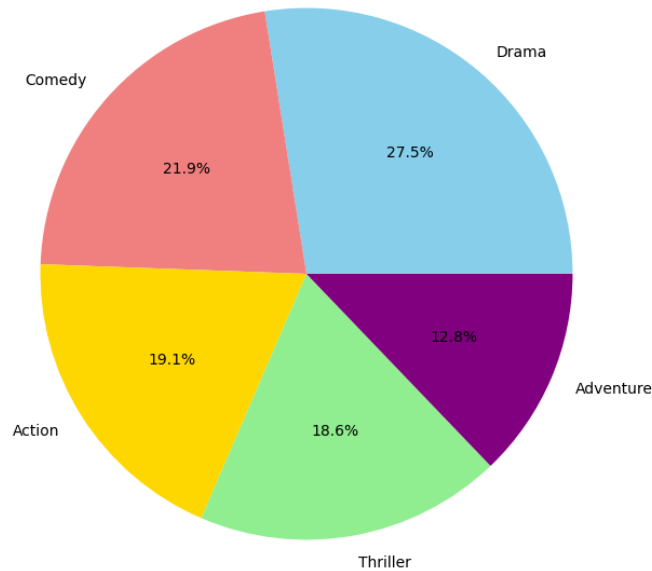
Año con más películas: 2021 (816 películas)

6. ¿Cuál es el género principal de las 20 películas más recientes? ¿Cuál es el género principal que predomina en el conjunto de datos? Representelo usando un gráfico.

El género principal de las 20 películas más recientes es el Drama

El género que predomina entre todos es drama

Distribución de Géneros (Top 5)



- a. ¿A qué género principal pertenecen las películas más largas?
Pertenecen a documentales

7. *¿Las películas de qué género principal obtuvieron mayores ganancias?*

Ingresos promedio por género principal:

Adventure	1.592496e+08
Action	9.284366e+07
Fantasy	9.257043e+07
Science Fiction	8.449515e+07

Aventura con 159,249,600\$ (aproximadamente 159.2 millones)

8. *¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?*

Los gráficos sugieren que la relación entre la cantidad de actores y los ingresos de una película es débil. Otros factores, como el género, la calidad de la película, el marketing, y el momento del estreno, tienen una influencia mucho mayor en los ingresos que el simple número de actores.

Sí, en los últimos diez años registrados ha aumentado la cantidad de actores presentes en las películas. Aunque hay momentos de ‘bajada’ se ve una tendencia a tener bastantes actores desde el año 2000.

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

No hay un patrón claro que sugiera que un mayor o menor ratio afecte significativamente el desempeño. La evidencia sugiere que el ratio de género en el reparto no tiene un impacto significativo en la popularidad o ingresos de las películas.

Correlación ratio género vs popularidad: -0.003

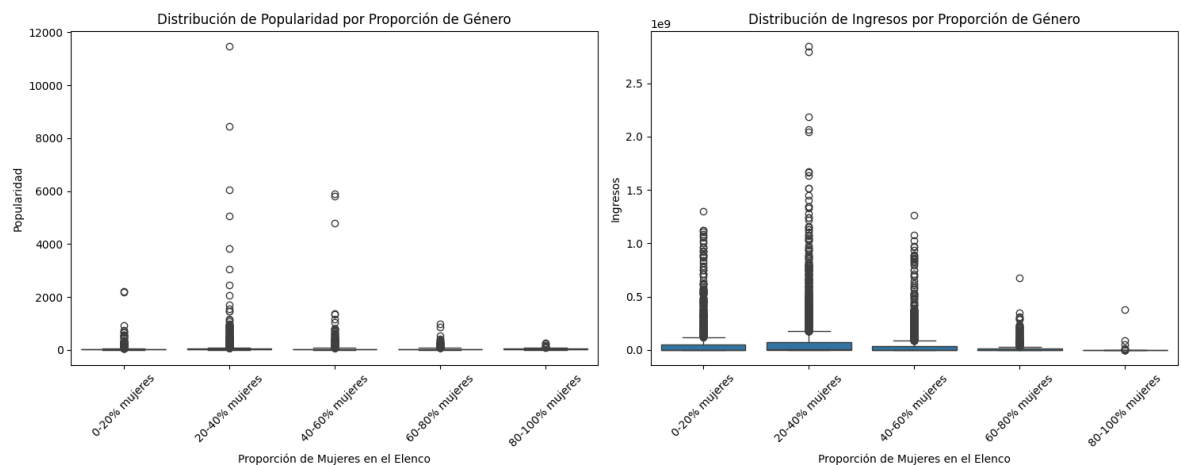
Correlación ratio género vs ingresos: -0.072

($|r| > 0.3$ para una correlación débil

$|r| > 0.5$ para una correlación moderada

$|r| > 0.7$ para una correlación fuerte)

Pero la gráfica si muestra que las películas con mayor popularidad y mayor ingreso se encuentran cuando hay un porcentaje del 20-40 porciento del mujeres por ende de 80-60 porciento de hombre.

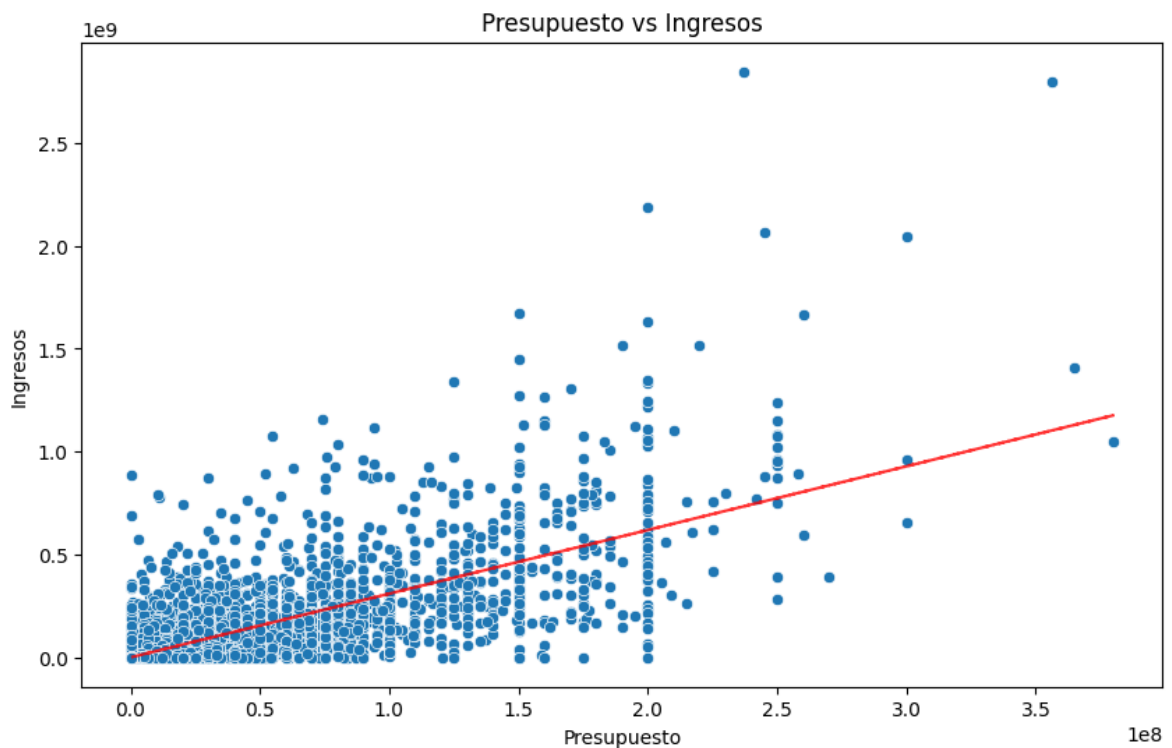


10. *¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?*

Laurent Bouzereau	10.0
Thomas Coven	10.0
Rebecca Sugar	10.0
Christin Baker	10.0
Víctor Barba Juan Olivares	10.0
Miguel Angel Zavala	10.0
Kaku Arakawa	10.0
Dave Bullock Troy Adomitis Victor Cook	9.6
Won Myeong-jun	9.5
Samuel Leong	9.5
Selena Quintanilla	9.4

Haruo Sotozaki	9.3
Haruo Sotozaki	9.3
Park Jun-soo	9.2
Kim Nam-joon Jeon Jung-kook Kim Tae-hyung Park...	9.2
Ulises Valencia	9.2
Igor Kopylov	9.2
Park Jun-soo	9.1
Michael Goguen	9.0
Cybil Richards	9.0

11. ¿Cómo se correlacionan los presupuestos con los ingresos? ¿Los altos presupuestos significan altos ingresos? Haga los gráficos que necesite, histograma, diagrama de dispersión



Si podemos observar que a mayor presupuesto mejor es el ingreso, ya que existen más puntos en la parte de arriba del diagrama de dispersión. Mas que eso podemos observar un Clustering en la parte inferior, indicando que la mayoría de las películas de nuestros datos fueron hechas con menos de 1 billón de dólares.

Con esto en mente si hay que aclarar que no hay una relación estable entre presupuesto e ingresos.

La correlación obtenida es la siguiente “Correlación entre presupuesto e ingresos: 0.756”. Podemos afirmar que existe una correlación fuerte y positiva entre presupuesto

e ingresos. Este valor sugiere que aproximadamente el 57% ($r^2 = 0.756^2 \approx 0.57$) de la variación en los ingresos puede explicarse por el presupuesto.

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

Sí, la tendencia nos indica que los meses de junio y diciembre son los que tienen mejores ingresos. Esto tiene sentido porque es en esas fechas que hay niños de vacaciones, por lo que hay más personas llegando al cine.

13. *¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuántas películas, en promedio, se han lanzado por mes?*

Junio

R//

Mes: Junio
Ingreso promedio: \$100,086,398.08
Cantidad de películas: 775

Mes: Mayo
Ingreso promedio: \$90,973,469.61
Cantidad de películas: 674

Mes: Julio
Ingreso promedio: \$78,144,684.91
Cantidad de películas: 790

Mes: Diciembre
Ingreso promedio: \$77,336,387.15
Cantidad de películas: 899

Mes: Noviembre
Ingreso promedio: \$73,928,377.88
Cantidad de películas: 780

Mes: Abril
Ingreso promedio: \$54,103,493.60
Cantidad de películas: 673

Mes: Marzo
Ingreso promedio: \$52,934,552.35
Cantidad de películas: 787

Mes: Febrero
Ingreso promedio: \$45,012,151.22
Cantidad de películas: 673

Mes: Octubre
Ingreso promedio: \$40,254,663.67
Cantidad de películas: 1027

Mes: Agosto
Ingreso promedio: \$36,899,643.31
Cantidad de películas: 890

Mes: Enero
Ingreso promedio: \$35,232,706.27
Cantidad de películas: 625

Mes: Septiembre
Ingreso promedio: \$33,109,915.48
Cantidad de películas: 1039

14. *¿Cómo se correlacionan las calificaciones con el éxito comercial?*

Según el resultado numérico (0.141) **no hay una relación lineal clara** entre las calificaciones y los ingresos.

15. *¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?*

Según los resultados

Correlación popularidad del elenco vs ingresos: 0.091

Correlación popularidad del elenco vs popularidad película: 0.084

No hay una correlación débil entre la popularidad del elenco y los ingresos al igual que la popularidad del elenco y la popularidad de la película.

16. *¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?*

Puede observarse que las páginas oficiales tienen mejores resultados. Han tenido mejores resultados aquellas películas con página oficial y sin videos.

V. Repositorio de github:

https://github.com/Fabiola-cc/CineVision_Studios.git