

Laboratorio 1. Análisis Exploratorio

INSTRUCCIONES:

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? Representelo usando un gráfico.
 1. ¿A qué género principal pertenecen las películas más largas?
7. ¿Las películas de qué género 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? ¿cuántas 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** (R o python) utilizado pues están pensando en contratar un analista de datos de planta que continúe con el análisis a lo largo del tiempo. Le están pidiendo que muestre todos los aportes del equipo en el repositorio de github.

DESCRIPCIÓN DEL DATASET

El dataset contiene datos de 19,883 películas obtenidos de la plataforma "[The movie DB](https://www.kaggle.com/datasets/tommhughes/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 generado por la película.

- originalTitle: 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 en minutos.
- 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.
- releaseYear: Año de lanzamiento

EJERCICIOS

1. **(3 puntos)** Haga una exploración rápida de sus datos, para eso haga un resumen de su conjunto de datos.
2. **(5 puntos)** Diga el tipo de cada una de las variables (cualitativa ordinal o nominal, cuantitativa continua, cuantitativa discreta)
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.**
4. Responda las siguientes preguntas:
 - 4.1. **(3 puntos)** ¿Cuáles son las 10 películas que contaron con más presupuesto?
 - 4.2. **(3 puntos)** ¿Cuáles son las 10 películas que más ingresos tuvieron?
 - 4.3. **(3 puntos)** ¿Cuál es la película que más votos tuvo?
 - 4.4. **(3 puntos)** ¿Cuál es la peor película de acuerdo a los votos de todos los usuarios?
 - 4.5. **(8 puntos)** ¿Cuántas películas se hicieron en cada año? ¿En qué año se hicieron más películas? Haga un gráfico de barras

- 4.6. **(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? Representelo usando un gráfico. ¿A qué género principal pertenecen las películas más largas?
- 4.7. **(8 puntos)** ¿Las películas de qué genero principal obtuvieron mayores ganancias?
- 4.8. **(3 puntos)** ¿La cantidad de actores influye en los ingresos de las películas? ¿se han hecho películas con más actores en los últimos años?
- 4.9. **(3 puntos)** ¿Es posible que la cantidad de hombres y mujeres en el reparto influya en la popularidad y los ingresos de las películas?
- 4.10. **(8 puntos)** ¿Quiénes son los directores que hicieron las 20 películas mejor calificadas?
- 4.11. **(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
- 4.12. **(5 puntos)** ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?
- 4.13. **(6 puntos)** ¿En qué meses se han visto los lanzamientos con mejores ingresos? ¿cuantas películas, en promedio, se han lanzado por mes?
- 4.14. **(7 puntos)** ¿Cómo se correlacionan las calificaciones con el éxito comercial?
- 4.15. **(5 puntos)** ¿Qué estrategias de marketing, como videos promocionales o páginas oficiales, generan mejores resultados?
- 4.16. **(4 puntos)** ¿La popularidad del elenco está directamente correlacionada con el éxito de taquilla?
5. **(¡10 puntos extras!)** Genere usted otras seis preguntas que le parezcan interesantes porque le permitan realizar otras exploraciones y respóndalas. No puede repetir ninguna de las instrucciones anteriores.

MATERIAL A ENTREGAR

- Vínculo del archivo de Google docs (se utilizará el historial para calificar) con el informe de análisis exploratorio que debería tener:
 - Enunciado de la pregunta que se está respondiendo.
 - Respuesta con su respectiva explicación
 - Gráfico, que le ayude a explicar la respuesta.
- Script de R o de Python que utilizó para responder las preguntas con el código utilizado
- Vinculo del controlador de versiones utilizado.

REFERENCIAS

Git en RStudio y R

- <https://www.youtube.com/watch?v=BAN4ErGF5ro>
- <https://rpubs.com/RonaldoAnticono/818156>

Git con Jupyter Notebook

- <https://deepnote.com/guides/versioning/how-to-use-git-with-jupyteripynb>
- <https://www.reviewnb.com/git-jupyter-notebook-ultimate-guide>