

## Hoja de Trabajo 1. Análisis Exploratorio

### INSTRUCCIONES:

Utilice el data set que se le adjunta en la tarea para resolver los ejercicios que se le presentan. Debe discutir cada uno de los hallazgos que haga al responder la pregunta. Genere un informe en pdf con las respuestas de cada uno de los ejercicios. Guarde el código que ha utilizado para resolver los ejercicios. Los lenguajes que tiene permitido usar para contestar las preguntas que se le hacen son R o Python. **La calificación de cada ejercicio tomará en cuenta tanto lo escrito en el informe como el código.**

### 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.
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? Represéntelo usando un gráfico
  - 4.7. **(8 puntos)** ¿Las películas de qué género 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. **(7 puntos)** ¿Se asocian ciertos meses de lanzamiento con mejores ingresos?
  - 4.13. **(8 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)** ¿A qué género principal pertenecen las películas más largas?
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, si aplica de acuerdo con la pregunta.
- Script de R o de Python que utilizó para responder las preguntas con el código utilizado
- Vinculo del controlador de versiones utilizado.