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 10866 películas obtenidos de la plataforma "The movie DB".

Variables:

- Id: Id de la película
- imdb_id: Id de la película en the movie data base (IMDB)
- popularity: Un índice de la popularidad de la película
- budget: El presupuesto para la película.
- Revenue: El ingreso de la película.
- original_title: El título original de la película.
- cast: Elenco de la película
- homepage: La página de inicio de la película
- director: Director de la película
- tagline: El eslogan de la película.
- keywords: Las palabras clave asociadas a la película.
- overview: Una breve trama de la película.
- runtime: La duración de la película.
- genres: El género de la película.
- production_companies: Las compañías productoras de la película.
- release_date: Fecha de lanzamiento de la película
- vote_count: El número de votos en la plataforma para la película.
- vote_average: El promedio de los votos en la plataforma para la película
- release_year: Año de lanzamiento 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 populares?
   - 4.7. **(8 puntos)** ¿Cuál es el género que predomina en el conjunto de datos? Represéntelo usando un gráfico
   - 4.8. **(3 puntos)** ¿Las películas de qué genero principal obtuvieron mayores ganancias?
   - 4.9. **(3 puntos)** ¿Las películas de qué genero principal necesitaron mayor presupuesto?
   - 4.10. **(8 puntos)** ¿Quiénes son los 20 mejores directores que hicieron películas altamente 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 máximos?
   - 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 tres preguntas que le parezcan interesantes porque le permitan realizar otras exploraciones y respóndalas. No puede repetir ninguna de las instrucciones anteriores.

MATERIAL A ENTREGAR

- Archivo .pdf 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