Ejercicios

1. **Carga del Dataset:**
    - Carga el archivo `movies2.csv` en un DataFrame de Spark y muestra los primeros 5 registros.
2. **Conteo de Registros:**
    - Calcula y muestra el número total de registros en el DataFrame.
3. **Películas por Año:**
    - Crea un nuevo DataFrame que contenga el número de películas lanzadas por año. Muestra los resultados ordenados por año en orden ascendente.
4. **Promedio de Popularidad por Año:**
    - Calcula el promedio de popularidad de las películas por año y muestra los resultados ordenados por año.
5. **Películas Más Populares:**
    - Encuentra y muestra las 10 películas con mayor popularidad.
6. **Promedio de Votos por Género:**
    - Explota la columna `genre_names` para que cada género tenga su propio registro y luego calcula el promedio de votos (`vote_average`) para cada género.
7. **Número de Películas por Género:**
    - Calcula el número de películas en cada género y muestra los resultados ordenados de mayor a menor.
8. **Películas con Más de 10000 Votos:**
    - Encuentra y muestra todas las películas que tienen más de 10,000 votos.
9. **Filtrar Películas por Año:**
    - Filtra y muestra todas las películas lanzadas después del año 2000.
10. **Películas con Título Más Largo:**
    - Encuentra y muestra las 5 películas con los títulos más largos (en términos de número de caracteres).
11. **Análisis de Sentimiento de Resumen:**
    - Realiza un análisis de sentimiento simple sobre la columna `overview` y muestra los 5 resúmenes más positivos y los 5 más negativos (esto requerirá un poco más de trabajo adicional con bibliotecas de procesamiento de texto).
12. **Distribución de la Popularidad:**
    - Crea un gráfico de la distribución de la popularidad de las películas.

Notas:

- Algunos ejercicios pueden requerir el uso de funciones avanzadas de Spark como `explode` para manejar la lista de géneros.
- Puedes usar la biblioteca de Spark SQL para escribir consultas SQL si prefieres ese enfoque.