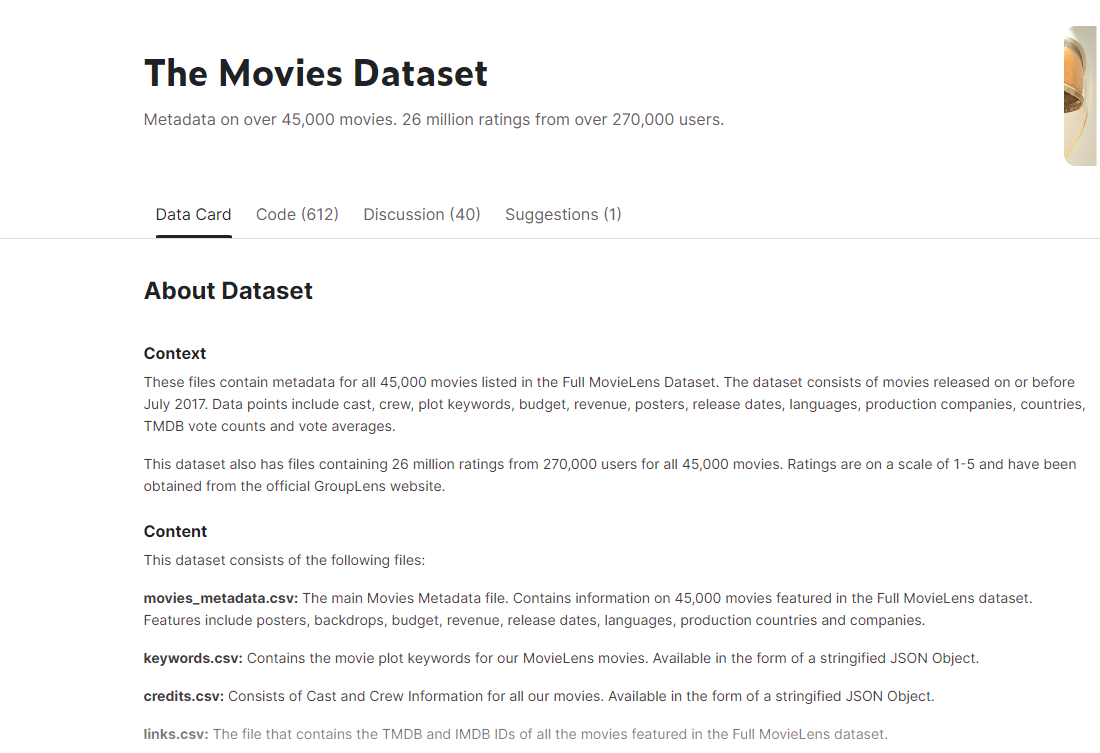
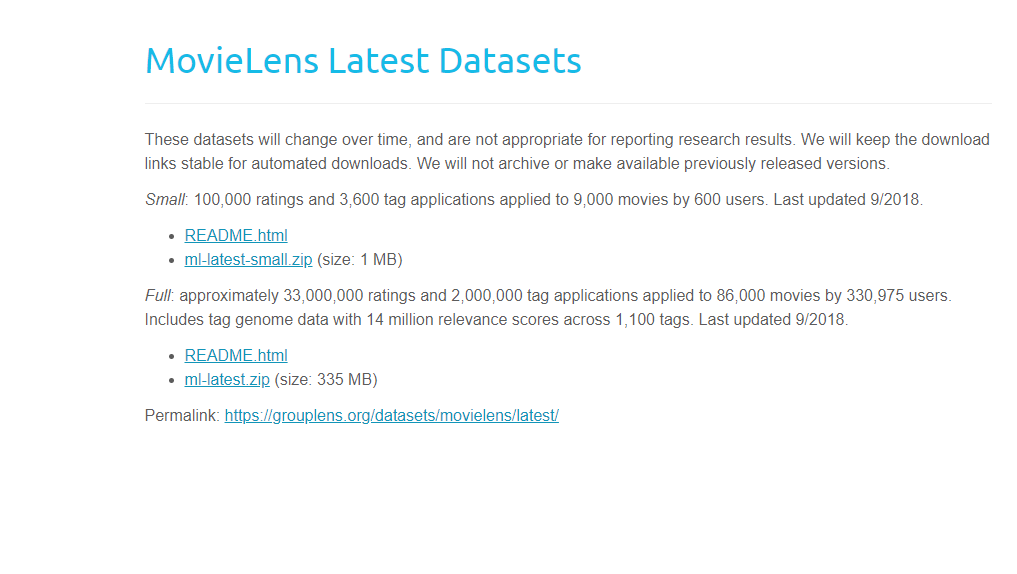
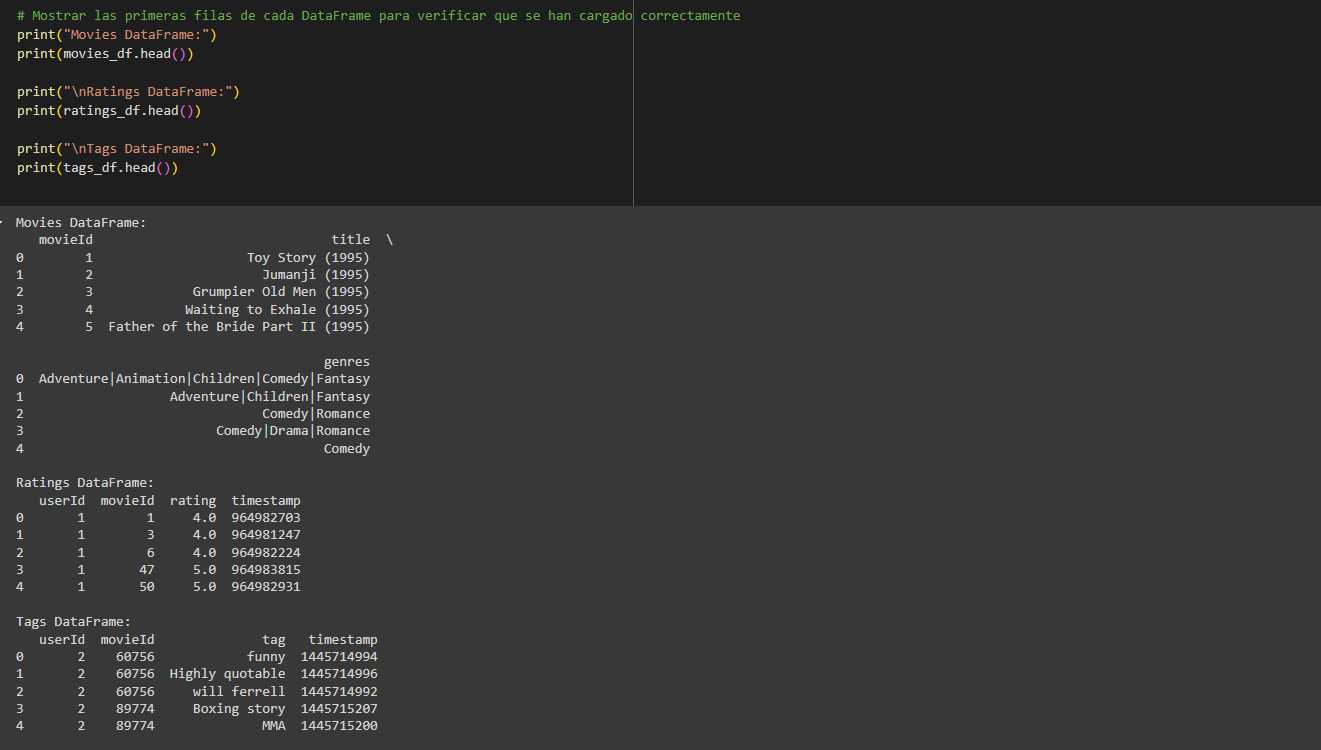
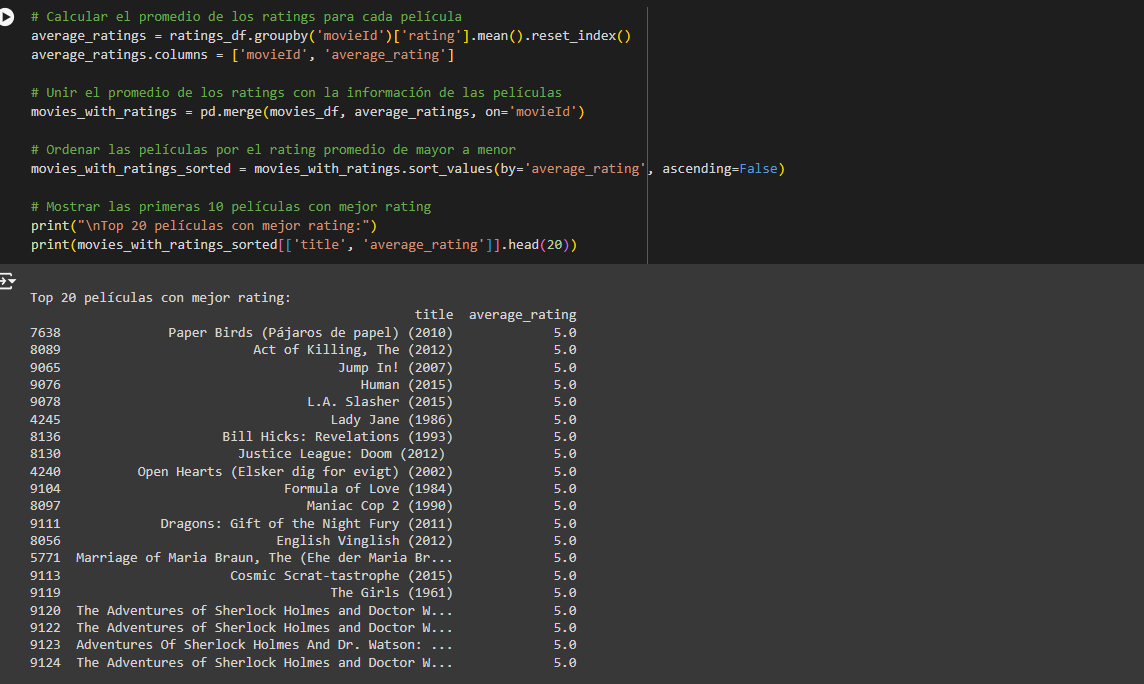
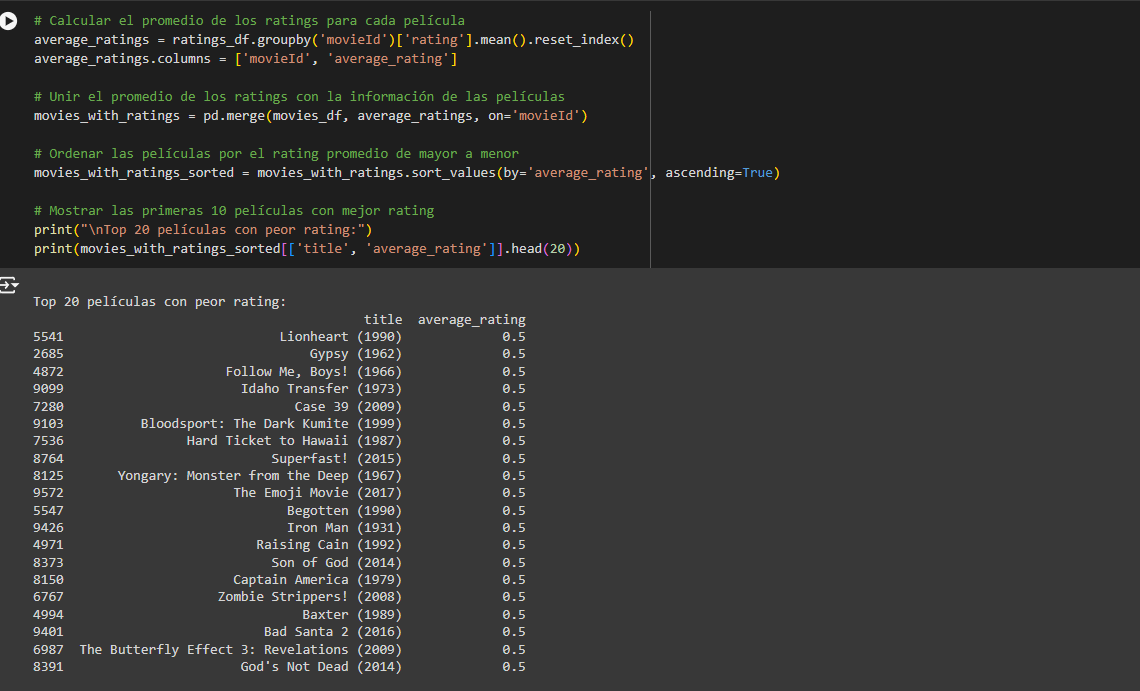
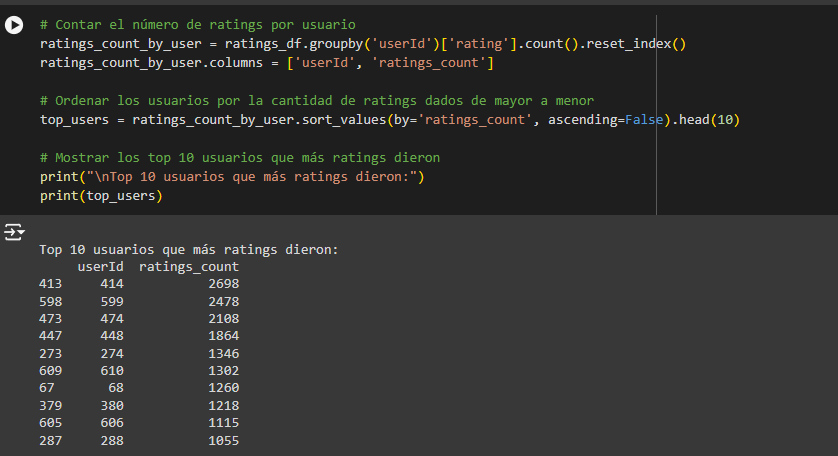
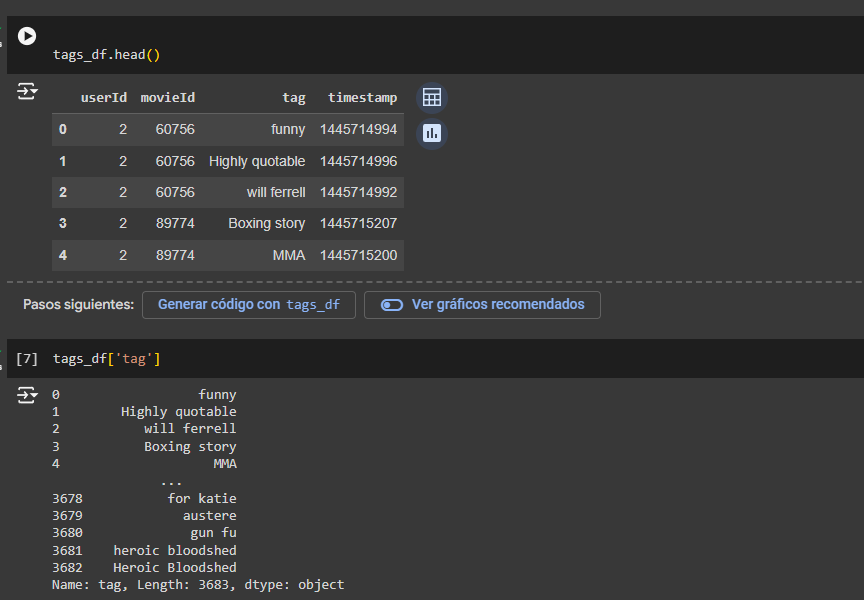
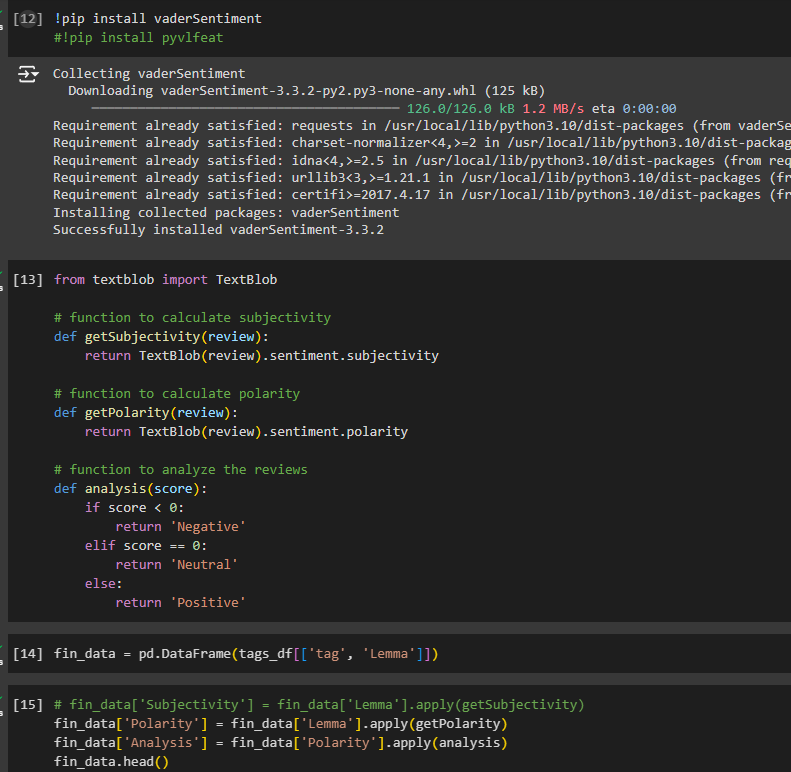
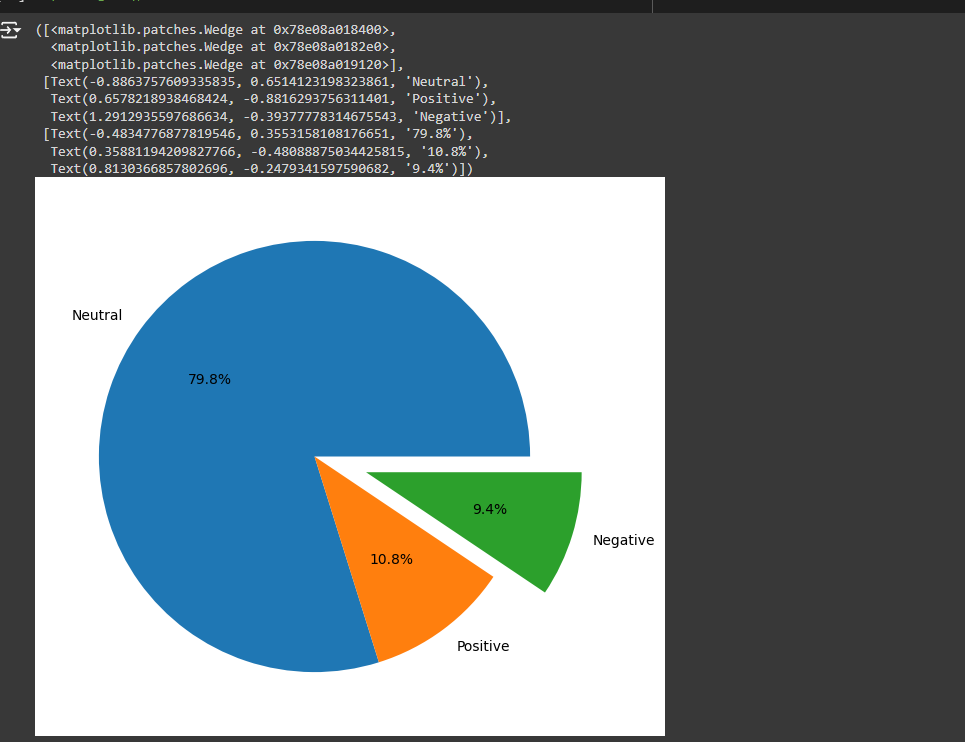
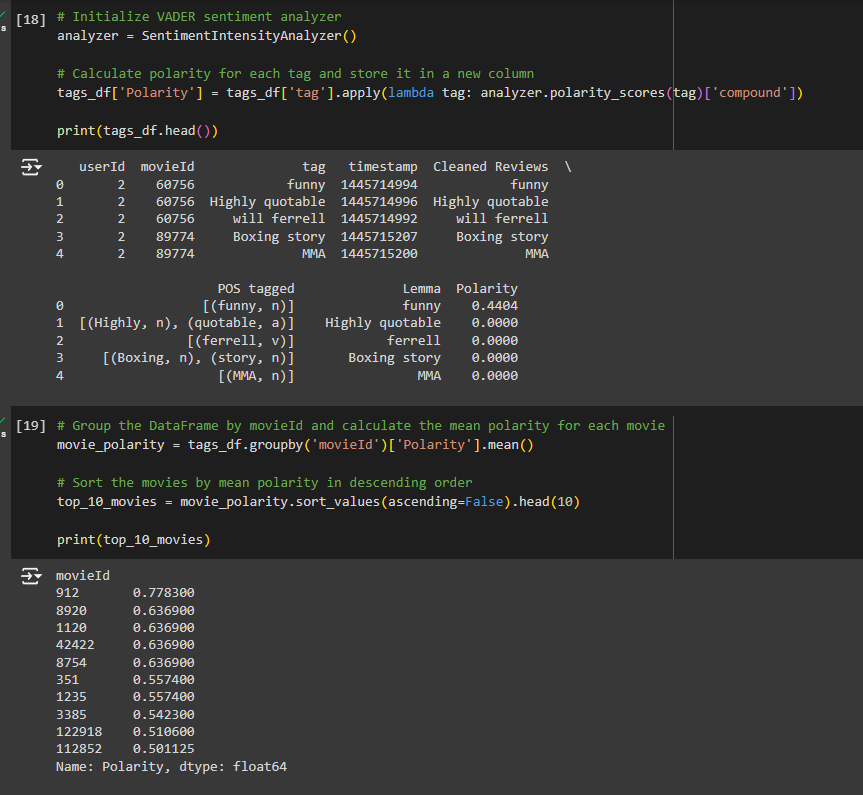
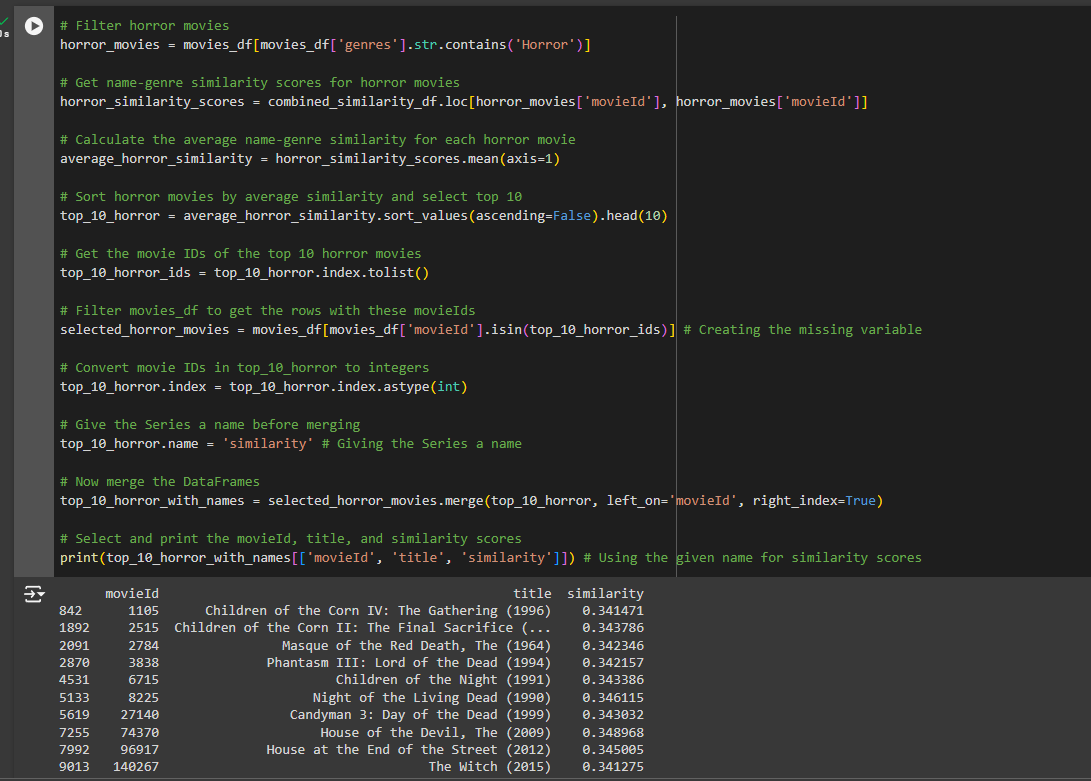
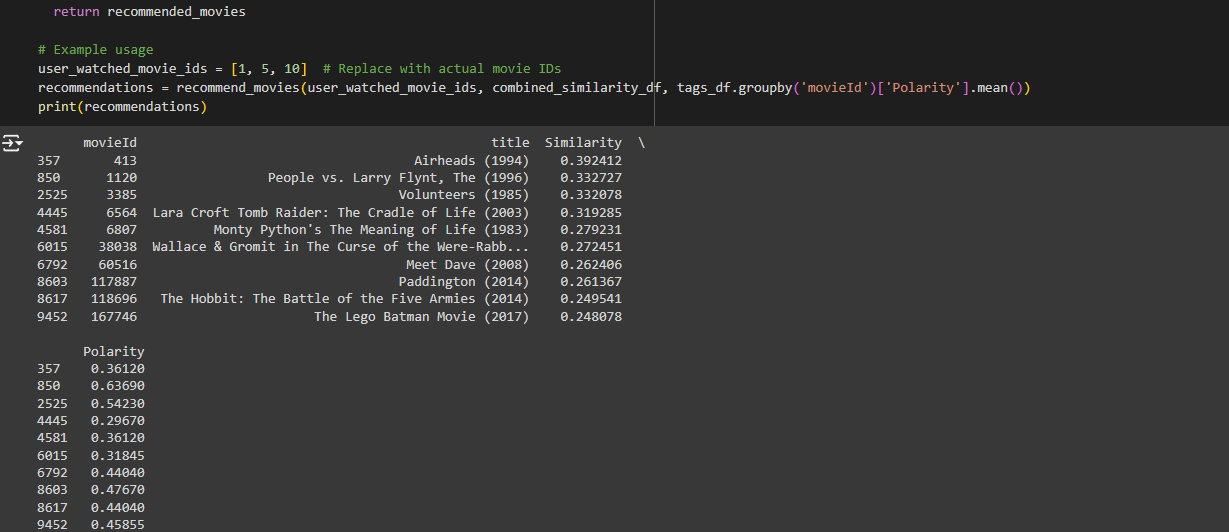
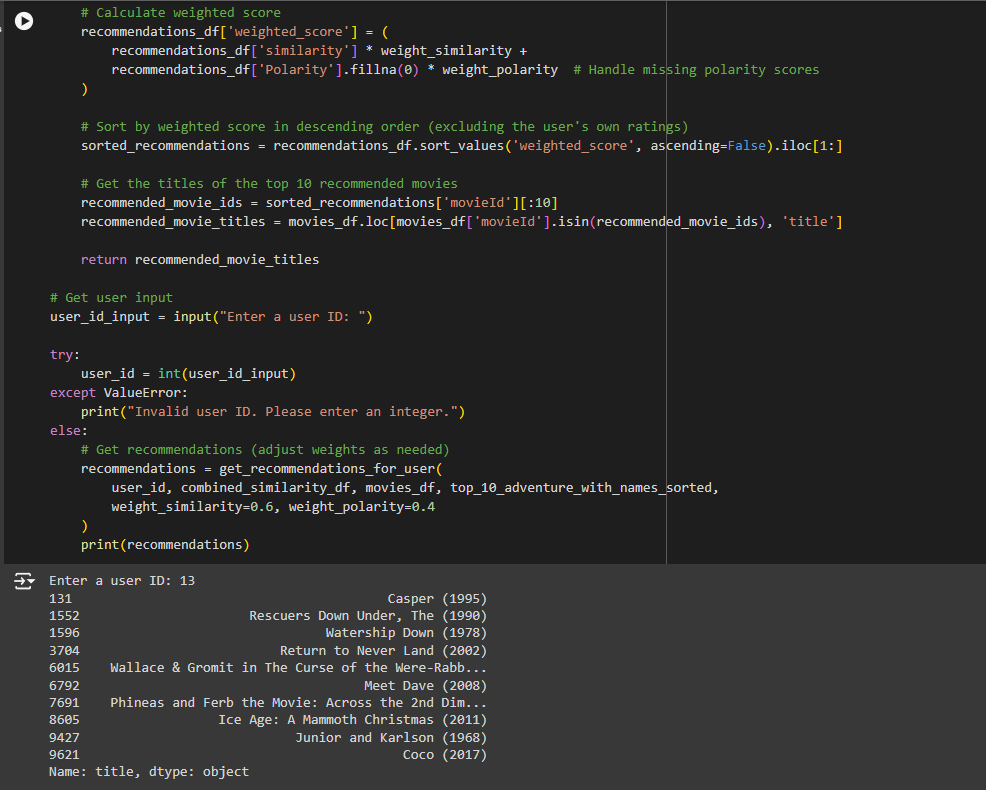
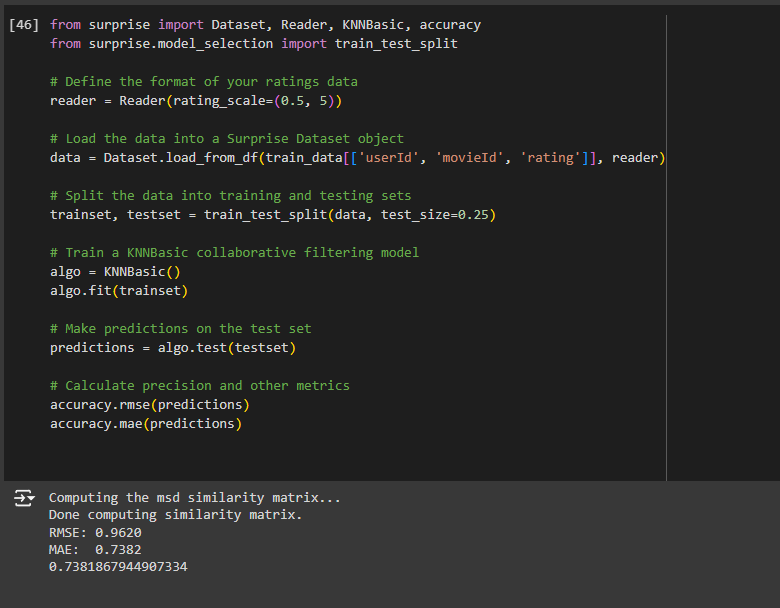
Redacta un informe claro y conciso que describe el proyecto, la metodología, los resultados y las

Conclusiones  
  
  
Para empezar el Proyecto decidí usar un dataset de Kaggle : The movies dataset

Link: https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset?select=ratings.csv  
  
  
Decidi usar solo un small sample para realizar el trabajo   
  
  
  
Una vez decidido que dataset usar era empezar a explorar los datos  
Primero Mire que tablas traía  
  
  
  
Despues mire las películas que mejor rating tenían  
  
Y las que peor rating tenían   
  
  
  
Aparte mire cuales son los usuarios que mas ratings dieron para darme una idea de que tan grande se podría ser el tamaño de películas por usuario  
  
  
  
Una vez terminada la exploración comenzó la limpieza y prepare los datos para análisis  
  
  
  
Debido a los ratings no tan realistas decidí usar los comentarios de las películas por los usuarios como guía para las recomendaciones  
  
Utilice las herramientas de sentiment análisis que era tarea para calcular la polaridad de los comentarios, en especifico use Vader por recomendaciones de la profesora de que es el mas preciso  
  
  
  
  
  
Una vez que hice la limpieza y prepare los datos, empecé con el calculo de similitud  
Decidí agrupar los películas en categorías con las cuales calculaba la similitud entre ellas lo cual me daba un valor de cuan similares eran las películas, probe con top 10 peliculas de horror   
  
  
  
Una vez ya calculada la similitud empecé con las recomendaciones,   
Le tire que el usuario miro las películas 1, 5, 10 lo cual toma como referencia para recomendarme las películas   
  
  
  
Senti que las recomendaciones tenían que ser mejores y le puse como parámetro que los valores de similarity tenían mas peso que la polaridad ya que muchas veces los usuarios tienen la polaridad en 0 porque esta se calcula en con los comentarios de los mismos usuarios y si no comentaron nunca este valor es nulo   
  
  
Para termine calcule la precisión del sistema de recomendación con ‘exactitud’ y me tiro que es 0.73 lo cual no esta mal pero podría ser mejor   
  
  
Conclusiones  
  
El Proyecto de Big Data estuvo bastante desafiante debido a que no querían salir las recomendaciones y complicaciones con el dataset por su gran tamaño, al final pudo recomendar películas según la similitud con películas que el usuario escribió una review positiva y el cálculo de similitud, si el mismo usuario coloco un review negativa en películas de terror lo mas probable que no le aparezcan y no se le recomienden más películas del género y si el escribió una review positiva en una de drama el top estaría plagado de este género, creo que sería mejor tomar más métricas para la recomendación de películas, como cuanto tiempo le prestó atención y cuantas emociones evocan las películas en el usuario, debido a la falta de estas métricas solamente se pueden analizar factores que son verbalizados por el usuario y que no siempre reflejan sus gustos específicos, creo que en un futuro con nuevas tecnologías un mejor sistema de recomendaciones será encontrado que refleje los gustos específicos del usuario.