Exercice 1 :

I built a new notebook based on "apache.access.log" dataset in databricks then I published the results you can find it in the link below:

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/1225248769674766/452581562825308/5550312523085441/latest.html

Exercice 2:

Task 1: User-based collaborative filtering
The band Coldplay receives much worse result from user 101 instead of receiving a value near to 2286. It receives 1067.

the reason may be due to the reduce the number of neighbours used for predictions, which could lead to less accurate recommendations. Fewer neighbours might not capture the full diversity of user preferences and could result in suboptimal suggestions.

setting k=10 may help to capture the full diversity of user preferences and could result in suboptimal suggestions.

Determining the proper value for k depends on various factors, such as the size of the dataset, the density of user ratings, and the specific application requirements. A small k may result in personalized recommendations but could be sensitive to noise in the data. On the other hand, a larger k might provide more diverse recommendations but could dilute the personalization aspect.

Task 2: Item-based collaborative filtering

Setting a threshold instead of choosing a fixed value for neighbours' selection in item-based collaborative filtering can also have implications on the recommendation process.

In item-based collaborative filtering, recommendations are generated based on the similarity between items rather than users. The selection process can be influenced by setting a threshold for item similarity instead of choosing a fixed number of neighbors:

- Using a threshold allows for flexible neighbor selection based on similarity. Instead of limiting the number of neighbors, the threshold determines which items are considered similar enough to be included in the recommendation process. This flexibility can be beneficial in capturing items with varying degrees of similarity.

- Sparse Item Relationship Handling: In some cases, the item-item similarity matrix can be sparse, especially if certain items have limited ratings or attributes in common. By using a threshold, items with low similarity to the target item can be filtered out, focusing only on the most relevant and similar items for recommendations.
- Impact on Recommendation Diversity.