DA512 Homework 2
Due: March 10th, 2021

Processing "Movielens Dataset" using Spark DataFrames / Spark SQL

Extract following informations from movielens dataset.
There are several versions of movielens dataset you can use the smallest one (ml-latest-small.zip)
https://grouplens.org/datasets/movielens/latest/

1. For each tag (not genre), find average ratings of movies and sort them according to average rating. (Sample results below are not correct !)

Tag	**Average Rating**
funny	3.8
pixar	3.5

2. Find top 10 (sorted by their similarity) most similar users for each user.
Similarity: You can use any similarity measure like Cosine, Manhattan, Euclidean ...etc. (or you can implement your own similarity measure)

Please create one python file for each question as following.

YourName-YourLastname-Question1.ipynb
YourName-YourLastname-Question2.ipynb

(Please do not use Turkish characters in your filenames)