Assignment 2 - MAP REDUCE
Fundamental of Big Data Spring 2023

Due Date:28th March 2023

Submission Details: *You must submit Python code and Jupyter notebook. No form of plagiarism will be tolerated.*

Question (5 marks each)

You are provided dataset "Movies.csv" that contains information about 1600 movies with properties such as year, length, leading actor and actress, director, and popularity. Your task is to write an efficient MAP REDUCE code to solve the following problems. Write combiner as well.

1. Find the total number of movies of each genre. Consider only the movies released after 1970 and have a length greater than 75 minutes.
2. For each actress, find the movie she acted in. You must print the movie's names and the movie's year.
3. Find the 25 most popular movies released during 1950-1990
4. Find the five most popular movies of each actor.
5. Find the average length of the movies of each genre.
6. Find the lead actor and lead actress pair who has acted in more than one movie together.
7. Find the names of directors who directed movies of both the 'Mystery' and 'Drama' Genre.
8. Find each director, find the average, max and min ranking of his movies.

Question 2 (10 marks)

In this example, you will use the given "Movies.csv" data file. Your task is to write an **efficient** MAP-Reduce code for finding movies (released after 1970) that might be similar or related to a movie made in the 1960's (1960-1969).

To keep things simple, for comparing two movies, we will use the following fields: Genre, director, lead actor, and lead actress.

So, we can say two movies are similar if they have the same genre, director, and same lead actor **or** actress.