Big Data Analytics Assignment 2

Instructions:

- Submit your Java code (MapReduce) and outputs on Google Classroom, clearly mentioning the question/part number against your answers. There should be a report showing the results of your analysis.
- Deadline to submit this assignment is Friday, 3 June 2022 11:59 pm. There will be no extension for this assignment.
- Enter your roll numbers (only numeric part) and other information (written in square brackets []) in the data files (.csv) where needed.

Dataset 1: US_AirlLines.txt //placed inside assignment folder

Dataset 2: Literacy_and_Literacy_Rate.csv //placed inside assignment folder

Task 1:

Run word count example (shared in the demo), dataset for the task is dataset 1. Take a positive number N>0 and find N top trending words in the corpus used above in task 1. The user will enter the value of N at runtime.
    (15)

Task 2:

Use data set 2 to perform the following tasks in a distributed environment   (45)

1. Count the total number of male literates and female literates in all locations.
2. Find the average literacy rate for both males and females separately.
3. Find the actual number of literate and illiterate men and women and their total. (For example, if there are 200 men and 200 women, and 60% men are literate and 55% women are literate, it means 120 men are literate 80 are illiterate while 110 women are literate and 90 women are illiterate. In total there are 230 literate people and 170 illiterate people)