MIS381N (Fall 2019) Homework4

For Homework4, you will use the MRJob library in Python to streamline the implementation of mapper and reduce functions for the Top10 MapReduce task.

The format of your program should be similar to the numerous examples that we have seen in the class to-date.

from mrjob.job import MRJob

#import any other files that you may need

class MRAverage(MRJob):

def mapper(self, \_, line):

# mapper code goes here

yield

def reducer(self, month, temp):

#reducer code goes here

yield

if \_\_name\_\_ == '\_\_main\_\_':

MRAverage.run()

Your input data will be provided to you in three different files bsalaries1.txt, bsalaries2.txt and bsalaries3.txt. Your code should output the overall top10 values across all the three files

What to submit:

1. Your source code Top10MapMR<uteid>.py

2. Screenshot of the stdout window which shows the answers of your program in a Word Document.

When I ran the program, the top10 salaries were

80431, 66551, 66111, 56781, 48201 47651, 31911, 28431, 27831, 16931