APPLICATION OF MAPREDUCE

Programming Paradigm for Processing

**Students Alcohol Consumption**

*DATASET 1: student-mat.csv*

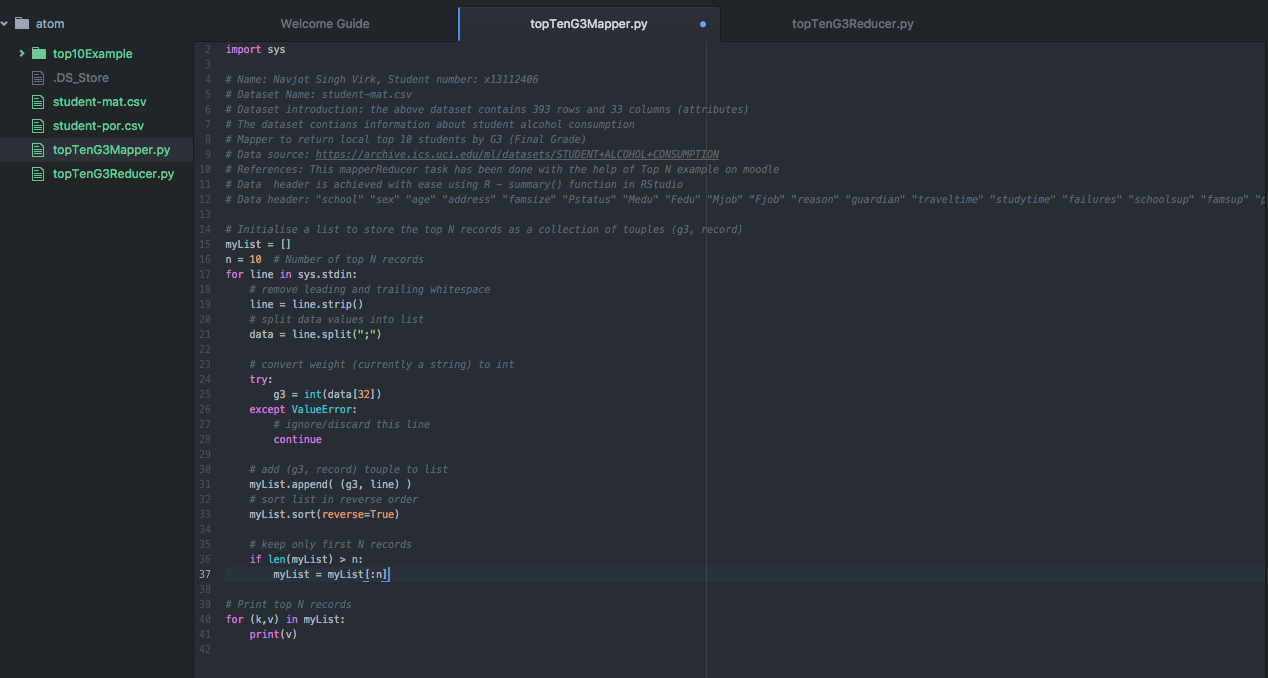
(Mathematics Students)

--------------------------------------------------------------------------------------------------------------------------

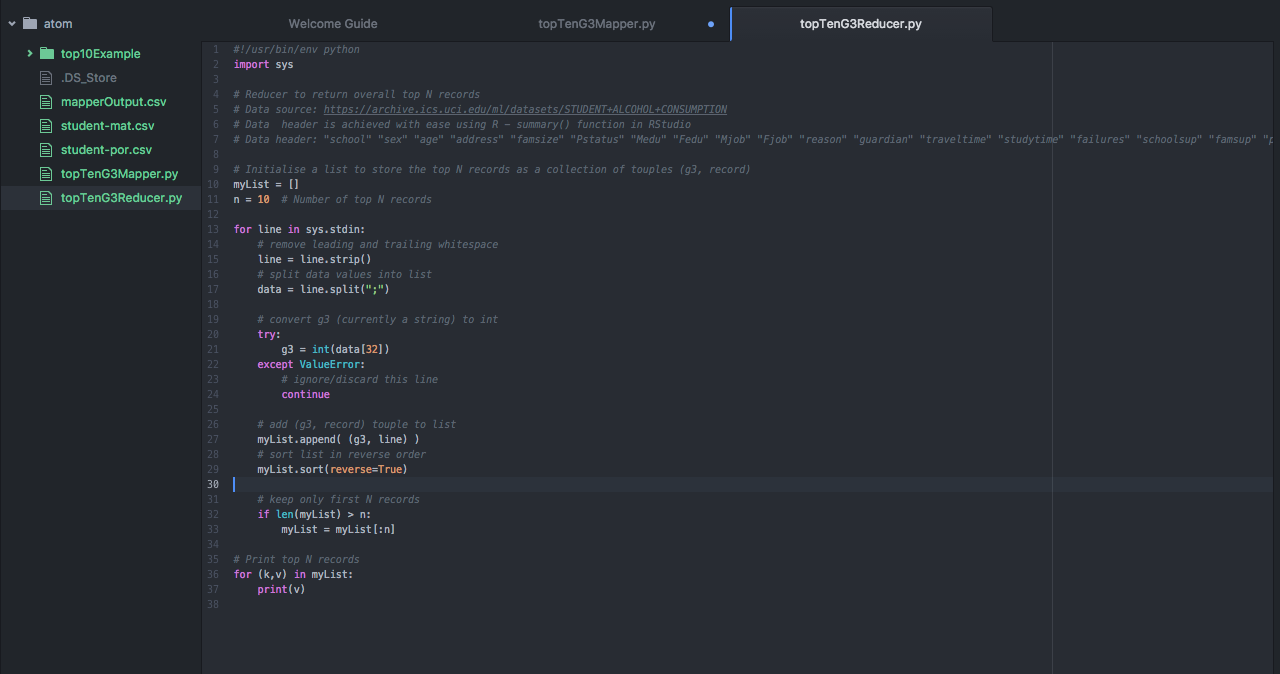
**Implement and Present the algorithms to process the dataset.**

The idea is to mapReduce and get top 10 student records with best G3 (Final Grade) and output a csv file and then use that csv to draw some meaningful plots/predictions.

Mapper (topTenG3Mapper.py)



Reducer (topTenG3Reducer.py)



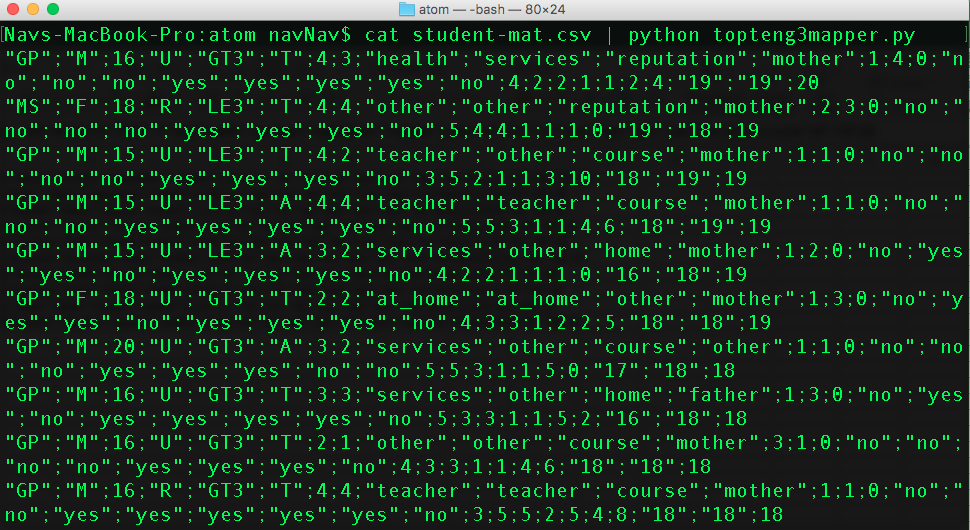
Reference: I am using Top N example available on moodle for this mapReduce processing

**Configuration Details**

For mapReduce processing, I used Atom editor for editing my .py files, and used mac terminal to run mapper and reducer on csv files and output csv result files and then I took final results csv file and used it in RStudio to plot some graphs.

**Present and Discuss your results**

**1 – Run mapper on the dataset**

****

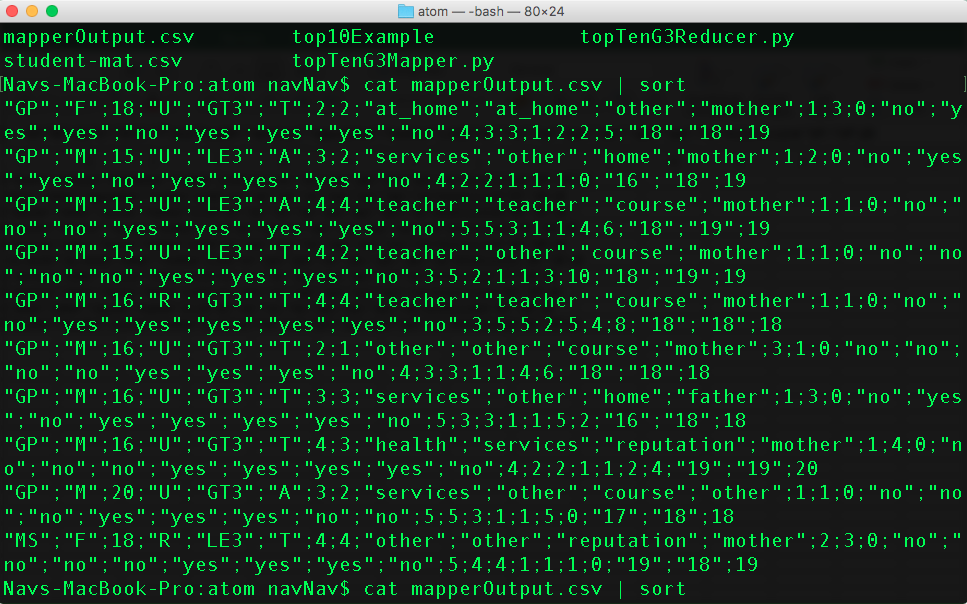
cat student-mat.csv | topteng3mapper.py

**2 – Outputting the mapper’s result into a csv file**

****

cat student-mat.csv | topteng3mapper.py > mapperOutput.csv

**3– Sorting**

****

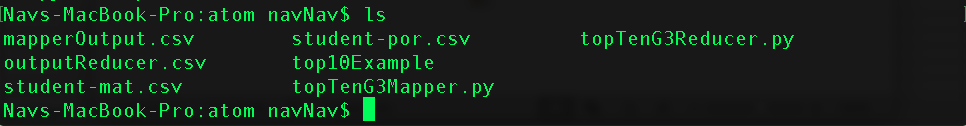
cat mapperOutput.csv | sort

**4 – Run reducer on the mapper results csv file**

****

cat mapperoutput.csv | sort | python topteng3reducer.py

**5 – Output the mapper results**

****

cat mapperoutput.csv | sort | python topteng3reducer.py > outputReducer.csv