

APPLICATION OF MAPREDUCE

Programming Paradigm for Processing

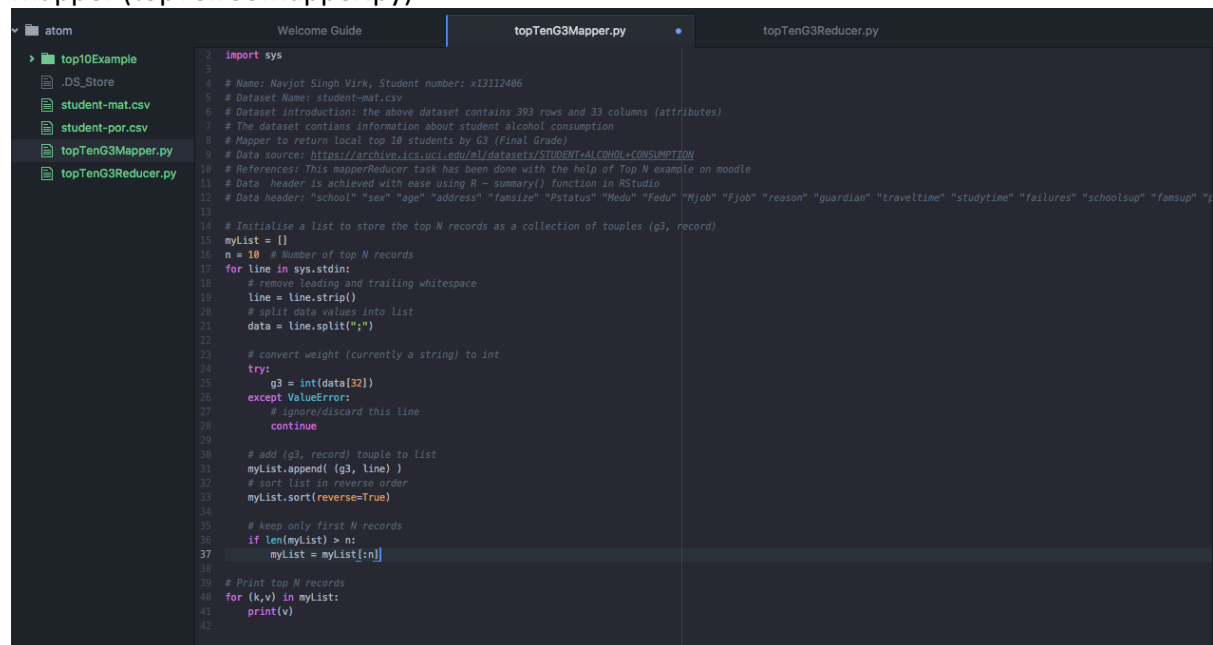
Students Alcohol Consumption

DATASET 1: student-mat.csv
(Mathematics Students)

Implement and Present the algorithms to process the dataset.

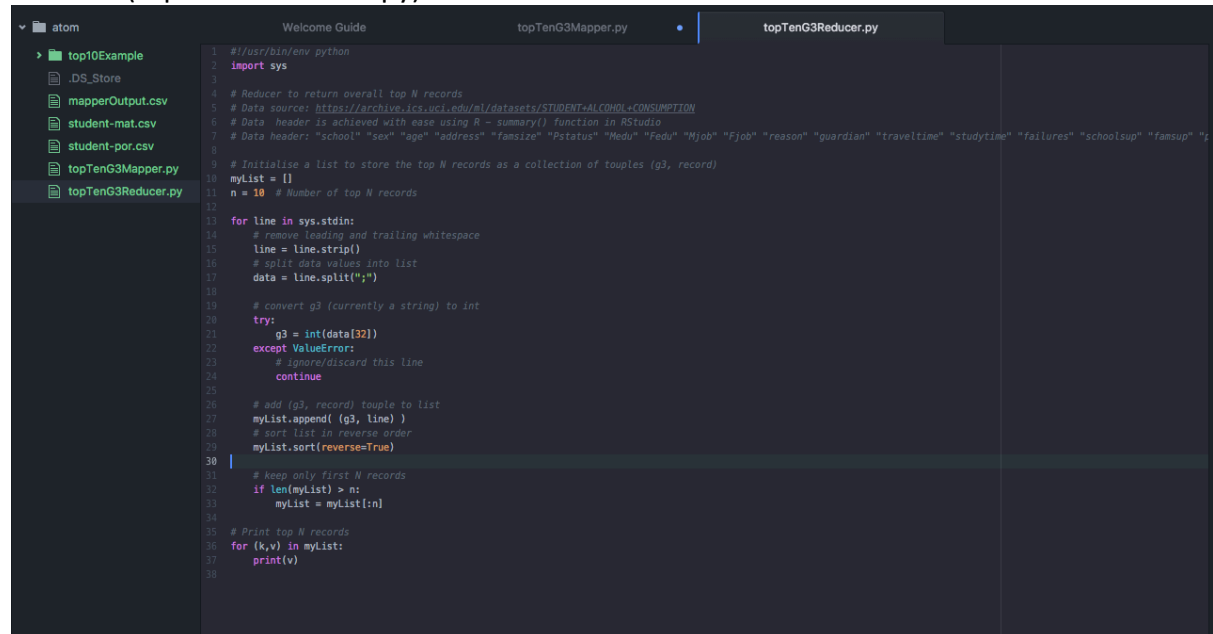
The idea is to mapReduce and get top 10 student records with best G3 (Final Grade) and output a csv file and then use that csv to draw some meaningful plots/predictions.

Mapper (topTenG3Mapper.py)



```
1 import sys
2
3 # Name: Navjot Singh Virk, Student number: x13112406
4 # Dataset Name: student-mat.csv
5 # Dataset introduction: the above dataset contains 393 rows and 33 columns (attributes)
6 # The dataset contains information about student alcohol consumption
7 # Mapper to return local top 10 students by G3 (Final Grade)
8 # Data source: https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION
9 # References: This mapperReducer task has been done with the help of Top N example on moodle
10 # Data header is achieved with ease using R ~ summary() function in RStudio
11 # Data header: "school" "sex" "age" "address" "famsize" "Pstatus" "Redu" "Fedu" "Rjob" "Fjob" "reason" "guardian" "traveltime" "studytime" "failures" "schoolsup" "famsup" "lunch" "paid"
12
13
14 # Initialise a list to store the top N records as a collection of tuples (g3, record)
15 myList = []
16 n = 10 # Number of top N records
17 for line in sys.stdin:
18     # remove leading and trailing whitespace
19     line = line.strip()
20     # split data values into list
21     data = line.split(";")
22     # convert weight (currently a string) to int
23     try:
24         g3 = int(data[32])
25     except ValueError:
26         # ignore/discard this line
27         continue
28     # add (g3, record) tuple to list
29     myList.append( (g3, line) )
30     # sort list in reverse order
31     myList.sort(reverse=True)
32     # keep only first N records
33     if len(myList) > n:
34         myList = myList[:n]
35
36 # Print top N records
37 for (k,v) in myList:
38     print(v)
```

Reducer (topTenG3Reducer.py)



```
1 #!/usr/bin/env python
2 import sys
3
4 # Reducer to return overall top N records
5 # Data source: https://archive.ics.wisc.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION
6 # Data header is achieved with ease using R - summary() function in RStudio
7 # Data header: "school" "sex" "age" "address" "fansize" "Pstatus" "Medu" "Fedu" "Mjob" "Fjob" "reason" "guardian" "traveltime" "studytime" "failures" "schoolsup" "famsup" "g3"
8
9 # Initialize a list to store the top N records as a collection of tuples (g3, record)
10 myList = []
11 n = 10 # Number of top N records
12
13 for line in sys.stdin:
14     # remove leading and trailing whitespace
15     line = line.strip()
16     # split data values into list
17     data = line.split(";")
18
19     # convert g3 (currently a string) to int
20     try:
21         g3 = int(data[32])
22     except ValueError:
23         # ignore/discard this line
24         continue
25
26     # add (g3, record) tuple to list
27     myList.append( (g3, line) )
28     # sort list in reverse order
29     myList.sort(reverse=True)
30
31     # keep only first N records
32     if len(myList) > n:
33         myList = myList[:n]
34
35 # Print top N records
36 for (k,v) in myList:
37     print(v)
```

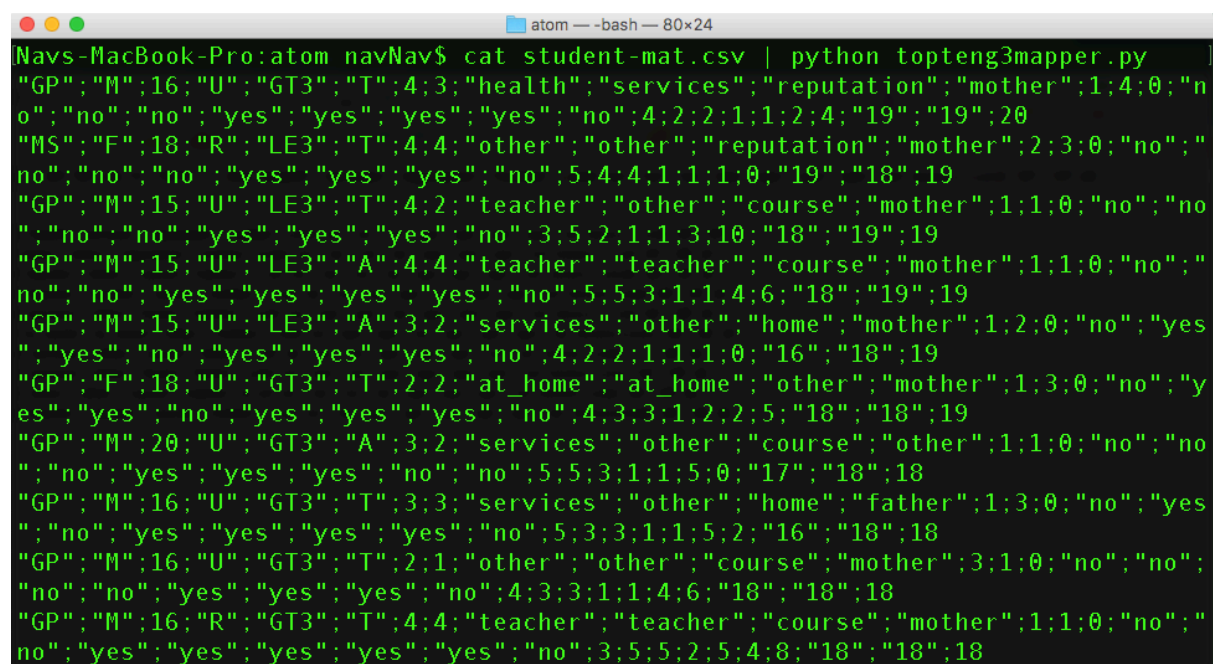
Reference: I am using Top N example available on moodle for this mapReduce processing

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



```
Navs-MacBook-Pro:atom navNav$ cat student-mat.csv | python topteng3mapper.py
"GP";"M";16;"U";"GT3";"T";4;3;"health";"services";"reputation";"mother";1;4;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";4;2;2;1;1;2;4;"19";"19";20
"MS";"F";18;"R";"LE3";"T";4;4;"other";"other";"reputation";"mother";2;3;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";5;4;4;1;1;1;0;"19";"18";19
"GP";"M";15;"U";"LE3";"T";4;2;"teacher";"other";"course";"mother";1;1;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";3;5;2;1;1;3;10;"18";"19";19
"GP";"M";15;"U";"LE3";"A";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";5;5;3;1;1;4;6;"18";"19";19
"GP";"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"yes";"no";4;2;2;1;1;1;0;"16";"18";19
"GP";"F";18;"U";"GT3";"T";2;2;"at_home";"at_home";"other";"mother";1;3;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;3;3;1;2;2;5;"18";"18";19
"GP";"M";20;"U";"GT3";"A";3;2;"services";"other";"course";"other";1;1;0;"no";"no";"no";"yes";"yes";"yes";"no";"no";5;5;3;1;1;5;0;"17";"18";18
"GP";"M";16;"U";"GT3";"T";3;3;"services";"other";"home";"father";1;3;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;3;3;1;1;5;2;"16";"18";18
"GP";"M";16;"U";"GT3";"T";2;1;"other";"other";"course";"mother";3;1;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;3;3;1;1;4;6;"18";"18";18
"GP";"M";16;"R";"GT3";"T";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"no";"yes";"yes";"yes";"yes";"yes";"no";3;5;5;2;5;4;8;"18";"18";18
```

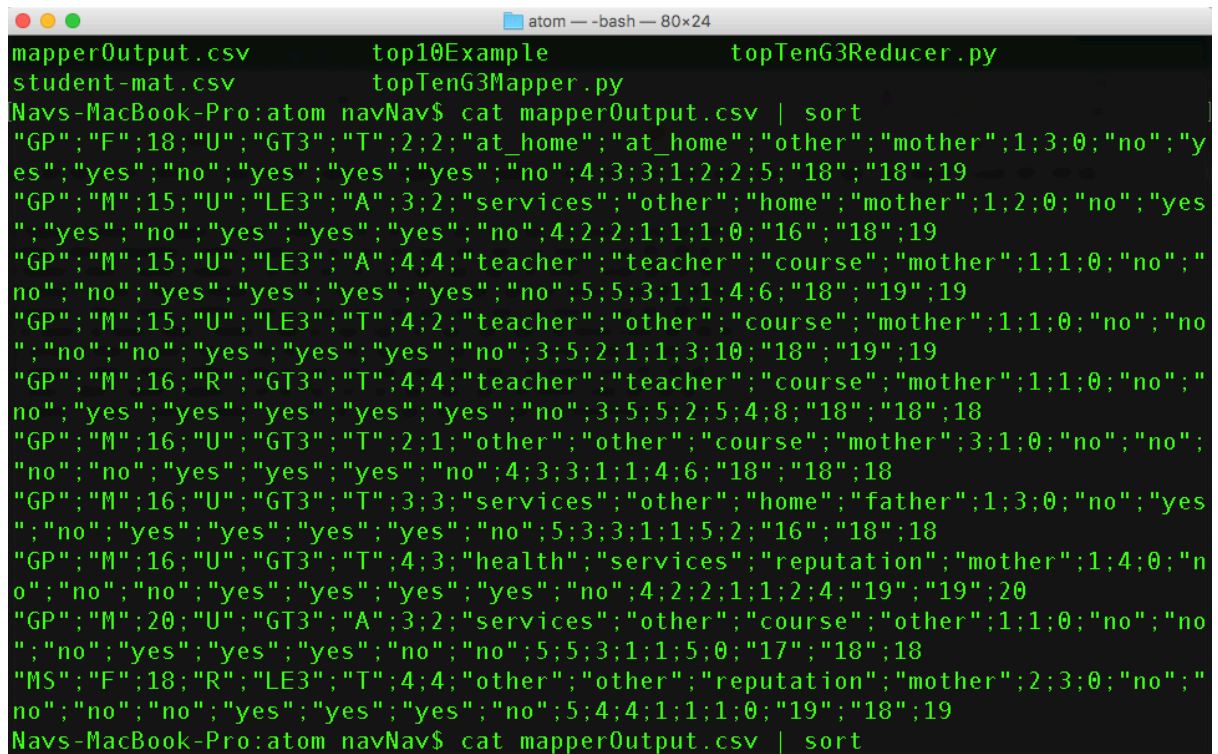
```
cat student-mat.csv |  
topteng3mapper.py
```

2 – Outputting the mapper's result into a csv file

```
Navs-MacBook-Pro:atom navNav$ cat student-mat.csv | python topteng3mapper.py > mapperOutput.csv
```

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

3– Sorting



```
atom — -bash — 80x24  
mapperOutput.csv      top10Example          topTenG3Reducer.py  
student-mat.csv       topTenG3Mapper.py  
Navs-MacBook-Pro:atom navNav$ cat mapperOutput.csv | sort  
"GP";"F";18;"U";"GT3";"T";2;2;"at_home";"at_home";"other";"mother";1;3;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;3;3;1;2;2;5;"18";"18";19  
"GP";"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;2;2;1;1;1;0;"16";"18";19  
"GP";"M";15;"U";"LE3";"A";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";5;5;3;1;1;4;6;"18";"19";19  
"GP";"M";15;"U";"LE3";"T";4;2;"teacher";"other";"course";"mother";1;1;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";3;5;2;1;1;3;10;"18";"19";19  
"GP";"M";16;"R";"GT3";"T";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"yes";"yes";"yes";"yes";"yes";"no";3;5;5;2;5;4;8;"18";"18";18  
"GP";"M";16;"U";"GT3";"T";2;1;"other";"other";"course";"mother";3;1;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;3;3;1;1;4;6;"18";"18";18  
"GP";"M";16;"U";"GT3";"T";3;3;"services";"other";"home";"father";1;3;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;3;3;1;1;5;2;"16";"18";18  
"GP";"M";16;"U";"GT3";"T";4;3;"health";"services";"reputation";"mother";1;4;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";4;2;2;1;1;2;4;"19";"19";20  
"GP";"M";20;"U";"GT3";"A";3;2;"services";"other";"course";"other";1;1;0;"no";"no";"no";"yes";"yes";"yes";"no";"no";5;5;3;1;1;5;0;"17";"18";18  
"MS";"F";18;"R";"LE3";"T";4;4;"other";"other";"reputation";"mother";2;3;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";5;4;4;1;1;1;0;"19";"18";19  
Navs-MacBook-Pro:atom navNav$ cat mapperOutput.csv | sort
```

```
cat mapperOutput.csv | sort
```

4 – Run reducer on the mapper results csv file

```
Navs-MacBook-Pro:atom navNav$ cat mapperoutput.csv | sort | python topteng3reducer.py
"GP";"M";16;"U";"GT3";"T";4;3;"health";"services";"reputation";"mother";1;4;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";4;2;2;1;1;2;4;"19";"19";20
"MS";"F";18;"R";"LE3";"T";4;4;"other";"other";"reputation";"mother";2;3;0;"no";"no";"no";"yes";"yes";"yes";"no";5;4;4;1;1;1;0;"19";"18";19
"GP";"M";15;"U";"LE3";"T";4;2;"teacher";"other";"course";"mother";1;1;0;"no";"no";"no";"yes";"yes";"yes";"no";3;5;2;1;1;3;10;"18";"19";19
"GP";"M";15;"U";"LE3";"A";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"no";"yes";"yes";"yes";"yes";"no";5;5;3;1;1;4;6;"18";"19";19
"GP";"M";15;"U";"LE3";"A";3;2;"services";"other";"home";"mother";1;2;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;2;2;1;1;1;0;"16";"18";19
"GP";"F";18;"U";"GT3";"T";2;2;"at_home";"at_home";"other";"mother";1;3;0;"no";"yes";"yes";"no";"yes";"yes";"yes";"no";4;3;3;1;2;2;5;"18";"18";19
"GP";"M";20;"U";"GT3";"A";3;2;"services";"other";"course";"other";1;1;0;"no";"no";"no";"yes";"yes";"yes";"no";"no";5;5;3;1;1;5;0;"17";"18";18
"GP";"M";16;"U";"GT3";"T";3;3;"services";"other";"home";"father";1;3;0;"no";"yes";"no";"yes";"yes";"yes";"yes";"no";5;3;3;1;1;5;2;"16";"18";18
"GP";"M";16;"U";"GT3";"T";2;1;"other";"other";"course";"mother";3;1;0;"no";"no";"no";"no";"yes";"yes";"yes";"no";4;3;3;1;1;4;6;"18";"18";18
"GP";"M";16;"R";"GT3";"T";4;4;"teacher";"teacher";"course";"mother";1;1;0;"no";"no";"yes";"yes";"yes";"yes";"yes";"no";3;5;5;2;5;4;8;"18";"18";18
```

```
cat mapperoutput.csv | sort | python
topteng3reducer.py
```

5 – Output the mapper results

```
Navs-MacBook-Pro:atom navNav$ ls
mapperOutput.csv      student-por.csv      topTenG3Reducer.py
outputReducer.csv     top10Example
student-mat.csv       topTenG3Mapper.py
Navs-MacBook-Pro:atom navNav$
```

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