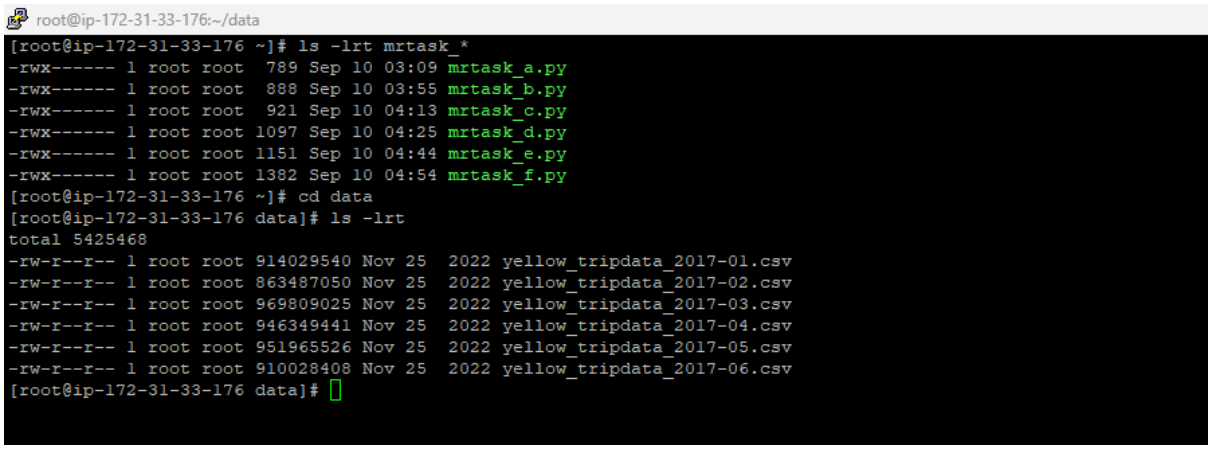
MapReduce Programing Assignment

Taks 4:

Python codes for 6 MapReduce Tasks created in EMR cluster. Yellow taxi data files downloaded and saved in data directory

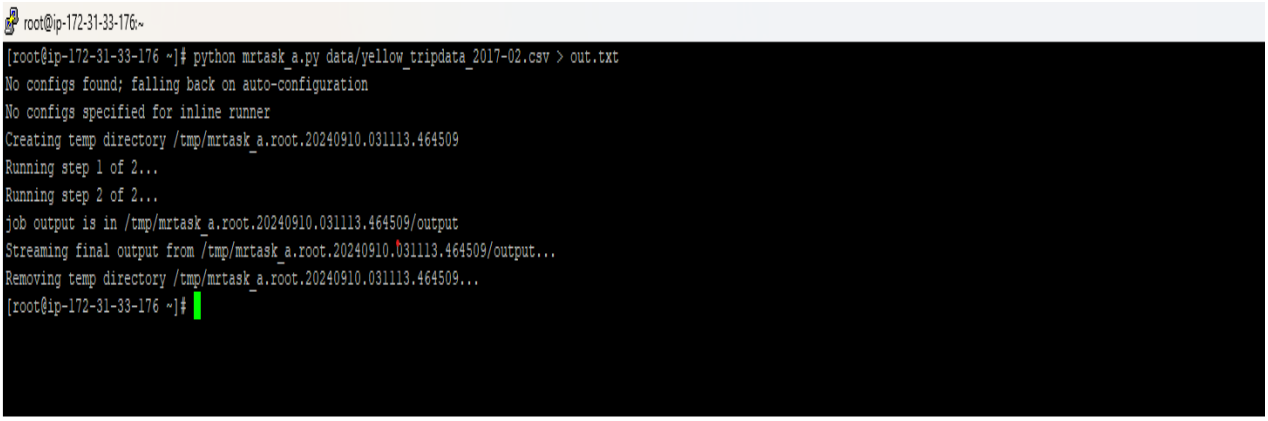


MapReduce tasks:

a. Which vendors have the most trips, and what is the total revenue generated by that vendor?

Mr job ran with file: yellow_tripdata_2017-02.csv.

Input:



Output:

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# ls -lrta out.txt
-rw-r--r-- 1 root root 22 Sep 10 03:13 out.txt
[root@ip-172-31-33-176 ~]# cat out.txt
"2"     78947447.68633662
[root@ip-172-31-33-176 ~]#
```

b. Which pickup location generates the most revenue?

Mr job ran with file: yellow_tripdata_2017-01.csv.

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# python mrtask_b.py data/yellow_tripdata_2017-01.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_b.root.20240910.035725.478439
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_b.root.20240910.035725.478439/output
Streaming final output from /tmp/mrtask_b.root.20240910.035725.478439/output...
Removing temp directory /tmp/mrtask_b.root.20240910.035725.478439...
[root@ip-172-31-33-176 ~]#
```

Output:

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# ls -lrt out.txt
-rw-r--r-- 1 root root 25 Sep 10 04:00 out.txt
[root@ip-172-31-33-176 ~]# cat out.txt
"132"   12660835.900003651
[root@ip-172-31-33-176 ~]#
```

c. What are the different payment types used by customers and their count? The final results should be in a sorted format.

Mr job ran with file: yellow_tripdata_2017-03.csv.

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# python mrtask_c.py data/yellow_tripdata_2017-03.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_c.root.20240910.041502.043637
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_c.root.20240910.041502.043637/output
Streaming final output from /tmp/mrtask_c.root.20240910.041502.043637/output...
Removing temp directory /tmp/mrtask_c.root.20240910.041502.043637...
[root@ip-172-31-33-176 ~]#
```

Output:

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# ls -lrt out.txt
-rw-r--r-- 1 root root 44 Sep 10 04:17 out.txt
[root@ip-172-31-33-176 ~]# cat out.txt
"1"     6994699
"2"     3231928
"3"     53815
"4"     14999
[root@ip-172-31-33-176 ~]#
```

d. What is the average trip time for different pickup locations?

Mr job ran with file: yellow_tripdata_2017-04.csv.

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# python mrtask_d.py data/yellow_tripdata_2017-04.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_d.root.20240910.042817.257046
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_d.root.20240910.042817.257046/output
Streaming final output from /tmp/mrtask_d.root.20240910.042817.257046/output...
Removing temp directory /tmp/mrtask_d.root.20240910.042817.257046...
[root@ip-172-31-33-176 ~]#
```

Output:

```
[root@ip-172-31-33-176 ~]# ls -lrt out.txt
-rw-r--r-- 1 root root 6065 Sep 10 04:31 out.txt
[root@ip-172-31-33-176 ~]# cat out.txt
"27"    25.29
"132"   15.933471289671525
"219"   15.55504690431519
"2"     14.982
"10"    14.962619305856842
"215"   14.868115501519737
"15"    13.109230769230768
"252"   11.790638297872341
"253"   11.610869565217387
"38"    11.373333333333333
"222"   11.172142857142859
"124"   11.056530612244897
"84"    10.776666666666666
"130"   10.610158924205377
"131"   10.601702127659573
"150"   10.47304347826087
"251"   10.397777777777778
"138"   9.849661122254648
"64"    9.667142857142855
"93"    9.637846534653459
"216"   9.113794162826423
"28"    9.002168874172192
"154"   8.817916666666667
"214"   8.795
"108"   8.65122448979592
"187"   8.3675
"221"   8.243157894736841
"70"    7.732855321861075
"203"   7.418125000000001
"240"   7.310645161290322
"176"   7.216666666666668
"117"   6.790833333333333
"101"   6.774848484848484
"259"   6.70779661016949
"208"   6.616338028169014
"99"    6.455
"31"    6.230731707317074
"194"   6.130171919770781
"57"    6.09
"184"   5.98375
"172"   5.887142857142856
"175"   5.861785714285714
"46"    5.695
"23"    5.685714285714285
"19"    5.608208955223882
```

root@ip-172-31-33-176:~

```
"189"   3.043386761650342
"168"   3.033163064833009
"167"   3.0105603448275862
"256"   3.0044359153291495
"69"    2.9899124726477018
"45"    2.98752998567847
"18"    2.982013422818792
"148"   2.978166283099227
"146"   2.956686853629
"235"   2.9525480769230756
"179"   2.9468698234099078
"105"   2.9433333333333334
"94"    2.9331999999999994
"37"    2.918742120343846
"231"   2.912988370635068
"62"    2.912773437499999
"34"    2.8912542372881362
"230"   2.8457236522871368
"152"   2.809593204341676
"49"    2.7977285992217884
"225"   2.7678237547892732
"159"   2.731895652173913
"7"     2.720229717615411
"136"   2.7162307692307692
"1"     2.7113372093023234
"166"   2.6929982572449758
"123"   2.687629629629628
"147"   2.683253012048194
"6"     2.681666666666667
"153"   2.6807339449541283
"145"   2.678940172291154
"264"   2.668597370869361
"4"     2.6590435613016803
"17"    2.654888670595942
"125"   2.636868486190723
"233"   2.5983989356322126
"242"   2.5755084745762717
"156"   2.551648745519714
"144"   2.549765053148823
"217"   2.542276785714283
"50"    2.524389251301712
"158"   2.496257010234165
"24"    2.4870291388572427
"79"    2.477318359993631
"211"   2.4741678513828225
"163"   2.469839950438182
"48"    2.4668400795820236
"74"    2.458514635806649
"162"   2.4482450055243072
"140"   2.4419126993495714
"114"   2.4401271977011203
"164"   2.427975650762789
"26"    2.4015873015873006
"58"    2.396666666666667
"118"   2.3951315789473684
"161"   2.3925397824842465
"42"    2.3841592134099376
"262"   2.378896222831806
"41"    2.3759898039004157
"224"   2.3753273739247747
"246"   2.3592965839026965
"68"    2.355144286004005
"151"   2.3480967888254045
```

e. Calculate the average tips to revenue ratio of the drivers for different pickup locations in sorted format.

Mr job ran with file: yellow_tripdata_2017-05.csv

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# python mrtask_e.py data/yellow_tripdata_2017-05.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_e.root.20240910.044555.325555
Running step 1 of 1...
job output is in /tmp/mrtask_e.root.20240910.044555.325555/output
Streaming final output from /tmp/mrtask_e.root.20240910.044555.325555/output...
Removing temp directory /tmp/mrtask_e.root.20240910.044555.325555...
[root@ip-172-31-33-176 ~]#
```

Output:

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# ls -lrt out.txt
-rw-r--r-- 1 root root 6609 Sep 10 04:49 out.txt
[root@ip-172-31-33-176 ~]# cat out.txt
"1"     0.11961918264336044
"10"    0.10313149532724888
"100"   0.10054811858030266
"101"   0.1538484771604267
"102"   0.09686867923894887
"105"   0.07175925925925926
"106"   0.11203391656570175
"107"   0.11929564967810695
"108"   0.12120884627098828
"109"   0.23737785016286642
"11"    0.058159884648828326
"111"   0.077659754354142
"112"   0.10900187462682069
"113"   0.11769752457713911
"114"   0.11571668143567702
"115"   0.1014877461571702
"116"   0.09140144834181262
"117"   0.029865191166550988
"118"   0.03324247058041353
"119"   0.06439098561727422
"12"    0.09033344274559964
"120"   0.14485426649668087
"121"   0.060253867786983323
"122"   0.035274937009041045
"123"   0.0750228876958997
"124"   0.09886020334918795
"125"   0.1223467893609015
"126"   0.05493403420192623
"127"   0.08833111068775
"128"   0.10511320192841678
"129"   0.06647065796213625
"13"    0.12390141102889671
"130"   0.1073558396219219
"131"   0.10384807204137783
"132"   0.1025660513460536
"133"   0.09147281347725217
"134"   0.10314438578934823
"135"   0.0785122214655781
"136"   0.032600578121018825
"137"   0.11253077758026406
"138"   0.1304834446631976
"139"   0.025278845473624637
"14"    0.09666329734829485
"140"   0.1110704470853594
"141"   0.11186974122790275
"142"   0.1131558975317282
"143"   0.11473162975887714
"144"   0.1105272721674073
"145"   0.09866671202718473
"146"   0.09229805528093973
"147"   0.04671430427439263
"148"   0.11347366987186253
"149"   0.12259087582337157
"15"    0.12542820971105154
"150"   0.10633375877472878
"151"   0.10794854174655912
"152"   0.08497213797829244
"153"   0.08409909578769388
"154"   0.07400473993394348
"155"   0.09130948944530835
```

```
"42"    0.07096965920546744
"43"    0.10407835112043434
"44"    0.07601977750309023
"45"    0.0983846147461719
"46"    0.12534064929126656
"47"    0.03340712045748017
"48"    0.10582758808855866
"49"    0.09681013151351817
"5"     0.711159737417943
"50"    0.10923919835836489
"51"    0.054687879641958155
"52"    0.13055804576734864
"53"    0.09939808622286242
"54"    0.11968717470048443
"55"    0.10775881095465412
"56"    0.09187656958656626
"57"    0.0982393882217696
"58"    0.023456558453743663
"59"    0.022958827634333566
"6"     0.04668850781897014
"60"    0.028830566964685292
"61"    0.09614549547751258
"62"    0.07257887627997188
"63"    0.07430008293013912
"64"    0.07777953179990528
"65"    0.1064885388809956
"66"    0.11474873960246652
"67"    0.080161361573226 79
"68"    0.1143179862016906
"69"    0.03888023995673477
"7"     0.08055330999297121
"70"    0.10558719618955761
"71"    0.057126398156593525
"72"    0.12979482386561036
"73"    0.06660273315751619
"74"    0.07309332094094213
"75"    0.09233854474158283
"76"    0.15699972823625333
"77"    0.03758990781571329
"78"    0.03340662069771519
"79"    0.11726023735794265
"8"     0.1128890978373831
"80"    0.11230199593086634
"81"    0.1314140628199353
"82"    0.0640252588354194
"83"    0.060647576771211555
"84"    0.0006641576267434137
"85"    0.0568739741017691
"86"    0.03836325514349365
"87"    0.12371478220911764
"88"    0.11535120965220538
"89"    0.08902658805206363
"9"     0.07875511622519114
"90"    0.11768850776024835
"91"    0.10308391768201007
"92"    0.056615381413745704
"93"    0.1071444344921823
"94"    0.038084346952212626
"95"    0.07928530356245143
"96"    0.09975598325818912
"97"    0.09868461512594295
"98"    0.06067388225420083
[root@ip-172-31-33-176 ~]#
```

f. How does revenue vary over time? Calculate the average trip revenue per month - analysing it by hour of the day (day vs night) and the day of the week (weekday vs weekend).

Mr job ran with file: yellow_tripdata_2017-02.csv

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# python mrtask_f.py data/yellow_tripdata_2017-02.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_f.root.20240910.054923.262334
Running step 1 of 1...
job output is in /tmp/mrtask_f.root.20240910.054923.262334/output
Streaming final output from /tmp/mrtask_f.root.20240910.054923.262334/output...
Removing temp directory /tmp/mrtask_f.root.20240910.054923.262334...
[root@ip-172-31-33-176 ~]#
```

Output:

```
root@ip-172-31-33-176:~
[root@ip-172-31-33-176 ~]# cat out.txt
[5, 0, 0]       20.67580628717184
[5, 0, 1]       20.222927816211197
[5, 0, 2]       17.475337946373955
[5, 0, 3]       16.500869375273254
[5, 0, 4]       17.72242385276702
[5, 0, 5]       17.214722279682466
[5, 0, 6]       15.535144889630047
[5, 1, 0]       21.031834002678092
[5, 1, 1]       21.433126875268695
[5, 1, 2]       15.673668375307125
[5, 1, 3]       16.01943694562365
[5, 1, 4]       17.82680830972634
[5, 1, 5]       16.98350939914392
[5, 1, 6]       14.909383101000635
[5, 10, 0]      15.729515147273062
[5, 10, 1]      16.458481953969176
[5, 10, 2]      16.25188047366245
[5, 10, 3]      16.69905101654024
[5, 10, 4]      15.561541607588083
[5, 10, 5]      14.090127353264979
[5, 10, 6]      15.917149462802318
[5, 11, 0]      16.501411351423833
[5, 11, 1]      17.057743057749477
[5, 11, 2]      16.678389769014256
[5, 11, 3]      16.971931701028566
[5, 11, 4]      15.936172298292535
[5, 11, 5]      14.186428367821614
[5, 11, 6]      15.179233698515443
[5, 12, 0]      16.674605377274567
[5, 12, 1]      17.57923643338651
[5, 12, 2]      16.922496229258133
[5, 12, 3]      17.642270375552705
[5, 12, 4]      17.397580914423866
[5, 12, 5]      14.46158955914949
[5, 12, 6]      15.622687109023158
[5, 13, 0]      16.996052168240148
[5, 13, 1]      17.32060245376047
[5, 13, 2]      17.4352844470701
[5, 13, 3]      17.80117183683161
[5, 13, 4]      18.1914882620306 95
[5, 13, 5]      14.938204334362416
[5, 13, 6]      15.724775792508087
[5, 14, 0]      16.934448932832197
[5, 14, 1]      17.59816293838024
[5, 14, 2]      18.1194581446035 38
[5, 14, 3]      18.37123266909594
[5, 14, 4]      17.836074949399933
[5, 14, 5]      15.152086137986176
[5, 15, 0]      16.44313240042794
[5, 15, 1]      16.995949565214467
[5, 15, 2]      17.367478529711896
[5, 15, 3]      17.72055704429827
[5, 15, 4]      18.213417171354976
[5, 15, 5]      15.153918250452595
[5, 16, 0]      17.48639502942853
[5, 16, 1]      18.755100234311122
[5, 16, 2]      18.927257542403005
[5, 16, 3]      18.71860042121102
[5, 16, 4]      19.279787511504303
[5, 16, 5]      15.294088397787366
[5, 17, 0]      16.819343818701725
[5, 17, 1]      17.772129760992215
```