MapReduce Tasks:

Task 4. Using MapReduce to perform data analysis on files downloaded to the EMR Instance

- ❖ Initially I set up the environment using commands sudo yum update, pip3 install mrjob etc,onces set up I proceeded
- ❖ I wrote python scripts, moved them to hadoop using WinSCP, gave python scripts execution permission using command chmod 700
- General code used for running the MapReduce jobs: python 'mr_script.py' 'input' >> 'output file'
- I used yellow_tripdata_2017-05.csv dataset for this task
 `wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-05.csv`
- **a.** Which vendors have the most trips, and what is the total revenue generated by that vendor?

Vendor 2 has the most trips, total revenue generated is 92896777.5452

```
[hadoop@ip-172-31-37-211 ~]$ python mrtask_a.py yellow_tripdata_2017-05.csv >> mrtask_a.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_a.hadoop.20241203.120703.945648
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_a.hadoop.20241203.120703.945648/output
Streaming final output from /tmp/mrtask_a.hadoop.20241203.120703.945648/output...
Removing temp directory /tmp/mrtask_a.hadoop.20241203.120703.945648...
[hadoop@ip-172-31-37-211 ~]$
[hadoop@ip-172-31-37-211 ~]$ cat mrtask_a.txt
"2" 92896777.54522054
```

b. Which pickup location generates the most revenue?

Pickup location 132 generates the most revenue: 14040591.2200

```
[hadoop@ip-172-31-37-211 ~]$ python mrtask_b.py yellow_tripdata_2017-05.csv >> mrtask_b.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_b.hadoop.20241203.123155.249088
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_b.hadoop.20241203.123155.249088/output
Streaming final output from /tmp/mrtask_b.hadoop.20241203.123155.249088/output...
Removing temp directory /tmp/mrtask_b.hadoop.20241203.123155.249088...
[hadoop@ip-172-31-37-211 ~]$ cat mrtask_b.txt
"132" 14040591.220016211
[hadoop@ip-172-31-37-211 ~]$
```

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

The different payment types used by customers and their count are shown below

Numeric code	Payment type	Count
1	Credit card	6780947
2	Cash	3250362
3	No charge	55027
4	Dispute	15791

```
[hadoop@ip-172-31-37-211 ~]$ python mrtask_c.py yellow_tripdata_2017-05.csv >> mrtask_c.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_c.hadoop.20241203.123814.195617
Running step 1 of 3...
Running step 2 of 3...
Running step 3 of 3...
gob output is in /tmp/mrtask_c.hadoop.20241203.123814.195617/output
Streaming final output from /tmp/mrtask_c.hadoop.20241203.123814.195617/output...
Removing temp directory /tmp/mrtask_c.hadoop.20241203.123814.195617...
[hadoop@ip-172-31-37-211 ~]$
[hadoop@ip-172-31-37-211 ~]$ cat mrtask_c.txt
"1" 6780947
"2" 3250362
"3" 55027
"4" 15791
[hadoop@ip-172-31-37-211 ~]$
```

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

The average trip time for different pickup locations and code used is shown below:

```
Inadoop@ip-172-31-37-211 -j$ python mrtask d.py yellow_tripdata_2017-05.csv >> mrtask_d.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_d.hadoop.20241203.130018.853355
Running step 1 of 1...

job output is in /tmp/mrtask_d.hadoop.20241203.130018.853355/output

Removing temp directory /tmp/mrtask_d.hadoop.20241203.130018.853355/output

Removing temp directory /tmp/mrtask_d.hadoop.20241203.130018.853355...

Removing temp directory /tmp/mrtask_d.had
```

```
17.629476052716164
15.57929292929293
14.331578947368426
20.914534001214356
129"
             46.805229740819875
17.532022849462365
16.221703296703296
15.368173258003768
134"
             41.79266819772832
18.82820512820513
18.910311918063318
14.544783434852864
'142"
'143"
             14.724678434186814
13.87881130774534
17.422410667367306
13.919851029794044
144"
'147"
'148"
149"
             13.134482758620688
14.868589743589745
             12.776791277258567
22.9303030303030333
154"
             23.84722222222218
             17.7777777777778
24.96510297482838
             17.056993321982915
12.977884615384616
             11.23930817610063
17.381399631675876
'162"
'163"
             15.5755374379993
17.007285407402943
16.135913909640667
164"
            23.146643518518516
'169"
             12.008282828282828
"17"
"170"
             15.23280819184737
             11.867179487179486
             12.247398589065257
             20.955774278215223
             8.291562500000001
15.1499705882<u>353</u>
             12.653225806451612
27.183506944444446
180"
             13.486574074074074
             13.586842105263157
             22.571194503171245
             18.036302681992343
             13.806195462478186
             0.49374999999999999
201"
```

```
19.28479532163743
'205"
'206"
            8.778993536472758
            16.197798742138364
18.476602564102567
17.805909853173578
            24.37777777777773
20.40786163522013
213"
214"
215"
            8.220833333333333
56.68466165413534
            34.080833333333333
           17.684177215189873
53.02001915708813
22.21085626911315
11.665315315315317
219"
            19.39444444444442
17.158481387478847
14.423218777619141
223"
           17.062662041086035
15.537459546925568
18.118900966183574
13.729634370452187
            18.16775177926668
17.842905702290793
            17.80712673224909
'234"
'235"
'236"
            21.51443736730361
13.552552898225786
            13.635731925509624
'239"
'24"
           14.086371798958675
16.91818181818182
            15.555228758169934
            17.764382566585954
            17.97972746905131
23.33055555555555
246"
247"
            25.680555555555554
            18.58691486586808
            15.882166210462294
16.62408789386401
256"
            12.558972943698416
            57.221774193548384
17.91923076923077
            12.903932584269663
            16.217423442449842
            105.5841666666668
            17.3040625
```

```
15.943539005698556
         29.257407407407406
15.23495145631068
         20.58723958333333
14.719335511982568
54"
         47.3187265917603
23.806329113924047
59"
        7.7902777777777
12.961559139784944
15.251830481496228
13.879747675962818
         65.07303921568626
         11.29107142857143
         13.053431372549017
         13.984743997175142
         26.37702202480104
         12.455172413793102
         22.723684210526315
         12.525644805036702
         13.519075992588938
         18.701871980676327
         16.11434108527132
         15.650715220659471
         22.969314641744553
"88
         16.722509667955716
'81"
         10.0166666666667
82"
         15.163276362823948
         16.827437223042836
84"
         5.575
         15.723731138545956
         14.700000000000001
         153.69927536231884
90"
         14.588625022901905
'91"
         23.28269720101781
92"
94"
         11.647380952380951
'96"
         18.58888888888888
"98"
         13.002500000000001
[hadoop@ip-172-31-37-211 ~]$
```

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

The average tips to revenue ratio of the drivers for different pickup locations in sorted format are shown below:

```
nadoop@ip-172-31-37-211 ~]$ python mrtask_e.py yellow_tripdata_2017
To configs found; falling back on auto-configuration
To configs specified for inline runner
Treating temp directory /tmp/mrtask_e.hadoop.20241203.131912.198301
                                                                                                                                                         llow_tripdata_2017-05.csv >> mrtask_e.txt
120" 0.14485426649668087
121" 0.06025386778698323
122" 0.035274937009041045
123" 0.0750228876958997
                  0.09886020334918795
0.1223467893609015
0.05493403420192623
                  0.08833111068775
0.10511320192841678
"130" 0.1073558396219219

"131" 0.1073558396219219

"131" 0.10384807204137783

"132" 0.1025660513460536

"133" 0.09147281347725217
                    0.10314438578934823
"134" 0.10314438578934823

"135" 0.0785122214655781

"136" 0.032600578121018825

"137" 0.11253077758026406

"138" 0.1304834446631976

"139" 0.025278845473624637

"14" 0.09666329734829485
 1141" 0.1110704470853594

1141" 0.1110704470853594

1141" 0.11186974122790275

1142" 0.1131558975317282

1143" 0.11473162975887714

1144" 0.1105272721674073
 144 0.0922360332603337

147" 0.04671430427439263

1148" 0.11347366987186253

1149" 0.12259087582337157
"15" 0.12542820971105154
"150" 0.10633375877472878
"151" 0.10794854174655912
"152" 0.08497213797829244
"153" 0.08409909578769388
"154" 0.07400473993394348
"155" 0.07400473993394383
"156" 0.07496251874062969
"157" 0.11154334990834797
"158" 0.11254334990834797
"158" 0.1296825635625094
"160" 0.09369565599255213
 160" 0.09369565599255213
                  0.11101476605573193
0.11138449806107008
```

```
0.0356042774025466
0.050410277807923884
0.11765956298165817
0.08304806476027186
'167"
'168"
169"
                  0.11737627919224174
0.04416689447785676
'171"
'172"
                 0.0662888212150767
0.09292588232864547
0.08816570904350345
0.19982277359326542
'173"
'174"
'176"
'177"
                 0.03941897087008663
0.08647112167836499
0.07413049841666887
0.10306335321889983
                 0.03374233128834356
0.07860584480628201
0.10968385132719934
0.03211303789338471
'184"
'185"
                  0.07172143661981017
0.10932535246060386
                  0.10117750519544694
0.05438239409744308
0.0720589495330306
                 0.09492841545380133
0.09427848086047703
0.10215364731577099
0.10078643815066789
'195"
'196"
                 0.132044713553796
0.09899140045991586
0.048087416271543615
0.0684931506849315
201"
                  0.07523246599344224
                  0.08833170762908359
0.0815078125388226
207"
                  0.11278732656864776
                  0.11425442486544808
224"
                  0.08099702729092222
0.08354577861128779
0.08721017303002598
                  0.10447949050358085
0.06993302518630319
```

242"

```
0.07890211996979177
0.09530565669975881
0.09546944114505117
                                                                                                                                                  0.07530565669975881
0.09546944114505117
0.0
0.11995315301676546
0.07335984971767694
0.052623959464350334
0.11923576897943587
0.11052464311894397
0.043282871250452847
0.0019544811624019544
0.08558204423759197
0.06711331976762965
0.0972480119905075
0.11601004961517414
0.1093043993422745
0.10842648999719887
0.15786461137400468
0.12318663572582067
0.0532139249601685
0.07205229212930792
0.10801526749738075
0.11380613560114887
0.1142608495023124
0.11230316290901997
0.1080869840823124
0.1123031629901997
0.10969817731785466
0.07044812927929389
0.10315928607679181
0.04496955778040933
0.0352269858440825
0.431189363995688
0.09241457441314165
0.0738668844623049
0.1204236725267228
0.10801598668844623049
0.1204236725267228
0.1169105842822479
0.17581828771112398
0.1041970427360184
0.103438070866592
0.0383247822455542
0.116973134993934789
0.10621226787883002
0.11953106379554157
             2444""
2446""
2447"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478"
2478
    36"
37"
38"
39"
4"
40"
                                                                                                                                                       0.11953106379554157
0.09274584418153507
0.09274584418153507
0.07096965920546744
0.10407837515030439423
0.0983846147461719
0.12534064929126656
0.03340712045748017
0.1058125898855866
0.0962810318513851864
0.096810318513851864
0.096817813858649
0.054687879641958155
0.1015939184584642
0.191678678641958156
0.09239580486726348644
0.09187656958656626
0.092395882251749366
0.024588827631837916
0.024588827631837916
0.024588827631837916
0.024588827631837916
0.024588827631837916
0.02458827631837916
0.024588366964685292
0.096157487672791258
0.074770787917999912
0.0757078823916794779178
0.0757078823916794779179
0.10488538809956
0.1147487396246652
0.0801631657322679
0.10488538809956
0.1147487360246652
0.0801631657322679
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
0.10488538809956
'41"
'42"
'43"
'44"
'46"
'46"
'47"
'48"
'49"
'50"
    66"
67"
68"
                                                                                                                                                            0.11726023735794265

0.1128890978373831

0.11230199593086634

0.1314140628199353

0.0640252588354194

0.060647576771211555

0.0006641576267434137

0.0568739741017691

0.03836325514349365

0.12371478220911764

0.11535120965220538

0.08902658805206363

0.07875511622519114

0.11768850776024835
'81"
'82"
    '83"
'84"
'85"
    '88"
'89"
    9"
                                                                                                                                                                         0.11768850776024835
0.10308391768201007
                                                                                                                                                                     0.10308391768201007

0.056615381413745704

0.1071444344921823

0.038084346952212626

0.07928530356245143

0.09975598325818912
'94"
'95"
                                                                                                                                                                     0.09868461512594295
0.06067388225420083
                                                                                                                                                            @ip-172-31-37-211 ~1$
```

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

Variation in revenue over time is shown below including code used :

```
18. 411963137349552
15. 359309303734426
17. 4439645341407
16. 5500203407324047
16. 5500203407324047
17. 21604263570683
16. 83762356868249
14. 73267798036185
16. 363879207901295
16. 117071206714605
15. 933338279383616
16. 270620128859278
16. 41614230776937226
16. 1614230776937226
16. 161423076937226
16. 161423076937226
16. 161423076937226
17. 069302560401354
17. 102999568033265
17. 069302560401354
17. 10299958033265
16. 490724900576904
15. 885527411273001
17. 51018618590916
15. 787822292178655
15. 219828879150024
15. 99229313006401
15. 99229313006401
15. 99229313006401
15. 99229314947
16. 363236084025758
16. 2782988820224
15. 99283947134947
16. 514845325711754
16. 81774921469818
16. 388015828166143
16. 23140652815097
17. 121041382564638
16. 579383172574527
17. 337623268018195
17. 00356693816382
16. 730750002516206
17. 69994119517573
16. 72930392121595
                                                                                                                                                            15.161717603507384
17.47747548027094
18.734166782523527
17.364087663335088
17.252883134436264
18.14697136778204
17.248078089078962
15.846046419400455
18.86490867580518
17.010436833469626
17.389016241913282
16.37661024166937
17.679739001270963
16.147548793518364
18.825056911276535
21.29299722297417
29.068907005970854
19.732507288628188
20.829568189597944
17.38188976038157
21.593018355659638
18.1002788103778
18.89491666663246
17.825468666663246
18.845702192891927
21.227005213121405
20.99248200309997
21.12531300337945
18.366698922269
14.978345859721552
14.767823020740611
15.28831117881104
16.389897123112565
19.32208822201787
19.992739240104342
15.307352827781946
14.66863340021126
14.66863340021126
14.66968724817286
14.99385951942333
16.300197449100637
                                                                                                                                                                  14.68863340021126
14.46050088621336
14.619068724817286
14.939835919422332
                                                                                                                                                                  14.939835919422332
16.300197449100637
18.066767322725
15.534355540424196
15.310034719953286
4.991955112543714
15.095595443793671
15.144876788509523
                                                                                                                                                                       14.67262836614724
15.685100979728976
                                                                                                                                                                       15.79006959000546
15.809253019981295
15.622612773623201
9, 2] 13.8226127/3625201

9, 3] 15.764076749004406

9, 4] 15.912008246957104

9, 5] 13.814643752241684

9, 6] 14.779713618189996

pop@ip-172-31-37-211 ~]$
```