TASK 04: Map Reduce Job Outputs:

Write MapReduce codes to perform the tasks using the files you've downloaded on your EMR Instance:

We have used data 5 (yellow_tripdata_2017-05) for this process. Make sure to install the python and MRJob before running these syntax

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

Vendor 2 has the most number of the trips and the revenue generated from it is the most. [Note: the output shown is in format : "VendorID", (total_revenue, total_trips, avg_distance_per_trip, avg_passengers_per_trip), to have more information, we need only 1st data]

B. Which pickup location generates the most revenue? Pickup spot 132 yields the highest revenue: 14040591.220016211

```
[hadoop8ip-10-0-26-104 ~]$ python mrtask_b.py /home/hadoop/yellow_tripdata_2017-05.csv > /home/hadoop/task_b_result.txt No configs found; falling back on auto-configuration No configs specified for inline runner

Creating temp directory /tmp/mrtask_b.hadoop.20250204.164216.878845

Running step 1 of 2...

Running step 2 of 2...

job output is in /tmp/mrtask_b.hadoop.20250204.164216.878845/output

Streaming final output from /tmp/mrtask_b.hadoop.20250204.164216.878845/output...

Removing temp directory /tmp/mrtask_b.hadoop.20250204.164216.878845...

[hadoop8ip-10-0-26-104 ~]$ ls

mrtask_b.py mrtask_b.py mrtask_c.py task_a_result.txt task_b_result.txt yellow_tripdata_2017-05.csv

[hadoop8ip-10-0-26-104 ~]$ cat task_b_result.txt

[hadoop8ip-10-0-26-104 ~]$ cat task_b_result.txt
```

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

The several payment types used by customers, together with their counts, are listed below:

1 Credit card, 2 Cash, 3 No Charge, 4 Dispute

```
[hadoop@ip-10-0-26-104 ~]$ ls

mrtask_a.py mrtask_b.py mrtask_c.py task_a_result.txt task_b_result.txt yellow_tripdata_2017-05.csv
[hadoop@ip-10-0-26-104 ~]$ python mrtask_c.py /home/hadoop/yellow_tripdata_2017-05.csv > /home/hadoop/task_c_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner

Creating temp directory /tmp/mrtask_c.hadoop.20250204.165926.948244
Running step 1 of 2...
Running step 2 of 2...

pub output is in /tmp/mrtask_c.hadoop.20250204.165926.948244/output

Streaming final output from /tmp/mrtask_c.hadoop.20250204.165926.948244/output...
Removing temp directory /tmp/mrtask_c.hadoop.20250204.165926.948244...

[hadoop@ip-10-0-26-104 ~]$ ls

mrtask_a.py mrtask_b.py mrtask_c.py task_a_result.txt task_b_result.txt task_c_result.txt yellow_tripdata_2017-05.csv
[hadoop@ip-10-0-26-104 ~]$ cat task_c_result.txt

"2" 3250362

"3" 55027

"4" 15791
```

D. What is the average trip time for different pickup locations? The average trip time for different pickup locations are:

```
hadoop@ip-10-0-26-104:
                                        17.629476052716164
15.57929292929293
14.331578947368426
    "128"
"129"
                                        20.914534001214356
37.98320691108302
  "131"
"132"
"133"
                                        11.690000000000003
46.805229740819875
17.532022849462365
                                      17.532022849462365
16.221703296703296
15.368173258003768
9.693303571428572
14.412591446828932
41.79266819772832
18.82820512820513
18.910311918063318
14.544783434852864
12.887731943471524
14.724678434186814
  "134"
"135"
    136"
   "137"
"138"
  "138"
"139"
"14"
"140"
"141"
                                        14.724678434186814
13.878881130774534
17.422410667367306
13.919851029794044
15.779948824343021
13.003051643192489
   "143"
    "145"
"146"
"147"
                                     15.//994824343021
13.003051643192489
16.994242069527864
13.134482758620688
14.868589743589745
22.067647058822532
13.72286961525267
15.034273318872025
12.776791277258567
22.930303030303033
23.84722222222218
17.777777777777
24.96510297482838
17.056993321982915
12.997884615384616
11.23930817610063
17.381399631675876
16.501390496486135
15.5755374379993
17.007285407402943
  "148"
"149"
   "155"
"156"
  "157"
"158"
  "160"
"160"
"161"
"162"
                                      17.007285407402943
16.135913909640667
```

```
hadoop@ip-10-0-26-104:
"164"
        16.135913909640667
        23.146643518518516
"165"
"166"
        16.128181641420127
        11.731709401709402
"168"
        11.912370135587249
"17"
        14.360986719350239
        15.23280819184737
        11.867179487179486
"172"
        29.295833333333333
"173"
        12.247398589065257
        12.091085271317832
"174"
"175"
        11.432758620689656
"176"
        0.6
"177"
        20.955774278215223
"178"
        8.291562500000001
"179"
        15.1499705882353
"18"
        12.653225806451612
"180"
        27.18350694444446
        15.849603785417022
"182"
        12.2456
        11.92777777777779
"183"
"184"
        7.34666666666666
        15.095833333333333
"186"
        17.3607280512866
"187"
        1.425
        15.907792044310169
"188"
        16.631431947577614
10.768354430379746
"189"
"19"
"190"
        26.461985335589397
"191"
        13.486574074074074
"192"
        13.586842105263157
"193"
        12.075183464619935
"194"
        27.354013503375846
        22.571194503171245
"196"
        18.181564465408805
        18.036302681992343
"198"
        13.806195462478186
"20"
        17.141091954022986
"200"
"201"
        0.4937499999999999
"202"
        15.435855263157896
```

```
hadoop@ip-10-0-26-104:~
"202"
"203"
          15.435855263157896
         19.28479532163743
0.4799999999999987
"204"
205"
"206"
"208"
"209"
"21"
         16.197798742138364
210"
         18.476602564102567
"211"
"213"
"214"
         20.40786163522013
215"
         56.68466165413534
"216"
"217"
         17.684177215189873
53.02001915708813
'22"
         22.21085626911315
"220"
"221"
         19.394444444444442
17.158481387478847
"222"
"223"
         14.423218777619141
224"
"225"
         15.393991912189485
"226"
"227"
         15.537459546925568
"228"
"229"
         13.729634370452187
'23"
         3.605128205128205
"230"
"231"
"233"
         16.130618672379082
15.740532219114375
"234"
235"
         12.943728784429947
"238"
         13.635731925509624
"239"
         13.766637412862702
"24"
         14.086371798958675
```

```
hadoop@ip-10-0-26-104:
           12.961559139784944
           15.251830481496228
          13.879747675962818
          65.07303921568626
          18.156792744967923
18.57747506019952
13.053431372549017
          15.859693387283983
          19.92417624521073
          13.984743997175142
26.37702202480104
12.455172413793102
          17.463793103448275
          22.723684210526315
          12.525644805036702
13.519075992588938
18.701871980676327
16.11434108527132
11.448753894080996
          15.650715220659471
          22.96931464174455
          16.722509667955716
          10.0166666666667
          15.163276362823948
          16.827437223042836
5.575
          15.723731138545956
          22.17575468660486
          17.87714633298575
          153.69927536231884
          14.588625022901905
          23.28269720101781
14.84195171026157
          38.06967213114755
          11.647380952380951
          18.848467302452317
           18.5888888888888
           16.081618563620673
           13.002500000000001
               10-0-26-104
```

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

The average trip to revenue ratio of the drivers for different pickup locations in sorted format is:

```
[hadoop@ip-10-0-26-104 ~]$ cat task_e_result.txt
"1"
"10"
        0.1031314953272488
"100"
        0.10054811858030266
        0.1538484771604267
0.09686867923894887
"102"
105"
        0.11203391656570175
"106"
"107"
        0.11929564967810695
"108"
        0.12120884627098828
"109"
        0.23737785016286642
"11"
        0.058159884648828326
"111"
        0.077659754354142
"112"
        0.10900187462682069
        0.11769752457713911
        0.11571668143567702
        0.1014877461571702
0.09140144834181262
"116"
        0.029865191166550988
        0.03324247058041353
        0.06439098561727422
"119"
        0.09033344274559964
"120"
        0.14485426649668087
"121"
"122"
        0.035274937009041045
"123"
"125"
"126"
"127"
        0.08833111068775
"128"
        0.10511320192841678
"129"
        0.06647065796213625
        0.12390141102889671
"13"
"130"
        0.1073558396219219
```

```
hadoop@ip-10-0-26-104:~
"131"
"132"
          0.10384807204137783
0.1025660513460536
"134"
"135"
          0.10314438578934823
0.0785122214655781
"136"
"137"
"138"
"139"
          0.025278845473624637
           0.09666329734829485
          0.1110704470853594
0.11186974122790275
'141"
142"
          0.1131558975317282
"143"
"144"
"145"
          0.1105272721674073
0.09866671202718473
"147"
"148"
          0.04671430427439263
0.11347366987186253
"149"
"150"
          0.10633375877472878
0.10794854174655912
"151"
           0.08497213797829244
154"
          0.07400473993394348
 '155"
          0.09130948944530835
"156"
          0.11154334990834797
0.1177428105347338
"158"
           0.04880437380957894
"16"
160"
          0.09369565599255213
"161"
"163"
          0.11101476605573193
0.11138449806107008
"164"
           0.07196634804086764
"166"
          0.0356042774025466
          0.050410277807923884
          0.11765956298165817
```

```
hadoop@ip-10-0-26-104:
 "17"
"170"
                    0.08304806476027186
0.11737627919224174
"171"
"172"
                    0.04416689447785676
0.27196083763937995
                    0.0662888212150767
0.09292588232864547
                    0.08816570904350345
0.19982277359326542
0.09851238280057963
0.03941897087008663
"175"
"176"
 "177"
"178"
"179"
"18"
                    0.08647112167836499
0.07413049841666887
"18"
"180"
"181"
"182"
"183"
                    0.10306335321889983
0.11383678943060177
                    0.04935382990636337
0.04058310746465461
"184"
"185"
"186"
"187"
                    0.03374233128834356
0.07860584480628201
                    0.10968385132719934
0.03211303789338471
0.07172143661981017
0.10932535246060386
"187"
"188"
"189"
"19"
                    0.11032972645756875
0.10117750519544694
 "190"
"190"
"191"
"192"
"193"
"194"
                    0.05438239409744308
0.0720589495330306
                    0.03922346473669742
0.12076731965739748
                    0.09492841545380133
0.09427848086047703
0.10215364731577099
0.10078643815066789
"196"
"197"
                    0.11989767545323099
0.0736013986013986
0.13187606166264348
0.132044713553796
 "2"
 "20"
 "200"
 "201"
                    0.09899140045991586
0.048087416271543615
0.0684931506849315
0.07523246599344224
 "203"
"204"
"205"
 "206"
                    0.08833170762908359
```

hadoop@ip-10-0-26-104:~

```
0.0815078125388226
"208"
"209"
          0.08291166272142231
0.11297948250644907
"21"
          0.06386360722984234
"210"
          0.07934380776340108
          0.11310320504679515
0.06409624702307627
"211"
"212"
"213"
          0.11780938978546256
          0.02993836219547989
          0.10353565292295142
0.11008942973165864
"215"
"216"
          0.07868790199337758
"217"
"218"
          0.19235819846724786
          0.1049626881741011
"22"
          0.05041866234795492
0.056836272960188075
"220"
"221"
          0.10844779603195978
"222"
          0.11278732656864776
"223"
          0.09822645175828154
"224"
          0.11425442486544808
"225"
          0.08099702729092222
"226"
          0.08354577861128779
          0.08721017303002598
0.09978962780363482
"227"
"228"
          0.11203502738122283
"229"
"23"
          0.08146106133700898
"230"
          0.11828076056384894
0.10577578661446632
"232"
          0.11514074932415078
"234"
          0.12096822968790891
          0.0557415553911836
0.11249379678942588
"235"
"236"
          0.1096331664258742
"237"
"238"
          0.10834063914173245
"239"
"24"
          0.11534544071014695
0.10648126258555479
          0.10447949050358085
"240"
          0.06993302518630319
"242"
          0.07890211996979177
"243"
          0.09530565669975881
"244"
          0.09546944114505117
"245"
```

```
♣ hadoop@ip-10-0-26-104:~
"245"
"246"
             0.11995315301676546
"247"
             0.07335984971767694
"248"
"249"
             0.11923576897943587
0.11052464311894397
"25"
             0.043282871250452847
0.0019544811624019544
0.08558204423759197
"250"
"251"
"252"
             0.06711331976762965
0.0972480119905075
"253"
"254"
"255"
             0.11601004961517414
0.1093043993422745
"256"
             0.10482648999719887
"258"
             0.15786461137400468
0.12318663572582067
"259"
"26"
             0.05332139249601685
"260"
"261"
"262"
             0.10801526749738075
0.11380613560114887
"263"
             0.1142608495023124
             0.1142603493023124
0.11230316290901997
0.10968911731785466
0.07044812927929389
0.10315928607679181
"264"
"265"
"27"
"28"
"29"
             0.04496955778040933
0.0352269585440825
"3"
"30"
             0.431189363995688
"32"
             0.07388668844623049
0.1204236725267228
"33"
             0.11469105842822479
0.17581828771112398
0.10412970427360184
"34"
"35"
"36"
             0.103438070866592
"39"
             0.11607313499394789
0.10621226787883002
"4"
"40"
             0.11953106379554157
"41"
"42"
             0.09274584418153507
0.07096965920546744
"43"
             0.10407835112043434
"44"
             0.07601977750309023
```

```
"44"
"45"
                        0.07601977750309023
                      0.07601977750309023
0.0983846147461719
0.12534064929126656
0.03340712045748017
0.10582758808855866
0.09681013151351817
0.711159737417943
0.10923919835836489
0.054687879641958155
"47"
"48"
"49"
"5"
"50"
"51"
"52"
                       0.13055804576734864
0.09939808622286242
"52"
"53"
"54"
"55"
                      0.11968717470048443
0.10775881095465412
0.09187656958656626
0.0982393882217696
0.023456558453743663
0.022958827634333566
"57"
"58"
"59"
"6"
                       0.04668850781897014
"60"
"61"
                      0.028830566964685292
0.09614549547751258
0.07257887627997188
0.07430008293013912
"62"
 "63"
"64"
"65"
                       0.07777953179990528
0.1064885388809956
                      0.1064883388809956
0.11474873960246652
0.08016136157322679
0.1143179862016906
0.03888023995673477
"66"
"67"
"68"
 69"
                       0.08055330999297121
0.10558719618955761
"71"
"72"
"73"
                      0.057126398156593525
0.12979482386561036
0.06660273315751619
0.07309332094094213
"74"
"75"
                       0.09233854474158283
0.15699972823625333
0.03758990781571329
"76"
"77"
"78"
"79"
                       0.11726023735794265
"8"
"80"
                       0.1128890978373831
0.11230199593086634
"81"
"82"
                             1314140628199353
0640252588354194
```

```
"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
[hadoop@ip-10-0-26-104 ~]$
```

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).

Revenue variation over time:

```
[hadoop@ip-10-0-26-104 ~] $ 1s

mrtask_a.py mrtask_c.py mrtask_e.py task_a_result.txt task_c_result.txt task_e_result.txt

mrtask_b.py mrtask_d.py mrtask_f.py task_b_result.txt task_d_result.txt yellow_tripdata_2017-05.csv

[hadoop@ip-10-0-26-104 ~] $ python mrtask_f.py yellow_tripdata_2017-05.csv > task_f_result.txt

No configs found; falling back on auto-configuration

No configs specified for inline runner
 Treating temp directory /tmp/mrtask_f.hadoop.20250204.173354.340505
 ob output is in /tmp/mrtask f.hadoop.20250204.173354.340505/output
Streaming final output from /tmp/mrtask_f.hadoop.20250204.173354.340505/output... Removing temp directory /tmp/mrtask_f.hadoop.20250204.173354.340505... [hadoop@ip-10-0-26-104 ~]$ ls
 mrtask a.py mrtask c.py mrtask_e.py task_a_result.txt task_c_result.txt task_e_result.txt yellow_tripdata_2017-05.csv
mrtask_b.py mrtask_d.py mrtask_f.py task_b_result.txt task_d_result.txt task_f_result.txt
[hadoop@ip-10-0-26-104 ~]$ cat task_f_result.txt
 0, 4]
0, 5]
0, 6]
1, 0]
                         18.099999027301898
17.105128026116333
                         17.87588900705367
19.429909785928686
                         16.436848288253056
18.217479700823546
                         16.275959370214274
                         15.141293067405417
16.241741228653577
                         16.639627234011154
                         16.499299606770162
16.48680345506514
                          14.908801744148995
      11, 0]
                          16.680433054406084
```

```
hadoop@ip-10-0-26-104:
                                                                                                                                                                26-104:~

16.92334503331534
18.596232180453956
16.728098769455887
14.021549507826318
14.98238805727666
16.480360929345835
17.22963705731718
17.05563095494162
17.975730053716557
17.69103689287293
14.499792055020734
15.27335388620518
16.795085865627826
17.27948235089487
17.774827762048346
18.595673825475238
18.205408289793773
14.99034880898968
15.937601258660894
17.20858440343585
17.34785479975938
18.110961353384074
                                       17.20858440343585
17.34785479975938
18.110961353384074
19.033494081619924
18.6133879091461994
15.336584485461453
17.128896489169726
16.91500182066798
17.307682451791745
17.75765307206797
18.750789042356082
18.479680845892585
15.509464318824842
17.056592424577616
18.26141013123625
18.859940648116382
19.6542708916958
20.04885999577759
19.8347954545141705
15.635747754243198
17.32405836477559
17.555959492507064
17.747241962602008
18.3960718339617
hadoop@ip-10-0-26-104:~
```

2] 17, 2] 17, 3] 17, 4] 17, 5] 18, 0] 18, 1] 18, 2] 18, 3] 18, 4] 18, 5] 18.3960718339617 18.887732509348027 18.411963137349552 [5, [5, 15.359309303734426 17.4439645341407 16.500203407324047 16.477160631363912 [5, 18, [5, 18, 17.053779035240108 [5, [5, 17.221604263570683 16.83762356868249 [5, 18, [5, 18, [5, 18, 5] [5, 18, 6] [5, 19, 0] [5, 19, 2] [5, 19, 3] [5, 19, 4] [5, 19, 5] [5, 19, 6] 16.363879207901295 16.117071206744605 15.933338279383616 16.270620125859278 16.472398718229226 16.161423076937226 14.486719040663449 16.246780024174765 19, 6] 2, 0] 2, 1] 2, 2] 2, 3] 2, 4] 2, 5] 2, 6] [5, 17.069302560401354 [5, 17.102999568033265 [5, 16.490724900576904 15.885527411273001 [5, 17.510186185909916 15.787822292178655 [5, 15.219828879150024 16.278329889820224 20, 1] 20, 2] 20, 3] 15.92229313006401 15.999283947134947 [5, 16.363236084025758 [5, [5, 16.041364614564298 15.008297507476271 20, 4] 5] 16.514845325711754 6] [5, [5, 16.81774921469818 0] [5, [5, 16.231406528150977 17.121041382564638 21, 21, 21, 22, 22, [5, [5, [5, 16.579383172574527 5] 15.030104427307885

16.83408532267027

17.337623268018195 17.00356693816382 16.730750002516206

6]

0] 1] 2]

```
hadoop@ip-10-0-26-104:~
[5, 4, 0]
                21.292999725297417
                 20.068907006970854
[5, 4, 1]
                 19.732507288628188
[5, 4, 2]
                20.829968819597944
[5, 4, 3]
[5, 4, 4]
                 20.915669272681466
[5, 4, 5]
                 17.825446810289648
[5, 4, 6]
                17.38188976038157
[5, 5, 0]
                 21.593018355659638
[5, 5, 1]
                 18.81002788103778
[5, 5, 2]
                18.894916666663246
[5, 5, 3]
                 19.845702192891927
[5, 5, 4]
                 21.227005213121405
                 20.992848200309997
[5, 5, 5]
                 21.125313003037945
[5, 5, 6]
                16.36666989228269
[5, 6, 0]
                 14.978345859721552
[5, 6, 1]
                 14.767823020740611
[5, 6, 2]
                 15.288311117881104
[5, 6, 3]
[5, 6, 4]
                 16.389897123112565
[5, 6, 5]
                 19.32208822201787
[5, 6, 6]
                 19.992739240104342
                 15.307352827781946
[5, 7, 1]
                 14.68863340021126
[5, 7, 2]
                 14.46050088621336
[5, 7, 3]
                 14.619068724817286
[5,
                 14.939835919422332
[5,
                 16.300197449100637
[5,
                 18.066767322725
[5,
       0]
                 15.534355540424196
[5,
   8, 1]
                 15.310034719953286
[5, 8, 2]
                 14.991955112543714
[5, 8, 3]
                 15.095595443793671
                 15.144876788509523
[5, 8, 4]
[5, 8, 5]
                 14.67262836614724
                 15.685100979728976
[5, 8, 6]
                 15.79006959000546
[5, 9, 0]
                 15.809253019981295
[5, 9, 1]
                 15.622612773623201
[5, 9, 2]
                 15.764076749004406
[5, 9, 3]
   9, 4]
                 15.912008246957104
[5,
                 13.814643752241684
[5,
   9, 5]
                 14.779713618189996
[5,
    9, 6]
[hadoop@ip-10-0-26-104 ~]$
```