MapReduce Assignment

MRJOB solutions

Inder Baldawa | Deepika Gupta | Shrawani Das

Task 4

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

```
[hadoop@ip-172-31-45-231 ~]$ cat output_task_a.txt
"2"      525037658.13640213
```

Vendor 2 has most trips and revenue generated is 525037658

b. Which pickup location generates the most revenue?

```
"132"     77196812.23975265
```

Pickup Location 132 Generates the most revenue with $77196812

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

```
                    WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/mrjob_c.hadoop.20230717.115343.126795/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/mrjob_c.hadoop.20230717.115343.126795/output...
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/mrjob_c.hadoop.20230717.115343.126795...
Removing temp directory /tmp/mrjob_c.hadoop.20230717.115343.126795...
[hadoop@ip-172-31-45-231 ~]$ cat output_task_c.txt
"1"      39754212
"2"      18832370
"3"      306912
"4"      88794
"5"      3
```

1.Credit card 39754212

2. Cash 18832370

3. No charge 306912

4. Dispute 88794

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

```
"102"    4.025800781249999
"105"    5.058235294117647
"108"    4.232330097087378
"111"    2.7176821192052976
"114"    2.4092175848490642
"117"    6.266162790697674
"12"     4.22755556510898
"120"    5.149235127478752
"123"    3.301801385681293
"126"    3.744306151645207
"129"    3.1253958451906656
"132"    15.887783997800563
"135"    4.284182879377431
"138"    9.844836102456174
"141"    2.010259795953193
"144"    2.5134206655499187
"147"    3.2264503816793897
"15"     7.018807339449541
"150"    8.205494505494507
"153"    3.2835104669887283
"156"    3.3472815533980587
"159"    2.737375928677564
"162"    2.4155778867382263
"165"    3.405671476137624
"168"    2.7102173913043477
"171"    3.755919540229887
"174"    4.017086330935251
"177"    3.98715500327011
"18"     3.1677553956834537
"180"    5.953006833712984
"183"    4.812004166666666
"186"    2.2789165963505145
"189"    2.892652954959646
```

```
"192"    4.500358890701467 6
"195"    4.599095548624894
"198"    3.251393697292499
"201"    4.783943661971831
"204"    0.4985714285714286
"207"    1.1052761123382773
"21"     4.677125
"210"    4.924309859154928
"213"    5.4949930843706785
"216"    9.264181914611509
"219"    14.861461512377078
"222"    9.963333333333335
"225"    2.8097743632336667
"228"    3.5022664756446984
"231"    2.8847471994521805
"234"    2.1032000673227036
"237"    1.7790478221062263
"24"     2.493792463996845
"240"    5.883370786516854
"243"    3.827098584868369
"246"    2.3835551947097535
"249"    2.2911436843866637
"252"    8.917625000000001
"255"    3.1175852499434256
"258"    4.654100905562742
"261"    4.01456261648351
"264"    2.645273063328758
"27"     4.665454545454545
"3"      4.457030303030304
"30"     6.850476190476191
"33"     3.9180095253736233
"36"     3.247495117187499
"39"     5.200807577268196
"42"     2.358211884915301
```

```
"57"     4.999497964949486 6
"6"      3.388038461538463
"60"     3.257266666666667
"63"     3.5291677588466563
"66"     3.848042112610515
"69"     2.909512022630835
"72"     3.945699248120301
"75"     2.3191676013766416
"78"     3.295724962630793
"81"     4.944974874371859
"84"     4.8385
"87"     4.331207490026939
"9"      4.3792178770949715
"90"     2.0503217018894024
"93"     9.505599709934735
"96"     4.752661870503597
"99"     6.928
"1"      2.2837912087912087
"10"     14.811629413374689
"100"    2.309335242067255
"106"    3.0220092576718676
"109"    6.427307692307693
"112"    3.09234724409449
"115"    3.4785892116182575
"118"    2.5742857142857147
"121"    4.802074074074076
"124"    8.91911147011309
"127"    4.001065356418656
"13"     3.9915872675816946
"130"    10.668824714573539
"133"    3.6310593637258113
"136"    3.2555158184319124
"139"    7.420576923076924
"142"    2.156600666407327
```

Complete solution Attached as a text file named "Output_task_d.txt

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

```
"102"    0.09286325213174113        "93"     0.11009789332101659        "248"    0.05406105607423323
"105"    0.08289275441720313        "96"     0.11215306381581798        "251"    0.08299641950457239
"108"    0.08561073480894167        "99"     0.06008865539320309        "254"    0.045390942684416416
"111"    0.101754582865694          "1"      0.11931462821703112        "257"    0.1131136567021167
"114"    0.11737171385745779        "10"     0.10761668826560952        "26"     0.06577859806429147
"117"    0.16546115321544122        "100"    0.10176112040929118        "260"    0.07000191673138242
"12"     0.0881691266641008         "106"    0.11182299192052228        "263"    0.11388129538454757
"120"    0.08303252748719825        "109"    0.17861970356364318        "29"     0.06573652675700856
"123"    0.11055898094072127        "112"    0.10939659242523393        "32"     0.06109787327761728
"126"    0.0722310684901339         "115"    0.09805645544907124        "35"     0.10022554243576677
"129"    0.06331017333926041        "118"    0.117289635611784          "38"     0.05791427287264455
"132"    0.10404480517960502        "121"    0.09104834906722972        "41"     0.09013710776489915
"135"    0.09112992092386224        "124"    0.08450445139029464        "44"     0.06402542001826438
"138"    0.13191299589436914        "127"    0.0883881479101126         "47"     0.03560717179556587
"141"    0.1076819360724295         "13"     0.12263264632357321        "5"      0.173567645642754
"144"    0.11283013521689307        "130"    0.10459971957672475        "50"     0.10916652254418897
"147"    0.04460414952837091        "133"    0.09428655857404536        "53"     0.07740291031997901
"15"     0.11600468842514833        "136"    0.040904378620177845       "56"     0.09883778991554348
"150"    0.1040620992466899         "139"    0.09244706142211939        "59"     0.011954905847373635
"153"    0.060563732535779416       "142"    0.11242578567788385        "62"     0.07661527162660449
"156"    0.0879253766696818         "145"    0.09530064643710734        "65"     0.10543895761087184
"159"    0.04663800562314161        "148"    0.11466547302905533        "68"     0.11443748108215406
"162"    0.12066632396973824        "151"    0.10736778554968535        "71"     0.09976058424195855
"165"    0.06797416237704262        "154"    0.09108097210043588        "74"     0.07135080378568034
"168"    0.049620759878179065       "157"    0.1007884941588729         "77"     0.09702969688085886
"171"    0.0805299331453214217      "16"     0.123473429808929          "8"      0.1086282452986951
"174"    0.06278143350842431        "160"    0.09647705819251776        "80"     0.1102718770090173
"177"    0.07641696352884511        "163"    0.11222164046426629        "83"     0.05736768211329472
"18"     0.05241971650085565        "166"    0.11278675284128852        "86"     0.10935811216524559
"180"    0.10523696750375819        "169"    0.06451074143663851        "89"     0.08958574112049193
"183"    0.058534099857434754       "172"    0.1723686242471645         "92"     0.06480851959104653
"186"    0.10977067019782545        "175"    0.1277230776008282         "95"     0.0833569398816454
"189"    0.11197515428640378        "178"    0.12194781013939289        "98"     0.08788658752859359
```

Complete solution Attached as a text file named "Output_task_e.txt

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).

```
[1, 0, 0]      17.519291696330246
[1, 0, 3]      16.626480239594823
[1, 0, 6]      15.44823421897282
[1, 1, 2]      16.55492861952621
[1, 1, 5]      14.981404492508563
[1, 10, 1]     15.127706071055114
[1, 10, 4]     15.164584937260672
[1, 11, 0]     14.480895520162157
[1, 11, 3]     15.248743102814249
[1, 11, 6]     13.477203683198983
[1, 12, 2]     15.195805597340291
[1, 12, 5]     13.17964793309867
[1, 13, 1]     15.129944553176106
[1, 13, 4]     15.731715024471264
[1, 14, 0]     15.17123758065054
[1, 14, 3]     16.288452460943006
[1, 14, 6]     15.353656693544734
[1, 15, 2]     15.808623447320434
[1, 15, 5]     14.286627006407732
[1, 16, 1]     16.79477246941084
[1, 16, 4]     17.140632285850383
[1, 17, 0]     15.603011225898907
[1, 17, 3]     16.493862512031395
[1, 17, 6]     15.842124390133883
[1, 18, 2]     15.838856512260497
[1, 18, 5]     13.799004771741297
[1, 19, 1]     15.184492458519966
[1, 19, 4]     15.181602676051245
[1, 2, 1]      16.653553023303136
[1, 2, 4]      15.440661491541219
[1, 20, 0]     15.381471427086986
[1, 20, 3]     15.787915693741633
[1, 20, 6]     16.05652368409678
```

```
[2, 1, 6]      15.080928222376645
[2, 10, 2]     15.584856505534193
[2, 10, 5]     13.235825072012585
[2, 11, 1]     15.137241340754445
[2, 11, 4]     15.81282871425917
[2, 12, 0]     14.997769908453845
[2, 12, 3]     15.09902583793382
[2, 12, 6]     13.83217228096712
[2, 13, 2]     16.214876939591328
[2, 13, 5]     14.139899300861922
[2, 14, 1]     15.76594127130329
[2, 14, 4]     16.988947496203075
[2, 15, 0]     15.391670236531361
[2, 15, 3]     15.68904171508541
[2, 15, 6]     15.989882115791604
[2, 16, 2]     17.706524452030504
[2, 16, 5]     14.637107029630501
[2, 17, 1]     16.018190446543887
[2, 17, 4]     16.885100841332783
[2, 18, 0]     15.211415337181817
[2, 18, 3]     15.424042083941371
[2, 18, 6]     15.117485791174863
[2, 19, 2]     15.4460226975043
[2, 19, 5]     14.126359078349951
[2, 2, 2]      15.231154903757654
[2, 2, 5]      15.293780018084675
[2, 20, 1]     15.249453542939467
[2, 20, 4]     15.531191802088282
[2, 21, 0]     16.226778214033512
[2, 21, 3]     16.01842724605061
[2, 21, 6]     16.746228025406182
```

```
[4, 14, 6]     16.840753872341526
[4, 15, 2]     16.831821811911965
[4, 15, 5]     15.83041643239797
[4, 16, 1]     17.303590020325895
[4, 16, 4]     18.66117987598206
[4, 17, 0]     16.935342234490466
[4, 17, 3]     18.247976802512632
[4, 17, 6]     17.193937591282953
[4, 18, 2]     16.369772137047704
[4, 18, 5]     14.624873769593135
[4, 19, 1]     15.497339795567646
[4, 19, 4]     15.762378440794592
[4, 2, 1]      16.34083949787082
[4, 2, 4]      18.132232268254125
[4, 20, 0]     15.806433938201556
[4, 20, 3]     16.359462410280326
[4, 20, 6]     16.98237080332982
[4, 21, 2]     16.449890988815383
[4, 21, 5]     15.33205354060174
[4, 22, 1]     16.76673100333906
[4, 22, 4]     16.520808734827668
[4, 23, 0]     17.74243445353905
[4, 23, 3]     18.103497884552784
[4, 23, 6]     19.53774978766057
[4, 3, 0]      17.035402713379195
[4, 3, 3]      17.008209039548387
[4, 3, 6]      15.842982230340015
[4, 4, 2]      20.618848994142695
[4, 4, 5]      17.593235454788704
[4, 5, 1]      18.900795007561847
[4, 5, 4]      21.08462074977878
[4, 6, 0]      16.050704128041698
[4, 6, 3]      15.132386678529734
```

Complete solution Attached as a text file named "Output_task_e.txt