Task 4. Write MapReduce codes to perform the tasks using the files you've downloaded on your EMR instance

Note : We pass the directory path as input to all the below MRJobs. The directory has all the 6 csv files and MRJob in-turn reads all the files for processing.

```
[[hadoop@ip-172-31-6-157 data]$ ls -lrt
total 5425468
-rw-rw-r-- 1 hadoop hadoop 914029540 Nov 25  2022 yellow_tripdata_2017-01.csv
-rw-rw-r-- 1 hadoop hadoop 863487050 Nov 25  2022 yellow_tripdata_2017-02.csv
-rw-rw-r-- 1 hadoop hadoop 969809025 Nov 25  2022 yellow_tripdata_2017-03.csv
-rw-rw-r-- 1 hadoop hadoop 946349441 Nov 25  2022 yellow_tripdata_2017-04.csv
-rw-rw-r-- 1 hadoop hadoop 951965526 Nov 25  2022 yellow_tripdata_2017-05.csv
-rw-rw-r-- 1 hadoop hadoop 910028408 Nov 25  2022 yellow_tripdata_2017-06.csv
[[hadoop@ip-172-31-6-157 data]$
```

a) **Which vendors have the most trips, and what is the total revenue generated by that vendor?**
>> Vendor 2 (VeriFone Inc.) has the most trips and total revenue is $525037658.14

```
     __|  __|_  )
     _|  (     /   Amazon Linux 2 AMI
    ___|\___|___|

https://aws.amazon.com/amazon-linux-2/
88 package(s) needed for security, out of 155 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::::E M:::::::M         M:::::::M R:::::RRRRRR:::::R
  E:::::E       EEEEE M:::::::::M       M:::::::::M RR::::R     R::::R
  E:::::E             M::::::M:::M     M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE    M::::::M M:::M M:::M M::::::M   R:::RRRRRR:::::R
  E:::::::::::::::E    M::::::M  M:::M:::M  M::::::M   R:::::::::::RR
  E:::::EEEEEEEEEE    M::::::M   M:::::M   M::::::M   R:::RRRRRR::::R
  E:::::E             M::::::M    M:::M    M::::::M   R:::R      R::::R
  E:::::E       EEEEE M::::::M     MMM     M::::::M   R:::R      R::::R
EE:::::EEEEEEEE::::E M::::::M             M::::::M   R:::R      R::::R
E::::::::::::::::::E M::::::M             M::::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM             MMMMMMMM RRRRRRR      RRRRRR

[[hadoop@ip-172-31-1-136 ~]$ python3 mrtask_a.py data/^C
[[hadoop@ip-172-31-1-136 ~]$ ps -efw | grep mrt
hadoop    27213 13606  0 04:06 pts/6    00:00:00 grep --color=auto mrt
[[hadoop@ip-172-31-1-136 ~]$ python3 mrtask_a.py data/
No configs found; falling back on auto-configuration
No configs specified for inline runner
```

```
"2"     525037658.13737655
```

b) Which pickup location generates the most revenue?

>> Pickup location 132 generated most revenue of $77196812.24

```
[hadoop@ip-172-31-1-136 ~]$ screen
[hadoop@ip-172-31-1-136 ~]$ python3 mrb.py data/
No configs found; falling back on auto-configuration
No configs specified for inline runner
```

```
"132"    77196812.23977433
```

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

>>> 5 different payment types are used by customer with the counts as shown in the output below

- 1 or Credit Card 39754212 times
- 2 or Cash   18832370 times
- 3 or Nor charge   306912 times
- 4 or Dispute   88794 times
- 5 or Unknown   3 times

```
[hadoop@ip-172-31-1-136 ~]$ python3 mrtask_c.py data/
No configs found; falling back on auto-configuration
No configs specified for inline runner
```

```
"1"     39754212
"2"     18832370
"3"     306912
"4"     88794
"5"     3
```

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

>>> Average trip time in seconds for pickup locations shown below in output

```
"1"      491.18705207835643
"10"     2986.1142231548215
"100"    927.1036323654433
"101"    1111.171568627451
"102"    1326.7421875
"104"    1417.0
"105"    1198.6666666666667
"106"    848.5432024687125
"107"    850.2061969948687
"108"    852.504854368932
"109"    3646.576923076923
"11"     1030.1524390243903
"110"    191.0
"111"    698.0794701986755
"112"    866.7667716535433
"113"    899.9909719158948
"114"    955.2777877786825
"115"    846.0082987551867
"116"    938.2699303995785
"117"    1159.0581395348838
"118"    735.5904761904762
"119"    979.2593184238551
"12"     1461.5228064141697
"120"    828.4872521246459
"121"    848.6666666666666
"122"    1629.77
"123"    930.8799076212471
"124"    1735.6284329563812
"125"    953.9664961793401
"126"    1113.7067238912732
"127"    933.9658399625644
"128"    893.8765743073047
"129"    855.5217985202049
"13"     1170.1200095775785
"130"    2110.0042534139243
"131"    846.7052341597796
"132"    2626.217323178503
"133"    996.3551984257133
"134"    950.9126246990024
"135"    1086.227626459144
"136"    683.5790921595599
"137"    827.1185944225588
"138"    2239.3492221442825
"139"    3579.403846153846
"14"     1012.0196145905024
"140"    835.7193930280911
"141"    735.7215745971303
"142"    831.7647130463778
"143"    786.8845666913322
"144"    1008.2267243981698
"145"    766.4283870790605
"146"    902.894640747503
"147"    789.2442748091603
"148"    1000.1113790439446
"149"    1003.7186440677966
"15"     872.5137614678899
"150"    1109.5384615384614
"151"    792.8776775438987
"152"    849.5668384259411
"153"    813.779388083736
"154"    1467.2142857142858
"155"    1621.730303030303
"156"    1164.2621359223301
```

.....

.....

"35"	1166.1373844121533
"36"	1012.40791015625
"37"	967.6872436529775
"38"	3059.3202614379084
"39"	1094.061814556331
"4"	899.679533415981
"40"	972.7623264127595
"41"	777.4007117536153
"42"	777.5933793433121
"43"	936.8280259262674
"44"	704.4285714285714
"45"	1107.9150232922204
"46"	1056.2631578947369
"47"	995.4081632653061
"48"	881.4218521084557
"49"	839.2991446028514
"5"	686.6923076923077
"50"	892.0096256834727
"51"	976.6214285714286
"52"	1039.9510570358952
"53"	1028.8467532467532
"54"	1021.9478827361563
"55"	1485.292156862745
"56"	1206.7722772277227
"57"	865.6976744186046
"58"	388.15384615384613
"59"	1009.0
"6"	450.11538461538464
"60"	754.3377777777778
"61"	920.1694269281408
"62"	918.1529535864979
"63"	1152.1068152031455
"64"	797.330985915493
"65"	1017.6511543837663
"66"	1068.923142546921
"67"	1162.696103896104
"68"	919.5494840005136
"69"	1004.2298444130128
"7"	795.8500817600714
"70"	1512.4559151785713
"71"	978.8518518518518
"72"	1180.5293233082707
"73"	1248.4758620689656
"74"	757.7213734172583
"75"	774.7370684258802
"76"	1303.3945912722802
"77"	1195.3758865248226
"78"	777.50822122571
"79"	911.428828476829
"8"	1049.4106145251396
"80"	909.4648496930906
"81"	781.9949748743719
"82"	878.4701140158932
"83"	952.6636375271983
"84"	706.45
"85"	1185.0851800554017
"86"	867.0769230769231
"87"	1226.5451736505456
"88"	1299.0926384989962
"89"	959.6696592158299
"9"	5069.413407821229
"90"	850.6428322112844
"91"	1251.828839390387
"92"	1051.3910096325367
"93"	2039.3028281363306
"94"	1160.1567164179105
"95"	1071.6730681298582
"96"	943.9640287769785
"97"	942.4900382701246
"98"	735.3614457831326
"99"	763.6

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

>>> sorted [pickup location <--> average tips to revenue ratio] shown below

```
"1"     0.11931462821703755
"10"    0.10761668826559809
"100"   0.10176112040806533
"101"   0.10391200650224494
"102"   0.09286325213174136
"104"   0.2000665778961385
"105"   0.08289275441720317
"106"   0.11182299192052901
"107"   0.11974753027213873
"108"   0.08561073480894177
"109"   0.1786197035636432
"11"    0.06611923513377366
"110"   0.0
"111"   0.10175458286569393
"112"   0.10939659242523861
"113"   0.11854772693334716
"114"   0.11737171385691483
"115"   0.09805645544907107
"116"   0.09022031542003062
"117"   0.16546115321544114
"118"   0.11728963561178403
"119"   0.0677823429510212
"12"    0.08816912666412315
"120"   0.08303252748719829
"121"   0.09104834906722978
"122"   0.13240255677287155
"123"   0.11055898094072264
"124"   0.08450445139029461
"125"   0.12389238738128613
"126"   0.07223106849013458
"127"   0.08838814791010641
"128"   0.10195419121098405
"129"   0.0633101733392787
"13"    0.12263264632333887
"130"   0.1045997195767284
"131"   0.08338783482815308
"132"   0.10404480517981093
"133"   0.09428655857404286
"134"   0.10749874267012843
"135"   0.09112992092386339
"136"   0.040904378620178095
"137"   0.11277472565910356
"138"   0.13191299589415384
"139"   0.09244706142211924
"14"    0.09044594848522367
"140"   0.1111844421180549
"141"   0.10768193607112014
"142"   0.11242578567608283
"143"   0.11408622408855171
"144"   0.11283013521682518
"145"   0.09530064643701493
"146"   0.08853658145669066
"147"   0.04460414952837101
"148"   0.11466547302828473
```

```
"202"    0.077220797914472
"203"    0.08922376129095652
"204"    0.06247952716655987
"205"    0.08302284089547893
"206"    0.0835728092968298
"207"    0.07416425281279042
"208"    0.0771871688501438
"209"    0.11438950071387221
"21"     0.08297404683405164
"210"    0.12696291028233796
"211"    0.11399355397937927
"212"    0.07459991612034597
"213"    0.09930916492713535
"214"    0.07745375602878152
"215"    0.108292216671091852
"216"    0.10479778546735218
"217"    0.07433903922530104
"218"    0.09787185841442574
"219"    0.10013130970242928
"22"     0.06303381261429293
"220"    0.0754805545138255
"221"    0.102966610645201111
"222"    0.09461551112794334
"223"    0.09657319511941766
"224"    0.115268800032060795
"225"    0.08262074885723784
"226"    0.08554438430845938
"227"    0.0806057192559698
"228"    0.09725404075366931
"229"    0.11256386704259114
"23"     0.11893642866783627
"230"    0.10760970767272766
"231"    0.11887862152233053
"232"    0.08364178523169383
"233"    0.11587691604669297
"234"    0.12013283079340241
"235"    0.05148549299609817
"236"    0.11108195526238974
"237"    0.10990442459364845
"238"    0.11044770762990037
"239"    0.11135965083688437
"24"     0.10590891911110777
"240"    0.07105544591130947
"241"    0.06142723563688872
"242"    0.07134362315468425
"243"    0.09649751817938466
"244"    0.09413161238210144
"245"    0.08812313347116933
"246"    0.119613178882515317
"247"    0.07294934365908352
"248"    0.05406105607423298
"249"    0.119990213646580888
"25"     0.10931197115088788
"250"    0.05704400870673731
"251"    0.0829964195045723
"252"    0.11791966686185068
"253"    0.09444941923267369
"254"    0.045390942684416194
"255"    0.1168569951546094
"256"    0.10853673778730212
"257"    0.11311365670211562
"258"    0.11397415106036027
"259"    0.08717567117240292
"26"     0.0657785980642925
"260"    0.07000191673138838
"261"    0.1076661924308677
"262"    0.11338221978066361
"263"    0.11388129538369067
"264"    0.112652426617702774
"265"    0.11224718082317552
"27"     0.08809887044827637
"28"     0.10225113334394173
```

```
"149"    0.08233859572151524
"15"     0.11600468842251483
"150"    0.1040620992466897
"151"    0.10736778555021802
"152"    0.08319834814824512
"153"    0.0605637325357794
"154"    0.09108097210043568
"155"    0.09633375755470457
"156"    0.08792537666968141
"157"    0.10078849415880525
"158"    0.11639038989027432
"159"    0.04663800562314045
"16"     0.1234734298089286
"160"    0.09647705819251869
"161"    0.11359049234222156
"162"    0.12066632396784141
"163"    0.11222164046260341
"164"    0.11217684253729482
"165"    0.06797416237704355
"166"    0.11278675284164902
"167"    0.06122252383032679
"168"    0.04962075987818404
"169"    0.06451074143663958
"17"     0.08317536901457533
"170"    0.11808315282090162
"171"    0.08052993314534176
"172"    0.17223686242471647
"173"    0.060081458043702254
"174"    0.06278143350842448
"175"    0.12772307760082793
"176"    0.15889955267208697
"177"    0.07641696352884644
"178"    0.1219478101393935
"179"    0.08477276212071912
"18"     0.052419716500855834
"180"    0.10523696750375865
"181"    0.11303400885954477
"182"    0.0495611836034771
"183"    0.05853409985743483
"184"    0.11633468693321357
"185"    0.07530969983194803
"186"    0.10977067019589502
"187"    0.17978913134704158
"188"    0.07438531115225966
"189"    0.11197515428642556
"19"     0.07943494943114222
"190"    0.10147578909371616
"191"    0.1248716806280989
"192"    0.07124383780662975
"193"    0.04602694182350332
"194"    0.12129343435403792
"195"    0.10463665842928094
"196"    0.08616922680963854
"197"    0.09057550505118056
"198"    0.08806313330584374
"199"    0.14024694882592245
"2"      0.10551996408687955
"20"     0.07539779360644304
"200"    0.10311919266153378
"201"    0.15164728494049579
"202"    0.09722093931447223
```

```
"35"   0.10022554243576758
"36"   0.105804890504773897
"37"   0.099211783789813363
"38"   0.05791427287264449
"39"   0.096517077268563887
"4"    0.10684873200924028
"40"   0.12031726919563045
"41"   0.090137107764995521
"42"   0.070362410836005937
"43"   0.10492270664830988
"44"   0.064025420018264438
"45"   0.099342745168855802
"46"   0.082343320887446263
"47"   0.035607171795566607
"48"   0.10627040728539182
"49"   0.099433009399017424
"5"    0.173567645642754
"50"   0.10916652254471039
"51"   0.089521524795662285
"52"   0.12903601946205442
"53"   0.077402910319977881
"54"   0.12031650057919829
"55"   0.094705513327400061
"56"   0.098837789915554232
"57"   0.096395566716488
"58"   0.14391826665236515
"59"   0.011954905847373635
"6"    0.048318188858848527
"60"   0.040872018894011143
"61"   0.0921603829004545
"62"   0.076615271626600291
"63"   0.10608349633094756
"64"   0.11356366975963506
"65"   0.105438957610779998
"66"   0.11716659766188296
"67"   0.10270469503747026
"68"   0.11443748108064815
"69"   0.047707317902001646
"7"    0.07831986411647032
"70"   0.107883809278128771
"71"   0.0997605842419591
"72"   0.088653796920270092
"73"   0.07567600767754308
"74"   0.071350803785759661
"75"   0.090137474126480446
"76"   0.11266074016140237
"77"   0.097029696880085883
"78"   0.049456821210188885
"79"   0.11800406819637435
"8"    0.10862824529869562
"80"   0.11027187700902322
"81"   0.0809078598125776
"82"   0.064743588258184684
"83"   0.05736768211328877
"84"   0.12333321182517411
"85"   0.066028545112660058
"86"   0.10935811216524567
"87"   0.12471673086151293
"88"   0.11647722217055859
"89"   0.089585741120486189
"9"    0.09026115108356442
"90"   0.11445563465162432
"91"   0.07770689984709578
"92"   0.06480851959104458
"93"   0.11009789332101962
"94"   0.032612401099441655
"95"   0.08335693988164992
"96"   0.1121530638158179
"97"   0.098603789667474414
"98"   0.087886587528593331
"99"   0.060088865539320309
```

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).

>>> The results
- Data seem to be only from Jan to June
- Tip amounts seem to increase from 9pm to midnight.
- Very early morning, like 3 to 6am has highest tips observed
- Summer (June shows up slightly higher tips to winter months)

output of average trip revenue by (month hour day_of_week) shown below

```
[1, 0, 0]        17.51929169633032
[1, 0, 1]        18.123630421432235
[1, 0, 2]        17.3589315046607
[1, 0, 3]        16.626480239594816
[1, 0, 4]        16.703014661253516
[1, 0, 5]        15.550549406444684
[1, 0, 6]        15.448234218972123
[1, 1, 0]        17.461791503066678
[1, 1, 1]        17.072226545057458
[1, 1, 2]        16.55492861952615
[1, 1, 3]        15.679302247938455
[1, 1, 4]        16.052486926635606
[1, 1, 5]        14.981404492508569
[1, 1, 6]        15.387799144324102
[1, 10, 0]       14.863052216909038
[1, 10, 1]       15.127706071054034
[1, 10, 2]       15.186626217363893
[1, 10, 3]       15.349964897412063
[1, 10, 4]       15.16458493725983
[1, 10, 5]       12.609842916276442
[1, 10, 6]       13.490205609938638
[1, 11, 0]       14.480895520161702
[1, 11, 1]       14.888644963766478
[1, 11, 2]       15.189027604271173
[1, 11, 3]       15.248743102812911
[1, 11, 4]       15.417036801711069
[1, 11, 5]       12.720721465880247
[1, 11, 6]       13.477203683199233
[1, 12, 0]       14.481984042562795
[1, 12, 1]       14.798213902435325
[1, 12, 2]       15.19580559733942
[1, 12, 3]       15.504904572604124
[1, 12, 4]       15.467102026936905
[1, 12, 5]       13.179647933098668
[1, 12, 6]       13.953226689445591
[1, 13, 0]       14.647866145163956
[1, 13, 1]       15.129944553175028
[1, 13, 2]       15.783343513633147
[1, 13, 3]       16.22223372993386
[1, 13, 4]       15.731715024471654
[1, 13, 5]       13.766750076335171
[1, 13, 6]       14.64248722087208
[1, 14, 0]       15.1712375806504
[1, 14, 1]       15.390354582919297
[1, 14, 2]       15.879764775837332
[1, 14, 3]       16.28845246094191
[1, 14, 4]       15.759659749888717
[1, 14, 5]       14.290606848792967
[1, 14, 6]       15.353656693543854
[1, 15, 0]       14.897552997614634
[1, 15, 1]       15.191981550355933
[1, 15, 2]       15.808623447320434
[1, 15, 3]       16.234477620663167
[1, 15, 4]       15.60063767838376
[1, 15, 5]       14.28662700640749
[1, 15, 6]       15.279192695092805
[1, 16, 0]       16.118353610150415
[1, 16, 1]       16.794772469409857
[1, 16, 2]       17.342044454087628
[1, 16, 3]       17.544082880554697
[1, 16, 4]       17.14063228585104
[1, 16, 5]       14.67454882591177
[1, 16, 6]       15.753284019435748
```

```
[1, 17, 0]      15.603011225898129
[1, 17, 1]      16.228117463370307
[1, 17, 2]      16.521664397999153
[1, 17, 3]      16.4938625120304
[1, 17, 4]      16.023100150993518
[1, 17, 5]      14.433303928129876
[1, 17, 6]      15.842124390133558
[1, 18, 0]      15.020223885270742
[1, 18, 1]      15.393338942252164
[1, 18, 2]      15.838856512259834
[1, 18, 3]      15.83556195932988
[1, 18, 4]      15.321434181036205
[1, 18, 5]      13.799004771741249
[1, 18, 6]      15.230707531144558
[1, 19, 0]      15.064448713868138
[1, 19, 1]      15.184492458519726
[1, 19, 2]      15.357075236500245
[1, 19, 3]      15.806505779916119
[1, 19, 4]      15.181602676051044
[1, 19, 5]      13.737655633714054
[1, 19, 6]      15.333196689466162
[1, 2, 0]       17.719516271533564
[1, 2, 1]       16.653553023303157
[1, 2, 2]       15.396756474519997
[1, 2, 3]       15.725557996112741
[1, 2, 4]       15.440661491541286
[1, 2, 5]       14.96469021695353
[1, 2, 6]       22.73650211790674
[1, 20, 0]      15.38147142708661
[1, 20, 1]      15.282768484545876
[1, 20, 2]      15.574260541230966
[1, 20, 3]      15.78791569374111
[1, 20, 4]      15.165278225817401
[1, 20, 5]      14.55086849557509
[1, 20, 6]      16.05652368409613
[1, 21, 0]      15.978846513350177
[1, 21, 1]      15.682161157934564
[1, 21, 2]      15.759659172121506
[1, 21, 3]      16.408154703040964
[1, 21, 4]      15.53635399306004
[1, 21, 5]      14.78243890060351
[1, 21, 6]      16.429218683100896
[1, 22, 0]      16.56790387709699
[1, 22, 1]      16.135728800168657
[1, 22, 2]      15.896375080256153
[1, 22, 3]      16.46820566457765
[1, 22, 4]      15.626564358000117
[1, 22, 5]      15.12343279486087
[1, 22, 6]      16.633329771485958
[1, 23, 0]      17.64719986594566
[1, 23, 1]      17.088389679197615
[1, 23, 2]      16.42834966012019
[1, 23, 3]      16.71372463316476
[1, 23, 4]      15.838564924691722
[1, 23, 5]      15.314807169882616
[1, 23, 6]      17.937987911385125
[1, 3, 0]       18.797614294269355
[1, 3, 1]       17.504767459029665
[1, 3, 2]       16.95794614331468
[1, 3, 3]       17.382108248058888
[1, 3, 4]       16.38426781057979
```

.....

....

```
[6, 4, 1]      20.34873959398353
[6, 4, 2]      20.50594324084606
[6, 4, 3]      20.47393085441388
[6, 4, 4]      19.93230409052137
[6, 4, 5]      17.177113617752504
[6, 4, 6]      17.160783623789715
[6, 5, 0]      21.28951760015908
[6, 5, 1]      18.27839578207133
[6, 5, 2]      18.535080427043987
[6, 5, 3]      19.574097962865523
[6, 5, 4]      20.595584381684304
[6, 5, 5]      20.50048413732902
[6, 5, 6]      20.694516962284176
[6, 6, 0]      16.344082818432714
[6, 6, 1]      14.433547435683215
[6, 6, 2]      14.786332033232927
[6, 6, 3]      15.27862272058213
[6, 6, 4]      16.038412687284865
[6, 6, 5]      18.14070623729715
[6, 6, 6]      21.122064983671518
[6, 7, 0]      15.372120455979731
[6, 7, 1]      14.3695591639776
[6, 7, 2]      14.431602173641124
[6, 7, 3]      14.6568928101514
[6, 7, 4]      14.776007032378336
[6, 7, 5]      15.366026978969314
[6, 7, 6]      18.74945233380108
[6, 8, 0]      15.437935202166893
[6, 8, 1]      15.191213893732293
[6, 8, 2]      15.262540927750129
[6, 8, 3]      15.091789660540393
[6, 8, 4]      15.22343207202625
[6, 8, 5]      14.709619247222852
[6, 8, 6]      15.881945982849937
[6, 9, 0]      16.211695392474443
[6, 9, 1]      15.74264209950844
[6, 9, 2]      15.763047256856828
[6, 9, 3]      15.793481283810515
[6, 9, 4]      15.694123389694342
[6, 9, 5]      13.891624417895672
[6, 9, 6]      14.797861061779749
```