TASK 4

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

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

This output was obtained from 1st database

```
[root@ip-172-31-62-227 ~] # python ql.py yellow_tripdata_2017-01.csv > ql.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/ql.root.20240409.145756.853555
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/ql.root.20240409.145756.853555/output
Streaming final output from /tmp/ql.root.20240409.145756.853555/output...
Removing temp directory /tmp/ql.root.20240409.145756.853555...
[root@ip-172-31-62-227 ~] # cat ql.txt
"2" 82859254.39585336
```

Vendor 2 has most trips and revenue generated is 82859254.39585336

B. Which pickup location generates the most revenue?

This output was obtained from 2nd database

```
[root@ip-172-31-62-227 ~]# vi q2.py
[root@ip-172-31-62-227 ~]# python q2.py yellow_tripdata_2017-02.csv > q2.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/q2.root.20240409.150348.904553
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/q2.root.20240409.150348.904553/output
Streaming final output from /tmp/q2.root.20240409.150348.904553/output...
Removing temp directory /tmp/q2.root.20240409.150348.904553...
[root@ip-172-31-62-227 ~]# cat q2.txt
"132" 10472838.379996026
[root@ip-172-31-62-227 ~]#
```

Pickup location 132 generate most revenue (10472838.38)

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

This output was obtained from 3rd database

```
[root@ip-172-31-62-227 ~] # vi q3.py
[root@ip-172-31-62-227 ~] # vi q3.txt
[root@ip-172-31-62-227 ~] # python q3.py yellow_tripdata_2017-03.csv > q3.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/q3.root.20240409.150943.147081
Running step 1 of 3...
Running step 2 of 3...
Running step 3 of 3...
job output is in /tmp/q3.root.20240409.150943.147081/output
Streaming final output from /tmp/q3.root.20240409.150943.147081/output...
Removing temp directory /tmp/q3.root.20240409.150943.147081...
[root@ip-172-31-62-227 ~] # cat q3.txt
"1" 6994699
"2" 3231928
"3" 53815
"4" 14999
[root@ip-172-31-62-227 ~] # ...
```

Different payment types used by customers with their count

- 1= Credit card 6994699 times
- 2= Cash 3231928 times
- 3= No charge 53815 times
- 4= Dispute 14999 times

Payment through credit card has done more often

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

This output was obtained from 4th database

```
root@ip-172-31-63-228 ~] # python q4.py yellow_tripdata_2017-04.csv > q4.txt To configs found; falling back on auto-configuration
No configs specified for inline runner
reating temp directory /tmp/q4.root.20240409.160125.397334
dunning step 1 of 1...
ob output is in /tmp/q4.root.20240409.160125.397334/output
treaming final output from /tmp/q4.root.20240409.160125.397334/output...
demoving temp directory /tmp/q4.root.20240409.160125.397334...
root@ip-172-31-63-228 ~]# cat q4.txt
       8.69859496124031
       46.56947758496023
       15.86489867520792
100"
       57.85808080808081
102"
       27.02078853046595
106"
       14.100608029318675
       14.39224446410785
       17.885714285714286
       11.4583333333333334
11"
        34.16747311827957
       8.989655172413793
112"
       13.645934065934068
       15.704456568326293
114"
       16.46732887097284
115"
       14.0958333333333333
        15.483265311788555
        15.8625000000000002
        13.819078947368425
118"
       13.650181159420288
```



```
1197 13.34063460809647
1109 12.962772966773
1.1.962772966773
1.2.962772966773
1.2.962772966773
1.2.962772966773
1.2.962676733
1.2.9626666666666
1.2.90 13.58651724317932
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026216
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.5743421026217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.57434217
1.2.90 14.5743421
```

```
| 12.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 | 13.1 |
```

```
5.19318060808080808

10.10518137522261

33* 14.10518137522261

34* 14.12223388080809

35* 26.12550551811

36* 14.3618227377500

37* 17.06511747851

38* 25.76678571428571

40* 13.2280585585555

40* 13.32805858585555

40* 13.32805858555555

40* 13.32805858555555

44* 14.281666666666667

44* 14.281666666666667

44* 28.1599173553713

45* 18.32337430819399999999

40* 14.02429150543955

58* 14.844666666666667

48* 14.84761804761906

54* 14.84761804761906

55* 31.63297872331333333333333

6 6 4.666666666667

57* 14.9515151515155

58* 5.79844444444444

6 18.9210663453166

6 4.61666666666667

6 4.6166666666667

6 4.61666666666667

6 4.61666666666667

6 4.61666666666667

6 4.6166666666667

6 4.61666666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.6166666666667

6 4.61666666666667

6 4.6166666666667

6 5 5 17.991346153846155

6 6 6 18.657041299932299

6 7 6 18.645821543105

6 7 7 12.6468821543105

6 6 7 13.4648821543105

6 7 13.1978557986074
```

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

This output was obtained from 5th database

```
7-46** 0.19954594[14505117
7-46** 0.19551530[67646
7-247** 0.0733894871747694
7-248** 0.19551530[67646
7-248** 0.19552550[6797943887
7-25** 0.19523576[97943887
7-25** 0.1052463118943497
7-25** 0.04328271250452847
7-25** 0.04328271250452847
7-25** 0.073281821350452847
7-25** 0.073281821350452847
7-25** 0.07328182183507
7-25** 0.073281818135367
7-26** 0.07328182480[2888
7-25** 0.1960100461517414
7-25** 0.1048264899511988
7-26** 0.1048264899511988
7-26** 0.1048264899511988
7-26** 0.1048264899511988
7-26** 0.1048264899511988
7-26** 0.1048264899511989
7-26** 0.104826489951989
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489951998
7-26** 0.104826489503124
7-26** 0.10480813213195460188
7-26** 0.10480813213195460188
7-26** 0.10480813213195460188
7-26** 0.10480813213195460188
7-26** 0.10480813213195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.1048081313195460188
7-26** 0.104808131319540188
7-26** 0.104808131319540188
7-26** 0.10480813319540188
7-26** 0.10480813131360184
7-36** 0.104808131331640188
7-37** 0.10480813133507
7-40** 0.104908131304344
7-40** 0.104908131304344
7-40** 0.104908131304344
```

```
"45" 0.098384617461719  
46" 0.1340712045748017  
0.03140712045748017  
0.03140712045748017  
0.031407120457480185866  
0.04" 0.058275880885866  
0.0507 0.058275880885866  
0.0507 0.05827586088586  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804576734864  
0.05085804586948648  
0.05085804586948648  
0.05085804586948648  
0.05085804586948648  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804586948  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.05085804645756267434137  
0.050858046457562674418048  
0.05085804645756267434137  
0.05085804645756267484187  
0.0508580464576626744187   
0.05085804645766267484187   
0.05085
```

```
"85" 0.5667389241017691
*66" 0.12371478220911764
*88" 0.12371478220911764
*88" 0.1655120658805206363
*99" 0.07975511622519114
*90" 0.11768850776024835
*99" 0.10308391768201007
*92" 0.056615381413745704
*93" 0.1071444344921823
*94" 0.1038808346582212626
*95" 0.07928530356245143
*96" 0.09955983258212626
*95" 0.0995598325818912
*97" 0.08668461512544295
*99" 0.06667388225420083
[root@ip-172-31-63-228 -]#
```

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)

This output was obtained from 5th database

```
[5,
[5,
[5,
         0]
1]
2]
3]
4]
5]
                      15.307352827781946
                      14.68863340021126
                      14.46050088621336
                      14.619068724817286
14.939835919422332
[5,
                      16.300197449100637
                      18.066767322725
[5,
[5,
[5,
[5,
[5,
[5,
[5,
         0]
1]
2]
3]
                      15.534355540424196
                      15.310034719953286
                      14.991955112543714
                      15.095595443793671
         4]
5]
6]
0]
1]
2]
3]
4]
                      15.144876788509523
                      14.67262836614724
15.685100979728976
                      15.79006959000546
[5,
[5,
                      15.809253019981295
                      15.764076749004406
                      15.912008246957104
                      13.814643752241684
```