**Task 4:** Implement MapReduce algorithms to execute operations utilizing the datasets available on your EMR Instance.

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

```
[hadoop@ip-172-31-65-22 ~]$ 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.20230711.080915.677828
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_a.hadoop.20230711.080915.677828/output
Streaming final output from /tmp/mrtask_a.hadoop.20230711.080915.677828/output...
Removing temp directory /tmp/mrtask_a.hadoop.20230711.080915.677828...
[hadoop@ip-172-31-34-221 ~]$ ls
mrtask_a.py  mrtask_a.txt  mrtask_b.py  mrtask_c.py  mrtask_d.py  mrtask_e.py  mrtask_f.py  yellow_tripdata_2017-05.csv
[hadoop@ip-172-31-65-22 ~]$ cat mrtask_a.txt
"2"     92896777.54522054
```

VeriFone Inc., the second vendor, has the highest number of trips, and the total revenue generated from their transactions amounts to $92,896,777.545.

B. Which pickup location generates the most revenue?

```
[hadoop@ip-172-31-65-22 ~]$ 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.20230711.081820.834912
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_b.hadoop.20230711.081820.834912/output
Streaming final output from /tmp/mrtask_b.hadoop.20230711.081820.834912/output...
Removing temp directory /tmp/mrtask_b.hadoop.20230711.081820.834912...
[hadoop@ip-172-31-65-22 ~]$ cat mrtask_b.txt
"132"   14040591.220016211
```

The pickup location with ID 132 yields the highest revenue, generating a total of $14,040,591.22.

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

```
[hadoop@ip-172-31-65-22 ~]$ python mrtask_c.py yellow_tripdata_2017-05.csv > mr
task_c.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_c.hadoop.20230711.082928.304167
Running step 1 of 3...
Running step 2 of 3...
Running step 3 of 3...
job output is in /tmp/mrtask_c.hadoop.20230711.082928.304167/output
Streaming final output from /tmp/mrtask_c.hadoop.20230711.082928.304167/output..
.
Removing temp directory /tmp/mrtask_c.hadoop.20230711.082928.304167...
[hadoop@ip-172-31-34-221 ~]$ cat mrtask_c.txt
"1"     6780947
"2"     3250362
"3"     55027
"4"     15791
```

Transactions processed via credit card are more prevalent.

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

```
[hadoop@ip-172-31-65-22. ~]$ 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.20230711.083318.497680
Running step 1 of 1...
job output is in /tmp/mrtask_d.hadoop.20230711.083318.497680/output
Streaming final output from /tmp/mrtask_d.hadoop.20230711.083318.497680/output...
Removing temp directory /tmp/mrtask_d.hadoop.20230711.083318.497680...
[hadoop@ip-172-31-65-22. ~]$ cat mrtask_d.txt
"1"     1.8963513513513515
"10"    14.784510175106485
"100"   2.3583416159916184
"101"   4.407368421052632
"102"   4.257391304347826
"105"   5.123333333333333
"106"   2.9348579970104627
"107"   2.1812257303114073
"108"   4.271621621621622
"109"   9.471666666666666
"11"    3.6414545454545455
"111"   4.137777777777778
"112"   3.204110059009193
"113"   2.1453345228565595
"114"   2.4499525294398175
"115"   7.224117647058824
"116"   3.2416491850895612
"117"   3.902
"118"   5.0325
```

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

```
[hadoop@ip-172-31-65-22 ~]$ python mrtask_e.py yellow_tripdata_2017-05.csv > mrtask_e.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_e.hadoop.20230711.084941.508793
Running step 1 of 1...
job output is in /tmp/mrtask_e.hadoop.20230711.084941.508793/output
Streaming final output from /tmp/mrtask_e.hadoop.20230711.084941.508793/output...
Removing temp directory /tmp/mrtask_e.hadoop.20230711.084941.508793...
[hadoop@ip-172-31-65-22 ~]$ cat mrtask_e.txt
"1"     0.11961918264336044
"10"    0.1031314953272488
"100"   0.10054811858030266
"101"   0.1538484771604267
"102"   0.09686867923894887
"105"   0.07175925925925926
"106"   0.11203391656570175
"107"   0.11929564967810695
"108"   0.12120884627098828
"109"   0.23737785016286642
"11"    0.058159884648828326
"111"   0.077659754354142
"112"   0.10900187462682069
"113"   0.11769752457713911
"114"   0.11571668143567702
"115"   0.1014877461571702
"116"   0.09140144834181262
"117"   0.029865191166550988
"118"   0.03324247058041353
```

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).

```
[hadoop@ip-172-31-65-22 ~]$ python mrtask_f.py yellow_tripdata_2017-05.csv > mrtask_f.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_f.ec2-user.20230711.092853.372635
Running step 1 of 1...
job output is in /tmp/mrtask_f.ec2-user.20230711.092853.372635/output
Streaming final output from /tmp/mrtask_f.ec2-user.20230711.092853.372635/output...
Removing temp directory /tmp/mrtask_f.ec2-user.20230711.092853.372635...
[hadoop@ip-172-31-65-22 ~]$ cat mrtask_f.txt
[5, 0, 0]       19.09149097120672
[5, 0, 1]       19.42010595117413
[5, 0, 2]       17.686158986637274
[5, 0, 3]       17.25599944664103
[5, 0, 4]       18.099999027301898
[5, 0, 5]       17.105128026116333
[5, 0, 6]       15.658808411947584
[5, 1, 0]       17.87588900705367
[5, 1, 1]       19.429909785928686
[5, 1, 2]       17.373021160869822
[5, 1, 3]       16.436848288253056
[5, 1, 4]       18.217479700823546
[5, 1, 5]       16.275959370214274
[5, 1, 6]       15.141293067405417
[5, 10, 0]      16.241741228653577
```