

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

All the 6 files provided in the provided in the problem statement were used for generating the output

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

```
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % ls
  input      mrtask_a.py      mrtask_b.py      mrtask_c.py      mrtask_d.py      mrtask_e.py      mrtask_f.py
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % ls input
yellow_tripdata_2017-01.csv      yellow_tripdata_2017-03.csv      yellow_tripdata_2017-05.csv
yellow_tripdata_2017-02.csv      yellow_tripdata_2017-04.csv      yellow_tripdata_2017-06.csv
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_a.py input > mrtask_a_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_a.hemanthreddyk.20230925.195342.333346
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_a.hemanthreddyk.20230925.195342.333346/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_a.hemanthreddyk.20230925.195342.333346/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_a.hemanthreddyk.20230925.195342.333346...
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_a_result.txt
 2      525037658.1383097
○ (base) hemanthreddyk@Hemanths-MacBook-Air MRJob %
```

Vendor-2 i.e VeriFone Inc. is the vendor with most trips and it generated total revenue of \$ 525037658.1383097

B. Which pickup location generates the most revenue?

```
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_b.py input > mrtask_b_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_b.hemanthreddyk.20230925.200401.549215
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_b.hemanthreddyk.20230925.200401.549215/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_b.hemanthreddyk.20230925.200401.549215/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_b.hemanthreddyk.20230925.200401.549215...
● (base) hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_b_result.txt
 132      77196812.24
○ (base) hemanthreddyk@Hemanths-MacBook-Air MRJob %
```

Pick up location 132 generates most revenue of \$ 77196812.24

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

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_c.py input > mrtask_c_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_c.hemanthreddyk.20230925.201015.349043
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_c.hemanthreddyk.20230925.201015.349043/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_c.hemanthreddyk.20230925.201015.349043/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_c.hemanthreddyk.20230925.201015.349043...
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_c_result.txt
"Credit card" 39754212
"Cash" 18832370
"No charge" 306912
"Dispute" 88794
"Unknown" 3
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob %
```

Payment Type	Count
Credit card	39754212
Cash	18832370
No charge	306912
Dispute	88794
Unknown	3

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

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_d.py input > mrtask_d_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_d.hemanthreddyk.20230925.203647.653100
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_d.hemanthreddyk.20230925.203647.653100/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_d.hemanthreddyk.20230925.203647.653100/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_d.hemanthreddyk.20230925.203647.653100...
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob %
```

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_d_result.txt
1      491.18705207835643
2      2289.9565217391305
3      1299.7575757575758
4      899.679533415981
5      686.6923076923077
6      450.11538461538464
7      795.8500817600714
8      1049.4106145251396
9      5069.413407821229
10     2993.996715628136
11     1030.1524390243903
12     1461.5228064141697
13     1170.1200095775785
14     1012.0196145905024
15     872.5137614678899
16     1315.3
17     841.7990659280982
18     859.5884892086331
19     747.044776119403
20     846.5162094763092
21     1229.890625
22     1315.2108036890645
23     765.9130434782609
24     813.3275934954767
25     929.0090624612716
26     718.9853281853282
27     526.7272727272727
28     1690.4115553121578
29     1314.3799621928167
30     1506.095238095238
31     1494.9198473282443
32     1208.4641148325359
33     1085.7095208946087
34     951.8898221343874
35     1166.1373844121533
36     1012.40791015625
37     967.6872436529775
38     3059.3202614379084
39     1094.061814556331
40     994.1455245428297
```

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
50	892.0096256834727
51	976.6214285714286
52	1044.7077185641929
53	1028.8467532467532
54	1021.9478827361563
55	1485.292156862745
56	1206.7722772277227
57	865.6976744186046
58	388.15384615384613
59	1009.0
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
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
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
90	850.6428322112844
91	1251.828839390387
92	1051.3910096325367
93	2039.3028281363306
94	1160.1567164179105
95	1071.6730681298582
96	943.9640287769785
97	941.0709229177274
98	735.3614457831326
99	763.6
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
110	191.0
111	698.0794701986755
112	866.7667716535433
113	899.1148330113725
114	955.2777877786825
115	846.0082987551867
116	938.2699303995785
117	1159.0581395348838
118	735.5904761904762
119	979.2593184238551
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
130	2110.0042534139243
131	846.7052341597796
132	2628.849705808318
133	996.3551984257133
134	950.9126246990024
135	1086.227626459144
136	683.5790921595599
137	827.1185944225588
138	2239.9364882573964
139	3579.403846153846
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	1000.5063450558392
149	1003.7186440677966
150	1109.5384615384614
151	793.2153074522814
152	849.5668384259411
153	813.779388083736
154	1467.2142857142858
155	1621.730303030303
156	1164.2621359223301
157	1230.2581005586592
158	994.7255508393539
159	852.2906389301635
160	1203.9138655462184
161	935.282605696834
162	906.2646635610922
163	951.7649994006813

164	925.6586131402529
165	1107.496115427303
166	890.6641412169023
167	899.6060388209921
168	771.3360398114196
169	947.9101899827289
170	882.2945248296076
171	772.3879310344828
172	958.6944444444445
173	879.1821631878557
174	784.0575539568346
175	767.75
176	1846.357142857143
177	1149.6396337475474
178	408.5500562429696
179	870.1715711625983
180	1802.2665148063782
181	950.9730341840722
182	836.2656488549618
183	724.8125
184	1002.8823529411765
185	1126.061320754717
186	1008.7132336796793
187	2339.478260869565
188	968.026302283808
189	909.9792762301483
190	1157.1567398119123
191	985.7132701421801
192	1094.9559543230016
193	850.7572725465509
194	1499.3031927857664
195	1249.815140345903
196	1059.4801938474504
197	939.7597208374875
198	794.982245894363
199	1047.095238095238
200	769.9037558685446
201	592.943661971831
202	972.7012844036698
203	1629.8995983935743
204	213.07142857142858

205	1506.3287037037037
206	816.4489795918367
207	507.9640265067845
208	1594.5827160493827
209	1148.1146539848555
210	1076.881690140845
211	1005.8916971033002
212	1267.416230366492
213	983.2531120331951
214	599.8636363636364
215	2912.769230769231
216	1705.113497745956
217	1054.9078863346106
218	1240.4111675126903
219	2727.1610715496777
220	724.2466926070039
221	862.4206008583691
222	1782.7380952380952
223	896.0254384374272
224	822.7864531678853
225	895.5908084163898
226	993.1989910155032
227	873.0696774193549
228	950.0332378223495
229	796.9896034262334
230	1015.8439600398245
231	1010.6449346098719
232	1027.2149090383814
233	915.019666675366
234	900.176481729967
235	1015.2363344051447
236	764.8343850963822
237	741.564881206982
238	774.4370287985316
239	787.5652921190596
240	844.561797752809
241	912.5617792421747
242	853.8853006681514
243	980.808996758637
244	1009.9856078617847
245	520.6333333333333

246	943.6957956238367
247	964.7215102778017
248	943.1546840958606
249	896.4203932478235
250	1005.76401179941
251	712.516129032258
252	1189.6425
253	1336.659574468085
254	1142.6480446927374
255	979.2834121416406
256	966.6904458775014
257	925.2052060737527
258	1229.0232858990944
259	909.4985163204748
260	930.4427896240932
261	1310.9905594478348
262	770.6135599773738
263	729.5529502822361
264	887.7241473468238
265	570.1504814587174

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

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_e.py input > mrtask_e_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_e.hemanthreddyk.20230925.201334.641227
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_e.hemanthreddyk.20230925.201334.641227/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_e.hemanthreddyk.20230925.201334.641227/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_e.hemanthreddyk.20230925.201334.641227...
```

```
[base] hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_e_result.txt
30      0.2561075875029364
104     0.2000665778961385
187     0.17978913134704155
109     0.17861970356364315
5       0.173567645642754
172     0.17223686242471645
117     0.16546115321544122
176     0.1588995526720869
201     0.1516472849404958
58      0.1439182666523652
199     0.14024694882592245
122     0.13240255677287172
138     0.13191299589443295
52      0.12903601946206134
175     0.12772307760082824
210     0.1269629102823797
191     0.12487168062809831
87      0.12471673086142389
125     0.12389238738091936
16      0.12347342980892896
84      0.12333321182517408
13      0.12263264632355549
178     0.1219478101393929
194     0.12129343435403593
33      0.12071128564971116
162     0.1206663239696391
40      0.12031726919563338
54      0.12031650579198613
234     0.12013283079525545
249     0.1199021364670282
107     0.1197475302735505
246     0.11961317882571391
1       0.11931462821703152
23      0.11893642866783642
231     0.1188786215233841
113     0.11854772693410352
170     0.11808315282263399
79      0.11800406819814163
252     0.1179196668618503
114     0.11737171385742014
118     0.11728963561178403
66      0.11716659766185752
255     0.11685699515469598
88      0.11647722217030364
158     0.11639038989042694
184     0.11633468693321347
15      0.11600468842514833
233     0.11587691604709018
224     0.11526880032062324
148     0.11466547302900158
90      0.11445563465266693
68      0.11443748108204026
209     0.11438950713881407
143     0.11408622408799579
211     0.11399355397892771
258     0.11397415106035942
263     0.11388129538448295
161     0.11359049234402926
64      0.11356366975963518
262     0.11338221978049286
```

262	0.11338221978049286
257	0.11311365670211662
181	0.11303400885962434
144	0.11283013521688436
166	0.11278675284131413
137	0.11277472565917154
76	0.11266074016140304
264	0.11265242617785655
229	0.11256386704360667
142	0.11242578567775695
265	0.11224718082319061
163	0.11222164046414991
164	0.11217684253869109
96	0.11215306381581801
189	0.1119751542864053
106	0.11182299192052278
239	0.11135965083840464
140	0.11118444211887941
236	0.11108195526441147
123	0.11055898094072138
238	0.11044770763087772
80	0.11027187700901675
93	0.11009789332101685
237	0.10990442459583612
186	0.10977067019772932
112	0.10939659242523363
86	0.10935811216524559
25	0.10931197115094941
50	0.10916652254420693
8	0.10862824529869514
256	0.1085367377873744
215	0.10829221671091675
70	0.10788380927811908
141	0.10768193607233408
261	0.10766619243052909
10	0.10761668826560974
230	0.1076097076743779
134	0.10749874267013107
151	0.107367785549714
4	0.10684873200928192
34	0.10658808178106198
48	0.10627040728714153
63	0.10608349633094598
24	0.105908919111111572
36	0.10580489050473628
2	0.10551996408687954
65	0.10543895761086589
180	0.1052369675037582
43	0.10492270664851001
216	0.10479778546735072
195	0.10463665842928394
130	0.10459971957672498
150	0.10406209924668995
132	0.10404480517972046
101	0.10391200650224515
200	0.10311919266153374
221	0.10296610645201129
67	0.10270469503747051
28	0.10225113334394048
128	0.10195419121098444
100	0.10176112040920149

100	0.10176112040920149
111	0.10175458286569397
190	0.10147578909371861
157	0.10078849415880711
35	0.10022554243576684
219	0.10013130970242769
71	0.0997605842419586
49	0.09943300939914784
45	0.0993427451689479
213	0.09930916492713449
37	0.09921178378979299
56	0.0983778991554337
97	0.09860378966752917
115	0.09805645544907131
218	0.09787185841442558
228	0.0972540407536761
202	0.09722093931447452
77	0.09702969688085884
223	0.09657319511944328
39	0.09651707726856314
243	0.09649751817937215
160	0.0964770581925178
57	0.09639556671648825
155	0.09633375755470429
145	0.09530064643710169
55	0.09470551332740033
222	0.09461551127943359
253	0.09444941923267383
133	0.09428655857404512
244	0.09413161238216916
102	0.09286325213174114
139	0.09244706142211938
61	0.09216038290044067
31	0.09122118342582707
135	0.09112992092386231
154	0.09108097210043588
121	0.09104834906722978
197	0.0905755050511818
14	0.09044594848522654
9	0.09026115108356463
116	0.09022031542010774
75	0.09013747412606656
41	0.09013710776490726
89	0.08958574112049165
51	0.08952152479566317
203	0.0892237612909564
72	0.0886537969202704
146	0.08853658145675274
127	0.08838814791011232
12	0.08816912666410132
245	0.0881231334711693
27	0.0880988704482764
198	0.08806313330584344
156	0.08792537666968181
98	0.08788658752859362
259	0.08717567117240324
196	0.08616922680964258
108	0.08561073480894169
226	0.08554438430856263
179	0.08477276212070443
124	0.08450445139029458

124	0.08450445139029458
232	0.08364178523168476
206	0.08357280929682974
131	0.08338783482815322
95	0.08335693988164575
152	0.08319834814829409
17	0.0831753690145568
120	0.08303252748719826
205	0.0830228408954791
251	0.08299641950457239
21	0.08297404683405192
105	0.08289275441720315
225	0.08262074885724466
46	0.08234332088746259
149	0.08233859572151554
81	0.08090785981257786
227	0.0806057192559692
171	0.08052993314534214
19	0.07943494943114211
7	0.07831986411657535
91	0.07770689984709486
214	0.07745375602878149
53	0.07740291031997903
208	0.07718716885014358
62	0.07661527162660442
177	0.07641696352884514
73	0.07567600767754318
220	0.07548055451382425
20	0.0753977936064434
185	0.07530969983194839
212	0.07459991612034526
188	0.07438531115226607
217	0.07433903922530674
207	0.0741642528127881
247	0.07294934365907783
126	0.07223106849013398
74	0.07135080378568823
242	0.07134362315468344
192	0.07124383780662928
240	0.07105544591130963
42	0.0703624108361557
260	0.07000191673138224
3	0.06985350809560525
165	0.06797416237704271
119	0.06778234295101977
11	0.06611923513377356
85	0.06602854511265939
26	0.06577859806429147
29	0.06573652675700856
92	0.06480851959104635
82	0.06474358825817085
169	0.06451074143663856
44	0.06402542001826439
129	0.06331017333926078
22	0.06303381261429238
174	0.06278143350842436
204	0.06247952716655986
241	0.06142723563688842
167	0.06122252383032549
32	0.06109787327761727
153	0.0605637325357794

153	0.0605637325357794
99	0.06008865539320309
173	0.06008145804370091
183	0.05853409985743476
38	0.05791427287264456
83	0.05736768211329448
250	0.05704400870673692
248	0.05406105607423324
18	0.05241971650085568
235	0.05148549299609722
168	0.04962075987817941
182	0.04956118360347683
78	0.04945682121018833
6	0.04831818885848519
69	0.04770731790200222
159	0.046638005623141514
193	0.04602694182348886
254	0.0453909426844164
147	0.04460414952837091
136	0.040904378620177845
60	0.04087201889401169
47	0.03560717179556588
94	0.032612401099441864
59	0.011954905847373637
110	0.0

(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % █

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).

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % python mrtask_f.py input > mrtask_f_result.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_f.hemanthreddyk.20230925.211254.040713
Running step 1 of 2...
Running step 2 of 2...
job output is in /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_f.hemanthreddyk.20230925.211254.040713/output
Streaming final output from /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_f.hemanthreddyk.20230925.211254.040713/output...
Removing temp directory /var/folders/gj/1f5gtygj3zv0ffsfvqywdj3r0000gn/T/mrtask_f.hemanthreddyk.20230925.211254.040713...
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob %
```

```
(base) hemanthreddyk@Hemanths-MacBook-Air MRJob % cat mrtask_f_result.txt
[["April", "weekday", "day"], [16.33992516832514, 3879800]
[["April", "weekday", "night"], [16.83881361847111, 2868555]
[["April", "weekend", "day"], [15.468942474531282, 1651279]
[["April", "weekend", "night"], [16.429201767053435, 1647501]
[["February", "weekday", "day"], [15.678661888195492, 3717426]
[["February", "weekday", "night"], [16.183219751484117, 2767875]
[["February", "weekend", "day"], [14.465304659335414, 1324513]
[["February", "weekend", "night"], [15.520779360795439, 1359961]
[["January", "weekday", "day"], [15.435724925965479, 4057046]
[["January", "weekday", "night"], [16.03961894503744, 2830564]
[["January", "weekend", "day"], [14.40274899156903, 1398924]
[["January", "weekend", "night"], [15.869888470663591, 1424286]
[["June", "weekday", "day"], [16.831393533058144, 4141694]
[["June", "weekday", "night"], [16.892620694858284, 3095442]
[["June", "weekend", "day"], [15.94084461289684, 1199461]
[["June", "weekend", "night"], [16.08689670421361, 1220396]
[["March", "weekday", "day"], [16.41823260804668, 4258174]
[["March", "weekday", "night"], [16.698247022584223, 3252119]
[["March", "weekend", "day"], [14.899300142798582, 1396213]
[["March", "weekend", "night"], [15.847244687684219, 1388935]
[["May", "weekday", "day"], [17.154349664704394, 4348838]
[["May", "weekday", "night"], [16.99594061861851, 3200653]
[["May", "weekend", "day"], [15.605799986023365, 1297598]
[["May", "weekend", "night"], [16.048686932392915, 1255038]
```