

MapReduce Tasks:

Task 4. Using MapReduce to perform data analysis on files downloaded to the EMR instance

- ❖ Initially I set up the environment using commands `sudo yum update`, `pip3 install mrjob` etc, once set up I proceeded
 - ❖ I wrote python scripts, moved them to hadoop using WinSCP, gave python scripts execution permission using command `chmod 700`
 - ❖ General code used for running the MapReduce jobs:
Python 'mr_script.py' 'input' > 'output file'
 - ❖ I used [yellow_tripdata_2017-05.csv](#) dataset for this task
- a. Which vendors have the most trips, and what is the total revenue generated by that vendor?

Vendor 2 has the most trips, total revenue generated is 92896777.5452

```
[hadoop@ip-172-31-60-246 ~]$ 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.20230906.144349.570870
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_a.hadoop.20230906.144349.570870/output
Streaming final output from /tmp/mrtask_a.hadoop.20230906.144349.570870/output...
Removing temp directory /tmp/mrtask_a.hadoop.20230906.144349.570870...
[hadoop@ip-172-31-60-246 ~]$ 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-172-31-60-246 ~]$ cat mrtask_a.txt
"2" 92896777.54522054
[hadoop@ip-172-31-60-246 ~]$
```

- b. Which pickup location generates the most revenue?

Pickup location 132 generates the most revenue : 14040591.2200

```
[hadoop@ip-172-31-60-246 ~]$ 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.20230906.145247.244760
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/mrtask_b.hadoop.20230906.145247.244760/output
Streaming final output from /tmp/mrtask_b.hadoop.20230906.145247.244760/output...
Removing temp directory /tmp/mrtask_b.hadoop.20230906.145247.244760...
[hadoop@ip-172-31-60-246 ~]$ ls
mrtask_a.py mrtask_a.txt mrtask_b.py mrtask_b.txt mrtask_c.py mrtask_d.py mrtask_e.py mrtask_f.py yellow_tripdata_2017-05.csv
[hadoop@ip-172-31-60-246 ~]$ cat mrtask_b.txt
"132" 14040591.220016211
[hadoop@ip-172-31-60-246 ~]$
```

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

The different payment types used by customers and their count are shown below:

Numeric code	Payment type	Count
1	Credit card	6780947
2	Cash	3250362
3	No charge	55027
4	Dispute	15791

```
[hadoop@ip-172-31-60-246 ~]$ python mrtask c.py yellow_tripdata_2017-05.csv > mrtask_c.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_c.hadoop.20230906.145839.932749
Running step 1 of 3...
Running step 2 of 3...
Running step 3 of 3...
job output is in /tmp/mrtask_c.hadoop.20230906.145839.932749/output
Streaming final output from /tmp/mrtask_c.hadoop.20230906.145839.932749/output...
Removing temp directory /tmp/mrtask_c.hadoop.20230906.145839.932749...
[hadoop@ip-172-31-60-246 ~]$ ls
mrtask_a.py mrtask_a.txt mrtask_b.py mrtask_b.txt mrtask_c.py mrtask_c.txt mrtask_d.py mrtask_e.py mrtask_f.py yellow_tripdata_2017-05.csv
[hadoop@ip-172-31-60-246 ~]$ cat mrtask_c.txt
"1" 6780947
"2" 3250362
"3" 55027
"4" 15791
[hadoop@ip-172-31-60-246 ~]$
```

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

The average trip time for different pickup locations and code used is shown below:

```
hadoop@ip-172-31-60-246:~$ 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.20230906.150447.118339
Running step 1 of 1...
job output is in /tmp/mrtask_d.hadoop.20230906.150447.118339/output
Streaming final output from /tmp/mrtask_d.hadoop.20230906.150447.118339/output...
Removing temp directory /tmp/mrtask_d.hadoop.20230906.150447.118339...
hadoop@ip-172-31-60-246:~$ ls
mrtask_a.py mrtask_a.txt mrtask_b.py mrtask_b.txt mrtask_c.py mrtask_c.txt mrtask_d.py mrtask_d.txt mrtask_e.py mrtask_f.py yellow_tripdata_2017-05.csv
hadoop@ip-172-31-60-246:~$ cat mrtask_d.txt
"1" 12.681244881244885
"10" 56.4414970815586
"100" 16.301056898951547
"101" 10.8640350877193
"102" 15.480072463768114
"105" 22.727777777777778
"106" 14.046902507889056
"107" 14.914703800523975
"108" 17.747474747474745
"109" 13.047222222222222
"11" 11.318787878787878
"111" 15.347685185185185
"112" 15.342635286583413
"113" 15.655663780443364
"114" 16.265089572862763
"115" 9.849019607843136
"116" 15.213805066967888
"117" 9.020000000000001
"118" 16.645833333333334
"119" 14.638700564971753
"12" 25.11584025215252
"120" 12.665972222222223
"121" 16.845833333333333
"122" 11.954761904761904
"123" 12.698253968253967
"124" 35.719855072463766
"125" 16.582152788143137
"126" 32.175581395348836
"127" 17.629476052716164
"128" 15.579292929292929
"129" 14.331578947368426
"13" 20.914534001214356
"130" 37.9820691108302
"131" 11.690000000000003
"132" 46.805228740818975
"133" 17.532022849462365
"134" 16.221703296703296
"135" 15.368173258003768
```

```
hadoop@ip-172-31-60-246:~$ cat mrtask_d.txt
"135" 15.368173258003768
"136" 9.693303571428572
"137" 14.412591446828932
"138" 41.79266819772832
"139" 18.82620512820513
"14" 18.910311918063318
"140" 14.544783434852864
"141" 12.887731943471524
"142" 14.724678434186814
"143" 13.87881130774534
"144" 17.422410667367306
"145" 13.519851029794044
"146" 15.77948824343021
"147" 13.008051643192489
"148" 16.99422069527864
"149" 13.13448276820688
"15" 14.868589743589745
"150" 22.067647058823532
"151" 13.72286961525267
"152" 15.034273318872025
"153" 12.776791277258567
"154" 22.930303030303033
"155" 23.8472222222222218
"156" 17.777777777777778
"157" 24.96510297482838
"158" 17.056993321982915
"159" 12.977084615384616
"16" 11.23930817610063
"160" 17.381399631675876
"161" 16.501390496486135
"162" 15.5755374379993
"163" 17.007285407402943
"164" 16.135913909640667
"165" 23.146643518518516
"166" 16.128181641420127
"167" 11.731709401709402
"168" 11.912370135587249
"169" 12.008282828282828
"17" 14.360986719350239
"170" 15.23280819184737
"171" 11.867179487179486
"172" 29.295833333333333
"173" 12.247398589065257
"174" 12.091085271317832
"175" 11.432758620689656
"176" 0.6
"177" 20.955774278215223
"178" 8.291562500000001
"179" 15.1499705882353
```

```
hadoop@ip-172-31-60-246:~$
"179" 15.1499705882353
"180" 12.653225806451612
"180" 27.183506944444446
"181" 15.849603785417022
"182" 12.2456
"183" 11.927777777777779
"184" 7.346666666666666
"185" 15.095833333333333
"186" 17.3607280512866
"187" 1.425
"188" 15.907792044810168
"189" 16.631431947577614
"190" 10.768354430379746
"190" 26.461985335589397
"191" 13.486574074074074
"192" 13.586842105263157
"193" 12.075103464819935
"194" 27.354013503375846
"195" 22.571194503171245
"196" 18.181564465408805
"197" 18.036302681992343
"198" 13.806195462478186
"200" 42.2
"200" 17.141091954022986
"200" 11.855392156862747
"201" 0.49374999999999997
"202" 15.435855263157896
"203" 19.28479532163743
"204" 0.47999999999999987
"205" 18.713008130081302
"206" 11.648809523809524
"207" 8.778993536472758
"208" 15.355907172995781
"209" 20.52999086340794
"210" 16.197798742138364
"210" 18.476602564102567
"211" 17.805909853173578
"212" 24.377777777777773
"213" 20.40786163522013
"214" 8.220833333333333
"215" 56.68466165413534
"216" 34.080833333333333
"217" 18.35507590132827
"218" 17.684177215189873
"219" 53.02001915708813
"220" 22.21085626911315
"220" 11.665315315315317
"221" 19.811904761904763
"222" 19.394444444444442
```

```
hadoop@ip-172-31-60-246:~$
"201" 0.49374999999999997
"202" 15.435855263157896
"203" 19.28479532163743
"204" 0.47999999999999987
"205" 18.713008130081302
"206" 11.648809523809524
"207" 8.778993536472758
"208" 15.355907172995781
"209" 20.52999086340794
"210" 16.197798742138364
"210" 18.476602564102567
"211" 17.805909853173578
"212" 24.377777777777773
"213" 20.40786163522013
"214" 8.220833333333333
"215" 56.68466165413534
"216" 34.080833333333333
"217" 18.35507590132827
"218" 17.684177215189873
"219" 53.02001915708813
"220" 22.21085626911315
"220" 11.665315315315317
"221" 19.811904761904763
"222" 19.394444444444442
"223" 17.158481387478847
"224" 14.423218777619141
"225" 15.393991912189485
"226" 17.062662041086035
"227" 15.537459546925568
"228" 18.118900966183574
"229" 13.729634370452187
"230" 3.605128205128205
"230" 18.16775177926668
"231" 17.842905702290793
"232" 17.80712673224909
"233" 16.130618672379082
"234" 15.740532219114375
"235" 21.51443736730361
"236" 13.552552898225786
"237" 12.943728784429947
"238" 13.635731925509624
"239" 13.766637412862702
"240" 14.086371798958675
"240" 16.918181818181818
"241" 15.555228758169934
"242" 13.33582822085891
"243" 17.764382566585954
"244" 17.97972746905131
"245" 23.330555555555552
```

```
hadoop@ip-172-31-60-246:~$
"245" 23.330555555555555552
"246" 16.162821706410472
"247" 16.876773493573694
"248" 13.5583333333333332
"249" 15.335284107155961
"250" 17.17143186440893
"251" 14.828939393939397
"252" 2.1999999999999997
"253" 22.431190476190476
"254" 25.6805555555555554
"255" 11.159439150439153
"256" 18.5869146586808
"257" 15.882166210462294
"258" 16.62408789386401
"259" 26.296310432569975
"260" 10.456666666666669
"261" 11.717295597464279
"262" 17.798491342362418
"263" 22.742345329876898
"264" 13.453498021610727
"265" 12.558972943698416
"266" 15.660701879388133
"267" 10.09781482973892
"268" 9.104166666666668
"269" 33.768
"270" 57.221774193548384
"271" 17.91923076923077
"272" 14.598923232323233
"273" 22.172916666666667
"274" 45.23925925925926
"275" 19.48258199014998
"276" 15.771012544802865
"277" 12.903929392469663
"278" 16.217423442449842
"279" 15.6778518308733
"280" 105.58416666666668
"281" 17.3040625
"282" 15.94539005698556
"283" 15.923511762519249
"284" 13.047298730655545
"285" 13.060410811865738
"286" 16.67567976383065
"287" 9.504166666666666
"288" 19.516391550098202
"289" 29.257407407407406
"290" 15.23495145631068
"291" 15.306237295695665
"292" 14.44308345010119
"293" 0.3
"294" 0.3
```

```
hadoop@ip-172-31-60-246:~$
"295" 0.3
"296" 15.42537099704824
"297" 11.96068376068376
"298" 18.217088839941262
"299" 20.58723958333333
"300" 14.719335511982568
"301" 47.3187265917603
"302" 23.806329113924047
"303" 15.616666666666667
"304" 5.333333333333333
"305" 19.616666666666664
"306" 7.790277777777777
"307" 12.961559139784944
"308" 15.251830481496228
"309" 13.879747675962818
"310" 65.07303921568626
"311" 11.29107142857143
"312" 18.156792744967923
"313" 18.57747506019952
"314" 13.053431372549017
"315" 15.859693387283983
"316" 19.92417624521073
"317" 13.984743997175142
"318" 26.37702202480104
"319" 12.455172413793102
"320" 17.463793103448275
"321" 22.728684210526315
"322" 12.525644805036702
"323" 13.519075992588938
"324" 18.701871980676327
"325" 16.11434108527132
"326" 11.448753894080996
"327" 15.650715220659471
"328" 22.969014641744553
"329" 16.722509667955716
"330" 10.016666666666667
"331" 15.163276362823948
"332" 16.827437223042836
"333" 5.575
"334" 15.723731138545956
"335" 14.700000000000001
"336" 22.17575468660486
"337" 23.09566049232271
"338" 17.87714632989595
"339" 153.69927536231884
"340" 14.588625022901905
"341" 23.28269720101781
"342" 14.84195171026157
"343" 38.06967213114755
```

```
hadoop@ip-172-31-60-246:~$  
"55" 47.3187265917603  
"56" 23.806329113924047  
"57" 15.616666666666667  
"58" 5.333333333333333  
"59" 19.616666666666664  
"60" 7.790277777777777  
"61" 12.961559139784944  
"62" 15.251830461496228  
"63" 13.879747675962818  
"64" 65.07303921568626  
"65" 11.29107142857143  
"66" 18.156792744967923  
"67" 18.57747606019952  
"68" 13.053431372549017  
"69" 15.859693387283983  
"70" 19.92417624521073  
"71" 13.984743997175142  
"72" 26.37702202480104  
"73" 12.458172413783102  
"74" 17.463793103448275  
"75" 22.723684210526315  
"76" 12.525644805036702  
"77" 13.519075992588938  
"78" 18.701871980676327  
"79" 16.1134408527132  
"80" 11.448753894080896  
"81" 15.650715220659471  
"82" 22.969314641744553  
"83" 16.722509667955716  
"84" 10.016666666666667  
"85" 15.163276362823948  
"86" 16.827437223042836  
"87" 5.575  
"88" 15.723731138545956  
"89" 14.700000000000001  
"90" 22.17575468660486  
"91" 23.09566049232271  
"92" 17.0771463299575  
"93" 153.69927536231884  
"94" 14.588625022901905  
"95" 23.28269720101781  
"96" 14.84195171026157  
"97" 38.06967213114755  
"98" 11.64738995230951  
"99" 18.848467302452317  
"00" 18.588888888888885  
"01" 16.08161856320673  
"02" 13.002500000000001  
[hadoop@ip-172-31-60-246 ~]$
```

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

The average tips to revenue ratio of the drivers for different pickup locations in sorted format are shown below:

```
PuTTY (inactive)  
[hadoop@ip-172-31-60-246 ~]$ 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.20230906.153139.645782  
Running step 1 of 1...  
Job output is in /tmp/mrtask_e.hadoop.20230906.153139.645782/output  
Streaming final output from /tmp/mrtask_e.hadoop.20230906.153139.645782/output...  
Removing temp directory /tmp/mrtask_e.hadoop.20230906.153139.645782...  
[hadoop@ip-172-31-60-246 ~]$ ls  
mrtask_a.py mrtask_e.txt mrtask_b.py mrtask_c.py mrtask_c.txt mrtask_d.py mrtask_d.txt mrtask_e.py mrtask_e.txt mrtask_f.py yellow_tripdata_2017-05.csv  
[hadoop@ip-172-31-60-246 ~]$ cat mrtask_e.txt  
"1" 0.11961518264336044  
"10" 0.1031314853272488  
"100" 0.10054811858030266  
"101" 0.1538484771604267  
"102" 0.09686867923894887  
"105" 0.0717592592925926  
"106" 0.1120391656570115  
"107" 0.11929564967810695  
"108" 0.12120884627098828  
"109" 0.23737785016286642  
"11" 0.058159884648828326  
"111" 0.07765954394142  
"112" 0.10900187462682069  
"113" 0.11769752457713911  
"114" 0.11571668143567702  
"115" 0.1014877461571702  
"116" 0.09140144834191262  
"117" 0.029865191166550988  
"118" 0.03324247058041353  
"119" 0.06439098561727422  
"12" 0.09033344274559964  
"120" 0.1448542649668087  
"121" 0.06025386778698323  
"122" 0.035274937009041045  
"123" 0.0750228876958997  
"124" 0.09886020334918795  
"125" 0.1223467893609015  
"126" 0.05493403420192623  
"127" 0.08833111068775  
"128" 0.10511320192841678  
"129" 0.06647065796213625  
"13" 0.12390141102889671  
"130" 0.1078558366219219  
"131" 0.10384807204137783  
"132" 0.1025660513460536  
"133" 0.09147281347725217  
"134" 0.10314438578934823  
"135" 0.0785122214655781
```

```
PutTY (inactive)
"128" 0.0785122214655781
"136" 0.032600578121018825
"137" 0.11253077758026406
"138" 0.1304834446631976
"139" 0.025278845473624637
"14" 0.09666329734829485
"140" 0.1110704470853594
"141" 0.11186974122790275
"142" 0.1131558975317282
"143" 0.11473162975887714
"144" 0.1105272721674073
"145" 0.09866671202718473
"146" 0.09229805528093973
"147" 0.04671430427439263
"148" 0.11347366987186253
"149" 0.12259087582337157
"15" 0.12542820971105154
"150" 0.106333758977472878
"151" 0.10794854174655912
"152" 0.08497213797829244
"153" 0.08409909578769388
"154" 0.07400473993394348
"155" 0.09130948944530835
"156" 0.07496251874062969
"157" 0.11154334990834797
"158" 0.1177428105347338
"159" 0.04880437380957894
"16" 0.1296825635625094
"160" 0.09369565599255213
"161" 0.09899106877197958
"162" 0.11962265897953572
"163" 0.11101476605573193
"164" 0.11138449806107008
"165" 0.07196634904086764
"166" 0.11344656333427505
"167" 0.0356042774025466
"168" 0.050410277807923884
"169" 0.11765956298165817
"17" 0.08304806476027186
"170" 0.11737627919224174
"171" 0.04416689447785676
"172" 0.27196083763397995
"173" 0.0662888212150767
"174" 0.09292588232864547
"175" 0.08816570904350345
"176" 0.19982277359326542
"177" 0.09851238280057963
"178" 0.03941897087008663
"179" 0.08647112167836499
```

```
PutTY (inactive)
"179" 0.08647112167836499
"18" 0.07413049841666887
"180" 0.10306335321889983
"181" 0.111383678943060177
"182" 0.04935382990636337
"183" 0.04058310746465461
"184" 0.03374233128834356
"185" 0.07860584480628201
"186" 0.10968385132719934
"187" 0.032113037893384771
"188" 0.07172143661981017
"189" 0.1092535246060386
"19" 0.1102972645758875
"190" 0.10117750613544694
"191" 0.05438239409741300
"192" 0.0720589495330306
"193" 0.03922346473668742
"194" 0.12076731965739748
"195" 0.09492841545380133
"196" 0.09427848086047703
"197" 0.10215364731577099
"198" 0.10078643815066789
"2" 0.11989767545332099
"20" 0.0736013986013986
"200" 0.13187606166264348
"201" 0.132044713553796
"202" 0.09899140045991586
"203" 0.048087416271543615
"204" 0.0684931506849315
"205" 0.07523246599344224
"206" 0.08833170762908359
"207" 0.0815078125388226
"208" 0.08291166272142231
"209" 0.11297948250644907
"21" 0.06386360722984234
"210" 0.07934380776340108
"211" 0.11310320504679515
"212" 0.06409624702307627
"213" 0.11780938978546256
"214" 0.02993836219547989
"215" 0.10353565292295142
"216" 0.11008942973165864
"217" 0.07868790199337758
"218" 0.19235819846724786
"219" 0.1049626881741011
"22" 0.05041866234795492
"220" 0.056836272960188075
"221" 0.10844779603195978
"222" 0.11278732656864776
```

```

PuTTY (inactive)
"222" 0.11278732656864776
"223" 0.09822645175828154
"224" 0.11425442486544808
"225" 0.08099702729092222
"226" 0.08354577861128779
"227" 0.08721017303002598
"228" 0.09978962780363482
"229" 0.11203502738122283
"23" 0.08146106133700898
"230" 0.10662409938743652
"231" 0.11828076056384894
"232" 0.10577578661446632
"233" 0.11514074932415078
"234" 0.12096922968790891
"235" 0.0557415553911836
"236" 0.11249379678942588
"237" 0.1096331644258742
"238" 0.10834063914173245
"239" 0.11534544071014695
"24" 0.10648126258555479
"240" 0.10447949050359085
"241" 0.06993302518630319
"242" 0.07890211996979177
"243" 0.09530565669975881
"244" 0.09546944114505117
"245" 0.0
"246" 0.11985315301676546
"247" 0.07335984971767694
"248" 0.052623959464350334
"249" 0.11923576897943587
"25" 0.11052464311894397
"250" 0.043282871250452847
"251" 0.0019544811624019544
"252" 0.08558204423759197
"253" 0.06711331976762965
"254" 0.0972480119905075
"255" 0.11601004961517414
"256" 0.1093043993422745
"257" 0.10482648999719887
"258" 0.15786461137400468
"259" 0.12318663572582067
"26" 0.05332139249601685
"260" 0.07205229212930792
"261" 0.10801526749738075
"262" 0.11380613560114887
"263" 0.1142608495022124
"264" 0.11220316280901997
"265" 0.10968911731785466
"27" 0.07044812927929389

```

```

PuTTY (inactive)
"27" 0.07044812927929389
"28" 0.10315928607678181
"29" 0.04496955778040933
"3" 0.0352269585440825
"30" 0.431189363995688
"31" 0.09241457441314165
"32" 0.07388668844623049
"33" 0.1204236725267228
"34" 0.11469105842822479
"35" 0.17581828771112398
"36" 0.10412970427360184
"37" 0.103438070866592
"38" 0.08832478224555542
"39" 0.11607313498394789
"4" 0.10621226787883002
"40" 0.11953106379554157
"41" 0.09274584418153507
"42" 0.0709665920546744
"43" 0.10407835112043434
"44" 0.07601977750309023
"45" 0.0983846147461719
"46" 0.12534064929126656
"47" 0.03340712045748017
"48" 0.1058275808855866
"49" 0.09681013151351817
"5" 0.711159737417943
"50" 0.10923919835836489
"51" 0.054687879641958155
"52" 0.13055804576734864
"53" 0.09939808622286242
"54" 0.11968717470048443
"55" 0.10775881095465412
"56" 0.0918765698656626
"57" 0.0982393882217696
"58" 0.023456558453743663
"59" 0.022958827634333566
"6" 0.04668850781897014
"60" 0.028830566964685292
"61" 0.09614549547751258
"62" 0.07257887627997188
"63" 0.07430008293013912
"64" 0.07777953179990528
"65" 0.106488538809956
"66" 0.11474873960246652
"67" 0.08016136157322679
"68" 0.1143179862016906
"69" 0.03888023995673477
"7" 0.08055330999297121
"70" 0.10550719618955761

```



```
Putty (inactive)
"54" 0.11968717470048443
"55" 0.10775801095465412
"56" 0.09187656958656626
"57" 0.0982393882217696
"58" 0.023456558453743663
"59" 0.022958827634333566
"60" 0.04668850781897014
"61" 0.028830566964685292
"62" 0.09614549547751258
"63" 0.07257887627897188
"64" 0.07430008293013912
"65" 0.07777953179990528
"66" 0.1064885388809956
"67" 0.11474873960246652
"68" 0.08016136157122879
"69" 0.114317862016906
"70" 0.03888023995673477
"71" 0.08055330999297121
"72" 0.10558719618955761
"73" 0.057126398156593525
"74" 0.1297482286561036
"75" 0.06660273315751619
"76" 0.07309332094094213
"77" 0.09233854474158283
"78" 0.15699972823625333
"79" 0.03758990781571329
"80" 0.03340662069771519
"81" 0.11726023735794265
"82" 0.1128890978373831
"83" 0.11230199593086634
"84" 0.1314140628199353
"85" 0.0640252588354194
"86" 0.06064757771211555
"87" 0.0006641576267434137
"88" 0.0568739741017691
"89" 0.03836325514349365
"90" 0.12371478220911764
"91" 0.11535120965220538
"92" 0.0890265805206363
"93" 0.07875511622519114
"94" 0.11768850776024835
"95" 0.10308391768201007
"96" 0.056615381413745704
"97" 0.1071444344921823
"98" 0.038084346952212626
"99" 0.07928530356245143
"00" 0.09975598325818912
"01" 0.08868461512594295
"02" 0.06067388225420083
```

- f. How does revenue vary over time? Calculate the average trip revenue per month - analyzing it by hour of the day (day vs night) and the day of the week (weekday vs weekend).

Variation in revenue over time is shown below including code used :

```
hadoop@ip-172-31-60-246:~$ python mrtask_f.py yellow_tripdata_2017-05.csv > mrtask_f.txt
No configs found: falling back to auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/mrtask_f.hadoop.20230906.154011.784359
Running step 1 of 1...
job output is in /tmp/mrtask_f.hadoop.20230906.154011.784359/output
Streaming final output from /tmp/mrtask_f.hadoop.20230906.154011.784359/output...
Removing temp directory /tmp/mrtask_f.hadoop.20230906.154011.784359...
hadoop@ip-172-31-60-246:~$ ls
mrtask_a.py  mrtask_b.py  mrtask_c.py  mrtask_d.py  mrtask_e.py  mrtask_f.py  yellow_tripdata_2017-05.csv
mrtask_a.txt mrtask_b.txt mrtask_c.txt mrtask_d.txt mrtask_e.txt mrtask_f.txt
hadoop@ip-172-31-60-246:~$ 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] 17.1051280026116338
[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
[5, 10, 1] 16.639627234011154
[5, 10, 2] 27.168365753878387
[5, 10, 3] 16.499299606770162
[5, 10, 4] 16.48680345506514
[5, 10, 5] 14.029692293163937
[5, 10, 6] 14.908801744148995
[5, 11, 0] 16.680433054406894
[5, 11, 1] 17.194492364612422
[5, 11, 2] 16.92334503331534
[5, 11, 3] 18.596232180453956
[5, 11, 4] 16.728098769455887
[5, 11, 5] 14.021549507826318
[5, 11, 6] 14.98238805727666
[5, 12, 0] 16.480360929345835
[5, 12, 1] 17.22869705731718
[5, 12, 2] 17.05563095494162
[5, 12, 3] 17.975730053716557
[5, 12, 4] 17.69103689287293
[5, 12, 5] 14.499792055020734
[5, 12, 6] 15.27335388620518
[5, 13, 0] 16.795085865627826
[5, 13, 1] 17.27948235089487
```

```
hadoop@ip-172-31-60-246:~$
[5, 13, 1] 17.27948235089487
[5, 13, 2] 17.774827762048346
[5, 13, 3] 18.595673825475238
[5, 13, 4] 18.205408289793773
[5, 13, 5] 14.99034880898968
[5, 13, 6] 15.937601258660894
[5, 14, 0] 17.20858440343585
[5, 14, 1] 17.34785479975938
[5, 14, 2] 18.110961353384074
[5, 14, 3] 19.033494081619924
[5, 14, 4] 18.613387909146194
[5, 14, 5] 15.336584485461453
[5, 14, 6] 17.128896489169726
[5, 15, 0] 16.91500182066798
[5, 15, 1] 17.307682451791745
[5, 15, 2] 17.75765307206797
[5, 15, 3] 18.7507890420356002
[5, 15, 4] 18.478680845892585
[5, 15, 5] 15.509464218824842
[5, 15, 6] 17.056592424577616
[5, 16, 0] 18.26141013123625
[5, 16, 1] 18.859940648116382
[5, 16, 2] 19.6542708916958
[5, 16, 3] 20.04885999577759
[5, 16, 4] 19.834795454141705
[5, 16, 5] 15.635747754243198
[5, 16, 6] 17.32405836477559
[5, 17, 0] 17.555959492507064
[5, 17, 1] 17.747241962602008
[5, 17, 2] 18.3960718339617
[5, 17, 3] 18.887732509348027
[5, 17, 4] 18.411963137349552
[5, 17, 5] 15.359309303734426
[5, 17, 6] 17.4439645341407
[5, 18, 0] 16.500203407324047
[5, 18, 1] 16.477160631363912
[5, 18, 2] 17.053779035240108
[5, 18, 3] 17.221604263570683
[5, 18, 4] 16.83762356868248
[5, 18, 5] 14.73267798036185
[5, 18, 6] 16.363879207801295
[5, 19, 0] 16.117071206744605
[5, 19, 1] 15.93338279383616
[5, 19, 2] 16.270620125859278
[5, 19, 3] 16.472398718229226
[5, 19, 4] 16.161423076937226
[5, 19, 5] 14.486719040663449
[5, 19, 6] 16.246780024174765
[5, 2, 0] 17.069302560401354
```

```
hadoop@ip-172-31-60-246:~$  
[5, 22, 3] 17.459894119517573  
[5, 22, 4] 16.72930392121595  
[5, 22, 5] 15.161717603507384  
[5, 22, 6] 17.47747548027094  
[5, 23, 0] 18.734166782523527  
[5, 23, 1] 17.364087663335088  
[5, 23, 2] 17.252863134436264  
[5, 23, 3] 18.14697156709204  
[5, 23, 4] 17.248078089078962  
[5, 23, 5] 15.846046419400455  
[5, 23, 6] 18.86490867580518  
[5, 3, 0] 17.010436833469626  
[5, 3, 1] 17.389016241913282  
[5, 3, 2] 16.62262594124502  
[5, 3, 3] 16.37661024166937  
[5, 3, 4] 17.679739001270963  
[5, 3, 5] 16.147548793518364  
[5, 3, 6] 15.822056911276535  
[5, 4, 0] 21.292999725297417  
[5, 4, 1] 20.068907006970854  
[5, 4, 2] 19.732507288628188  
[5, 4, 3] 20.829968819597944  
[5, 4, 4] 20.915669272681466  
[5, 4, 5] 17.825446810289648  
[5, 4, 6] 17.38188976038157  
[5, 5, 0] 21.593018355659638  
[5, 5, 1] 18.81002788103778  
[5, 5, 2] 18.894916666663246  
[5, 5, 3] 19.845702192891927  
[5, 5, 4] 21.227005213121405  
[5, 5, 5] 20.992848200309997  
[5, 5, 6] 21.125313003037945  
[5, 6, 0] 16.36666989228269  
[5, 6, 1] 14.978345859721552  
[5, 6, 2] 14.767823020740611  
[5, 6, 3] 15.288311117881104  
[5, 6, 4] 16.389897123112565  
[5, 6, 5] 19.3220822201787  
[5, 6, 6] 19.992739240104342  
[5, 7, 0] 15.307352827781946  
[5, 7, 1] 14.68863340021126  
[5, 7, 2] 14.46050088621336  
[5, 7, 3] 14.619068724817286  
[5, 7, 4] 14.939835919422332  
[5, 7, 5] 16.300197449100637  
[5, 7, 6] 18.066767322725  
[5, 8, 0] 15.534355540424196  
[5, 8, 1] 15.310034719953286  
[5, 8, 2] 14.991955112543714
```

```
hadoop@ip-172-31-60-246:~$  
[5, 3, 1] 17.389016241913282  
[5, 3, 2] 16.62262594124502  
[5, 3, 3] 16.37661024166937  
[5, 3, 4] 17.679739001270963  
[5, 3, 5] 16.147548793518364  
[5, 3, 6] 15.822056911276535  
[5, 4, 0] 21.292999725297417  
[5, 4, 1] 20.068907006970854  
[5, 4, 2] 19.732507288628188  
[5, 4, 3] 20.829968819597944  
[5, 4, 4] 20.915669272681466  
[5, 4, 5] 17.825446810289648  
[5, 4, 6] 17.38188976038157  
[5, 5, 0] 21.593018355659638  
[5, 5, 1] 18.81002788103778  
[5, 5, 2] 18.894916666663246  
[5, 5, 3] 19.845702192891927  
[5, 5, 4] 21.227005213121405  
[5, 5, 5] 20.992848200309997  
[5, 5, 6] 21.125313003037945  
[5, 6, 0] 16.36666989228269  
[5, 6, 1] 14.978345859721552  
[5, 6, 2] 14.767823020740611  
[5, 6, 3] 15.288311117881104  
[5, 6, 4] 16.389897123112565  
[5, 6, 5] 19.3220822201787  
[5, 6, 6] 19.992739240104342  
[5, 7, 0] 15.307352827781946  
[5, 7, 1] 14.68863340021126  
[5, 7, 2] 14.46050088621336  
[5, 7, 3] 14.619068724817286  
[5, 7, 4] 14.939835919422332  
[5, 7, 5] 16.300197449100637  
[5, 7, 6] 18.066767322725  
[5, 8, 0] 15.534355540424196  
[5, 8, 1] 15.310034719953286  
[5, 8, 2] 14.991955112543714  
[5, 8, 3] 15.095595443793671  
[5, 8, 4] 15.144876788509523  
[5, 8, 5] 14.67262836614724  
[5, 8, 6] 15.685100979728976  
[5, 9, 0] 15.79006959000546  
[5, 9, 1] 15.809253019981295  
[5, 9, 2] 15.622612773623201  
[5, 9, 3] 15.764076749004406  
[5, 9, 4] 15.912008246857104  
[5, 9, 5] 13.814643752241684  
[5, 9, 6] 14.778713618188996  
hadoop@ip-172-31-60-246:~$
```