TASK 4: MapReduce Tasks by Huang-Yin Tso, Arun Santhosh, Vipul Kumar and Gaurav Sharma

File used: yellow_tripdata_2017-01.csv

Task a: Which vendors have the most trips, and what is the total revenue generated by that vendor?

Out put: Vendor 2 has most trips

```
C:\Users\A_R_U_N\mrjob_project>python vendor_analysis.py yellow_tripdata_2017-01.csv

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\A_R_U_N\AppData\Local\Temp\vendor_analysis.A_R_U_N.20250208.105934.748527
Running step 1 of 2...
Running step 2 of 2...
job output is in C:\Users\A_R_U_N\AppData\Local\Temp\vendor_analysis.A_R_U_N.20250208.105934.748527\output
Streaming final output from C:\Users\A_R_U_N\AppData\Local\Temp\vendor_analysis.A_R_U_N.20250208.105934.748527\output
"Vendor with Most Trips: 2" [5312899,82859254.39585336]
Removing temp directory C:\Users\A_R_U_N\AppData\Local\Temp\vendor_analysis.A_R_U_N.20250208.105934.748527...
C:\Users\A_R_U_N\mrjob_project>
```

Task b: Which pickup location generates the most revenue?

Out put: Location 132 generates the most revenue.

```
C:\Users\A_R_U_N\mrjob_project>python mrtask_b.py yellow_tripdata_2017-01.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_b.A_R_U_N.20250209.040640.168390
Running step 1 of 2...
Running step 2 of 2...
job output is in C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_b.A_R_U_N.20250209.040640.168390\output
Streaming final output from C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_b.A_R_U_N.20250209.040640.168390\output...
"132" 12660835.900003651
Removing temp directory C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_b.A_R_U_N.20250209.040640.168390...
C:\Users\A_R_U_N\mrjob_project>
```

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

Out put: Payment through payment type 1(credit card) has done more often than others.

```
C:\Users\A_R_U_N\mrjob_project>python mrtask_c.py yellow_tripdata_2017-01.csv
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_c.A_R_U_N.20250209.042912.559891
Running step 1 of 2...
Running step 2 of 2...
job output is in C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_c.A_R_U_N.20250209.042912.559891\output
Streaming final output from C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_c.A_R_U_N.20250209.042912.559891\output
"1" 6506189
"2" 3144926
"3" 46257
"4" 13447
"5" 1
Removing temp directory C:\Users\A_R_U_N\AppData\Local\Temp\mrtask_c.A_R_U_N.20250209.042912.559891...
C:\Users\A_R_U_N\mrjob_project>
```

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

Out put:

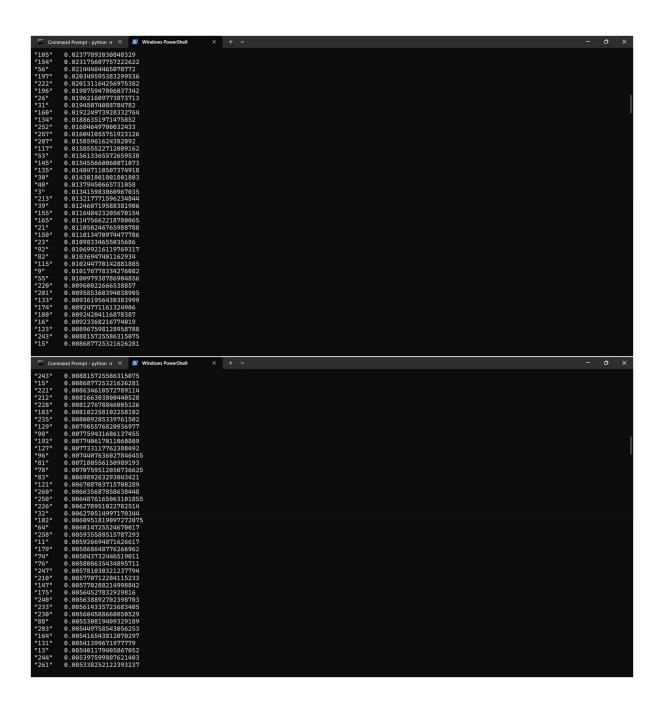
```
Command Prompt × + ~
```

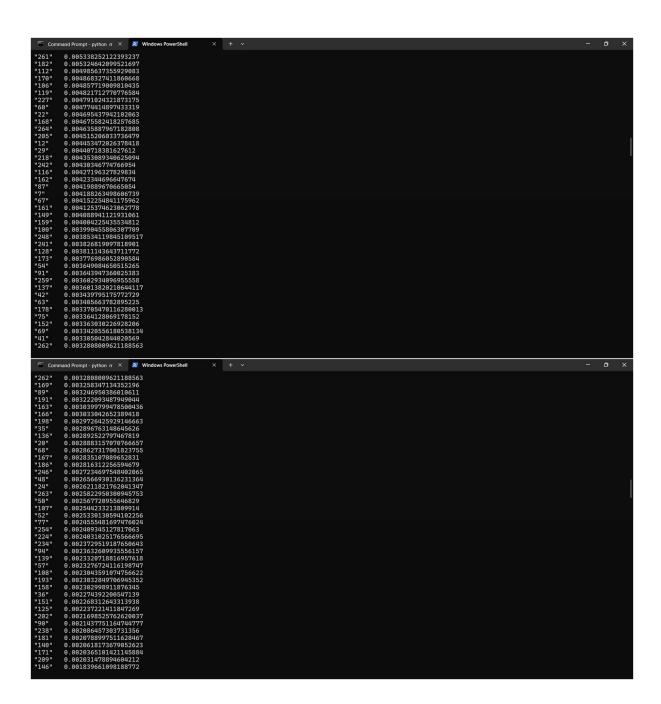


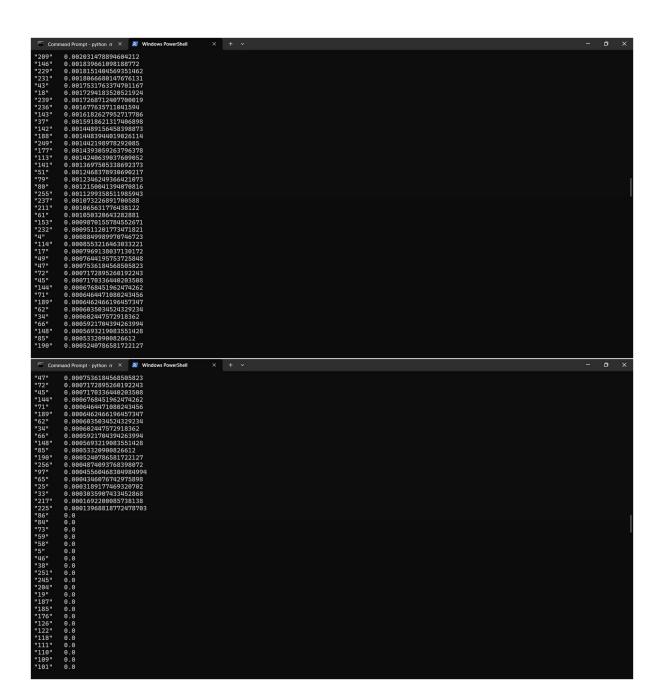


Task e: Calculate the average tips to revenue ratio of the drivers for different pickup locations in sorted format.

Out put:







Task 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).

Out put:

