

MapReduce Assignment

Map Reduce programs

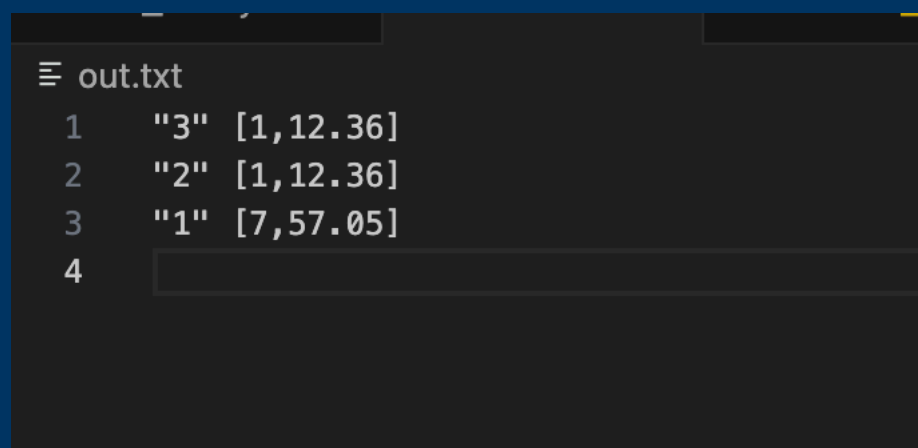
Siddhartha Ghoshal , 5 March 2024

Question 1 : Which vendors have the most trips, and what is the total revenue generated by that vendor?

Execution :

```
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % python mrtask_a.py < sampledata.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_a.Education.20240305.140018.538566
Running step 1 of 1...
reading from STDIN
job output is in /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_a.Education.20240305.140018.538566/output
Streaming final output from /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_a.Education.20240305.140018.538566/output...
Removing temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_a.Education.20240305.140018.538566...
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % █
```

Answer :



```
out.txt
1  "3" [1,12.36]
2  "2" [1,12.36]
3  "1" [7,57.05]
4
```

Conclusion :

:: vendor 1 has most count and max revenue

Question 2 : Which pickup location generates the most revenue?

Execution :

```
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % python mrtask_b.py < sampledata.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_b.Education.20240305.140226.483093
Running step 1 of 1...
reading from STDIN
job output is in /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_b.Education.20240305.140226.483093/output
Streaming final output from /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_b.Education.20240305.140226.483093/output...
Removing temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_b.Education.20240305.140226.483093...
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
```

Answer :

```
≡ out.txt
1  "249"  [7,57.05]
2  "141"  [2,24.72]
3  
```

Conclusion :

:: pickup location 249 generates the most revenue

Question 3 : What are the different payment types used by customers and their count? The final results should be in a sorted format.

Execution :

```
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % python mrtask_c.py < sampledata.csv > out.txt
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /var/folders/1f/1ymv6rl56mvd0yrxtt1sl2dr0000gp/T/mrtask_c.Education.20240305.141115.901044
Running step 1 of 2...
reading from STDIN
Running step 2 of 2...
job output is in /var/folders/1f/1ymv6rl56mvd0yrxtt1sl2dr0000gp/T/mrtask_c.Education.20240305.141115.901044/output
Streaming final output from /var/folders/1f/1ymv6rl56mvd0yrxtt1sl2dr0000gp/T/mrtask_c.Education.20240305.141115.901044/output...
Removing temp directory /var/folders/1f/1ymv6rl56mvd0yrxtt1sl2dr0000gp/T/mrtask_c.Education.20240305.141115.901044...
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
```

Answer :

```
out.txt
1 8.15
3 24.450000000000003
1 8.15
4 41.019999999999996
|
```

Conclusion :

:: payment_type 1 generates the most revenue

Question 4 : What is the average trip time for different pickup locations?

Execution :

```
○ (base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %  
● (base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % python mrtask_d.py < sampledata.csv > out.txt  
No configs found; falling back on auto-configuration  
No configs specified for inline runner  
Creating temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_d.Education.20240305.141429.761360  
Running step 1 of 1...  
reading from STDIN  
job output is in /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_d.Education.20240305.141429.761360/output  
Streaming final output from /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_d.Education.20240305.141429.761360/output...  
Removing temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_d.Education.20240305.141429.761360...  
○ (base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %
```

Answer :

```
out.txt  
"90"      [7,4.416666666666667]  
"142"     [2,9.683333333333334]  
|
```

Conclusion :

:: Average trip for different locations

Question 5 : Calculate the average tips to revenue ratio of the drivers for different pickup locations in sorted format.

Execution :

```
(base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment %  
● (base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % python mrtask_e.py < sampledata.csv > out.txt  
No configs found; falling back on auto-configuration  
No configs specified for inline runner  
Creating temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_e.Education.20240305.141546.905132  
Running step 1 of 1...  
reading from STDIN  
job output is in /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_e.Education.20240305.141546.905132/output  
Streaming final output from /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_e.Education.20240305.141546.905132/output...  
Removing temp directory /var/folders/1f/1ymv6rl56mvd0yrtxt1sl2dr0000gp/T/mrtask_e.Education.20240305.141546.905132...  
○ (base) Education@Siddharthas-MacBook-Pro-2 MapReduce_Assignment % $
```

Answer :

out.txt		
1	"249"	0.02366345311130587
2	"141"	0.08333333333333334
3		

Conclusion :

:: Ratio

Note :

*All Map reduce programs are done with mrjob , but executed locally on sample to csv.
The same needs to be executed in Hadoop cluster by the below method.*

```
python filename.py \  
-r hadoop \  
<input directory path> \  
--output-dir <output-directory-path>
```