CS777 – Week 1 Homework Submission

Andrew Marin – CS 777

**Task 1 – Top-10 Active Taxis**

Many different taxis have had multiple drivers. Write and execute a Spark Python program that computes the top ten taxis that have had the largest number of drivers. Your output should be a set of (medallion, number of drivers) pairs.

* Print a list of top 10 taxis having the largest number of drivers, and the amount of drivers (taxi ID and count)

**Task 1 Output:**

('11DC93DD66D8A9A3DD9223122CF99EFD', 352)

('EE06BD8A621CAC3B608ACFDF0585A76A', 348)

('6C1132EF70BC0A7DB02174592F9A64A1', 341)

('A10A65AFD9F401BF3BDB79C84D3549E7', 340)

('23DB792D3F7EBA03004E470B684F2738', 339)

('7DA8DF1E4414F81EBD3A0140073B2630', 337)

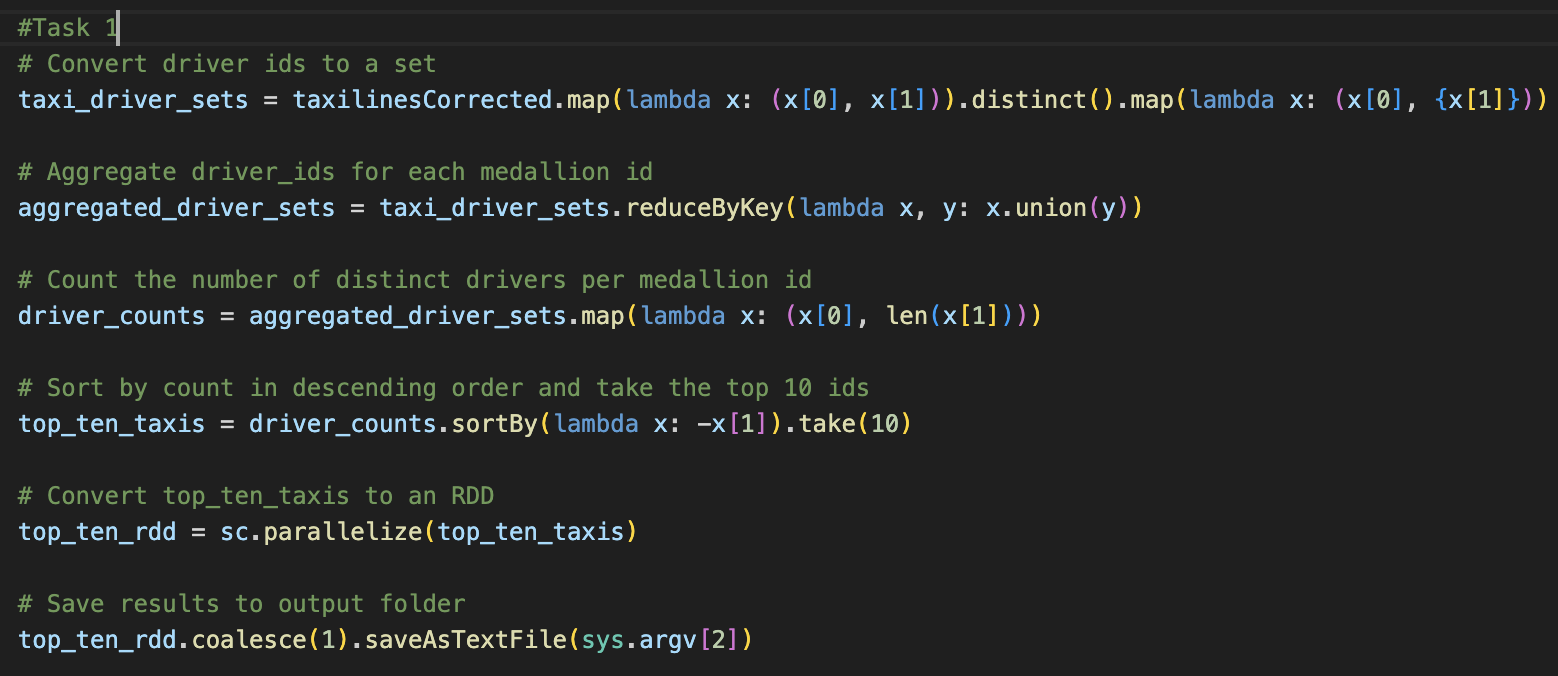
('0318F7BBB8FF48688698F04016E67F49', 335)

('738A62EEE9EC371689751A864C5EF811', 333)

('7D93E7FC4A7E4615A34B8286D92FF57F', 333)

('B07944BF31699A169091D2B16597A4A9', 333)

**Task 1 Code:**



**Task 2 – Top-10 Best Drivers**

We would like to figure out who the top 10 best drivers are in terms of their average earned money per minute spent carrying a customer. The total amount field is the total money earned on a trip. In the end, we are interested in computing a set of (driver, money per minute) pairs.

* Print a list of top 10 best drivers based on earned money per minute carrying a customer (Driver ID and average earning)

**Task 2 Output**

('E4F99C9ABE9861F18BCD38BC63D007A9', 29.984136139169387)

('664927CDE376A32789BA48BF55DFB7E3', 19.641591635123525)

('BA721F7DC14E1D7C93F30BB863E4B83C', 17.5)

('AD4660069108F79079A23D5E05358565', 13.125)

('32BB90E8976AAB5298D5DA10FE66F21D', 10.567005076142133)

('021FF8A9BECC2EFF3B3EC40A10B397E6', 10.5)

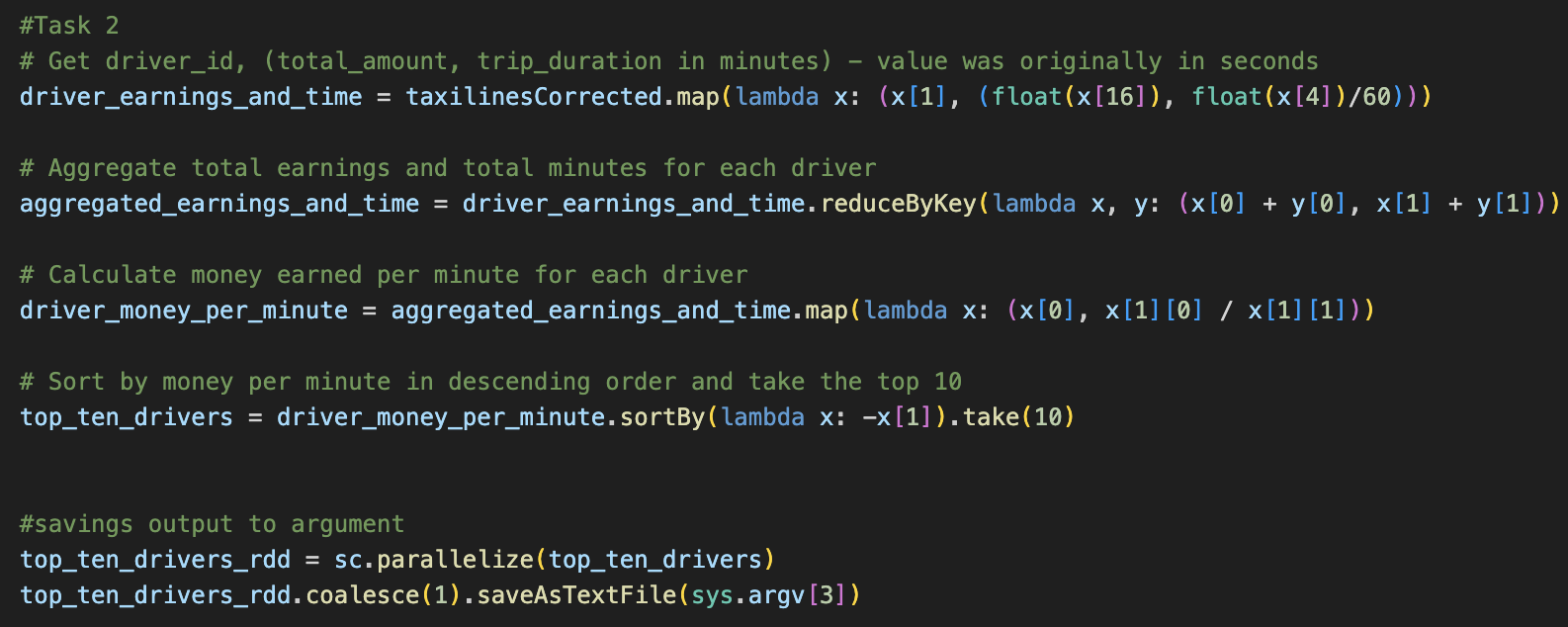
('19CA14E7EA6328A42E0EB13D585E4C22', 8.089887640449438)

('9C9540118D2725A4A63AF71BE096ABB4', 7.5)

('094B6D6E876E9B6DF60B84B3F5567FC9', 7.166666666666667)

('5C223A891DFB16D81E7E877CD17783C6', 6.75)

**Task 2 Code:**

****

**Spark History Output:**

To demonstrate that you did execute your code on the cloud it is important to include URLs in the screenshots. Otherwise, there is no way for us to verify if the code was executed in your cloud account.

**Task 1 & Task 2 (Submitted Together)**

A screenshot of a computer

Description automatically generated

A screenshot of a white and black table

Description automatically generated