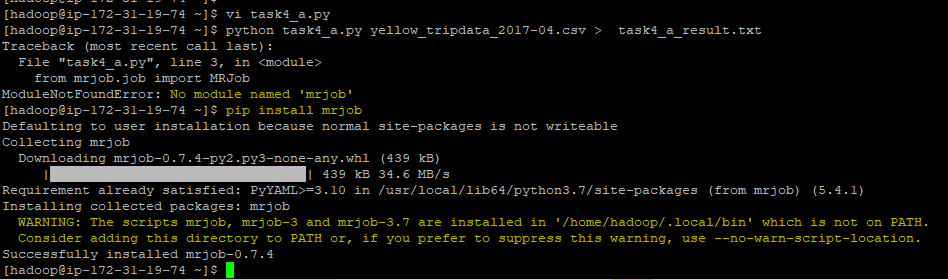
**MapReduce Tasks:**

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

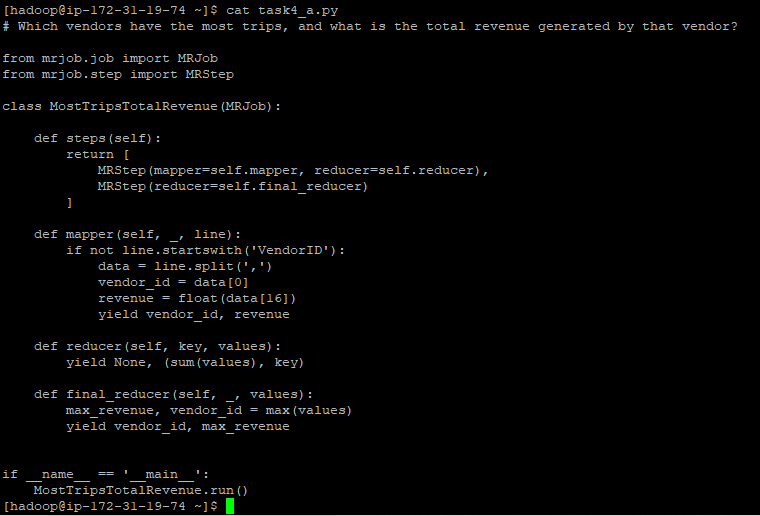


a) Which vendors have the most trips, and what is the total revenue generated by that vendor?

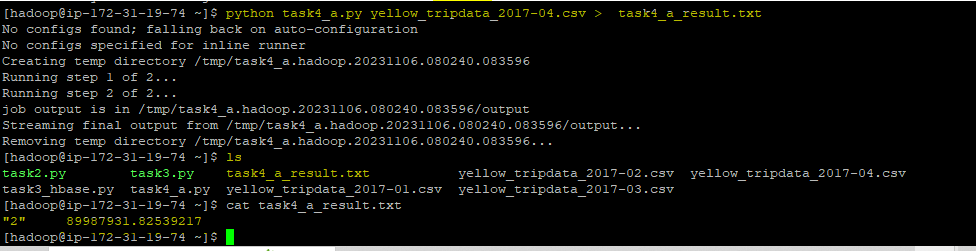
**Answer**:

**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

**Code snippet**:



**Result:**



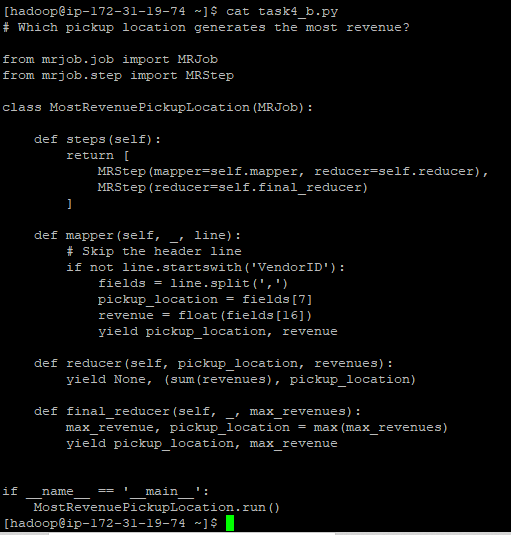
Based on the map reduce job result, in April 2017 Vendor Id = 2 **“VeriFone Inc.”** had the most trips and generated the total revenue **89987931.83 (sum of** Tolls\_amount ‘Total amount of all tolls paid in trip. “ field value).

b) Which pickup location generates the most revenue?

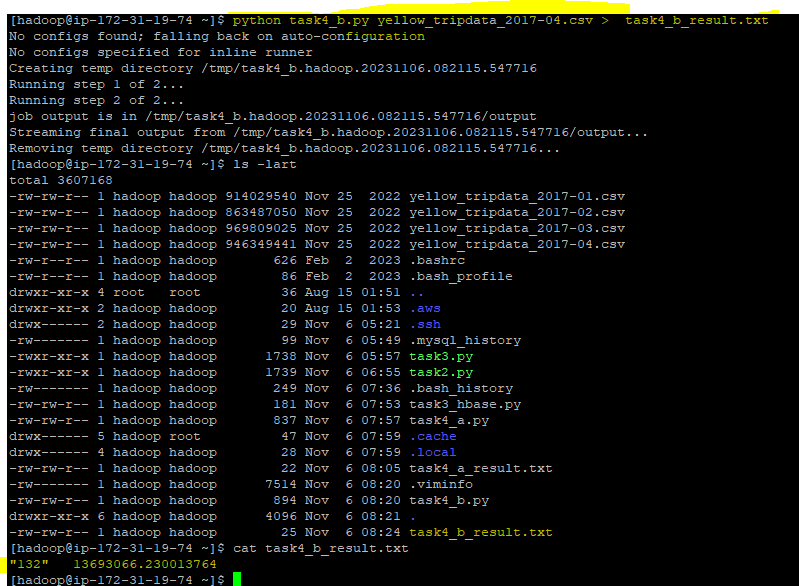
**Answer**:

**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

**Code snippet**:



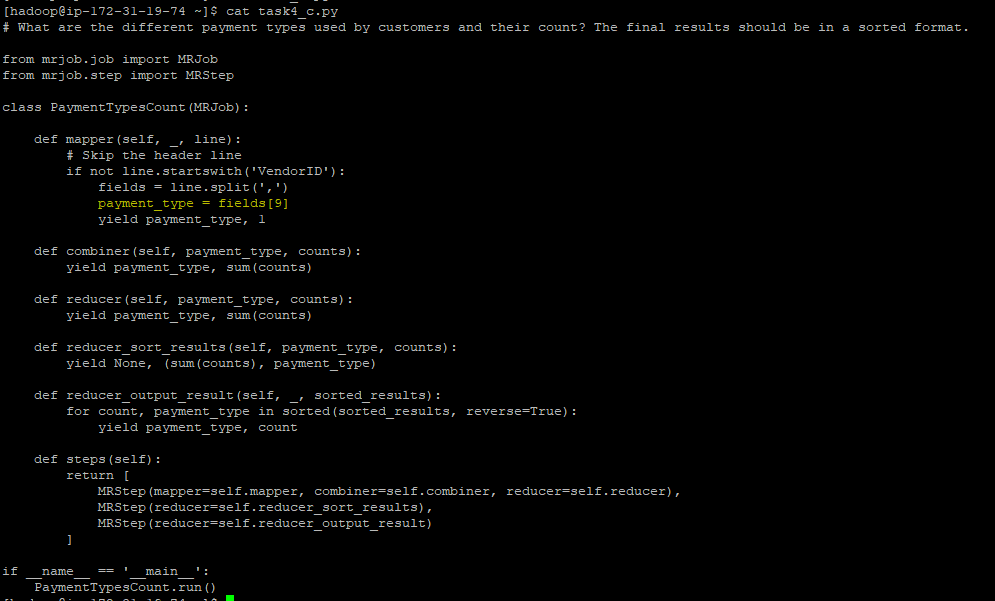
**Answer:**



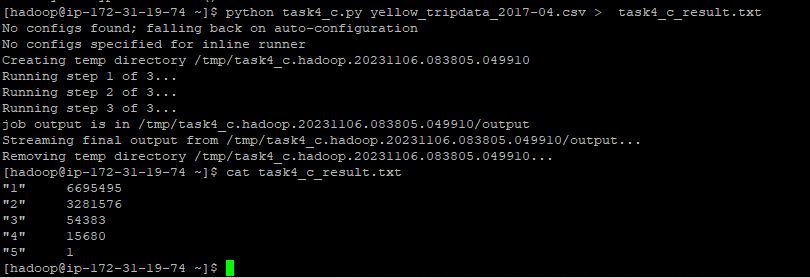
c) What are the different payment types used by customers and their count? The final results should be in a sorted format.

**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

**Code snippet**:



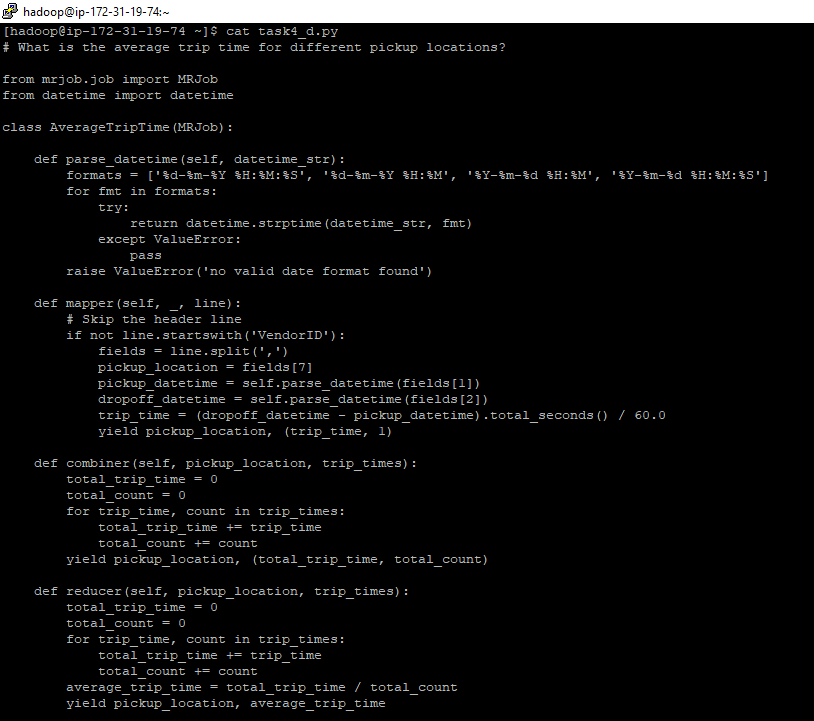
**Answer:**



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

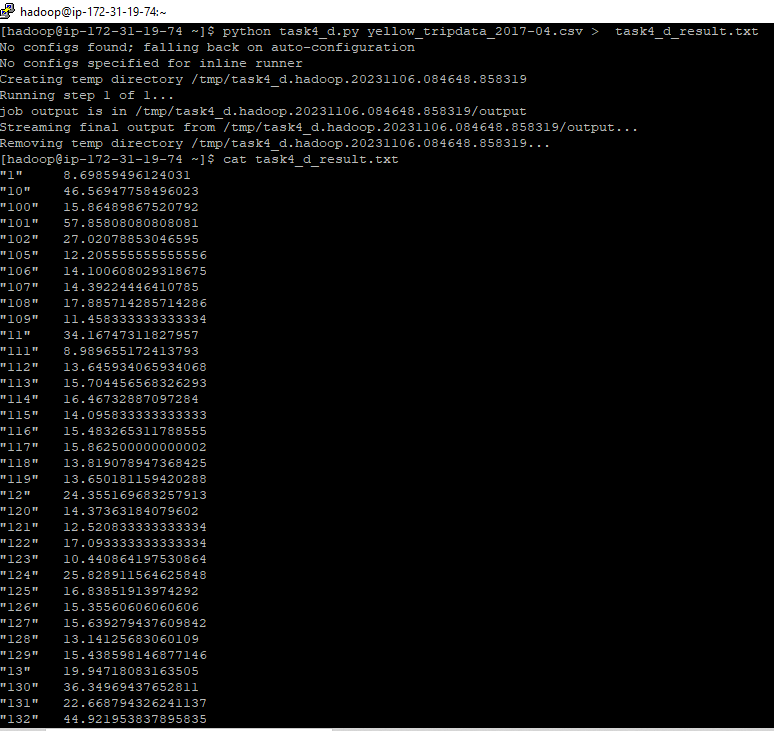
**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

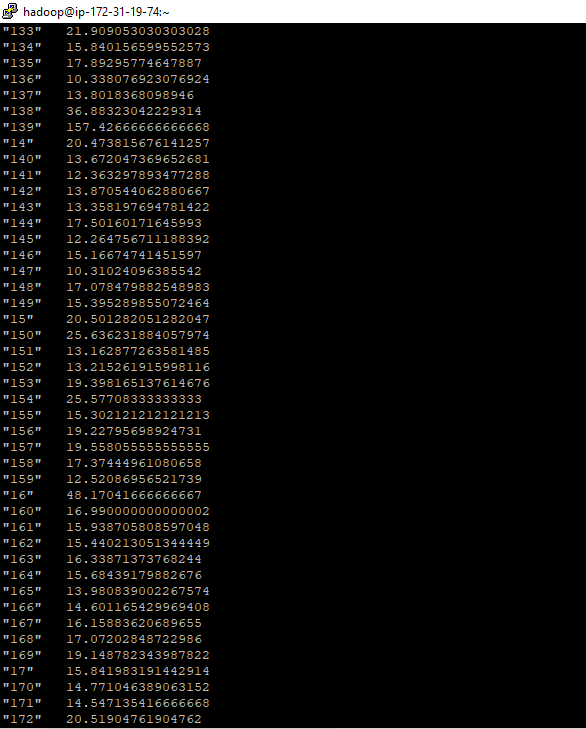
**Code snippet**:

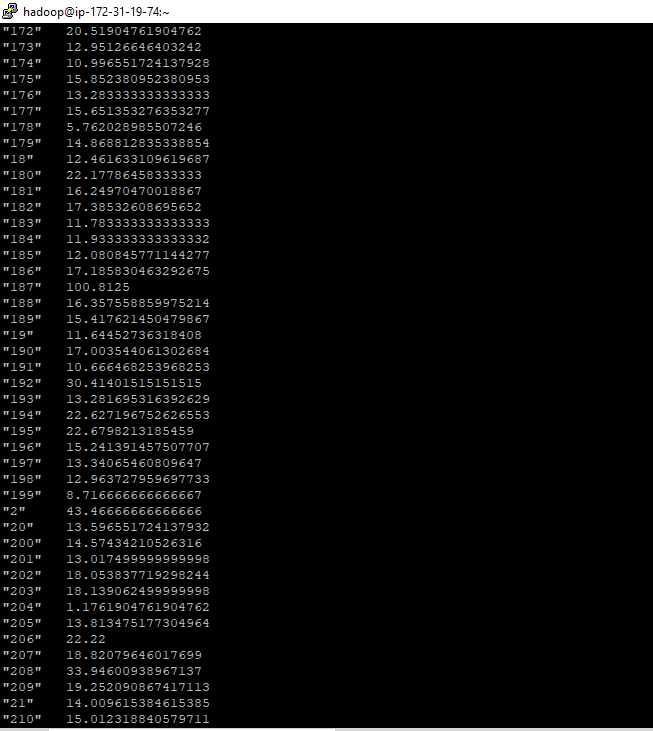


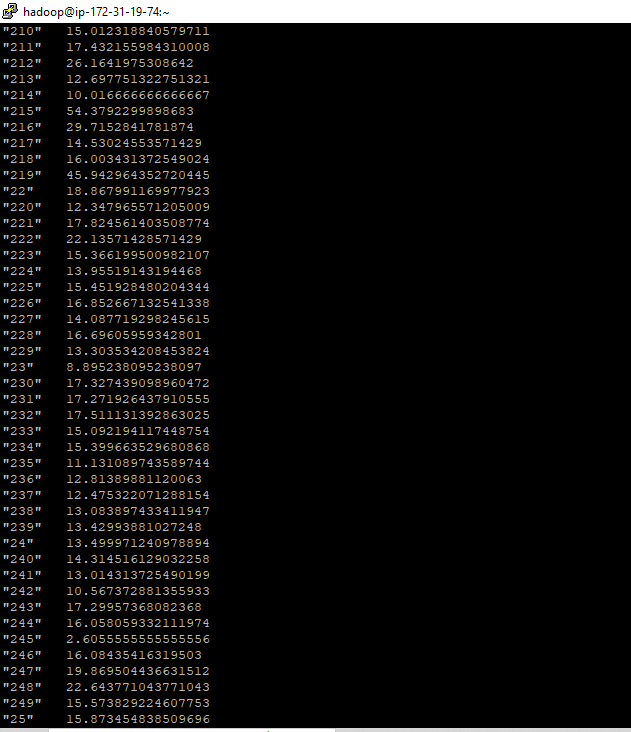


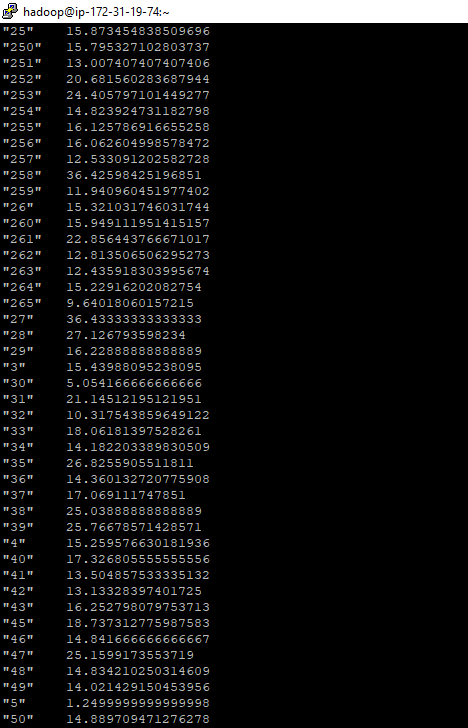
**Answer:**

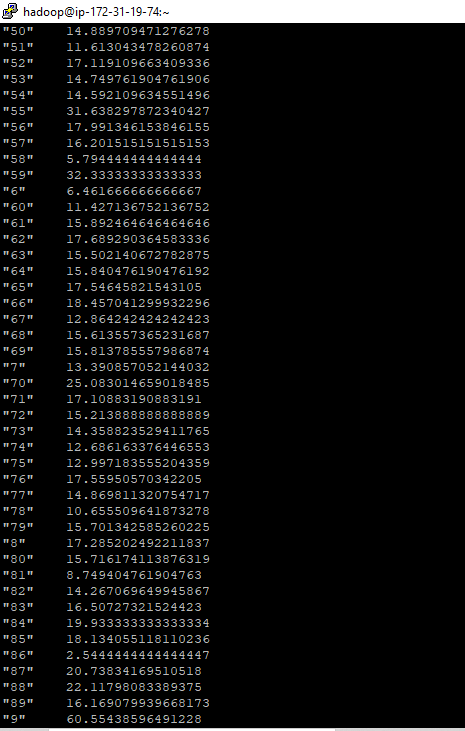


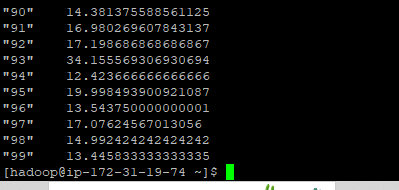








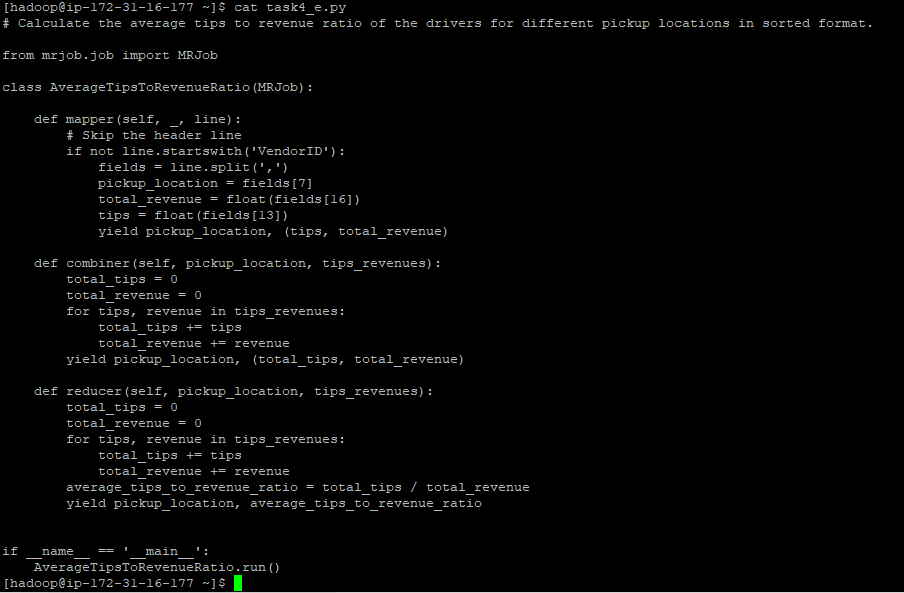




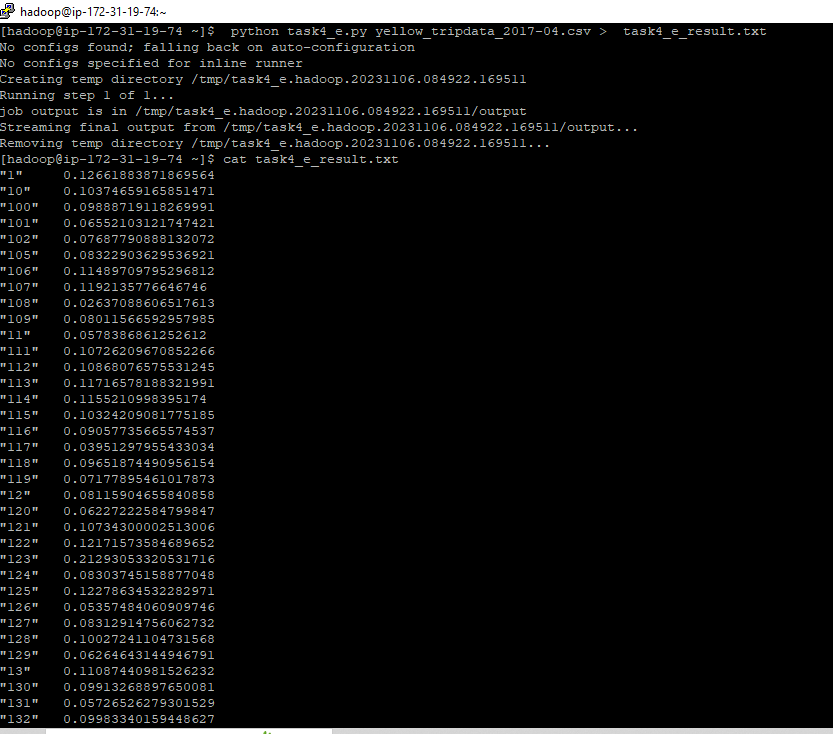
e) Calculate the average tips to revenue ratio of the drivers for different pickup locations in sorted format.

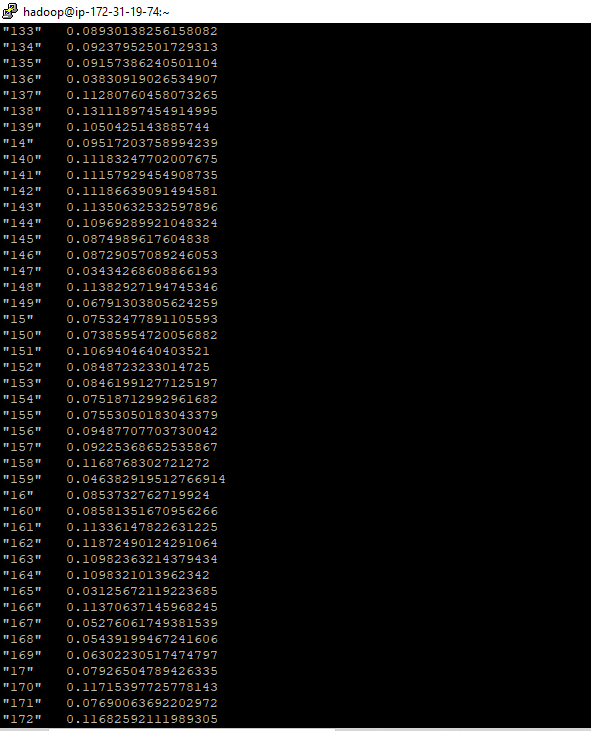
**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

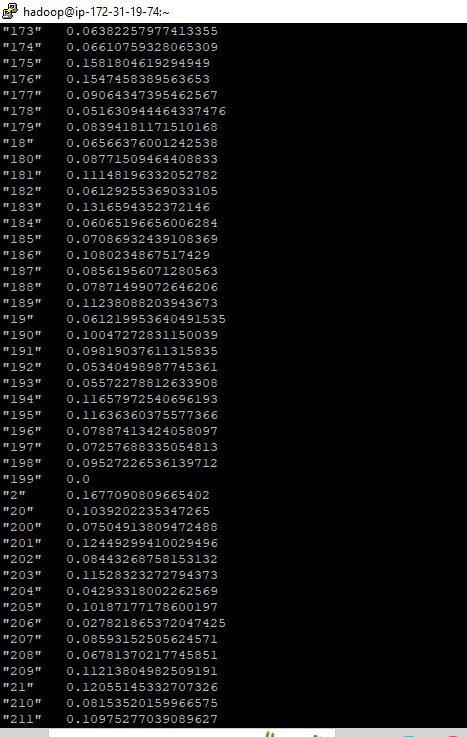
**Code snippet**:

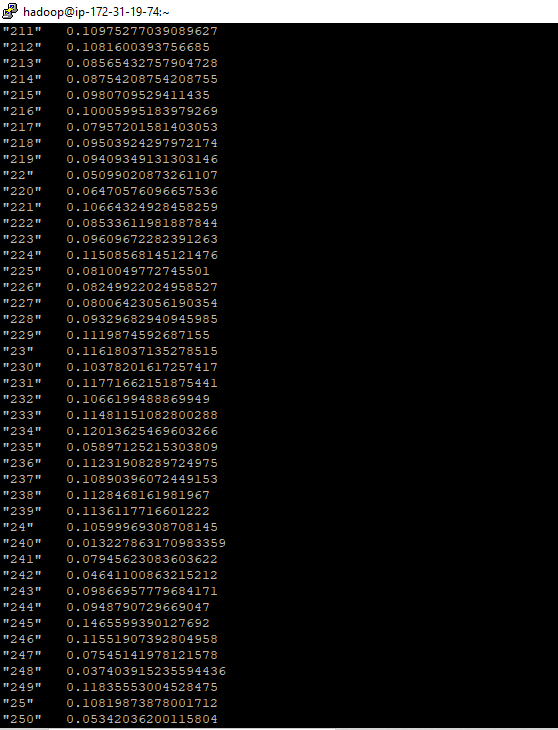


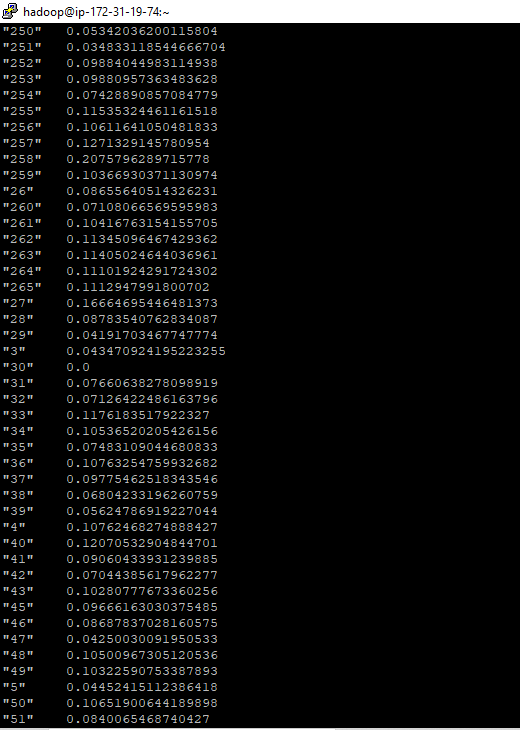
**Result:**

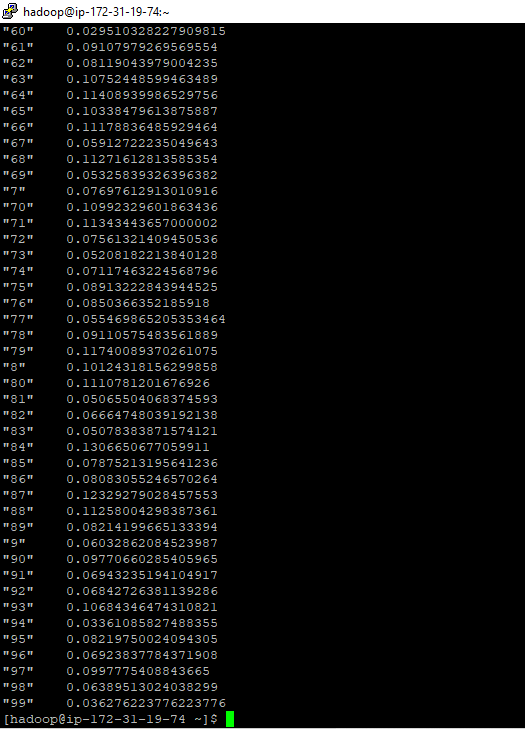








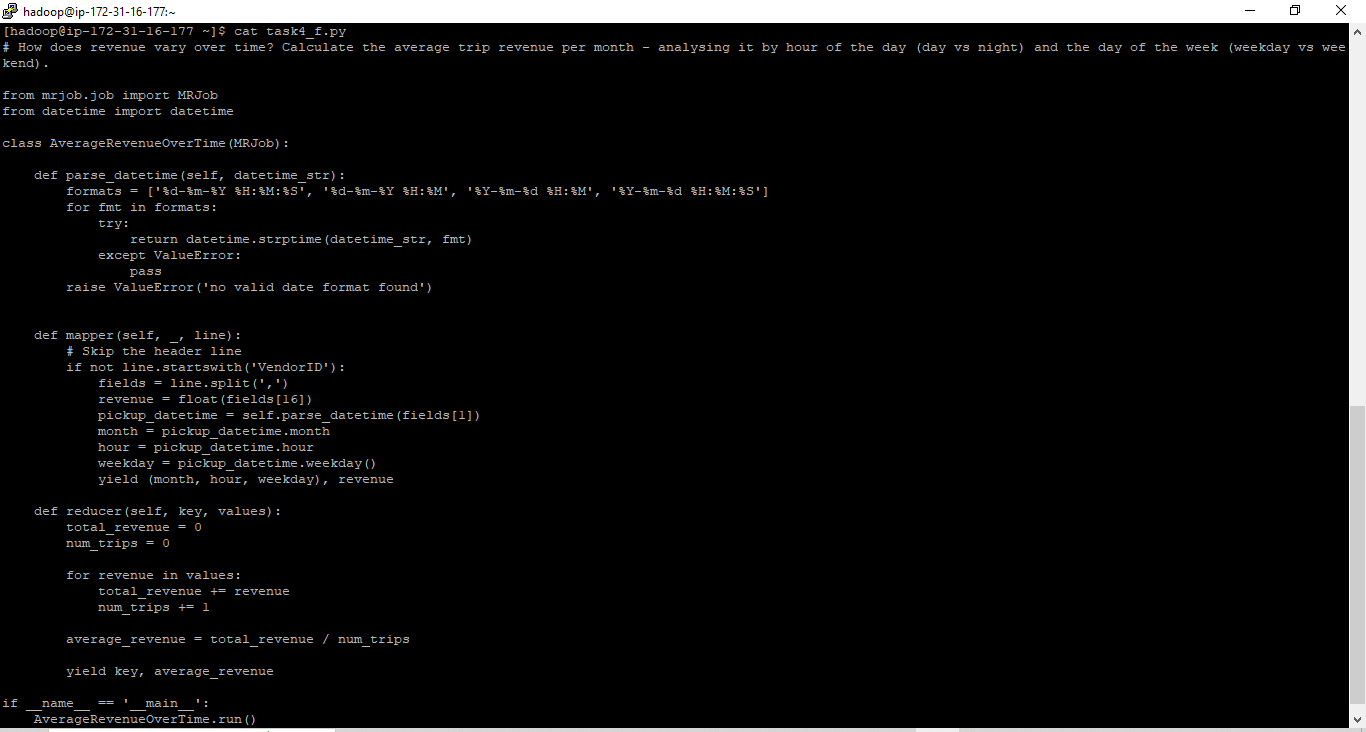




f) How does revenue vary over time? Calculate the average trip revenue per month - analyzing it by hour of the day (day vs night) and the day of the week (weekday vs weekend).

**Data file**: Using the file ‘**yellow\_tripdata\_2017-04.csv’** for this exercise.

**Code snippet**:



**Result:**

