

PySpark RDD Assignment

Given two separate datasets of a sports complex with the following shemas:

trans ID	player ID	player Name	player Age	game type
----------	-----------	-------------	------------	-----------

trans ID	date	player ID	player Name	player Age	game type	cost	state	code
----------	------	-----------	-------------	------------	-----------	------	-------	------

Load data from **trans240.csv** and **cust.csv** and perform the following queries:

1. For each month, show the number of distinct players, and the total cost, the results are sorted by month
 2. For each month, show the name three youngest player, the results are sorted by the month
 3. For each month, show the transaction detail including (transaction ID, date, player ID, player name, age, gametype, cost) of the transaction with highest cost, the results are sorted by month
 4. For each state, show the number of distinct players of each state, the results are sorted by the number of players
 5. For each state, show the name of three oldest player in each state, the results are sorted by the state
 6. For each state, show the average age of players in each state, the results are sorted by the average age
 7. Show the ID and all game types played by players who is more than 40 years old and don't play "Water Sports", the results are sorted by player IDs
 8. For each player ID, show the average number of game per month, the results are sorted by player ID
 9. For each player ID, show the list of three months with highest number of transaction, and the list of the number of transactions in these three months
 10. For each player ID, show the game with highest total cost, and the total cost of this game, the results are sorted by player ID
 11. For each player ID, show the number of places (states) that player has visited, sorted by player ID
 12. For each player ID, show the month with highest total cost, and the total cost in this month
 13. For each player ID, show the list of three games with most transactions, and the list of the number of transactions of these three games, the results are sorted by player IDs
-
14. For each game, show the number of transactions, the results are sorted by game
 15. For each game, show the number of transactions, and the total cost, the results are sorted by game
 16. For each month, show the number of transactions, and the total cost, the results are sorted by month