

Tạo spark dataframe từ hai file trans.txt và cust.txt và thực hiện các truy vấn sau:

1. Show tổng số tiền giao dịch (amount) và danh sách tên (firstname) người chơi không trùng nhau ở mỗi tháng

month	total amount	name list
1	30.57	[Hazel, Patrick]
2	395.22	[Paige, Dolores, Karen, Gretchen, Malcolm]
3	329.21000000000004	[Kristina, Dolores, Patrick]
4	265.90999999999997	[Hazel, Elsie, Gretchen]
5	897.27	[Kristina, Elsie, Hazel, Paige, Dolores, Patrick, Malcolm]

2. Show tên (firstname) của user có tổng số tiền giao dịch lớn nhất ở mỗi tháng. (gợi ý: dùng hàm Window.partitionBy)

month	firstname	sum(amount)
1	Hazel	20.13
2	Gretchen	193.63
3	Patrick	174.36
4	Hazel	185.26
5	Paige	198.44
6	Sherri	349.39
7	Dolores	144.59
8	Elsie	150.6
9	Elsie	157.94
10	Malcolm	195.41999999999996
11	Karen	165.1
12	Paige	293.67