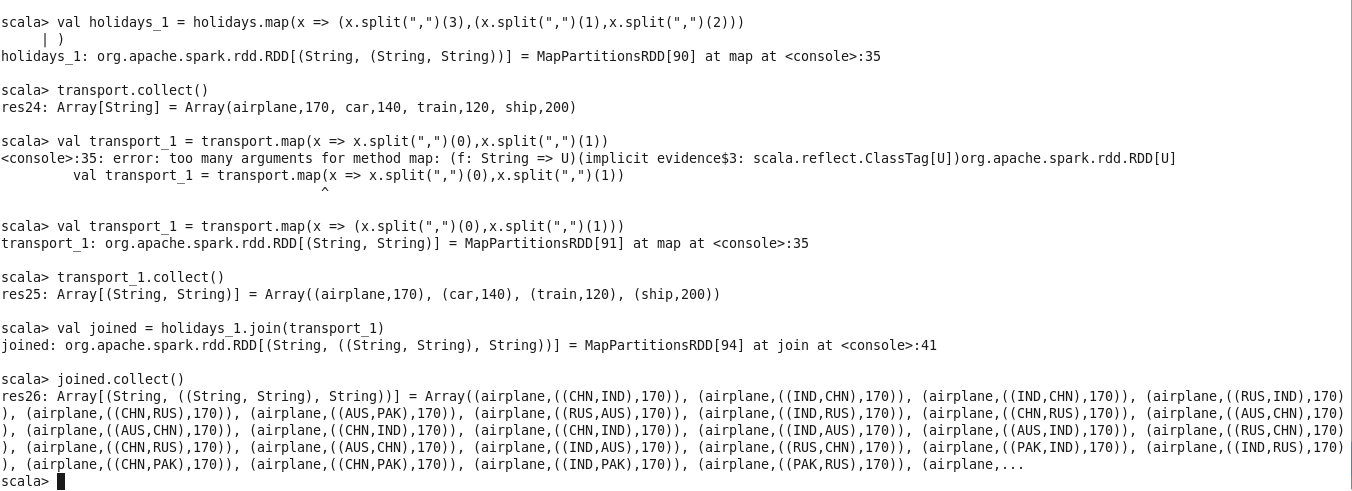
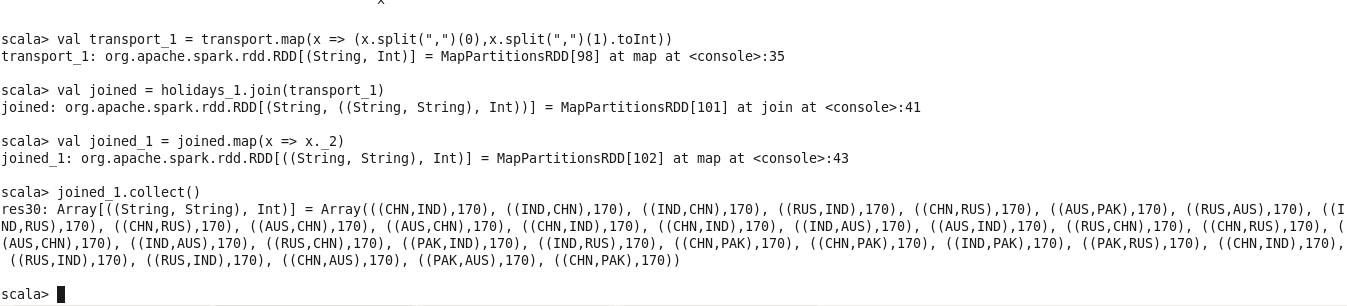
**Assignment18.2**

**Problem Statement**

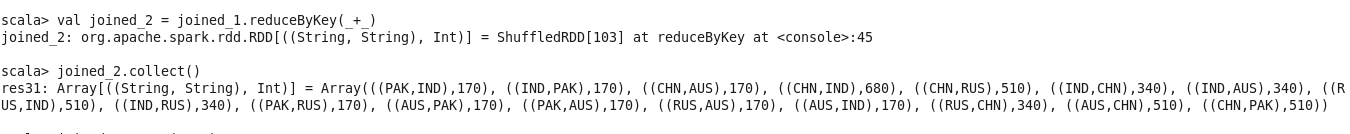
1. **Which route is generating the most revenue per year**

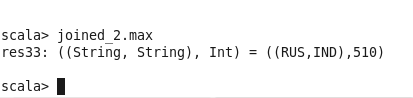
To solve this problem statement we have to join the holidays and transport rdd to get the cost per unit for airplane



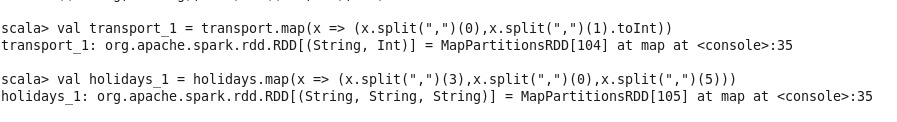


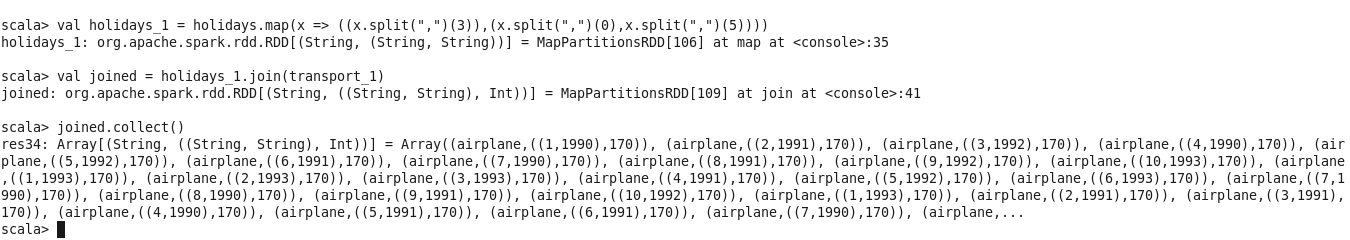
Now we can use reduceByKey and max on this key value pair to get the result.

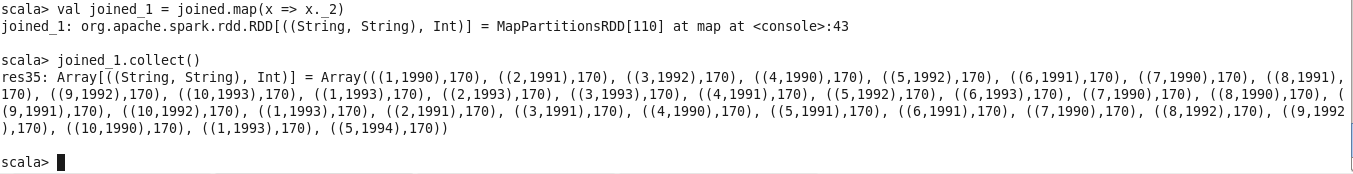




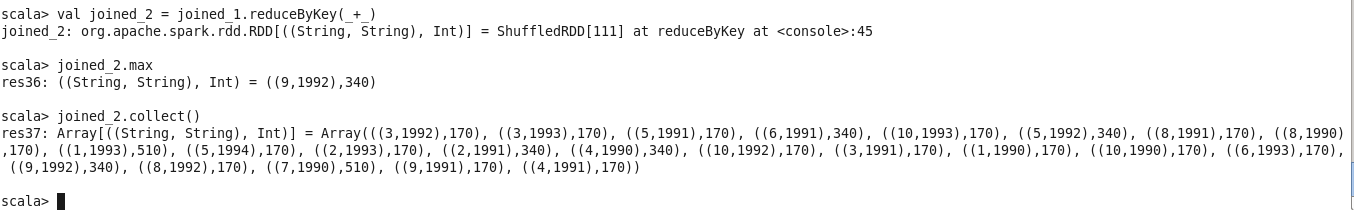
1. **What is the total amount spent by every user on air-travel per year**



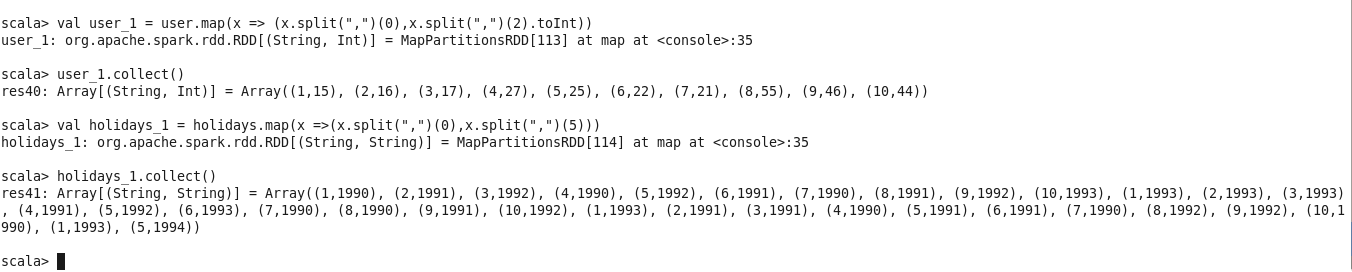


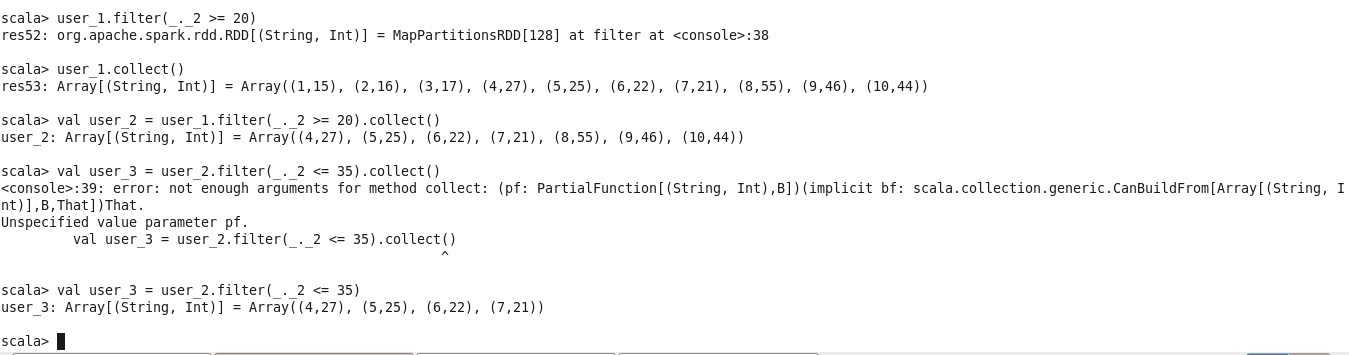


Again by reduceByKey we approach the result

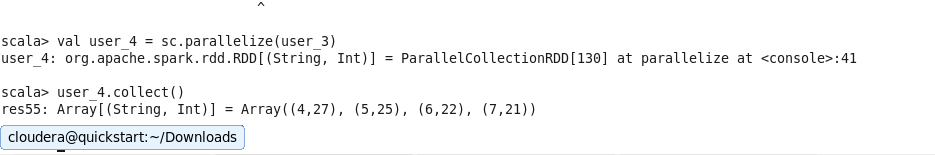


1. **Considering age groups of < 20 , 20-35, 35 > ,Which age group is travelling the most every year.**

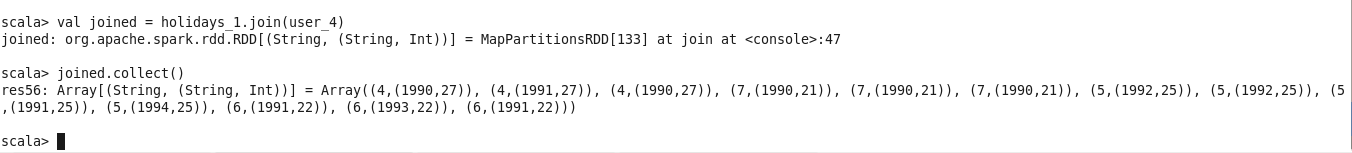


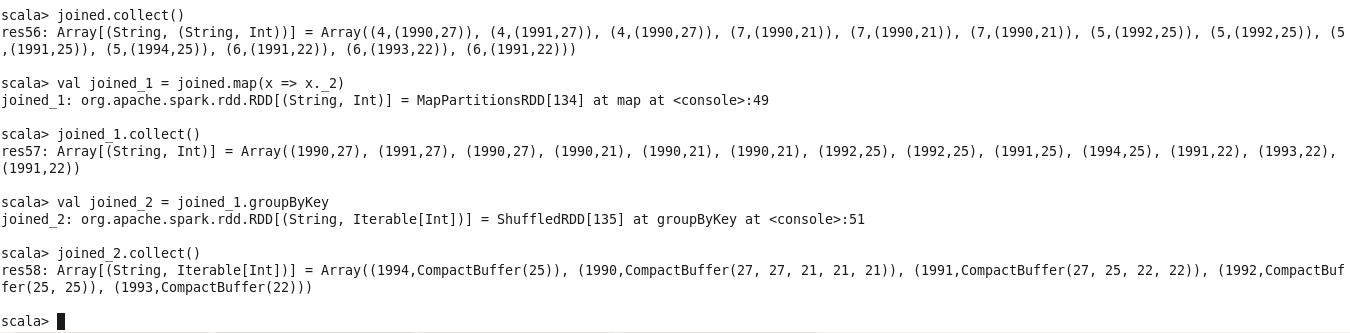


We have to convert array into rdd before joining



Now we have to join these two rdds





Hence for each year we get the age of the particular users.