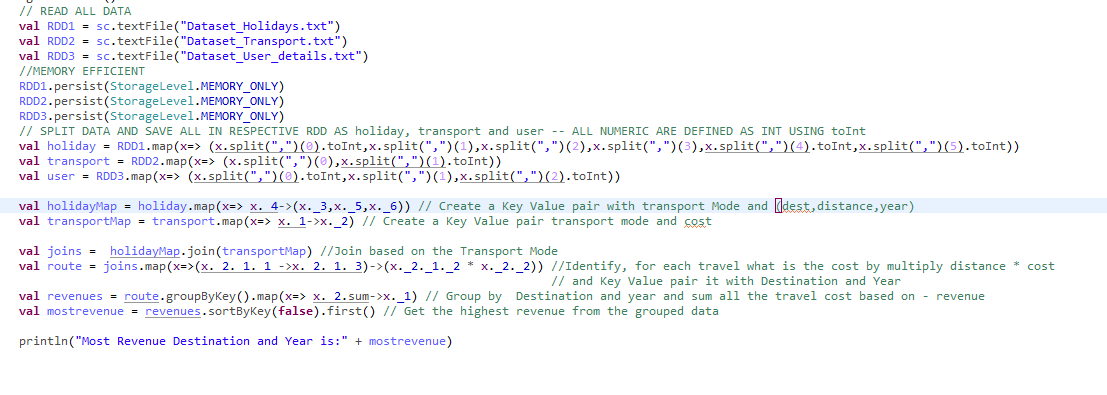
**DATAFRAME\_ASSIGNMENT**

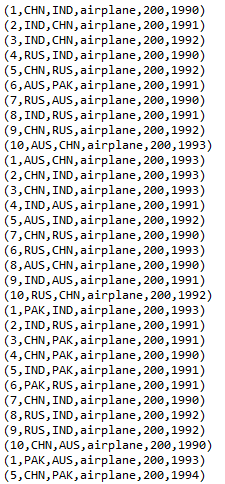
**All Code available in DataFrameAssignment.scala in GitHub**

1. Which route is generating the most revenue per year

CODE



holiday : Raw Holiday Data

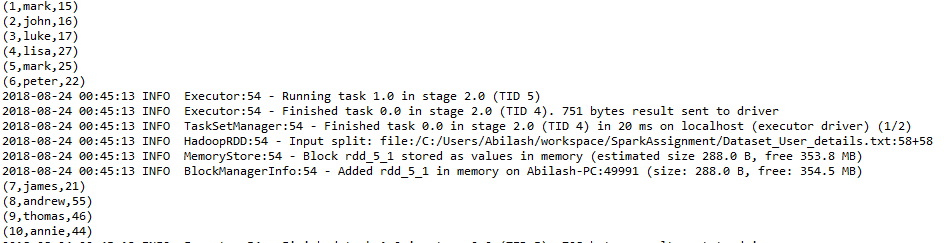


transport: Raw Transport Data

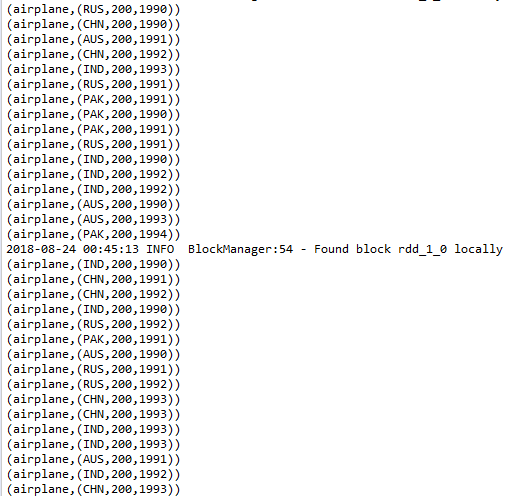




User: Raw User Data



holidaymap: Key Value pair with transport Mode and (dest,distance,year)

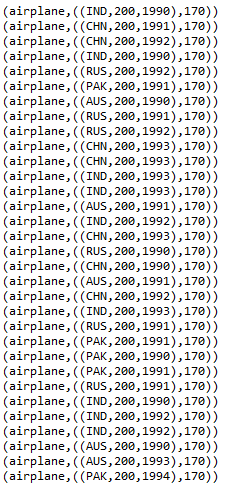


transportmap: Key Value pair transport mode and cost

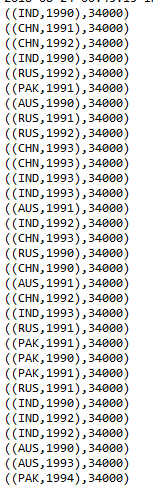




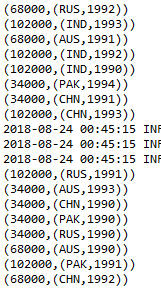
Joins: Join based on the Transport Mode (holidaymap and transportmap)



route: Cost of each route/trip

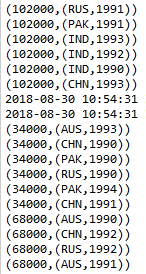


revenue: Group by Destination and year and sum all the travel cost

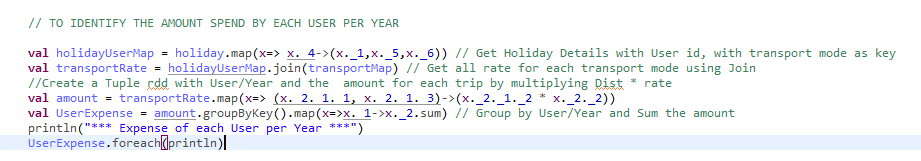


**OUTPUT:**



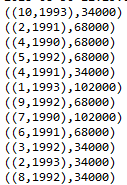


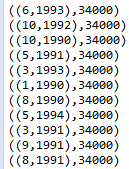
1. What is the total amount spent by every user on air-travel per year?



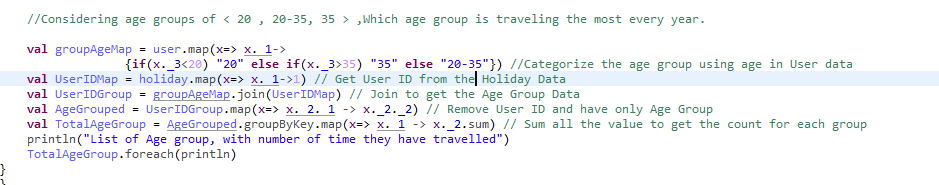
**OUTPUT**







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



**OUTPUT**





