**Assignment 18.2 Spark**

2. Problem Statement

1) Which route is generating the most revenue per year

2) What is the total amount spent by every user on air-travel per year

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

Use the dataset given below:

<https://drive.google.com/drive/folders/0B_P3pWagdIrrVThBaUdVSUtzbms>

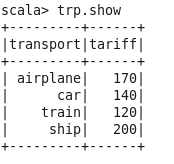
------\ data frame 1

case class TRANSPORT (transport: String , tariff: Int)

val tr = sc.textFile("Transport.txt").map(\_.split(",")).map(p => TRANSPORT(p(0),p(1).trim.toInt)).toDF()

tr.registerTempTable("Transport");

val trp = sqlContext.sql("SELECT transport, tariff FROM Transport")



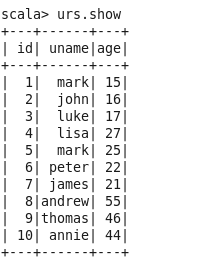
-----\ data frame 2

case class Users (id: Int, uname: String , age: Int)

val ur = sc.textFile("Users.txt").map(\_.split(",")).map(p => Users(p(0).toInt,p(1),p(2).trim.toInt)).toDF()

ur.registerTempTable("Users");

val urs = sqlContext.sql("SELECT id, uname, age FROM Users")



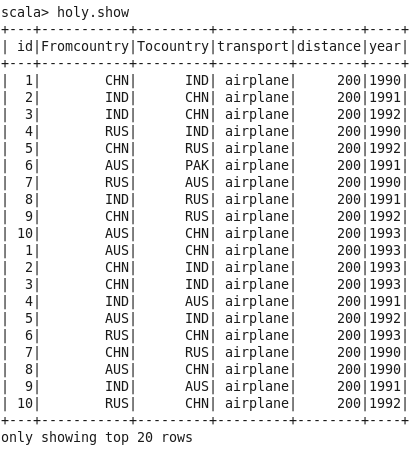
------\ data frame 3

case class Holyday (id: Int, Fromcountry: String ,Tocountry: String, transport: String, distance: Int, year: Int)

val hday = sc.textFile("Holidays.txt").map(\_.split(",")).map(p => Holyday(p(0).toInt, p(1), p(2), p(3),p(4).toInt ,p(5).trim.toInt)).toDF()

hday.registerTempTable("Holyday");

val holy = sqlContext.sql("SELECT id, Fromcountry, Tocountry, transport, distance, year FROM Holyday")

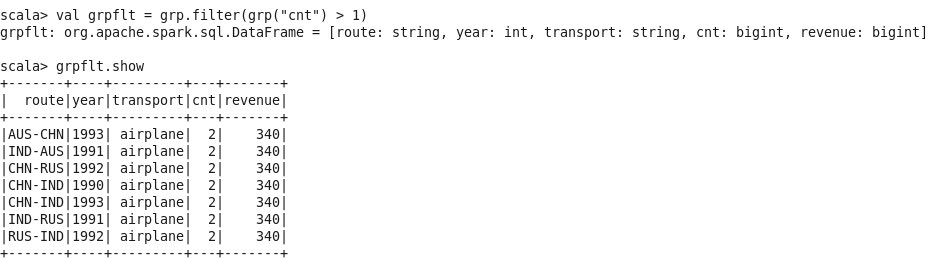


**Which route is generating the most revenue per year**

val grp = sqlContext.sql("SELECT CONCAT(Fromcountry, '-' ,Tocountry) route, year,min(H.transport) transport, count (CONCAT(Fromcountry, '-' ,Tocountry)) cnt, sum(tariff) revenue FROM Holyday H Join Transport T on T.transport=H.transport group by CONCAT(Fromcountry, '-' ,Tocountry) ,year order by count( CONCAT(Fromcountry, '-' ,Tocountry)) desc")

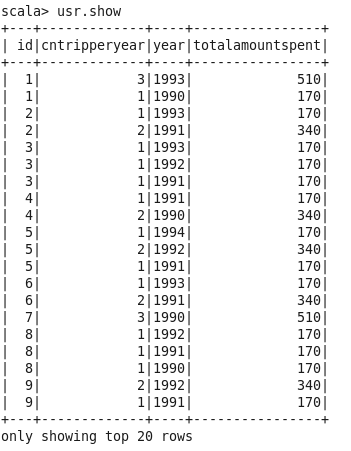
val grpflt = grp.filter(grp("cnt") > 1)

grpflt.show



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

val usr = sqlContext.sql("SELECT id ,Count(id) cntripperyear , year,sum(tariff) as totalamountspent FROM Holyday H join Transport T on T.transport=H.transport where H.transport == 'airplane' group by id,year order by id,year desc")



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

**val usr30 = sqlContext.sql(""" Select count(id) idd , agelimit from (SELECT H.id ,case when age <= 20 then '<20' when age >= 21 and age <= 35 then '21-35' when age >= 36 then '>35' else 0 end as agelimit FROM Holyday H join Users U on U.id=H.id) A group by agelimit """).limit(1)**

