**ASSIGNMENT-18.2**

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

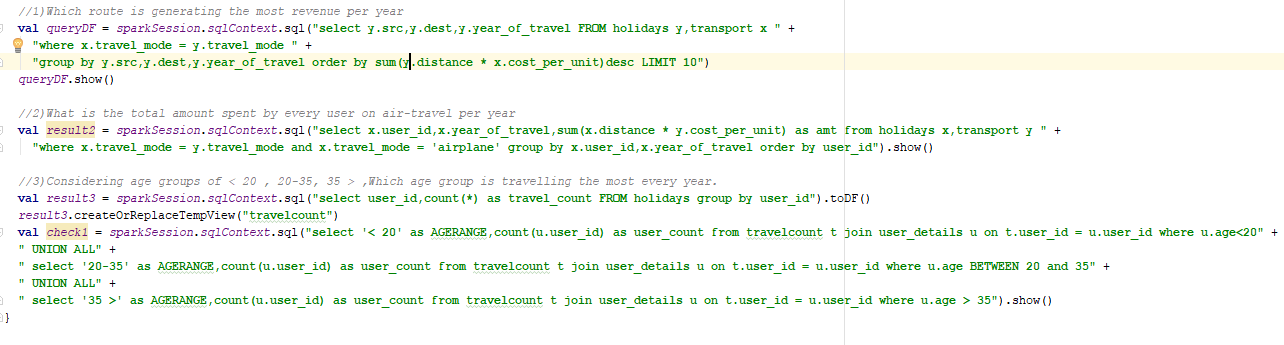
3) Considering age groups of < 20 , 20-35, 35 > ,Which age group is travelling the most

every year.

**PROGRAM –** *Text file data is stored in a view for analysis*



**PROGRAM –***Analysis Queries for the questions provided*

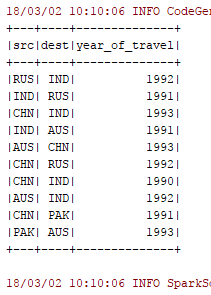


1. Which route is generating the most revenue per year

**QUERY:**

**val** *queryDF* = *sparkSession*.*sqlContext*.sql(**"select y.src,y.dest,y.year\_of\_travel FROM holidays y,transport x "** +  
 **"where x.travel\_mode = y.travel\_mode "** +  
 **"group by y.src,y.dest,y.year\_of\_travel order by sum(y.distance \* x.cost\_per\_unit)desc LIMIT 10"**)  
*queryDF*.show()

**RESULT:**

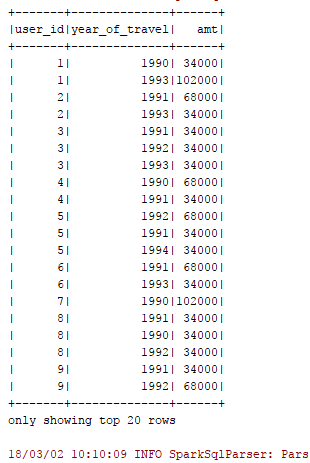
****

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

**QUERY:**

**val** *result2* = *sparkSession*.*sqlContext*.sql(**"select x.user\_id,x.year\_of\_travel,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y "** +  
 **"where x.travel\_mode = y.travel\_mode and x.travel\_mode = 'airplane' group by x.user\_id,x.year\_of\_travel order by user\_id"**).show()

**RESULT:**

****

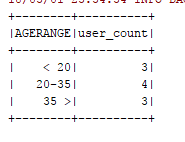
1. Considering age groups of < 20 , 20-35, 35 > ,Which age group is travelling the most

every year.

**QUERY:**

**val** *result3* = *sparkSession*.*sqlContext*.sql(**"select user\_id,count(\*) as travel\_count FROM holidays group by user\_id"**).toDF()  
*result3*.createOrReplaceTempView(**"travelcount"**)  
**val** *check1* = *sparkSession*.*sqlContext*.sql(**"select '< 20' as AGERANGE,count(u.user\_id) as user\_count from travelcount t join user\_details u on t.user\_id = u.user\_id where u.age<20"** +  
**" UNION ALL"** +  
**" select '20-35' as AGERANGE,count(u.user\_id) as user\_count from travelcount t join user\_details u on t.user\_id = u.user\_id where u.age BETWEEN 20 and 35"** +  
**" UNION ALL"** +  
**" select '35 >' as AGERANGE,count(u.user\_id) as user\_count from travelcount t join user\_details u on t.user\_id = u.user\_id where u.age > 35"**).show()

**RESULT:**

****

*Attached the below file for reference:*

**