**ASSIGNMENT-18.3**

1) Considering age groups of < 20 , 20-35, 35 > ,Which age group spends the most

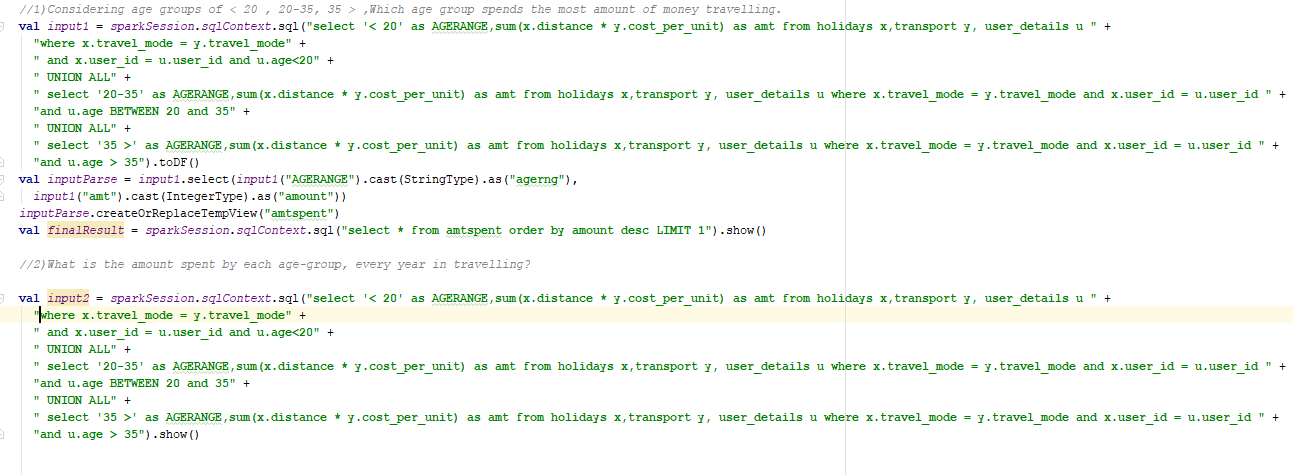
amount of money travelling.

2) What is the amount spent by each age-group, every year in travelling?

**PROGRAM -** *file data is stored in a view for analysis*



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



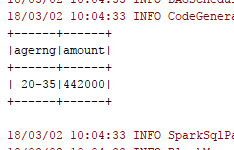
1. Considering age groups of < 20 , 20-35, 35 > ,Which age group spends the most

amount of money travelling.

**QUERY:**

**val** *input1* = *sparkSession*.*sqlContext*.sql(**"select '< 20' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u "** +  
 **"where x.travel\_mode = y.travel\_mode"** +  
 **" and x.user\_id = u.user\_id and u.age<20"** +  
 **" UNION ALL"** +  
 **" select '20-35' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u where x.travel\_mode = y.travel\_mode and x.user\_id = u.user\_id "** +  
 **"and u.age BETWEEN 20 and 35"** +  
 **" UNION ALL"** +  
 **" select '35 >' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u where x.travel\_mode = y.travel\_mode and x.user\_id = u.user\_id "** +  
 **"and u.age > 35"**).toDF()  
**val** *inputParse* = *input1*.select(*input1*(**"AGERANGE"**).cast(StringType).as(**"agerng"**),  
 *input1*(**"amt"**).cast(IntegerType).as(**"amount"**))  
*inputParse*.createOrReplaceTempView(**"amtspent"**)  
**val** *finalResult* = *sparkSession*.*sqlContext*.sql(**"select \* from amtspent order by amount desc LIMIT 1"**).show()

**RESULT:**

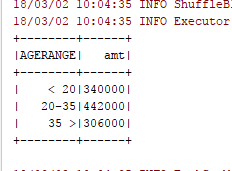
****

1. What is the amount spent by each age-group, every year in travelling?

**QUERY:**

**val** *input2* = *sparkSession*.*sqlContext*.sql(**"select '< 20' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u "** +  
 **"where x.travel\_mode = y.travel\_mode"** +  
 **" and x.user\_id = u.user\_id and u.age<20"** +  
 **" UNION ALL"** +  
 **" select '20-35' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u where x.travel\_mode = y.travel\_mode and x.user\_id = u.user\_id "** +  
 **"and u.age BETWEEN 20 and 35"** +  
 **" UNION ALL"** +  
 **" select '35 >' as AGERANGE,sum(x.distance \* y.cost\_per\_unit) as amt from holidays x,transport y, user\_details u where x.travel\_mode = y.travel\_mode and x.user\_id = u.user\_id "** +  
 **"and u.age > 35"**).show()

**RESULT:**

****

*Attached the file used for reference:*

**