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

amount of money travelling.

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,amount:Int,year:Int)**

**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**case class UserDetails(userid:String,name:String,age:Int)**

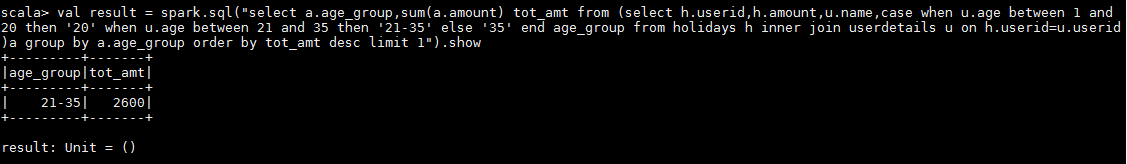
**val userdetailsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_User\_details.txt").map(line => line.split(","))**

**val userdetailssmapRDD = userdetailsRDD.map(u=>UserDetails(u(0),u(1),u(2).toInt))**

**val userdetailsDF = userdetailssmapRDD.toDF**

**userdetailsDF.registerTempTable("userdetails")**

**val result = spark.sql("select a.age\_group,sum(a.amount) tot\_amt from (select h.userid,h.amount,u.name,case when u.age between 1 and 20 then '20' when u.age between 21 and 35 then '21-35' else '35' end age\_group from holidays h inner join userdetails u on h.userid=u.userid)a group by a.age\_group order by tot\_amt desc limit 1").show**



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

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,amount:Int,year:Int)**

**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**case class UserDetails(userid:String,name:String,age:Int)**

**val userdetailsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_User\_details.txt").map(line => line.split(","))**

**val userdetailssmapRDD = userdetailsRDD.map(u=>UserDetails(u(0),u(1),u(2).toInt))**

**val userdetailsDF = userdetailssmapRDD.toDF**

**userdetailsDF.registerTempTable("userdetails")**

**val result = spark.sql("select a.age\_group,sum(a.amount) tot\_amt from (select h.userid,h.amount,u.name,case when u.age between 1 and 20 then '20' when u.age between 21 and 35 then '21-35' else '35' end age\_group from holidays h inner join userdetails u on h.userid=u.userid)a group by a.age\_group").show**

