

Considering age groups of < 20 , $20-35$, $35 >$,Which age group spends the most amount of money travelling.

```

cloudera@quickstart:~$ scala
scala>
scala>
scala>
scala>
scala> autoDF.join(UserDF,autoDF("user_id") === UserDF("user_id")).join(TransDF,autoDF("transport") === TransDF("Transport")).groupBy(UserDF("age")<20,UserDF("age")>20 && UserDF("age")<35,UserDF("age")>35,autoDF("year")).sum("amount").show()
+-----+-----+-----+-----+
| (age < 20)|((age > 20) && (age < 35))|(age > 35)|year|sum(amount)|
+-----+-----+-----+-----+
| false|true|false|1990|1000|
| false|true|false|1991|800|
| false|true|false|1992|400|
| false|true|false|1993|200|
| false|true|false|1994|200|
| true|false|false|1990|200|
| true|false|false|1991|600|
| true|false|false|1992|200|
| true|false|false|1993|1000|
| false|false|true|1990|400|
| false|false|true|1991|400|
| false|false|true|1992|800|
| false|false|true|1993|200|
+-----+-----+-----+-----+
scala>

```

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

```

cloudera@quickstart:~$ scala
scala> autoDF.join(UserDF,autoDF("user_id") === UserDF("user_id")).join(TransDF,autoDF("transport") === TransDF("Transport")).groupBy(UserDF("age"),autoDF("year")).sum("amount").show()
+-----+-----+-----+
|age|year|sum(amount)|
+-----+-----+-----+
| 15|1993|600|
| 21|1990|600|
| 16|1991|400|
| 16|1993|200|
| 27|1990|400|
| 27|1991|200|
| 22|1991|400|
| 22|1993|200|
| 44|1990|200|
| 17|1991|200|
| 44|1992|200|
| 17|1992|200|
| 44|1993|200|
| 17|1993|200|
| 55|1990|200|
| 55|1991|200|
| 55|1992|200|
| 46|1991|200|
| 46|1992|400|
| 25|1991|200|
+-----+-----+-----+
only showing top 20 rows
scala>

```