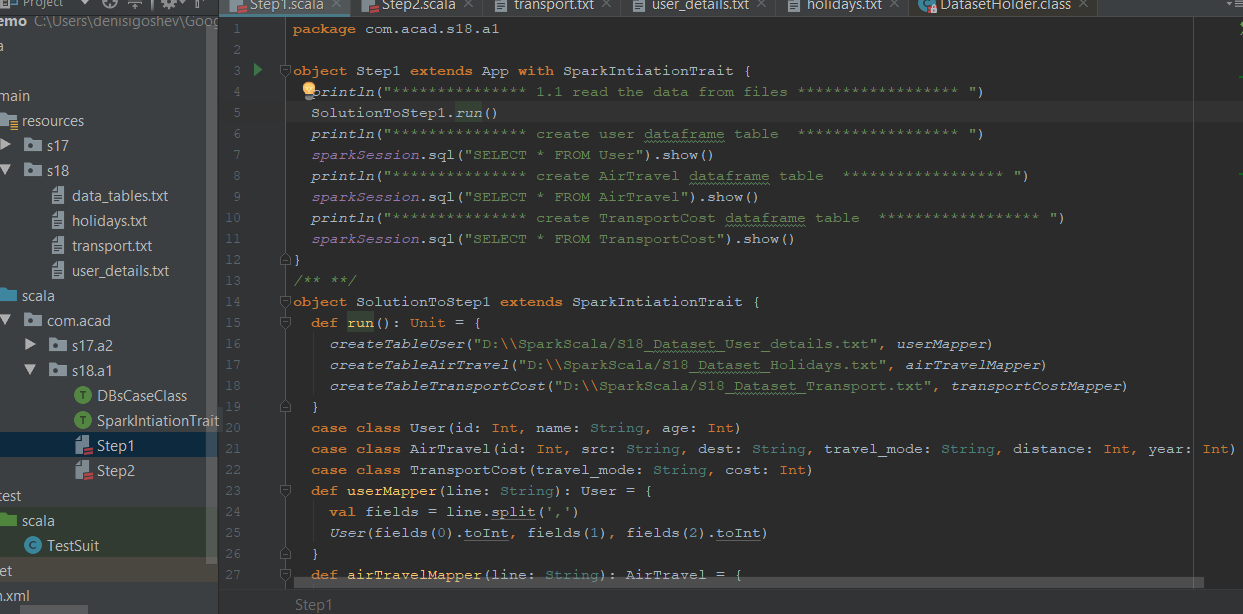
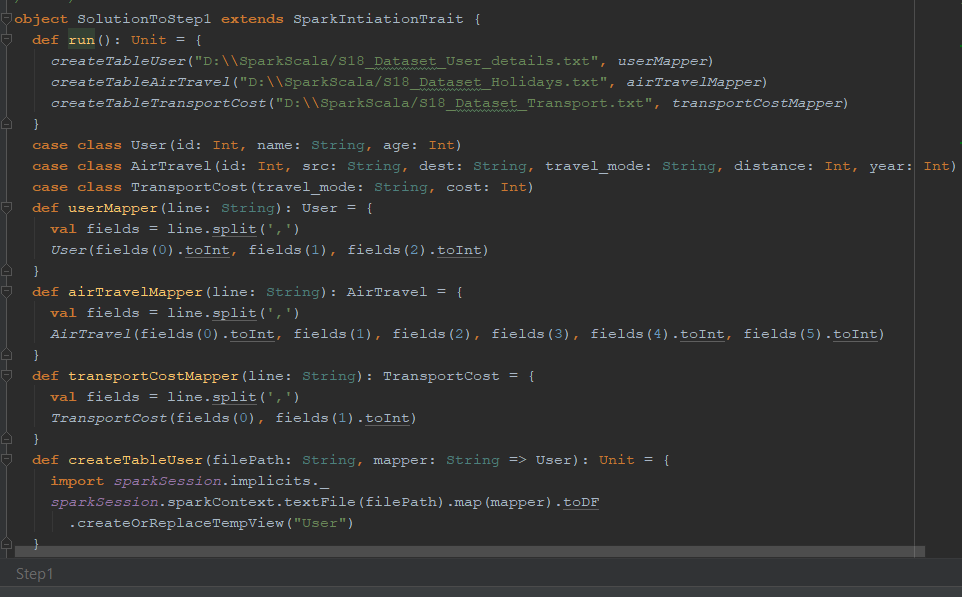


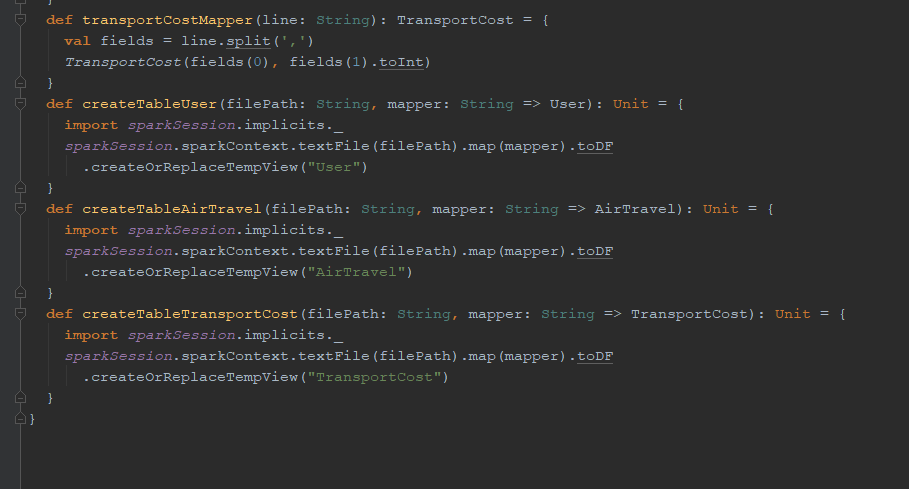
<https://drive.google.com/drive/folders/0B_P3pWagdIrrVThBaUdVSUtzbms>

The preparation step:

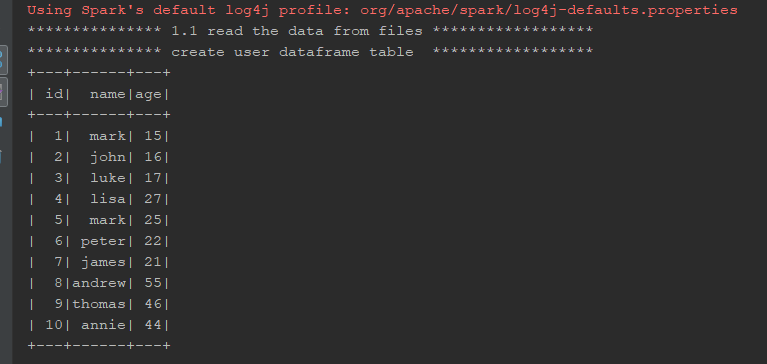
Let’s create the dataframe from the files

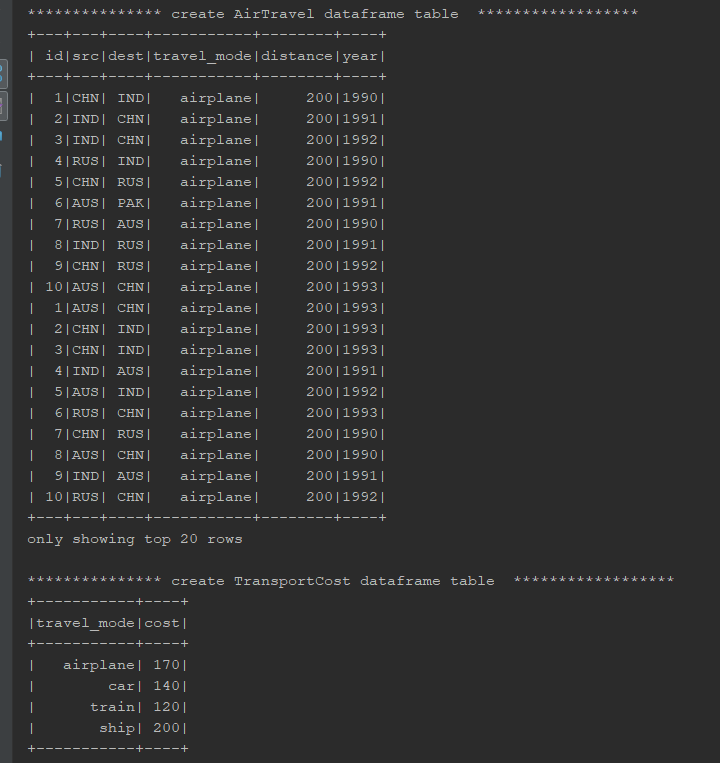




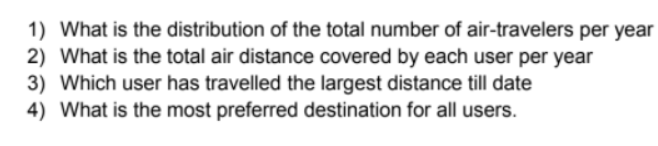


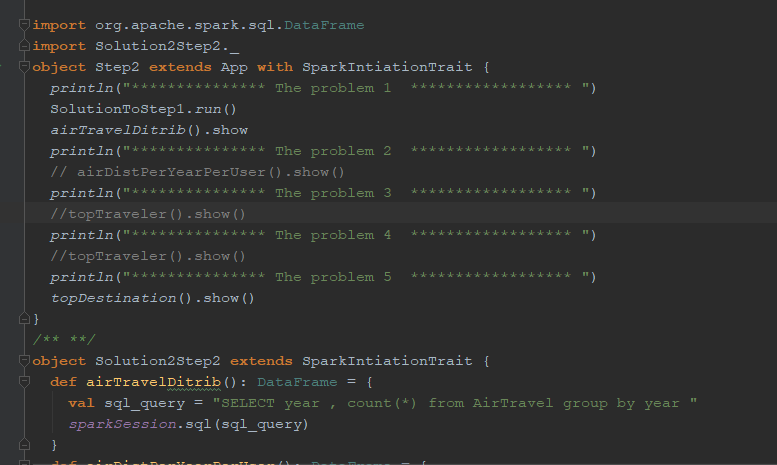
The result :

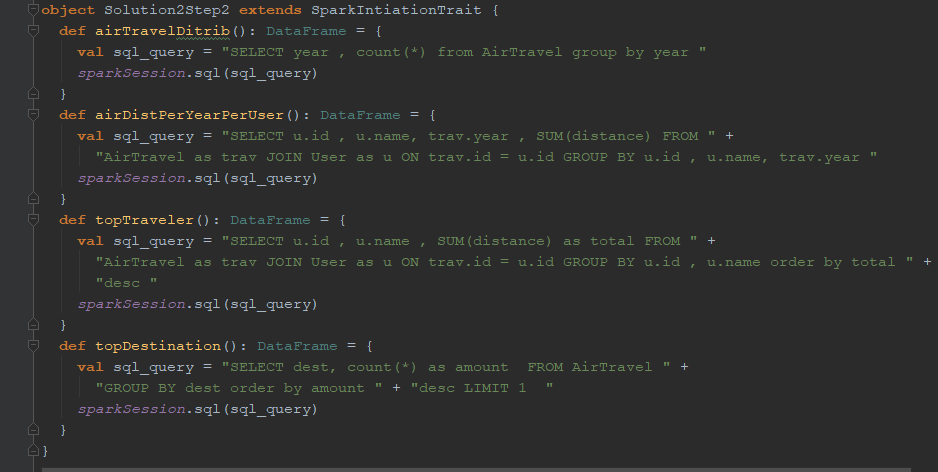




The Problems 1,2,3,4







**The results :**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* The problem 1 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**+----+--------+**

**|year|count(1)|**

**+----+--------+**

**|1990| 8|**

**|1994| 1|**

**|1991| 9|**

**|1992| 7|**

**|1993| 7|**

**+----+--------+**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* The problem 2 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**+---+------+----+-------------+**

**| id| name|year|sum(distance)|**

**+---+------+----+-------------+**

**| 1| mark|1990| 200|**

**| 1| mark|1993| 600|**

**| 6| peter|1991| 400|**

**| 6| peter|1993| 200|**

**| 3| luke|1992| 200|**

**| 3| luke|1993| 200|**

**| 3| luke|1991| 200|**

**| 5| mark|1992| 400|**

**| 5| mark|1991| 200|**

**| 5| mark|1994| 200|**

**| 9|thomas|1992| 400|**

**| 9|thomas|1991| 200|**

**| 4| lisa|1990| 400|**

**| 4| lisa|1991| 200|**

**| 8|andrew|1991| 200|**

**| 8|andrew|1990| 200|**

**| 8|andrew|1992| 200|**

**| 7| james|1990| 600|**

**| 10| annie|1993| 200|**

**| 10| annie|1992| 200|**

**+---+------+----+-------------+**

**only showing top 20 rows**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* The problem 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**+---+------+-----+**

**| id| name|total|**

**+---+------+-----+**

**| 1| mark| 800|**

**| 5| mark| 800|**

**| 3| luke| 600|**

**| 9|thomas| 600|**

**| 4| lisa| 600|**

**| 8|andrew| 600|**

**| 6| peter| 600|**

**| 7| james| 600|**

**| 10| annie| 600|**

**| 2| john| 600|**

**+---+------+-----+**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* The problem 4 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**+----+------+**

**|dest|amount|**

**+----+------+**

**| IND| 9|**

**+----+------+**