**ASSIGNMENT-18.1**

1. **What is the distribution of the total number of air-travelers per year**

**val baseRDD = sc.textFile("/home/acadgild/Assignment-18/S18\_Dataset\_Holidays.txt")**

**import org.apache.spark.storage.StorageLevel**

**baseRDD.persist(StorageLevel.MEMORY\_ONLY)**

**val splitRDD = baseRDD.map(x => (x.split(",")(5).toInt,1))**

**val countSplit = splitRDD.reduceByKey((x,y) => (x + y))**

**countSplit.foreach(println)**

**2)What is the total air distance covered by each user per year?**

**val splitRDD = baseRDD.map(x => ((x.split(",")(0),x.split(",")(5)),x.split(",")(4).toInt))**

**val distRDD = splitRDD.reduceByKey((x,y) => (x + y))**

**distRDD.foreach(println)**

**3)Which user has travelled the largest distance till date?**

**val userRDD = baseRDD.map(x=> (x.split(",")(0),x.split(",")(4).toInt))**

**val totaldistRDD = userRDD.reduceByKey((x,y) => (x+y))**

**val maxRDD = totaldistRDD.takeOrdered(1)**

**4) What is the most preferred destination for all users?**

**val destRDD = baseRDD.map(x => (x.split(",")(2),1))**

**val destreduceRDD = destRDD.reduceByKey((x,y) => (x + y))**

**val maxRDD = destreduceRDD.takeOrdered(1)(Ordering[Int].reverse.on(\_.\_2))**