```
package sparkAssignmentRDD2;

import java.util.List;
import java.util.Arrays;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaPairRDD;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;

import scala.Tuple2;

public class saprkRDD2 {
	public static void main (String args[])
	{
		//SparkConf conf = new SparkConf().setAppName("Spark1").setMaster("local[*]");
		SparkConf conf = new SparkConf().setAppName("Spark1");
		JavaSparkContext sc = new JavaSparkContext(conf);

		//JavaRDD<String> dataset = sc.textFile("data/yellow_tripdata_2017-12.csv");

		//args[0] take the input file
		JavaRDD<String> dataset = sc.textFile(args[0]);
		JavaRDD<String> records= dataset.filter(
				line -> {
					String[] values = line.trim().split(",");
					//to check if there are any blank lines or not
					int size = values.length;
					if(size == 17)
						if (values[5].contentEquals("4"))
```

```
                                                {
                                                        return true;
                                                }
                                        return false;
                                });


                System.out.println(records);
                records.foreach(x->System.out.println(x));
                //args[1] stores the output in the location which the user will define
                records.saveAsTextFile(args[1]);
        }
}
```